

SPRINT 4 – Data Analytics

A continuación, se presentan los resultados del Sprint #4 de la especialización Data Analytics de IT Academy, realizados por Rossemary Castellanos entregado el día 14/10/2025. Se realizaron los 3 niveles del Sprint.

Partiendo de algunos archivos CSV diseñarás y crearás tu base de datos.

Los archivos .csv son los siguiente:

- "american_users.csv"
- "companies.csv"
- "credit_cards.csv"
- "european_users.csv"
- "products.csv"
- "transactions.csv"

NIVEL 1

Descarga los archivos CSV, estúdialos y diseña una base de datos con un esquema de estrella que contenga, al menos 4 tablas de las que puedas realizar las siguientes consultas:

Lo primero a realizar es un análisis del tipo de datos presentes en los archivos de manera de poder definir las tablas a crear y tipo de relación entre ellas, para ello se visualizaron en Excel aprovechando la fácil analogía con el formato CSV.

Resalta el hecho al ver los nombres de los archivos, que hay parecidos con el tipo de base de datos con la que se trabajó en los sprints anteriores: *transactions.csv*, *american_users.csv*, *companies.csv*, *credit_cards.csv*, *european_users.csv*, *products.csv*. Esta observación permite inferir que probablemente se trata de una plataforma de ventas digitales, por lo cual se tomara como nombre de la base de datos **marketplace**.

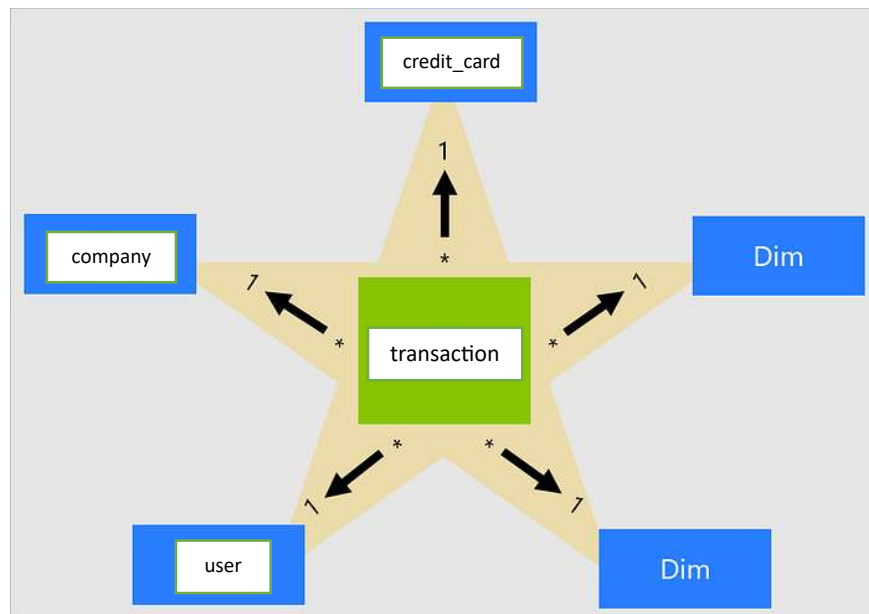
Basados en estas similitudes si analizamos la basa de datos **transactions** de los Sprint 2 y 3, la tabla central era la tabla **transaction**, esto se asocia a una tabla de hechos la cual se rodea de otras tablas (tablas de dimensiones), que proporcionan el contexto para analizar, permitiendo filtrar, agrupar y explorar los datos.

Por tal motivo nos enfocamos primero en visualizar las columnas en particular del archivo *trasactions*, donde se observan campos como *id*, *card_id*, *business_id*, *product_ids* y *user_id*; que se relacionan precisamente con los nombres del resto de archivos .csv.

De esta manera se establece que la tabla de hechos seria la formada por el archivo *trasactions* y en torno a la misma se diseñan y estructuran el resto de las tablas de dimensiones derivadas de forma general por los archivos *credit_cards*, *companies*, *users* tanto americanos como europeos y finalmente *products*.

Lo descrito antes permite diseñar la base de datos en un esquema de estrella:

- Tabla **transaction**, tabla de hecho proveniente de *transactions.csv*.
- Tabla **credit_card**, tabla de dimensión proveniente de *credit_card.csv*.
- Tabla **company**, tabla de dimensión proveniente de *companies.csv*.
- Tabla **user**, tabla de dimensión proveniente de *american_users.csv*, *european_users.csv*.



Para comenzar primero se procede a crear la base de datos **marketplace** con el comando CREATE DATABASE. Luego se crea la estructura de la tabla de hecho **transaction**, basados en los formatos usados en sprint anteriores como punto de inicio. A continuación, se listan los campos de la tabla:

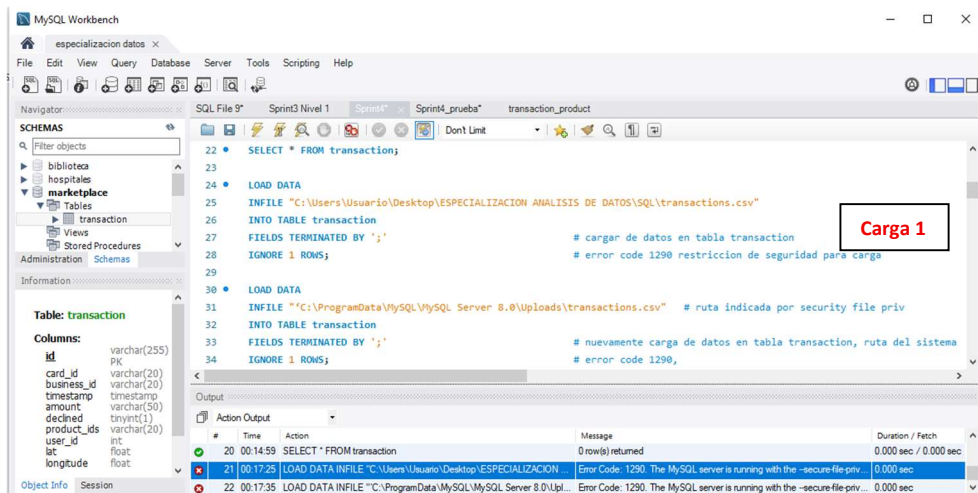
- *id*: identificación de la transacción como clave primaria y única,
- *credit_card_id*: número identificador de la tarjeta de crédito utilizada,
- *business_id*: identificador de la compañía que vende,
- *timestamp*: el momento en que se realiza la operación (fecha y hora),
- *amount*: el importe de la venta,
- *declined*: verificación de proceder o no con la venta,
- *product_ids*: números identificadores de los productos vendidos,
- *user_id*: identificador de usuario,
- *lat, longitude*: ubicación referida como latitud y longitud respectivamente.

```

1  # SPRINT 4
2
3  # Nivel 1.
4  # Descarga los archivos CSV, estudiales y diseña una base de datos con un esquema de estrella que contenga,
5  # al menos 4 tablas de las que puedas realizar las siguientes consultas:
6
7  # Se visualizaron los archivos csv y los campos que contenian, inferiendo que se trata de una plataforma de ventas digitales
8  • CREATE DATABASE marketplace;
9
10 • CREATE TABLE transaction (
11     id VARCHAR(255) PRIMARY KEY,
12     card_id VARCHAR(20),
13     business_id VARCHAR(20),
14     timestamp TIMESTAMP,
15     amount VARCHAR(50),
16     declined BOOLEAN,
17     product_ids VARCHAR(20),
18     user_id INTEGER,
19     lat FLOAT,
20     longitude FLOAT);
  
```

#	Time	Action	Message	Duration / Fetch
14	12:23:49	CREATE DATABASE marketplace	1 row(s) affected	0.000 sec
15	12:30:02	CREATE TABLE transaction (# el archivo transactions genera la tabla de hecho...	0 row(s) affected	0.046 sec

Ahora procedemos a llenar la tabla **transaction** con los datos del archivo *transactions.csv*, para lo cual se usa el comando **LOAD DATA INFILE**, se identifica la ruta donde se ubica el archivo, la tabla donde serán cargados los datos, así como el tipo de delimitador entre registros (en este caso el delimitador es “;”).

