Curso de Especialización en Inteligencia Artificial y BigData

Sistemas de Big Data

Tarea Evaluable 4.1

Autor: Carlos Sánchez Recio.
07 / 01 / 2025

Índice

1. Crea la tabla vehículos en HBase. 1
2. Inserta mediante órdenes `put` las tres primeras filas de la tabla. 1
3. Haz un archivo CSV con las 3 últimas filas y cárgalo en la tabla. 2
4. Muestra todos los datos de la fila 4 con un `get`. 4
5. Muestra los datos de información general de la tabla con un `scan`. 4
6. Muestra todos los datos de los vehículos con estado *bo*. 5
7. Muestra los datos de los propietarios de los vehículos de Mallorca. 6
8. Muestra la información general de los vehículos que sean posteriores a 2017 (incluido) con menos de 50000 km. 6
9. Modifica la fila 2 y ponle el teléfono 656123321. 6
10. Borra todos los datos en la fila 5. 6
11. Crea una tabla externa en Hive vinculada a la tabla vehículos de HBase. 7
12. Haz una consulta en HiveQL que devuelva los datos de información general de los vehículos de Palma. 8
13. Haz una consulta en Impala que devuelva todos los datos de los 3 vehículos más caros. 8
14. Haz una consulta en HiveQL que agrupe a los vehículos por estado y muestre la media de km de cada estado. 9
15. Haz una consulta en Impala que agrupe a los vehículos por isla y muestre el número de vehículos de cada isla. 9

En esta tarea vamos a trabajar con HBase. Tenemos una tabla *vehicles* para gestionar un sistema de venta de vehículos de segunda mano .

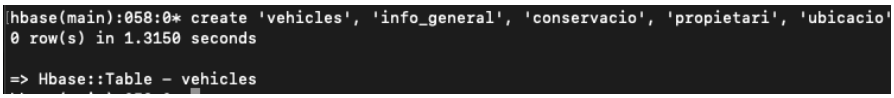
Esta tabla muestra los datos que debe contener la tabla *vehicles*:

clau	info_general				conservacio			propietari				ubicacio			
	marca	model	matricula	preu	any	km	estat	nif	nom	telefon	mail	ciutat	illa	longitud	latitud
1	Toyota	Corolla	1234ABC	12000	2015	120000	bo	12345678A	Joan Martínez	623456789	joan.martinez@email.com	Palma	Mallorca	2.650160	39.569600
2	Ford	Focus	5678DEF	15500	2018	55000	molt bo	87654321B	Ana García		ana.garcia@email.com	Palma	Mallorca		
3	Honda	Civic	9012GHI	18000	2020	20000	bo	11223344C	David López	655123456		Eivissa	Eivissa	1.452850	38.916280
4	Mercedes-Benz	A-Class	3456JKL	30000	2019	40000	bo	22334455D	Maria Sánchez	672345678		Manacor	Mallorca	3.703790	40.416780
5	BMW	X5	7890JKL	25000	2017	70000	bo	33445566E	Jordi Pérez		jordi.perez@email.com	Ciutadella	Menorca	0.376288	39.469907
6	Audi	A4	1122XYZ	20000	2016	80000	molt bo	44556677F	Laura Martín	687654321	laura.martin@email.com	Maó	Menorca		

Tienes que seguir los siguientes pasos:

