



**Instituto Infnet**

MIT em Engenharia de Dados: BigData  
Trabalho Da Disciplina Infraestrutura Hadoop

Aluno: Charles de Araujo Melo

Atividade da disciplina Infraestrutura Hadoop  
Prof.: Andre Ormastroni Victor do curso  
MIT em Engenharia de Dados

## **1. Introdução**

Hadoop destaca-se como uma tecnologia aberta, baseada no paradigma MapReduce, que utiliza a computação paralela e distribuída para resolver o problema da escalabilidade no processamento de BigData, com garantias de tolerância a falhas. Das vantagens em adotá-lo, está o fato de se utilizar aglomerados de máquinas convencionais, tornando-o eficaz como solução de baixo custo de implantação. Ademais, com ele as empresas podem conquistar uma grande vantagem competitiva, dispondo de um mecanismo que possibilita avaliar grandes quantidades de dados em busca de novas informações.

Através dos procedimentos das praticas laboratoriais se utilizando como ferramenta de trabalho um dos serviços da nuvem da google, google cloud, serão discutidos e elaborados um estudo prático utilizando a ferramenta Hive para responder um conjunto de perguntas elaboradas na prática laboratorial. As Bases de dados utilizadas no trabalho são: demogra\_cliente, end\_cliente e transacoes.

## **2. Objetivos**

Objetivo é por em prática no dia a dia, planejando e implementando projetos que envolvam a infraestrutura Hadoop e como utilizá-la na solução de problemas que envolvam grande quantidades de dados.

## **3. Descrição**

Minha base de dados é de uma loja de venda de bicicletas onde utilizarei para identificar vantagens empresariais para futuro aumento de vendas e conhecer melhor as características dos clientes para maiores captações.

# 1. Visão geral do esquema de dados (caracterização das tabelas, dicionário de dados, etc.)

describe demogra\_cliente;

```
+-----+
| tab_name |
+-----+
| demogra_cliente |
| end_cliente |
| transacoes |
+-----+
3 rows selected (0.203 seconds)
0: jdbc:hive2://localhost:10000/default> describe demogra_cliente;
INFO : Compiling command(queryId=hive_20230826152853_6ccca393-862f-418d-a07c-86201e993516): describe demogra_cliente
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Semantic Analysis Completed (retail = false)
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:col_name, type:string, comment:from deserializer), FieldSchema(name:data_type, type:string, comment:from deserializer), FieldSchema(name:comment, type:string, comment:from deserializer)], properties:null)
INFO : Completed compiling command(queryId=hive_20230826152853_6ccca393-862f-418d-a07c-86201e993516): Time taken: 0.034 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20230826152853_6ccca393-862f-418d-a07c-86201e993516): describe demogra_cliente
INFO : Starting task [Stage-0:DML] in serial mode
INFO : Completed executing command(queryId=hive_20230826152853_6ccca393-862f-418d-a07c-86201e993516): Time taken: 0.017 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager
+-----+-----+-----+
| col_name | data type | comment |
+-----+-----+-----+
| customer_id | int | |
| nome | string | |
| sobrenome | string | |
| sexo | string | |
| compra_bike_ultimo_3_ano | int | |
| ano | int | |
| mes | string | |
| dia | int | |
| cargo | string | |
| categoria_cargo | string | |
| segmento | string | |
| possui_carro | string | |
| falecido | string | |
+-----+-----+-----+
13 rows selected (0.078 seconds)
0: jdbc:hive2://localhost:10000/default> 
```

describe end\_cliente;

```
+-----+
| tab_name |
+-----+
| demogra_cliente |
| end_cliente |
| transacoes |
+-----+
3 rows selected (0.084 seconds)
0: jdbc:hive2://localhost:10000/default> describe end_cliente;
INFO : Compiling command(queryId=hive_20230826153113_45926caf-71c1-4248-a7e1-2f2a5b06c623): describe end_cliente
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Semantic Analysis Completed (retail = false)
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:col_name, type:string, comment:from deserializer), FieldSchema(name:data_type, type:string, comment:from deserializer), FieldSchema(name:comment, type:string, comment:from deserializer)], properties:null)
INFO : Completed compiling command(queryId=hive_20230826153113_45926caf-71c1-4248-a7e1-2f2a5b06c623): Time taken: 0.058 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20230826153113_45926caf-71c1-4248-a7e1-2f2a5b06c623): describe end_cliente
INFO : Starting task [Stage-0:DML] in serial mode
INFO : Completed executing command(queryId=hive_20230826153113_45926caf-71c1-4248-a7e1-2f2a5b06c623): Time taken: 0.021 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager
+-----+-----+-----+
| col_name | data type | comment |
+-----+-----+-----+
| customer_id | int | |
| endereco | string | |
| estado | string | |
| pais | string | |
| codigo_postal | int | |
+-----+-----+-----+
5 rows selected (0.118 seconds)
0: jdbc:hive2://localhost:10000/default> 
```

describe transacoes;

```
+-----+
| tab_name |
+-----+
| demogra_cliente |
| end_cliente |
| transacoes |
+-----+
3 rows selected (0.068 seconds)
0: jdbc:hive2://localhost:10000/default> describe transacoes;
INFO : Compiling command(queryId=hive_20230826153314_00e3085d-ff73-47bd-a415-1c121d8e29fe): describe transacoes
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Semantic Analysis Completed (retail = false)
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:col_name, type:string, comment:from deserializer), FieldSchema(name:data_type, type:string, comment:from deserializer), FieldSchema(name:comment, type:string, comment:from deserializer)], properties:null)
INFO : Completed compiling command(queryId=hive_20230826153314_00e3085d-ff73-47bd-a415-1c121d8e29fe): Time taken: 0.04 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20230826153314_00e3085d-ff73-47bd-a415-1c121d8e29fe): describe transacoes
INFO : Starting task [Stage-0:DML] in serial mode
INFO : Completed executing command(queryId=hive_20230826153314_00e3085d-ff73-47bd-a415-1c121d8e29fe): Time taken: 0.014 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager
+-----+-----+-----+
| col_name | data type | comment |
+-----+-----+-----+
| transaction_id | int | |
| product_id | int | |
| customer_id | int | |
| ano | int | |
| mes | string | |
| dia | int | |
| pedido_online | string | |
| status_pedido | string | |
| marca | string | |
| linha_produto | string | |
| classe_produto | string | |
| preco_lista | decimal(10,2) | |
| custo_padrao | decimal(10,2) | |
+-----+-----+-----+
13 rows selected (0.083 seconds)
0: jdbc:hive2://localhost:10000/default> 
```

## 2. Identificação das fontes de onde os dados foram obtidos

<https://www.kaggle.com/datasets/rohithsahoo/bicycle-store-dataset>

## 3. Construção de um *database* no Hive que implemente o esquema proposto

### Conectando no Hive

create database charles\_hadoop;

beeline -u jdbc:hive2://localhost:10000/default -n charlesmeloti@cluster-charles-m -d org.apache.hive.jdbc.HiveDriver

```
Linux cluster-charles-m 5.10.0-25-cloud-amd64 #1 SMP Debian 5.10.191-1 (2023-08-16) x86_64

The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
Last login: Sat Aug 26 15:27:56 2023 from 35.235.241.18
charlesmeloti@cluster-charles-m:~$ beeline -u jdbc:hive2://localhost:10000/default -n charlesmeloti@cluster-charles-m -d org.apache.hive.jdbc.HiveDriver
Connecting to jdbc:hive2://localhost:10000/default
Connected to: Apache Hive (version 3.1.3)
Driver: Hive JDBC (version 3.1.3)
Transaction isolation: TRANSACTION_REPEATABLE_READ
Beeline version 3.1.3 by Apache Hive
0: jdbc:hive2://localhost:10000/default>
```

