

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

Global Solution – Arquitetura de BI & Big Data

São Paulo

2021

1 – Conexão com o banco ORCL no servidor oracle.fiap.com.br

```
oracle@b
File Edit View Search Terminal Help
[oracle@bigdatalite ~]$ sqlplus RM88964/250292@oracle.fiap.com.br:1521/orcl.fiap.com.br

SQL*Plus: Release 12.1.0.2.0 Production on Thu Jun 3 17:41:08 2021

Copyright (c) 1982, 2014, Oracle. All rights reserved.

Last Successful login time: Thu Jun 03 2021 17:40:35 -04:00

Connected to:
Oracle Database 19c Enterprise Edition Release 19.0.0.0.0 - Production
```

2 – Criação das tabelas CLIENTE, AVALIACAO e de suas chaves primárias (retiradas do script gerado a partir do Data Modeler).

```
SQL> CREATE TABLE t_trl_avaliacao(
  2   cd_avaliacao      number(3) not null,
  3   nr_quarto         number(3) not null,
  4   cd_cliente       number(5) not null,
  5   cd_hotel          number(2) not null,
  6   ds_avaliacao      varchar2(500),
  7   qt_estrelas_avaliacao number(1)
  8 );
```

Table created.

```
SQL> CREATE TABLE t_trl_cliente(
  2   cd_cliente       NUMBER(5) NOT NULL,
  3   nm_cliente       VARCHAR2(50) NOT NULL,
  4   nr_rg            NUMBER(8) NOT NULL,
  5   nr_rg_digito     CHAR(1) NOT NULL,
  6   nr_cpf           NUMBER(9) NOT NULL,
  7   nr_cpf_digito    NUMBER(2) NOT NULL,
  8   nr_idade_cliente NUMBER(2) NOT NULL,
  9   nr_celular       NUMBER(11) NOT NULL,
 10   ds_endereco      VARCHAR2(100) NOT NULL,
 11   ds_email         VARCHAR2(50) NOT NULL);
```

Table created.

```
SQL> █
```

```
SQL> ALTER TABLE t_trl_avaliacao
  2   ADD CONSTRAINT pk_trl_avaliacao PRIMARY KEY ( cd_avaliacao,
  3   nr_quarto, cd_hotel, cd_cliente);
```

Table altered.

Table created.

```
SQL> ALTER TABLE t_trl_cliente ADD CONSTRAINT pk_trl_cliente PRIMARY KEY ( cd_cliente )
;
```

Table altered.

