

## OUTPUT SCREENSHOTS

## Shared Memory

### 1GB with 1 Thread

The screenshot shows a Linux terminal window and a web browser window. The terminal window displays the output of a Hadoop command:

```
ubuntu@ip-172-31-13-243:~$ sh run.sh
File Creation!
1.80user 0.00system 0:14.76elapsed 12%CPU (0avgtext+0avgdata 61160maxresident)k
0inputs+195312outputs (0major+3679minor)pagefaults 0swaps
Program Run!
Split Files: 25secs.
Quick Sort: 35 secs.
Merge Sort: 90 secs.
Validate File.
Records: 10000000
Checksum: 4c499aae14ae2b
Duplicate keys: 0
SUCCESS: 10000000 records are in order
1.52user 0.00system 0:16.52elapsed 9%CPU (0avgtext+0avgdata 25500maxresident)k
195312inputs+0outputs (0major+3551minor)pagefaults 0swaps
```

The browser window shows the Hadoop Metrics page for the cluster, displaying status checks, alarm status, public DNS, and a detailed events section.

### 1GB with 2 Threads

### 1GB with 4 Threads

### 1GB – 8 Threads

### 10GB - 1 Thread

## 10 GB – 2 Threads

### 10GB – 4 Threads

### 10GB - 8 Threads

Cloud - ubuntu@ip-172-31-13-243: ~/64 - Xshell 5 (Free for Home/School)

File Edit View Tools Tab Window Help

ssh://ubuntu@ec2-52-38-10-112.us-west-2.compute.amazonaws.com:22

To add the current session, click on the left arrow button.

Cloud Cloud +

```
ubuntuip-172-31-13-243:~$ sudo java SharedMemorySort
Split Files: 20secs,
Quick Sort: 37 secs,
Merge Sort: 53 secs.
ubuntuip-172-31-13-243:~$ vi SharedMemorySort.java
ubuntuip-172-31-13-243:~$ javac SharedMemorySort.java
ubuntuip-172-31-13-243:~$ sudo java SharedMemorySort
Split Files: 289secs,
Quick Sort: 453 secs,
java.lang.IndexOutOfBoundsException: Index: 0, Size: 0
    at java.util.ArrayList.rangeCheck(ArrayList.java:635)
    at java.util.ArrayList.get(ArrayList.java:411)
    at SharedMemorySort.mergeSort(SharedMemorySort.java:110)
    at SharedMemorySort.main(SharedMemorySort.java:66)
Merge Sort: 1878 secs.
ubuntuip-172-31-13-243:~$ cd 64
ubuntuip-172-31-13-243:~/64$ ./valsort Ou-bash: cannot create temp file for here-document: No space left on dev
ubuntuip-172-31-13-243:~/64$ sudo ./valsort Output
Records: 100000000
Checksum: 2faf20d0000e8d8
Duplicate Keys: 0
SUCCESS - all records are in order
ubuntuip-172-31-13-243:~/64$ sudo time ./valsort Output
Records: 100000000
Checksum: 2faf20d0000e8d8
Duplicate Keys: 0
SUCCESS - all records are in order
15.56user 0.04system 2:41.06elapsed 9%CPU (0avgtext+0avgdata 25500maxresident)k
19531256inputs+outputs (1major+3351minor)pagefaults 0swaps
ubuntuip-172-31-13-243:~/64$ sudo time ./gensort -a 100000000 New1
17.46user 0.16system 2:41.59elapsed 10%CPU (0avgtext+0avgdata 61170maxresident)k
5521inputs+19531256outputs (1major+4180minor)pagefaults 0swaps
ubuntuip-172-31-13-243:~/64$
```

## Hadoop – 1GB on single node

The screenshot shows two terminal sessions in Xshell. Session 1 (C1) displays the execution of a Hadoop sort command:

```
ubuntu@ip-172-31-26-176:~/64$ time ./gensort -a 10000000 Newl
real   0m14.769s
user   0m1.802s
sys    0m0.026s
ubuntu@ip-172-31-26-176:~/64$ hadoop fs -put /Newl /Newl
16/03/26 21:50:15 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
put: '/Newl': No such file or directory
ubuntu@ip-172-31-26-176:~/64$ hadoop fs -put Newl /Newl
16/03/26 21:50:23 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
ubuntu@ip-172-31-26-176:~/64$ hadoop fs -ls /
16/03/26 21:50:50 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 3 items
-rw-r--r-- 1 ubuntu supergroup 100000000 2016-03-26 21:50 /Newl
drwxr-xr-x 1 ubuntu supergroup 0 2016-03-26 20:22 /tmp
drwxr-xr-x 1 ubuntu supergroup 0 2016-03-26 20:22 /user
ubuntu@ip-172-31-26-176:~/64$ hadoop jar HadoopSort.jar HadoopSort /Newl /output
16/03/26 21:51:22 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
16/03/26 21:51:22 INFO client.RMProxy: Connecting to ResourceManager at ec2-52-32-186-114.us-west-2.compute.amazonaws.com/172.31.26.176:8032
16/03/26 21:51:22 INFO client.RMProxy: Connecting to ResourceManager at ec2-52-32-186-114.us-west-2.compute.amazonaws.com/172.31.26.176:8032
16/03/26 21:51:23 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this
...
16/03/26 21:51:23 INFO mapred.FileInputFormat: Total input paths to process : 1
16/03/26 21:51:23 INFO mapreduce.JobSubmitter: number of splits:8
16/03/26 21:51:23 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1459023875973_0006
16/03/26 21:51:24 INFO impl.YarnClientImpl: Submitted application application_1459023875973_0006
16/03/26 21:51:24 INFO mapreduce.Job: The url to track the job: http://ec2-52-32-186-114.us-west-2.compute.amazonaws.com:8088/proxy/application_1459023875973_0006/
16/03/26 21:51:24 INFO mapreduce.Job: Running job: job_1459023875973_0006
```

Session 2 (C2) displays the HDFS and MapReduce metrics for the job:

```
HDFS: Number of read operations=27
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
Job Counters
  Killed map tasks=1
  Launched map tasks=9
  Launched reduce tasks=1
  Data-local map tasks=9
  Total time spent by all maps in occupied slots (ms)=407133
  Total time spent by all reduces in occupied slots (ms)=43781
  Total time spent by all map tasks (ms)=407133
  Total time spent by all reduce tasks (ms)=43781
  Total vcore-milliseconds taken by all map tasks=407133
  Total vcore-milliseconds taken by all reduce tasks=407133
  Total megabyte-milliseconds taken by all map tasks=416904192
  Total megabyte-milliseconds taken by all reduce tasks=44831744
Map-Reduce Framework
  Map input records=10000000
  Map output records=10000000
  Map output bytes=1000000000
  Map output materialized bytes=1020000048
  Input split bytes=944
  Combine input records=0
  Combine output records=0
  Reduce input groups=10000000
  Reduce input records=10000000
  Reduce shuffle bytes=1020000048
  Reduce input records=10000000
  Reduce output records=10000000
  Spilled records=9995241
  Shuffled Maps=8
  Failed Shuffles=0
  Merged Map outputs=8
  GC time elapsed (ms)=7000
  CPU time spent (ms)=103840
  Physical memory (bytes) snapshot=2234245120
  Virtual memory (bytes) snapshot=17166135296
  Total committed heap usage (bytes)=1741684736
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=1000028672
File Output Format Counters
  Bytes Written=1000000000
Time: 121 secs
ubuntu@ip-172-31-26-176:~/64$
```

```

Shuffled Maps =8
Failed Shuffles=0
Merged Map outputs=8
GC time elapsed (ms)=7000
CPU time spent (ms)=103840
Physical memory (bytes) snapshot=2234245120
Virtual memory (bytes) snapshot=17166135296
Total committed heap usage (bytes)=1741684736
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=1000002872
File Output Format Counters
  Bytes Written=10000000000
Time: 1:21 secs
ubuntu@ip-172-31-26-176:~/>hadoop fs -ls /a/
16/03/26 21:54:03 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
ls: '/a': No such file or directory
ubuntu@ip-172-31-26-176:~/>hadoop fs -ls /a
16/03/26 21:54:07 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
ls: '/a': No such file or directory
ubuntu@ip-172-31-26-176:~/>hadoop fs -ls /output/
16/03/26 21:54:16 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r-- 1 ubuntu supergroup 0 2016-03-26 21:53 /output/_SUCCESS
-rw-r--r-- 1 ubuntu supergroup 1000000000 2016-03-26 21:53 /output/part-00000
ubuntu@ip-172-31-26-176:~/>hadoop fs -get /output/part-00000 ./output
16/03/26 21:55:12 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
ubuntu@ip-172-31-26-176:~/>hadoop fs -ls /h
total 1.9G
-rwxrwxr-x 1 ubuntu 500 138K Mar 17 2013 gnsort
-rw-r--r-- 1 ubuntu ubuntu 4.4K Mar 26 21:47 HadoopSort.jar
-rwxr--r-- 1 ubuntu ubuntu 954M Mar 26 21:49 Newl
-rw-r--r-- 1 ubuntu ubuntu 954M Mar 26 21:55 output
-rwxrwxr-x 1 ubuntu 500 132K Mar 17 2013 valsort
ubuntu@ip-172-31-26-176:~/>time ./valsort output
Records: 10000000
Checksum: 4c499aae14ae2b
Duplicate keys: 0
SUCCESS - all records are in order

real    0m14.094s
user    0m1.492s
sys     0m0.027s
ubuntu@ip-172-31-26-176:~/>64$ █

