Projektmunka 1

Gyurcsán Dávid Lóránt

Tartalomjegyzék

Bevezetés	4
A projekt témája	4
Az adatok forrása	4
Tervezés	5
EK diagram	5
Az adatbázis táblái és tárolt adatai	6
car_brands	6
car_company	6
car_customers	6
car_employees	6
car junyard	7
car_mechanics	7
car_models	7
car_serviceprice	7
car_services	8
car_shops	8
A végleges táblák és kapcsolataik	9
Megvalósítás	10
A táblák feltöltése adatokkal	10
car_brands	10
car_company	10
car_customers	10
car_employees	10
car_junyard	10
car_mechanics	10
car_models	10
car_serviceprice	11
car_services	11
<u>car_shops</u>	11

Lekérdezések elemzése, optimalizáció	12
Elemzés	12
Optimalizáció	13
Elemzés	14
<u>Optimalizáció</u>	14
NoSQL adatbázis kezelés	15
Az adatbázis tábláinak átalakítása	15
Az adatok átvitele MongoDB-be	15-17
Lekérdezések	18
Mely autókereskedések vannak Olaszországban?	18
Melyik autó kereskedések vannak kapcsolatban az "a mi" nevezetű autószerelő 1műhellyel?	18
Legdrágábban javított autó.	18
Legolcsóbban javított autó.	19
Melyik aza az autó amit Gambiaban készítettek és ott is adták el?	19
Melyik autót vásárolta meg egy német állampolgár egy németországi kereskedést	:ől? 20
Kik dolgoznak a "Mauris" autószerelő műhelybe?	20
Átlagárak autószerelő műhelyeként	21
Melyek a legdrágább autószerelő mühelyek?	21
Hány vásárlót hívnak "Joshua"-nak?	21
Mellékletek	22
Táblák létrehozása	22

Bevezetés

A projekt témája

A projekt egy kitalált autó kereskedés, márkaszervíz és roncstelep adatainak kezelését valósítja meg. Az autókat megvásárolhatják a vevők az autó kereskedésben, probléma esetén az autót márkaszervízben javítják az autószerelők.

Nagyobb kár esetén, amelyik autót már nem lehet megjavítani az autót kivonják a forgalomból és a roncstelepre kerül.

Az autókat cégek számára is árusítják.

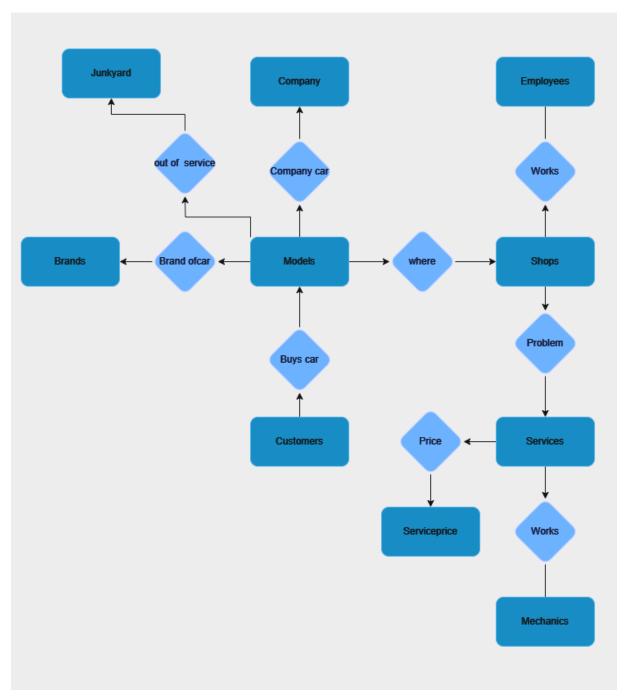
Az adatok forrása

Az adatok generáltak, a spawner.exe program használatával készültek, importálva lettek az adatbázisba.

Tervezés

EK diagram

Az egyed-kapcsolat diagram megtervezéséhez a draw.io web alkalmazását használtam.



Az adatbázis táblái és tárolt adatai

car_brands

Ebben a táblában az autók márkáit tároljuk.