3 – Inserção de dados dentro das tabelas CLIENTE e AVALIACAO.

```
SQL> insert into t_trl_cliente values(
  2  1,'nome1',11111111,1,11111111,11,01,1111111111,'rua1','email1');

1 row created.

SQL> select * from t_trl_cliente;

CD_CLIENTE NM_CLIENTE NR_RG N
-----
NR_CPF NR_CPF_DIGITO NR_IDADE_CLIENTE NR_CELULAR
-----
DS_ENDERECO
-----
DS_EMAIL
-----
          1 nome1 11111111 1
11111111 11 1 1.1111E+10
rua1
email1

SQL> insert into t_trl_cliente values(
  2  00002,'nome2',22222222,2,22222222,22,02,2222222222,'rua2','email2');

1 row created.

SQL> insert into t_trl_cliente values(
  2  00003,'nome3',33333333,3,33333333,33,03,3333333333,'rua3','email3');

1 row created.

SQL> insert into t_trl_cliente values(
  2  00004,'nome4',44444444,4,44444444,44,04,4444444444,'rua4','email4');

1 row created.

SQL> insert into t_trl_cliente values(
  2  00005,'nome5',55555555,5,55555555,55,05,5555555555,'rua5','email5');

1 row created.

SQL> insert into t_trl_cliente values(
  2  00006,'nome6',66666666,6,66666666,66,06,6666666666,'rua6','email6');

1 row created.

SQL> insert into t_trl_cliente values(
  2  00007,'nome7',77777777,7,77777777,77,07,7777777777,'rua7','email7');

1 row created.

SQL> insert into t_trl_cliente values(
  2  00008,'nome8',88888888,8,88888888,88,08,8888888888,'rua8','email8');

1 row created.

SQL> insert into t_trl_cliente values(
  2  00009,'nome9',99999999,9,99999999,99,09,9999999999,'rua9','email9');

1 row created.

SQL> insert into t_trl_cliente values(
  2  00010,'nome10',10101010,1,10101010,10,50,1010101010,'rua10','email10');

1 row created.
```

```
SQL> INSERT INTO t_trl_avaliacao VALUES (001, 001, 00001, 01, 'hotel1',1);
1 row created.

SQL> INSERT INTO t_trl_avaliacao VALUES (002, 002, 00002, 02, 'hotel2',2);
1 row created.

SQL> INSERT INTO t_trl_avaliacao VALUES (003,003,00003,03,'hotel3',3);
1 row created.

SQL> INSERT INTO t_trl_avaliacao VALUES (004,004,00004,04,'hotel4',4);
1 row created.

SQL> INSERT INTO t_trl_avaliacao VALUES (005,005,00005,05,'hotel5',5);
1 row created.

SQL> INSERT INTO t_trl_avaliacao VALUES (006,006,00006,06,'hotel6',1);
1 row created.

SQL> INSERT INTO t_trl_avaliacao VALUES (007,007,00007,07,'hotel7',2);
1 row created.

SQL> INSERT INTO t_trl_avaliacao VALUES (008,008,00008,08,'hotel8',3);
1 row created.

SQL> INSERT INTO t_trl_avaliacao VALUES (009,009,00009,09,'hotel9',4);
1 row created.

SQL> INSERT INTO t_trl_avaliacao VALUES (010,010,00010,10,'hotel10',5);
1 row created.
```

```
SQL> set linesize 1000
SQL> select * from t_trl_avaliacao;
```

	1	1	1	1 hotel1
1	2	2	2	2 hotel2
2	3	3	3	3 hotel3
3	4	4	4	4 hotel4
4	5	5	5	5 hotel5
5	6	6	6	6 hotel6
1	7	7	7	7 hotel7
2	8	8	8	8 hotel8
3	9	9	9	9 hotel9
4	10	10	10	10 hotel10
5				

```
SQL> set linesize 1000
SQL> select * from t_trl_cliente;
```

CD_CLIENTE	NM_CLIENTE	NR_RG_N	NR_CPF	NR_CPF_DIGITO	NR_IDADE_CLIENTE	NR_CELULAR	DS_ENDERECO
DS_EMAIL							
1	nome1	11111111 1	111111111	11	1	1.1111E+10	rua1
email1							
2	nome2	22222222 2	222222222	22	2	2.2222E+10	rua2
email2							
3	nome3	33333333 3	333333333	33	3	3.3333E+10	rua3
email3							
4	nome4	44444444 4	444444444	44	4	4.4444E+10	rua4
email4							
5	nome5	55555555 5	555555555	55	5	5.5556E+10	rua5
email5							
6	nome6	66666666 6	666666666	66	6	6.6667E+10	rua6
email6							
7	nome7	77777777 7	777777777	77	7	7.7778E+10	rua7
email7							
8	nome8	88888888 8	888888888	88	80	8.8889E+10	rua8
email8							
9	nome9	99999999 9	999999999	99	99	1.0000E+11	rua9
email9							
10	nome10	10101010 1	101010101	10	50	1.0101E+10	rua10
email10							

5 – Transferência dos dados das tabelas CLIENTE e AVALIACAO para o HADOOP.