```

## Hadoop – 10GB on single node

```

16/03/27 05:12:02 INFO mapreduce.Job: map 17% reduce 0%
16/03/27 05:12:05 INFO mapreduce.Job: map 18% reduce 0%
16/03/27 05:12:10 INFO mapreduce.Job: map 19% reduce 0%
16/03/27 05:12:12 INFO mapreduce.Job: map 20% reduce 0%
16/03/27 05:12:24 INFO mapreduce.Job: map 21% reduce 0%
16/03/27 05:12:34 INFO mapreduce.Job: map 22% reduce 0%
16/03/27 05:12:37 INFO mapreduce.Job: map 23% reduce 0%
16/03/27 05:12:42 INFO mapreduce.Job: map 24% reduce 0%
16/03/27 05:13:09 INFO mapreduce.Job: map 25% reduce 0%
16/03/27 05:13:13 INFO mapreduce.Job: map 26% reduce 0%
16/03/27 05:13:14 INFO mapreduce.Job: map 26% reduce 1%
16/03/27 05:13:15 INFO mapreduce.Job: map 27% reduce 1%
16/03/27 05:13:21 INFO mapreduce.Job: map 27% reduce 2%
16/03/27 05:13:25 INFO mapreduce.Job: map 28% reduce 2%
16/03/27 05:13:29 INFO mapreduce.Job: map 29% reduce 2%
16/03/27 05:13:34 INFO mapreduce.Job: map 30% reduce 2%
16/03/27 05:13:44 INFO mapreduce.Job: map 31% reduce 2%
16/03/27 05:13:45 INFO mapreduce.Job: map 31% reduce 3%
16/03/27 05:13:52 INFO mapreduce.Job: map 31% reduce 4%
16/03/27 05:13:58 INFO mapreduce.Job: map 31% reduce 5%
16/03/27 05:14:01 INFO mapreduce.Job: map 31% reduce 6%
16/03/27 05:14:07 INFO mapreduce.Job: map 31% reduce 7%
16/03/27 05:14:14 INFO mapreduce.Job: map 32% reduce 7%
16/03/27 05:14:16 INFO mapreduce.Job: map 32% reduce 8%
16/03/27 05:14:20 INFO mapreduce.Job: map 33% reduce 8%
16/03/27 05:14:21 INFO mapreduce.Job: map 34% reduce 8%
16/03/27 05:14:31 INFO mapreduce.Job: map 34% reduce 9%
16/03/27 05:14:32 INFO mapreduce.Job: map 34% reduce 9%
16/03/27 05:14:39 INFO mapreduce.Job: map 35% reduce 9%
16/03/27 05:14:48 INFO mapreduce.Job: map 36% reduce 9%
16/03/27 05:14:48 INFO mapreduce.Job: map 37% reduce 10%
16/03/27 05:15:18 INFO mapreduce.Job: map 37% reduce 11%
16/03/27 05:15:24 INFO mapreduce.Job: map 37% reduce 12%
16/03/27 05:15:27 INFO mapreduce.Job: map 37% reduce 12%
16/03/27 05:15:40 INFO mapreduce.Job: map 38% reduce 12%
16/03/27 05:15:49 INFO mapreduce.Job: map 40% reduce 12%
16/03/27 05:16:02 INFO mapreduce.Job: map 41% reduce 12%
16/03/27 05:16:06 INFO mapreduce.Job: map 42% reduce 12%
16/03/27 05:16:24 INFO mapreduce.Job: map 43% reduce 12%
16/03/27 05:16:28 INFO mapreduce.Job: map 44% reduce 12%
16/03/27 05:16:34 INFO mapreduce.Job: map 44% reduce 13%
16/03/27 05:16:37 INFO mapreduce.Job: map 44% reduce 14%
16/03/27 05:16:40 INFO mapreduce.Job: map 44% reduce 15%
16/03/27 05:16:59 INFO mapreduce.Job: map 45% reduce 15%
16/03/27 05:17:02 INFO mapreduce.Job: map 47% reduce 15%
16/03/27 05:17:12 INFO mapreduce.Job: map 48% reduce 15%
16/03/27 05:17:21 INFO mapreduce.Job: map 49% reduce 15%
16/03/27 05:17:24 INFO mapreduce.Job: map 50% reduce 15%
16/03/27 05:17:30 INFO mapreduce.Job: map 51% reduce 15%
16/03/27 05:17:35 INFO mapreduce.Job: map 51% reduce 16%

