

**FIAP - FACULDADE DE INFORMÁTICA E ADMINISTRAÇÃO PAULISTA**  
**DATA SCIENCE**

Fábio Pereira de Lima – RM98803  
Giovanna Cardoso Satorres – RM99944  
Giullia Bianca Rocha Souza – RM552108  
Gustavo Semenuk – RM550472  
Rafael Luiz Custódio Guimarães - RM98304

**SPRINT 3: HIVE PARA O**  
**CHALLENGE MINSAIT**

BIG DATA ARCHITECTURE & DATA INTEGRATION

## ★ Carregando os arquivos gerados pelo PIG para o HIVE

- Para dar continuidade às operações, o próximo passo foi realizar o carregamento dos arquivos gerados pelo PIG para o HIVE.
- Para iniciar o HIVE, executamos o comando “**hive**” no terminal.

```
[oracle@bigdatalite ~]$ hive
```

```
Logging initialized using configuration in jar:file:/usr/lib/hive/lib/hive-common-1.1.0-cdh5.13.1.jar!/hive-log4j.properties  
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.  
hive> █
```

- Após isso, criamos um novo banco de dados chamado 'maracuja' com o comando “**CREATE DATABASE maracuja;**” e o selecionamos para uso com o comando “**use maracuja;**”.

```
hive> CREATE DATABASE maracuja;  
OK  
Time taken: 12.094 seconds  
hive> █
```

```
hive> use maracuja;  
OK  
Time taken: 0.238 seconds  
hive> █
```

- Em seguida, criamos uma tabela chamada 'dados\_maracuja\_bahia', definindo os campos apropriados e o formato das colunas. Para isso, utilizamos o comando “**CREATE TABLE dados\_maracuja\_bahia (Municipio STRING, Area\_colhida INT, Quantidade\_produzida INT, Rendimento\_Medio INT) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\;';**”

```
hive> CREATE TABLE dados_maracuja_bahia (  
  >   Municipio STRING,  
  >   Area_colhida INT,  
  >   Quantidade_produzida INT,  
  >   Rendimento_Medio INT  
  > )  
  > ROW FORMAT DELIMITED  
  > FIELDS TERMINATED BY '\;';  
OK  
Time taken: 0.434 seconds
```

- Para carregar os dados do arquivo gerado pelo PIG para a tabela 'dados\_maracuja\_bahia' executamos os comando:  
“**LOAD DATA INPATH '/user/oracle/challenge/dados\_transformados/RM550472' INTO TABLE dados\_maracuja\_bahia;**”.

```
hive> LOAD DATA INPATH '/user/oracle/challenge/dados_transformados/RM550472' INTO TABLE dados_maracuja_bahia;
Loading data to table maracuja.dados_maracuja_bahia
OK
Time taken: 2.171 seconds
hive> █
```

## ★ Listando o total de registros do arquivo utilizando o HIVE

- Após o carregamento dos dados, o próximo passo foi utilizar o HIVE para realizar operações de análise.
- Executamos o comando “**show tables;**” para exibir a lista de tabelas no banco de dados 'maracuja'.

```
hive> show tables;
OK
dados_maracuja_bahia
Time taken: 0.585 seconds, Fetched: 1 row(s)
hive> █
```

- Executamos o comando “**show columns from dados\_maracuja\_bahia;**” para listar as colunas da tabela 'dados\_maracuja\_bahia'.

```
hive> show columns from dados_maracuja_bahia;
OK
municipio
area_colhida
quantidade_produzida
rendimento_medio
Time taken: 0.242 seconds, Fetched: 4 row(s)
hive> █
```

- Por fim, para listar todos os registros presentes na tabela 'dados\_maracuja\_bahia', executamos o comando "**SELECT \* FROM dados\_maracuja\_bahia;**".

```
hive> SELECT * FROM dados_maracuja_bahia;
OK
Ituacu 700      14000    20
Itirucu 416     7051     17
Itapicuru      505    6363     13
Jaguaquara     606    6272     10
Mucuge 240     6160     26
Ibicoara       300    6000     20
Brumado 550    5500     10
Tanhacu 300    4500     15
Juazeiro       257    4429     17
Carinhanha     206    3502     17
Curaca 235    2870     12
Jussiape       250    2624     10
Prado 120     2519     21
Itaquara       206    2266     11
Jequie 160     1920     12
Maracas 306    1831      6
Brejoes 260    1243      5
Bonito 180     1210      7
Caravelas       77    1078     14
Mirante 100    1000      10
Planaltino     192    857      4
Iramaia 80     797      10
Irajuba 101     780       8
Pocoas 75      765      10
Laje 95       726       8
Time taken: 1.078 seconds, Fetched: 25 row(s)
hive> █
```