

Week 4

Step 1

```
ubuntu-xenial-large-521 — vagrant@pleasechangeme: /vagrant_data — ssh • vagrant ssh — 153x41
vagrant@pleasechangeme:/vagrant_data$ hadoop jar wc.jar WordCount /user/$USER/sampletext /user/$USER/output-wc
17/02/15 14:11:42 INFO client.RMPProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
17/02/15 14:11:43 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
17/02/15 14:11:43 INFO input.FileInputFormat: Total input paths to process : 7
17/02/15 14:11:43 INFO mapreduce.JobSubmitter: number of splits:7
17/02/15 14:11:43 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1487188071701_0001
17/02/15 14:11:44 INFO impl.YarnClientImpl: Submitted application application_1487188071701_0001
17/02/15 14:11:44 INFO mapreduce.Job: The url to track the job: http://pleasechangeme:8088/proxy/application_1487188071701_0001/
17/02/15 14:11:52 INFO mapreduce.Job: Running job: job_1487188071701_0001
17/02/15 14:11:52 INFO mapreduce.Job: Job job_1487188071701_0001 running in uber mode : false
17/02/15 14:11:52 INFO mapreduce.Job:  map 0% reduce 0%
17/02/15 14:12:16 INFO mapreduce.Job:  map 86% reduce 0%
17/02/15 14:12:23 INFO mapreduce.Job:  map 100% reduce 0%
17/02/15 14:12:25 INFO mapreduce.Job:  map 100% reduce 100%
17/02/15 14:12:25 INFO mapreduce.Job: Job job_1487188071701_0001 completed successfully
17/02/15 14:12:25 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=170702
    FILE: Number of bytes written=1116305
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=227308
    HDFS: Number of bytes written=71011
    HDFS: Number of read operations=24
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
  Job Counters
    Launched map tasks=7
    Launched reduce tasks=1
    Data-local map tasks=7
    Total time spent by all maps in occupied slots (ms)=135715
    Total time spent by all reduces in occupied slots (ms)=6443
    Total time spent by all map tasks (ms)=135715
    Total time spent by all reduce tasks (ms)=6443
    Total vcore-seconds taken by all map tasks=135715
    Total vcore-seconds taken by all reduce tasks=6443
    Total megabyte-seconds taken by all map tasks=138972160
    Total megabyte-seconds taken by all reduce tasks=6597632
  Map-Reduce Framework
```

```
ubuntu-xenial-large-521 — vagrant@pleasechangeme: /vagrant_data — ssh • vagrant ssh — 153x41
Virtual memory (bytes) snapshot=15033733120
Total committed heap usage (bytes)=1179123712
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=226475
File Output Format Counters
  Bytes Written=71011
vagrant@pleasechangeme:/vagrant_data$ hadoop fs -cat /user/$USER/output-wc/part-r-00000
Each 1
First 1
Hearing 1
If 1
Ordinary 1
Tell 1
This 1
Thou 1
Two 1
We 1
Well, 1
When 2
evil 1
the 1
What 2
$1 2
$1,000 1
$1,500 1
$1,600 3
$1.5 1
$10 3
$10,000 1
$14 1
$14,500 1
$140 2
$15 1
$17 1
```