```

```
16/03/27 05:20:48 INFO mapreduce.Job: map 66% reduce 21%
16/03/27 05:20:51 INFO mapreduce.Job: map 67% reduce 21%
16/03/27 05:21:04 INFO mapreduce.Job: map 68% reduce 21%
16/03/27 05:21:13 INFO mapreduce.Job: map 69% reduce 21%
16/03/27 05:21:15 INFO mapreduce.Job: map 70% reduce 21%
16/03/27 05:21:17 INFO mapreduce.Job: map 71% reduce 21%
16/03/27 05:21:25 INFO mapreduce.Job: map 71% reduce 22%
16/03/27 05:21:28 INFO mapreduce.Job: map 71% reduce 23%
16/03/27 05:21:31 INFO mapreduce.Job: map 71% reduce 24%
16/03/27 05:21:50 INFO mapreduce.Job: map 72% reduce 24%
16/03/27 05:21:53 INFO mapreduce.Job: map 73% reduce 24%
16/03/27 05:21:55 INFO mapreduce.Job: map 74% reduce 24%
16/03/27 05:22:06 INFO mapreduce.Job: map 75% reduce 24%
16/03/27 05:22:14 INFO mapreduce.Job: map 76% reduce 24%
16/03/27 05:22:18 INFO mapreduce.Job: map 77% reduce 24%
16/03/27 05:22:53 INFO mapreduce.Job: map 77% reduce 25%
16/03/27 05:22:56 INFO mapreduce.Job: map 77% reduce 26%
16/03/27 05:23:13 INFO mapreduce.Job: map 79% reduce 26%
16/03/27 05:23:15 INFO mapreduce.Job: map 80% reduce 26%
16/03/27 05:23:25 INFO mapreduce.Job: map 81% reduce 26%
16/03/27 05:23:29 INFO mapreduce.Job: map 82% reduce 26%
16/03/27 05:23:36 INFO mapreduce.Job: map 83% reduce 26%
16/03/27 05:23:41 INFO mapreduce.Job: map 84% reduce 26%
16/03/27 05:23:51 INFO mapreduce.Job: map 84% reduce 27%
16/03/27 05:23:54 INFO mapreduce.Job: map 84% reduce 28%
16/03/27 05:24:15 INFO mapreduce.Job: map 86% reduce 28%
16/03/27 05:24:18 INFO mapreduce.Job: map 87% reduce 28%
16/03/27 05:24:27 INFO mapreduce.Job: map 88% reduce 28%
16/03/27 05:24:37 INFO mapreduce.Job: map 89% reduce 28%
16/03/27 05:24:42 INFO mapreduce.Job: map 90% reduce 28%
16/03/27 05:24:46 INFO mapreduce.Job: map 91% reduce 28%
16/03/27 05:25:19 INFO mapreduce.Job: map 91% reduce 29%
16/03/27 05:25:22 INFO mapreduce.Job: map 91% reduce 30%
16/03/27 05:25:38 INFO mapreduce.Job: map 92% reduce 30%
16/03/27 05:25:41 INFO mapreduce.Job: map 93% reduce 30%
16/03/27 05:25:44 INFO mapreduce.Job: map 94% reduce 30%
16/03/27 05:25:53 INFO mapreduce.Job: map 95% reduce 30%
16/03/27 05:26:01 INFO mapreduce.Job: map 96% reduce 30%
16/03/27 05:26:06 INFO mapreduce.Job: map 97% reduce 30%
16/03/27 05:26:11 INFO mapreduce.Job: map 97% reduce 31%
16/03/27 05:26:11 INFO mapreduce.Job: map 97% reduce 32%
16/03/27 05:26:12 INFO mapreduce.Job: map 99% reduce 32%
16/03/27 05:27:03 INFO mapreduce.Job: map 99% reduce 33%
16/03/27 05:27:06 INFO mapreduce.Job: map 100% reduce 33%
16/03/27 05:27:12 INFO mapreduce.Job: map 100% reduce 46%
16/03/27 05:27:16 INFO mapreduce.Job: map 100% reduce 55%
16/03/27 05:27:19 INFO mapreduce.Job: map 100% reduce 67%
16/03/27 05:27:25 INFO mapreduce.Job: map 100% reduce 68%
16/03/27 05:27:28 INFO mapreduce.Job: map 100% reduce 69%
```

```
HDFS: Number of read operations=228
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
Job Counters
  Killed map tasks=2
  Launched map tasks=77
  Launched reduce tasks=1
  Data-local map tasks=77
  Total time spent by all maps in occupied slots (ms)=5240272
  Total time spent by all reduces in occupied slots (ms)=1039311
  Total time spent by all map tasks (ms)=5240272
  Total time spent by all reduce tasks (ms)=1039311
  Total vcore-milliseconds taken by all map tasks=5240272
  Total vcore-milliseconds taken by all reduce tasks=1039311
  Total megabyte-milliseconds taken by all map tasks=5366038520
  Total megabyte-milliseconds taken by all reduce tasks=106425446
Map -Reduce Framework
  Map input records=1000000000
  Map output records=1000000000
  Map output bytes=100000000000
  Map output materialized bytes=10200000450
  Input split bytes=8775
  Combine input records=0
  Combine output records=0
  Reduce input groups=1000000000
  Reduce shuffle bytes=10200000450
  Reduce input records=1000000000
  Reduce output records=1000000000
  Spilled Records=396636763
  Shuffled Maps =75
  Failed Shuffles=0
  Merged Map outputs=75
  GC time elapsed (ms)=69520
  CPU time spent (ms)=157050
  Physical memory (bytes) snapshot=19137937408
  Virtual memory (bytes) snapshot=144727474176
  Total committed heap usage (bytes)=15635316736
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=1000003030104
File Output Format Counters
  Bytes Written=10000000000
Time: 1256 secs
ubuntu@ip-172-31-28-27:~/64$
```

```

Map output records=1000000000
Map output bytes=10000000000
Map output materialized bytes=10200000450
Input split bytes=8775
Combine input records=0
Combine output records=0
Reduce input groups=100000000
Reduce shuffle bytes=10200000450
Reduce input records=100000000
Reduce output records=100000000
Spilled Records=396056763
Shuffled Maps =75
Failed Shuffles=0
Merged Map outputs=75
GC time elapsed (ms)=89520
CPU time elapsed (ms)=157050
Physical memory (bytes) snapshot=19137937408
Virtual memory (bytes) snapshot=14477474176
Total committed heap usage (bytes)=15635316736

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=100000303104
File Output Format Counters
  Bytes Written=100000000000

Time: 1256 secs
ubuntu@ip-172-31-28-27:~/64$ hadoop fs -ls /output
16/03/27 05:30:33 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--  1 ubuntu  supergroup          0 2016-03-27 05:30 /output/_SUCCESS
-rw-r--r--  1 ubuntu  supergroup 100000000000 2016-03-27 05:30 /output/part-00000
ubuntu@ip-172-31-28-27:~/64$ hadoop fs -get /output/part-00000 ./output
16/03/27 05:31:02 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
ubuntu@ip-172-31-28-27:~/64$ ls
gensort HadoopSort.jar output valsort
ubuntu@ip-172-31-28-27:~/64$ time ./valsort output
Records: 100000000
Checksum: 2faf20d000e08d8
Duplicate keys: 0
SUCCESS - all records are in order

real    2m43.765s
user    0m15.19s
sys     0m0.078s
ubuntu@ip-172-31-28-27:~/64$ 

```

## Hadoop 100GB on 16 Nodes

```

ssh://ubuntu@ec2-52-38-22-193.us-west-2.compute.amazonaws.com:22
To add the current session, click on the left arrow button.
Send keyboard input to current session only. ON
-x 3 Server0 x 4 Server1 x 5 Server10 x 6 Server11 x 7 Server12 x 8 Server13 x 9 Server14 x 10 Server15 x 11 Server2 x 12 Server3 x 13 Server4 x

-rw-r--r--  1 ubuntu  supergroup 100000000000 2016-03-28 04:39 /Input
ubuntu@ip-172-31-18-255:~/64$ rm /mnt/raid/Input
ubuntu@ip-172-31-18-255:~/64$ ls
gensort valsort
ubuntu@ip-172-31-18-255:~/64$ df -h
Filesystem      Size  Used Avail Use% Mounted on
udev            7.4G   7.4G    0K  1% /dev
tmpfs           1.5G  372K  1.5G  1% /run
/dev/xvda1       7.8G  7.4G  6.4M 100% /
none            4.0K    0  4.0K  0% /sys/fs/cgroup
none            5.0M    0  5.0M  0% /run/lock
none            7.4G    0  7.4G  0% /run/shm
none           100M    0  100M  0% /run/user
/dev/md127      453G  95G  336G  22% /mnt/raid
ubuntu@ip-172-31-18-255:~/64$ hadoop jar /home/ubuntu/HadoopSort.jar HadoopSort /Input /Output
16/03/28 04:42:46 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
16/03/28 04:42:47 INFO client.RMProxy: Connecting to ResourceManager at ec2-52-38-22-193.us-west-2.compute.amazonaws.com/172.31.18.255:8032
16/03/28 04:42:47 INFO client.RMProxy: Connecting to ResourceManager at ec2-52-38-22-193.us-west-2.compute.amazonaws.com/172.31.18.255:8032
16/03/28 04:42:47 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
16/03/28 04:42:48 INFO mapred.FileInputFormat: Total input paths to process : 1
16/03/28 04:42:48 INFO net.NetworkTopology: Adding a new node: /default-rack/172.31.18.255:50010
16/03/28 04:42:49 INFO mapreduce.JobSubmitter: number of splits:745
16/03/28 04:42:49 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1459137888961_0001
16/03/28 04:42:49 INFO impl.YarnClientImpl: Submitted application application_1459137888961_0001
16/03/28 04:42:49 INFO mapreduce.Job: The url to track the job: http://ec2-52-38-22-193.us-west-2.compute.amazonaws.com:8088/proxy/application_1459137888961_0001/
16/03/28 04:42:49 INFO mapreduce.Job: Running job: job_1459137888961_0001
16/03/28 04:42:57 INFO mapreduce.Job: Job job_1459137888961_0001 running in uber mode : false
16/03/28 04:42:57 INFO mapreduce.Job: map 0% reduce 0%

```

Send text to all sessions within current window

ssh://ubuntu@ec2-52-38-22-193.us-west-2.compute.amazonaws.com:22

SSH2 xterm T 191x30 30,1 18 sessions CAP NUM

Windows Taskbar icons: File Explorer, Google Chrome, FileZilla, Google Sheets, Lock Screen, Task View, Microsoft Edge.