- Transferência da tabela AVALIACAO.

```
[oracle@bigdatalite Downloads]$ sqoop import --connect jdbc:oracle:thin:@oracle.fiap.com.br:1521/ORCL.fiap.com.br --username RM88964 --password 250292 --table T_TRL_AVALIACAO --input-fields-terminated-by ';' --lines-terminated-by '\n' --as-textfile --target-dir avaliacao.txt --m 1
Warning: /usr/lib/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
21/06/03 15:54:24 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.7.0
21/06/03 15:54:24 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
21/06/03 15:54:24 INFO oracle.OracleManagerFactory: Data Connector for Oracle and Hadoop is disabled.
21/06/03 15:54:24 INFO manager.SqlManager: Using default fetchSize of 1000
21/06/03 15:54:24 INFO tool.CodeGenTool: Beginning code generation
21/06/03 15:54:25 INFO manager.OracleManager: Time zone has been set to GMT
21/06/03 15:54:25 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM T_TRL_AVALIACAO t WHERE 1=0
21/06/03 15:54:25 INFO orm.CompilationManager: HADOOP MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-oracle/compile/41f749976fd90fa9b0edf17f9df5ae99/T_TRL_AVALIACAO.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
21/06/03 15:54:26 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-oracle/compile/41f749976fd90fa9b0edf17f9df5ae99/T_TRL_AVALIACAO.jar
21/06/03 15:54:26 INFO manager.OracleManager: Time zone has been set to GMT
21/06/03 15:54:26 WARN manager.OracleManager: The table T_TRL_AVALIACAO contains a multi-column primary key. Sqoop will default to the column CD_AVALIACAO only for this job.
21/06/03 15:54:26 INFO manager.OracleManager: Time zone has been set to GMT
21/06/03 15:54:26 WARN manager.OracleManager: The table T_TRL_AVALIACAO contains a multi-column primary key. Sqoop will default to the column CD_AVALIACAO only for this job.
21/06/03 15:54:26 INFO mapreduce.ImportJobBase: Beginning import of T_TRL_AVALIACAO
21/06/03 15:54:26 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
21/06/03 15:54:26 INFO manager.OracleManager: Time zone has been set to GMT
21/06/03 15:54:27 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
21/06/03 15:54:27 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
21/06/03 15:54:28 INFO db.DBInputFormat: Using read committed transaction isolation
21/06/03 15:54:28 INFO mapreduce.JobSubmitter: number of splits=1

HDFS: Number of bytes read=87
HDFS: Number of bytes written=175
HDFS: Number of read operations=4
HDFS: Number of large read operations=0
HDFS: Number of write operations=2

Job Counters
  Launched map tasks=1
  Other local map tasks=1
  Total time spent by all maps in occupied slots (ms)=2246
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=2246
  Total vcore-seconds taken by all map tasks=2246
  Total megabyte-seconds taken by all map tasks=574976

Map-Reduce Framework
  Map input records=10
  Map output records=10
  Input split bytes=87
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=38
  CPU time spent (ms)=1500
  Physical memory (bytes) snapshot=203005952
  Virtual memory (bytes) snapshot=2118328320
  Total committed heap usage (bytes)=119537664

File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=175
21/06/03 15:54:38 INFO mapreduce.ImportJobBase: Transferred 175 bytes in 10.8462 seconds (16.1346 bytes/sec)
21/06/03 15:54:38 INFO mapreduce.ImportJobBase: Retrieved 10 records.
```

