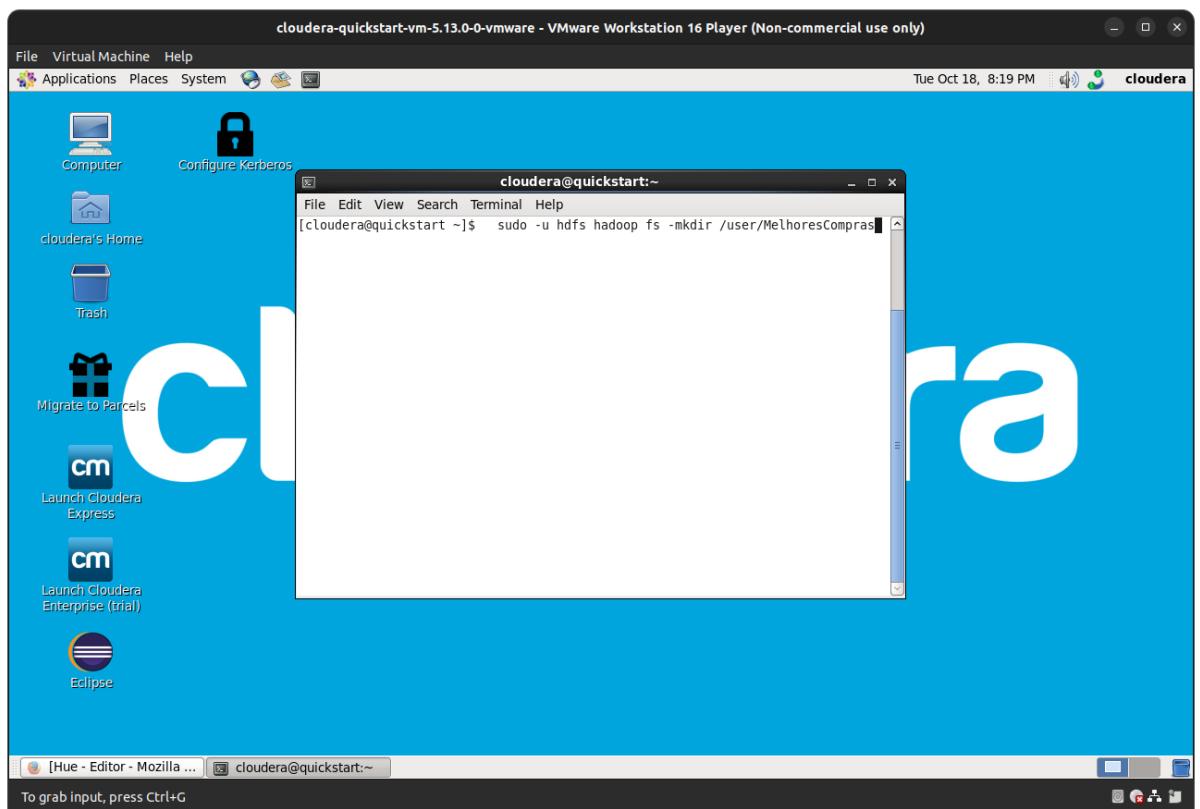


Perguntas práticas

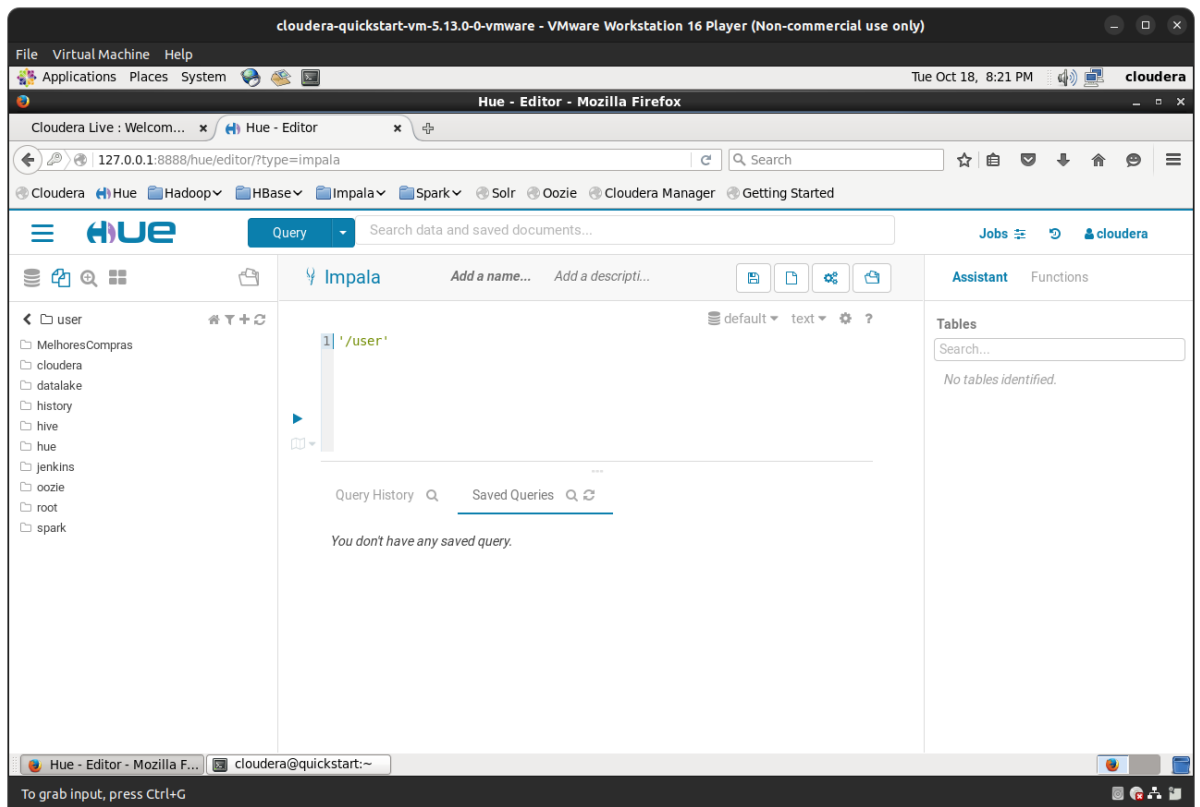
A) Dentro da VM Big Data, criar o diretório MelhoresCompras dentro do HDFS.

Execute o comando a seguir no terminal da VM Big Data:

```
sudo -u hdfs hadoop fs -mkdir /user/MelhoresCompras
```



Confira no Hue pelo Firefox se o diretório foi criado:

