**Procedimentos do Enunciado**

1. Iniciar os 5 serviços do Hadoop

**/usr/local/hadoop/sbin/start-all.sh**

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

**/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.

**/usr/local/hadoop/bin/hdfs dfs -put /usr/local/hadoop/Dados/covidData.txt /Desafio**

**/usr/local/hadoop/bin/hdfs dfs -ls /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”.

Seguir o video **Manipulando dados com o Hive**.

**/usr/local/hadoop/bin/hdfs dfs -mkdir /hive**

**cd /usr/local/hive**

**rm -rf metastore\_db**

**bin/schematool -initSchema -dbType derby**

**/usr/local/hive/bin/hive**

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

**create database dbDesafio;**

6. Crie uma tabela chamada DadosCovid

**use dbDesafio;**

**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';**

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

**LOAD DATA INPATH '/Desafio/covidData.txt' INTO TABLE DadosCovid;**

8. Execute uma sentença SQL que conte todos os registros da tabela DadosCovid.

**select count(\*) from dadoscovid;**

**Query ID = root\_20200813031414\_4bd04d68-5ffa-4ef7-99cf-5e6e3360d7dd**

**Total jobs = 1**

**Launching Job 1 out of 1**

**Number of reduce tasks determined at compile time: 1**

**In order to change the average load for a reducer (in bytes):**

**set hive.exec.reducers.bytes.per.reducer=<number>**

**In order to limit the maximum number of reducers:**

**set hive.exec.reducers.max=<number>**

**In order to set a constant number of reducers:**

**set mapreduce.job.reduces=<number>**

**Job running in-process (local Hadoop)**

**2020-08-13 03:14:48,550 Stage-1 map = 100%, reduce = 100%**

**Ended Job = job\_local1057330258\_0001**

**MapReduce Jobs Launched:**

**Stage-Stage-1: HDFS Read: 1858648 HDFS Write: 0 SUCCESS**

**Total MapReduce CPU Time Spent: 0 msec**

**OK**

**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

**select descpais,count(\*) from dadoscovid where descpais in ('Brazil','Uruguay') group by descpais;**

**Query ID = root\_20200813031717\_45932ba9-8cba-4fdb-875c-795f42939893**

**Total jobs = 1**

**Launching Job 1 out of 1**

**Number of reduce tasks not specified. Estimated from input data size: 1**

**In order to change the average load for a reducer (in bytes):**

**set hive.exec.reducers.bytes.per.reducer=<number>**

**In order to limit the maximum number of reducers:**

**set hive.exec.reducers.max=<number>**

**In order to set a constant number of reducers:**

**set mapreduce.job.reduces=<number>**

**Job running in-process (local Hadoop)**

**2020-08-13 03:17:59,149 Stage-1 map = 100%, reduce = 100%**

**Ended Job = job\_local766509246\_0002**

**MapReduce Jobs Launched:**

**Stage-Stage-1: HDFS Read: 3733666 HDFS Write: 0 SUCCESS**

**Total MapReduce CPU Time Spent: 0 msec**

**OK**

**Brazil 92**

**Uruguay 74**

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

**select avg(novosCasos) from DadosCovid where descPais = "France";**

**Query ID = root\_20200813032020\_3f879fa7-761c-4b49-b175-9680bab2db7b**

**Total jobs = 1**

**Launching Job 1 out of 1**

**Number of reduce tasks determined at compile time: 1**

**In order to change the average load for a reducer (in bytes):**

**set hive.exec.reducers.bytes.per.reducer=<number>**

**In order to limit the maximum number of reducers:**

**set hive.exec.reducers.max=<number>**

**In order to set a constant number of reducers:**

**set mapreduce.job.reduces=<number>**

**Job running in-process (local Hadoop)**

**2020-08-13 03:20:16,479 Stage-1 map = 100%, reduce = 100%**

**Ended Job = job\_local191167568\_0003**

**MapReduce Jobs Launched:**

**Stage-Stage-1: HDFS Read: 5592300 HDFS Write: 0 SUCCESS**

**Total MapReduce CPU Time Spent: 0 msec**

**OK**

**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.