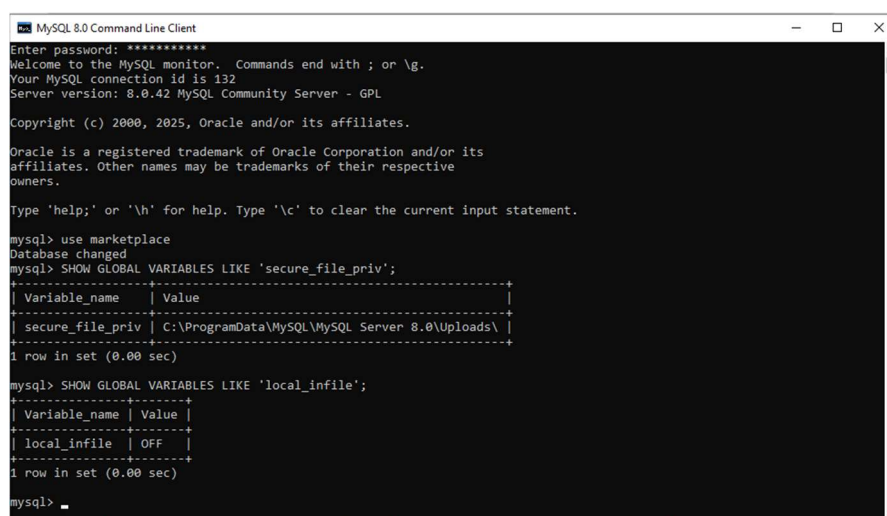


Al ejecutar la sentencia de carga identificada en la imagen como carga 1, se genera el *error code 1290*, asociado a la opción de configuración *--secure-file-priv*, parámetro de seguridad que controla el directorio desde el cual MySQL cargar y descargar archivos, restringiendo la ruta para interactuar con archivos a una definida como medida de protección.

Primero debemos identificar la ruta definida por el sistema para cargar y descargar archivo a través de un terminal mysql donde:

- Se introduce el *password* para MySQL,
- ingresa a la base de datos **marketplace** con el comando **USE**,
- ejecuta el comando **SHOW GLOBAL VARIABLES LIKE 'secure_file_priv'**, para identificar el directorio permitido por el sistema para cargar y descargar archivos.

La terminal muestra la ruta permitida para interactuar con archivos, se puede utilizar este directorio y copiar aquí los archivos o la otra opción es modificar el directorio donde se desea que apunten la ubicación de los archivos.



Para cambiar la ruta se accede como administrador para editar el archivo *my.ini* que contiene los datos de configuración de MySQL, se modifica lo relacionado a la ruta en *--secure-file-priv*.

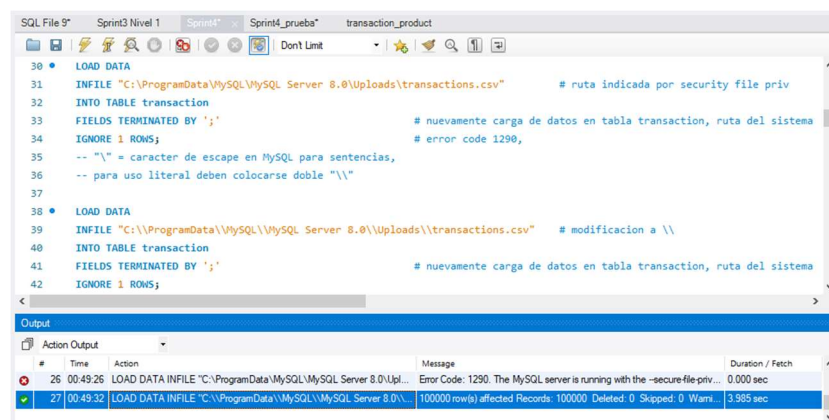
Existe otra variable de seguridad que se debe verificar '*local_infile*', esta controla que MySQL permita o rechace carga de datos locales desde los clientes, se ejecuta el comando *SHOW GLOBAL VARIABLES LIKE 'local_infile'*, para identificar su status el cual cambia según la versión de MySQL que se tenga instalada. En OFF esta deshabilitada y en ON habilitada. Se debe activar en caso de estar apagada modificando el archivo *my.ini*, (insertando *LOCAL INFILE = 1*).

Para este Sprint se decidió mover los archivos al directorio establecido que indica la terminal: '*C:\ProgramData\MySQL\MySQL Server 8.0\Uploads\archivo.csv*'. Esto con la finalidad de intentar no alterar la configuración de seguridad que establece el sistema en la medida de lo posible.

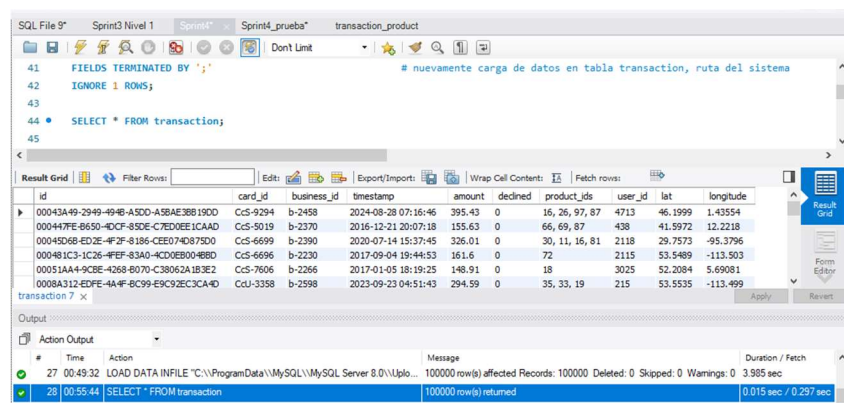
Se procedió a modificar la ruta en la sentencia, generando nuevamente error como se aprecia en la imagen anexa, según la documentación "La barra invertida es el carácter de escape de MySQL dentro de las cadenas en las sentencias SQL... para especificar una barra invertida literal, se deben especificar dos barras invertidas para que el valor se interprete como una sola".

Los valores predeterminados para el comando *LOAD DATA* hace que al leer las entradas interprete los caracteres precedidos por el carácter de escape (**), como secuencias de escape.

Se aplican los cambios a la ruta del directorio duplicando las barras invertidas, "*C:\\ProgramData\\MySQL\\MySQL Server 8.0\\Uploads\\transactions.csv*", realizándose la carga de forma exitosa como lo indica la imagen.



A continuación, se presenta la imagen de la tabla **transaction** con los datos provenientes del archivo *transactions.csv*.



Para las siguientes tablas y archivos se realizó de forma similar a lo visto anteriormente, se diseñaron las estructuras de las tablas a partir de los archivos .csv que serían cargados a posterior.

Se inicio con la primera tabla de dimensiones, la tabla **credit_card**, que contendrá los datos de las tarjetas usadas para pagar las transacciones, se listan a continuación los campos diseñados a partir del archivo *credit_cards.csv*:

- *id*: número identificador del código de la tarjeta que sería la clave primaria y única, debemos revisar que el formato coincida con *credit_card_id* de la tabla **transaction**,
- *user_id*: número identificador del usuario,
- *iban*: alfanumérico, corresponde a la cuenta iban asociado a la tarjeta,
- *pan*: serial identificador único de la tarjeta,
- *pin*: código clave de la tarjeta,
- *cvv*: código de seguridad de la tarjeta,
- *track1*, *track2*: campos desconocidos se asociaron a formato VARCHAR (255),
- *expiring_date*: fecha de caducidad de la tarjeta.

