FIAP – FACULDADE DE INFORMÁTICA E ADMINISTRAÇÃO PAULISTA

BRUNO BIANCCHI – RM 84351

LUIS HENRIQUE CALDAS ALTERO – RM 88670

VICTOR LAMPRECHT – RM 86691

Checkpoint 2 – Arquitetura de BI & Big Data

São Paulo

2021

cd Downloads

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

ls

```
[oracle@bigdatalite Downloads]$ ls
AAA-r0_LQkT0ze54KnzYzhDMa?d1=0    pig_1632143168230.log
alunos.csv                       pig_1632145316643.log
armas.txt                        pig_1632150587213.log
Avaliação_MongoDB.docx           population.csv
Cassandra.pdf                    Redis.pdf
catalogo_de_galaxias.csv         reducer.py
Cidades.csv                      sentiment_analysis
csv_gen.sh                       sentiment_analysis_challenge.pig
dictionary.tsv                   sentiment_analysis.pig
empregado.csv                    sentiment_analysis_teste
empregados.csv                   sfpd.csv
empregado.txt                    shakespeare.txt
mapper.py                        SumReducer.java
MAPREDUCE em Mongodb.txt         united_airlines_tweets.tsv
ml-latest                        upgrade pig.txt
ml-latest.zip                    WordCount.java
Movies.csv                       WordMapper.java
[oracle@bigdatalite Downloads]$
```

hadoop fs -ls

```
[oracle@bigdatalite Downloads]$ hadoop fs -ls
Found 10 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 2021-09-20 11:00 .staging
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
```

hadoop fs -copyFromLocal armas.txt dados

```
[oracle@bigdatalite Downloads]$ hadoop fs -copyFromLocal armas.txt dados
```

hadoop fs -ls dados

```
[oracle@bigdatalite Downloads]$ hadoop fs -ls dados
Found 1 items
-rw-r--r--   1 oracle oracle      30926 2021-09-21 07:15 dados/armas.txt
```

hadoop fs -chmod 664 dados/armas.txt

```
[oracle@bigdatalite Downloads]$ hadoop fs -chmod 664 dados/armas.txt
[oracle@bigdatalite Downloads]$
```

spark-shell

```
[oracle@bigdatalite Downloads]$ spark-shell
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel).
Welcome to


      ____              __
     / __/__  ___ _____/ /__
    _\ \/ _ \/ _ `/ __/  '_/
   /___/ .__/\_,_/_/ /_/\_\   version 1.6.0
      /_/

Using Scala version 2.10.5 (Java HotSpot(TM) 64-Bit Server VM, Java 1.8.0_91)
Type in expressions to have them evaluated.
Type :help for more information.
21/09/21 07:19:48 WARN util.Utils: Your hostname, bigdatalite.localdomain resolv
es to a loopback address: 127.0.0.1; using 10.0.2.15 instead (on interface eth1)
21/09/21 07:19:48 WARN util.Utils: Set SPARK_LOCAL_IP if you need to bind to ano
ther address
Spark context available as sc (master = yarn-client, app id = application_163162
2112260_0009).
SQL context available as sqlContext.

scala>
```

val textFile=sc.textFile("dados/armas.txt")

```
scala> val textFile=sc.textFile("dados/armas.txt")
textFile: org.apache.spark.rdd.RDD[String] = dados/armas.txt MapPartitionsRDD[1]
 at textFile at <console>:27

scala>
```

textFile.coun

```
scala> textFile.count
res0: Long = 953

scala>
```

textFIle.first

```
scala> textFile.first
res1: String = As armas e os barões assinalados,

scala>
```

val linhascomtodo = textFile.filter(line=>line.contains("todo"))

```
scala> val linhascomtodo = textFile.filter(line=>line.contains("todo"))
linhascomtodo: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[2] at filter
at <console>:29

scala>
```