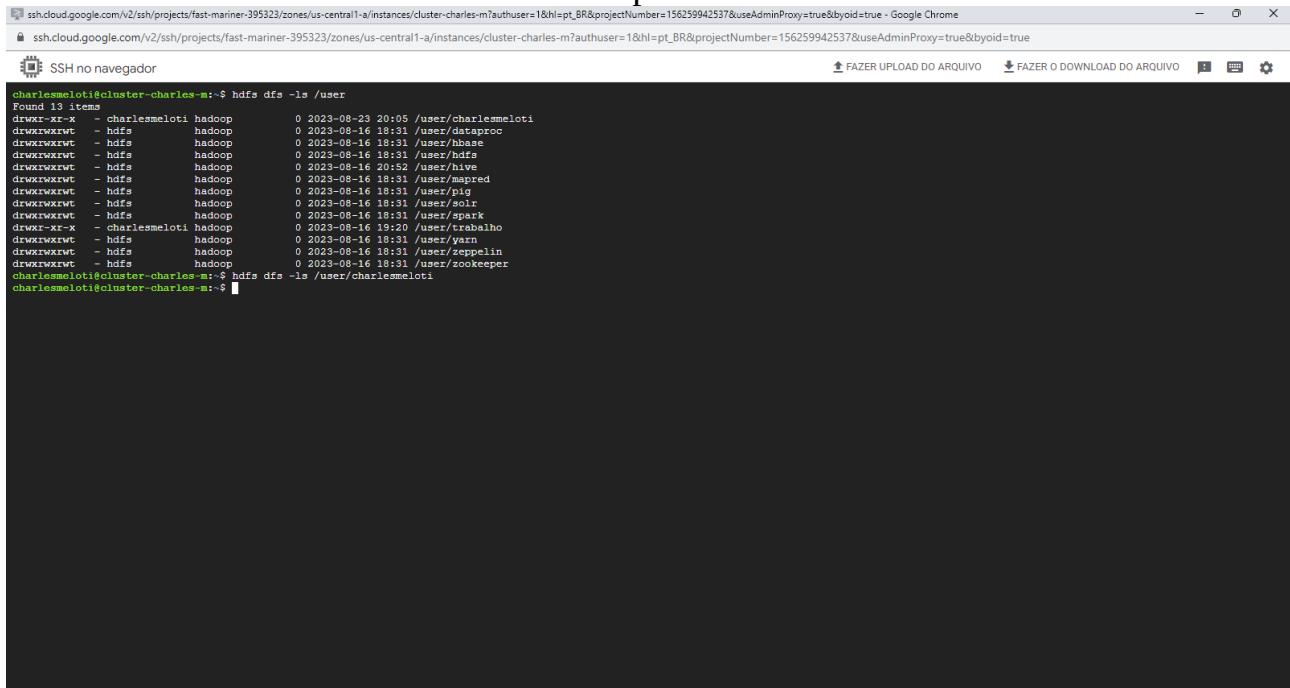
### Verificando tabelas no diretório

The screenshot shows a Google Cloud Shell terminal window. The terminal output shows the user running the `ls` command, which lists the files `demogra_cliente.csv`, `end_cliente.csv`, and `transacoes.csv`. A transfer confirmation overlay is visible in the bottom right corner, titled "3 itens transferidos", listing the same three files with green checkmarks indicating successful transfer.

```
charlesmeloti@cluster-charles-m:~$ ls
charlesmeloti@cluster-charles-m:~$ ls
demogra_cliente.csv  end_cliente.csv  transacoes.csv
charlesmeloti@cluster-charles-m:~$
```

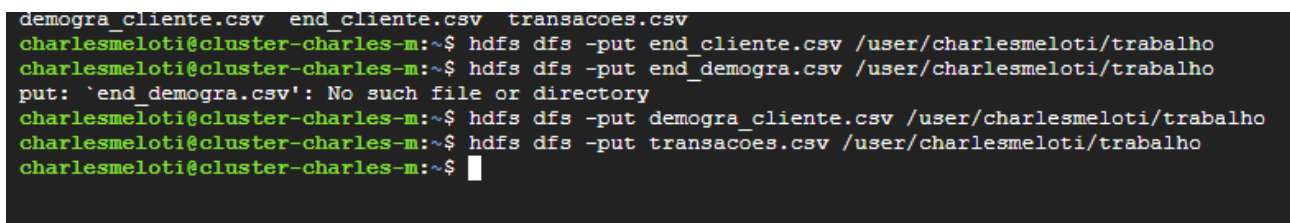
3 itens transferidos	
demogra_cliente.c...	✓
end_cliente.csv	✓
transacoes.csv	✓

## Verificando a pasta do usuário



```
ssh.cloud.google.com/v2/ssh/projects/fast-mariner-395323/zones/us-central1-a/instances/cluster-charles-m?authuser=1&hl=pt_BR&projectNumber=156259942537&useAdminProxy=true&byoid=true - Google Chrome
ssh.cloud.google.com/v2/ssh/projects/fast-mariner-395323/zones/us-central1-a/instances/cluster-charles-m?authuser=1&hl=pt_BR&projectNumber=156259942537&useAdminProxy=true&byoid=true
SSH no navegador
FAZER UPLOAD DO ARQUIVO FAZER O DOWNLOAD DO ARQUIVO
charlesmeloti@cluster-charles-m:~$ hdfs dfs -ls /user
Found 13 items
drwxr-xr-x - charlesmeloti hadoop 0 2023-08-23 20:05 /user/charlesmeloti
drwxrwxrwt - hdfs hadoop 0 2023-08-16 18:31 /user/dataproc
drwxrwxrwt - hdfs hadoop 0 2023-08-16 18:31 /user/hbase
drwxrwxrwt - hdfs hadoop 0 2023-08-16 18:31 /user/hdfs
drwxrwxrwt - hdfs hadoop 0 2023-08-16 20:52 /user/hive
drwxrwxrwt - hdfs hadoop 0 2023-08-16 18:31 /user/mapred
drwxrwxrwt - hdfs hadoop 0 2023-08-16 18:31 /user/pig
drwxrwxrwt - hdfs hadoop 0 2023-08-16 18:31 /user/solr
drwxrwxrwt - hdfs hadoop 0 2023-08-16 18:31 /user/spark
drwxr-xr-x - charlesmeloti hadoop 0 2023-08-16 19:20 /user/trabalho
drwxrwxrwt - hdfs hadoop 0 2023-08-16 18:31 /user/yarn
drwxrwxrwt - hdfs hadoop 0 2023-08-16 18:31 /user/zeppelin
drwxrwxrwt - hdfs hadoop 0 2023-08-16 18:31 /user/zookeeper
charlesmeloti@cluster-charles-m:~$ hdfs dfs -ls /user/charlesmeloti
charlesmeloti@cluster-charles-m:~$
```

## Enviando os arquivos para a partição



```
demogra_cliente.csv end_cliente.csv transacoes.csv
charlesmeloti@cluster-charles-m:~$ hdfs dfs -put end_cliente.csv /user/charlesmeloti/trabalho
charlesmeloti@cluster-charles-m:~$ hdfs dfs -put end_demogra.csv /user/charlesmeloti/trabalho
put: 'end_demogra.csv': No such file or directory
charlesmeloti@cluster-charles-m:~$ hdfs dfs -put demogra_cliente.csv /user/charlesmeloti/trabalho
charlesmeloti@cluster-charles-m:~$ hdfs dfs -put transacoes.csv /user/charlesmeloti/trabalho
charlesmeloti@cluster-charles-m:~$
```

```
hdfs dfs -put end_cliente.csv /user/charlesmeloti/trabalho
hdfs dfs -put demogra_cliente.csv /user/charlesmeloti/trabalho
hdfs dfs -put transacoes.csv /user/charlesmeloti/trabalho
```

