

## 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.

```
14/10/07 03:06:08 INFO input.FileInputFormat: Total input paths to process : 2
14/10/07 03:06:09 INFO mapred.JobClient: Running job: job_201410070252_0001
14/10/07 03:06:10 INFO mapred.JobClient: map 0% reduce 0%
14/10/07 03:06:21 INFO mapred.JobClient: map 50% reduce 0%
14/10/07 03:06:23 INFO mapred.JobClient: map 100% reduce 0%
14/10/07 03:06:28 INFO mapred.JobClient: map 100% reduce 100%
14/10/07 03:06:31 INFO mapred.JobClient: Job complete: job_201410070252_0001
14/10/07 03:06:31 INFO mapred.JobClient: Counters: 33
14/10/07 03:06:31 INFO mapred.JobClient: File System Counters
14/10/07 03:06:31 INFO mapred.JobClient: FILE: Number of bytes read=30573
14/10/07 03:06:31 INFO mapred.JobClient: FILE: Number of bytes written=564617
14/10/07 03:06:31 INFO mapred.JobClient: FILE: Number of read operations=0
14/10/07 03:06:31 INFO mapred.JobClient: FILE: Number of large read operations=0
14/10/07 03:06:31 INFO mapred.JobClient: FILE: Number of write operations=0
14/10/07 03:06:31 INFO mapred.JobClient: HDFS: Number of bytes read=46759
14/10/07 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
14/10/07 03:06:31 INFO mapred.JobClient: HDFS: Number of large read operations=0
14/10/07 03:06:31 INFO mapred.JobClient: HDFS: Number of write operations=1
14/10/07 03:06:31 INFO mapred.JobClient: Job Counters
14/10/07 03:06:31 INFO mapred.JobClient: Launched map tasks=2
14/10/07 03:06:31 INFO mapred.JobClient: Launched reduce tasks=1
14/10/07 03:06:31 INFO mapred.JobClient: Data-local map tasks=1
14/10/07 03:06:31 INFO mapred.JobClient: Rack-local map tasks=1
14/10/07 03:06:31 INFO mapred.JobClient: Total time spent by all maps in occupied slots (ms)=16914
14/10/07 03:06:31 INFO mapred.JobClient: Total time spent by all reduces in occupied slots (ms)=6181
14/10/07 03:06:31 INFO mapred.JobClient: Total time spent by all maps waiting after reserving slots (ms)=0
14/10/07 03:06:31 INFO mapred.JobClient: Total time spent by all reduces waiting after reserving slots (ms)=0
14/10/07 03:06:31 INFO mapred.JobClient: Map-Reduce Framework
14/10/07 03:06:31 INFO mapred.JobClient: Map input records=876
14/10/07 03:06:31 INFO mapred.JobClient: Map output records=7225
14/10/07 03:06:31 INFO mapred.JobClient: Map output bytes=73404
14/10/07 03:06:31 INFO mapred.JobClient: Input split bytes=254
14/10/07 03:06:31 INFO mapred.JobClient: Combine input records=7225
14/10/07 03:06:31 INFO mapred.JobClient: Combine output records=2152
14/10/07 03:06:31 INFO mapred.JobClient: Reduce input groups=1851
14/10/07 03:06:31 INFO mapred.JobClient: Reduce shuffle bytes=30579
14/10/07 03:06:31 INFO mapred.JobClient: Reduce input records=2152
14/10/07 03:06:31 INFO mapred.JobClient: Reduce output records=1851
14/10/07 03:06:31 INFO mapred.JobClient: Spilled Records=4304
14/10/07 03:06:31 INFO mapred.JobClient: CPU time spent (ms)=2760
14/10/07 03:06:31 INFO mapred.JobClient: Physical memory (bytes) snapshot=460414976
14/10/07 03:06:31 INFO mapred.JobClient: Virtual memory (bytes) snapshot=1860259840
14/10/07 03:06:31 INFO mapred.JobClient: Total committed heap usage (bytes)=270934016
[cloudera@elephant ~]$
```

3. Go to JobTracker web

<http://horse:50030/jobtracker.jsp>

horse Hadoop Map/Reduce Administration - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Cloudera VM - Hadoop, mad... Hadoop NameNode elephant... horse Hadoop Map/Reduce A...

horse:50030/jobtracker.jsp

Most Visited Cloudera Cloudera Manager Hue HDFS NameNode Hadoop JobTracker HBase Master Solr