linhascomtodo.collect()

```
scala> linhascomtodo.collect()
res2: Array[String] = Array(Dada ao mundo por Deus, que todo o mande,, (Que pelo
 mundo todo fa�a espanto), T�tis todo o cer�leo senhorio, Os outros Deuses todos
 assentados,, Entre todos as partes em porfia,, E n�ctar sobre todos esparziu.,
T�o querido de todos, e benquisto,, De todos os que as ondas navegamos, Tanto po
r todo o mundo se estenderam,, Aquele que criou todo o Hemisf�rio,, Tudo o que s
ente, o todo o insens�bil;, Com gesto ledo a todos, e fingido., Com que seja de
todo destru�do., Que quase todo o mar t�m destru�do, Contra n�s; e que todos seu
s intentos, Sagaz, astuto, e s�bio em todo o dano,, J� a ilha e todo o mais dese
mparando,)

scala>
```

linhascomtodo.collect().foreach(println)

```
scala> linhascomtodo.collect().foreach(println)
Dada ao mundo por Deus, que todo o mande,
(Que pelo mundo todo fa�a espanto)
T�tis todo o cer�leo senhorio
Os outros Deuses todos assentados,
Entre todos as partes em porfia,
E n�ctar sobre todos esparziu.
T�o querido de todos, e benquisto,
De todos os que as ondas navegamos
Tanto por todo o mundo se estenderam,
Aquele que criou todo o Hemisf�rio,
Tudo o que sente, o todo o insens�bil;
Com gesto ledo a todos, e fingido.
Com que seja de todo destru�do.
Que quase todo o mar t�m destru�do
Contra n�s; e que todos seus intentos
Sagaz, astuto, e s�bio em todo o dano,
J� a ilha e todo o mais desemparando,

scala>
```

textFile.filter(line => line.contains("todo")).count

```
scala> textFile.filter(line => line.contains("todo")).count
res4: Long = 17

scala>
```

val wordArray = linhascomtodo.flatMap(line => line.split(" "))

```
scala> val wordArray = linhascomtodo.flatMap(line => line.split(" "))
wordArray: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[4] at flatMap at
<console>:31

scala>
```

wordArray.collect

```
scala> wordArray.collect
res5: Array[String] = Array(Dada, ao, mundo, por, Deus,, que, todo, o, mande,, (
Que, pelo, mundo, todo, fa�a, espanto), T�tis, todo, o, cer�leo, senhorio, Os, o
utros, Deuses, todos, assentados,, Entre, todos, as, partes, em, porfia,, E, n�c
tar, sobre, todos, esparziu., T�o, querido, de, todos,, e, benquisto, De, todos
, os, que, as, ondas, navegamos, Tanto, por, todo, o, mundo, se, estenderam,, Aq
uele, que, criou, todo, o, Hemisf�rio,, Tudo, o, que, sente,, o, todo, o, insens
�bil;, Com, gesto, ledo, a, todos,, e, fingido., Com, que, seja, de, todo, destr
u�do., Que, quase, todo, o, mar, t�m, destru�do, Contra, n�s;, e, que, todos, se
us, intentos, Sagaz,, astuto,, e, s�bio, em, todo, o, dano,, J�, a, ilha, e, tod
o, o, mais, desemparando,)

scala>
```

val wordcountPairs = wordArray.map(word => (word,1))

```
scala> val wordcountPairs = wordArray.map(word => (word,1))
wordcountPairs: org.apache.spark.rdd.RDD[(String, Int)] = MapPartitionsRDD[5] at
 map at <console>:33

scala>
```