1. Crea la tabla vehículos en HBase.

```
create 'vehicles', 'info_general', 'conservacio', 'propietari', 'ubicacio'
```



2. Inserta mediante órdenes put las tres primeras filas de la tabla.

```
put 'vehicles', '1', 'info_general:marca', 'Toyota'
put 'vehicles', '1', 'info_general:model', 'Corolla'
put 'vehicles', '1', 'info_general:matricula', '1234ABC'
put 'vehicles', '1', 'info_general:preu', '12000'
put 'vehicles', '1', 'conservacio:any', '2015'
put 'vehicles', '1', 'conservacio:km', '120000'
put 'vehicles', '1', 'conservacio:estat', 'bo'
put 'vehicles', '1', 'propietari:nif', '12345678A'
put 'vehicles', '1', 'propietari:nom', 'Joan Mart nez'
put 'vehicles', '1', 'propietari:telefon', '623456789'
put 'vehicles', '1', 'propietari:mail', 'joan.martinez@email.com'
put 'vehicles', '1', 'ubicacio:ciutat', 'Palma'
put 'vehicles', '1', 'ubicacio:illa', 'Mallorca'
put 'vehicles', '1', 'ubicacio:longitud', '2.650160'
put 'vehicles', '1', 'ubicacio:latitud', '39.569600'

put 'vehicles', '2', 'info_general:marca', 'Ford'
put 'vehicles', '2', 'info_general:model', 'Focus'
put 'vehicles', '2', 'info_general:matricula', '5678DEF'
put 'vehicles', '2', 'info_general:preu', '15500'
put 'vehicles', '2', 'conservacio:any', '2018'
put 'vehicles', '2', 'conservacio:km', '55000'
put 'vehicles', '2', 'conservacio:estat', 'molt bo'
put 'vehicles', '2', 'propietari:nif', '87654321B'
put 'vehicles', '2', 'propietari:nom', 'Ana Garc a'
put 'vehicles', '2', 'propietari:telefon', ''
put 'vehicles', '2', 'propietari:mail', 'ana.garcia@email.com'
put 'vehicles', '2', 'ubicacio:ciutat', 'Palma'
```

```
put 'vehicles', '2', 'ubicacio:illa', 'Mallorca'

put 'vehicles', '3', 'info_general:marca', 'Honda'
put 'vehicles', '3', 'info_general:model', 'Civic'
put 'vehicles', '3', 'info_general:matricula', '9012GHI'
put 'vehicles', '3', 'info_general:preu', '18000'
put 'vehicles', '3', 'conservacio:any', '2020'
put 'vehicles', '3', 'conservacio:km', '20000'
put 'vehicles', '3', 'conservacio:estat', 'bo'
put 'vehicles', '3', 'propietari:nif', '11223344C'
put 'vehicles', '3', 'propietari:nom', 'David L pez'
put 'vehicles', '3', 'propietari:telefon', '655123456'
put 'vehicles', '3', 'ubicacio:ciutat', 'Eivissa'
put 'vehicles', '3', 'ubicacio:illa', 'Eivissa'
put 'vehicles', '3', 'ubicacio:longitud', '1.452850'
put 'vehicles', '3', 'ubicacio:latitud', '38.916280'

# test
scan 'vehicles'
```

```
hbase(main):060:0> put 'vehicles', '1', 'info_general:model', 'Corolla'
0 row(s) in 0.0440 seconds

hbase(main):061:0> put 'vehicles', '1', 'info_general:matricula', '1234ABC'
0 row(s) in 0.1220 seconds

hbase(main):062:0> put 'vehicles', '1', 'info_general:preu', '12000'
0 row(s) in 0.2370 seconds

hbase(main):063:0> put 'vehicles', '1', 'conservacio:any', '2015'
0 row(s) in 0.0800 seconds

hbase(main):064:0> put 'vehicles', '1', 'conservacio:km', '120000'
0 row(s) in 0.0160 seconds

hbase(main):065:0> put 'vehicles', '1', 'conservacio:estat', 'bo'
0 row(s) in 0.0140 seconds

hbase(main):066:0> put 'vehicles', '1', 'propietari:nif', '12345678A'
0 row(s) in 0.0050 seconds

hbase(main):067:0> put 'vehicles', '1', 'propietari:nom', 'Joan Mart?nez'
0 row(s) in 0.0040 seconds

hbase(main):068:0> put 'vehicles', '1', 'propietari:telefon', '623456789'
0 row(s) in 0.0040 seconds

hbase(main):069:0> put 'vehicles', '1', 'propietari:mail', 'joan.martinez@email.com'
0 row(s) in 0.0110 seconds

hbase(main):070:0> put 'vehicles', '1', 'ubicacio:ciutat', 'Palma'
0 row(s) in 0.0480 seconds

hbase(main):071:0> put 'vehicles', '1', 'ubicacio:illa', 'Mallorca'
0 row(s) in 0.0040 seconds

hbase(main):072:0> put 'vehicles', '1', 'ubicacio:longitud', '2.650160'
0 row(s) in 0.0050 seconds

hbase(main):073:0> put 'vehicles', '1', 'ubicacio:latitud', '39.569600'
0 row(s) in 0.0080 seconds

hbase(main):074:0>
hbase(main):075:0* put 'vehicles', '2', 'info_general:marca', 'Ford'
0 row(s) in 0.1280 seconds

hbase(main):076:0> put 'vehicles', '2', 'info_general:model', 'Focus'
0 row(s) in 0.0070 seconds

hbase(main):077:0> put 'vehicles', '2', 'info_general:matricula', '5678DEF'
0 row(s) in 0.0050 seconds

hbase(main):078:0> put 'vehicles', '2', 'info_general:preu', '15500'
0 row(s) in 0.0040 seconds

hbase(main):079:0> put 'vehicles', '2', 'conservacio:any', '2018'
0 row(s) in 0.0040 seconds
```