En la imagen siguiente se muestran las sentencias para la creación y carga de datos de la tabla **credit_card** del archivo *credit_cards.csv*. Se identifico la Primary Key como *id*, verificando su similitud con el campo *card_id* de la tabla de hechos **transaction**. Para los formatos como ya se mencionó se basó en los usados en sprint anteriores y para campos desconocidos se utilizó VARCHAR (255). En la carga de datos el delimitador cambio para este archivo csv (",").

```

45
46 CREATE TABLE credit_card (          -- Creamos la tabla credit_card
47     id VARCHAR(15) PRIMARY KEY,
48     user_id VARCHAR(20),
49     iban VARCHAR(50),                  # basados en los formatos utilizados en sprint anteriores
50     pan VARCHAR(50),                  # se definieron estos formatos para el archivo .csv
51     pin VARCHAR(4),                  # para campos nuevos se fijo como estandar VARCHAR(255)
52     cvv VARCHAR(3),
53     track1 VARCHAR(255),
54     track2 VARCHAR(255),
55     expiring_date VARCHAR(15)
56 );
57
58 LOAD DATA INFILE 'C:\\ProgramData\\MySQL\\MySQL Server 8.0\\Uploads\\credit_cards.csv' # cargar de datos en tabla credit_card
59 INTO TABLE credit_card # ruta indicada por security file priv
60 FIELDS TERMINATED BY ',' # cambia el delimitador respecto al archivo anterior
61 IGNORE 1 ROWS;
62
63

```

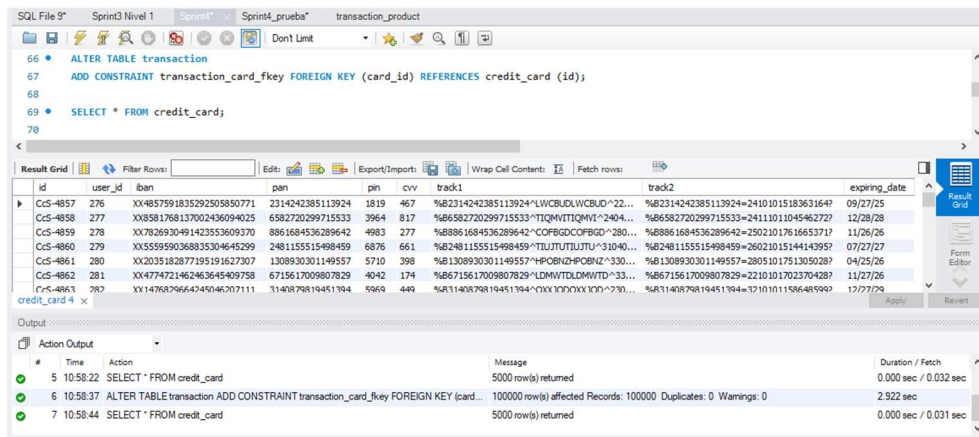
#	Time	Action	Message	Duration / Fetch
27	00:49:32	LOAD DATA INFILE 'C:\\ProgramData\\MySQL\\MySQL Server 8.0\\Uploads\\tr...	100000 row(s) affected Records: 100000 Deleted: 0 Skipped: 0 Warnings: 0	3.985 sec
28	00:55:44	SELECT * FROM transaction	100000 row(s) returned	0.015 sec / 0.297 sec
29	01:06:51	CREATE TABLE credit_card (- Creamos la tabla credit_card id VARCHAR(15) PRI...	0 row(s) affected	0.062 sec
30	01:09:01	LOAD DATA INFILE 'C:\\ProgramData\\MySQL\\MySQL Server 8.0\\Uploads\\credit_cards.csv'	5000 row(s) affected Records: 5000 Deleted: 0 Skipped: 0 Warnings: 0	0.250 sec

Siendo esta la primera tabla de dimensiones generada, se debe establecer la relación con la tabla de hechos **transaction**.

La relación entre ambas tablas es de 1: N, uno a muchos, donde con una tarjeta (*credit_card_id*) puedo realizar varios pedidos o transacciones, pero una transacción es gestionada por una única tarjeta.

Se establece el vínculo entre las tablas (**credit_card** y **transaction**) a través de la sentencia conformada por los comandos ALTER TABLE, ADD CONSTRAINT, FOREIGN KEY y REFERENCES. De esta forma se crea una restricción de clave foránea (*transaction_card_fkey*) para enlazar las dos tablas, garantizando que los valores de una columna en una de las tablas existan como clave

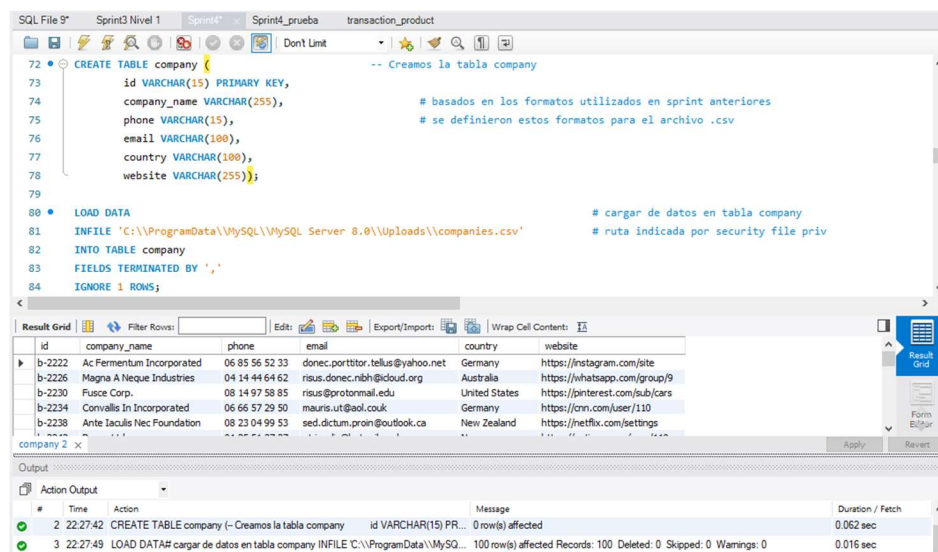
principal en la otra tabla, específicamente, como se asocia a la recién definida Foreign Key (**transaction.credit_card_id**) a la Primary Key (**credit_card.id**) de otra tabla. La imagen siguiente muestra la ejecución de la sentencia.



La siguiente tabla de dimensiones a crear es la tabla **company** que contendrá los datos de las empresas que venden u ofrecen productos, donde los campos definidos por medio del archivo **companies.csv** son:

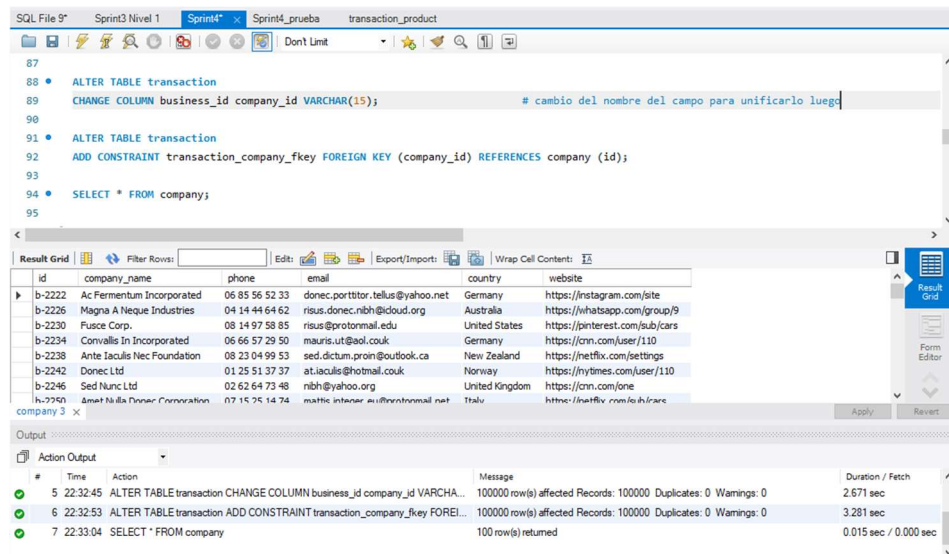
- **id**: número identificador del código de la empresa como clave primaria y única,
- **company_name**: nombre de la empresa,
- **phone**, **email**: referencias de contacto como teléfono y correo electrónico,
- **country**: información del país de ubicación, y
- **website**: su página web.

Al igual que la tabla anterior se creó la estructura de la tabla **company** y se cargaron los datos para llenar la tabla. Se identificó la Primary Key como **id**, verificando su similitud con un campo en la tabla de hechos **transaction**. En este caso hay coincidencia con el campo **business_id**, así se puede definir que la relación entre ambas tablas es de 1: N, uno a muchos, donde una empresa puede generar diversas ventas, pero una transacción producto de una única empresa.



Lo primero es unificar el nombre del campo que será el vínculo entre ambas tablas, cambiando con el comando ALTER y RENAME de **business_id** a **company_id**.

En la sentencia ADD CONSTRAINT añade la restricción (*transaction_company_fkey*), FOREIGN KEY especifica la columna que actúa como foránea en la tabla de hechos (*transaction.company_id*), y REFERENCES indica la tabla y columna principal con la que se está relacionando (*company.id*).