A következő adatokat tároljuk:

• brand_id: a márka azonosítója, az elsődleges kulcs

• brand: az autó márkája

car_company

Ebben a táblában azokat a vevőket tároljuk, akik céges autóként vásároltak.

A következő adatokat tároljuk:

• company_id: a cég azonosítoja, az elsődleges kulcs

• name: a cég neve

• owner: a cég tulajdonosa

• model_id: vásárolt autó azonoítója, az idegen kulcs

car_customers

Ebben a táblában tároljuk el az ügyfeleket.

A következő adatok kerülnek eltárolásra:

- customer_id: az ügyfél azonosítója, ez lesz az elsődleges kulcs,
- first_name: az ügyfél keresztneve,
- last name: az ügyfél vezetékneve,
- email: az ügyfél e-mail címe,
- phone: az ügyfél telefonszáma,
- city: az ügyfél lakhelye település szinten,
- country: az ügyfél lakhelye ország szinten,
- birth_date: az ügyfél születési dátuma.

car_employees

Ebben a táblában a munkásokat tároljuk.

A következő adatokat tároljuk:

- employee_id: az eladó azonosítója, az elsődleges kulcs
- name: az eladó neve
- shop_id: autó kereskedés azonosítója, az idegen kulcs

car_junyard

Ebben a táblában a roncstelepeket tároljuk.

A következő adatokat tároljuk:

- junkyard_id: roncstelep azonosítója, az elsődleges kulcs
- name: roncstelep neve
- junked_date: autó ronccsá nyilvánításának dátuma
- model_id: autó azonosítója, az idegen kulcs

car_mechanics

Ebben a táblában az autószerelőket tároljuk.

A következő adatokat tároljuk:

- mechanic_id: autószerelő azonosítója, az elsődleges kulcs
- name: az autószerelő neve
- service_id: az autószerelő műhely azonosítója, az idegen kulcs

car_models

Ebben a táblában az autók modelljeit tároljuk.

A következő adatokat tároljuk:

- model_id: autó modell azonosító száma
- model: autó modell neve
- city: az autó gyártási városa
- country: az autó gyártási országa
- made_date: gyártási idő
- customer_id: vásárló azonosítója
- brand id: márka azonosító
- shop_id: autó kereskedés azonosító

car_serviceprice

Ebben a táblában az autószerelés árait tároljuk autónkként.

A következő adatokat tároljuk:

- Price_id: ár azonosító
- Usd: ár
- model_id: autó azonosító
- service_id: autószerelő műhely azonosító

car_services

Ebben a táblában az autószerelő műhelyeket tároljuk.

A következő adatokat tároljuk:

- service_id: az autószerelő műhely azonosítója
- name: az autószerelő műhely neve

car_shops

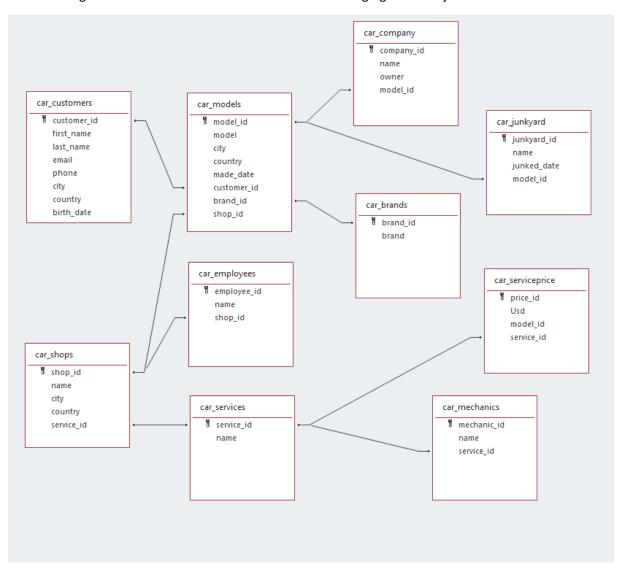
Ebben a táblában az autó kereskedéseket tároljuk.

A következő adatokat tároljuk:

- shop_id: autó kereskedés azonosítója
- name: autó kereskedés megnevezése
- city: autó kereskedés címe város szinten
- country: autó kereskedés címe ország szinten
- service_id: autószerelő műhely azonosítója

A végleges táblák és kapcsolataik

A táblák megtervezése után a következő lett az adatbázis végleges modellje:



Megvalósítás

A táblák feltöltése adatokkal

car_brands

A car_brands tábla elkészítéséhez a spawner.exe programot használtam.