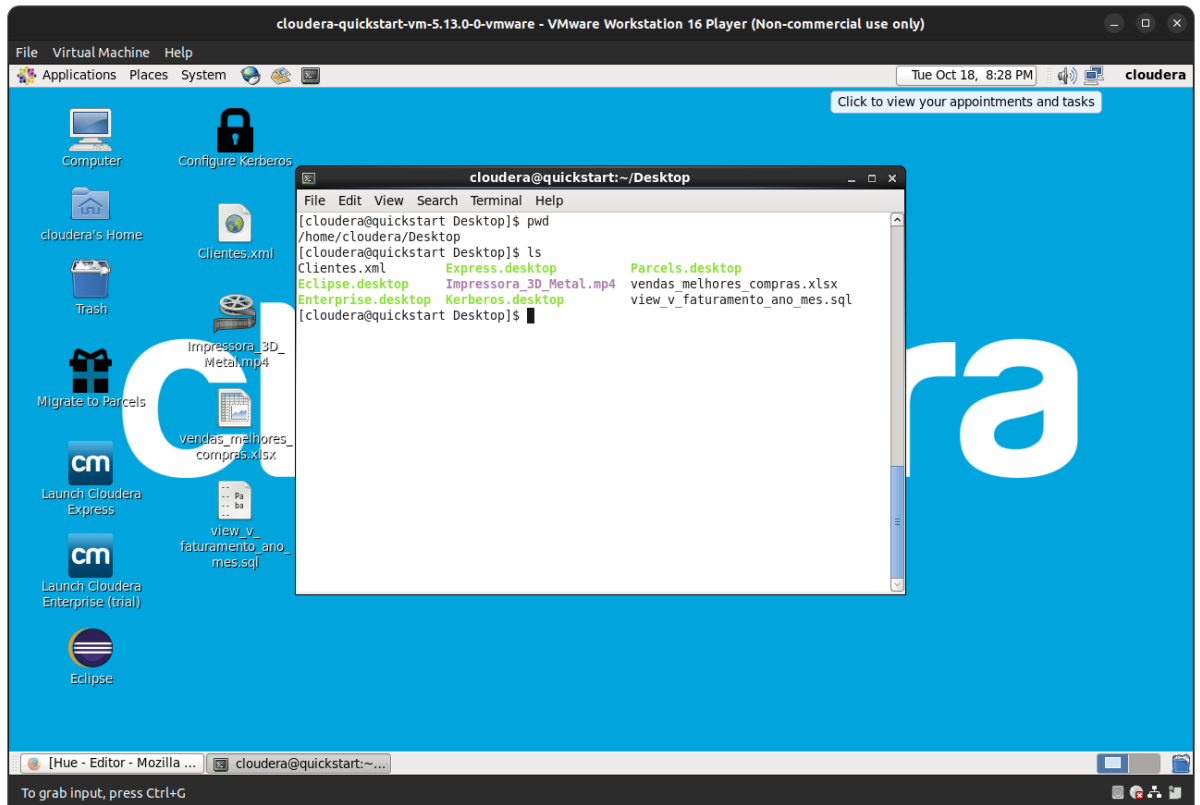


B) transferir os arquivos vendas_melhores_compras.xlsx, clientes.xml e Impressora_3D_Metal.mp4 para o diretório criado no item a (*não se esqueça de apresentar os comandos de transferência dos arquivos e a evidência de que os arquivos foram transferidos com sucesso).