**select sum(novoscasos),sum(novosobitos) from dadoscovid where dataocorrencia="2020-05-26T00:00:00Z";**

**Query ID = root\_20200813032323\_d64a90cd-2e98-43e1-82b2-ac6c4a3d74f0**

**Total jobs = 1**

**Launching Job 1 out of 1**

**Number of reduce tasks determined at compile time: 1**

**In order to change the average load for a reducer (in bytes):**

**set hive.exec.reducers.bytes.per.reducer=<number>**

**In order to limit the maximum number of reducers:**

**set hive.exec.reducers.max=<number>**

**In order to set a constant number of reducers:**

**set mapreduce.job.reduces=<number>**

**Job running in-process (local Hadoop)**

**2020-08-13 03:23:07,545 Stage-1 map = 100%, reduce = 100%**

**Ended Job = job\_local1971672529\_0004**

**MapReduce Jobs Launched:**

**Stage-Stage-1: HDFS Read: 7450934 HDFS Write: 0 SUCCESS**

**Total MapReduce CPU Time Spent: 0 msec**

**OK**

**99023 1493**

12. Execute o seguinte comando: describe extended DadosCovid

**describe extended DadosCovid;**

**OK**

**dataocorrencia string**

**siglapais string**

**descpais string**

**regiao string**

**novoscasos int**

**casosacumulados int**

**novosobitos int**

**obitosacumulados int**

**Detailed Table Information Table(tableName:dadoscovid, dbName:dbdesafio, owner:root, createTime:1597288257, 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://0.0.0.0:8020/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=,, field.delim=,, line.delim=**

**}), bucketCols:[], sortCols:[], parameters:{}, skewedInfo:SkewedInfo(skewedColNames:[], skewedColValues:[], skewedColValueLocationMaps:{}), storedAsSubDirectories:false), partitionKeys:[], parameters:{numFiles=0, transient\_lastDdlTime=1597288405, COLUMN\_STATS\_ACCURATE=true, totalSize=0, numRows=0, rawDataSize=0}, viewOriginalText:null, viewExpandedText:null, tableType:MANAGED\_TABLE)**

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

**select concat(dataOcorrencia, " ", siglaPais, " ", descPais) from DadosCovid where novosCasos = 501;**

**OK**

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

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

**select regiao, count(1) from DadosCovid group by regiao order by regiao;**

**Query ID = root\_20200813032626\_fb409c30-24c8-44cc-83f7-356e8f66c6c4**

**Total jobs = 2**

**Launching Job 1 out of 2**

**Number of reduce tasks not specified. Estimated from input data size: 1**

**In order to change the average load for a reducer (in bytes):**

**set hive.exec.reducers.bytes.per.reducer=<number>**

**In order to limit the maximum number of reducers:**

**set hive.exec.reducers.max=<number>**

**In order to set a constant number of reducers:**

**set mapreduce.job.reduces=<number>**

**Job running in-process (local Hadoop)**

**2020-08-13 03:26:53,980 Stage-1 map = 100%, reduce = 100%**

**Ended Job = job\_local383883991\_0005**

**Launching Job 2 out of 2**

**Number of reduce tasks determined at compile time: 1**

**In order to change the average load for a reducer (in bytes):**

**set hive.exec.reducers.bytes.per.reducer=<number>**

**In order to limit the maximum number of reducers:**

**set hive.exec.reducers.max=<number>**

**In order to set a constant number of reducers:**

**set mapreduce.job.reduces=<number>**

**Job running in-process (local Hadoop)**

**2020-08-13 03:26:55,301 Stage-2 map = 100%, reduce = 100%**

**Ended Job = job\_local665818614\_0006**

**MapReduce Jobs Launched:**

**Stage-Stage-1: HDFS Read: 11168202 HDFS Write: 0 SUCCESS**

**Stage-Stage-2: HDFS Read: 11168202 HDFS Write: 0 SUCCESS**

**Total MapReduce CPU Time Spent: 0 msec**

**OK**

**118**

**AFRO 3417**

**AMRO 4079**

**EMRO 1816**

**EURO 5431**

**SEARO 970**

**WPRO 1874**

**Pergunta 1**

Quantos registros existem na tabela DadosCovid, criada no Hive?

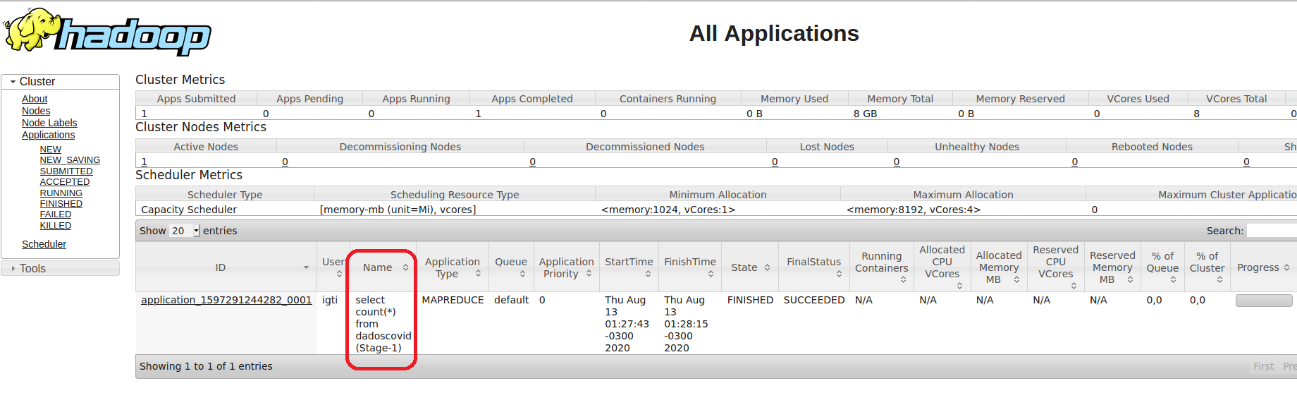
Resposta no passo 8.

