

Estudante: Clístenes Grizafis Bento

Professor: João Eugênio Marynowski

Atividade 4

Dado o conjunto de programas exemplos disponível em:

```
$ hadoop jar /usr/lib/hadoop-mapreduce/hadoop-mapreduce-examples.jar
```

Comente brevemente justificando e apresentando a aplicação de algum outro programa exemplo.

Resposta: O programa escolhido foi o grep, que conta as correspondências de uma REGEX no arquivo de entrada, pode ser usado para identificar a existência de determinada palavra em emails ou arquivos enviados para pessoas, ajudando a identificação de documentos span.

Abaixo segue exemplo de identificação em um documento .js de um jogo sobre capivaras com a palavra “capivara”.

root@sandbox:~

```
.jar grep /user/hue/main.js /atividade_4 "capivara"
22/11/19 10:21:39 INFO client.RMPProxy: Connecting to ResourceManager at sandbox.
hortonworks.com/10.0.2.15:8050
22/11/19 10:21:40 WARN mapreduce.JobSubmitter: No job jar file set. User classe
s may not be found. See Job or Job#setJar(String).
22/11/19 10:21:40 INFO input.FileInputFormat: Total input paths to process : 1
22/11/19 10:21:40 INFO mapreduce.JobSubmitter: number of splits:1
22/11/19 10:21:41 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_16
68876909585_0003
22/11/19 10:21:41 INFO mapred.YARNRunner: Job jar is not present. Not adding any
jar to the list of resources.
22/11/19 10:21:41 INFO impl.YarnClientImpl: Submitted application application_16
68876909585_0003
22/11/19 10:21:41 INFO mapreduce.Job: The url to track the job: http://sandbox.h
ortonworks.com:8088/proxy/application_1668876909585_0003/
22/11/19 10:21:41 INFO mapreduce.Job: Running job: job_1668876909585_0003
22/11/19 10:21:48 INFO mapreduce.Job: Job job_1668876909585_0003 running in uber
mode : false
22/11/19 10:21:48 INFO mapreduce.Job: map 0% reduce 0%
22/11/19 10:21:53 INFO mapreduce.Job: map 100% reduce 0%
22/11/19 10:22:00 INFO mapreduce.Job: map 100% reduce 100%
22/11/19 10:22:00 INFO mapreduce.Job: Job job_1668876909585_0003 completed succe
ssfully
22/11/19 10:22:00 INFO mapreduce.Job: Counters: 49
    File System Counters
        FILE: Number of bytes read=25
        FILE: Number of bytes written=198827
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=32165
        HDFS: Number of bytes written=111
        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)=3571
        Total time spent by all reduces in occupied slots (ms)=3785
        Total time spent by all map tasks (ms)=3571
        Total time spent by all reduce tasks (ms)=3785
        Total vcore-seconds taken by all map tasks=3571
        Total vcore-seconds taken by all reduce tasks=3785
        Total megabyte-seconds taken by all map tasks=892750
        Total megabyte-seconds taken by all reduce tasks=946250
Map-Reduce Framework
```

root@sandbox:~

```
Map input records=928
Map output records=81
Map output bytes=1377
Map output materialized bytes=25
Input split bytes=117
Combine input records=81
Combine output records=1
Reduce input groups=1
Reduce shuffle bytes=25
Reduce input records=1
Reduce output records=1
Spilled Records=2
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=96
CPU time spent (ms)=1350
Physical memory (bytes) snapshot=339984384
Virtual memory (bytes) snapshot=1802756096
Total committed heap usage (bytes)=209190912

Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0

File Input Format Counters
  Bytes Read=32048
File Output Format Counters
  Bytes Written=111
```

```
22/11/19 10:22:00 INFO client.RMPProxy: Connecting to ResourceManager at sandbox.
hortonworks.com/10.0.2.15:8050
22/11/19 10:22:00 WARN mapreduce.JobSubmitter: No job jar file set.  User classe
s may not be found.  See Job or Job#setJar(String).
22/11/19 10:22:00 INFO input.FileInputFormat: Total input paths to process : 1
22/11/19 10:22:00 INFO mapreduce.JobSubmitter: number of splits:1
22/11/19 10:22:00 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_16
68876909585_0004
22/11/19 10:22:00 INFO mapred.YARNRunner: Job jar is not present.  Not adding any
jar to the list of resources.
22/11/19 10:22:00 INFO impl.YarnClientImpl: Submitted application application_16
68876909585_0004
22/11/19 10:22:00 INFO mapreduce.Job: The url to track the job: http://sandbox.h
ortonworks.com:8088/proxy/application_1668876909585_0004/
22/11/19 10:22:00 INFO mapreduce.Job: Running job: job_1668876909585_0004
22/11/19 10:22:07 INFO mapreduce.Job: Job job_1668876909585_0004 running in uber
mode : false
```

```
root@sandbox:~  
mode : false  
22/11/19 10:22:07 INFO mapreduce.Job: map 0% reduce 0%  
22/11/19 10:22:14 INFO mapreduce.Job: map 100% reduce 0%  
22/11/19 10:22:22 INFO mapreduce.Job: map 100% reduce 100%  
22/11/19 10:22:22 INFO mapreduce.Job: Job job_1668876909585_0004 completed successfully  
22/11/19 10:22:23 INFO mapreduce.Job: Counters: 49  
  File System Counters  
    FILE: Number of bytes read=25  
    FILE: Number of bytes written=197793  
    FILE: Number of read operations=0  
    FILE: Number of large read operations=0  
    FILE: Number of write operations=0  
    HDFS: Number of bytes read=255  
    HDFS: Number of bytes written=12  
    HDFS: Number of read operations=7  
    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)=5373  
    Total time spent by all reduces in occupied slots (ms)=5751  
    Total time spent by all map tasks (ms)=5373  
    Total time spent by all reduce tasks (ms)=5751  
    Total vcore-seconds taken by all map tasks=5373  
    Total vcore-seconds taken by all reduce tasks=5751  
    Total megabyte-seconds taken by all map tasks=1343250  
    Total megabyte-seconds taken by all reduce tasks=1437750  
  Map-Reduce Framework  
    Map input records=1  
    Map output records=1  
    Map output bytes=17  
    Map output materialized bytes=25  
    Input split bytes=144  
    Combine input records=0  
    Combine output records=0  
    Reduce input groups=1  
    Reduce shuffle bytes=25  
    Reduce input records=1  
    Reduce output records=1  
    Spilled Records=2  
    Shuffled Maps =1  
    Failed Shuffles=0  
    Merged Map outputs=1  
    GC time elapsed (ms)=55  
    CPU time spent (ms)=2140
```

```
CPU time spent (ms)=2140
Physical memory (bytes) snapshot=337547264
Virtual memory (bytes) snapshot=1795878912
Total committed heap usage (bytes)=207618048

Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0

File Input Format Counters
  Bytes Read=111

File Output Format Counters
  Bytes Written=12
[root@sandbox ~]# hdfs dfs -cat /atividade_4/part*
81      capivara
[root@sandbox ~]#
```