

Estudante: Clístenes Grizafis Bento

Professor: João Eugenio Marynowski

Atividade 2 – Executar o Contador de Palavras (wordcount)

```
$ hdfs dfs -ls /  
$ hdfs dfs -put /var/log/mysql/mysql.log /  
$ hdfs dfs -ls /  
$ hadoop jar /usr/lib/hadoop-mapreduce/hadoop-  
mapreduce-examples.jar wordcount /mysql.log /saida  
$ hdfs dfs -cat /saida/part*  
$ hdfs dfs -get /saida /root/  
$ ls -l /root/saida  
$ cat /root/saida/part*
```

root@sandbox:~

```
[root@sandbox ~]# hdfs dfs -ls /
```

Found 6 items

```
drwxrwxrwx - yarn  hadoop          0 2022-11-19 05:55 /app-logs
drwxr-xr-x - hdfs  hdfs           0 2014-04-21 07:23 /apps
drwxr-xr-x - mapred hdfs           0 2014-04-21 07:16 /mapred
drwxr-xr-x - hdfs  hdfs           0 2014-04-21 07:16 /mr-history
drwxrwxrwx - hdfs  hdfs           0 2014-04-22 07:21 /tmp
drwxr-xr-x - hdfs  hdfs           0 2014-04-22 07:21 /user
```

```
[root@sandbox ~]# hdfs dfs -put /va
```

vagrant/ var/

```
[root@sandbox ~]# hdfs dfs -put /var/log/mysql.log
```

```
[root@sandbox ~]# hdfs dfs -ls /
```

Found 6 items

```
drwxrwxrwx - yarn  hadoop          0 2022-11-19 05:55 /app-logs
drwxr-xr-x - hdfs  hdfs           0 2014-04-21 07:23 /apps
drwxr-xr-x - mapred hdfs           0 2014-04-21 07:16 /mapred
drwxr-xr-x - hdfs  hdfs           0 2014-04-21 07:16 /mr-history
drwxrwxrwx - hdfs  hdfs           0 2014-04-22 07:21 /tmp
drwxr-xr-x - hdfs  hdfs           0 2014-04-22 07:21 /user
```

```
[root@sandbox ~]# hadoop jar /usr/lib/hadoop-mapreduce/hadoop-mapreduce-examples.jar wordcount /mysql.log /saida
```

22/11/19 06:48:16 INFO client.RMProxy: Connecting to ResourceManager at sandbox.hortonworks.com/10.0.2.15:8050

22/11/19 06:48:17 INFO mapreduce.JobSubmitter: Cleaning up the staging area /user/root/.staging/job_1668866818960_0004

org.apache.hadoop.mapreduce.lib.input.InvalidInputException: Input path does not exist: hdfs://sandbox.hortonworks.com:8020/mysql.log

at org.apache.hadoop.mapreduce.lib.input.FileInputFormat.singleThreadedListStatus(FileInputFormat.java:320)

at org.apache.hadoop.mapreduce.lib.input.FileInputFormat.listStatus(FileInputFormat.java:263)

at org.apache.hadoop.mapreduce.lib.input.FileInputFormat.get_splits(FileInputFormat.java:375)

at org.apache.hadoop.mapreduce.JobSubmitter.writeNew_splits(JobSubmitter.java:493)

at org.apache.hadoop.mapreduce.JobSubmitter.write_splits(JobSubmitter.java:510)

at org.apache.hadoop.mapreduce.JobSubmitter.submitJobInternal(JobSubmitter.java:394)

at org.apache.hadoop.mapreduce.Job\$10.run(Job.java:1285)

at org.apache.hadoop.mapreduce.Job\$10.run(Job.java:1282)

at java.security.AccessController.doPrivileged(Native Method)

at javax.security.auth.Subject.doAs(Subject.java:415)

at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1557)

at org.apache.hadoop.mapreduce.Job.submit(Job.java:1282)

at org.apache.hadoop.mapreduce.Job.waitForCompletion(Job.java:1303)

at org.apache.hadoop.examples.WordCount.main(WordCount.java:84)

at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)

at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:57)