Generáltattam 100000 sornyi adatot ehhez a táblához, az általam megadott táblák elnevezéseivel és a megfelelő formátumok megadásával.

car_company

Acar_company tábla elkészítéséhez a spawner.exe programot használtam.

Generáltattam 50000 sornyi adatot ehhez a táblához, az általam megadott táblák elnevezéseivel és a megfelelő formátumok megadásával.

car_customers

A car_customers tábla elkészítéséhez a spawner.exe programot használtam.

Generáltattam 1000000 sornyi adatot ehhez a táblához, az általam megadott táblák elnevezéseivel és a megfelelő formátumok megadásával.

car_employees

A car_employees tábla elkészítéséhez a spawner.exe programot használtam.

Generáltattam 5000 sornyi adatot ehhez a táblához, az általam megadott táblák elnevezéseivel és a megfelelő formátumok megadásával.

car_junyard

A car_junyard tábla elkészítéséhez a spawner.exe programot használtam.

Generáltattam 5000 sornyi adatot ehhez a táblához, az általam megadott táblák elnevezéseivel és a megfelelő formátumok megadásával.

car_mechanics

A car_mechanics tábla elkészítéséhez a spawner.exe programot használtam.

Generáltattam 500 sornyi adatot ehhez a táblához, az általam megadott táblák elnevezéseivel és a megfelelő formátumok megadásával.

car_models

A car_models tábla elkészítéséhez a spawner.exe programot használtam.

Generáltattam 1000000 sornyi adatot ehhez a táblához, az általam megadott táblák elnevezéseivel és a megfelelő formátumok megadásával.

car_serviceprice

A car_serviceprice tábla elkészítéséhez a spawner.exe programot használtam.

Generáltattam 10000 sornyi adatot ehhez a táblához, az általam megadott táblák elnevezéseivel és a megfelelő formátumok megadásával.

car_services

A car_services tábla elkészítéséhez a spawner.exe programot használtam. Generáltattam 100 sornyi adatot ehhez a táblához, az általam megadott táblák elnevezéseivel és a megfelelő formátumok megadásával.

car_shops

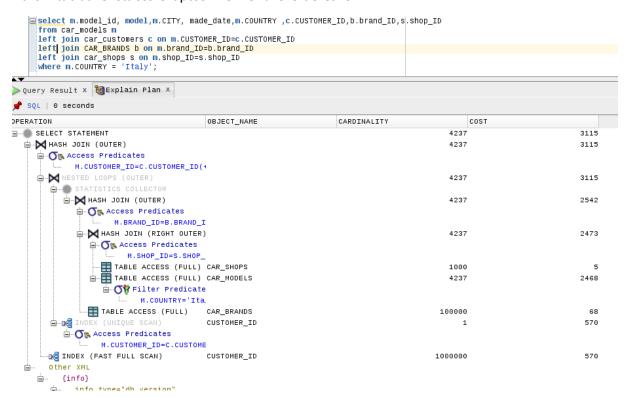
A car_shops tábla elkészítéséhez a spawner.exe programot használtam. Generáltattam vele 1000 sornyi adatot ehhez a táblához, az általam megadott táblák elnevezéseivel és a megfelelő formátumok megadásával.

Lekérdezések elemzése, optimalizáció

Elemzés

A lekérdezésben megszeretnénk tudni hogy melyik autót ki vásárolta meg, mi a márkája és, hogy melyik autó kereskedésben vásárolták.

Három táblát kellett összekapcsolni ehhez a lekérdezéshez.



Hash join (Right Outer)-t hajt végre a model és az autókereskedés táblára.

Hash join (Outer)-al összeköti a model és a márka táblákat.

Hash join (Outer)-al összeköti végül a model és a vásárló táblákat is.

A becsült cardinality 4237, és a becsült cost pedig 3115.