- Transferência da tabela CLIENTE.

```
[oracle@bigdatalite Downloads]$ sqoop import --connect jdbc:oracle:thin:@oracle.fiap.com.br:1521/ORCL.fiap.com.br --username RM88964 --password 250292 --table T_TRL_CLIENTE --input-fi
elds-terminated-by ';' --lines-terminated-by '\n' --as-textfile --target-dir cliente.txt
Warning: /usr/lib/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
21/06/03 15:46:28 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.7.0
21/06/03 15:46:28 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
21/06/03 15:46:28 INFO oracle.OracleManagerFactory: Data Connector for Oracle and Hadoop is disabled.
21/06/03 15:46:28 INFO manager.SqlManager: Using default fetchSize of 1000
21/06/03 15:46:28 INFO tool.CodeGenTool: Beginning code generation
21/06/03 15:46:29 INFO manager.OracleManager: Time zone has been set to GMT
21/06/03 15:46:29 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM T_TRL_CLIENTE t WHERE t=0
21/06/03 15:46:29 INFO orm.CompilationManager: HADOOP MAPRED HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-oracle/compile/06120366ba49cbf4f01bf2bf5a34e1f1/T_TRL_CLIENTE.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
21/06/03 15:46:30 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-oracle/compile/06120366ba49cbf4f01bf2bf5a34e1f1/T_TRL_CLIENTE.jar
21/06/03 15:46:30 INFO manager.OracleManager: Time zone has been set to GMT
21/06/03 15:46:30 INFO manager.OracleManager: Time zone has been set to GMT
21/06/03 15:46:30 INFO mapreduce.ImportJobBase: Beginning import of T_TRL_CLIENTE
21/06/03 15:46:30 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
21/06/03 15:46:30 INFO manager.OracleManager: Time zone has been set to GMT
21/06/03 15:46:31 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
21/06/03 15:46:31 INFO client.RMProxy: Connecting to Resource Manager at /0.0.0.0:8032
21/06/03 15:46:32 INFO db.DBInputFormat: Using read committed transaction isolation
21/06/03 15:46:32 INFO db.DataDrivenDBInputFormat: BoundingValsQuery: SELECT MIN(CD_CLIENTE), MAX(CD_CLIENTE) FROM T_TRL_CLIENTE
21/06/03 15:46:32 INFO mapreduce.JobSubmitter: number of splits:4
21/06/03 15:46:32 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1617712126533_0020
21/06/03 15:46:32 INFO impl.YarnClientImpl: Submitted application application_1617712126533_0020
21/06/03 15:46:32 INFO mapreduce.Job: The url to track the job: http://bigdatalite.localdomain:8088/proxy/application_1617712126533_0020/
21/06/03 15:46:32 INFO mapreduce.Job: Running job: job_1617712126533_0020
21/06/03 15:46:37 INFO mapreduce.Job: Job job_1617712126533_0020 running in uber mode : false
21/06/03 15:46:37 INFO mapreduce.Job: map 0% reduce 0%
21/06/03 15:46:42 INFO mapreduce.Job: map 25% reduce 0%
21/06/03 15:46:45 INFO mapreduce.Job: map 50% reduce 0%
21/06/03 15:46:46 INFO mapreduce.Job: map 75% reduce 0%
21/06/03 15:46:52 INFO mapreduce.Job: map 100% reduce 0%
21/06/03 15:46:52 INFO mapreduce.Job: Job job_1617712126533_0020 completed successfully
21/06/03 15:46:52 INFO mapreduce.Job: Counters: 30
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=549788
  FILE: Number of read operations=0
```

```
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=456
HDFS: Number of bytes written=646
HDFS: Number of read operations=16
HDFS: Number of large read operations=0
HDFS: Number of write operations=8
```

Job Counters

```
Launched map tasks=4
Other local map tasks=4
Total time spent by all maps in occupied slots (ms)=19008
Total time spent by all reduces in occupied slots (ms)=0
Total time spent by all map tasks (ms)=19008
Total vcore-seconds taken by all map tasks=19008
Total megabyte-seconds taken by all map tasks=4866048
```

Map-Reduce Framework

```
Map input records=11
Map output records=11
Input split bytes=456
Spilled Records=0
Failed Shuffles=0
Merged Map outputs=0
GC time elapsed (ms)=254
CPU time spent (ms)=7140
Physical memory (bytes) snapshot=862867456
Virtual memory (bytes) snapshot=8504492032
Total committed heap usage (bytes)=470810624
```

File Input Format Counters

```
Bytes Read=0
```

File Output Format Counters

```
Bytes Written=646
```

```
21/06/03 15:46:52 INFO mapreduce.ImportJobBase: Transferred 646 bytes in 20.8964 seconds (30.9145 bytes/sec)
```

```
21/06/03 15:46:52 INFO mapreduce.ImportJobBase: Retrieved 11 records.
```