## hdfs dfs -cat /user/charlesmeloti/trabalho/demogra\_cliente.csv

```
ssh.cloud.google.com/v2/ssh/projects/fast-mariner-395323/zones/us-central1-a/instances/cluster-charles-m7authuser=1&hl=pt_BR&projectNumber=156259942537&useAdminProxy=true&byoid=true&pageViewId=6C87502E-0446-4E0D-B135-7DEE5A386538 - Google...
ssh.cloud.google.com/v2/ssh/projects/fast-mariner-395323/zones/us-central1-a/instances/cluster-charles-m7authuser=1&hl=pt_BR&projectNumber=156259942537&useAdminProxy=true&byoid=true&pageViewId=6C87502E-0446-4E0D-B135...
SSH no navegador
FAZER UPLOAD DO ARQUIVO FAZER O DOWNLOAD DO ARQUIVO
charlesmeloti@cluster-charles-m:~$ ls
demogra_cliente.csv end_cliente.csv transacoes.csv
charlesmeloti@cluster-charles-m:~$ hdfs dfs -put demogra_cliente.csv /user/charlesmeloti/trabalho
charlesmeloti@cluster-charles-m:~$ hdfs dfs -put end_cliente.csv /user/charlesmeloti/trabalho
charlesmeloti@cluster-charles-m:~$ hdfs dfs -put transacoes.csv /user/charlesmeloti/trabalho
charlesmeloti@cluster-charles-m:~$ hdfs dfs -cat /user/charlesmeloti/trabalho/demogra_cliente.csv
customer_id,name,sobrenome,sexo,compra_bike,ultimo_3_anos,Ano,Mes,Dia,cargo,categoria_cargo,segmento,posui_carro,falecido
1,Laraine,Medendorp,Female,93,1953,outubro,12,Executive Secretary,Health,Mass Customer,Yes,N
2,Eli,Bockman,Male,81,1980,dezembro,16,Administrative Officer,Financial Services,Mass Customer,Yes,N
3,Arlin,Dearle,Male,61,1954,janeiro,20,Recruiting Manager,Property,Mass Customer,Yes,N
4,Talbot,"",Male,33,1961,outubro,3,"",IT,Mass Customer,No,N
5,Sheila-kathryn,Catton,Female,56,1977,maio,13,Senior Editor,n/a,Affluent Customer,Yes,N
6,Curr,Duckhouse,Male,35,1966,setembro,16,"",Retail,High Net Worth,Yes,N
7,Fina,Merrill,Female,6,1976,fevereiro,23,"",Financial Services,Affluent Customer,Yes,N
8,Rod,Inder,Male,31,1962,março,30,Media Manager,I,n/a,Mass Customer,No,N
9,Mala,Lind,Female,97,1973,março,10,Business Systems Development Analyst,Agriculture,Affluent Customer,Yes,N
10,Florence,Birdall,Female,49,1988,outubro,11,Senior Quality Engineer,Financial Services,Mass Customer,Yes,N
11,Utiah,Bisatt,Male,99,1954,abril,30,"",Property,Mass Customer,No,N
12,Sweeney,Plattman,Male,58,1994,julho,21,Nuclear Power Engineer,Manufacturing,Mass Customer,No,N
13,Gabriele,Morcross,Male,38,1955,fevereiro,15,Developer I,Financial Services,High Net Worth,Yes,N
14,Rayshell,Kitteman,Female,85,1983,março,25,Account Executive,Financial Services,Affluent Customer,No,N
15,Erolli,Radage,Male,91,2000,julho,13,Junior Executive,Manufacturing,Mass Customer,No,N
16,Hastin,Raz,Male,36,1977,fevereiro,27,Media Manager,IV,n/a,Mass Customer,Yes,N
17,Heath,Paraday,Male,57,1962,março,19,Sales Associate,n/a,Affluent Customer,Yes,N
18,Margie,Measham,Female,79,1967,julho,6,Professor,n/a,Affluent Customer,No,N
19,Sorcha,Keyson,Female,76,2001,abril,15,Geological Engineer,Manufacturing,High Net Worth,No,N
20,Bastile,Fitch,Male,72,1980,agosto,13,Project Manager,Manufacturing,Mass Customer,No,N
21,Mile,Cammocke,Male,74,1980,setembro,20,Safety Technician I,Manufacturing,Affluent Customer,Yes,N
22,Deeanne,Durnell,Female,79,1962,dezembro,10,"",IT,Mass Customer,No,N
23,Olav,Polak,Male,43,1995,fevereiro,10,"",n/a,High Net Worth,Yes,N
24,Kim,Spivey,Female,55,1977,dezembro,3,Research Assistant I,Agriculture,Mass Customer,Yes,N
25,Geoff,Assaf,Male,72,1976,dezembro,2,Accounting Assistant III,Financial Services,Mass Customer,Yes,N
26,Trixie,Ginnelly,Female,12,1978,junho,10,Editor,Financial Services,Mass Customer,Yes,N
27,Garvin,Klees,Male,37,1978,setembro,25,Research Nurse,Health,Mass Customer,Yes,N
28,Fee,Zeilmer,Male,5,1978,setembro,30,Senior Quality Engineer,Health,Affluent Customer,Yes,N
29,Mona,Sancraft,Female,62,1968,junho,22,Safety Technician III,Manufacturing,Mass Customer,No,N
30,Darrick,Helleckas,Male,18,1961,outubro,18,"",IT,Affluent Customer,Yes,N
31,Star,Praton,Female,3,1962,novembro,24,Staff Accountant III,Telecommunications,High Net Worth,Yes,N
32,Martion,Vanichkin,Female,17,1995,abril,20,Legal Assistant,Manufacturing,Affluent Customer,No,N
33,Ernst,Hacon,Male,44,1957,junho,25,Product Engineer,n/a,Affluent Customer,Yes,N
34,Jephthah,Bachmann,Gay,59,1843,dezembro,21,Legal Assistant,IT,Affluent Customer,No,N
35,Margaretha,Strettie,Female,40,1963,setembro,28,Information Systems Manager,Health,High Net Worth,Yes,N
36,Lurette,Stonell,Female,33,1977,novembro,9,VP Quality Control,n/a,Affluent Customer,No,N
37,Laurie,Derryhouse,Female,46,1985,dezembro,22,Social Worker,Health,High Net Worth,No,N
38,Cordi,Merman,Female,37,1955,outubro,29,Senior Cost Accountant,Financial Services,Affluent Customer,No,N
39,Humfredo,Smalley,Male,5,1979,abril,16,Assistant Media Planner,Entertainment,Mass Customer,No,N
40,Tomasine,Jerchin,Female,91,1981,outubro,27,Payment Adjustment Coordinator,Manufacturing,Affluent Customer,No,N
41,Basilus,Coupe,Male,64,1976,abril,14,Food Chemist,Health,Mass Customer,No,N
42,Chiquita,Durnell,Female,24,1977,março,28,Accountant III,Manufacturing,Mass Customer,No,N
```

## Verificando se tem database criado

```
ssh.cloud.google.com/v2/ssh/projects/fast-mariner-395323/zones/us-central1-a/instances/cluster-charles-m7authuser=1&hl=pt_BR&projectNumber=156259942537&useAdminProxy=true&byoid=true - Google Chrome
ssh.cloud.google.com/v2/ssh/projects/fast-mariner-395323/zones/us-central1-a/instances/cluster-charles-m7authuser=1&hl=pt_BR&projectNumber=156259942537&useAdminProxy=true&byoid=true
SSH no navegador
FAZER UPLOAD DO ARQUIVO FAZER O DOWNLOAD DO ARQUIVO
0: jdbc:hive2://localhost:10000/default> show databases;
INFO : Compiling command(queryId=hive_20230823213931_9beaf53-d9e-493d-b8d6-e0920b10a6d9): show databases
INFO : Concurrent mode is disabled, not creating a lock manager
INFO : Semantic Analysis Completed (retail = false)
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:database_name, type:string, comment:from deserializer)], properties:null)
INFO : Completed compiling command(queryId=hive_20230823213931_9beaf53-d9e-493d-b8d6-e0920b10a6d9): Time taken: 0.066 seconds
INFO : Concurrent mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20230823213931_9beaf53-d9e-493d-b8d6-e0920b10a6d9): show databases
INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryId=hive_20230823213931_9beaf53-d9e-493d-b8d6-e0920b10a6d9): Time taken: 0.035 seconds
INFO : OK
INFO : Concurrent mode is disabled, not creating a lock manager
+-----+
| database_name |
+-----+
| default       |
+-----+
1 row selected (0.161 seconds)
0: jdbc:hive2://localhost:10000/default>
```

## Criando database

```
sshcloud.google.com/v2/ssh/projects/fast-mariner-395323/zones/us-central1-a/instances/cluster-charles-m7authuser=1&hl=pt_BR&projectNumber=156259942537&useAdminProxy=true&byoid=true - Google Chrome
sshcloud.google.com/v2/ssh/projects/fast-mariner-395323/zones/us-central1-a/instances/cluster-charles-m7authuser=1&hl=pt_BR&projectNumber=156259942537&useAdminProxy=true&byoid=true

SSH no navegador

FAZER UPLOAD DO ARQUIVO
FAZER O DOWNLOAD DO ARQUIVO

INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Semantic Analysis Completed (retrial = false)
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:database_name, type:string, comment:from deserializer)], properties:null)
INFO : Completed compiling command(queryId=hive_20230823214827_6599b999-657a-4a6c-a739-36b5a9alc3d4); Time taken: 0.025 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20230823214827_6599b999-657a-4a6c-a739-36b5a9alc3d4): show databases
INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryId=hive_20230823214827_6599b999-657a-4a6c-a739-36b5a9alc3d4); Time taken: 0.005 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager
+-----+
| database_name |
+-----+
| default       |
+-----+
1 row selected (0.054 seconds)
0: jdbc:hive2://localhost:10000/default> create database charles_hadoop;
INFO : Compiling command(queryId=hive_20230823214847_69aba4f1-f620-4ccc-a91d-a88499dd074f): create database charles_hadoop
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Semantic Analysis Completed (retrial = false)
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:database_name, type:string, comment:from deserializer)], properties:null)
INFO : Completed compiling command(queryId=hive_20230823214847_69aba4f1-f620-4ccc-a91d-a88499dd074f); Time taken: 0.018 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20230823214847_69aba4f1-f620-4ccc-a91d-a88499dd074f): create database charles_hadoop
INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryId=hive_20230823214847_69aba4f1-f620-4ccc-a91d-a88499dd074f); Time taken: 0.048 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager
No rows affected (0.09 seconds)
0: jdbc:hive2://localhost:10000/default> show databases;
INFO : Compiling command(queryId=hive_20230823214850_77ea405d-49e6-4c92-839e-827576994203): show databases
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Semantic Analysis Completed (retrial = false)
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:database_name, type:string, comment:from deserializer)], properties:null)
INFO : Completed compiling command(queryId=hive_20230823214850_77ea405d-49e6-4c92-839e-827576994203); Time taken: 0.047 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20230823214850_77ea405d-49e6-4c92-839e-827576994203): show databases
INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryId=hive_20230823214850_77ea405d-49e6-4c92-839e-827576994203); Time taken: 0.009 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager
+-----+
| database_name |
+-----+
| charles_hadoop |
| default       |
+-----+
2 rows selected (0.122 seconds)
0: jdbc:hive2://localhost:10000/default>
```

## Criando tabela end\_cliente e usando describe

```
CREATE EXTERNAL TABLE end_cliente (customer_id int, endereco STRING, estado String,
pais STRING, codigo_postal int)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE;
LOCATION '/user/charlesmeloti/trabalho';
```

```
hive> CREATE EXTERNAL TABLE end_cliente (customer_id int, endereco STRING, estado int, pais STRING, codigo_postal int)
> ROW FORMAT DELIMITED
> FIELDS TERMINATED BY ','
> STORED AS TEXTFILE
> LOCATION '/user/charlesmeloti/trabalho';
OK
Time taken: 0.314 seconds
hive>
```