System tray icons: Battery, Signal, ENG 11:43 PM, Date 3/27/2016.

Send keyboard input to current session only. ON

```

16/03/28 04:49:55 INFO mapreduce.Job: map 48% reduce 7%
16/03/28 04:50:03 INFO mapreduce.Job: map 49% reduce 7%
16/03/28 04:50:11 INFO mapreduce.Job: map 50% reduce 7%
16/03/28 04:50:20 INFO mapreduce.Job: map 51% reduce 7%
16/03/28 04:50:27 INFO mapreduce.Job: map 52% reduce 7%
16/03/28 04:50:29 INFO mapreduce.Job: map 53% reduce 8%
16/03/28 04:50:34 INFO mapreduce.Job: map 53% reduce 8%
16/03/28 04:50:44 INFO mapreduce.Job: map 54% reduce 8%
16/03/28 04:50:53 INFO mapreduce.Job: map 55% reduce 8%
16/03/28 04:51:02 INFO mapreduce.Job: map 56% reduce 8%
16/03/28 04:51:11 INFO mapreduce.Job: map 57% reduce 8%
16/03/28 04:51:19 INFO mapreduce.Job: map 58% reduce 8%
16/03/28 04:51:28 INFO mapreduce.Job: map 59% reduce 8%
16/03/28 04:51:33 INFO mapreduce.Job: map 59% reduce 9%
16/03/28 04:51:38 INFO mapreduce.Job: map 60% reduce 9%
16/03/28 04:51:44 INFO mapreduce.Job: map 61% reduce 9%
16/03/28 04:51:53 INFO mapreduce.Job: map 62% reduce 9%
16/03/28 04:52:00 INFO mapreduce.Job: map 63% reduce 9%
16/03/28 04:52:08 INFO mapreduce.Job: map 64% reduce 9%
16/03/28 04:52:15 INFO mapreduce.Job: map 65% reduce 9%
16/03/28 04:52:23 INFO mapreduce.Job: map 66% reduce 9%
16/03/28 04:52:31 INFO mapreduce.Job: map 67% reduce 9%
16/03/28 04:52:37 INFO mapreduce.Job: map 67% reduce 10%
16/03/28 04:52:40 INFO mapreduce.Job: map 68% reduce 10%
16/03/28 04:52:50 INFO mapreduce.Job: map 69% reduce 10%
16/03/28 04:52:58 INFO mapreduce.Job: map 70% reduce 10%
16/03/28 04:53:06 INFO mapreduce.Job: map 71% reduce 10%
16/03/28 04:53:13 INFO mapreduce.Job: map 72% reduce 10%
16/03/28 04:53:21 INFO mapreduce.Job: map 73% reduce 10%
16/03/28 04:53:27 INFO mapreduce.Job: map 73% reduce 11%
16/03/28 04:53:39 INFO mapreduce.Job: map 74% reduce 11%
16/03/28 04:53:37 INFO mapreduce.Job: map 75% reduce 11%
16/03/28 04:53:46 INFO mapreduce.Job: map 76% reduce 11%
16/03/28 04:53:53 INFO mapreduce.Job: map 77% reduce 11%
16/03/28 04:54:01 INFO mapreduce.Job: map 78% reduce 11%
16/03/28 04:54:11 INFO mapreduce.Job: map 79% reduce 11%
16/03/28 04:54:19 INFO mapreduce.Job: map 80% reduce 11%
16/03/28 04:54:25 INFO mapreduce.Job: map 80% reduce 12%
16/03/28 04:54:27 INFO mapreduce.Job: map 81% reduce 12%
16/03/28 04:54:37 INFO mapreduce.Job: map 82% reduce 12%
16/03/28 04:54:47 INFO mapreduce.Job: map 83% reduce 12%
16/03/28 04:54:55 INFO mapreduce.Job: map 84% reduce 12%
16/03/28 04:55:04 INFO mapreduce.Job: map 85% reduce 12%
16/03/28 04:55:13 INFO mapreduce.Job: map 86% reduce 12%
16/03/28 04:55:18 INFO mapreduce.Job: map 86% reduce 13%
16/03/28 04:55:33 INFO mapreduce.Job: map 87% reduce 13%

```

<http://ec2-52-38-22-193.us-west-2.compute.amazonaws.com:50070/dfshealth.html#tab-datanode>

Node	Last contact	Admin State	Capacity	Used	Non DFS Used	Remaining	Blocks	Block pool used	Failed Volumes	Version
ip-172-31-1-177.us-west-2.compute.internal:50010 (172.31.1.177:50010)	0	In Service	49.72 GB	48 KB	8.7 GB	41.01 GB	1	48 KB (0%)	0	2.7.2
ip-172-31-6-0.us-west-2.compute.internal:50010 (172.31.6.0:50010)	2	In Service	49.72 GB	132 KB	13.36 GB	36.36 GB	1	132 KB (0%)	0	2.7.2
ip-172-31-1-0.us-west-2.compute.internal:50010 (172.31.1.0:50010)	0	In Service	49.72 GB	4.13 MB	10.24 GB	39.48 GB	2	4.13 MB (0.01%)	0	2.7.2
ip-172-31-11-211.us-west-2.compute.internal:50010 (172.31.11.211:50010)	0	In Service	49.72 GB	48 KB	8.06 GB	41.65 GB	1	48 KB (0%)	0	2.7.2
ip-172-31-4-5.us-west-2.compute.internal:50010 (172.31.4.5:50010)	0	In Service	49.72 GB	48 KB	9.6 GB	40.12 GB	1	48 KB (0%)	0	2.7.2
ip-172-31-12-18.us-west-2.compute.internal:50010 (172.31.12.18:50010)	0	In Service	49.72 GB	48 KB	8.83 GB	40.89 GB	1	48 KB (0%)	0	2.7.2
ip-172-31-11-33.us-west-2.compute.internal:50010 (172.31.11.33:50010)	0	In Service	49.72 GB	144 KB	8.58 GB	41.14 GB	2	144 KB (0%)	0	2.7.2
ip-172-31-0-131.us-west-2.compute.internal:50010 (172.31.0.131:50010)	0	In Service	49.72 GB	144 KB	9.98 GB	39.74 GB	2	144 KB (0%)	0	2.7.2
ip-172-31-1-5.us-west-2.compute.internal:50010 (172.31.1.5:50010)	0	In Service	49.72 GB	132 KB	9.34 GB	40.38 GB	1	132 KB (0%)	0	2.7.2
ip-172-31-15-231.us-west-2.compute.internal:50010 (172.31.15.231:50010)	0	In Service	49.72 GB	132 KB	10.11 GB	39.61 GB	1	132 KB (0%)	0	2.7.2
ip-172-31-4-11.us-west-2.compute.internal:50010 (172.31.4.11:50010)	0	In Service	49.72 GB	48 KB	8.7 GB	41.01 GB	1	48 KB (0%)	0	2.7.2
ip-172-31-14-1.us-west-2.compute.internal:50010	0	In Service	49.72 GB	144 KB	9.34 GB	40.38 GB	2	144 KB (0%)	0	2.7.2

MapReduce Job job\_1459137888961\_0001

Job Overview

Job Name:	HadoopSort
State:	RUNNING
Uberized:	false
Started:	Mon Mar 28 04:42:56 UTC 2016
Elapsed:	17mins, 31sec

ApplicationMaster

Attempt Number	Start Time	Node	Logs
1	Mon Mar 28 04:42:52 UTC 2016	ip-172-31-1-0.us-west-2.compute.internal:8042	logs

Task Type	Progress	Total	Pending	Running	Complete
Map	<div style="width: 745px;"></div>	745	0	0	745
Reduce	<div style="width: 1px;"></div>	1	0	1	0

Attempt Type	New	Running	Failed	Killed	Successful
Maps	0	0	0	44	745
Reduces	0	1	0	1	0

Server0 - ubuntu@ip-172-31-18-255: ~/64 - Xshell 5 (Free for Home/School)

File Edit View Tools Tab Window Help

ssh://ubuntu@ec2-52-38-22-193.us-west-2.compute.amazonaws.com:22

To add the current session, click on the left arrow button.