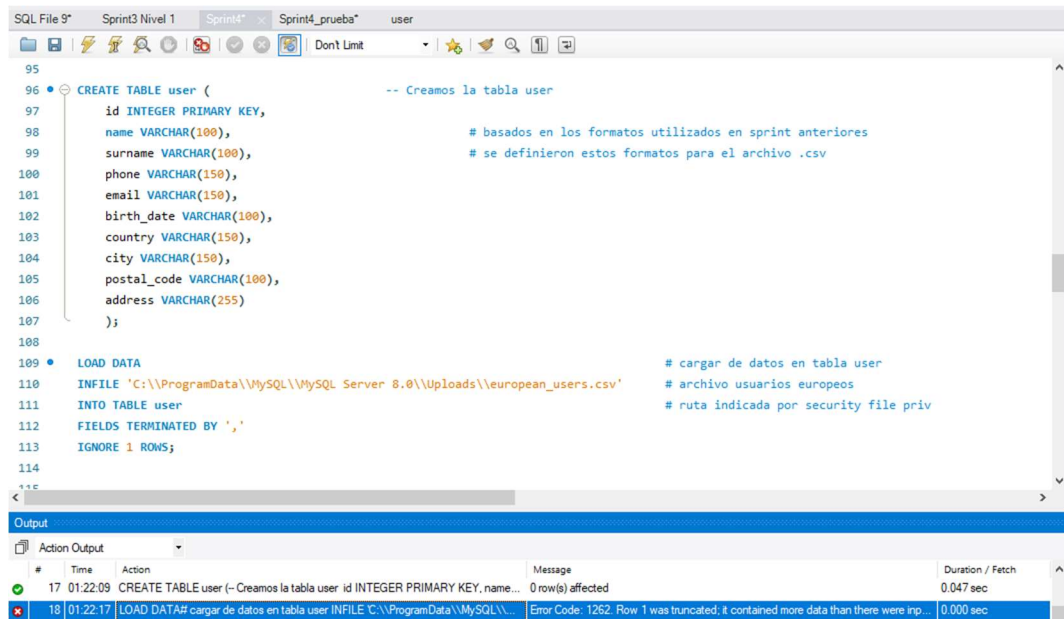


Ahora se creará la última tabla de dimensiones relacionada a los datos de los usuarios. En este caso se tienen 2 archivos de usuarios separados en región geográfica, *american_users.csv* y *european_users.csv*, por lo que se pueden realizar estructurar de diversas maneras. Analizando primero ambos archivos se observa que el nombre de los campos es igual en los 2 archivos:

- *id*: número identificador del código del usuario, asignado como Primary Key, se debe verificar que el formato coincida con *user_id* de la tabla **transaction**.
- *name*: nombre del usuario,
- *surname*: apellido del usuario,
- *phone*: teléfono del usuario,
- *email*: correo electrónico del usuario,
- *birth_date*: fecha de nacimiento del usuario,
- *country*: país donde se localiza o reside el usuario,
- *city*: ciudad donde se localiza o reside el usuario,
- *postal_code*: código postal del lugar donde se localiza o reside el usuario y
- *address*: dirección del sitio donde se localiza o reside el usuario.

Lo segundo es definir si se establecen 2 tablas por separado o se unifican en una única tabla **user**, esto último será lo que se realice, debido a que se tienen campos coincidentes, la numeración del identificador *id* es diferente en ambas tablas por lo que no se tendría problemas al momento de unirlos. Otro factor que se puede considerar es crear una columna adicional que identifique el continente, por si se desea realizar a futuro en la base de datos búsquedas o análisis por área geográfica.

Se utilizar el mismo procedimiento que antes, se crea la tabla **user** que tendrá los datos de todos los usuarios, considerando los formatos de los campos usados en sprint anteriores. Después se generan 2 sentencias para cargar los archivos *american_users.csv* y *european_users.csv*. En la imagen anexa se muestra el resultado.



```

95
96 CREATE TABLE user (
97     id INTEGER PRIMARY KEY,
98     name VARCHAR(100),
99     surname VARCHAR(100),
100    phone VARCHAR(150),
101    email VARCHAR(150),
102    birth_date VARCHAR(100),
103    country VARCHAR(150),
104    city VARCHAR(150),
105    postal_code VARCHAR(100),
106    address VARCHAR(255)
107 );
108
109 LOAD DATA
110 INFILE 'C:\\ProgramData\\MySQL\\MySQL Server 8.0\\Uploads\\european_users.csv'
111 INTO TABLE user
112 FIELDS TERMINATED BY ','
113 IGNORE 1 ROWS;
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2380
2381
2382
2383
2384
2385
2386
2387
2388
2389
2390
2391
2392
2393
2394
2395
2396
2397
2398
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2439
2440
2441
2442
2443
2444
2445
2446
2447
2448
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2459
2460
2461
2462
2463
2464
2465
2466
2467
2468
2469
2470
2471
2472
2473
2474
2475
2476
2477
2478
2479
2480
2481
2482
2483
2484
2485
2486
2487
2488
2489
2490
2491
2492
2493
2494
2495
2496
2497
2498
2499
2500
2501
2502
2503
2504
2505
2506
2507
2508
2509
2510
2511
2512
2513
2514
2515
2516
2517
2518
2519
2520
2521
2522
2523
2524
2525
2526
2527
2528
2529
2530
2531
2532
2533
2534
2535
2536
2537
2538
2539
2540
2541
2542
2543
2544
2545
2546
2547
2548
2549
2550
2551
2552
2553
2554
2555
2556
2557
2558
2559
2560
2561
2562
2563
2564
2565
2566
2567
2568
2569
2570
2571
2572
2573
2574
2575
2576
2577
2578
2579
2580
2581
2582
2583
2584
2585
2586
2587
2588
2589
2590
2591
2592
2593
2594
2595
2596
2597
2598
2599
2600
2601
2602
2603
2604
2605
2606
2607
2608
2609
2610
2611
2612
2613
2614
2615
2616
2617
2618
2619
2620
2621
2622
2623
2624
2625
2626
2627
2628
2629
2630
2631
2632
2633
2634
2635
2636
2637
2638
2639
2640
2641
2642
2643
2644
2645
2646
2647
2648
2649
2650
2651
2652
2653
2654
2655
2656
2657
2658
2659
2660
2661
2662
2663
2664
2665
```


The screenshot shows a SQL IDE with a query window containing the following SQL commands:

```

131 • SELECT * FROM user;
132
133 • ALTER TABLE transaction
134 ADD CONSTRAINT transaction_user_fkey FOREIGN KEY (user_id) REFERENCES user (id);
135
136 • SELECT * FROM user;
137

```

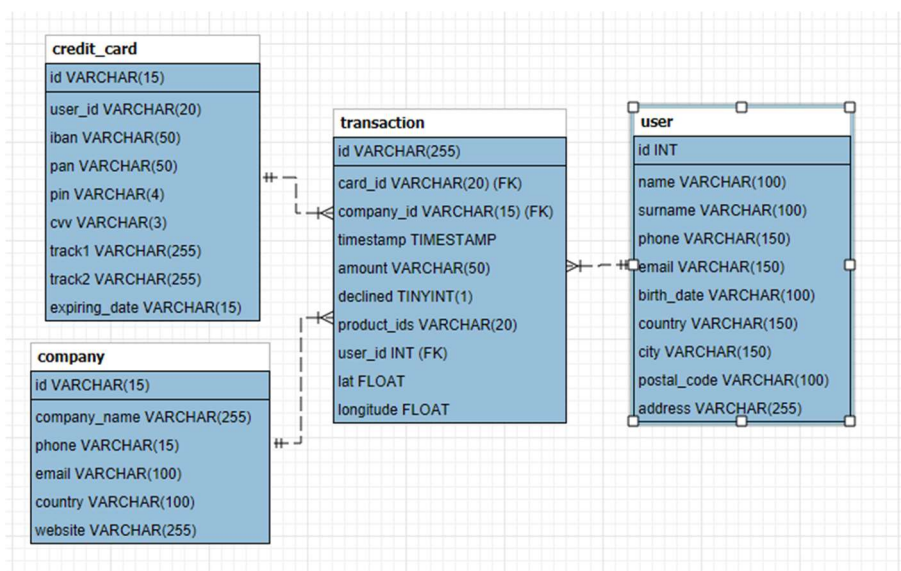
Below the query window is the 'Result Grid' showing a table with 10 rows of user data. The columns are: id, name, surname, phone, email, birth_date, country, city, postal_code, and address.

