

TIAGO DALTRO DUARTE – RESPOSTAS EXERCICIOS HIVE

docker-compose up -d

1 – docker exec -it hive-server bash

Mkdir -p /home/input/dados

Curl -O

https://raw.githubusercontent.com/caiuafranca/dados_curso/main/cursos.csv

2 – No diretório root@hive_server:/opt# foi criado um diretório HDFS:

Hdfs dfs -mkdir -p user/aluno/dados/curso/

Depois ir ao diretório que está o arquivo cursos.csv e movê-lo para o diretório hdfs criado:

hdfs dfs -moveFromLocal curso.csv /user/aluno/dados/curso

3 – beeline -u jdbc:hive2://localhost:10000

show databases;

create database treinamento;

4 – CREATE EXTERNAL TABLE IF NOT EXISTS cursos_stg(

`id_curso` int,

`id_unidade` int,

`codigo` int,

`nome` string,

`nivel` string,

`id_modalidade_educacao` int,

`id_municipio` int,

`id_tipo_oferta_curso` int,

`id_area_curso` int,

`id_grau_academico` int,

`id_eixo_conhecimento` int,

`ativo` string)

COMMENT 'Tabela Externa de Cursos'

row format delimited

FIELDS TERMINATED BY ','

STORED AS TEXTFILE

LOCATION '/user/aluno/dados/curso/'

5 – select * from cursos_stg; (DICIDIR NÃO PRINT DA TABELA PORQUE FICOU MUITO GRANDE).

6 – COMANDO PARA VER AS 5 PRIMEIRAS LINHAS:

```
SELECT * FROM cursos_stg LIMIT 5;
```

7 – select count(*) from cursos_stg;

8 – create table cursos(

 `id_curso` int,

 `id_unidade` int,

 `codigo` int,

 `nome` string,

 `nivel` string,

 `id_modalidade_educacao` int,

 `id_municipio` int,

 `id_tipo_oferta_curso` int,

 `id_area_curso` int,

 `id_grau_academico` int,

 `id_eixo_conhecimento` int,

 `ativo` string)

PARTITIONED BY (dt_foto STRING)

ROW FORMAT SERDE 'org.apache.hadoop.hive.ql.io.orc.OrcSerde'

STORED AS INPUTFORMAT

'org.apache.hadoop.hive.ql.io.orc.OrcInputFormat'

OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.orc.OrcOutputFormat'

TBLPROPERTIES ('orc.compress'='SNAPPY');

SET hive.exec.dynamic.partition=true;

SET hive.exec.dynamic.partition.mode=nonstrict;

9 –

10 –

11 –