

**FIAP – FACULDADE DE INFORMÁTICA E ADMINISTRAÇÃO PAULISTA**

BRUNO BIANCCHI – RM 84351

DOUGLAS ANDRADE SANTOS FILHO – 87004

LUIS HENRIQUE CALDAS ALTERO – RM 88670

PEDRO GUILHERME POLLONI BARRETO - RM 88964

VICTOR LAMPRECHT – RM 86691

**Challenge Sprint 3 – Consulta da Análise de Sentimento – Arquitetura de  
BI & Big Data**

**São Paulo  
2021**

**Comando:** cd Downloads

```
[oracle@bigdatalite ~]$ cd Downloads
[oracle@bigdatalite Downloads]$
```

**Comando:** pig -x local sentiment\_analysis\_challenge.pig

```
[oracle@bigdatalite Downloads]$ pig -x local sentiment_analysis_challenge.pig
log4j:WARN No appenders could be found for logger (org.apache.hadoop.util.Shell)
.
log4j:WARN Please initialize the log4j system properly.
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more in
fo.
2021-09-20 11:09:47,229 [main] INFO org.apache.pig.Main - Apache Pig version 0.
12.0-cdh5.7.0 (rexpoted) compiled Mar 23 2016, 11:34:31
2021-09-20 11:09:47,229 [main] INFO org.apache.pig.Main - Logging error message
s to: /home/oracle/Downloads/pig_1632150587213.log
2021-09-20 11:09:47,243 [main] INFO org.apache.hadoop.conf.Configuration.deprec
ation - user.name is deprecated. Instead, use mapreduce.job.user.name
2021-09-20 11:09:47,387 [main] INFO org.apache.pig.impl.util.Utils - Default bo
otup file /home/oracle/.pigbootup not found
2021-09-20 11:09:47,438 [main] INFO org.apache.hadoop.conf.Configuration.deprec
ation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-09-20 11:09:47,439 [main] INFO org.apache.hadoop.conf.Configuration.deprec
ation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.addr
ess
2021-09-20 11:09:47,440 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.HExecutionEngine - Connecting to hadoop file system at: file:///
```

```
Success!

Job Stats (time in seconds):
JobId  Alias  Feature Outputs
job_local1588913684_0001  clean_tokens,dictionary,english_tweets,token_sen
timent,tokenized,tweets HASH_JOIN  file:///home/oracle/Downloads/sentiment_
analysis_teste,

Input(s):
Successfully read records from: "file:///home/oracle/Downloads/united_airlines_t
weets.tsv"
Successfully read records from: "file:///home/oracle/Downloads/dictionary.tsv"

Output(s):
Successfully stored records in: "file:///home/oracle/Downloads/sentiment_analysi
s_teste"

Job DAG:
job_local1588913684_0001

2021-09-20 11:10:50,032 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.mapReduceLayer.MapReduceLauncher - Success!
[oracle@bigdatalite Downloads]$
```

**Comando:** hadoop fs -copyFromLocal sentiment\_analysis\_teste/part-r-00000  
analise\_sentimento

```
[oracle@bigdatalite Downloads]$ hadoop fs -copyFromLocal sentiment_analysis_test
e/part-r-00000 analise_sentimento
```

**Comando:** `hadoop fs -ls`

```
[oracle@bigdatalite Downloads]$ hadoop fs -ls
Found 11 items
drwxr-xr-x  - oracle oracle          0 2016-05-23 20:42 .Trash
drwxr-xr-x  - oracle oracle          0 2021-09-14 08:26 .sparkStaging
drwx----- - oracle oracle          0 2021-08-31 09:06 .staging
-rw-r--r--  1 oracle oracle      12085 2021-09-20 09:52 analyse_sentimento
drwxr-xr-x  - oracle oracle          0 2021-09-14 08:11 dados
drwxr-xr-x  - oracle oracle          0 2016-06-01 18:59 mediademo
drwxr-xr-x  - oracle oracle          0 2016-05-15 12:02 moviedemo
drwxr-xr-x  - oracle oracle          0 2016-05-15 12:03 moviework
drwxr-xr-x  - oracle oracle          0 2016-05-15 12:03 oggdemo
drwxr-xr-x  - oracle oracle          0 2016-05-15 12:03 oozie-oozi
-rw-r--r--  1 oracle oracle     10563 2021-09-14 08:14 population.csv
[oracle@bigdatalite Downloads]$
```

**Comando:** `hive`

```
[oracle@bigdatalite Downloads]$ hive
2021-09-20 09:59:40,370 WARN [main] mapreduce.TableMapReduceUtil: The hbase-prefix-tree module jar containing PrefixTreeCodec is not present. Continuing without it.

Logging initialized using configuration in jar:file:/usr/lib/hive/lib/hive-common-1.1.0-cdh5.7.0.jar!/hive-log4j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive>
```