1. Com os arquivos de Assets baixados localmente, vamos selecionar e arrastá-los para a janela do VMWare Player:



2. Após, teremos os arquivos presentes no caminho da VM
/home/cloudera/Desktop

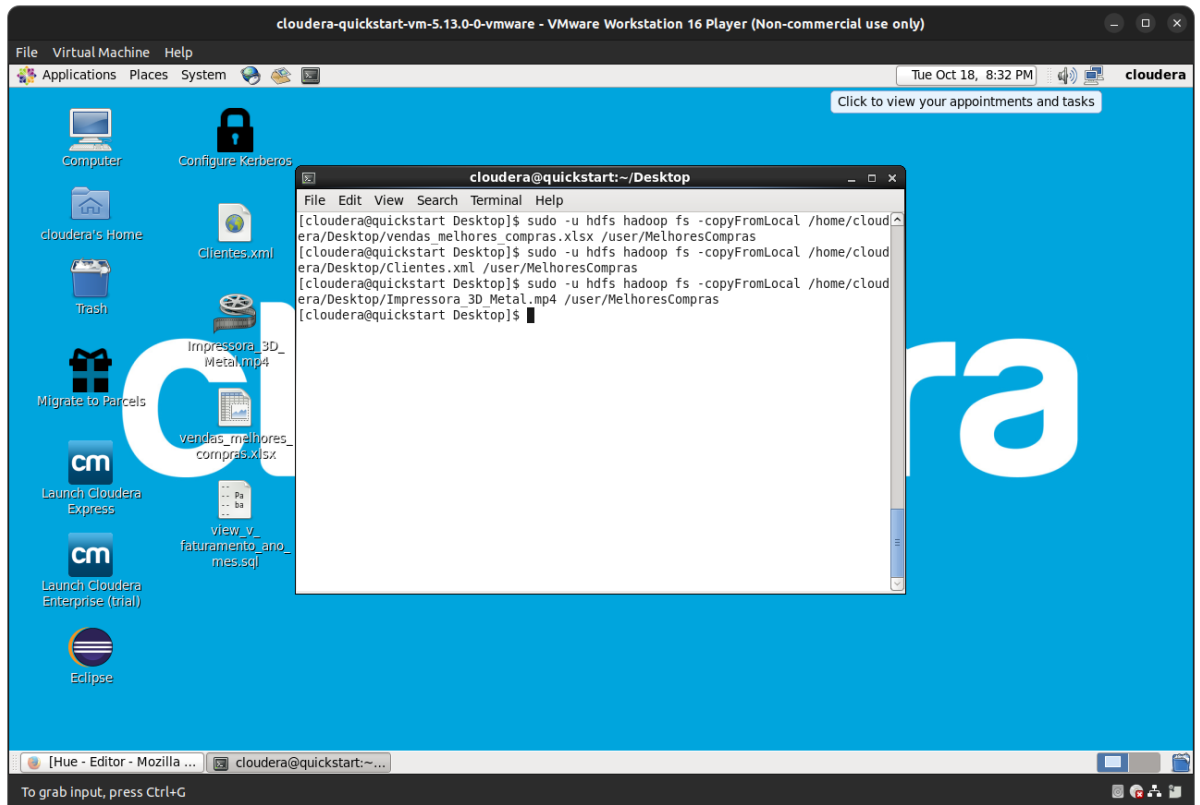


3. Vamos usar os seguintes comandos para mover os arquivos para
o diretório MelhoresCompras:

```
sudo -u hdfs hadoop fs -copyFromLocal /home/cloudera/Desktop/vendas_melhores_compras.xlsx  
/user/MelhoresCompras
```