id	name	surname	phone	email	birth_date	country	city	postal_code	address
1	Zeus	Gamble	1-282-581-0551	interdum.enim@protonmail.edu	Nov 17, 1985	United States	New York	10001	348-7818 Sagittis St.
2	Garrett	Mcconnell	(718) 257-2412	integer.vitae.nibh@protonmail.org	Aug 23, 1992	United States	Philadelphia	19101	903 Sit Ave
3	Caran	Harrison	(522) 598-1365	interdum.feugiat@aol.org	Apr 29, 1998	United States	Houston	77001	736-2063 Tellus St.
4	Howard	Stafford	1-411-740-3269	ornare.egestas@cloud.edu	Feb 18, 1989	United States	Phoenix	85001	Ap #545-2244 Erat. Rd.
5	Hayfa	Pierce	1-554-541-2077	et.malesuada.fames@hotmail.org	Sep 26, 1998	United States	Philadelphia	19101	341-2821 Ultrices Av.
6	Joel	Tyson	(718) 288-8020	gravida.nunc.sed@yahoo.ca	Oct 15, 1989	United States	San Jose	95101	888-2799 Amet Street
7	Rafael	Jimenez	(817) 689-0478	egest@outlook.ca	Dec 4, 1981	United States	Chicago	60601	8627 Malesuada Rd.
8	Nissim	Franks	(692) 157-3469	egestas aliquam.fringilla@google.ca	Aug 1, 1993	United States	New York	10001	Ap #251-7144 Integer St.
9	Mennix	Mcclain	(590) 883-2184	aliquam.nisi@outlook.com	Jan 24, 1987	United States	San Antonio	78201	647-3080 Lacus. St.
10	Robert	Mccarthy	(324) 746-6771	fermentum@protonmail.com	Apr 30, 1984	United States	San Jose	95101	P.O. Box 773

Below the result grid is the 'Output' window showing the execution of the SQL commands:

#	Time	Action	Message	Duration / Fetch
21	01:29:51	SELECT * FROM user	5000 row(s) returned	0.000 sec / 0.015 sec
22	01:32:52	ALTER TABLE transaction ADD CONSTRAINT transaction_user_fkey FOREIGN K...	100000 row(s) affected Records: 100000 Duplicates: 0 Warnings: 0	4.297 sec
23	01:32:59	SELECT * FROM user	5000 row(s) returned	0.000 sec / 0.016 sec

Ya con las tablas cargadas en la base de datos **marketplace** y las relaciones entre las tablas definidas se muestra el esquema finalizado.



Ejercicio 1

Realiza una subconsulta que muestre a todos los usuarios con más de 80 transacciones utilizando al menos 2 tablas.

El enunciado pide una subconsulta, sin embargo, si se puede evitar el uso de subconsulta y aplicar un JOIN es lo más acertado. Para cumplir con lo solicitado se realiza subconsulta sin aplicar JOIN usando las tablas **transaction** y **user**. Como se pide aplicar una subconsulta se hace en el FROM con los datos requeridos de la tabla **transaction** donde se agrupa por usuarios y se aplica filtro de restricción a > 80 transacciones. Luego se realiza otra sentencia con la subconsulta sobre la tabla **user** en el SELECT de la consulta principal.

Se presenta en la imagen la consulta y resultados, 4 usuarios con más de 80 transacciones.

Ejercicio 1
Realiza una subconsulta que muestre a todos los usuarios con más de 80 transacciones utilizando al menos 2 tablas.

```

141 • SELECT id, t.num_transacciones, name, surname
142 FROM user, (SELECT user_id, COUNT(id) AS num_transacciones FROM transaction
143              GROUP BY user_id HAVING COUNT(id) > 80) AS t
144 WHERE t.user_id = id;
145
146 • SELECT user_id, COUNT(id) AS num_transacciones, (SELECT name FROM user
147              WHERE id = t.user_id) AS user_name
148 FROM transaction AS t
149 GROUP BY user_id
150 HAVING COUNT(id) > 80 ;

```

Result Grid

user_id	num_transacciones	user_name
185	110	Molly
289	94	Dxvngl
318	91	Bnyr
454	81	Sfztoh

Action Output

#	Time	Action	Message	Duration / Fetch
30	22:24:59	SELECT id, t.num_transacciones, name, surname FROM user, (SELECT user...	4 row(s) returned	0.047 sec / 0.000 sec
31	22:25:04	SELECT user_id, COUNT(id) AS num_transacciones, (SELECT name FROM ...	4 row(s) returned	0.109 sec / 0.000 sec

Ejercicio 2

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

Para esta consulta se requiere unir 3 tablas:

- **company** para filtrar por la empresa “Donec Ltd.”,
- **credit_card** para obtener los *iban* de las tarjetas y agrupar por este campo, y
- **transaction** para calcular las ventas promedio (*avg_ventas*) y vincular las otras tablas.

En la imagen anexa se presenta la consulta y los resultados, siendo 371 tarjetas con transacciones en la empresa Donec Ltd.

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

```

160 • SELECT ROUND(AVG(t.amount), 2) AS avg_ventas, ca.iban AS iban_card_DonecLtd
161 FROM transaction AS t
162 JOIN company AS co
163 ON co.id = t.company_id
164 JOIN credit_card AS ca
165 ON ca.id = t.card_id
166 WHERE co.company_name = "Donec Ltd"
167 GROUP BY ca.iban
168 ;

```

Result Grid

avg_ventas	iban_card_DonecLtd
356.25	XX911406401125586307586805
142.96	SK9446370242474562577506
257.37	XX77675291784595297555640
139.59	XX413827362289719304908990
240.41	XX347787246070769610780308
188.58	XX688768436543090894854602
439.39	MX28368851538688349

Action Output

#	Time	Action	Message	Duration / Fetch
63	00:39:41	SELECT ROUND(AVG(t.amount)) AS avg_ventas, ca.iban AS iban_card_DonecLtd FROM tran...	371 row(s) returned	0.016 sec / 0.000 sec

NIVEL 2

Crea una nueva tabla que refleje el estado de las tarjetas de crédito basado en: si las tres últimas transacciones han sido declinadas entonces es inactivo, si al menos una no es rechazada entonces es activo. Partiendo de esta tabla responde:

Ejercicio 1

¿Cuántas tarjetas están activas?

Para crear la tabla primero se definen que campos están involucrados *card_id*, *id*, *declined*, *timestamp*; todos de la tabla **transaction**. Ahora se requiere un comando que clasifique con una condición si son activas o inactivas (CASE WHEN). Pero se necesita establecer bien la condición ya que incluyen el ordenar cronológicamente cada transacción por tarjeta. Para estas búsquedas la función Ventana es adecuada (OVER).

```

170 # Nivel 2 - Nivel 2 - Nivel 2 - Nivel 2 - Nivel 2 - Nivel 2 - Nivel 2 - Nivel 2 - Nivel 2
171
172 # Crea una nueva tabla que refleje el estado de las tarjetas de crédito basado en:
173 # si las tres últimas transacciones han sido declinadas entonces es inactivo, si al menos una no es rechazada entonces es activo.
174 # Partiendo de esta tabla responde:
175
176 -- Se requieren 2 cosas para crear la nueva tabla:
177 # 1. Clasificar las tarjetas en activas e inactivas con un condicional, se usa CASE WHEN THEN
178 # 2. Definir la condición que permite separar las tarjetas, se usa la función ventana OVER, RANK, ROW_NUMBER
179
180 SELECT id, timestamp, card_id, declined, COUNT(declined)
181 OVER (PARTITION BY card_id ORDER BY timestamp DESC) AS orden
182 FROM transaction;
  
```

id	timestamp	card_id	declined	orden
B00988A9-5189-4ED8-9E4C-648B1D654F5B	2024-10-25 13:11:54	CcS-4857	0	1
B954A0B3-9314-4615-ACB8-831E0245D8D1	2024-10-07 16:43:17	CcS-4857	0	2
2C537AD9-D80E-402A-A79F-7D155138F47C	2024-07-27 10:50:49	CcS-4857	0	3
40A0B056-3438-40A9-8F5B-8CDE5660047D	2024-02-08 14:55:10	CcS-4857	1	4
FD857386-28F9-49DE-8F32-E2458CDE1752	2024-01-15 20:18:34	CcS-4857	0	5
8B8CF288-420E-4EAE-83C0-F222CB7F5863	2022-09-21 22:52:27	CcS-4857	0	6