Send keyboard input to current session only.  ON

```
16/03/28 05:18:04 INFO mapreduce.Job: map 98% reduce 1%
16/03/28 05:18:10 INFO mapreduce.Job: map 99% reduce 1%
16/03/28 05:18:19 INFO mapreduce.Job: map 100% reduce 1%
16/03/28 05:18:49 INFO mapreduce.Job: map 100% reduce 2%
16/03/28 05:19:34 INFO mapreduce.Job: map 100% reduce 3%
16/03/28 05:20:29 INFO mapreduce.Job: map 100% reduce 4%
16/03/28 05:21:14 INFO mapreduce.Job: map 100% reduce 5%
16/03/28 05:21:52 INFO mapreduce.Job: map 100% reduce 6%
16/03/28 05:22:46 INFO mapreduce.Job: map 100% reduce 7%
16/03/28 05:23:25 INFO mapreduce.Job: map 100% reduce 8%
16/03/28 05:24:07 INFO mapreduce.Job: map 100% reduce 9%
16/03/28 05:24:56 INFO mapreduce.Job: map 100% reduce 10%
16/03/28 05:25:38 INFO mapreduce.Job: map 100% reduce 11%
16/03/28 05:26:12 INFO mapreduce.Job: map 100% reduce 12%
16/03/28 05:26:54 INFO mapreduce.Job: map 92% reduce 0%
16/03/28 05:27:16 INFO mapreduce.Job: map 93% reduce 0%
16/03/28 05:27:23 INFO mapreduce.Job: map 93% reduce 1%
16/03/28 05:27:27 INFO mapreduce.Job: map 94% reduce 1%
16/03/28 05:27:38 INFO mapreduce.Job: map 95% reduce 1%
16/03/28 05:27:49 INFO mapreduce.Job: map 96% reduce 1%
16/03/28 05:27:53 INFO mapreduce.Job: map 96% reduce 2%
16/03/28 05:27:59 INFO mapreduce.Job: map 97% reduce 2%
16/03/28 05:28:08 INFO mapreduce.Job: map 98% reduce 2%
16/03/28 05:28:15 INFO mapreduce.Job: map 99% reduce 2%
16/03/28 05:28:23 INFO mapreduce.Job: map 100% reduce 2%
16/03/28 05:28:30 INFO mapreduce.Job: map 100% reduce 3%
16/03/28 05:29:09 INFO mapreduce.Job: map 100% reduce 4%
16/03/28 05:29:42 INFO mapreduce.Job: map 100% reduce 5%
16/03/28 05:30:18 INFO mapreduce.Job: map 100% reduce 6%
```

Send text to all sessions within current window

ssh://ubuntu@ec2-52-38-22-193.us-west-2.compute.amazonaws.com:22

SSH2 xterm I: 191x30 L: 30,1 17 sessions CAP NUM

```

HDFS: Number of read operations=2238
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
Job Counters
  Killed map tasks=1
  Launched map tasks=746
  Launched reduce tasks=1
  Data-local map tasks=746
  Total time spent by all maps in occupied slots (ms)=9693018
  Total time spent by all reduces in occupied slots (ms)=494275
  Total time spent by all map tasks (ms)=9693018
  Total vcore-milliseconds taken by all map tasks=9693018
  Total vcore-milliseconds taken by all reduce tasks=494275
  Total megabyte-milliseconds taken by all map tasks=9025650432
  Total megabyte-milliseconds taken by all reduce tasks=9720089600
Map-Reduce Framework
  Map input records=1000000000
  Map output records=1000000000
  Map output bytes=100000000000
  Map output materialized bytes=102000000470
  Input split bytes=87910
  Combined input records=0
  Combined output records=0
  Reduce input groups=1000000000
  Reduce shuffle bytes=102000000470
  Reduce input records=1000000000
  Reduce output records=1000000000
  Spilled Records =6174638893
  Shuffled Maps =745
  Failed Shuffles=0
  Merged Map outputs=745
  GC time elapsed (ms)=448099
  CPU time spent (ms)=11916950
  Physical memory (bytes) snapshot=201843904512
  Virtual memory (bytes) snapshot=1439412592640
  Total committed heap usage (bytes)=154459439104
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=1000000047424
File Output Format Counters
  Bytes Written=1000000000000
Time: 4922 secs

