## **Exercises MapReduce**

- 1. Check files
   \$ hadoop fs -ls wordcount/input
- 2. Launch Job WordCount MapReduce

\$ hadoop jar /usr/lib/hadoop-0.20-mapreduce/hadoop-examples.jar wordcount
wordcount/input wordcount/output

Note: wordcount/input cannot exists or the job will fail.

```
.4/10/07 03:06:09 INFO mapred.JobClient: Running job: job_201410070252_0001
.4/10/07 03:06:10 INFO mapred.JobClient: map 0% reduce 0%
.4/10/07 03:06:21 INFO mapred.JobClient: map 50% reduce 0%
.4/10/07 03:06:23 INFO mapred.JobClient: map 100% reduce 0%
  4/10/07 03:06:28 INFO mapred.Jobclient: map 100% reduce 100%
4/10/07 03:06:31 INFO mapred.Jobclient: Job complete: job_201410070252_0001

      14/10/07 03:06:31 INFO mapred.JobClient: Job complete:

      14/10/07 03:06:31 INFO mapred.JobClient: Counters:
      33

      14/10/07 03:06:31 INFO mapred.JobClient: File System
      File System

      14/10/07 03:06:31 INFO mapred.JobClient: FILE: Num
      FILE: Num

                                                                                                                                 File System Counters

FILE: Number of bytes read=30573

FILE: Number of bytes written=564617

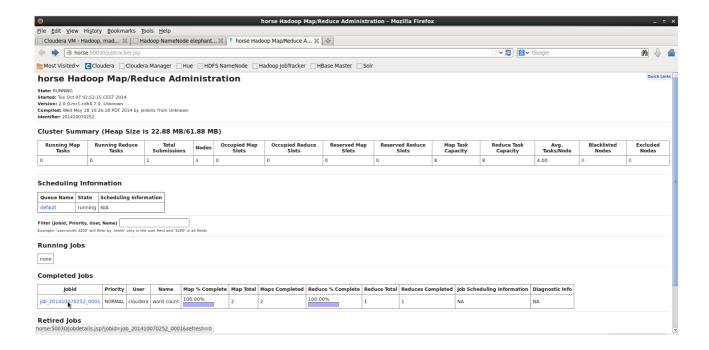
FILE: Number of read operations=0

FILE: Number of large read operations=0

HDFS: Number of bytes read=46759
   4/10/07 03:06:31 INFO mapred.JobClient:
4/10/07 03:06:31 INFO mapred.JobClient:
                                                                                                                                          HDFS: Number of bytes written=19283
 14/10/07 03:06:31 INFO mapred.JobClient:
                                                                                                                                         HDFS: Number of read operations=4
HDFS: Number of large read operations=0
                                                                                                                                         HDFS: Number of write operations=1
                                                                                                                                       Counters
Launched map tasks=2
Launched reduce tasks=1
Data-local map tasks=1
Rack-local map tasks=1
Total time spent by all maps in occupied slots (ms)=16914
Total time spent by all reduces in occupied slots (ms)=6181
Total time spent by all maps waiting after reserving slots (ms)=0
Total time spent by all reduces waiting after reserving slots (ms)=0
   4/10/07 03:06:31 INFO mapred.JobClient:
   4/10/07 03:06:31 INFO mapred.JobClient:
   4/10/07 03:06:31 INFO mapred.JobClient:
4/10/07 03:06:31 INFO mapred.JobClient:
 4/10/07 03:06:31 INFO mapred.Jobclient:
4/10/07 03:06:31 INFO mapred.Jobclient:
4/10/07 03:06:31 INFO mapred.Jobclient:
4/10/07 03:06:31 INFO mapred.Jobclient:
                                                                                                                                          Total time spent by all reduces waiting after reserving slots (ms)=0
                                                                                                                                   Map-Reduce Framework
                                                                                                                                         Map output records=7225
Map output bytes=73404
Input split bytes=254
Combine input records=7225
Combine output records=2152
 .4/10/07 03:06:31 INFO mapred.JobClient:
                                                                                                                                         Reduce input groups=1851
Reduce shuffle bytes=30579
  4/10/07 03:06:31 INFO mapred.JobClient:
4/10/07 03:06:31 INFO mapred.JobClient:
  4/10/07 03:06:31 INFO mapred.JobClient:
4/10/07 03:06:31 INFO mapred.JobClient:
                                                                                                                                          Reduce output records=1851
                                                                                                                                          Spilled Records=4304
 .4/10/07 03:06:31 INFO mapred.Jobclient:
.4/10/07 03:06:31 INFO mapred.Jobclient:
.4/10/07 03:06:31 INFO mapred.Jobclient:
.4/10/07 03:06:31 INFO mapred.Jobclient:
                                                                                                                                         Physical memory (bytes) snapshot=460414976
Virtual memory (bytes) snapshot=1860259840
Total committed heap usage (bytes)=270934016
      loudera@elephant ~]$
horse Hadoop Map/Re... 🔲 cloudera@elephant:~
```

3. Go to JobTracker web

http://horse:50030/jobtracker.jsp

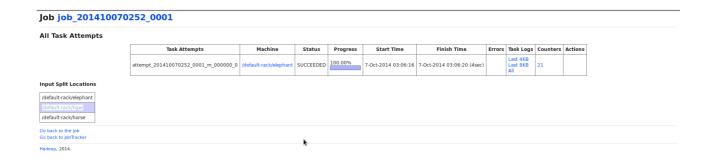


4. View job on "Completed Jobs" or "Running Jobs"



- 5. On Job File you can view metadata configuration for the job
- 6. Click on map link
- 7. Click on one from two tasks

Total time spent by all maps waiting after reserving slots (ms) Total time spent by all reduces waiting after recenting slots (ms) horse:50030/jobtasks.jsp?jobid=job\_201410070252\_0001&type=map&pagenum=1&state=completed



Which datanode read the block?