Aula 3 - Hive - Prática

Lucio Monteiro

Ambiente Hadoop - Exercício 1

1) Entrar no site https://databricks.com//
2) Crie um login - Versão gratuita - Community Edition
3) Clique em criar -> Cluster -> Escolher um nome e esperar executar
4) Criar um arquivo com estas dados:

```
id,name,age
1,jose,10
2,maria,10
3,joao,11
4,antonio,11
5,helio,15
6,dan,20
7,ze,20
```

5) Salvar e fazer upload no databricks → Create -> Table -> Upload File
6) Verificar o nome que foi feito o upload.

Ambiente Hadoop - Exercício 1

1) Verificar se criou a tabela: display(dbutils.fs.ls("/FileStore/tables/"))
2) show databases
3) desc database default
4) show tables
5) select * from NOMETABELA
6) create table NOMETABELAHIVE (id INT, name STRING) row format delimited fields terminated by ',' lines terminated by '\n' stored as textfile partitioned by (age INT) tblproperties("skip.header.line.count"="1");
7) desc NOMETABELAHIVE
8) describe formatted NOMETABELAHIVE
9) insert overwrite table NOMETABELAHIVE partition(age = 10) select id, name from NOMETABELA where age=10;
10) select * from NOMETABELAHIVE where age = 10
11) insert overwrite table NOMETABELAHIVE partition(age = 20) select id, name from NOMETABELA where age=20;
12) insert overwrite table NOMETABELAHIVE partition(age = 11) select id, name from NOMETABELA where age=11;
13) insert overwrite table student_ori partition(age = 15) select id, name from student_ori_csv where age=15;
14) select * from student_ori
15) select * from student_ori where age = 15

Ambiente Hadoop - Exercício 2

1) Entrar no site https://databricks.com//
2) Crie um login - Versão gratuita - Community Edition
3) Clique em criar -> Cluster -> Escolher um nome e esperar executar
4) Copiar os 3 arquivos para o databricks:
    a) yob2015.txt, yob2016.txt, yob2017.txt
    b) Create -> Table -> Upload File
5) Verificar o nome que foi feito o upload.

Ambiente Hadoop - Exercício 2

1) show tables
2) select * from yob2015_1_txt limit 1
3) select count(*) from yob2015_1_txt
4) select count(*) from yob2016_1_txt
5) select count(*) from yob2017_1_txt
6) create table Nomes_Ano (nome string, sexo STRING, quant int) row format delimited fields terminated by ','
   lines terminated by '\n' stored as textfile partitioned by (ano INT);
7) desc Nomes_Ano
8) describe formatted Nomes_Ano
9) insert overwrite table Nomes_Ano partition(ano = 2015) select nome, sexo, quant from yob2015_1_txt;
10) insert overwrite table Nomes_Ano partition(ano = 2016) select nome, sexo, quant from yob2016_1_txt;
11) insert overwrite table Nomes_Ano partition(ano = 2017) select nome, sexo, quant from yob2017_1_txt;
12) select count(*) from Nomes_Ano
13) select count(*) from Nomes_Ano where ano= 2015
14) select * from Nomes_Ano where ano=2015 limit 10
15) select * from Nomes_Ano where ano=2016 limit 10

Ambiente Hadoop - Exercício 2

1) select * from Nomes_Ano where ano=2017 limit 10
2) select count(nome) as qtd from Nomes_Ano where ano=2017;
3) select sum(nome) as qtd from Nomes_Ano where ano=2017;
4) select sexo, sum(quant) as qtd from Nomes_Ano where ano=2015 group by sexo;
5) select ano, sexo, sum(quant) as qtd from Nomes_Ano group by ano, sexo order by ano desc;
6) select ano, sexo, sum(quant) as qtd from Nomes_Ano where nome like 'A%' group by ano, sexo order by ano desc;
7) select nome, max(quant) as qtd from Nomes_Ano where ano=2016 group by nome order by qtd desc limit 5;
8) select nome, max(quant) as qtd,sexo from Nomes_Ano where ano=2016 group by nome,sexo order by qtd desc limit 10;

Obrig.ada