**Comando:** `create database challenge;`

`Use challenge;`

```
hive> create database challenge;
OK
Time taken: 0.068 seconds
hive> use challenge;
OK
Time taken: 0.012 seconds
hive>
```

**Comando:** `CREATE TABLE IF NOT EXISTS analise (id STRING, palavra STRING, dicionario STRING, peso INT) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LINES TERMINATED BY '\n' STORED AS TEXTFILE;`

```
hive> CREATE TABLE
> IF NOT EXISTS analise
> (id STRING,
> palavra STRING,
> dicionario STRING,
> peso INT)
> ROW FORMAT DELIMITED
> FIELDS TERMINATED BY '\t'
> LINES TERMINATED BY '\n'
> STORED AS TEXTFILE;
OK
Time taken: 0.043 seconds
hive>
```

**Comando:** `LOAD DATA INPATH 'analyse_sentimento' INTO TABLE analise;`

```
hive> LOAD DATA INPATH 'analise_sentimento' INTO TABLE analise;
Loading data to table challenge.analise
OK
Time taken: 0.107 seconds
hive>
```

**Comando:** SELECT \* FROM analise;

```
hive> SELECT * FROM analise;
OK
475641126129369088      wow      wow      1
475114512215863297      wow      wow      1
474435643792060416      wow      wow      1
473931319794085888      deal     deal     1
473938364240896000      deal     deal     1
473930423324520449      deal     deal     1
473940138884464640      deal     deal     1
473940073352683520      deal     deal     1
473919711797727232      deal     deal     1
473918819950272513      deal     deal     1
473938323686162433      deal     deal     1
473940070831898624      deal     deal     1
473940141241663489      deal     deal     1
473939067961221120      deal     deal     1
473939069806731265      deal     deal     1
473930419448971264      deal     deal     1
473939067961221120      earn     earn     1
473940073352683520      earn     earn     1
473938323686162433      earn     earn     1
473940070831898624      earn     earn     1

475580179986010112      luxurytravel luxurytravel 1
475580069197672448      luxurytravel luxurytravel 1
475579246749159426      luxurytravel luxurytravel 1
475582520210489344      luxurytravel luxurytravel 1
475581551770869760      luxurytravel luxurytravel 1
475579605848694784      luxurytravel luxurytravel 1
475580342108430336      luxurytravel luxurytravel 1
475579345952866304      luxurytravel luxurytravel 1
475641304077328384      luxurytravel luxurytravel 1
475579728750182400      luxurytravel luxurytravel 1
475581669848915968      luxurytravel luxurytravel 1
473412900577873920      luxurytravel luxurytravel 1
475578976795389952      luxurytravel luxurytravel 1
475579614308622336      luxurytravel luxurytravel 1
475581769002270720      luxurytravel luxurytravel 1
475581681664262144      luxurytravel luxurytravel 1
475579315292483585      luxurytravel luxurytravel 1
475579349119537152      luxurytravel luxurytravel 1
473233757961723904      delayedflight delayedflight -1
474021915175034881      flyerfriendly flyerfriendly 1
473607792331206657      flyerfriendly flyerfriendly 1
474022355157520386      flyerfriendly flyerfriendly 1
Time taken: 0.061 seconds, Fetched: 334 row(s)
hive>
```

**Comando:** SELECT palavra, count (\*) palavra FROM analise where palavra = 'thanks' group by palavra;

```
hive> SELECT palavra, count (*) palavra FROM analise where palavra = 'thanks' group by palavra;
```

```

MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.07 sec HDFS Read: 19888 HDFS Write: 10 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 70 msec
OK
thanks 36
Time taken: 15.086 seconds, Fetched: 1 row(s)
hive>

```

**Comando:** SELECT palavra, count (\*) palavra FROM analise where palavra = 'thanks' or palavra = 'luxurytravel' group by palavra;

```

hive> SELECT palavra, count (*) palavra FROM analise where palavra = 'thanks' or
palavra = 'luxurytravel' group by palavra;

```

```

MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.05 sec HDFS Read: 19476 HDFS Write: 26 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 50 msec
OK
luxurytravel 28
thanks 36
Time taken: 17.362 seconds, Fetched: 2 row(s)
hive>

```

**Comando:** SELECT palavra, count (\*) palavra FROM analise where palavra = 'thanks' or palavra = 'luxurytravel' or palavra = 'great' group by palavra;

```

hive> SELECT palavra, count (*) palavra FROM analise where palavra = 'thanks' or
palavra = 'luxurytravel' or palavra = 'great' group by palavra;

```

```

MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.34 sec HDFS Read: 19610 HDFS Write: 34 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 340 msec
OK
great 9
luxurytravel 28
thanks 36
Time taken: 15.481 seconds, Fetched: 3 row(s)
hive>

```