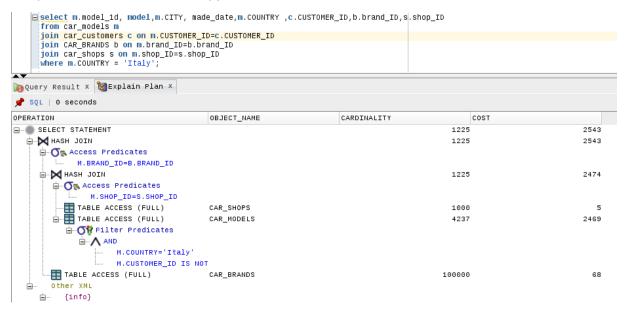
Optimalizáció

A lekérdezés optimalizálásához nem kötöttem meg a tábla összekötés fajtáját.

A lekérdezéshez az oracle továbbra is Hash Join-okat alkalmaz.

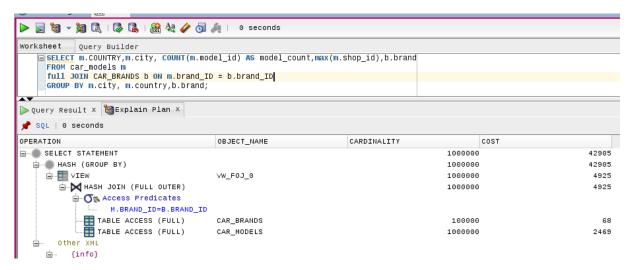
A where feltételben megadott "Italy" szűrést előbb elvégezte így is csökkentve a táblák összekapcsolásához kellő sorok számát.

Az optimalizáció után a cardinality jelentősen csökkent 1225-re és a cost is csökkent 2543-ra.



Elemzés

Ebben a lekérdezésben az országot, a várost, az ott készült modellek számát, a legnagyobb id-vel rendelkező autó kereskedés id-jét és az autó márkáját tudhatjuk meg csoportosítva város, ország és márka szerint.



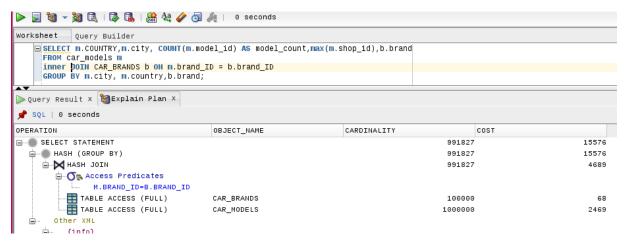
Hash Join (full outer) használatával összeköti a models táblát a brands táblával.

Majd Group By használatával megcsinálja a csoportosítást.

A lekérdezés Cardinalty száma meg egyezik a models tábla sorainak számával ami 1000000, a Cost 42905.

Optimalizáció

Az optimalizációhoz a két tábla között meglett változtatva a full join, inner join-ra. Hash joint segítségével megint összeköti a 2db táblát. Utána pedig Group By-al csoportosít. A cardinality 991827 re csökkent és a cost pedig több mint a felére 15576-ra.



NoSQL adatbázis kezelés

Az adatbázis tábláinak átalakítása

A NoSQL adatbázisként a MongoDB-t választottam.

Az adatok átvitele MongoDB-be

Json-ba exportáltam a táblákat, majd importáltam a MongDB-be, aggregate parancsal összeépítettem a azokat.

```
db.getCollection('mechanics').aggregate([
$lookup: {
from: "services",
localField: "service_id",
foreignField: "service id",
as: "services"
   $out: "mechanics2"
db.getCollection('shops').aggregate([
 $lookup: {
 from: "services",
 localField: "service id",
 foreignField: "service_id",
 as: "services"
    $out: "shops2"
db.getCollection('employees').aggregate([
$lookup: {
from: "shops2",
localField: "shop id",
foreignField: "shop id",
as: "shops2"
    $out: "employeesv2"
```

Lekérdezések

Mely autókereskedések vannak Olaszországban?

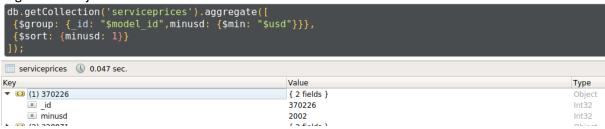


Melyik autó kereskedések vannak kapcsolatban az "a mi" nevezetű autószerelő műhellyel?