- Captura de tela exibindo as tabelas transferidas no HADOOP.

```
[oracle@bigdatalite Downloads]$ hadoop fs -ls
Found 30 items
drwxr-xr-x - oracle oracle 0 2016-05-23 20:42 .Trash
drwxr-xr-x - oracle oracle 0 2016-05-23 20:39 .sparkStaging
drwx----- - oracle oracle 0 2021-06-03 15:54 .staging
drwxr-xr-x - oracle oracle 0 2021-04-27 07:35 ACTIVITY
drwxr-xr-x - oracle oracle 0 2021-04-27 08:06 DADOS_ETL
drwxr-xr-x - oracle oracle 0 2021-06-03 14:17 Documents
drwxr-xr-x - oracle oracle 0 2021-06-03 15:39 Downloads
drwxr-xr-x - oracle oracle 0 2021-04-27 07:40 FILTERED_ACTIVITIES
drwxr-xr-x - oracle oracle 0 2021-06-03 14:30 T_TRL_AVALIACAO
drwxr-xr-x - oracle oracle 0 2021-06-03 13:55 T_TRL_CLIENTE
drwxr-xr-x - oracle oracle 0 2021-04-27 07:40 sqoop
drwxr-xr-x - oracle oracle 0 2021-06-03 15:54 avaliacao.txt
drwxr-xr-x - oracle oracle 0 2021-06-03 15:46 cliente.txt
-rw-r--r-- 1 oracle oracle 495 2021-04-26 20:52 csv_generator.csv
-rw-r--r-- 1 oracle oracle 20837 2021-04-26 20:57 dados.csv
-rw-r--r-- 1 oracle oracle 14122015 2021-04-13 07:38 empregado.txt
drwxr-xr-x - oracle oracle 0 2021-04-27 07:36 etl
-rw-r--r-- 1 oracle oracle 495 2021-04-26 20:56 lalalala.csv
-rw-r--r-- 1 oracle oracle 20837 2021-04-26 20:55 log_2021_04_26__20_45_55.csv
drwxr-xr-x - oracle oracle 0 2016-06-01 18:59 mediademo
drwxr-xr-x - oracle oracle 0 2021-03-16 08:32 meunome
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
drwxr-xr-x - oracle oracle 0 2021-03-30 07:51 pedro
drwxr-xr-x - oracle oracle 0 2021-03-30 07:28 pedro.txt
drwxr-xr-x - oracle oracle 0 2021-03-30 07:39 pedro2
drwxr-xr-x - oracle oracle 0 2021-03-30 07:40 pedroo
-rw-r--r-- 1 oracle oracle 572 2021-03-30 07:51 reducer.py
[oracle@bigdatalite Downloads]$
```

6 – Junção das partições das tabelas CLIENTE e AVALIACAO do HADOOP para o sistema local.

```
[oracle@bigdatalite Downloads]$ hadoop fs -copyToLocal Downloads cliente.txt
```

```
[oracle@bigdatalite ~]$ cd Downloads
[oracle@bigdatalite Downloads]$ ls
ITBDA_AKATSUKI_SCRIPT_DDL.sql      Cidades.csv      empregados.csv      mr_armas.txt      pig_1620130481190.log      prop.txt      tabelas_trl
alunos.csv                         cliente.txt       empregado.txt       pedro              pig_1620735576228.log      Redis.pdf      T_TRL_AVALIACAO.java
armas.txt                          csv_generator.csv estado.csv          pedro1.pig        pig_1620735670951.log      reducer.py     T_TRL_CLIENTE.java
Avaliação_MongoDB.docx            csv_gen.sh       fil.txt            pedro2.pig        pig_1620735829740.log      saidapop.csv  united_airlines_tweets.tsv
avaliacao.txt                     Dados.zip        log_2021_04_26__20_45_55.csv      pedro3.pig        pig_1620736131853.log      sentiment_analysis.pig  upgrade_pig.txt
campinas                          departamento.csv  mapper.py          pedro4.pig        pig_1620736848066.log      sfpd.csv      WordCount.java
Cassandra.pdf                     dictionary.tsv    MAPREDUCE em MongoDB.txt      pedro5.pig        pig_1621335771026.log      shakespeare.txt      WordMapper.java
catalogo_de_galaxias.csv          empregado.csv    Movies.csv         pedro.txt         population.csv              SumReducer.java
[oracle@bigdatalite Downloads]$ cd cliente.txt
[oracle@bigdatalite cliente.txt]$ cat part-m-00000 part-m-00001 part-m-00002 part-m-00003 > cliente_concatenada.txt
[oracle@bigdatalite cliente.txt]$
[oracle@bigdatalite Downloads]$ hadoop fs -copyToLocal Downloads avaliacao.txt
[oracle@bigdatalite Downloads]$ ls
ITBDA_AKATSUKI_SCRIPT_DDL.sql      Cidades.csv      empregados.csv      mr_armas.txt      pig_1620130481190.log      prop.txt      tabelas_trl
alunos.csv                         cliente.txt       empregado.txt       pedro              pig_1620735576228.log      Redis.pdf      T_TRL_AVALIACAO.java
armas.txt                          csv_generator.csv estado.csv          pedro1.pig        pig_1620735670951.log      reducer.py     T_TRL_CLIENTE.java
Avaliação_MongoDB.docx            csv_gen.sh       fil.txt            pedro2.pig        pig_1620735829740.log      saidapop.csv  united_airlines_tweets.tsv
avaliacao.txt                     Dados.zip        log_2021_04_26__20_45_55.csv      pedro3.pig        pig_1620736131853.log      sentiment_analysis.pig  upgrade_pig.txt
campinas                          departamento.csv  mapper.py          pedro4.pig        pig_1620736848066.log      sfpd.csv      WordCount.java
Cassandra.pdf                     dictionary.tsv    MAPREDUCE em MongoDB.txt      pedro5.pig        pig_1621335771026.log      shakespeare.txt      WordMapper.java
catalogo_de_galaxias.csv          empregado.csv    Movies.csv         pedro.txt         population.csv              SumReducer.java
[oracle@bigdatalite ~]$ cd Downloads
[oracle@bigdatalite Downloads]$ cd avaliacao.txt
[oracle@bigdatalite avaliacao.txt]$ ls
cliente1.txt cliente.txt part-m-00000 part-m-00001 part-m-00002 part-m-00003 _SUCCESS
[oracle@bigdatalite avaliacao.txt]$ cat part-m-00000 part-m-00001 part-m-00002 part-m-00003 > avaliacao_concatenada.txt
[oracle@bigdatalite avaliacao.txt]$ ls
avaliacao_concatenada.txt cliente1.txt cliente.txt part-m-00000 part-m-00001 part-m-00002 part-m-00003 _SUCCESS
```