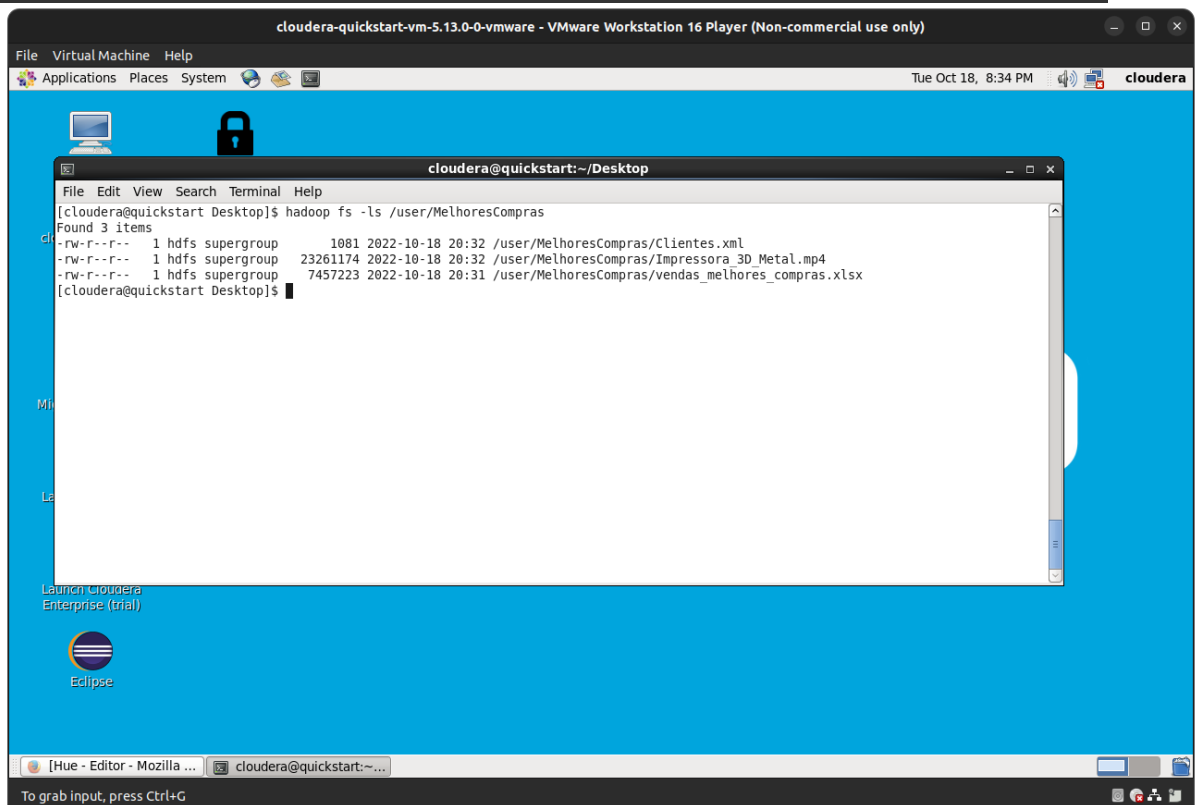
```
sudo -u hdfs hadoop fs -copyFromLocal /home/cloudera/Desktop/Clientes.xml  
/user/MelhoresCompras
```

```
sudo -u hdfs hadoop fs -copyFromLocal /home/cloudera/Desktop/Impressora_3D_Metal.mp4  
/user/MelhoresCompras
```