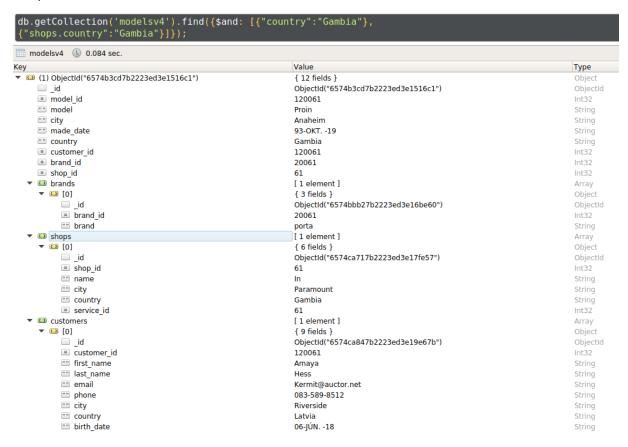


Legdrágábban javított autó.

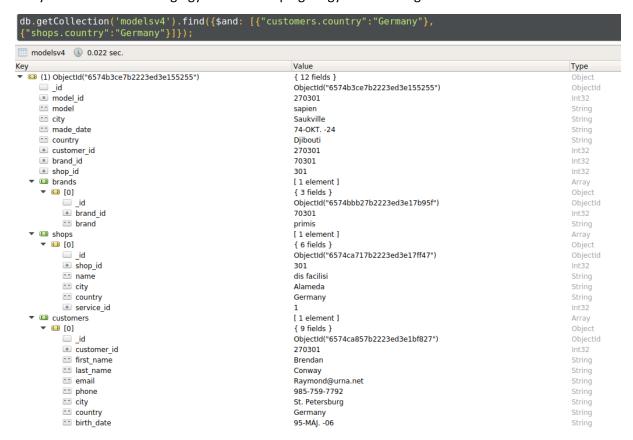
Legolcsóbban javított autó.



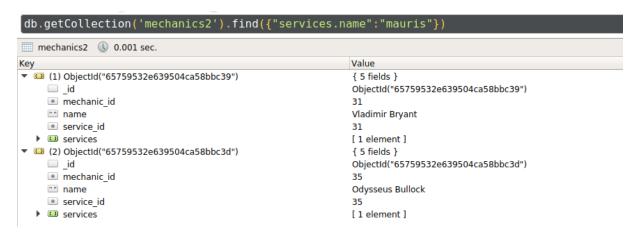
Melyik aza az autó amit Gambiaban készítettek és ott is adták el?



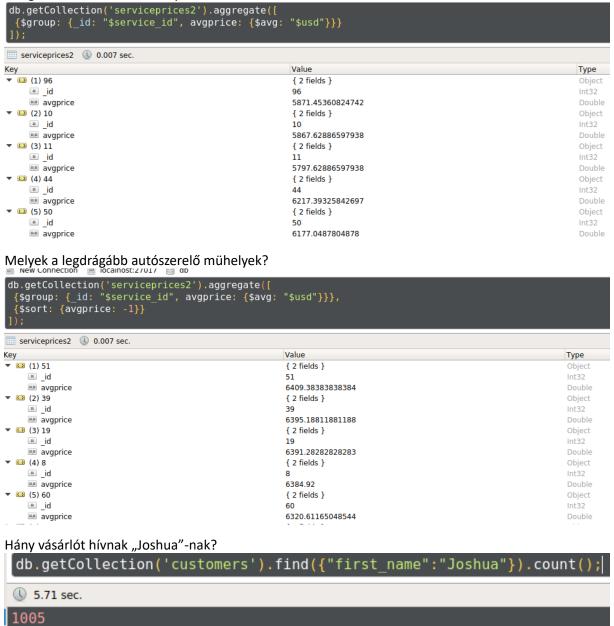
Melyik autót vásárolta meg egy német állampolgár egy németországi kereskedéstől?



Kik dolgoznak a "Mauris" autószerelő műhelybe?



Átlagárak autószerelő műhelyeként.