```

## Spark 1GB on Single Node

```

Send keyboard input to current session only. [ ] ON
scala> val input = sc.textFile("/Input")
16/03/28 06:52:01 INFO storage.MemoryStore: Block broadcast_0 stored as values in memory (estimated size 26.5 KB, free 26.5 KB)
16/03/28 06:52:02 INFO storage.MemoryStore: Block broadcast_0_piece0 stored as bytes in memory (estimated size 5.6 KB, free 32.1 KB)
16/03/28 06:52:02 INFO storage.BlockManagerInfo: Added broadcast_0_piece0 in memory on 172.31.15.32:47604 (size: 5.6 KB, free: 511.5 MB)
16/03/28 06:52:02 INFO spark.SparkContext: Created broadcast 0 from textfile at <console>:27
input: org.apache.spark.rdd.RDD[String] = /input MapPartitionsRDD[1] at textfile at <console>:27
scala> val data = input.flatMap(line => line.split("\n")).map(line => (line, "")).sortByKey().coalesce(1, false).map(case (k, v) => k + v)
16/03/28 06:53:02 WARN snappy.LoadSnappy: Snappy native library not loaded
16/03/28 06:53:03 INFO mapred.FileInputFormat: Total input paths to process : 1
16/03/28 06:53:03 INFO spark.SparkContext: Starting job: sortByKey at <console>:29
16/03/28 06:53:03 INFO scheduler.DAGScheduler: Got job 0 (sortByKey at <console>:29) with 8 output partitions
16/03/28 06:53:03 INFO scheduler.DAGScheduler: Final stage: ResultStage 0 (sortByKey at <console>:29)
16/03/28 06:53:03 INFO scheduler.DAGScheduler: Parents of final stage: List()
16/03/28 06:53:03 INFO scheduler.DAGScheduler: Missing parents: List()
16/03/28 06:53:03 INFO scheduler.DAGScheduler: Submitting ResultStage 0 (MapPartitionsRDD[5] at sortByKey at <console>:29), which has no missing parents
16/03/28 06:53:03 INFO storage.MemoryStore: Block broadcast_1 stored as values in memory (estimated size 3.9 KB, free 36.0 KB)
16/03/28 06:53:03 INFO storage.MemoryStore: Block broadcast_1_piece0 stored as bytes in memory (estimated size 2.1 KB, free 38.1 KB)
16/03/28 06:53:03 INFO storage.BlockManagerInfo: Added broadcast_1_piece0 in memory on 172.31.15.32:47604 (size: 2.1 KB, free: 511.5 MB)
16/03/28 06:53:03 INFO spark.SparkContext: Created broadcast 1 from broadcast at DAGScheduler.scala:1006
16/03/28 06:53:03 INFO scheduler.DAGScheduler: Submitting 8 missing tasks for stage: ResultStage 0 (MapPartitionsRDD[5] at sortByKey at <console>:29)
16/03/28 06:53:03 INFO scheduler.TaskSetManager: Adding task set 0 with 8 tasks
16/03/28 06:53:03 INFO scheduler.TaskSetManager: Starting Task 0.0 in stage 0 (TID 0, ip: 172.31.6-35.us-west-2.compute.internal, partition 0, NODE_LOCAL, 2170 bytes)
16/03/28 06:53:03 INFO scheduler.TaskSetManager: Starting Task 1.0 in stage 0 (TID 1, ip: 172.31.6-35.us-west-2.compute.internal, partition 1, NODE_LOCAL, 2170 bytes)
16/03/28 06:53:04 INFO storage.BlockManagerInfo: Added broadcast_1_piece0 in memory on ip-172-31-6-35.us-west-2.compute.internal:38679 (size: 2.1 KB, free: 1539.0 MB)
16/03/28 06:53:04 INFO storage.BlockManagerInfo: Added broadcast_0_piece0 in memory on ip-172-31-6-35.us-west-2.compute.internal:38679 (size: 5.6 KB, free: 1539.0 MB)
16/03/28 06:53:07 INFO scheduler.TaskSetManager: Starting Task 2.0 in stage 0 (TID 2, ip: 172.31.6-35.us-west-2.compute.internal, partition 2, NODE_LOCAL, 2170 bytes)
16/03/28 06:53:07 INFO scheduler.TaskSetManager: Finished Task 0.0 in stage 0 (TID 0, in 3788 ms on ip-172-31-6-35.us-west-2.compute.internal (1/8))
16/03/28 06:53:07 INFO scheduler.TaskSetManager: Starting Task 3.0 in stage 0 (TID 3, ip: 172.31.6-35.us-west-2.compute.internal, partition 3, NODE_LOCAL, 2170 bytes)
16/03/28 06:53:07 INFO scheduler.TaskSetManager: Finished Task 1.0 in stage 0 (TID 1, in 3867 ms on ip-172-31-6-35.us-west-2.compute.internal (2/8))
16/03/28 06:53:09 INFO scheduler.TaskSetManager: Starting Task 4.0 in stage 0 (TID 4, ip: 172.31.6-35.us-west-2.compute.internal, partition 4, NODE_LOCAL, 2170 bytes)
16/03/28 06:53:09 INFO scheduler.TaskSetManager: Finished Task 2.0 in stage 0 (TID 2, in 1764 ms on ip-172-31-6-35.us-west-2.compute.internal (3/8))
16/03/28 06:53:09 INFO scheduler.TaskSetManager: Starting Task 5.0 in stage 0.0 (TID 5, ip: 172.31.6-35.us-west-2.compute.internal, partition 5, NODE_LOCAL, 2170 bytes)
16/03/28 06:53:09 INFO scheduler.TaskSetManager: Finished Task 3.0 in stage 0.0 (TID 3, in 1729 ms on ip-172-31-6-35.us-west-2.compute.internal (4/8))
16/03/28 06:53:10 INFO scheduler.TaskSetManager: Starting Task 6.0 in stage 0.0 (TID 6, ip: 172.31.6-35.us-west-2.compute.internal, partition 6, NODE_LOCAL, 2170 bytes)
16/03/28 06:53:10 INFO scheduler.TaskSetManager: Finished Task 4.0 in stage 0.0 (TID 4, in 1796 ms on ip-172-31-6-35.us-west-2.compute.internal (5/8))
16/03/28 06:53:11 INFO scheduler.TaskSetManager: Starting Task 7.0 in stage 0.0 (TID 7, ip: 172.31.6-35.us-west-2.compute.internal, partition 7, NODE_LOCAL, 2170 bytes)
16/03/28 06:53:11 INFO scheduler.TaskSetManager: Finished Task 5.0 in stage 0.0 (TID 5, in 1771 ms on ip-172-31-6-35.us-west-2.compute.internal (6/8))
16/03/28 06:53:12 INFO scheduler.TaskSetManager: Starting Task 8.0 in stage 0.0 (TID 8, in 1724 ms on ip-172-31-6-35.us-west-2.compute.internal (7/8))
16/03/28 06:53:12 INFO scheduler.TaskSetManager: Finished Task 6.0 in stage 0.0 (TID 6, in 1217 ms on ip-172-31-6-35.us-west-2.compute.internal (8/8))
16/03/28 06:53:12 INFO scheduler.DAGScheduler: ResultStage 0 (sortByKey at <console>:29) finished in 8.453 s
16/03/28 06:53:12 INFO scheduler.TaskSetManager: Removed TaskSet 0.0, whose tasks have all completed, from pool
16/03/28 06:53:12 INFO scheduler.DAGScheduler: Job 0 finished: sortByKey at <console>:29, took 8.51921 s
data: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[6] at map at <console>:29
scala>