**R.: 17705**

**Pergunta 2**

Ao realizar a contagem de registros na tabela DadosCovid, o Hive instanciou um job MapReduce. Qual foi o nome dado a esse job? Para descobrir, acesse a ferramenta administrativa do Hadoop.

Após o passo 8 acessar <http://localhost:8088/cluster/apps>



**R.: select count(\*) from DadosCovid (Stage-1).**

**Pergunta 3**

Quantas comunicações foram realizadas pelo país Uruguay?

Resposta no passo 9

**R.: 74.**

**Pergunta 4**

Quantas comunicações foram realizadas pelo país Brazil? Quantas comunicações foram realizadas pelo país Brazil?

Resposta no passo 9

**R.: 92.**

**Pergunta 5**

Ao inserir o comando:

select avg(novosCasos) from DadosCovid where descPais = “France”;

Qual é o resultado apresentado pelo Hive?

Resposta no passo 10

**R.: 1141.632.**

**Pergunta 6**

Considerando todos os países, quantos novos casos foram informados no dia

26/05/2020?

Resposta no passo 11

**R.: 99023.**

**Pergunta 7**

Considerando todos os países, quantos novos óbitos foram informados no dia

26/05/2020?

Resposta no passo 11

**R.: 1493.**

**Pergunta 8**

O que o Hive faz quando executa a sentença: select count(1) from DadosCovid?.

**select count(1) from DadosCovid;**

**Query ID = root\_20200813050202\_e3df674b-736e-440e-b29f-4e1a8cc90376**

**Total jobs = 1**

**Launching Job 1 out of 1**

**Number of reduce tasks determined at compile time: 1**

**In order to change the average load for a reducer (in bytes):**

**set hive.exec.reducers.bytes.per.reducer=<number>**

**In order to limit the maximum number of reducers:**

**set hive.exec.reducers.max=<number>**

**In order to set a constant number of reducers:**

**set mapreduce.job.reduces=<number>**

**Job running in-process (local Hadoop)**

**2020-08-13 05:02:26,559 Stage-1 map = 100%, reduce = 100%**

**Ended Job = job\_local766133032\_0001**

**MapReduce Jobs Launched:**

**Stage-Stage-1: HDFS Read: 1858648 HDFS Write: 0 SUCCESS**

**Total MapReduce CPU Time Spent: 0 msec**

**OK**

**17705**

**R.: Retorna o número total de registros, após executar funções Map e Reduce.**

**Pergunta 9**

É correto afirmar que, quando utilizamos o comando describe extended

DadosCovid, os seguintes dados são apresentados na tela:

Resposta no passo 12

**R.: line.delim e field.delim.**

**Pergunta 10**

Marque a alternativa que melhor representa o resultado do seguinte comando:

select concat(dataOcorrencia, “ “, siglaPais, “ “, descPais) from DadosCovid

where novosCasos = 501;

Resposta no passo 13

**R.: 2020-03-11T00:00:00Z ES Spain.**

**Pergunta 11**

São características do Apache Hive, EXCETO:

**R.: É uma ferramenta construída para transferir dados entre o cluster Hadoop e um banco de dados SQL**

**Pergunta 12**

Quantos registros de comunicação foram feitos para a região EURO?

Resposta no passo 14

**R.: 5431.**

**Pergunta 13**

Marque a opção que representa o número de jobs MapReduce que foram criados ao executar a sentença:

select região, count(1) from DadosCovid group by regiao order by regiao;

Resposta no passo 14

**MapReduce Jobs Launched:**

**Stage-Stage-1: HDFS Read: 11168202 HDFS Write: 0 SUCCESS**

**Stage-Stage-2: HDFS Read: 11168202 HDFS Write: 0 SUCCESS**

**R.: 2.**

**Pergunta 14**

Marque a alternativa que NÃO representa comandos do Hive:

**R.: HDFS e YARN.**