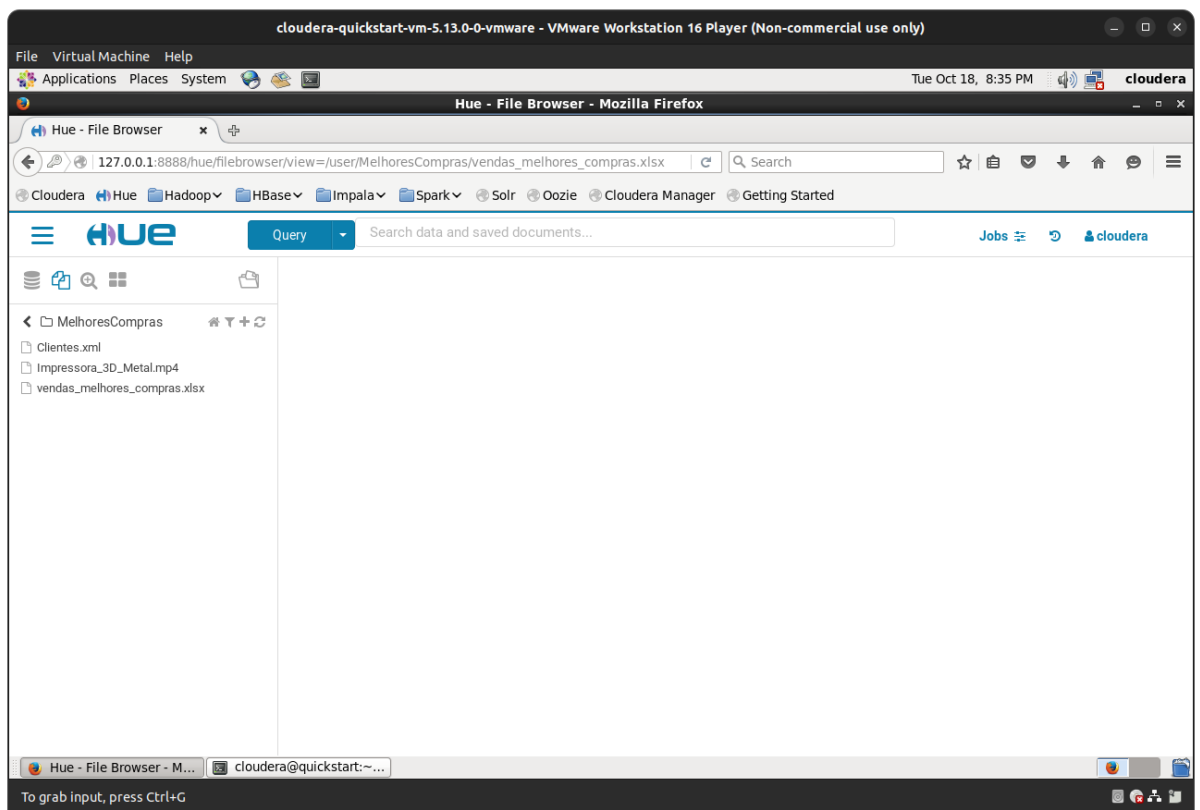


4. Podemos conferir a existência dos arquivos no diretório com o seguinte comando:

```
hadoop fs -ls /user/MelhoresCompras
```



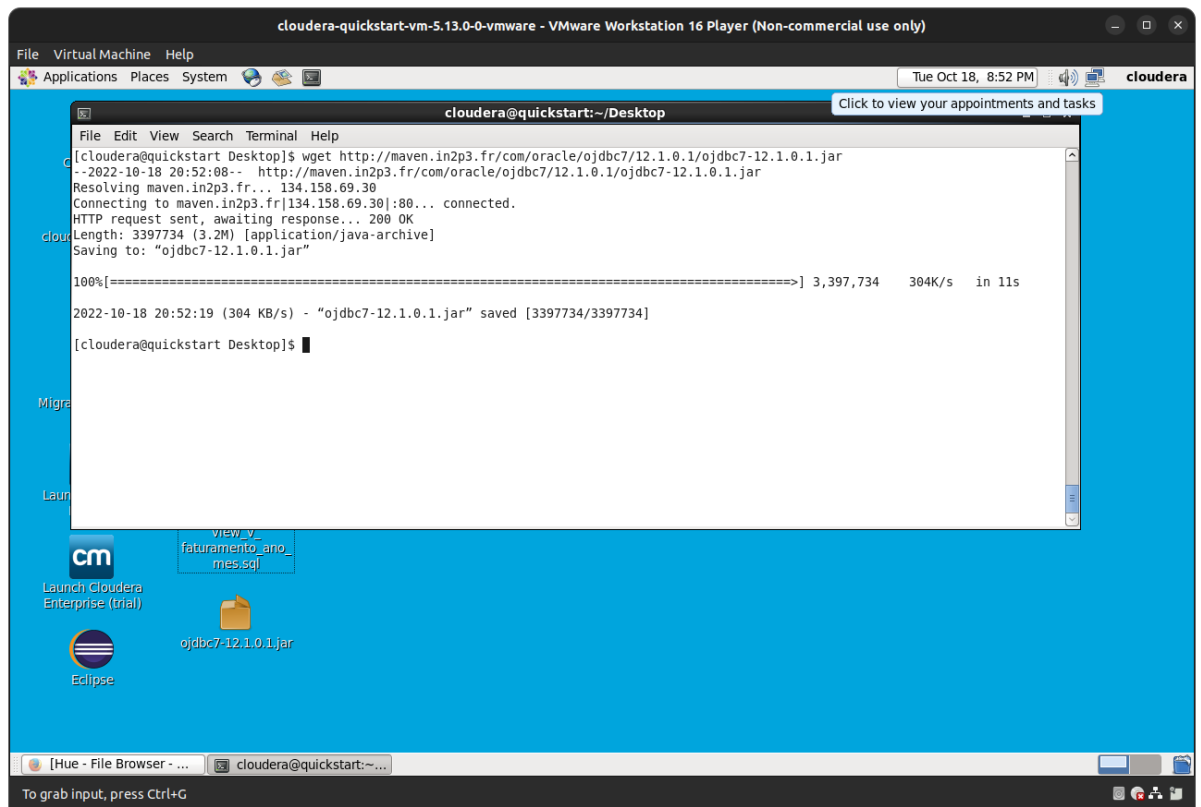
5. Também podemos conferir na interface do Hue:



C) Utilizando o sqoop, transferir os dados da view pf0110.v_faturamento_ano_mes para o HDFS.