wordcountPairs.collect

```
scala> wordcountPairs.collect
res6: Array[(String, Int)] = Array((Dada,1), (ao,1), (mundo,1), (por,1), (Deus,,
1), (que,1), (todo,1), (o,1), (mande,,1), ((Que,1), (pelo,1), (mundo,1), (todo,1
), (fa�a,1), (espanto),1), (T�tis,1), (todo,1), (o,1), (cer�leo,1), (senhorio,1)
, (Os,1), (outros,1), (Deuses,1), (todos,1), (assentados,,1), (Entre,1), (todos,
1), (as,1), (partes,1), (em,1), (porfia,,1), (E,1), (n�ctar,1), (sobre,1), (todo
s,1), (esparziu.,1), (T�o,1), (querido,1), (de,1), (todos,,1), (e,1), (benquisto
,,1), (De,1), (todos,1), (os,1), (que,1), (as,1), (ondas,1), (navegamos,1), (Tan
to,1), (por,1), (todo,1), (o,1), (mundo,1), (se,1), (estenderam,,1), (Aquele,1),
 (que,1), (criou,1), (todo,1), (o,1), (Hemisf�rio,,1), (Tudo,1), (o,1), (que,1),
 (sente,,1), (o,1), (todo,1), (o,1), (insens�bil;,1), (Com,1), (gesto,1), (le...
scala>
```

val wordcounts = wordcountPairs.reduceByKey((x,y) => x+y)

```
scala> wordcounts.collect
res7: Array[(String, Int)] = Array((Dada,1), (ilha,1), (os,1), (em,2), (Os,1), (
t�m,1), (Sagaz,,1), (gesto,1), (todo,10), (estenderam,,1), (Tanto,1), (Entre,1),
 (as,2), (pelo,1), (dano,,1), (criou,1), (T�o,1), (mais,1), (benquisto,,1), (fin
gido.,1), (ledo,1), (seus,1), (astuto,,1), (outros,1), (Tudo,1), (se,1), (intent
os,1), (s�bio,1), (mar,1), (ao,1), (J�,1), (insens�bil;,1), (e,5), (T�tis,1), (q
ue,6), (n�ctar,1), (porfia,,1), (de,2), (a,2), (cer�leo,1), (Aquele,1), (navegam
os,1), (quase,1), (sobre,1), (esparziu.,1), (o,10), (por,2), (destru�do.,1), (He
misf�rio,,1), (destru�do,1), (mande,,1), (Contra,1), (Deus,,1), (desemparando,,1
), (querido,1), (Com,2), (partes,1), (n�s;,1), (espanto),1), (todos,,2), (sente,
,1), (E,1), (todos,5), (seja,1), (senhorio,1), (Que,1), ((Que,1), (fa�a,1), (...
scala>
```

wordcounts.toDF().show()

```
scala> wordcounts.toDF().show()
+----------+---+
|        _1| _2|
+----------+---+
|      Dada|  1|
|      ilha|  1|
|        os|  1|
|        em|  2|
|        Os|  1|
|       t�m|  1|
|    Sagaz,|  1|
|     gesto|  1|
|      todo| 10|
|estenderam,|  1|
|     Tanto|  1|
|     Entre|  1|
|        as|  2|
|      pelo|  1|
|     dano,|  1|
|     criou|  1|
|       T�o|  1|
|      mais|  1|
| benquisto,|  1|
```

Exercício:

ls

```
[oracle@bigdatalite Downloads]$ ls
AAA-r0_LQkT0ze54KnzYzhDMa?d1=0  pig_1632143168230.log
alunos.csv                      pig_1632145316643.log
armas.txt                       pig_1632150587213.log
Avaliação_MongoDB.docx          population.csv
Cassandra.pdf                   Redis.pdf
catalogo_de_galaxias.csv        reducer.py
Cidades.csv                     sentiment_analysis
csv_gen.sh                      sentiment_analysis_challenge.pig
dictionary.tsv                  sentiment_analysis.pig
empregado.csv                   sentiment_analysis_teste
empregados.csv                  sfpd.csv
empregado.txt                   shakespeare.txt
mapper.py                       SumReducer.java
MAPREDUCE em Mongodb.txt        united_airlines_tweets.tsv
ml-latest                       upgrade pig.txt
ml-latest.zip                   WordCount.java
Movies.csv                      WordMapper.java
```