## describe end\_cliente;

```
INFO : Compiling command(queryId=hive_20230825013721_b17f4eb2-Sb19-4ce0-b1a7-48a5ffa38ae7): describe end_cliente
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Semantic Analysis Completed (retrial = false)
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:col name, type:string, comment:from deserializer), FieldSchema(name:data_type, type:string, comment:from deserializer), FieldSchema(name:comment, type:string, comment:from deserializer)], properties:null)
INFO : Completed compiling command(queryId=hive_20230825013721_b17f4eb2-Sb19-4ce0-b1a7-48a5ffa38ae7); Time taken: 0.051 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20230825013721_b17f4eb2-Sb19-4ce0-b1a7-48a5ffa38ae7): describe end_cliente
INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryId=hive_20230825013721_b17f4eb2-Sb19-4ce0-b1a7-48a5ffa38ae7); Time taken: 0.027 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager
+-----+
| col_name | data_type | comment |
+-----+
| customer_id | int | |
| endereco | string | |
| estado | string | |
| pais | string | |
| codigo_postal | int | |
+-----+
5 rows selected (0.118 seconds)
0: jdbc:hive2://localhost:10000/default>
```

## Selecionando toda tabela end\_cliente

ssh.cloud.google.com/v2/ssh/projects/fast-mariner-395323/zones/us-central1-a/instances/cluster-charles-m?authuser=1&hl=pt_BR&projectNumber=156259942537&useAdminProxy=true&byoid=true - Google Chrome				
ssh.cloud.google.com/v2/ssh/projects/fast-mariner-395323/zones/us-central1-a/instances/cluster-charles-m?authuser=1&hl=pt_BR&projectNumber=156259942537&useAdminProxy=true&byoid=true				
SSH no navegador				
3901	8458 Stang Parkway	NSW	Australia	2671
3902	60 Towne Pass	NSW	Australia	2877
3903	5 Rutledge Center	QLD	Australia	4311
end_cliente.customer_id	end_cliente.endereco	end_cliente.estado	end_cliente.pais	end_cliente.codigo_postal
3904	9 Blaine Place	VIC	Australia	3756
3905	24688 Hollow Ridge Way	VIC	Australia	3064
3906	32876 American Ash Avenue	NSW	Australia	2099
3907	18 Bay Place	NSW	Australia	2176
3908	4 Hanover Junction	VIC	Australia	3192
3909	0 Darwin Junction	NSW	Australia	2747
3910	5211 Merrick Court	VIC	Australia	3103
3911	801 Shasta Crossing	NSW	Australia	2539
3912	55 Anzinger Drive	QLD	Australia	4152
3913	13768 Commercial Trail	NSW	Australia	2155
3914	109 Rippling Road	NSW	Australia	2156
3915	47 Spenser Circle	NSW	Australia	2770
3916	15051 Monterey Street	NSW	Australia	2756
3917	7365 Cariboca Hill	QLD	Australia	4680
3918	42 Maryland Plaza	NSW	Australia	2486
3919	802 Bashford Hill	VIC	Australia	3150
3920	12548 Eagan Court	NSW	Australia	2768
3921	24167 Montana Place	NSW	Australia	2065
3922	0 Southridge Drive	QLD	Australia	4311
3923	1 Banding Way	QLD	Australia	4218
3924	259 Union Junction	VIC	Australia	3107
3925	7727 Merrick Parkway	NSW	Australia	2099
3926	3015 Bowman Alley	NSW	Australia	2777
3927	2 Annamark Park	NSW	Australia	2170
3928	38 Steensland Circle	NSW	Australia	2234
3929	80 Northland Place	QLD	Australia	4178
3930	104 Kingsford Park	NSW	Australia	2154
3931	22 Mifflin Junction	NSW	Australia	2096
3932	60272 Montana Drive	NSW	Australia	2223
3933	730 Judy Drive	VIC	Australia	3959
3934	1269 Esch Parkway	QLD	Australia	4078
3935	3 Kedzie Center	VIC	Australia	3106
3936	67473 Nova Avenue	VIC	Australia	3089
3937	866 Canary Drive	NSW	Australia	2750
3938	20 Dwight Junction	NSW	Australia	2768
3939	0900 Northport Point	NSW	Australia	2085
3940	572 Buena Vista Plaza	NSW	Australia	2111
3941	695 Anthes Crossing	QLD	Australia	4350
3942	62 Luster Avenue	NSW	Australia	2036
3943	819 Basil Plaza	QLD	Australia	4053
3944	81 Monument Avenue	NSW	Australia	2126
3945	093 Mallard Hill	QLD	Australia	4034
3946	903 Southoff Plaza	NSW	Australia	2565

select \* from end\_cliente limit 50;

ssh.cloud.google.com/v2/ssh/projects/fast-mariner-395323/zones/us-central1-a/instances/cluster-charles-m?authuser=1&hl=pt_BR&projectNumber=156259942537&useAdminProxy=true&byoid=true - Google Chrome				
ssh.cloud.google.com/v2/ssh/projects/fast-mariner-395323/zones/us-central1-a/instances/cluster-charles-m?authuser=1&hl=pt_BR&projectNumber=156259942537&useAdminProxy=true&byoid=true				
SSH no navegador				
end_cliente.customer_id	end_cliente.endereco	end_cliente.estado	end_cliente.pais	end_cliente.codigo_postal
NULL	endereco	estado	pais	NULL
1	060 Morning Avenue	New South Wales	Australia	2016
2	6 Meadow Vale Court	New South Wales	Australia	2153
4	0 Holy Cross Court	QLD	Australia	4211
5	17979 Del Mar Point	New South Wales	Australia	2448
6	9 Oakridge Court	VIC	Australia	3216
7	4 Delaware Trail	New South Wales	Australia	2210
8	49 Londonderry Lane	New South Wales	Australia	2650
9	97736 7th Trail	New South Wales	Australia	2023
11	93405 Ludington Park	VIC	Australia	3044
12	44339 Golden Leaf Alley	QLD	Australia	4557
13	2 Sutherland Street	VIC	Australia	3799
14	9 Mcbride Trail	New South Wales	Australia	2760
15	9861 New Castle Avenue	New South Wales	Australia	2428
16	52 Moland Street	VIC	Australia	3331
17	82391 Kensington Lane	VIC	Australia	3058
18	092 2nd Alley	New South Wales	Australia	2135
19	59 Spaight Circle	New South Wales	Australia	2233
20	032 Bartelt Crossing	New South Wales	Australia	2444
21	18 Jenna Center	New South Wales	Australia	2650
24	3 Cordelia Plaza	New South Wales	Australia	2153
25	28 5th Center	QLD	Australia	4413
26	52 Carey Alley	QLD	Australia	4740
27	96 Texas Plaza	VIC	Australia	3218
28	48 Eagan Avenue	QLD	Australia	4868
29	9 Buell Park	QLD	Australia	4116
30	02663 Buell Parkway	New South Wales	Australia	2519
31	2294 Pleasure Place	New South Wales	Australia	2135
32	2351 Petterle Place	New South Wales	Australia	2756
33	63 Lukken Drive	New South Wales	Australia	2170
34	833 Luster Way	QLD	Australia	4005
35	34 Russell Plaza	QLD	Australia	4127
36	78 Del Sol Hill	New South Wales	Australia	2535
37	54 Sege Plaza	QLD	Australia	4726
38	90 Lawn Parkway	QLD	Australia	4805
39	38726 Ilene Crossing	New South Wales	Australia	2212
40	3496 Brown Drive	New South Wales	Australia	2207
41	7188 Cody Way	VIC	Australia	3350
42	1 Claremont Park	VIC	Australia	3931
43	60 Eagan Circle	QLD	Australia	4018
44	89314 Eagle Crest Center	QLD	Australia	4670
45	2028 Lakeswood Place	New South Wales	Australia	2480
46	70875 Hudson Point	VIC	Australia	3212
47	543 Killdeer Park	New South Wales	Australia	2322
48	22125 Ramsey Drive	VIC	Australia	3188
49	3107 Calypso Terrace	New South Wales	Australia	2210



Criando a tabela demogra\_cliente pelo Beeline

```
CREATE TABLE demogra_cliente (customer_id int, nome STRING, sobrenome String, sexo
                                STRING, compra_bike_ultimo_3_ano int,
                                Ano int, Mes String, Dia int, cargo String, categoria_cargo String, segmento String, possui_carro
                                String, falecido String
                                )
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE;
```