3. Haz un archivo CSV con las 3 últimas filas y cárgalo en la tabla.

```
hbase(main):103:0> scan 'vehicles'
ROW                                    COLUMN+CELL
 1                                     column=conservacio:any, timestamp=1736272409702, value=2015
 1                                     column=conservacio:estat, timestamp=1736272409811, value=bo
 1                                     column=conservacio:km, timestamp=1736272409784, value=120000
 1                                     column=info_general:marca, timestamp=1736272408894, value=Toyota
 1                                     column=info_general:matricula, timestamp=1736272409273, value=1234ABC
 1                                     column=info_general:model, timestamp=1736272408980, value=Corolla
 1                                     column=info_general:preu, timestamp=1736272409492, value=12000
 1                                     column=propietari:mail, timestamp=1736272409903, value=joan.martinez@email.com
 1                                     column=propietari:nif, timestamp=1736272409826, value=12345678A
 1                                     column=propietari:nom, timestamp=1736272409867, value=Joan Mart\xC3\xADnez
 1                                     column=propietari:telefon, timestamp=1736272409882, value=623456789
 1                                     column=ubicacio:ciutat, timestamp=1736272409960, value=Palma
 1                                     column=ubicacio:illa, timestamp=1736272409978, value=Mallorca
 1                                     column=ubicacio:latitud, timestamp=1736272410016, value=39.569600
 1                                     column=ubicacio:longitud, timestamp=1736272409997, value=2.650160
 2                                     column=conservacio:any, timestamp=1736272410282, value=2018
 2                                     column=conservacio:estat, timestamp=1736272410328, value=molt bo
 2                                     column=conservacio:km, timestamp=1736272410314, value=55000
 2                                     column=info_general:marca, timestamp=1736272410159, value=Ford
 2                                     column=info_general:matricula, timestamp=1736272410252, value=5678DEF
 2                                     column=info_general:model, timestamp=1736272410178, value=Focus
 2                                     column=info_general:preu, timestamp=1736272410267, value=15500
 2                                     column=propietari:mail, timestamp=1736272410441, value=ana.garcia@email.com
 2                                     column=propietari:nif, timestamp=1736272410341, value=87654321B
 2                                     column=propietari:nom, timestamp=1736272410402, value=Ana Garc\xC3\xADa
 2                                     column=propietari:telefon, timestamp=1736272410416, value=
 2                                     column=ubicacio:ciutat, timestamp=1736272410464, value=Palma
 2                                     column=ubicacio:illa, timestamp=1736272410478, value=Mallorca
 3                                     column=conservacio:any, timestamp=1736272410586, value=2020
 3                                     column=conservacio:estat, timestamp=1736272410643, value=bo
 3                                     column=conservacio:km, timestamp=1736272410628, value=20000
 3                                     column=info_general:marca, timestamp=1736272410527, value=Honda
 3                                     column=info_general:matricula, timestamp=1736272410554, value=9012GHI
 3                                     column=info_general:model, timestamp=1736272410541, value=Civic
 3                                     column=info_general:preu, timestamp=1736272410568, value=18000
 3                                     column=propietari:nif, timestamp=1736272410658, value=11223344C
 3                                     column=propietari:nom, timestamp=1736272410671, value=David L\xC3\xB3pez
 3                                     column=propietari:telefon, timestamp=1736272410685, value=655123456
 3                                     column=ubicacio:ciutat, timestamp=1736272410706, value=Eivissa
 3                                     column=ubicacio:illa, timestamp=1736272410721, value=Eivissa
 3                                     column=ubicacio:latitud, timestamp=1736272413333, value=38.916280
 3                                     column=ubicacio:longitud, timestamp=1736272410735, value=1.452850
3 row(s) in 0.0820 seconds

hbase(main):104:0>
```

```
hdfs dfs -mkdir vehicles
hdfs dfs -put vehicles.csv  vehicles/
cat vehicles.csv # Just to test if needed
hdfs dfs -ls /user/cloudera/vehicles # Just to test if needed
hbase org.apache.hadoop.hbase.mapreduce.ImportTsv \
-Dimporttsv.columns="HBASE_ROW_KEY,info_general:marca,info_general:model,
info_general:matricula,info_general:preu,conservacio:any,conservacio:km,
conservacio:estat,propietari:nif,propietari:nom,propietari:telefon,propietari:mail,
ubicacio:ciutat,ubicacio:illa,ubicacio:longitud,ubicacio:latitud" \
-Dimporttsv.separator=',' vehicles /user/cloudera/vehicles/vehicles.csv
```

```
# test
scan 'vehicles'
```

```
[cloudera@quickstart hbase]$ hdfs dfs -mkdir vehicles
[cloudera@quickstart hbase]$ hdfs dfs -put vehicles.csv  vehicles/
[cloudera@quickstart hbase]$ cat vehicles.csv
:key,info_general:marca,info_general:model,info_general:matricula,info_general:preu,conservacio:any,conservacio:km,conser