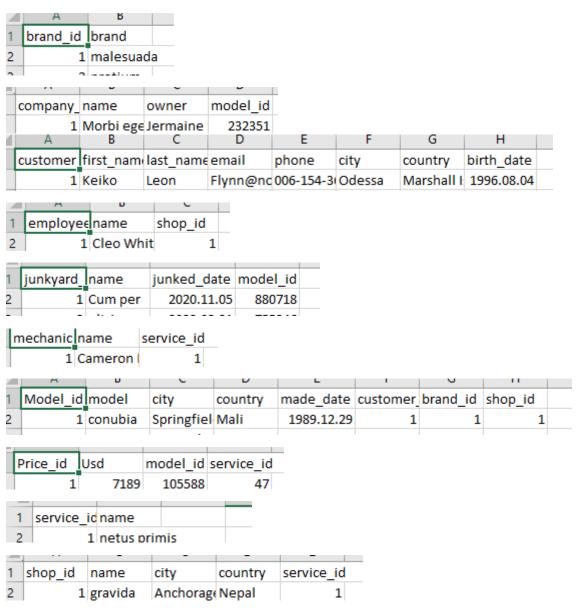
Mellékletek

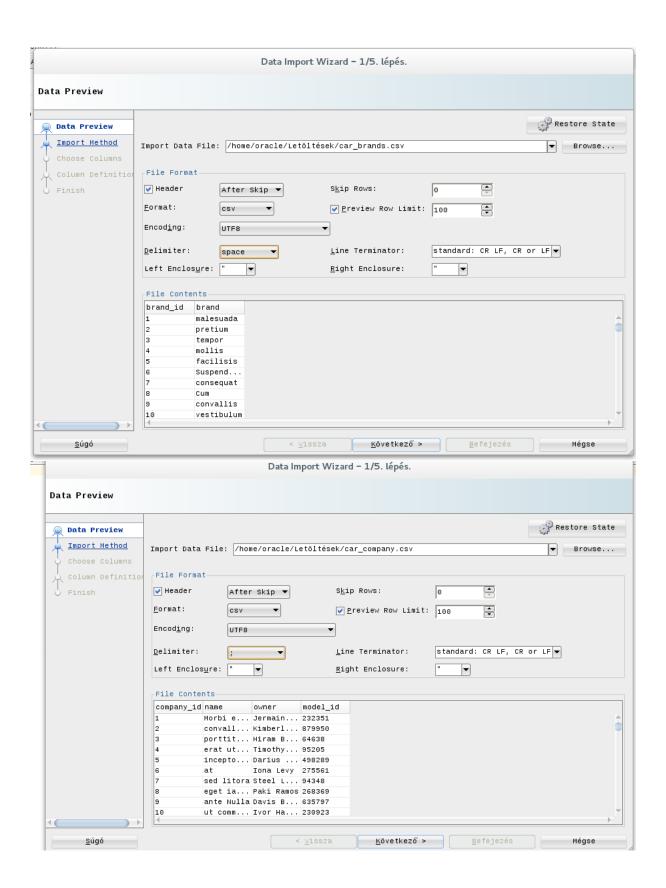
Táblák létrehozása

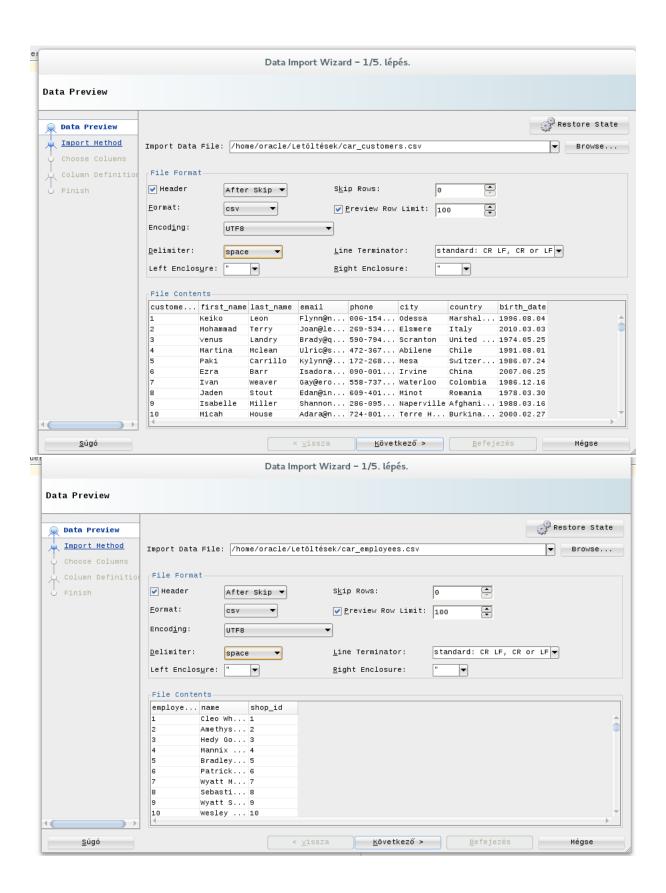
A táblákat Excelben kiegészítettem egy sorral, ami tartalmazza az oszlopok neveit.