Jogando os dados na tabela demogra\_cliente

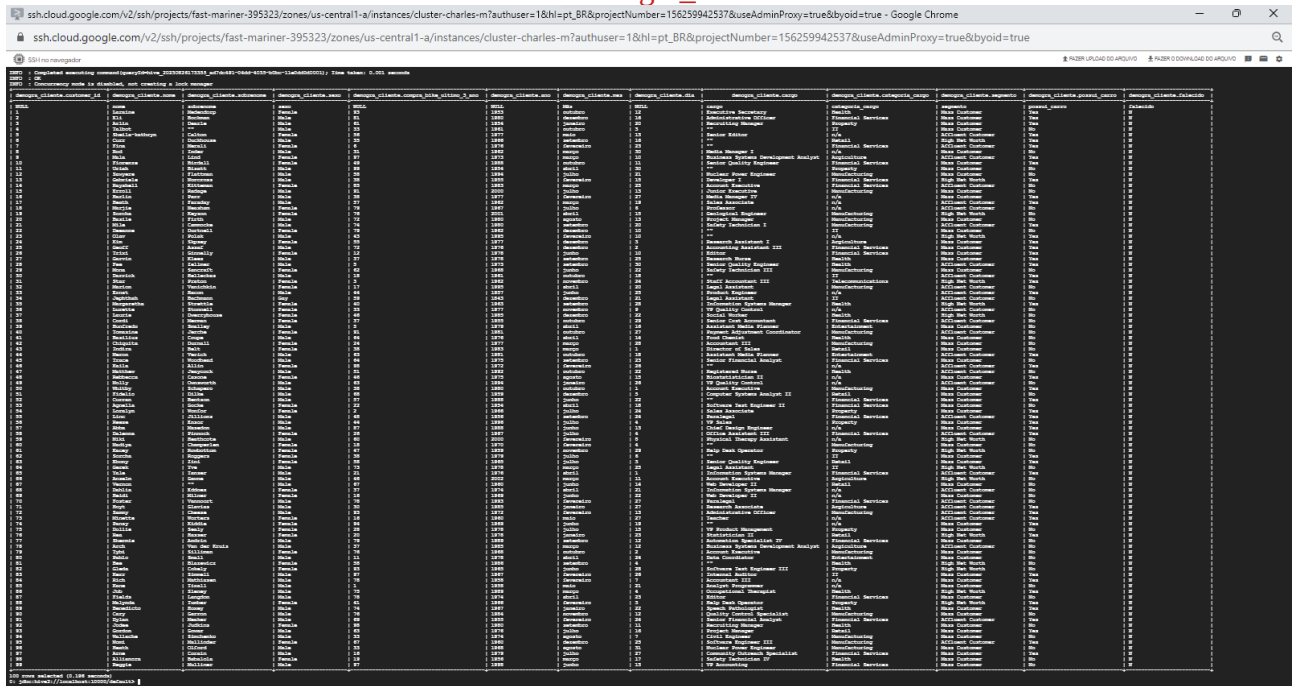
```
LOAD DATA INPATH '/user/charlesmeloti/trabalho/demogra_cliente.csv' overwrite into table
demogra_cliente;
```

```
INFO : Compiling command(queryId=hive_20230825181527_f9844212-373a-41ea-a5d3-dd0fc89c9459): CREATE TABLE demogra_cliente (customer_id int, nome STRING, sobrenome String, sexo STRING, compra_bike_ulti
mo_3_ano int,
Ano int, Mes String, Dia int, cargo String, categoria_cargo String, segmento String, possui_carro String, falecido String
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Semantic Analysis Completed (retrial = false)
INFO : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO : Completed compiling command(queryId=hive_20230825181527_f9844212-373a-41ea-a5d3-dd0fc89c9459): Time taken: 0.782 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20230825181527_f9844212-373a-41ea-a5d3-dd0fc89c9459): CREATE TABLE demogra_cliente (customer_id int, nome STRING, sobrenome String, sexo STRING, compra_bike_ulti
mo_3_ano int,
Ano int, Mes String, Dia int, cargo String, categoria_cargo String, segmento String, possui_carro String, falecido String
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE
INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryId=hive_20230825181527_f9844212-373a-41ea-a5d3-dd0fc89c9459): Time taken: 2.67 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager
No rows affected (5.668 seconds)
0: jdbc:hive2://localhost:10000/default>
```

describe demogra\_cliente;

```
0: jdbc:hive2://localhost:10000/default> describe demogra_cliente;
INFO : Compiling command(queryId=hive_20230826180454_2af3e4a6-e60b-44a9-a848-4aecd293acd): describe demogra_cliente
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Semantic Analysis Completed (retrial = false)
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:col_name, type:string, comment:from deserializer), FieldSchema(name:data_type, type:string, comment:from deserializer), FieldSchema
(name:comment, type:string, comment:from deserializer)], properties:null)
INFO : Completed compiling command(queryId=hive_20230826180454_2af3e4a6-e60b-44a9-a848-4aecd293acd): Time taken: 0.031 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20230826180454_2af3e4a6-e60b-44a9-a848-4aecd293acd): describe demogra_cliente
INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryId=hive_20230826180454_2af3e4a6-e60b-44a9-a848-4aecd293acd): Time taken: 0.015 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager
+-----+-----+-----+
|      col_name      | data type | comment |
+-----+-----+-----+
| customer_id        | int       |         |
| nome                | string    |         |
| sobrenome           | string    |         |
| sexo                | string    |         |
| compra_bike_ultimo_3_ano | int       |         |
| ano                 | int       |         |
| mes                 | string    |         |
| dia                 | int       |         |
| cargo               | string    |         |
| categoria_cargo     | string    |         |
| segmento            | string    |         |
| possui_carro        | string    |         |
| falecido            | string    |         |
+-----+-----+-----+
13 rows selected (0.071 seconds)
0: jdbc:hive2://localhost:10000/default>
```

select \* from demogra\_cliente limit 100



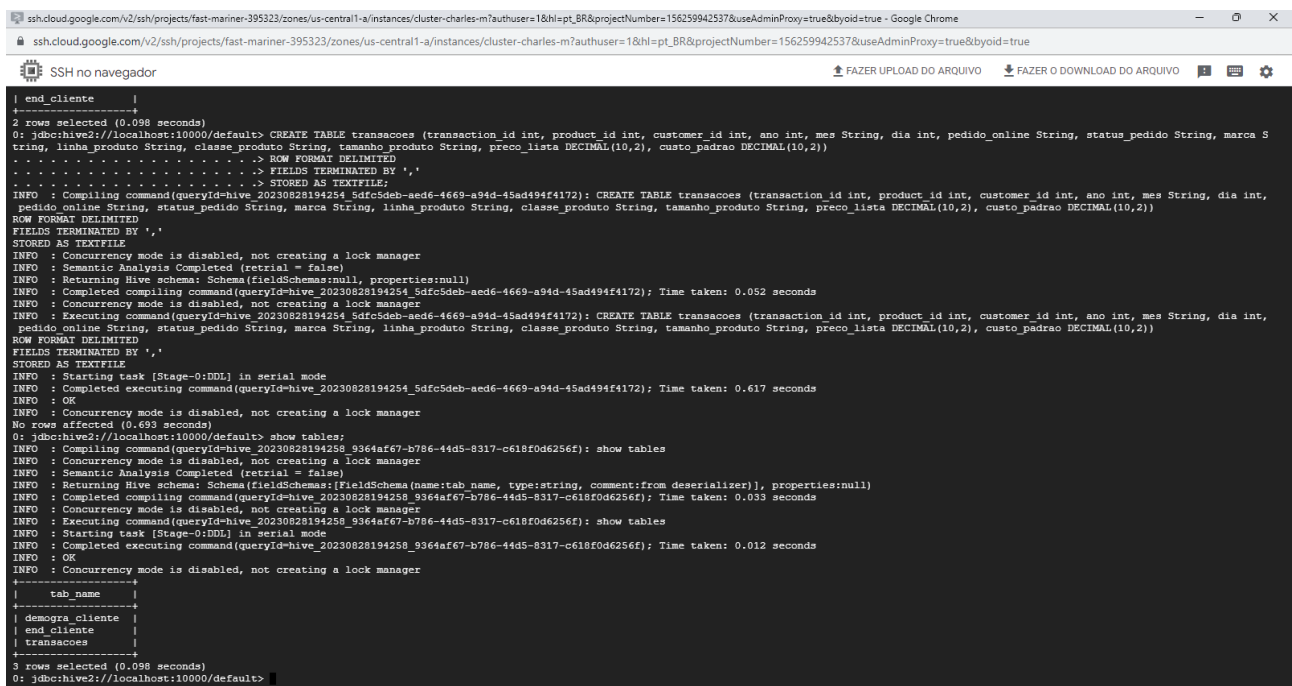
Criando a tabela transacoes

CREATE TABLE transacoes (transaction\_id int, product\_id int, customer\_id int, ano int, mes String, dia int, pedido\_online String, status\_pedido String, marca String, linha\_produto String, classe\_produto String, tamanho\_produto String, preco\_lista DECIMAL(10,2), custo\_padrao DECIMAL(10,2))

ROW FORMAT DELIMITED  
FIELDS TERMINATED BY ','  
STORED AS TEXTFILE

Jogando os dados na tabela transacoes\_cliente

LOAD DATA INPATH '/user/charlesmeloti/trabalho/transacoes.csv' overwrite into table transacoes;



describe transacoes;

```
INFO : Concurrency mode is disabled, not creating a lock manager
No rows affected (0.51 seconds)
0: jdbc:hive2://localhost:10000/default> describe transacoes;
INFO : Compiling command(queryId=hive_20230828194556_98d49a9a-f623-4e43-9aeb-9e7276a8be6c): describe transacoes
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Semantic Analysis Completed (retired = false)
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:col_name, type:string, comment:from deserializer), FieldSchema(name:data_type, type:string, comment:from deserializer), FieldSchema(name:comment, type:string, comment:from deserializer)], properties:null)
INFO : Completed compiling command(queryId=hive_20230828194556_98d49a9a-f623-4e43-9aeb-9e7276a8be6c): Time taken: 0.058 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20230828194556_98d49a9a-f623-4e43-9aeb-9e7276a8be6c): describe transacoes
INFO : Starting task [Stage-0:DML] in serial mode
INFO : Completed executing command(queryId=hive_20230828194556_98d49a9a-f623-4e43-9aeb-9e7276a8be6c): Time taken: 0.046 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager

+-----+-----+-----+
| col_name | data_type | comment |
+-----+-----+-----+
| transaction_id | int | |
| product_id | int | |
| customer_id | int | |
| ano | int | |
| mes | string | |
| dia | int | |
| pedido_online | string | |
| status_pedido | string | |
| marca | string | |
| linha_produto | string | |
| classe_produto | string | |
| tamanho_produto | string | |
| preco_lista | decimal(10,2) | |
| custo_padrao | decimal(10,2) | |
+-----+-----+-----+

14 rows selected (0.161 seconds)
0: jdbc:hive2://localhost:10000/default>
```