vacio:estat,propietari:nif,propietari:nom,propietari:telefon,propietari:mail,ubicacio:ciutat,ubicacio:illa,ubicacio:longi
tud,ubicacio:latitud
4,Mercedes-Benz,A-Class,3456JKL,30000,2019,40000,bo,22334455D,Maria Sánchez,672345678,,Manacor,Mallorca,3.703790,40.41678
0
5,BMW,X5,7890JKL,25000,2017,70000,bo,33445566E,Jordi Pérez,,jordi.perez@email.com,Ciutadella,Menorca,0.376288,39.469907
6,Audi,A4,1122XYZ,20000,2016,80000,molt bo,44556677F,Laura Martín,687654321,laura.martin@email.com,Maó,Menorca
[cloudera@quickstart hbase]$ hdfs dfs -ls /cloudera/vehicles
ls: `/cloudera/vehicles': No such file or directory
[cloudera@quickstart hbase]$ hdfs dfs -ls /user/cloudera/vehicles
Found 1 items
-rw-r--r--   1 cloudera cloudera        621 2025-01-07 09:59 /user/cloudera/vehicles/vehicles.csv
[cloudera@quickstart hbase]$ hbase org.apache.hadoop.hbase.mapreduce.ImportTsv \
> -Dimporttsv.columns=HBASE_ROW_KEY,info_general:marca,info_general:model,info_general:matricula,info_general:preu,conser
vacio:any,conservacio:km,conservacio:estat,propietari:nif,propietari:nom,propietari:telefon,propietari:mail,ubicacio:ciut
at,ubicacio:illa,ubicacio:longitud,ubicacio:latitud \
> -Dimporttsv.separator=',' vehicles /user/cloudera/vehicles/vehicles.csv
```

```
3                           column=ubicacio:latitud, timestamp=1736274013182, value=38.916280
3                           column=ubicacio:longitud, timestamp=1736274012537, value=1.452850
4                           column=conservacio:any, timestamp=1736274149148, value=2019
4                           column=conservacio:estat, timestamp=1736274149148, value=bo
4                           column=conservacio:km, timestamp=1736274149148, value=40000
4                           column=info_general:marca, timestamp=1736274149148, value=Mercedes-Benz
4                           column=info_general:matricula, timestamp=1736274149148, value=3456JKL
4                           column=info_general:model, timestamp=1736274149148, value=A-Class
4                           column=info_general:preu, timestamp=1736274149148, value=30000
4                           column=propietari:mail, timestamp=1736274149148, value=
4                           column=propietari:nif, timestamp=1736274149148, value=22334455D
4                           column=propietari:nom, timestamp=1736274149148, value=Maria S\xC3\xA1nchez
4                           column=propietari:telefon, timestamp=1736274149148, value=672345678
4                           column=ubicacio:ciutat, timestamp=1736274149148, value=Manacor
4                           column=ubicacio:illa, timestamp=1736274149148, value=Mallorca
4                           column=ubicacio:latitud, timestamp=1736274149148, value=40.416780
4                           column=ubicacio:longitud, timestamp=1736274149148, value=3.703790
5                           column=conservacio:any, timestamp=1736274149148, value=2017
5                           column=conservacio:estat, timestamp=1736274149148, value=bo
5                           column=conservacio:km, timestamp=1736274149148, value=70000
5                           column=info_general:marca, timestamp=1736274149148, value=BMW
5                           column=info_general:matricula, timestamp=1736274149148, value=7890JKL
5                           column=info_general:model, timestamp=1736274149148, value=X5
5                           column=info_general:preu, timestamp=1736274149148, value=25000
5                           column=propietari:mail, timestamp=1736274149148, value=jordi.perez@email.com
5                           column=propietari:nif, timestamp=1736274149148, value=33445566E
5                           column=propietari:nom, timestamp=1736274149148, value=Jordi P\xC3\xA9rez
5                           column=propietari:telefon, timestamp=1736274149148, value=
5                           column=ubicacio:ciutat, timestamp=1736274149148, value=Ciutadella
5                           column=ubicacio:illa, timestamp=1736274149148, value=Menorca
5                           column=ubicacio:latitud, timestamp=1736274149148, value=39.469907
5                           column=ubicacio:longitud, timestamp=1736274149148, value=0.376288
6                           column=conservacio:any, timestamp=1736274149148, value=2016
6                           column=conservacio:estat, timestamp=1736274149148, value=molt bo
6                           column=conservacio:km, timestamp=1736274149148, value=80000
6                           column=info_general:marca, timestamp=1736274149148, value=Audi
6                           column=info_general:matricula, timestamp=1736274149148, value=1122XYZ
6                           column=info_general:model, timestamp=1736274149148, value=A4
6                           column=info_general:preu, timestamp=1736274149148, value=20000
6                           column=propietari:mail, timestamp=1736274149148, value=laura.martin@email.com
6                           column=propietari:nif, timestamp=1736274149148, value=44556677F
6                           column=propietari:nom, timestamp=1736274149148, value=Laura Mart\xC3\xADn
6                           column=propietari:telefon, timestamp=1736274149148, value=687654321
6                           column=ubicacio:ciutat, timestamp=1736274149148, value=Ma\xC3\xB3
6                           column=ubicacio:illa, timestamp=1736274149148, value=Menorca
6 row(s) in 0.3120 seconds
```

