

Como realizar consultas de maneira simples no ambiente complexo de Big Data com Hive e Impala

• HIVE

Sair do modo safemode do Hive

```
sudo -u hdfs hadoop dfsadmin -safemode leave
```

Acessar o hive

```
hive
```

Mostra os bancos de dados

```
show databases;
```

Coloca o banco de dados em uso

```
use bancoteste;
```

Criando um database

```
create database bancoteste;
```

```
create database teste if exists teste;
```

Cria tabela ("bancoteste." não é obrigatorio, porem precisa por o banco que deseja em uso)

```
create table bancoteste.tabela (id int);
```

Cria uma tabela do tipo external

```
create table external tabela (id int);
```

Mostra todas tabelas criadas em um banco de dados

show tables;

Mostra os headers da tabela

set hive.cli.print.header=true;

Mostra qual banco de dados esta em uso

set hive.cli.print.current.db=true;

Descreve a tabela, nome e tipo de dados

desc tabela

Inserir registro

insert into table teste values("registro");

Local onde esta armazenado as tabelas e banco de dados do hive

hdfs dfs -ls /user/hive/warehouse

- Cria tabela TB_EXT_EMPLOYEE

CREATE EXTERNAL TABLE TB_EXT_EMPLOYEE(

id STRING,

groups STRING,

age STRING,

active_lifestyle STRING,

salary STRING)

ROW FORMAT DELIMITED FIELDS

TERMINATED BY '\;'

STORED AS TEXTFILE

LOCATION '/user/hive/warehouse/external/tableas/employee'

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

Pega o arquivo employee.txt e coloca no hdfs

```
hdfs dfs -put /home/everis/employee.txt /user/hive/warehouse/everis/employee
```

Muda os privilegios da pasta no hdfs

```
hdfs dfs -chmod 775 /home/everis/employee.txt  
/user/hive/warehouse/everis/employee
```

Muda os privilegios do arquivo no hdfs

```
hdfs dfs -chmod 775 /home/everis/employee.txt  
/user/hive/warehouse/everis/employee/employee.txt
```

Visualizo o arquivo no hdfs

```
hdfs dfs -cat /home/everis/employee.txt  
/user/hive/warehouse/everis/employee/employee.txt
```

Copio arquivo do hdfs para maquina local

```
hdfs dfs -copyToLocal /user/hive/warehouse/everis/employee/employee.txt
```

Faz um select count na tabela

```
select count(*) from TB_EXT_EMPLOYEE;
```

- Cria uma tabela particionada

```
CREATE EXTERNAL TABLE TB_EMPLOYEE(  
id INT,  
groups STRING,  
age INT,  
active_lifestyle STRING,  
salary DOUBLE)  
PARTITIONED BY(dt_processamento STRING)  
ROW FORMAT DELIMITED FIELDS TERMINATED BY '|'   
STORED AS PARQUET TBLPROPERTIES ("parquet.compression"="SNAPPY");
```

- Insere registros baseado em um select

```
insert into table TB_EMPLOYEE partition (dt_processamento='20201118')
```

```
select
```

```
id,
```

```
groups,
```

```
age,
```

```
active_lifestyle,
```

```
salary
```

```
from TB_EXT_EMPLOYEE;
```

- Cria tabela external chamada localidade

```
CREATE EXTERNAL TABLE localidade(
```

```
street string,
```

```
city string,
```

```
zip string,
```

```
state string,
```

```
beds string,
```

```
baths string,
```

```
sq_ft string,
```

```
type string,
```

```
sale_date string,
```

```
price string,
```

```
latitude string,
```

```
longitude string)
```

```
PARTITIONED BY (particao STRING)
```

```
ROW FORMAT DELIMITED FIELDS TERMINATED BY ","
```

```
STORED AS TEXTFILE
```

```
location '/user/hive/warehouse/external/tabelas/localidade'
```

```
tblproperties ("skip.header.line.count"="1");
```

- Carrega tabela a partir de arquivo na maquina local

```
load data local inpath '/home/everis/base_localidade.csv'
```

into table teste.localidade partition (particao='2021-01-21');

- Cria tabela melhor formatada

CREATE TABLE tb_localidade_parquet(

street string,

city string,

zip string,

state string,

beds string,

baths string,

sq_ft string,

type string,

sale_date string,

price string,

latitude string,

longitude string)

PARTITIONED BY (PARTICAO STRING)

STORED AS PARQUET;

- Insere na tabela tb_localidade_parquet baseado na tabela localidade

INSERT into TABLE tb_localidade_parquet

PARTITION(PARTICAO='01')

SELECT

street,

city,

zip,

state,

beds,

baths,

sq_ft,

```
type,  
sale_date,  
price,  
latitude,  
longitude  
FROM localidade;
```

Executando um JOIN entre tabelas

```
select  
tb01.id,  
tb02.zip  
from tb_ext_employee tb01  
full outer join tb_localidade_parquet tb02  
on tb01.id = tb02.zip;
```

- Limpa terminal hive (o ! permite chamar um comando linux na sessão do hive)
!clear;

Acessa o Hive no modo silencioso (executa algo sem entrar no hive em si e sai em seguida)

```
hive -S -e "select count(*) from banco.tabela;"
```

• IMPALA

Acessar o Impala

```
impala-shell
```

Caso não apareça o banco de dados ou tabelas. Invalidar metadata antigo armazenado no Impala

```
INVALIDATE METADATA banco.tabela;
```

Lista os bancos de dados

show databases;

Coloca banco de dados em uso

use bancodedados;

Lista as tabelas

show tables;

• Anotações

Mostra algumas funções/comandos do hadoop

hadoop -h

Exibe o manual do hadoop

man hadoop

Mostra algumas funções/comandos do HDFS

hdfs -h

Exibe o manual do HDFS

man hdfs