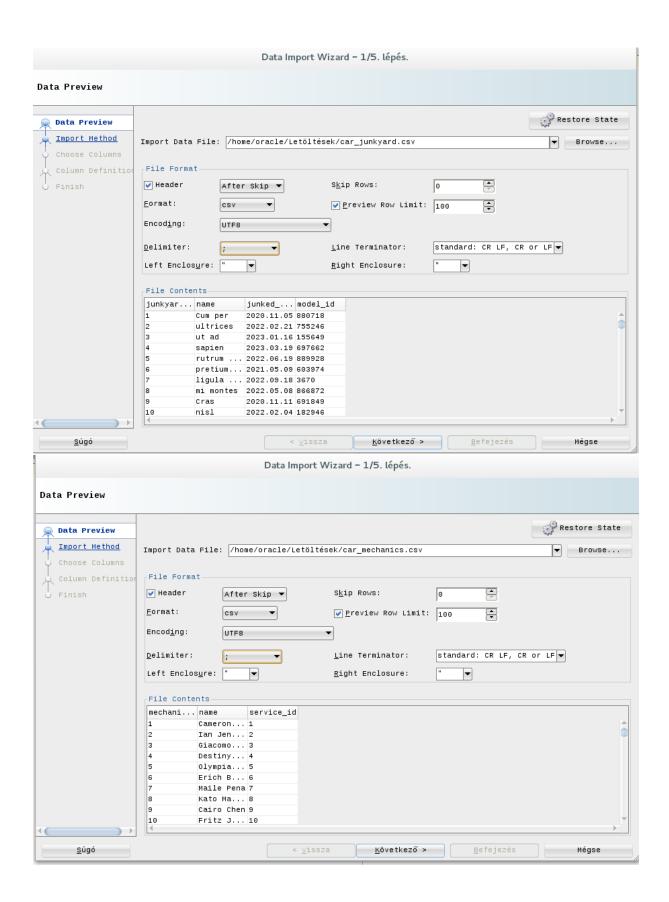
Csv formátumba elmentettem és átmásoltam a virtuális gépre.

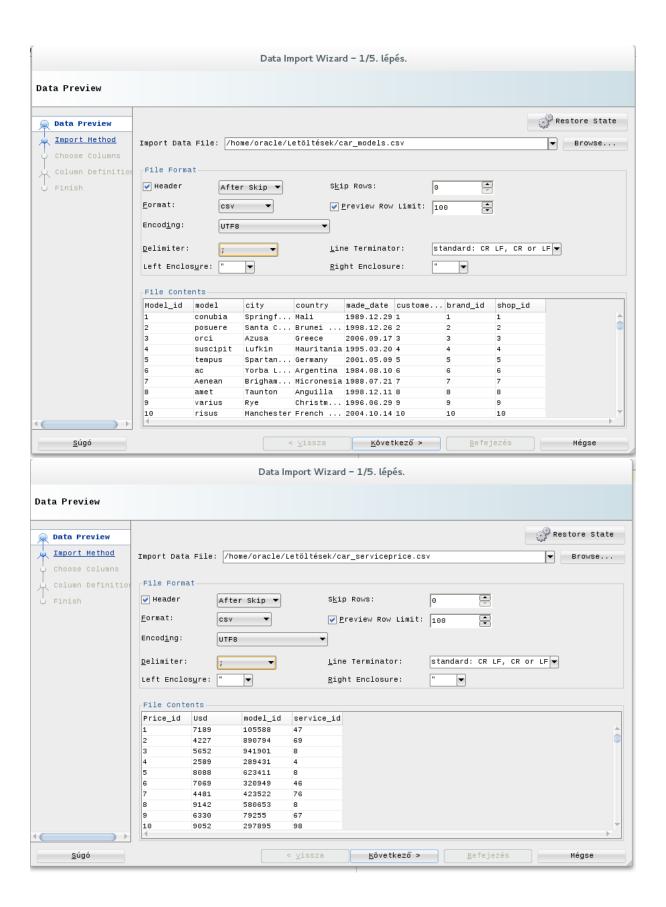
Oracle adatbázisba a Data Import Wizard segítségével importáltam a táblákat.