1. Para iniciar, precisamos instalar os drivers que possibilitam o sqoop se conectar ao banco Oracle. Vamos baixar o driver OJDB usando o comando:

```
wget http://maven.in2p3.fr/com/oracle/ojdbc7/12.1.0.1/ojdbc7-12.1.0.1.jar
```

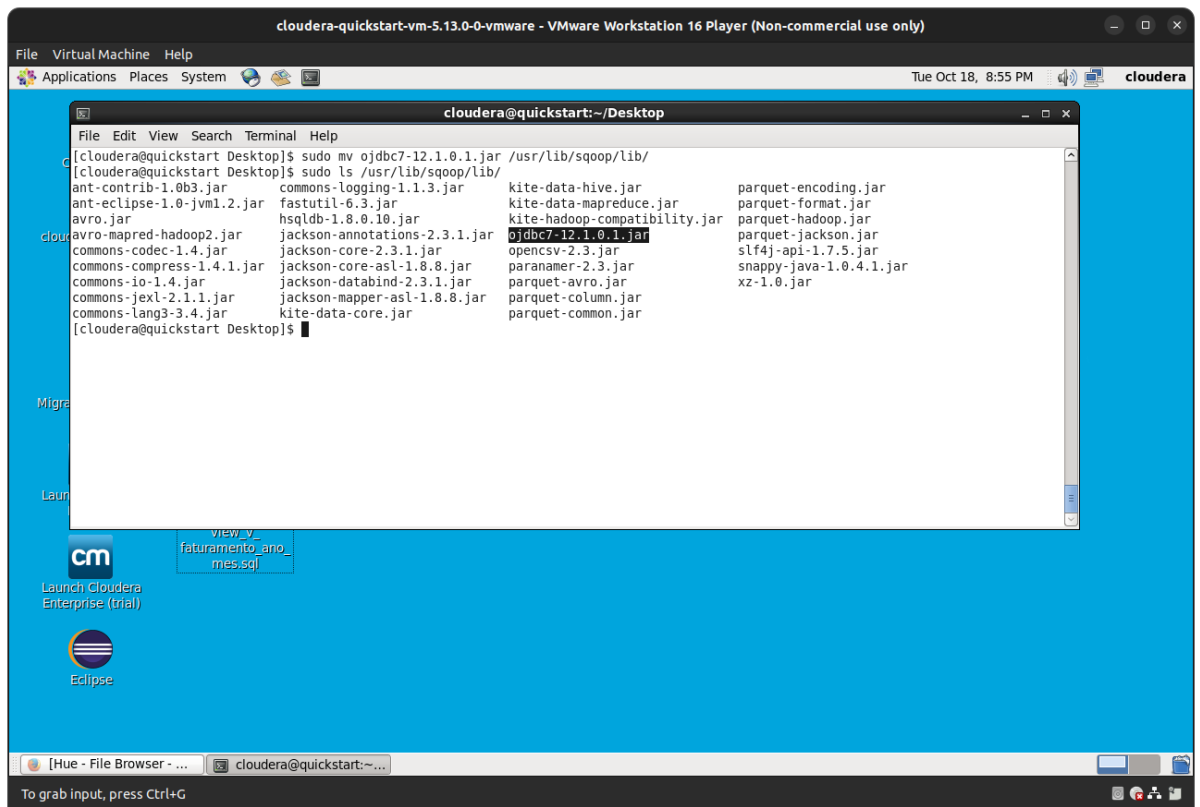


2. Iremos mover o arquivo do driver para o diretório da instalação do sqoop:

```
sudo mv ojdbc7-12.1.0.1.jar /usr/lib/sqoop/lib/
```