at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:57)

```

root@sandbox:~
  at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:57)
  at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:
43)
  at java.lang.reflect.Method.invoke(Method.java:606)
  at org.apache.hadoop.util.ProgramDriver$ProgramDescription.invoke(ProgramDriver.java:
72)
  at org.apache.hadoop.util.ProgramDriver.run(ProgramDriver.java:145)
  at org.apache.hadoop.examples.ExampleDriver.main(ExampleDriver.java:74)
  at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
  at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:57)
  at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:
43)
  at java.lang.reflect.Method.invoke(Method.java:606)
  at org.apache.hadoop.util.RunJar.main(RunJar.java:212)
[root@sandbox ~]# hadoop jar /usr/lib/hadoop-mapreduce/hadoop-mapreduce-examples.jar wordcoun
t /
ambari-views/   etc/           media/         selinux/       var/
.autofsck       hadoop/        mnt/           srv/           virtualization
.autorelabel    home/          opt/           sys/
bin/            lib/           proc/          tmp/
boot/           lib64/         root/          usr/
dev/            lost+found/    sbin/          vagrant/
[root@sandbox ~]# hadoop jar /usr/lib/hadoop-mapreduce/hadoop-mapred^C
[root@sandbox ~]# clear
[root@sandbox ~]# hdfs dfs -ls /
Found 6 items
drwxrwxrwx - yarn  hadoop      0 2022-11-19 05:55 /app-logs
drwxr-xr-x - hdfs  hdfs      0 2014-04-21 07:23 /apps
drwxr-xr-x - mapred hdfs      0 2014-04-21 07:16 /mapred
drwxr-xr-x - hdfs  hdfs      0 2014-04-21 07:16 /mr-history
drwxrwxrwx - hdfs  hdfs      0 2014-04-22 07:21 /tmp
drwxr-xr-x - hdfs  hdfs      0 2014-04-22 07:21 /user
[root@sandbox ~]# hdfs dfs -put /var/log/mysqld.log /
[root@sandbox ~]# hdfs dfs -ls /
Found 7 items
drwxrwxrwx - yarn  hadoop      0 2022-11-19 05:55 /app-logs
drwxr-xr-x - hdfs  hdfs      0 2014-04-21 07:23 /apps
drwxr-xr-x - mapred hdfs      0 2014-04-21 07:16 /mapred
drwxr-xr-x - hdfs  hdfs      0 2014-04-21 07:16 /mr-history
-rw-r--r-- 1 root  hdfs    4773 2022-11-19 06:50 /mysqld.log
drwxrwxrwx - hdfs  hdfs      0 2014-04-22 07:21 /tmp
drwxr-xr-x - hdfs  hdfs      0 2014-04-22 07:21 /user
[root@sandbox ~]# hadoop jar /usr/lib/hadoop-mapreduce/hadoop-mapreduce-examples.jar wordcoun
t /mysqld.log /saida
22/11/19 06:51:22 INFO client.RMProxy: Connecting to ResourceManager at sandbox.hortonworks.c
om/10.0.2.15:8050
22/11/19 06:51:23 INFO input.FileInputFormat: Total input paths to process : 1
22/11/19 06:51:23 INFO mapreduce.JobSubmitter: number of splits:1

```

root@sandbox:~

```
22/11/19 06:51:23 INFO mapreduce.JobSubmitter: number of splits:1
22/11/19 06:51:24 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1668866818960_0005
22/11/19 06:51:24 INFO impl.YarnClientImpl: Submitted application application_1668866818960_0005
22/11/19 06:51:24 INFO mapreduce.Job: The url to track the job: http://sandbox.hortonworks.com:8088/proxy/application_1668866818960_0005/
22/11/19 06:51:24 INFO mapreduce.Job: Running job: job_1668866818960_0005
22/11/19 06:51:33 INFO mapreduce.Job: Job job_1668866818960_0005 running in uber mode : false
22/11/19 06:51:33 INFO mapreduce.Job: map 0% reduce 0%
22/11/19 06:51:40 INFO mapreduce.Job: map 100% reduce 0%
22/11/19 06:51:48 INFO mapreduce.Job: map 100% reduce 100%
22/11/19 06:51:48 INFO mapreduce.Job: Job job_1668866818960_0005 completed successfully
22/11/19 06:51:48 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=1977
    FILE: Number of bytes written=202335
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=4884
    HDFS: Number of bytes written=1421
    HDFS: Number of read operations=6
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
  Job Counters
    Launched map tasks=1
    Launched reduce tasks=1
    Data-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=4891
    Total time spent by all reduces in occupied slots (ms)=5287
    Total time spent by all map tasks (ms)=4891
    Total time spent by all reduce tasks (ms)=5287
    Total vcore-seconds taken by all map tasks=4891
    Total vcore-seconds taken by all reduce tasks=5287
    Total megabyte-seconds taken by all map tasks=1222750
    Total megabyte-seconds taken by all reduce tasks=1321750
  Map-Reduce Framework
    Map input records=80
    Map output records=604
    Map output bytes=7099
    Map output materialized bytes=1977
    Input split bytes=111
    Combine input records=604
    Combine output records=141
    Reduce input groups=141
    Reduce shuffle bytes=1977
    Reduce input records=141
```

