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;

INSERT OVERWRITE TABLE treinamento.cursos

PARTITION(dt\_foto)

SELECT

id\_curso,

id\_unidade,

codigo,

nome,

nivel,

id\_modalidade\_educacao,

id\_municipio,

id\_tipo\_oferta\_curso,

id\_area\_curso,

id\_grau\_academico,

id\_eixo\_conhecimento,

ativo,

'21052022' as dt\_foto

from treinamento.cursos\_stg;

9 – show partitions treinamento.cursos;

10 – select count(\*) from treinamento.cursos;

11 – select count(\*) from treinamento.cursos where ativo = 1;