4. Muestra todos los datos de la fila 4 con un `get`.

```
get 'vehicles', '4'
```

```
hbase(main):002:0> get 'vehicles', '4'
COLUMN                          CELL
 conservacio:any                timestamp=1736274149148, value=2019
 conservacio:estat              timestamp=1736274149148, value=bo
 conservacio:km                 timestamp=1736274149148, value=40000
 info_general:marca             timestamp=1736274149148, value=Mercedes-Benz
 info_general:matricula         timestamp=1736274149148, value=3456JKL
 info_general:model             timestamp=1736274149148, value=A-Class
 info_general:preu              timestamp=1736274149148, value=30000
 propietari:mail                timestamp=1736274149148, value=
 propietari:nif                 timestamp=1736274149148, value=22334455D
 propietari:nom                 timestamp=1736274149148, value=Maria S\xC3\xA1nchez
 propietari:telefon             timestamp=1736274149148, value=672345678
 ubicacio:ciutat                timestamp=1736274149148, value=Manacor
 ubicacio:illa                  timestamp=1736274149148, value=Mallorca
 ubicacio:latitud               timestamp=1736274149148, value=40.416780
 ubicacio:longitud              timestamp=1736274149148, value=3.703790
15 row(s) in 0.0210 seconds
```

5. Muestra los datos de información general de la tabla con un `scan`.

```
scan 'vehicles', {COLUMNS => 'info_general'}
```

```
hbase(main):003:0> scan 'vehicles', {COLUMNS => 'info_general'}
ROW                                    COLUMN+CELL
 1                                     column=info_general:marca, timestamp=1736274011838, value=Toyota
 1                                     column=info_general:matricula, timestamp=1736274011913, value=1234ABC
 1                                     column=info_general:model, timestamp=1736274011878, value=Corolla
 1                                     column=info_general:preu, timestamp=1736274011945, value=12000
 2                                     column=info_general:marca, timestamp=1736274012189, value=Ford
 2                                     column=info_general:matricula, timestamp=1736274012218, value=5678DEF
 2                                     column=info_general:model, timestamp=1736274012204, value=Focus
 2                                     column=info_general:preu, timestamp=1736274012231, value=15500
 3                                     column=info_general:marca, timestamp=1736274012378, value=Honda
 3                                     column=info_general:matricula, timestamp=1736274012409, value=9012GHI
 3                                     column=info_general:model, timestamp=1736274012395, value=Civic
 3                                     column=info_general:preu, timestamp=1736274012421, value=18000
 4                                     column=info_general:marca, timestamp=1736274149148, value=Mercedes-Benz
 4                                     column=info_general:matricula, timestamp=1736274149148, value=3456JKL
 4                                     column=info_general:model, timestamp=1736274149148, value=A-Class
 4                                     column=info_general:preu, timestamp=1736274149148, value=30000
 5                                     column=info_general:marca, timestamp=1736274149148, value=BMW
 5                                     column=info_general:matricula, timestamp=1736274149148, value=7890JKL
 5                                     column=info_general:model, timestamp=1736274149148, value=X5
 5                                     column=info_general:preu, timestamp=1736274149148, value=25000
 6                                     column=info_general:marca, timestamp=1736274149148, value=Audi
 6                                     column=info_general:matricula, timestamp=1736274149148, value=1122XYZ
 6                                     column=info_general:model, timestamp=1736274149148, value=A4
 6                                     column=info_general:preu, timestamp=1736274149148, value=20000
6 row(s) in 0.0250 seconds
```

6. Muestra todos los datos de los vehículos con estado *bo*.

```
scan 'vehicles', {FILTER => "SingleColumnValueFilter('conservacio',
'estat', =, 'binary:bo')"}
```

```
hbase(main):004:0> scan 'vehicles', {FILTER => "SingleColumnValueFilter('conservacio', 'estat', =, 'binary:bo')"}
ROW                                    COLUMN+CELL
 1                                     column=conservacio:any, timestamp=1736274011963, value=2015
 1                                     column=conservacio:estat, timestamp=1736274011994, value=bo
 1                                     column=conservacio:km, timestamp=1736274011979, value=120000
 1                                     column=info_general:marca, timestamp=1736274011838, value=Toyota
 1                                     column=info_general:matricula, timestamp=1736274011913, value=1234ABC
 1                                     column=info_general:model, timestamp=1736274011878, value=Corolla
 1                                     column=info_general:preu, timestamp=1736274011945, value=12000
 1                                     column=propietari:mail, timestamp=1736274012059, value=joan.martinez@email.com
 1                                     column=propietari:nif, timestamp=1736274012009, value=12345678A
 1                                     column=propietari:nom, timestamp=1736274012025, value=Joan Mart\xC3\xADnez
 1                                     column=propietari:telefon, timestamp=1736274012041, value=623456789
 1                                     column=ubicacio:ciutat, timestamp=1736274012084, value=Palma
 1                                     column=ubicacio:illa, timestamp=1736274012115, value=Mallorca
 1                                     column=ubicacio:latitud, timestamp=1736274012169, value=39.569600
 1                                     column=ubicacio:longitud, timestamp=1736274012134, value=2.650160
 3                                     column=conservacio:any, timestamp=1736274012434, value=2020
 3                                     column=conservacio:estat, timestamp=1736274012460, value=bo
 3                                     column=conservacio:km, timestamp=1736274012447, value=20000
 3                                     column=info_general:marca, timestamp=1736274012378, value=Honda
 3                                     column=info_general:matricula, timestamp=1736274012409, value=9012GHI
 3                                     column=info_general:model, timestamp=1736274012395, value=Civic
 3                                     column=info_general:preu, timestamp=1736274012421, value=18000
 3                                     column=propietari:nif, timestamp=1736274012473, value=11223344C
 3                                     column=propietari:nom, timestamp=1736274012487, value=David L\xC3\xB3pez
 3                                     column=propietari:telefon, timestamp=1736274012500, value=655123456
 3                                     column=ubicacio:ciutat, timestamp=1736274012513, value=Eivissa
 3                                     column=ubicacio:illa, timestamp=1736274012525, value=Eivissa
 3                                     column=ubicacio:latitud, timestamp=1736274013182, value=38.916280
 3                                     column=ubicacio:longitud, timestamp=1736274012537, value=1.452850
 4                                     column=conservacio:any, timestamp=1736274149148, value=2019
 4                                     column=conservacio:estat, timestamp=1736274149148, value=bo
 4                                     column=conservacio:km, timestamp=1736274149148, value=40000
 4                                     column=info_general:marca, timestamp=1736274149148, value=Mercedes-Benz
 4                                     column=info_general:matricula, timestamp=1736274149148, value=3456JKL
 4                                     column=info_general:model, timestamp=1736274149148, value=A-Class
 4                                     column=info_general:preu, timestamp=1736274149148, value=30000
 4                                     column=propietari:mail, timestamp=1736274149148, value=
 4                                     column=propietari:nif, timestamp=1736274149148, value=22334455D
 4                                     column=propietari:nom, timestamp=1736274149148, value=Maria S\xC3\xA1nchez
 4                                     column=propietari:telefon, timestamp=1736274149148, value=672345678
 4                                     column=ubicacio:ciutat, timestamp=1736274149148, value=Manacor
 4                                     column=ubicacio:illa, timestamp=1736274149148, value=Mallorca
 4                                     column=ubicacio:latitud, timestamp=1736274149148, value=40.416780
 4                                     column=ubicacio:longitud, timestamp=1736274149148, value=3.703790
 5                                     column=conservacio:any, timestamp=1736274149148, value=2017
 5                                     column=conservacio:estat, timestamp=1736274149148, value=bo
 5                                     column=conservacio:km, timestamp=1736274149148, value=70000
 5                                     column=info_general:marca, timestamp=1736274149148, value=BMW
 5                                     column=info_general:matricula, timestamp=1736274149148, value=7890JKL
 5                                     column=info_general:model, timestamp=1736274149148, value=X5
 5                                     column=info_general:preu, timestamp=1736274149148, value=25000
 5                                     column=propietari:mail, timestamp=1736274149148, value=jordi.perez@email.com
 5                                     column=propietari:nif, timestamp=1736274149148, value=33445566E
 5                                     column=propietari:nom, timestamp=1736274149148, value=Jordi P\xC3\xA9rez
 5                                     column=propietari:telefon, timestamp=1736274149148, value=
 5                                     column=ubicacio:ciutat, timestamp=1736274149148, value=Ciutadella
 5                                     column=ubicacio:illa, timestamp=1736274149148, value=Menorca
 5                                     column=ubicacio:latitud, timestamp=1736274149148, value=39.469907
 5                                     column=ubicacio:longitud, timestamp=1736274149148, value=0.376288
4 row(s) in 0.0970 seconds
```