7 – Criação Script da tabela CLIENTE e execução do script no PIG.

```
[oracle@bigdatalite Downloads]$ vi bag_1.pig
```

```
cliente = LOAD 'cliente_concatenada.txt' USING PigStorage(',') AS (cd_cliente: int, nm_cliente: chararray, nr_rg: int, nr_rg_digito: chararray, nr_cpf: int, nr_cpf_digito: int, nr_idade_cliente: int, nr_celular: int, ds_endereco: chararray, ds_email: chararray);
```

```
group_c = GROUP cliente BY cd_cliente;
```

```
contagem = FOREACH group_c GENERATE SUM(cliente.cd_cliente);
```

```
DUMP contagem;
```

```
STORE contagem INTO 'bag_1';
```

```
~  
~  
~  
~  
~  
~  
~  
~  
~
```

```
-- INSERT --
```

```
9,29
```

```
All
```

```
[oracle@bigdatalite Downloads]$ pig -x local bag_1.pig
```

```
Length = 587  
Locations:
```

```
-----
```

```
2021-06-04 18:54:38,008 [LocalJobRunner Map Task Executor #0] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.PigRecordReader - Current split being processed file:/home/oracle/Downloads/cliente_concatenada.txt:0+587  
2021-06-04 18:54:38,031 [LocalJobRunner Map Task Executor #0] INFO org.apache.hadoop.mapred.MapTask - (EQUATOR) 0 kvi 26214396(104857584)  
2021-06-04 18:54:38,031 [LocalJobRunner Map Task Executor #0] INFO org.apache.hadoop.mapred.MapTask - mapreduce.task.io.sort.mb: 100  
2021-06-04 18:54:38,031 [LocalJobRunner Map Task Executor #0] INFO org.apache.hadoop.mapred.MapTask - soft limit at 83886080  
2021-06-04 18:54:38,031 [LocalJobRunner Map Task Executor #0] INFO org.apache.hadoop.mapred.MapTask - bufstart = 0; bufvoid = 104857600  
2021-06-04 18:54:38,031 [LocalJobRunner Map Task Executor #0] INFO org.apache.hadoop.mapred.MapTask - kvstart = 26214396; length = 6553600  
2021-06-04 18:54:38,035 [LocalJobRunner Map Task Executor #0] INFO org.apache.hadoop.mapred.MapTask - Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer  
2021-06-04 18:54:38,046 [LocalJobRunner Map Task Executor #0] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.  
2021-06-04 18:54:38,065 [LocalJobRunner Map Task Executor #0] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.PigGenericMapReduce$Map - Aliases being processed per job phase (AliasName[line,offset]): M: cliente[1,10], cliente[-1,-1], contagem[5,11], group_c[3,10] C: contagem[5,11], group_c[3,10] R: contagem[5,11]
```

```
s - Detected Local mode. Stats reported below may be incomplete
2021-06-04 18:54:39,031 [main] INFO org.apache.pig.tools.pigstats.SimplePigStat
s - Script Statistics:
```