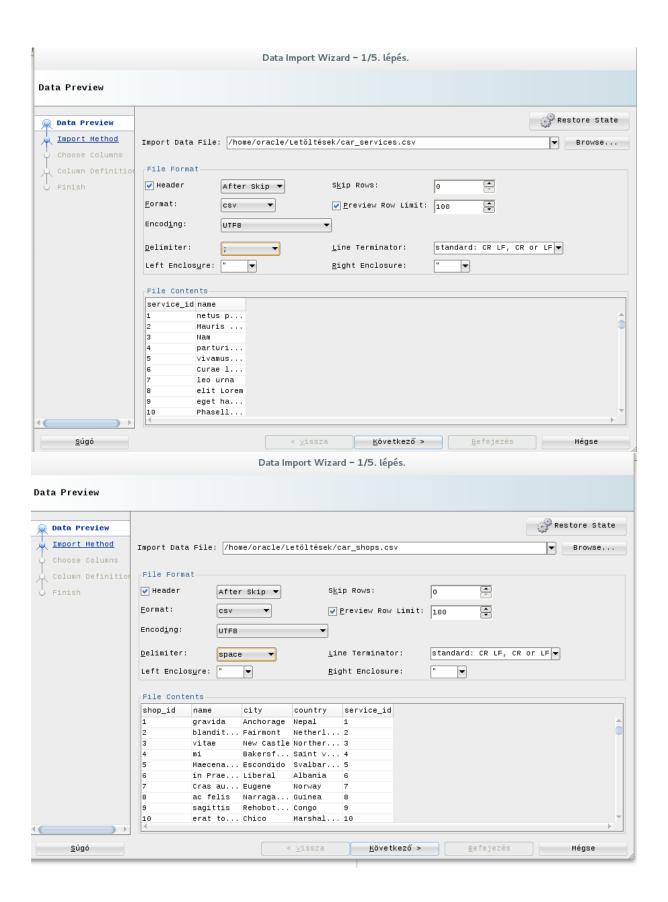












```
itsh@ubuntu:~$ mongoimport --db db --collection company --file ~/exportcompany.json
2023-12-10T02:37:31.646-0800
                                 connected to: localhost
2023-12-10T02:37:32.300-0800
                                  imported 50000 documents
itsh@ubuntu:~$ mongoimport --db db --collection employees --file ~/exportemployees.json
2023-12-10T02:37:56.336-0800
                                 connected to: localhost
2023-12-10T02:37:56.487-0800
                                 imported 5000 documents
itsh@ubuntu:~$ mongoimport --db db --collection junkyard --file ~/exportjunkyard.json
                                 connected to: localhost imported 5000 documents
2023-12-10T02:38:20.767-0800
2023-12-10T02:38:20.890-0800
itsh@ubuntu:~$ mongoimport --db db --collection mechanics --file ~/exportmechanics.json
2023-12-10T02:38:42.559-0800
                                 connected to: localhost
                                 imported 500 documents
2023-12-10T02:38:42.567-0800
itsh@ubuntu:~$ mongoimport --db db --collection services --file ~/exportservices.json
2023-12-10T02:39:06.319-0800
                                 connected to: localhost
2023-12-10T02:39:06.325-0800
                                  imported 100 documents
itsh@ubuntu:~$ mongoimport --db db --collection serviceprices --file ~/exportserviceprices.json
                                 connected to: localhost imported 10000 documents
2023-12-10T02:39:23.065-0800
2023-12-10T02:3<u>9</u>:23.207-0800
itsh@ubuntu:~$
```