7. Muestra los datos de los propietarios de los vehículos de Mallorca.

```
# Selected manually due to version incompatibilities
scan 'vehicles', {FILTER => "RowFilter(=,'binary:1') OR RowFilter(=,'binary:2')
OR RowFilter(=,'binary:4')", COLUMNS => ['propietari']}
```

```
hbase(main):010:0> scan 'vehicles', {FILTER => "RowFilter(=,'binary:1') OR RowFilter(=,'binary:2') OR RowFilter(=,'binary
:4')", COLUMNS => ['propietari']}
ROW                          COLUMN+CELL
 1                           column=propietari:mail, timestamp=1736274012059, value=joan.martinez@email.com
 1                           column=propietari:nif, timestamp=1736274012009, value=12345678A
 1                           column=propietari:nom, timestamp=1736274012025, value=Joan Mart\xC3\xADnez
 1                           column=propietari:telefon, timestamp=1736274012041, value=623456789
 2                           column=propietari:mail, timestamp=1736274012331, value=ana.garcia@email.com
 2                           column=propietari:nif, timestamp=1736274012287, value=87654321B
 2                           column=propietari:nom, timestamp=1736274012302, value=Ana Garc\xC3\xADa
 2                           column=propietari:telefon, timestamp=1736274012317, value=
 4                           column=propietari:mail, timestamp=1736274149148, value=
 4                           column=propietari:nif, timestamp=1736274149148, value=22334455D
 4                           column=propietari:nom, timestamp=1736274149148, value=Maria S\xC3\xA1nchez
 4                           column=propietari:telefon, timestamp=1736274149148, value=672345678
3 row(s) in 0.0730 seconds
```

8. Muestra la información general de los vehículos que sean posteriores a 2017 (incluido) con menos de 50000 km.

```
# Selected manually due to version incompatibilities
scan 'vehicles', {FILTER => "RowFilter(=,'binary:3') OR RowFilter(=,'binary:4')",
COLUMNS => ['info_general']}
```

```
hbase(main):008:0> scan 'vehicles', {FILTER => "RowFilter(=,'binary:3') OR RowFilter(=,'binary:4')", COLUMNS => ['info_ge
neral']}
ROW                          COLUMN+CELL
 3                           column=info_general:marca, timestamp=1736274012378, value=Honda
 3                           column=info_general:matricula, timestamp=1736274012409, value=9012GHI
 3                           column=info_general:model, timestamp=1736274012395, value=Civic
 3                           column=info_general:preu, timestamp=1736274012421, value=18000
 4                           column=info_general:marca, timestamp=1736274149148, value=Mercedes-Benz
 4                           column=info_general:matricula, timestamp=1736274149148, value=3456JKL
 4                           column=info_general:model, timestamp=1736274149148, value=A-Class
 4                           column=info_general:preu, timestamp=1736274149148, value=30000
2 row(s) in 0.0210 seconds
```