HadoopVersion	PigVersion	UserId	StartedAt	FinishedAt	Features
2.6.0-cdh5.7.0	0.12.0-cdh5.7.0	oracle	2021-06-04 18:54:38	2021-06-04 18:54:39	GROUP_BY

Success!

Job Stats (time in seconds):

JobId	Alias	Feature	Outputs
job_local815760992_0002	cliente,contagem,group_c	GROUP_BY,COMBINER	f

ile:///home/oracle/Downloads/bag_1,

Input(s):

Successfully read records from: "file:///home/oracle/Downloads/cliente_concatenada.txt"

Output(s):

Successfully stored records in: "file:///home/oracle/Downloads/bag_1"

Job DAG:

job_local815760992_0002

```
2021-06-04 18:54:39,031 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
```

- Resultado do DUMP.

Output(s):

Successfully stored records in: "file:/tmp/temp-431530782/tmp-1930839090"

Job DAG:

job_local1867993209_0001

```
2021-06-04 18:54:38,449 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
```

```
2021-06-04 18:54:38,452 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
```

```
2021-06-04 18:54:38,452 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
```

```
2021-06-04 18:54:38,460 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
```

```
2021-06-04 18:54:38,461 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
```

```
(1)
(2)
(3)
(4)
(5)
(6)
(7)
(8)
(9)
(10)
```

- Passagem do bag_1 para o HADOOP.

```
[oracle@bigdatalite Downloads]$ hadoop fs -copyFromLocal bag_1
```

8 –Criação Script da tabela AVALIACAO e execução do script no PIG.

```
[oracle@bigdatalite Downloads]$ vi bag_2.pig|
```

```
avaliacao = LOAD 'avaliacao_concatenada.txt' USING PigStorage(',') AS (cd_avaliacao: int, nr_quarto: int, cd_cliente: int, cd_hotel: int, ds_avaliacao: chararray, qt_estrelas_avaliacao: int);
```

```
group_a = GROUP avaliacao BY qt_estrelas_avaliacao;
```

```
contagem = FOREACH group_a GENERATE AVG(avaliacao.cd_avaliacao);
```

```
ILLUSTRATE contagem;
```

```
STORE contagem INTO 'bag_2';
```

```
~  
~  
~
```

1,1

All

```
ne.mapReduceLayer.InputSizeReducerEstimator - BytesPerReducer=1000000000 maxReducers=999 totalInputFileSize=175
```

```
2021-06-04 19:13:15,549 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting Parallelism to 1
```

```
2021-06-04 19:13:15,559 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
```

```
2021-06-04 19:13:15,562 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.PigGenericMapReduce$Map - Aliases being processed per job phase (AliasName[line,offset]): M: avaliacao[1,12],avaliacao[-1,-1],group_a[3,10] C: R: contagem[5,11]
```

```
2021-06-04 19:13:15,565 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
```

```
2021-06-04 19:13:15,568 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.PigMapReduce$Reduce - Aliases being processed per job phase (AliasName[line,offset]): M: avaliacao[1,12],avaliacao[-1,-1],group_a[3,10] C: R: contagem[5,11]
```

```
2021-06-04 19:13:15,568 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
```

```
2021-06-04 19:13:15,570 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
```

```
2021-06-04 19:13:15,570 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
```

```
2021-06-04 19:13:15,570 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.AccumulatorOptimizer - Reducer is to run in accumulative mode.
```

```
2021-06-04 19:13:15,570 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
```

```
2021-06-04 19:13:15,571 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent
```

```
s - Detected Local mode. Stats reported below may be incomplete
2021-06-04 19:13:16,805 [main] INFO org.apache.pig.tools.pigstats.SimplePigStat
s - Script Statistics:
```

```
HadoopVersion  PigVersion      UserId  StartedAt      FinishedAt      Features
2.6.0-cdh5.7.0  0.12.0-cdh5.7.0 oracle  2021-06-04 19:13:15      2021-06-04 19:13
:16      GROUP_BY
```