select \* from transacoes limit 100;

ssh.cloud.google.com/v2/ssh/projects/fast-mariner-395323/zones/us-central1-a/instances/cluster-charles-m?authuser=1&hl=pt_BR&projectNumbers=156259942537&useAdminProxy=true&byoid=true - Google Chrome													
ssh.cloud.google.com/v2/ssh/projects/fast-mariner-395323/zones/us-central1-a/instances/cluster-charles-m?authuser=1&hl=pt_BR&projectNumbers=156259942537&useAdminProxy=true&byoid=true													
CUI no navegador													
Completed executing command(queryId=hive_20230828194556_98d49a9a-f623-4e43-9aeb-9e7276a8be6c): Time taken: 0.058 seconds													
Concurrency mode is disabled, not creating a lock manager													
transacoes.transaction_id	transacoes.product_id	transacoes.customer_id	transacoes.ano	transacoes.mes	transacoes.dia	transacoes.pedido_online	transacoes.status_pedido	transacoes.marca	transacoes.linha_produto	transacoes.classe_produto	transacoes.tamanho_produto	transacoes.preco_lista	transacoes.custo_padrao
0	0	0	2023	08	01	0	0	0	0	0	0	0	0
1	1	1	2023	08	01	1	1	1	1	1	1	1	1
2	2	2	2023	08	01	2	2	2	2	2	2	2	2
3	3	3	2023	08	01	3	3	3	3	3	3	3	3
4	4	4	2023	08	01	4	4	4	4	4	4	4	4
5	5	5	2023	08	01	5	5	5	5	5	5	5	5
6	6	6	2023	08	01	6	6	6	6	6	6	6	6
7	7	7	2023	08	01	7	7	7	7	7	7	7	7
8	8	8	2023	08	01	8	8	8	8	8	8	8	8
9	9	9	2023	08	01	9	9	9	9	9	9	9	9
10	10	10	2023	08	01	10	10	10	10	10	10	10	10
11	11	11	2023	08	01	11	11	11	11	11	11	11	11
12	12	12	2023	08	01	12	12	12	12	12	12	12	12
13	13	13	2023	08	01	13	13	13	13	13	13	13	13
14	14	14	2023	08	01	14	14	14	14	14	14	14	14
15	15	15	2023	08	01	15	15	15	15	15	15	15	15
16	16	16	2023	08	01	16	16	16	16	16	16	16	16
17	17	17	2023	08	01	17	17	17	17	17	17	17	17
18	18	18	2023	08	01	18	18	18	18	18	18	18	18
19	19	19	2023	08	01	19	19	19	19	19	19	19	19
20	20	20	2023	08	01	20	20	20	20	20	20	20	20
21	21	21	2023	08	01	21	21	21	21	21	21	21	21
22	22	22	2023	08	01	22	22	22	22	22	22	22	22
23	23	23	2023	08	01	23	23	23	23	23	23	23	23
24	24	24	2023	08	01	24	24	24	24	24	24	24	24
25	25	25	2023	08	01	25	25	25	25	25	25	25	25
26	26	26	2023	08	01	26	26	26	26	26	26	26	26
27	27	27	2023	08	01	27	27	27	27	27	27	27	27
28	28	28	2023	08	01	28	28	28	28	28	28	28	28
29	29	29	2023	08	01	29	29	29	29	29	29	29	29
30	30	30	2023	08	01	30	30	30	30	30	30	30	30
31	31	31	2023	08	01	31	31	31	31	31	31	31	31
32	32	32	2023	08	01	32	32	32	32	32	32	32	32
33	33	33	2023	08	01	33	33	33	33	33	33	33	33
34	34	34	2023	08	01	34	34	34	34	34	34	34	34
35	35	35	2023	08	01	35	35	35	35	35	35	35	35
36	36	36	2023	08	01	36	36	36	36	36	36	36	36
37	37	37	2023	08	01	37	37	37	37	37	37	37	37
38	38	38	2023	08	01	38	38	38	38	38	38	38	38
39	39	39	2023	08	01	39	39	39	39	39	39	39	39
40	40	40	2023	08	01	40	40	40	40	40	40	40	40
41	41	41	2023	08	01	41	41	41	41	41	41	41	41
42	42	42	2023	08	01	42	42	42	42	42	42	42	42
43	43	43	2023	08	01	43	43	43	43	43	43	43	43
44	44	44	2023	08	01	44	44	44	44	44	44	44	44
45	45	45	2023	08	01	45	45	45	45	45	45	45	45
46	46	46	2023	08	01	46	46	46	46	46	46	46	46
47	47	47	2023	08	01	47	47	47	47	47	47	47	47
48	48	48	2023	08	01	48	48	48	48	48	48	48	48
49	49	49	2023	08	01	49	49	49	49	49	49	49	49
50	50	50	2023	08	01	50	50	50	50	50	50	50	50
51	51	51	2023	08	01	51	51	51	51	51	51	51	51
52	52	52	2023	08	01	52	52	52	52	52	52	52	52
53	53	53	2023	08	01	53	53	53	53	53	53	53	53
54	54	54	2023	08	01	54	54	54	54	54	54	54	54
55	55	55	2023	08	01	55	55	55	55	55	55	55	55
56	56	56	2023	08	01	56	56	56	56	56	56	56	56
57	57	57	2023	08	01	57	57	57	57	57	57	57	57
58	58	58	2023	08	01	58	58	58	58	58	58	58	58
59	59	59	2023	08	01	59	59	59	59	59	59	59	59
60	60	60	2023	08	01	60	60	60	60	60	60	60	60
61	61	61	2023	08	01	61	61	61	61	61	61	61	61
62	62	62	2023	08	01	62	62	62	62	62	62	62	62
63	63	63	2023	08	01	63	63	63	63	63	63	63	63
64	64	64	2023	08	01	64	64	64	64	64	64	64	64
65	65	65	2023	08	01	65	65	65	65	65	65	65	65
66	66	66	2023	08	01	66	66	66	66	66	66	66	66
67	67	67	2023	08	01	67	67	67	67	67	67	67	67
68	68	68	2023	08	01	68	68	68	68	68	68	68	68
69	69	69	2023	08	01	69	69	69	69	69	69	69	69
70	70	70	2023	08	01	70	70	70	70	70	70	70	70
71	71	71	2023	08	01	71	71	71	71	71	71	71	71
72	72	72	2023	08	01	72	72	72	72	72	72	72	72
73	73	73	2023	08	01	73	73	73	73	73	73	73	73
74	74	74	2023	08	01	74	74	74	74	74	74	74	74
75	75	75	2023	08	01	75	75	75	75	75	75	75	75
76	76	76	2023	08	01	76	76	76	76	76	76	76	76
77	77	77	2023	08	01	77	77	77	77	77	77	77	77
78	78	78	2023	08	01	78	78	78	78	78	78	78	78
79	79	79	2023	08	01	79	79	79	79	79	79	79	79
80	80	80	2023	08	01	80	80	80	80	80	80	80	80
81	81	81	2023	08	01	81	81	81	81	81	81	81	81
82	82	82	2023	08	01	82	82	82	82	82	82	82	82
83	83	83	2023	08	01	83	83	83	83	83	83	83	83
84	84	84	2023	08	01	84	84	84	84	84	84	84	84
85	85	85	2023	08	01	85	85	85	85	85	85	85	85
86	86	86	2023	08	01	86	86	86	86	86	86	86	86
87	87	87	2023	08	01	87	87	87	87	87	87	87	87
88	88	88	2023	08	01	88	88	88	88	88	88	88	88
89	89	89	2023	08	01	89	89	89	89	89	89	89	89
90	90	90	2023	08	01	90	90	90	90	90	90	90	90
91	91	91	2023	08	01	91	91	91	91	91	91	91	91
92	92	92	2023	08	01	92	92	92	92	92	92	92	92
93	93	93	2023	08	01	93	93	93	93	93	93	93	93
94	94	94	2023	08	01	94	94	94	94	94	94	94	94
95	95	95	2023	08	01	95	95	95	95	95	95	95	95
96	96	96	2023	08	01	96	96	96	96	96	96	96	96
97	97	97	2023	08	01	97	97	97	97	97	97	97	97
98	98	98	2023	08	01	98	98	98	98	98	98	98	98
99	99	99	2023	08	01	99	99	99	99	99	99	99	99