9. Modifica la fila 2 y ponle el teléfono 656123321.

```
put 'vehicles', '2', 'propietari:telefon', '656123321'
get 'vehicles', '2' # Check
```

```
hbase(main):002:0> put 'vehicles', '2', 'propietari:telefon', '656123321'
0 row(s) in 0.1230 seconds

hbase(main):003:0> get 'vehicles', '2'
COLUMN                       CELL
 conservacio:any             timestamp=1736274012245, value=2018
 conservacio:estat           timestamp=1736274012271, value=molt bo
 conservacio:km              timestamp=1736274012258, value=55000
 info_general:marca          timestamp=1736274012189, value=Ford
 info_general:matricula      timestamp=1736274012218, value=5678DEF
 info_general:model          timestamp=1736274012204, value=Focus
 info_general:preu           timestamp=1736274012231, value=15500
 propietari:mail             timestamp=1736274012331, value=ana.garcia@email.com
 propietari:nif              timestamp=1736274012287, value=87654321B
 propietari:nom              timestamp=1736274012302, value=Ana Garc\xC3\xADa
 propietari:telefon          timestamp=1736276748645, value=656123321
 ubicacio:ciutat             timestamp=1736274012344, value=Palma
 ubicacio:illa               timestamp=1736274012360, value=Mallorca
13 row(s) in 0.0230 seconds
```

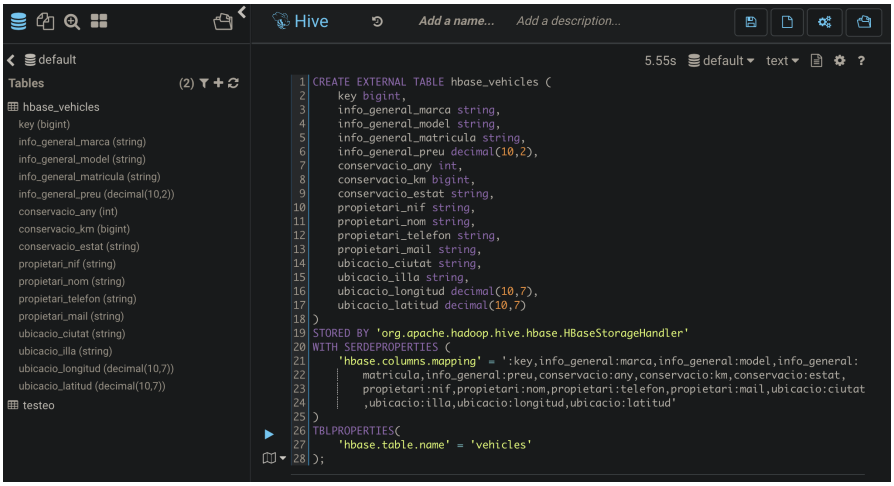
10. Borra todos los datos en la fila 5.

```
deleteall 'vehicles', '5'
scan 'vehicles' # Check
```

```
4                                   column=info_general:model, timestamp=1736274149148, value=A-Class
4                                   column=info_general:preu, timestamp=1736274149148, value=30000
4                                   column=propietari:mail, timestamp=1736274149148, value=
4                                   column=propietari:nif, timestamp=1736274149148, value=22334455D
4                                   column=propietari:nom, timestamp=1736274149148, value=Maria S\xC3\x
4                                   column=propietari:telefon, timestamp=1736274149148, value=672345678
4                                   column=ubicacio:ciutat, timestamp=1736274149148, value=Manacor
4                                   column=ubicacio:illa, timestamp=1736274149148, value=Mallorca
4                                   column=ubicacio:latitud, timestamp=1736274149148, value=40.416780
4                                   column=ubicacio:longitud, timestamp=1736274149148, value=3.703790
6                                   column=conservacio:any, timestamp=1736274149148, value=2016
6                                   column=conservacio:estat, timestamp=1736274149148, value=molt bo
6                                   column=conservacio:km, timestamp=1736274149148, value=80000
6                                   column=info_general:marca, timestamp=1736274149148, value=Audi
6                                   column=info_general:matricula, timestamp=1736274149148, value=1122X
6                                   column=info_general:model, timestamp=1736274149148, value=A4
6                                   column=info_general:preu, timestamp=1736274149148, value=20000
6                                   column=propietari:mail, timestamp=1736274149148, value=laura.martin
6                                   column=propietari:nif, timestamp=1736274149148, value=44556677F
6                                   column=propietari:nom, timestamp=1736274149148, value=Laura Mart\xC
6                                   column=propietari:telefon, timestamp=1736274149148, value=687654321
6                                   column=ubicacio:ciutat, timestamp=1736274149148, value=Ma\xC3\xB3
```

