**PASSO A PASSO**

HDFS

hdfs dfs -mkdir /gas/

hdfs dfs -put gas.csv /gas/gas.csv

hdfs dfs -setrep 3 /gas/gas.csv

HIVE

beeline

!connect jdbc:hive2://

create schema gas;

use gas;

CREATE EXTERNAL TABLE GAS\_TOTAL

(

regiao string,

estado string,

produto string,

unidade\_medida string,

preco\_medio\_revenda double,

preco\_minimo\_revenda double,

preco\_maximo\_revenda double,

mes int,

ano int

)

ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.OpenCSVSerde'

WITH SERDEPROPERTIES (

"separatorChar" = ",",

"quoteChar" = "\"",

"escapeChar" = "\\"

)

STORED AS TextFile

LOCATION '/gas/'

TBLPROPERTIES("skip.header.line.count" = "1");

create table gas\_orc like gas\_total stored as orcfile;

insert into gas\_orc select \* from gas\_total;

PIG

pig -x mapreduce

gas\_min = LOAD '/user/cloudera/gasmin/' as (produto:chararray, preco:float, estado:chararray, ano:int);

dim10 = LIMIT gas\_min 10;

dump dim10;

gas\_etanol\_pe = FILTER gas\_min BY produto == 'ETANOL HIDRATADO' AND estado == 'PERNAMBUCO';

dump gas\_etanol\_pe

STORE gas\_etanol\_pe INTO '/user/cloudera/pig/' USING PigStorage (',');