Se inicia como se observa en la imagen aplicando la función ventana, esta permite usar funciones de agregación manteniendo el detalle de cada fila, lo cual es ideal para la clasificar datos. En la sentencia se usa el comando OVER para agrupar las tarjetas por transacciones declinadas o no (PARTITION BY), y ordenando por fecha (ORDER BY); así se puede identificar las ultimas transacciones y saber si fueron rechazadas o no. Esto ayudara a definir el condicional para crear la tabla.

```

187 # función ventana en select (para confirmar el orden de uso de tarjeta) y en condición del CASE para definir que separar
188 SELECT card_id, declined, COUNT(declined)
189 OVER (PARTITION BY card_id ORDER BY timestamp DESC) AS orden,
190 CASE
191   WHEN declined = 1 AND COUNT(declined)
192   OVER (partition by card_id order by timestamp DESC) IN (1, 2, 3) THEN "inactiva"
193   ELSE "activa"
194 END AS status
195 FROM transaction;
  
```

card_id	declined	orden	status
CcS-4898	0	49	activa
CcS-4898	0	50	activa
CcS-4898	0	51	activa
CcS-4898	0	52	activa
CcS-4898	0	53	activa
CcS-4898	0	54	activa
CcS-4899	1	1	inactiva
CcS-4899	1	2	inactiva
CcS-4899	1	3	inactiva
CcS-4899	0	4	activa
CcS-4899	0	5	activa
CcS-4899	0	6	activa

Ahora se procede como se ve en la imagen anterior a armar la tabla, introduciendo el comando CASE para definir las condiciones de tarjeta activa y tarjeta inactiva. Aquí se indica la función ventana que detalla la condición que se necesita cumplir.

A continuación, como se aprecia en la imagen siguiente se intenta agrupar por tarjetas, esto genera el error code 1055, asociado a incompatibilidad con el modo sql “only_full_group_by”, para solucionar se debe limitar los elementos del select o desactivar el modo. Se intenta limitar los campos del select, obteniendo el mismo resultado error code 1055.

```

185 SELECT card_id, declined, COUNT(declined)
186 OVER (PARTITION BY card_id ORDER BY timestamp DESC) AS orden,
187 CASE WHEN declined = 1 AND COUNT(declined)
188 OVER (partition by card_id order by timestamp DESC) IN (1, 2, 3) THEN "inactiva" # condiciones para inactivo
189 ELSE "activa" # condiciones para activo
190 END AS status
191 FROM transaction
192 GROUP BY card_id;
193
194 # al usar GROUP BY genera error 1055, indica no compatible con sql_mode = only_full_group_by
195 # se soluciona limitando el select o desactivando este modo
196 # incluso abajo limitando el select a solo el campo que agrupa y la funcion agregada
197
198 SELECT card_id, COUNT(declined)
199 OVER (PARTITION BY card_id ORDER BY timestamp DESC) AS orden,
200 CASE WHEN declined = 1 AND COUNT(declined)
201 OVER (partition by card_id order by timestamp DESC) IN (1, 2, 3) THEN "inactiva" # condiciones para inactivo
202 ELSE "activa" # condiciones para activo
203 END AS status
204 FROM transaction
205 GROUP BY card_id;
  
```

Output:

#	Time	Action	Message	Duration / Fetch
23	11:27:27	SELECT card_id, declined, COUNT(declined) OVER (PARTITION BY card...	Error Code: 1055. Expression #2 of SELECT list is not in GROUP BY clause a...	0.000 sec
24	11:28:54	SELECT card_id, COUNT(declined) OVER (PARTITION BY card_id ORDE...	Error Code: 1055. Expression #2 of SELECT list is not in GROUP BY clause a...	0.000 sec

Por tal motivo se decide desactivar temporalmente el modo “only_full_group_by” para generar la tabla **card_status**. Se identifico en documentación los comandos para activar y desactivar el modo. Al desactivarlo se pudo ejecutar la sentencia para crear la tabla, adicionalmente se limito a presenta la tarjeta y el status, ya que el uso del COUNT (*declined*) en función ventana era solo para poder realizar la condición de clasificación en el CASE.

```

207 # Se procede a desactivar el modo ONLY_FULL_GROUP_BY para aplicar la consulta que genera la tabla
208 SET sql_mode=(SELECT REPLACE(@@sql_mode,'ONLY_FULL_GROUP_BY','')); # DESACTIVAR
209
210 CREATE TABLE card_status AS # se crea una tabla que permita revisar a futuro la consulta
211 SELECT card_id, #COUNT(declined) OVER (PARTITION BY card_id ORDER BY timestamp DESC) AS orden,
212 CASE WHEN declined = 1 AND COUNT(declined)
213 OVER (partition by card_id order by timestamp DESC) IN (1, 2, 3) THEN "inactiva" # condiciones para inactivo
214 ELSE "activa" # condiciones para activo
215 END AS status
216 FROM transaction
217 GROUP BY card_id;
218
  
```

Result Grid:

card_id	status
CS-4893	activa
CS-4894	activa
CS-4895	inactiva
CS-4896	inactiva
CS-4897	activa

card_status 16 x

Output:

#	Time	Action	Message	Duration / Fetch
68	11:57:02	SET sql_mode=(SELECT REPLACE(@@sql_mode,'ONLY_FULL_GROUP_B...	0 row(s) affected	0.000 sec
69	11:57:08	CREATE TABLE card_status AS# se crea una tabla que permita revisar a fut...	5000 row(s) affected Records: 5000 Duplicates: 0 Warnings: 0	0.516 sec
70	11:57:21	SELECT * FROM card_status	5000 row(s) returned	0.000 sec / 0.016 sec

En la imagen siguiente se presenta la activación de modo sql “only_full_group_by” y la confirmación de consultar la nueva tabla. Se intento crear una vista (VIEW) en lugar de una tabla, pero al activar nuevamente el modo sql, generaba error al realizar cualquier consulta.

Finalmente, en la imagen se muestra el resultado de la consulta solicitada en el ejercicio, acerca del numero de tarjetas activas.

The screenshot shows a SQL query window with the following code:

```
221 # Se activa nuevamente el modo
222 SET SESSION sql_mode = CONCAT(@@sql_mode, ',ONLY_FULL_GROUP_BY'); # ACTIVAR
223
224 SELECT * FROM card_status; # se confirma que se puede consultar la tabla creada
225
226 # Ejercicio 1
227 # ¿Cuántas tarjetas están activas?
228
229 SELECT COUNT(status) # se consulta la tabla card_status filtrando status activo
230 FROM status_card
231 WHERE status = "activa";
232
```

The results pane shows a single row with the value 4983 for the COUNT(status) query.

#	Time	Action	Message	Duration / Fetch
79	12:04:31	SET SESSION sql_mode = CONCAT(@@sql_mode, ',ONLY_FULL_GROUP_...	0 row(s) affected	0.000 sec
80	12:04:33	SELECT * FROM card_status	5000 row(s) returned	0.000 sec / 0.016 sec
81	12:04:36	SELECT COUNT(status) # se consulta la tabla card_status filtrando status act...	1 row(s) returned	0.375 sec / 0.000 sec

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:

Primero se crea la estructura de la tabla **product** basados en el archivo *products.csv* que contiene los datos, donde se identifica:

- *id*: número identificador del código del producto, asignado como Primary Key,
- *product_name*: nombre del producto,
- *price*: precio del producto, en dólar (\$),
- *colour*: color de producto, en formato hexadecimal,
- *weight*: peso del producto,
- *warehouse_id*: se supone el almacén donde se localiza el producto,

The screenshot shows a SQL query window with the following code:

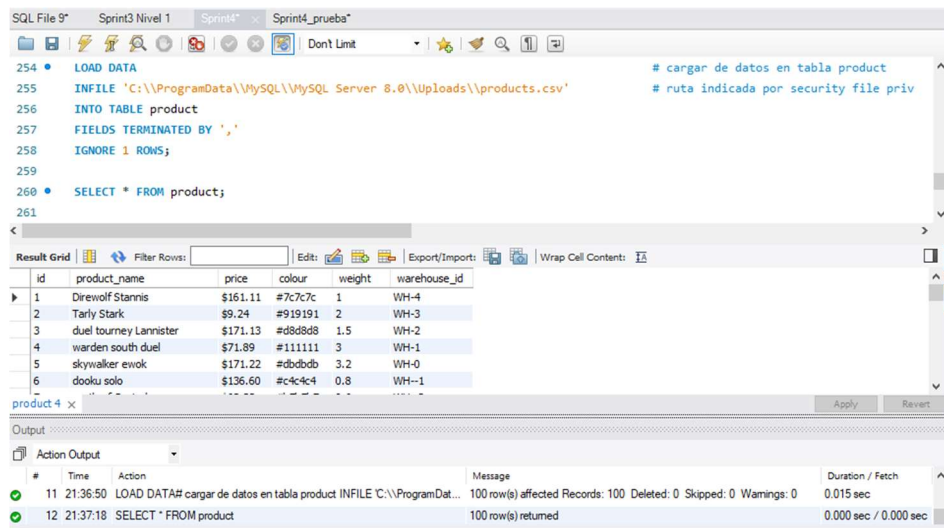
```
239 # Nivel 3 - Nivel 3 - Nivel 3 - Nivel 3 - Nivel 3 - Nivel 3 - Nivel 3 - Nivel 3 - Nivel 3
240
241 # Crea una tabla con la que podamos unir los datos del nuevo archivo products.csv con la base de datos creada,
242 # teniendo en cuenta que desde transaction tienes product_ids. Genera la siguiente consulta:
243
244 # primero se crea la tabla product definiendo los campos a partir del archivo products.csv
245
246 CREATE TABLE product ( -- Creamos la tabla product
247     id INTEGER PRIMARY KEY,
248     product_name VARCHAR(255), # se definieron estos formatos
249     price VARCHAR(20), # en funcion de los registros del archivo .csv
250     colour VARCHAR(15),
251     weight FLOAT,
252     warehouse_id VARCHAR(15));
253
254 LOAD DATA
255 INFILE 'C:\\ProgramData\\MySQL\\MySQL Server 8.0\\Uploads\\products.csv' # ruta indicada por security file priv
256 INTO TABLE product
257 FIELDS TERMINATED BY ','
258 IGNORE 1 ROWS;
259
```

The results pane shows the execution of the CREATE TABLE and LOAD DATA statements.

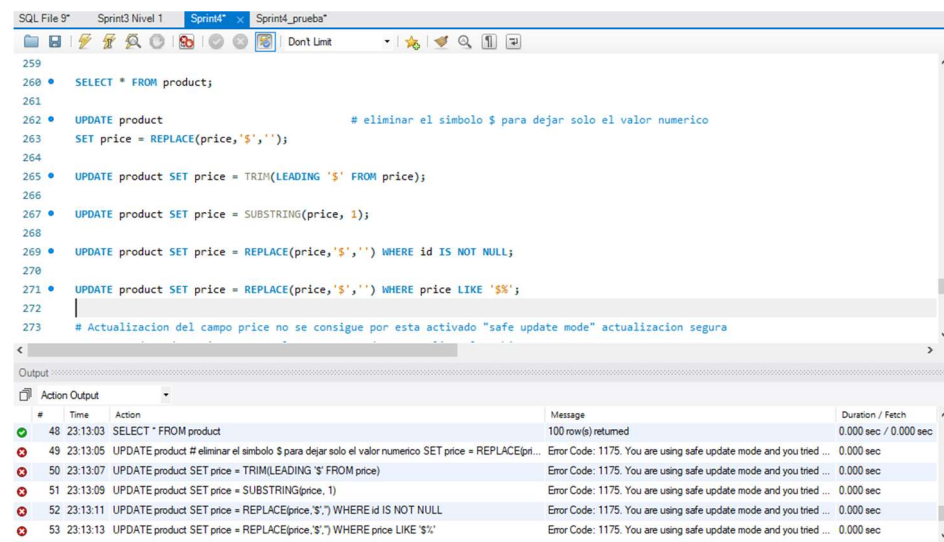
#	Time	Action	Message	Duration / Fetch
10	21:36:45	CREATE TABLE product (- Creamos la tabla product id INTEGER PRIMARY K...	0 row(s) affected	0.031 sec
11	21:36:50	LOAD DATA# cargar de datos en tabla product INFILE C:\\ProgramData\\My...	100 row(s) affected Records: 100 Deleted: 0 Skipped: 0 Warnings: 0	0.015 sec

Los formatos en los campos se determinan del tipo de registros vistos en el archivo .csv y la longitud aproximada de los mismos. En la imagen se aprecia las sentencias ejecutadas para crear y cargar datos en la tabla **product**.

A continuación, se anexa como quedan los datos cargados en la tabla, se identifica que para la columna precio los registros tiene delante el símbolo dólar (\$), considerando que es un campo de métricas con el que realizaran cálculos a futuro, se debe retirar el símbolo para dejar solo el valor numérico.

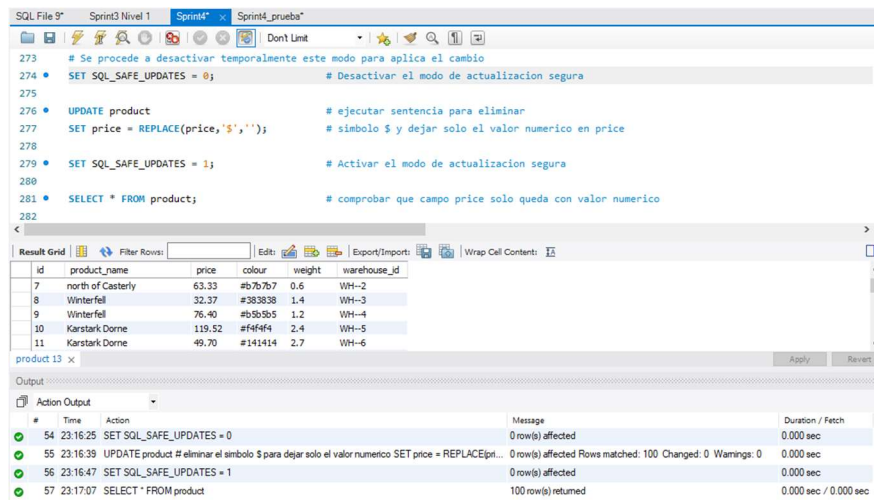


Para retirar el símbolo se utiliza el comando UPDATE y REPLACE para sustituir el símbolo \$ por un espacio vacío. En la imagen siguiente se muestra el error 1175, generado al intentar modificar la columna, indica que el modo de actualización segura (safe update mode) está activo. Este modo evita la eliminación o actualización accidental de datos. Para solucionar se puede modificar la consulta para incluir una clave o desactivar temporalmente el modo seguro.



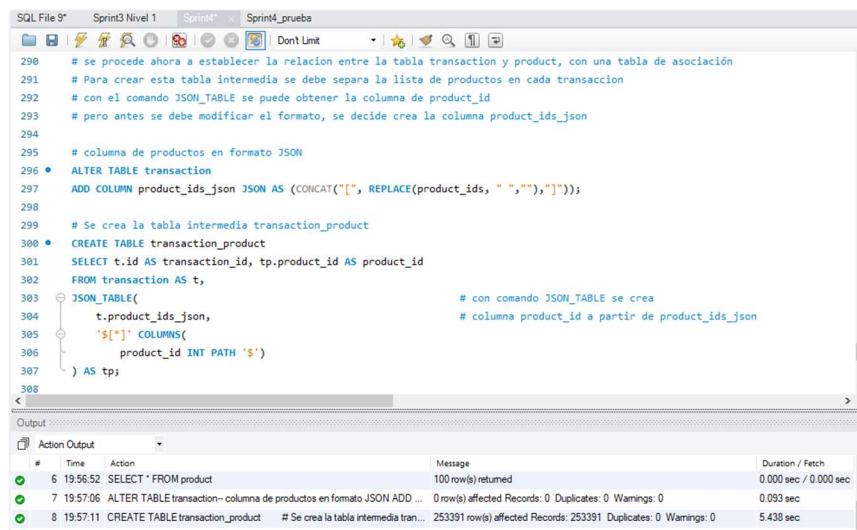
Con el fin de evitar alterar la seguridad se usaron diferentes comandos para intentar el cambio en la columna *price* (UPDATE con WHERE, TRIM y SUBSTRING), obteniendo el mismo error 1175.

Se decide desactivar el modo seguro aplicar el cambio y luego activarlo nuevamente. Como se observa en la imagen siguiente se puede modificar la columna *price* en la tabla **product**.