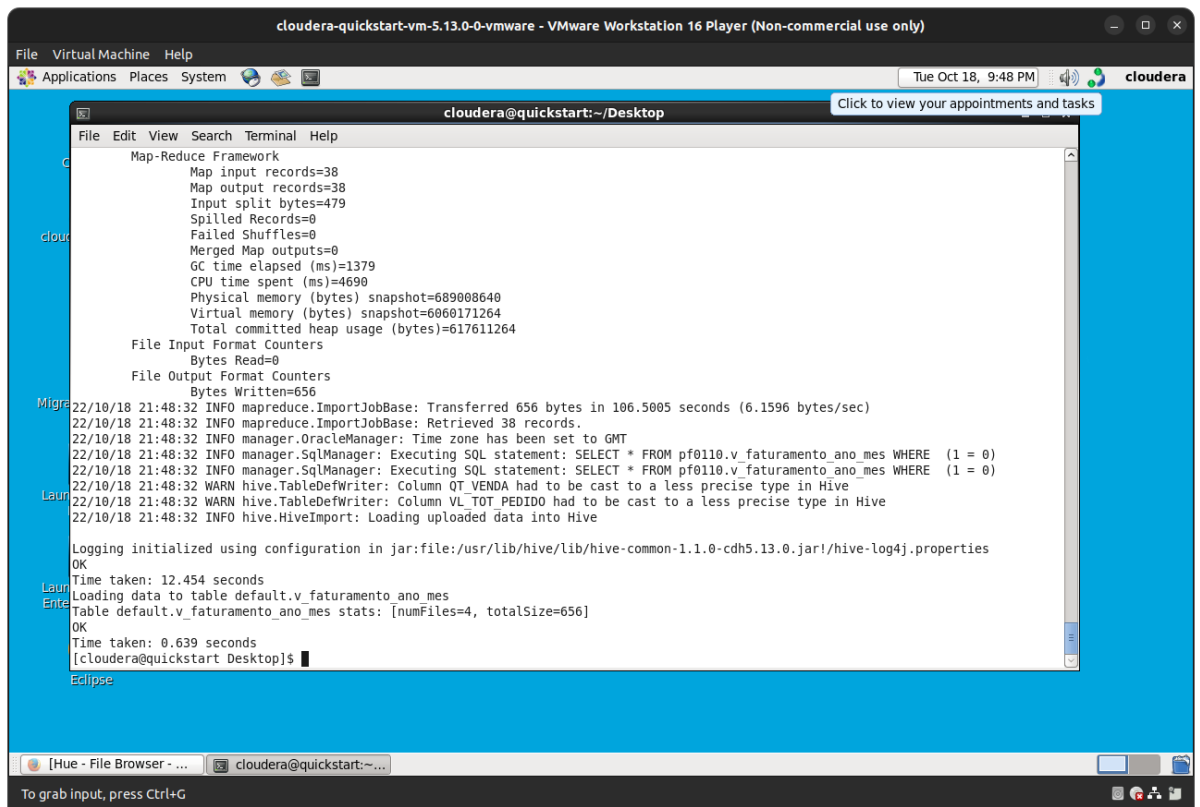
3. Podemos confirmar listando os arquivos no diretório:

```
sudo ls /usr/lib/sqoop/lib/
```