## 4. Formulação de perguntas analíticas a serem respondidas com os dados

## 5. Consultas SQL que respondem às perguntas elaboradas

1. Visualizar a classe dos produtos comprados de todos os clientes que compraram bicicletas com o cargo de ‘Teacher’ e que possuem carro.

```
select
dc.cargo,
dc.nome,
dc.sobrenome,
dc.possui_carro,
t.classe_produto
from demogra_cliente as dc
JOIN transacoes as t
ON dc.customer_id = t.customer_id
where cargo = 'Teacher' and possui_carro = 'Yes';
```

```
ssh.cloud.google.com/v2/ssh/projects/fast-mainer-395323/zones/us-central1-a/instances/cluster-charles-m?authuser=1&hl=pt_BR&projectNumbers=156259942537&useAdminProxy=true&byoid=true - Google Chrome
ssh.cloud.google.com/v2/ssh/projects/fast-mainer-395323/zones/us-central1-a/instances/cluster-charles-m?authuser=1&hl=pt_BR&projectNumbers=156259942537&useAdminProxy=true&byoid=true

SSH no navegador FAZER UPLOAD DO ARQUIVO FAZER O DOWNLOAD DO ARQUIVO

VERTICES MODE STATUS TOTAL COMPLETED RUNNING PENDING FAILED KILLED
Map 1 ..... container SUCCEEDED 1 1 0 0 0 0
Map 2 ..... container SUCCEEDED 1 1 0 0 0 0
VERTICES: 02/02 [=====] 100% ELAPSED TIME: 9.80 s
INFO : Completed executing command(queryId=hive_20230828204036_a505afe9-ffdf-47c3-83ac-d04387452f2); Time taken: 10.169 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager
+-----+
| dc.cargo | dc.nome | dc.sobrenome | dc.possui_carro | t.classe_produto |
+-----+
| Teacher | Judah | Prime | Yes | low |
| Teacher | Lettie | McGenis | Yes | medium |
| Teacher | Judah | Prime | Yes | high |
| Teacher | Judah | Prime | Yes | medium |
| Teacher | Lettie | McGenis | Yes | low |
| Teacher | Lettie | McGenis | Yes | medium |
| Teacher | Judah | Prime | Yes | low |
| Teacher | Lettie | McGenis | Yes | medium |
| Teacher | Judah | Prime | Yes | low |
| Teacher | Lettie | McGenis | Yes | medium |
| Teacher | Darcey | Annis | Yes | high |
| Teacher | Kalinda | "" | Yes | medium |
| Teacher | Kalinda | "" | Yes | medium |
| Teacher | Darcey | Annis | Yes | low |
| Teacher | Kalinda | "" | Yes | high |
| Teacher | Kalinda | "" | Yes | medium |
| Teacher | Kalinda | "" | Yes | low |
| Teacher | Kalinda | "" | Yes | medium |
| Teacher | Darcey | Annis | Yes | medium |
| Teacher | Darcey | Annis | Yes | high |
| Teacher | Kalinda | "" | Yes | low |
| Teacher | Dmitry | Kunkler | Yes | medium |
| Teacher | Dmitry | Kunkler | Yes | high |
| Teacher | Dmitry | Kunkler | Yes | medium |
| Teacher | Dmitry | Kunkler | Yes | medium |
| Teacher | Julie | Kennagh | Yes | medium |
| Teacher | Julie | Kennagh | Yes | high |
| Teacher | Julie | Kennagh | Yes | high |
| Teacher | Julie | Kennagh | Yes | low |
| Teacher | Julie | Kennagh | Yes | low |
```

2. Visualizar a soma do valor de todos os produtos vendidos no mês de ‘dezembro’ do ano de ‘2017’.

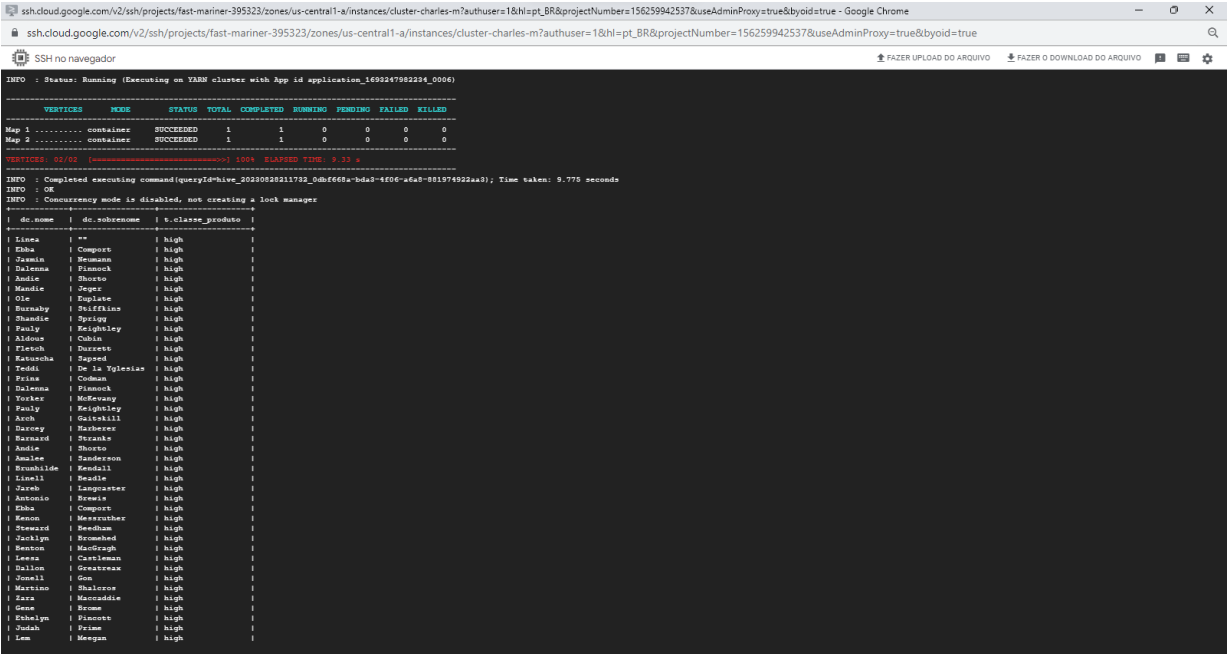
```
select sum (preco_lista)
from transacoes
where mes = 'dezembro';
```

```
0: jdbc:hive2://localhost:10000/default> select sum (preco_lista) as soma_preco from transacoes where mes = 'dezembro';
INFO : Compiling command(queryId=hive_20230828210300_45ae1f63-74ac-474c-abf5-aa27d4c78a5c): select sum (preco_lista) as soma_preco from transacoes where mes = 'dezembro'
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Semantic Analysis Completed (retiral = false)
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:soma_preco, type:decimal(20,2), comment:null)], properties:null)
INFO : Completed compiling command(queryId=hive_20230828210300_45ae1f63-74ac-474c-abf5-aa27d4c78a5c); Time taken: 0.146 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20230828210300_45ae1f63-74ac-474c-abf5-aa27d4c78a5c): select sum (preco_lista) as soma_preco from transacoes where mes = 'dezembro'
INFO : Query ID = hive_20230828210300_45ae1f63-74ac-474c-abf5-aa27d4c78a5c
INFO : Total jobs = 1
INFO : Launching Job 1 out of 1
INFO : Starting task [Stage-1:MAPRED] in serial mode
INFO : Subscribed to counters: {} for queryId: hive_20230828210300_45ae1f63-74ac-474c-abf5-aa27d4c78a5c
INFO : Session is already open
INFO : Dag name: select sum (preco_lista) as scm...'dezembro' (Stage-1)
INFO : Status: Running (Executing on YARN cluster with App id application_1693247982234_0005)

VERTICES MODE STATUS TOTAL COMPLETED RUNNING PENDING FAILED KILLED
Map 1 ..... container SUCCEEDED 1 1 0 0 0 0
Reducer 2 ..... container SUCCEEDED 1 1 0 0 0 0
VERTICES: 02/02 [=====] 100% ELAPSED TIME: 9.55 s
INFO : Completed executing command(queryId=hive_20230828210300_45ae1f63-74ac-474c-abf5-aa27d4c78a5c); Time taken: 9.786 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager
+-----+
| soma_preco |
+-----+
| 127424366.00 |
+-----+
1 row selected (9.991 seconds)
0: jdbc:hive2://localhost:10000/default>
```