— □ ×

[illegible]

root@sandbox:~

```
5      2
6:06:13 6
787981  2
7:04:46 8
7:04:47 3
7:04:52 2
7:14:51 5
7:18:07 5
7:22:27 2
7:22:29 1
7:22:33 2
7:36:13 3
7:36:18 2
8.0M    5
802111  2
804306  1
=        5
Completed      5
Creating       1
Database       4
Doublewrite    2
Event         8
Foreign        1
Initializing    5
InnoDB: 42
Loaded         5
Log            2
MB             3
Normal         3
Purging        3
Reading        1
Restoring      1
Scheduler:     8
Setting        3
Shutdown       6
Source         5
Started;       5
Starting       9
The            2
Version:      5
[Note] 19
a             1
be           3
buffer 12
buffer...     1
complete      3
completed;    3
connections.  5
```

root@sandbox:~

```
connections. 5
constraint 2
crash 1
created 4
created! 1
creating 1
daemon 5
data 2
database 1
databases 5
did 3
distribution 5
does 1
doublewrite 1
down 1
ended 3
events 8
exist: 3
file 12
files 1
files... 1
first 1
for 5
foreign 1
found: 1
from 10
full: 3
half-written 1
ib_logfiles! 1
ibdata 1
in 2
information 1
initialization 5
key 2
log 12
match 1
mysqld 8
mysqld_safe 8
new 4
normally! 1
not 6
number 10
of 5
pages 1
physically 3
pid 3
pool 5
pool, 5
```

root@sandbox:~

```
pool, 5
port: 5
possible 1
queue. 3
ready 5
recovery. 1
sequence 10
shut 1
shutdown 3
shutdown... 3
size 8
socket: 5
specified 1
system 2
tables 2
tablespace 1
the 10
to 6
wait... 3
was 1
with 5
writes 3
[root@sandbox ~]# hdfsdfs -get /saida /root/
-bash: hdfsdfs: command not found
[root@sandbox ~]# ls -l /root/saida
ls: cannot access /root/saida: No such file or directory
[root@sandbox ~]# hdfsdfs -get /saida /root/
-bash: hdfsdfs: command not found
[root@sandbox ~]# hdfsdfs -get
.bash_history .bashrc .pki/ start_ambari.sh
.bash_logout .cshrc .rnd start_hbase.sh
.bash_profile .hdfs_prepared .ssh/ .tcshrc
[root@sandbox ~]# hdfs dfs -get /saida /root/
[root@sandbox ~]# ls -l /root/saida/
total 4
-rw-r--r-- 1 root root 1421 Nov 19 06:54 part-r-00000
-rw-r--r-- 1 root root 0 Nov 19 06:54 _SUCCESS
[root@sandbox ~]# cat /root/saida/part*
'/var/lib/mysql/mysql.sock' 5
'5.1.73' 5
./ib_logfile0 2
./ib_logfile1 2
./ibdata1 2
libd 1
/usr/libexec/mysqld: 11
/var/lib/mysql 5
/var/run/mysqld/mysqld.pid 3
0 17
```



```
root@sandbox:~  
0      17  
04:05:55      1  
06:06:13      1  
07:04:46      1  
07:04:52      1  
07:14:51      1  
07:18:06      1  
07:22:33      1  
07:36:18      1  
10      1  
140421  27  
140422  12  
221119  13  
3306    5  
44233   2  
4:05:55 5  
5        2  
6:06:13 6  
787981  2  
7:04:46 8  
7:04:47 3  
7:04:52 2  
7:14:51 5  
7:18:07 5  
7:22:27 2  
7:22:29 1  
7:22:33 2  
7:36:13 3  
7:36:18 2  
8.0M    5  
802111  2  
804306  1  
=        5  
Completed      5  
Creating       1  
Database       4  
Doublewrite    2  
Event          8  
Foreign        1  
Initializing    5  
InnoDB: 42  
Loaded         5  
Log            2  
MB             3  
Normal         3  
Purging        3  
Reading        1  
Restoring      1
```

root@sandbox:~

```
Restoring      1
Scheduler:     8
Setting 3
Shutdown       6
Source 5
Started;       5
Starting       9
The 2
Version:       5
[Note] 19
a 1
be 3
buffer 12
buffer...     1
complete      3
completed;    3
connections.  5
constraint    2
crash 1
created 4
created!      1
creating      1
daemon 5
data 2
database      1
databases     5
did 3
distribution  5
does 1
doublewrite   1
down 1
ended 3
events 8
exist: 3
file 12
files 1
files...      1
first 1
for 5
foreign 1
found: 1
from 10
full: 3
half-written  1
ib_logfiles!  1
ibdata 1
in 2
information   1
```

root@sandbox:~

```
for 5
foreign 1
found: 1
from 10
full: 3
half-written 1
ib_logfiles! 1
ibdata 1
in 2
information 1
initialization 5
key 2
log 12
match 1
mysqld 8
mysqld_safe 8
new 4
normally! 1
not 6
number 10
of 5
pages 1
physically 3
pid 3
pool 5
pool, 5
port: 5
possible 1
queue. 3
ready 5
recovery. 1
sequence 10
shut 1
shutdown 3
shutdown... 3
size 8
socket: 5
specified 1
system 2
tables 2
tablespace 1
the 10
to 6
wait... 3
was 1
with 5
writes 3
[root@sandbox ~]#
```