Ahora se debe realizar la relación entre la tabla de hecho **transaction** y la tabla de dimensión recién generada **product**. En este caso la relación es de N:M, de muchos a muchos, un conjunto de productos puede estar involucrados en diferentes transacciones. Para definir el vínculo se debe crear una tabla intermedia o de asociación, donde estén presentes las 2 Foreign Key representando a las Primary Key de ambas tablas.

Para la tabla **transaction** se tiene el campo *product_ids*, donde se listan los productos vendidos en cada transacción. Para crear la tabla intermedia es necesario separar estas listas de productos. Para esto se usará el comando **JSON_TABLE** que permite separar los elementos, pero necesita como entrada un archivo json, para esto se creará una columna adicional.



En la imagen anterior se aprecia la creación de la columna *product_ids_json* en formato json (línea 296), a partir de *product_ids* e insertando los corchetes ([]) propios de los archivos json.

Luego se crea la tabla intermedia de asociación **transaction_product** con los campos **transaction.id** y la nueva columna *product_id*, generada dentro de la sentencia a partir de **JSON_TABLE** usando la columna creada previamente en formato json.

Ya con la tabla de unión **transaction_product** creada, se generan los vínculos de esta con las tablas **transaction** y **product**. Se define las restricciones (CONSTRAIN) y Foreign key, junto con las referencias a cada Primary key. La sentencia se aplica para cada tabla por separado. De esta manera se asocia:

- **transaction.id = transaction_product.transaction_id** y
- **product.id = transaction_product.product_id**.

En la imagen siguiente se muestra las sentencias para generar las relaciones y la tabla de asociación creada.

```

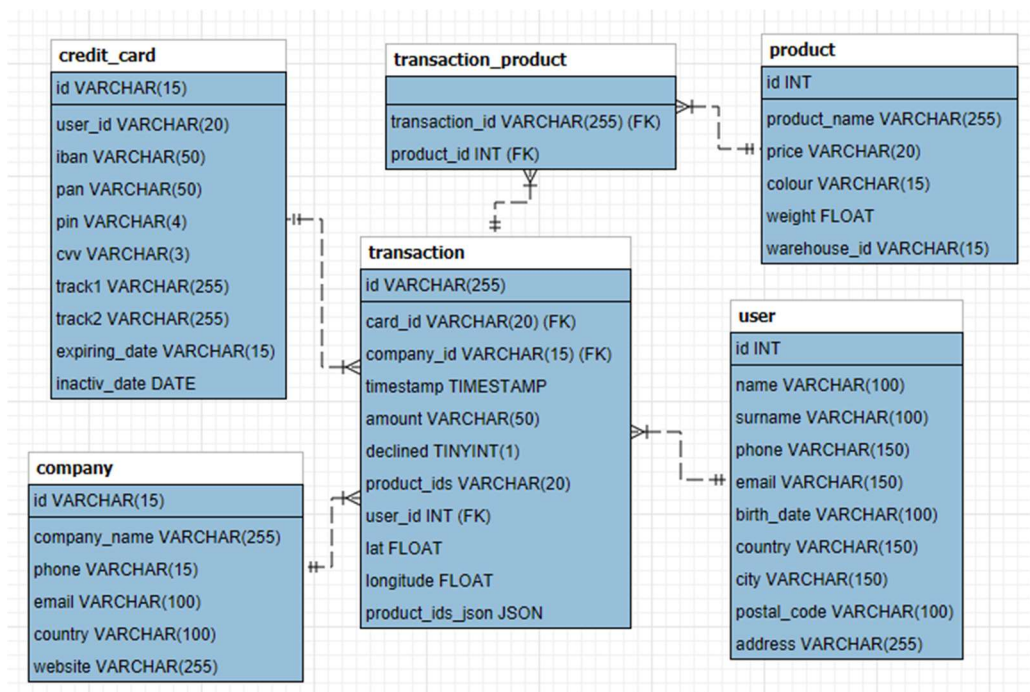
311 # Relacion entre la tabla transaction_product a traves de restricción y Foreign key con tabla producto y transaction
312
313 • ALTER TABLE transaction_product
314   ADD CONSTRAINT product_fkey FOREIGN KEY (product_id) REFERENCES product (id); -- referencia a traves de product.id
315
316 • ALTER TABLE transaction_product
317   ADD CONSTRAINT transaction_fkey FOREIGN KEY (transaction_id) REFERENCES transaction (id); -- referencia a traves de product.id
318
319 • SELECT * FROM transaction_product;
320

```

transaction_id	product_id
00043A49-2949-494B-A5DD-A5BAE388190D	16
00043A49-2949-494B-A5DD-A5BAE388190D	26
00043A49-2949-494B-A5DD-A5BAE388190D	97
00043A49-2949-494B-A5DD-A5BAE388190D	87
000447FE-8650-4DCF-8SDE-C7ED0EE1CAAD	66
000447FE-8650-4DCF-8SDE-C7ED0EE1CAAD	69
000447FE-8650-4DCF-8SDE-C7ED0EE1CAAD	87
00045D6B-ED2E-4F2F-8186-CEE074D875D0	30

#	Time	Action	Message	Duration / Fetch
13	20:13:56	ALTER TABLE transaction_product ADD CONSTRAINT product_fkey FOREIGN KEY (product_id) REFERENCES product (id);	253391 row(s) affected Records: 253391 Duplicates: 0 Warnings: 0	2.968 sec
14	20:14:04	ALTER TABLE transaction_product ADD CONSTRAINT transaction_fkey FOREIGN KEY (transaction_id) REFERENCES transaction (id);	253391 row(s) affected Records: 253391 Duplicates: 0 Warnings: 0	5.266 sec
15	20:15:03	SELECT * FROM transaction_product	253391 row(s) returned	0.000 sec / 0.391 sec

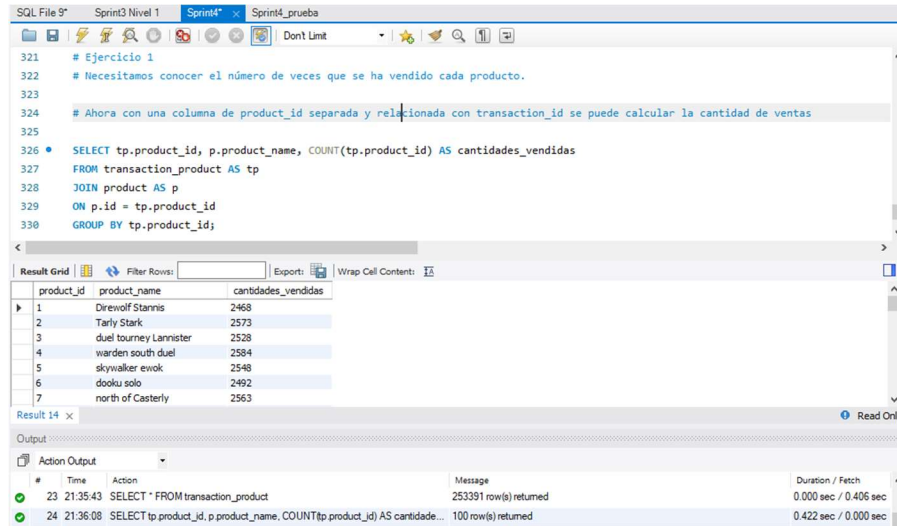
Como resultado final se presenta el diagrama de la base de datos **marketplace**.



Ejercicio 1

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

Para saber cuántas veces se han vendido los productos se realiza la consulta con un COUNT de productos sobre las transacciones realizadas y agrupando por productos, se adiciono el nombre del producto con un JOIN a la tabla **product**. En la imagen se presenta la consulta con el resultado para cada producto.



The screenshot shows a SQL IDE window with a query editor and a results grid. The query is as follows:

```
321 # Ejercicio 1
322 # Necesitamos conocer el número de veces que se ha vendido cada producto.
323
324 # Ahora con una columna de product_id separada y relacionada con transaction_id se puede calcular la cantidad de ventas
325
326 • SELECT tp.product_id, p.product_name, COUNT(tp.product_id) AS cantidades_vendidas
327 FROM transaction_product AS tp
328 JOIN product AS p
329 ON p.id = tp.product_id
330 GROUP BY tp.product_id;
```

The results grid shows the following data:

product_id	product_name	cantidades_vendidas
1	Direwolf Stannis	2468
2	Tarly Stark	2573
3	duel tourney Lannister	2528
4	warden south duel	2584
5	skywalker evok	2548
6	dooku solo	2492
7	north of Casterly	2563

The output pane shows the execution of the query, indicating that 100 rows were returned.