### horse Hadoop Map/Reduce Administration

State: RUNNING  
Started: Tue Oct 07 02:52:15 CEST 2014  
Version: 2.0.0-mr1-cdh4.7.0. Unknown  
Compiled: Wed May 28 10:26:18 PDT 2014 by jenkins from Unknown  
Identifier: 201410070252

#### Cluster Summary (Heap Size is 22.88 MB/61.88 MB)

Running Map Tasks	Running Reduce Tasks	Total Submissions	Nodes	Occupied Map Slots	Occupied Reduce Slots	Reserved Map Slots	Reserved Reduce Slots	Map Task Capacity	Reduce Task Capacity	Avg. Tasks/Node	Blacklisted Nodes	Excluded Nodes
0	0	1	4	0	0	0	0	8	8	4.00	0	0

#### Scheduling Information

Queue Name	State	Scheduling Information
default	running	N/A

Filter (Jobid, Priority, User, Name)

Example: 'usersmith 3200' will filter by 'smith' only in the user field and '3200' in all fields

#### Running Jobs

none

#### Completed Jobs

Jobid	Priority	User	Name	Map % Complete	Map Total	Maps Completed	Reduce % Complete	Reduce Total	Reduces Completed	Job Scheduling Information	Diagnostic Info
<a href="#">job_201410070252_0001</a>	NORMAL	cloudera	word count	100.00%	2	2	100.00%	1	1	NA	NA

#### Retired Jobs

horse:50030/jobdetails.jsp?jobid=job\_201410070252\_0001&refresh=0

## 4. View job on “Completed Jobs” or “Running Jobs”

### Hadoop job\_201410070252\_0001 on horse

User: cloudera  
Job Name: word count  
Job File: [https://elephant.8020/user/cloudera/staging/job\\_201410070252\\_0001/job.xml](https://elephant.8020/user/cloudera/staging/job_201410070252_0001/job.xml)  
Submit Host: elephant  
Submit Host Address: 192.168.1.1  
Job-ACLs: All users are allowed  
Job Setup: Successful  
Status: Succeeded  
Started at: Tue Oct 07 03:06:09 CEST 2014  
Finished at: Tue Oct 07 03:06:31 CEST 2014  
Finished in: 22sec  
Job Cleanup: Successful

Kind	% Complete	Num Tasks	Pending	Running	Complete	Killed	Failed/Killed Task Attempts
map	100.00%	2	0	0	2	0	0 / 0
reduce	100.00%	1	0	0	1	0	0 / 0

Counter		Map	Reduce	Total
File System Counters	FILE: Number of bytes read	0	30,573	30,573
	FILE: Number of bytes written	366,451	198,166	564,617
	FILE: Number of read operations	0	0	0
	FILE: Number of large read operations	0	0	0
	FILE: Number of write operations	0	0	0
	HDFS: Number of bytes read	46,759	0	46,759
	HDFS: Number of bytes written	0	19,283	19,283
	HDFS: Number of read operations	4	0	4
	HDFS: Number of large read operations	0	0	0
	HDFS: Number of write operations	0	1	1
Job Counters	Launched map tasks	0	0	2
	Launched reduce tasks	0	0	1
	Data-local map tasks	0	0	1
	Rack-local map tasks	0	0	1
	Total time spent by all maps in occupied slots (ms)	0	0	16,914
	Total time spent by all reduces in occupied slots (ms)	0	0	6,181
	Total time spent by all maps waiting after reserving slots (ms)	0	0	0
	Total time spent by all reduces waiting after reserving slots (ms)	0	0	0

horse:50030/jobtasks.jsp?jobid=job\_201410070252\_0001&type=map&pagenum=1&state=completed

## 5. On Job File you can view metadata configuration for the job

## 6. Click on *map* link

## 7. Click on one from two tasks

Job [job\\_201410070252\\_0001](#)

All Task Attempts

Task Attempts	Machine	Status	Progress	Start Time	Finish Time	Errors	Task Logs	Counters	Actions
attempt_201410070252_0001_m_000000_0	<a href="#">/default-rack/elephant</a>	SUCCEEDED	100.00% <div><div></div></div>	7-Oct-2014 03:06:16	7-Oct-2014 03:06:20 (4sec)		<a href="#">Last 4KB</a> <a href="#">Last 8KB</a> <a href="#">All</a>	<a href="#">21</a>	

Input Split Locations

<a href="#">/default-rack/elephant</a>
<a href="#">/default-rack/tiger</a>
<a href="#">/default-rack/horse</a>

[Go back to the job](#)  
[Go back to JobTracker](#)  
[Hadoop, 2014.](#)

Which datanode read the block?