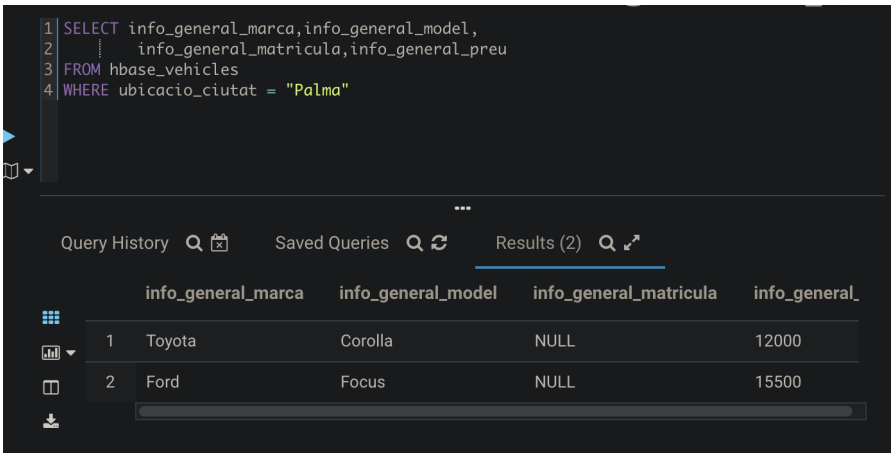
11. Crea una tabla externa en Hive vinculada a la tabla vehículos de HBase.

```sql
CREATE EXTERNAL TABLE hbase_vehicles (
    key bigint,
    info_general_marca string,
    info_general_model string,
    info_general_matricula string,
    info_general_preu decimal(10,2),
    conservacio_any int,
    conservacio_km bigint,
    conservacio_estat string,
    propietari_nif string,
    propietari_nom string,
    propietari_telefon string,
    propietari_mail string,
    ubicacio_ciutat string,
    ubicacio_illa string,
    ubicacio_longitud decimal(10,7),
    ubicacio_latitud decimal(10,7)
)
STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
WITH SERDEPROPERTIES (
    'hbase.columns.mapping' = ':key,info_general:marca,info_general:model,info_general:
        matricula,info_general:preu,conservacio:any,conservacio:km,conservacio:estat,
        propietari:nif,propietari:nom,propietari:telefon,propietari:mail,ubicacio:ciutat
        ,ubicacio:illa,ubicacio:longitud,ubicacio:latitud'
)
TBLPROPERTIES(
    'hbase.table.name' = 'vehicles'
);
```



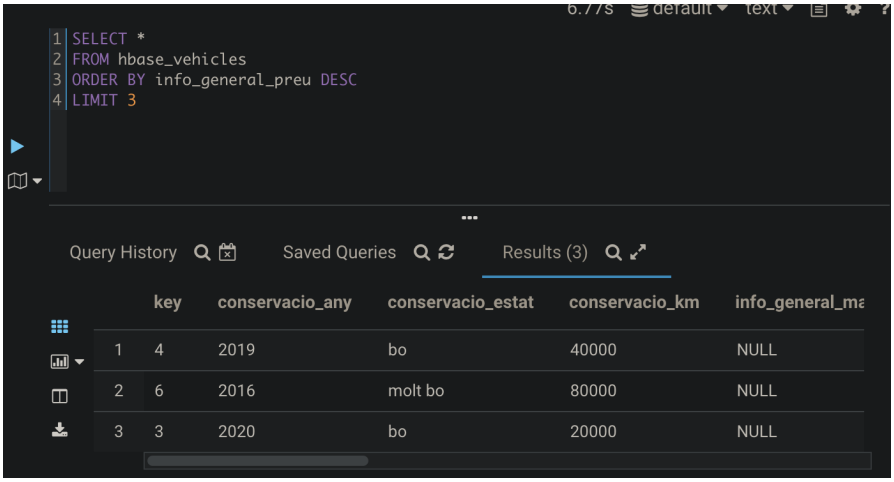
12. Haz una consulta en HiveQL que devuelva los datos de información general de los vehículos de Palma.

```
SELECT info_general_marca,info_general_model,
        info_general_matricula,info_general_preu
FROM hbase_vehicles
WHERE ubicacio_ciutat = "Palma";
```



13. Haz una consulta en Impala que devuelva todos los datos de los 3 vehículos más caros.

```
INVALIDATE METADATA;
SELECT *
FROM hbase_vehicles
ORDER BY info_general_preu DESC
LIMIT 3;
```



14. Haz una consulta en HiveQL que agrupe a los vehículos por estado y muestre la media de km de cada estado.

```
SELECT conservacio_estat, avg(conservacio_km) AS avg_km
FROM hbase_vehicles
GROUP BY conservacio_estat
ORDER BY avg_km DESC;
```

15. Haz una consulta en Impala que agrupe a los vehículos por isla y muestre el número de vehículos de cada isla.

```
INVALIDATE METADATA;
SELECT COUNT(*) AS coches, ubicacio_illa
FROM hbase_vehicles
GROUP BY ubicacio_illa;
```

SELECT COUNT(*) AS coches, ubicacio_illa
FROM hbase_vehicles
GROUP BY ubicacio_illa
Query History
Saved Queries
Results (3)
coches
ubicacio_illa
1 1 Menorca
2 3 Mallorca
3 1 Eivissa