Success!

Job Stats (time in seconds):

```
JobId  Alias  Feature Outputs
job_local1168666297_0001      avaliacao,contagem,group_a      GROUP_BY,COMBINE
R      file:///home/oracle/Downloads/bag_2,
```

Input(s):

Successfully read records from: "file:///home/oracle/Downloads/avaliacao_concate
nada.txt"

Output(s):

Successfully stored records in: "file:///home/oracle/Downloads/bag_2"

Job DAG:

job_local1168666297_0001

```
2021-06-04 19:13:16,806 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.mapReduceLayer.MapReduceLauncher - Success!
[oracle@bigdatalite Downloads]$ █
```

- Resultado do ILLUSTRATE.

```
| avaliacao      | cd_avaliacao:int | nr_quarto:int | cd_cliente:int | c
d_hotel:int      | ds_avaliacao:chararray | qt_estrelas_avaliacao:int |
-----
|                | 9                | 9             | 9              | 9
|                | hotel9           | 4             |                |
|                | 4                | 4             | 4              | 4
|                | hotel4           | 4             |                |
-----
|
|
|
-----
| group_a      | group:int | avaliacao:bag{tuple(cd_avaliacao:int,nr_quarto:i
nt,cd_cliente:int,cd_hotel:int,ds_avaliacao:chararray,qt_estrelas_avaliacao:int)
}
|
-----
|                | 4                | {(9, ..., 4), (4, ..., 4)}
|
|
|
-----
| contagem      | :double      |
|                | 6.5          |
```

- Passagem do bag_2 para o HADOOP.

```
[oracle@bigdatalite Downloads]$ hadoop fs -copyFromLocal bag_2
```

9 –Resultado das operações STORE no HADOOP.

```
[oracle@bigdatalite Downloads]$ hadoop fs -ls
Found 22 items
drwxr-xr-x - oracle oracle      0 2016-05-23 20:42 .Trash
drwxr-xr-x - oracle oracle      0 2016-05-23 20:39 .sparkStaging
drwx----- - oracle oracle      0 2021-04-27 08:05 .staging
drwxr-xr-x - oracle oracle      0 2021-04-27 07:17 ACTIVITY
drwxr-xr-x - oracle oracle      0 2021-04-21 22:49 EMPREGADO_100
drwxr-xr-x - oracle oracle      0 2021-04-27 07:44 FILTERED_ACTIVITIES
drwxr-xr-x - oracle oracle      0 2021-04-27 07:43 sqoop
drwxr-xr-x - oracle oracle      0 2021-06-04 18:21 bag_1
drwxr-xr-x - oracle oracle      0 2021-06-04 19:20 bag_2
drwxr-xr-x - oracle oracle      0 2021-04-27 08:05 dados
-rw-r--r-- 1 oracle oracle    33763 2021-04-26 21:21 dados.csv
drwxr-xr-x - oracle oracle      0 2021-04-21 22:52 dois
-rw-r--r-- 1 oracle oracle  14122015 2021-04-20 07:17 empregado.txt
-rw-r--r-- 1 oracle oracle  9338880 2021-05-13 20:17 empregado_mp
drwxr-xr-x - oracle oracle      0 2021-04-27 07:19 etl
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    572 2021-05-13 20:03 reducer.py
drwxr-xr-x - oracle oracle      0 2021-04-21 22:54 tres
[oracle@bigdatalite Downloads]$
```