**1 - Iniciar os 5 serviços do Hadoop, por meio do comando /usr/local/hadoop/sbin/start-all.sh.**

igti@igti-VirtualBox:/usr/local/hadoop$ bin/sbin/start-all.sh

**2 - Por meio do comando -mkdir, criar um diretório chamado Desafio no HDFS.**

igti@igti-VirtualBox:/usr/local/hadoop$ bin/hdfs dfs -mkdir /Desafio

**3. Inserir no diretório Desafio, o arquivo covidData.txt. Esse arquivo se encontra no seguinte endereço do sistema de arquivos do sistema operacional da máquina virtual: /usr/local/hadoop/Dados. Para inserir o arquivo, utilize o comando –put.**

igti@igti-VirtualBox:/usr/local/hadoop$ bin/hdfs dfs -put /usr/local/hadoop/Dados/covidData.txt /Desafio

**4. Iniciar o Hive em /usr/local/hive/bin/hive. Se houver algum erro de Schema, seguir os passos para correção, apresentados no vídeo “Manipulação de dados com o Hive”.**

-> igti@igti-VirtualBox:/usr/local/hive$ rm -rf metastore\_db/

-> igti@igti-VirtualBox:/usr/local/hive$ bin/schematool -initSchema -dbType derby

-> igti@igti-VirtualBox:/usr/local/hive$ bin/hive

**5. Criar um *database* chamado dbDesafio, por meio do comando create database.**

hive> create database dbDesafio;

**6. Em seguida, por meio do comando create table, crie uma tabela chamada DadosCovid, que armazene os seguintes campos do arquivo covidData.txt: dataOcorrencia String; siglaPais String; descPais String; regiao String; novosCasos int; casosAcumulados int; novosObitos int; obitosAcumulados int**

hive> create table DadosCovid (dataOcorrencia STRING, siglaPais STRING, descPais STRING, regiao STRING, novosCasos INT, casosAcumulados INT, novosObitos INT, obitosAcumulados INT) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' STORED AS TEXTFILE LOCATION '/Desafio';

Verificar se a tabela foi criada: hive> select \* from DadosCovid;

**7. Faça a importação dos dados que estão no HDFS para a nova tabela, usando o comando LOAD DATA INPATH.**

hive> LOAD DATA INPATH 'hdfs://localhost:54310/Desafio/covidData.txt' INTO TABLE DadosCovid;

Teste: hive> select \* from DadosCovid;

**8. Execute uma sentença SQL que conte todos os registros da tabela DadosCovid. Para isso utilize a função count(\*) do SQL. Anote o resultado.**

hive> SELECT COUNT(\*) FROM DadosCovid; Resultado = 17705;

**9. Execute uma sentença SQL que verifique quantas comunicações os países Uruguay e Brazil fizeram cada um durante o período de apuração do arquivo. Lembre-se, cada linha do arquivo é uma comunicação. Para isso, utilize a cláusula where. Anote o resultado.**

hive> SELECT COUNT(\*) FROM DadosCovid WHERE descPais='Brazil'; Resultado = 92;

hive> SELECT COUNT(\*) FROM DadosCovid WHERE descPais='Uruguay'; Resultado = 74;

**10. Execute a seguinte sentença: select avg(novosCasos) from DadosCovid where descPais = “France”; Anote o resultado.**

hive> SELECT AVG(novosCasos) FROM DadosCovid WHERE descPais='France'; Resultado = 1141.632;

**11. Execute uma sentença que apure quantos novos casos e quantos novos óbitos foram comunicadas no dia 26/05/2020, considerando todos os países. Anote o resultado**.

hive> SELECT SUM(novosCasos) FROM DadosCovid WHERE dataOcorrencia LIKE '2020-05-26%'; Resultado = 99023;

hive> SELECT SUM(novosObitos) FROM DadosCovid WHERE dataOcorrencia LIKE '2020-05-26%'; Resultado = 1493;

**12. Execute o seguinte comando: describe extended DadosCovid. Copie o resultado apresentado em tela.**

hive> describe extended DadosCovid;

Resultado: OK

dataocorrencia string

siglapais string

descpais string

regiao string

novoscasos int

casosacumulados int

novosobitos int

obitosacumulados int

Detailed Table Information Table(tableName:dadoscovid, dbName:default, owner:igti, createTime:1591667439, lastAccessTime:0, retention:0, sd:StorageDescriptor(cols:[FieldSchema(name:dataocorrencia, type:string, comment:null), FieldSchema(name:siglapais, type:string, comment:null), FieldSchema(name:descpais, type:string, comment:null), FieldSchema(name:regiao, type:string, comment:null), FieldSchema(name:novoscasos, type:int, comment:null), FieldSchema(name:casosacumulados, type:int, comment:null), FieldSchema(name:novosobitos, type:int, comment:null), FieldSchema(name:obitosacumulados, type:int, comment:null)], location:hdfs://localhost:54310/Desafio, inputFormat:org.apache.hadoop.mapred.TextInputFormat, outputFormat:org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat, compressed:false, numBuckets:-1, serdeInfo:SerDeInfo(name:null, serializationLib:org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe, parameters:{serialization.format=,, line.delim=\n, field.delim=,}), bucketCols:[], sortCols:[], parameters:{}, skewedInfo:SkewedInfo(skewedColNames:[], skewedColValues:[], skewedColValueLocationMaps:{}), storedAsSubDirectories:false), partitionKeys:[], parameters:{transient\_lastDdlTime=1591667957, bucketing\_version=2, totalSize=929317, numFiles=1}, viewOriginalText:null, viewExpandedText:null, tableType:MANAGED\_TABLE, rewriteEnabled:false, catName:hive, ownerType:USER)

Time taken: 0.255 seconds, Fetched: 10 row(s)

**13. Execute a sentença: select concat(dataOcorrencia, “ “, siglaPais, “ “, descPais) from DadosCovid where novosCasos = 501; Anote o resultado.**

16 - hive> SELECT CONCAT(dataOcorrencia, " ", siglaPais, " ", descPais) FROM DadosCovid WHERE novosCasos=501;

Resultado: 2020-03-11T00:00:00Z ES Spain

**14. Execute a sentença: select região, count(1) from DadosCovid group by regiao order by regiao; Anote os resultados.**

hive> SELECT regiao, COUNT(1) FROM DadosCovid GROUP BY regiao ORDER BY regiao;

Resultado:

OK

118

AFRO 3417

AMRO 4079

EMRO 1816

EURO 5431

SEARO 970

WPRO 1874