hadoop fs -copyFromLocal shakespeare.txt dados

```
[oracle@bigdatalite Downloads]$ hadoop fs -copyFromLocal shakespeare.txt dados
[oracle@bigdatalite Downloads]$ █
```

hadoop fs -ls dados

```
[oracle@bigdatalite Downloads]$ hadoop fs -ls dados
Found 2 items
-rw-rw-r--   1 oracle oracle      30926 2021-09-21 07:15 dados/armas.txt
-rw-r--r--   1 oracle oracle    8877968 2021-09-21 07:41 dados/shakespeare.txt
```

val textFile=sc.textFile("dados/shakespeare")

```
scala> val textFile=sc.textFile("dados/shakespeare.txt")
textFile: org.apache.spark.rdd.RDD[String] = dados/shakespeare.txt MapPartitions
RDD[3] at textFile at <console>:27

scala> textFile.count
res1: Long = 175436

scala>
```

textFile.count

```
scala> textFile.count
res1: Long = 175436
```

textFile.first

```
scala> textFile.first
res3: String = hamlet@0          HAMLET
```

val linhascomthe = textFile.filter(line => line.contains("the"))

```
scala> val linhascomthe = textFile.filter(line => line.contains("the"))
linhascomthe: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[4] at filter a
t <console>:29

scala>
```

linhascomthe.collect()

```
scala> linhascomthe.collect()
res5: Array[String] = Array(hamlet@75   HAMLET  son to the late, and nephew to t
he present king., hamlet@230    LUCIANUS        nephew to the king., hamlet@726G
ERTRUDE queen of Denmark, and mother to Hamlet., hamlet@883              and othe
r Attendants. (Lord:), hamlet@946               Ghost of Hamlet's Father. (Ghost
:), hamlet@1023 SCENE I Elsinore. A platform before the castle., hamlet@1122    B
ERNARDO Who's there?, hamlet@1200       BERNARDO        Long live the king!, ham
let@1316        BERNARDO        'Tis now struck twelve; get thee to bed, Francis
co., hamlet@1596                The rivals of my watch, bid them make haste., ha
mlet@1643       FRANCISCO       I think I hear them. Stand, ho! Who's there?, ha
mlet@1764       MARCELLUS       And liegemen to the Dane., hamlet@2004          W
hat, is Horatio there?, hamlet@2369             With us to watch the minutes of
this night;, hamlet@2771                When yond sa...
scala>
```

linhascomthe.collect().foreach(println)

```
scala> linhascomthe.collect().foreach(println)
loverscomplaint@11949    Feeling it break, with bleeding groans they pine;
loverscomplaint@11999    And supplicant their sighs to you extend,
loverscomplaint@12041    To leave the battery that you make 'gainst mine,
loverscomplaint@12265    Whose sights till then were levell'd on my face;
loverscomplaint@12398    O, how the channel to the stream gave grace!
loverscomplaint@12443    Who glazed with crystal gate the glowing roses
loverscomplaint@12490    That flame through water which their hue encloses.
loverscomplaint@12542    'O father, what a hell of witchcraft lies
loverscomplaint@12584    In the small orb of one particular tear!
loverscomplaint@12625    But with the inundation of the eyes
loverscomplaint@12881    Even there resolved my reason into tears;
loverscomplaint@13315    In either's aptness, as it best deceives,
loverscomplaint@13487    Could 'scape the hail of his all-hurting aim,
loverscomplaint@13576    And, veil'd in them, did win whom he would maim:
loverscomplaint@13625    Against the thing he sought he would exclaim;
loverscomplaint@13766    'Thus merely with the garment of a Grace
loverscomplaint@13849    That th' unexperient gave the tempter place,
loverscomplaint@13894    Which like a cherubin above them hover'd.
loverscomplaint@14288    Would yet again betray the fore-betray'd,
README@477       The Taming of the Shrew
README@878       Othello
README@886       Timon of Athens
```