3. Seleccionar nomes e sobrenomes de 50 clientes que compraram bicicletas da classe ‘High’.

```
select
dc.nome,
dc.sobrenome,
t.classe_produto
from demogra_cliente as dc
JOIN transacoes as t
ON dc.customer_id = t.customer_id
where classe_produto = 'high' limit 50;
```



```
INFO : Status: Running (Executing on YARN cluster with App id application_1693247982234_0006)

-----
VERTICES      NAME      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container  SUCCEEDED  1      1      0      0      0      0
Map 2 ..... container  SUCCEEDED  1      1      0      0      0      0

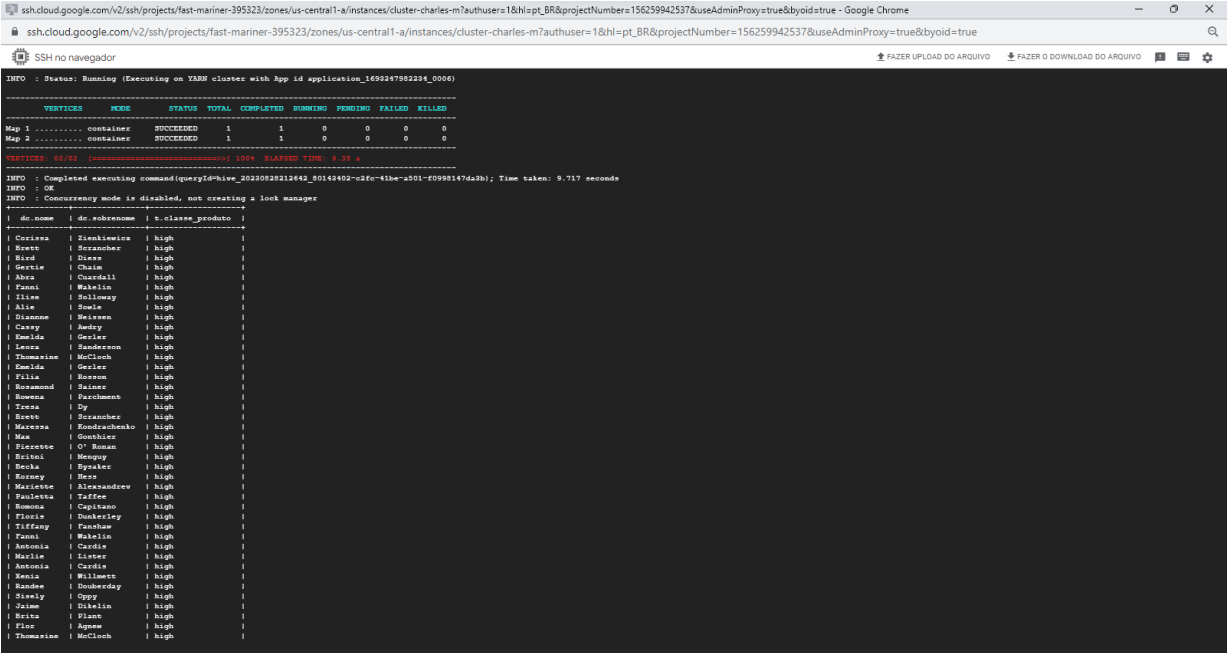
VERTICES 0/62 (=====) 100% ELAPSED TIME 9.77 s

INFO : Completed executing command(queryId=hive_2023082811732_0dbf66fa-bda3-4f06-afaf-80197492aa83); Time taken: 9.775 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager

+-----+
| dc.nome | dc.sobrenome | t.classe_produto |
+-----+
| Linda  | Lee         | high             |
| Edna  | Compost    | high             |
| Jasmin | Beumann    | high             |
| Dalena | Finnock    | high             |
| Andie  | Shusto     | high             |
| Mandie | Dejer      | high             |
| Ols    | Eplacet    | high             |
| Burnaby | Stiffline  | high             |
| Shondie | Spragg     | high             |
| Paulie | Neighley   | high             |
| Aldous | Cuhin      | high             |
| Fitch  | Ducrest    | high             |
| Katucha | Sagoed     | high             |
| Teddi  | De la Tglenias | high             |
| Trina  | Cuhin      | high             |
| Dalena | Finnock    | high             |
| Esther | McCreary   | high             |
| Paulie | Neighley   | high             |
| Arch  | Galtkill   | high             |
| Barney | Radheaux   | high             |
| Bernard | Stranks    | high             |
| Andie  | Shusto     | high             |
| Amalie | Sanderson  | high             |
| Brunkilde | Kendall  | high             |
| Linell | Beadie     | high             |
| Japh  | Langraeter | high             |
| Antonio | Rerwin     | high             |
| Edna  | Compost    | high             |
| Xenon  | Meersruther | high             |
| Steward | Renthlow   | high             |
| Jashlyn | Bromhead   | high             |
| Swanton | MacGragh   | high             |
| Leora  | Castlemann | high             |
| Dalina | Greatstream | high             |
| Jonell | Gert       | high             |
| Martin | Shalence   | high             |
| Izra  | Macnoddie  | high             |
| Gert  | Brown      | high             |
| Scholym | Fincrest  | high             |
| Judah  | Prime      | high             |
| Lem    | Kengen     | high             |
+-----+
```

4. Seleccionar nomes das 50 clientes do sexo ‘female’ que compraram bicicletas da classe ‘High’.

```
select
dc.nome,
dc.sobrenome,
dc.sexo,
t.classe_produto
from demogra_cliente as dc
JOIN transacoes as t
ON dc.customer_id = t.customer_id
where sexo = 'Female' and classe_produto = 'High' limit 50;
```



```
INFO : Status: Running (Executing on YARN cluster with App id application_1693247982234_0006)

-----
VERTICES      NAME      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container  SUCCEEDED  1      1      0      0      0      0
Map 2 ..... container  SUCCEEDED  1      1      0      0      0      0

VERTICES 0/62 (=====) 100% ELAPSED TIME 9.71 s

INFO : Completed executing command(queryId=hive_2023082812442_80143402-c2fc-41be-a501-f0988147da3b); Time taken: 9.717 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager

+-----+
| dc.nome | dc.sobrenome | dc.sexo | t.classe_produto |
+-----+
| Corineea | Elenkiewicz | female | high             |
| Berna  | Seranacher  | female | high             |
| Bird   | Diera       | female | high             |
| Gertrude | Chaim      | female | high             |
| Abba  | Cuardall    | female | high             |
| Fanni  | Makelin     | female | high             |
| Ilise  | Sellmeyer   | female | high             |
| Alice  | Soule       | female | high             |
| Dianne | Baker       | female | high             |
| Casey  | Andry       | female | high             |
| Emilda | Geisler     | female | high             |
| Leora  | Sanderson   | female | high             |
| Thomasine | McCleish  | female | high             |
| Emilda | Geisler     | female | high             |
| Filia  | Boston     | female | high             |
| Rosamund | Rainer     | female | high             |
| Rosena | Faruchmont  | female | high             |
| Tressa | Dy          | female | high             |
| Bivott | Scranacher  | female | high             |
| Maxene | Emofschewals | female | high             |
| Max  | Gombhizer   | female | high             |
| Elzabeth | O' Roman   | female | high             |
| Ertini | Monroy      | female | high             |
| Becha  | Bysaaler    | female | high             |
| Estney | Rice        | female | high             |
| Mariette | Alexandrev  | female | high             |
| Pauletta | Taffee     | female | high             |
| Emma  | Capitano    | female | high             |
| Florin | Dunkerley   | female | high             |
| Tiffany | Fashow     | female | high             |
| Fanni  | Makelin     | female | high             |
| Antonia | Cardie     | female | high             |
| Marlie | Litter      | female | high             |
| Antonia | Cardie     | female | high             |
| Xenia  | Willmettt  | female | high             |
| Randee | Donherday   | female | high             |
| Sisely | Oppy        | female | high             |
| Jason  | Schalin     | female | high             |
| Brita  | Vlach       | female | high             |
| Flor  | Agnew       | female | high             |
| Thomasine | McCleish  | female | high             |
+-----+
```

5. Pegar todos os clientes masculino com cargo de TI que não possui carro e compraram bicicletas da classe 'low'.

```
select
dc.nome,
dc.categoria_cargo,
t.classe_produto,
dc.possui_carro
from demogra_cliente as dc
JOIN transacoes as t
ON dc.customer_id = t.customer_id
where categoria_cargo = 'TI' and possui_carro = 'No' and classe_produto = 'low' and sexo= 'Male';
```

```
-----
VERTICES      MODE      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container  SUCCEEDED    1         1         0         0         0         0
Map 2 ..... container  SUCCEEDED    1         1         0         0         0         0
-----
VERTICES: 02/02 [=====] 100% ELAPSED TIME: 9.87 s
-----
INFO  : Completed executing command(queryId=hive_20230828214708_4d68950d-76ce-4d63-be65-fcc661df94df); Time taken: 10.273 seconds
INFO  : OK
INFO  : Concurrency mode is disabled, not creating a lock manager
-----
+-----+-----+-----+-----+
| dc.nome | dc.categoria_cargo | t.classe_produto | dc.possui_carro |
+-----+-----+-----+-----+
| Jaymie  | IT                  | low              | No              |
| Gibby   | IT                  | low              | No              |
| Murdoch | IT                  | low              | No              |
| D'arcy  | IT                  | low              | No              |
| Garrek  | IT                  | low              | No              |
| Garrek  | IT                  | low              | No              |
| Victor  | IT                  | low              | No              |
| Wainwright | IT                | low              | No              |
| Iver    | IT                  | low              | No              |
| Wainwright | IT                | low              | No              |
| Mortimer | IT                 | low              | No              |
| Iver    | IT                  | low              | No              |
| Victor  | IT                  | low              | No              |
| Edmund  | IT                  | low              | No              |
| Gereh   | IT                  | low              | No              |
| Irwin   | IT                  | low              | No              |
| Yard    | IT                  | low              | No              |
| Yard    | IT                  | low              | No              |
| Yard    | IT                  | low              | No              |
| Tremain | IT                  | low              | No              |
| Bradley | IT                  | low              | No              |
| Bradley | IT                  | low              | No              |
| Bradley | IT                  | low              | No              |
| Neil    | IT                  | low              | No              |
| Colas   | IT                  | low              | No              |
+-----+-----+-----+-----+
25 rows selected (10.548 seconds)
0: jdbc:hive2://localhost:10000/default>
```

## 5. Proposta de evolução do trabalho

- Exibir um gráfico com a evolução das vendas por ano ou mês.
- Criar um dashboard para melhor visualização das perguntas analíticas.
- Aumentar o número de clusters para eficiência das consultas.

## **Link do trabalho com repositório no github:**

<https://github.com/Charles-Melo-TI/trabalho-hadoop>

## **Referências:**

<https://www.devmedia.com.br/hadoop-fundamentos-e-instalacao/29466>  
<https://www.kaggle.com/datasets/rohitsahoo/bicycle-store-dataset>

Rio de Janeiro, 28/08/2023