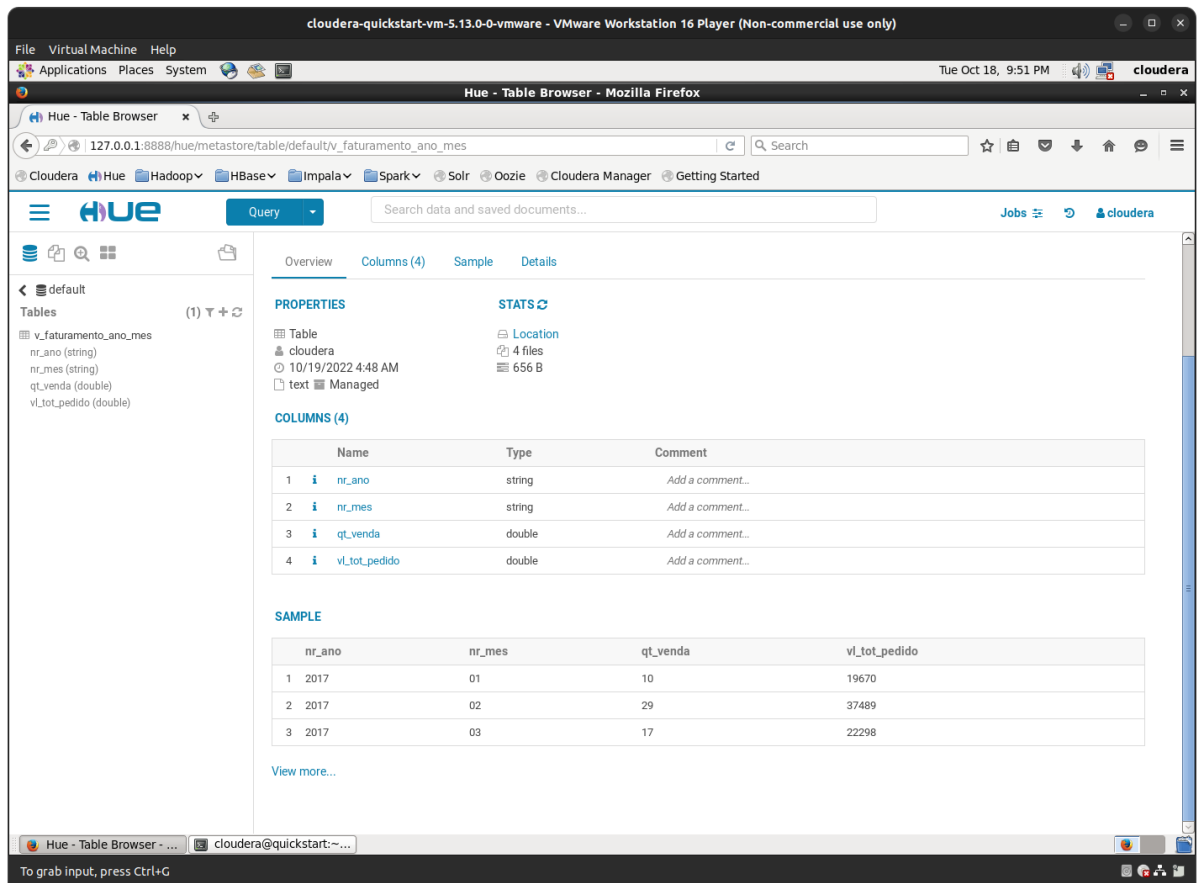
4. Vamos utilizar o seguinte comando para importar a view para nosso HDFS e garantir que a teremos como uma tabela no HIVE (usuário e senha foram mascarados):

```
sqoop import --connect jdbc:oracle:thin:@oracle.fiap.com.br:1521:orcl --username RM00000 --password 'DtNasc#000000' --query 'SELECT * FROM pf0110.v_faturamento_ano_mes WHERE $CONDITIONS' --target-dir /user/MelhoresCompras/v_faturamento_ano_mes --split-by NR_ANO --hive-import --hive-table v_faturamento_ano_mes
```



```
cloudera-quickstart-vm-5.13.0-0-vmware - VMware Workstation 16 Player (Non-commercial use only)
File Virtual Machine Help
Applications Places System
Tue Oct 18, 9:48 PM cloudera
cloudera@quickstart:~/Desktop
File Edit View Search Terminal Help
Map-Reduce Framework
  Map input records=38
  Map output records=38
  Input split bytes=479
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=1379
  CPU time spent (ms)=4690
  Physical memory (bytes) snapshot=689080640
  Virtual memory (bytes) snapshot=6060171264
  Total committed heap usage (bytes)=617611264
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=656
22/10/18 21:48:32 INFO mapreduce.ImportJobBase: Transferred 656 bytes in 106.5005 seconds (6.1596 bytes/sec)
22/10/18 21:48:32 INFO mapreduce.ImportJobBase: Retrieved 38 records.
22/10/18 21:48:32 INFO manager.OracleManager: Time zone has been set to GMT
22/10/18 21:48:32 INFO manager.SqlManager: Executing SQL statement: SELECT * FROM pf0110.v_faturamento_ano_mes WHERE (1 = 0)
22/10/18 21:48:32 INFO manager.SqlManager: Executing SQL statement: SELECT * FROM pf0110.v_faturamento_ano_mes WHERE (1 = 0)
22/10/18 21:48:32 WARN hive.TableDefWriter: Column QT_VENDA had to be cast to a less precise type in Hive
22/10/18 21:48:32 WARN hive.TableDefWriter: Column VL_TOT_PEDIDO had to be cast to a less precise type in Hive
22/10/18 21:48:32 INFO hive.HiveImport: Loading uploaded data into Hive
Logging initialized using configuration in jar:file:/usr/lib/hive/lib/hive-common-1.1.0-cdh5.13.0.jar!/hive-log4j.properties
OK
Time taken: 12.454 seconds
Loading data to table default.v_faturamento_ano_mes
Table default.v_faturamento_ano_mes stats: [numFiles=4, totalSize=656]
OK
Time taken: 0.639 seconds
[cloudera@quickstart Desktop]$
Eclipse
```

5. Pelo screenshot acima podemos perceber que o sqoop foi capaz de trazer 38 registros. Podemos confirmar que a view foi importada no HIVE a partir da interface visual do Hue:



Conclusão do Grupo

O desafio proposto na atividade contribuiu para entendermos que ao se trabalhar com Big Data são necessárias ferramentas específicas, que são preparadas para lidar com grandes volumes, velocidade e variedade de dados, de forma que toda e qualquer informação disponível nesses dados possa ser encontrada, analisada e aproveitada em tempo hábil, ou seja, com baixa latência.

A experiência de trabalharmos com algumas dessas ferramentas nos proporcionou um entendimento quanto a eficiência e importância das mesmas nas etapas de integração de dados. Dados esses, que podem ser originados de várias fontes.

Para além disso, essa experiência nos ajudou a refletir que quanto maior a capacidade de integrar os dados dessas várias fontes, maior será o valor agregado ao negócio, visto que essa integração possibilita uma combinação de informações que irão fomentar tomadas de decisões, busca por padrões e obtenção de insights.

Deste modo, entendemos que a integração de dados é fundamental para estruturar um ambiente Data-Driven, e temos percebido que a transformação digital trabalhada nessa fase e em fases anteriores, da empresa Melhores Compras, tem se desenvolvido para atingir esse objetivo, ou seja, uma empresa orientada a dados.