textFile.filter(line => line.contains("the")).count

```
scala> textFile.filter(line => line.contains("the")).count
res7: Long = 50962

scala>
```

val wordArray = linhascomthe.flatMap(line => line.split(" "))

```
scala> val wordArray = linhascomthe.flatMap(line => line.split(" "))
wordArray: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[6] at flatMap at
<console>:31

scala>
```

wordArray.collect

```
scala> wordArray.collect
res8: Array[String] = Array(hamlet@75    HAMLET  son, to, the, late,, and, nephew
, to, the, present, king., hamlet@230    LUCIANUS        nephew, to, the, king.,
hamlet@726       GERTRUDE        queen, of, Denmark,, and, mother, to, Hamlet., h
amlet@883               and, other, Attendants., (Lord:), hamlet@946            G
host, of, Hamlet's, Father., (Ghost:), hamlet@1023       SCENE, I        Elsinore
., A, platform, before, the, castle., hamlet@1122       BERNARDO        Who's, t
here?, hamlet@1200       BERNARDO        Long, live, the, king!, hamlet@1316    B
ERNARDO 'Tis, now, struck, twelve;, get, thee, to, bed,, Francisco., hamlet@1596
The, rivals, of, my, watch,, bid, them, make, haste., hamlet@1643        FRANCISC
O       I, think, I, hear, them., Stand,, ho!, Who's, there?, hamlet@1764       M
ARCELLUS        And, liegemen, to, the, Dane., hamlet@2004              What,, i
s, Horatio, there?, hamlet@2369        With...
scala>
```

val wordcountPairs = wordArray.map(word => (word,1))

```
scala> val wordcountPairs = wordArray.map(word => (word,1))
wordcountPairs: org.apache.spark.rdd.RDD[(String, Int)] = MapPartitionsRDD[7] at
 map at <console>:33

scala>
```

wordcountPairs.collect

```
scala> wordcountPairs.collect
res9: Array[(String, Int)] = Array((hamlet@75    HAMLET   son,1), (to,1), (the,1),
 (late,,1), (and,1), (nephew,1), (to,1), (the,1), (present,1), (king.,1), (hamle
t@230   LUCIANUS       nephew,1), (to,1), (the,1), (king.,1), (hamlet@726      G
ERTRUDE queen,1), (of,1), (Denmark,,1), (and,1), (mother,1), (to,1), (Hamlet.,1)
, (hamlet@883         and,1), (other,1), (Attendants.,1), ((Lord:),1), (hamlet
@946          Ghost,1), (of,1), (Hamlet's,1), (Father.,1), ((Ghost:),1), (haml
et@1023 SCENE,1), (I     Elsinore.,1), (A,1), (platform,1), (before,1), (the,1),
(castle.,1), (hamlet@1122       BERNARDO       Who's,1), (there?,1), (hamlet@12
00      BERNARDO       Long,1), (live,1), (the,1), (king!,1), (hamlet@1316    B
ERNARDO 'Tis,1), (now,1), (struck,1), (twelve;,1), (get,1), (thee,1), (to,1), (b
ed,,1), (Francisco.,1), (hamlet@1596              The,1), (rivals,1), (of,1...
scala> █
```

val wordcounts = wordcountPairs.reduceByKey((x,y) => x+y)

```
scala> val wordcounts = wordcountPairs.reduceByKey((x,y) => x+y)
wordcounts: org.apache.spark.rdd.RDD[(String, Int)] = ShuffledRDD[8] at reduceBy
Key at <console>:35

scala> █
```