```

```
Send keyboard input to current session only. [ ] - 5 ON

16/03/28 06:53:35 INFO scheduler.DAGScheduler: Registering RDD 3 (map at <console>:29)
16/03/28 06:53:35 INFO scheduler.DAGScheduler: Got job 1 (saveAsTextFile at <console>:32) with 1 output partitions
16/03/28 06:53:35 INFO scheduler.DAGScheduler: Final stage: ResultStage 2 (saveAsTextFile at <console>:32)
16/03/28 06:53:35 INFO scheduler.DAGScheduler: Parents of final stage: List(ShuffleMapStage 1)
16/03/28 06:53:35 INFO scheduler.DAGScheduler: Missing parents: List(ShuffleMapStage 1)
16/03/28 06:53:35 INFO scheduler.DAGScheduler: Submitting ShuffleMapStage 1 (MapPartitionsRDD[3] at map at <console>:29), which has no missing parents
16/03/28 06:53:35 INFO storage.MemoryStore: Block broadcast_2 stored as values in memory (estimated size 5.1 KB, free 43.3 KB)
16/03/28 06:53:35 INFO storage.MemoryStore: Block broadcast_2_piece0 stored as bytes in memory (estimated size 2.8 KB, free 46.1 KB)
16/03/28 06:53:35 INFO storage.BlockManagerInfo: Added broadcast_2_piece0 in memory on 172.31.15.32:47604 (size: 2.8 KB, free: 511.5 MB)
16/03/28 06:53:35 INFO spark.SparkContext: Created broadcast 2 from broadcast at DAGScheduler, scala:1006
16/03/28 06:53:35 INFO scheduler.DAGScheduler: Submitting 8 missing tasks from ShuffleMapStage 1 (MapPartitionsRDD[3] at map at <console>:29)
16/03/28 06:53:35 INFO scheduler.TaskSchedulerImpl: Adding task set 1.0 with 8 tasks
16/03/28 06:53:35 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 1.0 (TID 8, ip-172-31-6-35.us-west-2.compute.internal, partition 0.NODE.LOCAL, 2159 bytes)
16/03/28 06:53:35 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 1.0 (TID 9, ip-172-31-6-35.us-west-2.compute.internal, partition 1.NODE.LOCAL, 2159 bytes)
16/03/28 06:53:35 INFO storage.BlockManagerInfo: Added broadcast_2_piece0 in memory on 172.31.15.32:47604 (size: 2.8 KB, free: 1539.0 MB)
16/03/28 06:53:40 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 1.0 (TID 10, ip-172-31-6-35.us-west-2.compute.internal, partition 2.NODE.LOCAL, 2159 bytes)
16/03/28 06:53:40 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 1.0 (TID 9) in 5520 ms on ip-172-31-6-35.us-west-2.compute.internal (1/8)
16/03/28 06:53:41 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 1.0 (TID 10) in 5509 ms on ip-172-31-6-35.us-west-2.compute.internal (2/8)
16/03/28 06:53:41 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 1.0 (TID 8) in 5509 ms on ip-172-31-6-35.us-west-2.compute.internal (1/8)
16/03/28 06:53:46 INFO scheduler.TaskSetManager: Starting task 4.0 in stage 1.0 (TID 10) in 5286 ms on ip-172-31-6-35.us-west-2.compute.internal, partition 4.NODE.LOCAL, 2159 bytes
16/03/28 06:53:46 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 1.0 (TID 9) in 5286 ms on ip-172-31-6-35.us-west-2.compute.internal (2/8)
16/03/28 06:53:46 INFO scheduler.TaskSetManager: Starting task 5.0 in stage 1.0 (TID 10) in 5286 ms on ip-172-31-6-35.us-west-2.compute.internal, partition 5.NODE.LOCAL, 2159 bytes
16/03/28 06:53:46 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 1.0 (TID 8) in 5300 ms on ip-172-31-6-35.us-west-2.compute.internal (4/8)
16/03/28 06:53:51 INFO scheduler.TaskSetManager: Starting task 6.0 in stage 1.0 (TID 14, ip-172-31-6-35.us-west-2.compute.internal, partition 6.NODE.LOCAL, 2159 bytes)
16/03/28 06:53:51 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 1.0 (TID 12) in 5146 ms on ip-172-31-6-35.us-west-2.compute.internal (5/8)
16/03/28 06:53:51 INFO scheduler.TaskSetManager: Starting task 7.0 in stage 1.0 (TID 15, ip-172-31-6-35.us-west-2.compute.internal, partition 7.NODE.LOCAL, 2159 bytes)
16/03/28 06:53:51 INFO scheduler.TaskSetManager: Finished task 5.0 in stage 1.0 (TID 13) in 5192 ms on ip-172-31-6-35.us-west-2.compute.internal (6/8)
16/03/28 06:53:51 INFO scheduler.TaskSetManager: Finished task 7.0 in stage 1.0 (TID 15) in 2204 ms on ip-172-31-6-35.us-west-2.compute.internal (7/8)
16/03/28 06:53:55 INFO scheduler.DAGScheduler: ShuffleMapStage 1 (map at <console>:29) finished in 19.870 s
16/03/28 06:53:55 INFO scheduler.DAGScheduler: looking for newly runnable stages
16/03/28 06:53:55 INFO scheduler.DAGScheduler: running: Set()
16/03/28 06:53:55 INFO scheduler.DAGScheduler: waiting: Set(ResultStage 2)
16/03/28 06:53:55 INFO scheduler.DAGScheduler: failed: Set()
16/03/28 06:53:55 INFO scheduler.DAGScheduler: Submitting ResultStage 2 (MapPartitionsRDD[9] at saveAsTextFile at <console>:32), which has no missing parents
16/03/28 06:53:55 INFO scheduler.TaskSetManager: Finished task 6.0 in stage 1.0 (TID 14) in 3928 ms on ip-172-31-6-35.us-west-2.compute.internal (8/8)
16/03/28 06:53:55 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 1.0, whose tasks have all completed, from pool
16/03/28 06:53:55 INFO storage.MemoryStore: Block broadcast_3 stored as values in memory (estimated size 16.5 KB, free 62.5 KB)
16/03/28 06:53:55 INFO storage.MemoryStore: Block broadcast_3_piece0 stored as bytes in memory (estimated size 8.6 KB, free 71.1 KB)
16/03/28 06:53:55 INFO storage.BlockManagerInfo: Added broadcast_3_piece0 in memory on 172.31.15.32:47604 (size: 8.6 KB, free: 511.5 MB)
16/03/28 06:53:55 INFO spark.SparkContext: Created broadcast 3 from broadcast at DAGScheduler, scala:1006
16/03/28 06:53:55 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from ResultStage 2 (MapPartitionsRDD[9] at saveAsTextFile at <console>:32)
16/03/28 06:53:55 INFO scheduler.TaskSchedulerImpl: Adding task set 2.0 with 1 tasks
16/03/28 06:53:55 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 2.0 (TID 16, ip-172-31-6-35.us-west-2.compute.internal, partition 0.NODE.LOCAL, 2289 bytes)
16/03/28 06:53:55 INFO storage.BlockManagerInfo: Added broadcast_3_piece0 in memory on 172.31.15.32:47604 (size: 8.6 KB, free: 1539.0 MB)
16/03/28 06:53:55 INFO spark.MapOutputTrackerMasterEndpoint: Asked to send map output locations for shuffle 0 to ip-172-31-6-35.us-west-2.compute.internal:35603
16/03/28 06:53:55 INFO spark.MapOutputTrackerMaster: Size of output statuses for shuffle 0 is 204 bytes
```

```
[Send keyboard input to current session only.]
ON

16/03/28 06:53:35 INFO scheduler.DAGScheduler: Submitting ShuffleMapStage 1 (MapPartitionsRDD[3] at map at <console>:29), which has no missing parents
16/03/28 06:53:35 INFO storage.MemoryStore: Block broadcast_2 stored as values in memory (estimated size 5.1 KB, free 43.3 KB)
16/03/28 06:53:35 INFO storage.MemoryStore: Block broadcast_2_piece0 stored as bytes in memory (estimated size 2.8 KB, free 46.1 KB)
16/03/28 06:53:35 INFO storage.BlockManagerInfo: Added broadcast_2_piece0 in memory on 172.31.15.32:47604 (size: 2.8 KB, free: 511.5 MB)
16/03/28 06:53:35 INFO spark.SparkContext: Created broadcast 2 from broadcast at DAGScheduler.scala:1006
16/03/28 06:53:35 INFO scheduler.DAGScheduler: Submitting 8 missing tasks from ShuffleMapStage 1 (MapPartitionsRDD[3] at map at <console>:29)
16/03/28 06:53:35 INFO scheduler.TaskSchedulerImpl: Adding task set 1.0 with 8 tasks
16/03/28 06:53:35 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 1.0 (TID 8, ip-172-31-6-35.us-west-2.compute.internal, partition 0.NODE_LOCAL, 2159 bytes)
16/03/28 06:53:35 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 1.0 (TID 9, ip-172-31-6-35.us-west-2.compute.internal, partition 1.NODE_LOCAL, 2159 bytes)
16/03/28 06:53:35 INFO storage.BlockManagerInfo: Added broadcast_2_piece0 in memory on ip-172-31-6-35.us-west-2.compute.internal:38679 (size: 2.8 KB, free: 1539.0 KB)
16/03/28 06:53:41 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 1.0 (TID 10, ip-172-31-6-35.us-west-2.compute.internal, partition 2.NODE_LOCAL, 2159 bytes)
16/03/28 06:53:41 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 1.0 (TID 9) in 5520 ms on ip-172-31-6-35.us-west-2.compute.internal (1/8)
16/03/28 06:53:41 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 1.0 (TID 11, ip-172-31-6-35.us-west-2.compute.internal, partition 3.NODE_LOCAL, 2159 bytes)
16/03/28 06:53:41 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 1.0 (TID 8) in 5599 ms on ip-172-31-6-35.us-west-2.compute.internal (2/8)