Step 2

```
ubuntu-xenial-large-521 — vagrant@pleasechangeme: /vagrant_data — ssh • vagrant ssh — 153x41
[vagrant@pleasechangeme:/vagrant_data$ hadoop jar wc2.jar WordCount2 /user/$USER/sampletext /user/$USER/output-wc22
17/02/15 14:21:57 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
17/02/15 14:21:58 INFO input.FileInputFormat: Total input paths to process : 7
17/02/15 14:21:58 INFO mapreduce.JobSubmitter: number of splits:7
17/02/15 14:21:58 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1487188071701_0003
17/02/15 14:21:58 INFO impl.YarnClientImpl: Submitted application application_1487188071701_0003
17/02/15 14:21:58 INFO mapreduce.Job: The url to track the job: http://pleasechangeme:8088/proxy/application_1487188071701_0003/
17/02/15 14:21:58 INFO mapreduce.Job: Running job: job_1487188071701_0003
17/02/15 14:22:05 INFO mapreduce.Job: Job job_1487188071701_0003 running in uber mode : false
17/02/15 14:22:05 INFO mapreduce.Job: map 0% reduce 0%
17/02/15 14:22:26 INFO mapreduce.Job: map 57% reduce 0%
17/02/15 14:22:27 INFO mapreduce.Job: map 86% reduce 0%
17/02/15 14:22:31 INFO mapreduce.Job: map 100% reduce 0%
17/02/15 14:22:33 INFO mapreduce.Job: map 100% reduce 100%
17/02/15 14:22:34 INFO mapreduce.Job: Job job_1487188071701_0003 completed successfully
17/02/15 14:22:34 INFO mapreduce.Job: Counters: 50
  File System Counters
    FILE: Number of bytes read=170702
    FILE: Number of bytes written=1117561
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=227308
    HDFS: Number of bytes written=71011
    HDFS: Number of read operations=24
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
  Job Counters
    Launched map tasks=7
    Launched reduce tasks=1
    Data-local map tasks=7
    Total time spent by all maps in occupied slots (ms)=121869
    Total time spent by all reduces in occupied slots (ms)=4398
    Total time spent by all map tasks (ms)=121869
    Total time spent by all reduce tasks (ms)=4398
    Total vcore-seconds taken by all map tasks=121869
    Total vcore-seconds taken by all reduce tasks=4398
    Total megabyte-seconds taken by all map tasks=124793856
    Total megabyte-seconds taken by all reduce tasks=4503552
  Map-Reduce Framework
    Map input records=1096
  File Output Format Counters
    Bytes Written=71011
[vagrant@pleasechangeme:/vagrant_data$ hadoop fs -cat /user/$USER/output-wc22/part-r-00000
"Each 1
"First 1
"Hearing 1
"If 1
"Ordinary 1
"Tell 1
"This 1
"Thou 1
"Two 1
"We 1
"Well, 1
"When 2
"evil 1
"the 1
"what 2
$1 2
$1,000 1
$1,500 1
$1,600 3
$1.5 1
$10 3
$10,000 1
$14 1
$14,500 1
$140 2
$15 1
$17 1
$180,000 1
$2 2
$2,000 1
$2,000. 1
$2,500 1
$2.5 1
$2.6 1
$20 1
$200 1
$238 1
$246 1
```

Step 3

```
ubuntu-xenial-large-521 — vagrant@pleasechangeme: /vagrant_data — ssh * vagrant ssh — 153x41
CPU time spent (ms)=3870
Physical memory (bytes) snapshot=1635606528
Virtual memory (bytes) snapshot=15033733120
Total committed heap usage (bytes)=1179123712
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=226475
File Output Format Counters
  Bytes Written=4194
[vagrant@pleasechangeme:/vagrant_data$ hadoop fs -cat /user/$USER/output-w3/part-r-00000
--      174
10      8
20      5
21st    9
50      8
A       5
Act     5
After   5
America 92
America's 12
America. 23
American 87
Americans 61
Americans, 5
And      215
As       12
Because 14
But      79
Cold     6
Congress 57
Federal  9
For      18
Government 21
Great    5
His      6
```

Step 4

```
ubuntu-xenial-large-521 — vagrant@pleasechangeme: /vagrant_data — -bash — ...
File Output Format Counters
  Bytes Written=43517
[vagrant@pleasechangeme:/vagrant_data$ hadoop fs -cat /user/$USER/output9/part-r-]
00000
$1      2
$10     3
$1000   1
$10000  1
$14     1
$140    2
$14500  1
$15     2
$1500   1
$1600   3
$17     1
$180000 1
$2      2
$20     1
$200    1
$2000   2
$238    1
$246    1
$25     1
$2500   1
```

```
ubuntu-xenial-large-521 — vagrant@pleasechangeme: /vagrant_data — -bash —...  
a 777  
aband 2  
abb 1  
abess 1  
abidg 2  
ability 11  
able 27  
ablebodied 1  
ablest 1  
aboliti 1  
abraham 1  
abroad 6  
absence 1  
absolutely 4  
abuse 2  
abuses 1  
academy 3  
accelere 2  
accelerg 2  
accept 5  
acceptance 1  
accepted 2  
access 12  
accompanied 1
```