wordcounts.collect

```
scala> wordcounts.collect
res10: Array[(String, Int)] = Array((timonofathens@72808              [March,1
), (durst,,1), (merrywivesofwindsor@109291       SIMPLE  Why,,1), (CAESAR       N
ay,,1), (bone,1), (bombast,2), (periclesprinceoftyre@23485            The,1),
(comedyoferrors@14062          Are,1), (sonnets@68826  To,1), (tamingoftheshrew
@120236 LUCENTIO       Here,1), (signal.,1), (hem,1), (Friend,1), (merrywivesof
windsor@27026          for,1), (kingjohn@79908          To,1), (timonofathens@77
198     PHRYNIA |,1), (hamlet@170612         liberal-conceited,1), (fuller,2)
, (muchadoaboutnothing@56100          For,1), (kinghenryv@93912            n
e'er,1), (2kinghenryvi@22254     QUEEN,1), (kinglear@15097            that,1),
 (othello@69344         To,1), (jade,4), (venusandadonis@37929  Having,1), (coun
tervail,1), (crying,7), (soon;,1), (kinghenryv@90679           his,1), (breath,
57), (coriolanus@158646 Second,1), (venusandadonis@11872         W...
scala> █
```

wordcounts.toDF().show()

```
scala> wordcounts.toDF().show()
+--------------------+---+
|                  _1| _2|
+--------------------+---+
|timonofathens@728...|  1|
|              durst,|  1|
|merrywivesofwinds...|  1|
|      CAESAR     Nay,|  1|
|                bone|  1|
|             bombast|  2|
|periclesprinceoft...|  1|
|comedyoferrors@14...|  1|
|    sonnets@68826  To|  1|
|tamingoftheshrew@...|  1|
|             signal.|  1|
|                 hem|  1|
|              Friend|  1|
|merrywivesofwinds...|  1|
|  kingjohn@79908   To|  1|
|timonofathens@771...|  1|
|hamlet@170612    li...|  1|
|              fuller|  2|
|muchadoaboutnothi...|  1|
```

hadoop fs -copyFromLocal alunos.csv dados

```
[oracle@bigdatalite Downloads]$ hadoop fs -copyFromLocal alunos.csv dados
[oracle@bigdatalite Downloads]$ ▮
```

hadoop fs -ls dados

```
[oracle@bigdatalite Downloads]$ hadoop fs -ls dados
Found 3 items
-rw-r--r--   1 oracle oracle       1146 2021-09-21 08:05 dados/alunos.csv
-rw-rw-r--   1 oracle oracle      30926 2021-09-21 07:15 dados/armas.txt
-rw-rw-r--   1 oracle oracle    8877968 2021-09-21 07:41 dados/shakespeare.txt
```

hadoop fs -chmod 664 dados/alunos.csv

```
[oracle@bigdatalite Downloads]$ hadoop fs -chmod 664 dados/alunos.csv
[oracle@bigdatalite Downloads]$ ▮
```

val alunosFile=sc.textFile("dados/alunos.csv")

```
scala> val alunosFile=sc.textFile("dados/alunos.csv")
alunosFile: org.apache.spark.rdd.RDD[String] = dados/alunos.csv MapPartitionsRDD
[12] at textFile at <console>:27

scala> ▮
```

alunosFile.count()

```
scala> alunosFile.count()
res12: Long = 30

scala> ▮
```

alunosFile.first()

```
scala> alunosFile.first()
res13: String = 1,Humberto,Silva,36,929451722,Aracaju

scala> ▮
```

val linhascomrosa = alunosFile.filter(line => line.contains("Rosa"))

```
scala> val linhascomrosa = alunosFile.filter(line => line.contains("Rosa"))
linhascomrosa: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[14] at filter
 at <console>:29

scala> ▮
```

linhascomrosa.collect()

```
scala> linhascomrosa.collect()
res15: Array[String] = Array(6,Rosa,Drauzio,35,988782874,Barra)

scala> ▮
```

alunosFile.filter(line => line.contains("Maria")).count

```
scala> alunosFile.filter(line => line.contains("Maria")).count
res16: Long = 1

scala> ▮
```