16/03/28 06:53:46 INFO scheduler.TaskSetManager: Starting task 4.0 in stage 1.0 (TID 12, ip-172-31-6-35.us-west-2.compute.internal, partition 4.NODE_LOCAL, 2159 bytes)
16/03/28 06:53:46 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 1.0 (TID 10) in 5286 ms on ip-172-31-6-35.us-west-2.compute.internal (3/8)
16/03/28 06:53:44 INFO scheduler.TaskSetManager: Starting task 5.0 in stage 1.0 (TID 13, ip-172-31-6-35.us-west-2.compute.internal, partition 5.NODE_LOCAL, 2159 bytes)
16/03/28 06:53:44 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 1.0 (TID 11) in 5309 ms on ip-172-31-6-35.us-west-2.compute.internal (4/8)
16/03/28 06:53:51 INFO scheduler.TaskSetManager: Starting task 6.0 in stage 1.0 (TID 14, ip-172-31-6-35.us-west-2.compute.internal, partition 6.NODE_LOCAL, 2159 bytes)
16/03/28 06:53:51 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 1.0 (TID 12) in 5146 ms on ip-172-31-6-35.us-west-2.compute.internal (5/8)
16/03/28 06:53:51 INFO scheduler.TaskSetManager: Starting task 7.0 in stage 1.0 (TID 15, ip-172-31-6-35.us-west-2.compute.internal, partition 7.NODE_LOCAL, 2159 bytes)
16/03/28 06:53:51 INFO scheduler.TaskSetManager: Finished task 5.0 in stage 1.0 (TID 13) in 5192 ms on ip-172-31-6-35.us-west-2.compute.internal (6/8)
16/03/28 06:53:53 INFO scheduler.TaskSetManager: Finished task 7.0 in stage 1.0 (TID 15) in 2204 ms on ip-172-31-6-35.us-west-2.compute.internal (7/8)
16/03/28 06:53:53 INFO scheduler.DAGScheduler: ShuffleMapStage 1 (map at <console>:29) finished in 19.879 s
16/03/28 06:53:53 INFO scheduler.DAGScheduler: looking for newly runnable stages
16/03/28 06:53:53 INFO scheduler.DAGScheduler: running: Set()
16/03/28 06:53:53 INFO scheduler.DAGScheduler: waiting: Set(ResultStage 2)
16/03/28 06:53:55 INFO scheduler.DAGScheduler: failed: Set()
16/03/28 06:53:55 INFO scheduler.DAGScheduler: Submitting ResultStage 2 (MapPartitionsRDD[9] at saveAsTextFile at <console>:32), which has no missing parents
16/03/28 06:53:55 INFO scheduler.TaskSetManager: Filling task set 1.0 in stage 1.0 (TID 14) in 3928 ms on ip-172-31-6-35.us-west-2.compute.internal (8/8)
16/03/28 06:53:55 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks have all completed, from pool
16/03/28 06:53:55 INFO storage.MemoryStore: Block broadcast_3 stored as values in memory (estimated size 16.5 KB, free 62.5 KB)
16/03/28 06:53:55 INFO storage.MemoryStore: Block broadcast_3_piece0 stored as bytes in memory (estimated size 8.6 KB, free 71.1 KB)
16/03/28 06:53:55 INFO storage.BlockManagerInfo: Added broadcast_3_piece0 in memory on 172.31.15.32:47604 (size: 8.6 KB, free: 511.5 MB)
16/03/28 06:53:55 INFO spark.SparkContext: Created broadcast 3 from broadcast at DAGScheduler.scala:1006
16/03/28 06:53:55 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from ResultStage 2 (MapPartitionsRDD[9] at saveAsTextFile at <console>:32)
16/03/28 06:53:55 INFO scheduler.TaskSchedulerImpl: Adding task set 2.0 with 1 tasks
16/03/28 06:53:55 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 2.0 (TID 16, ip-172-31-6-35.us-west-2.compute.internal, partition 0.NODE_LOCAL, 2289 bytes)
16/03/28 06:53:55 INFO storage.BlockManagerInfo: Added broadcast_3_piece0 in memory on ip-172-31-6-35.us-west-2.compute.internal:38679 (size: 8.6 KB, free: 1539.0 KB)
16/03/28 06:53:55 INFO spark.MapOutputTrackerMasterEndpoint: Asked to send map output locations for shuffle 0 to ip-172-31-6-35.us-west-2.compute.internal:35603
16/03/28 06:53:55 INFO spark.MapOutputTrackerMaster: Size of output statuses for shuffle 0 is 204 bytes
16/03/28 06:54:40 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 2.0 (TID 16) in 45573 ms on ip-172-31-6-35.us-west-2.compute.internal (1/1)
16/03/28 06:54:40 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 2.0, whose tasks have all completed, from pool
16/03/28 06:54:40 INFO scheduler.DAGScheduler: ResultStage 2 (saveAsTextFile at <console>:32) finished in 45.379 s
16/03/28 06:54:40 INFO scheduler.DAGScheduler: Job 1 finished: saveAsTextFile at <console>:32, took 65.298582 s

scala> [OK]
```



```

Send keyboard input to current session only. [+] ON
16/03/28 07:44:24 INFO storage.BlockManagerInfo: Added broadcast_13_piece0 in memory on 172.31.15.32:47604 (size: 5.6 KB, free: 511.5 MB)
16/03/28 07:44:24 INFO spark.SparkContext: Created broadcast 13 from textFile at <console>:27
input: org.apache.spark.rdd.RDD[String] = /input MapPartitionsRDD[32] at textFile at <console>:27

scale> val data = input.flatMap(line => line.split("\n")).map(line => (line, "")).sortByKey().map(case(k,v)=>k+v)

16/03/28 07:44:44 INFO mapred.FileInputFormat: Total input paths to process : 1
16/03/28 07:44:44 INFO spark.SparkContext: Starting job: sortByKey at <console>:29
16/03/28 07:44:44 INFO scheduler.DAGScheduler: Got job 8 (sortByKey at <console>:29) with 75 output partitions
16/03/28 07:44:44 INFO scheduler.DAGScheduler: Final stage: ResultStage 12 (sortByKey at <console>:29)
16/03/28 07:44:44 INFO scheduler.DAGScheduler: Parents of final stage: List()
16/03/28 07:44:44 INFO scheduler.DAGScheduler: Missing parents: List()
16/03/28 07:44:44 INFO scheduler.DAGScheduler: Submitting ResultStage 12 (MapPartitionsRDD[36] at sortByKey at <console>:29), which has no missing parents
16/03/28 07:44:44 INFO storage.MemoryStore: Block broadcast_14 stored as values in memory (estimated size 3.9 KB, free 168.4 KB)
16/03/28 07:44:44 INFO storage.MemoryStore: Block broadcast_14_piece0 stored as bytes in memory (estimated size 2.1 KB, free 170.5 KB)
16/03/28 07:44:44 INFO storage.BlockManagerInfo: Added broadcast_14_piece0 in memory on 172.31.15.32:47604 (size: 2.1 KB, free: 511.5 MB)
16/03/28 07:44:44 INFO spark.SparkContext: Created broadcast 14 from broadcast at DAGScheduler.scala:1096
16/03/28 07:44:44 INFO scheduler.DAGScheduler: Submitting 75 missing tasks from ResultStage 12 (MapPartitionsRDD[36] at sortByKey at <console>:29)
16/03/28 07:44:44 INFO scheduler.TaskSchedulerImpl: Adding task set 12.0 with 75 tasks
16/03/28 07:44:44 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 12.0 (TID 89, ip:172-31-6-35.us-west-2.compute.internal, partition 0.NODE_LOCAL, 2170 bytes)
16/03/28 07:44:44 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 12.0 (TID 90, ip:172-31-6-35.us-west-2.compute.internal, partition 1.NODE_LOCAL, 2170 bytes)
16/03/28 07:44:44 INFO storage.BlockManagerInfo: Added broadcast_14_piece0 in memory on ip:172-31-6-35.us-west-2.compute.internal:38679 (size: 2.1 KB, free: 1539.0 MB)
16/03/28 07:44:44 INFO storage.BlockManagerInfo: Added broadcast_14_piece0 in memory on ip:172-31-6-35.us-west-2.compute.internal:38679 (size: 5.6 KB, free: 1539.0 MB)
16/03/28 07:44:44 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 12.0 (TID 91, ip:172-31-6-35.us-west-2.compute.internal, partition 2.NODE_LOCAL, 2170 bytes)
16/03/28 07:44:44 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 12.0 (TID 89) in 1877 ms on ip:172-31-6-35.us-west-2.compute.internal (1/75)
16/03/28 07:44:44 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 12.0 (TID 92, ip:172-31-6-35.us-west-2.compute.internal, partition 3.NODE_LOCAL, 2170 bytes)
16/03/28 07:44:44 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 12.0 (TID 90) in 1988 ms on ip:172-31-6-35.us-west-2.compute.internal (2/75)
16/03/28 07:44:44 INFO scheduler.TaskSetManager: Starting task 4.0 in stage 12.0 (TID 93, ip:172-31-6-35.us-west-2.compute.internal, partition 4.NODE_LOCAL, 2170 bytes)
16/03/28 07:44:44 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 12.0 (TID 91) in 1715 ms on ip:172-31-6-35.us-west-2.compute.internal (3/75)
16/03/28 07:44:44 INFO scheduler.TaskSetManager: Starting task 5.0 in stage 12.0 (TID 94, ip:172-31-6-35.us-west-2.compute.internal, partition 5.NODE_LOCAL, 2170 bytes)
16/03/28 07:44:44 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 12.0 (TID 92) in 1786 ms on ip:172-31-6-35.us-west-2.compute.internal (4/75)
16/03/28 07:44:44 INFO scheduler.TaskSetManager: Starting task 6.0 in stage 12.0 (TID 95, ip:172-31-6-35.us-west-2.compute.internal, partition 6.NODE_LOCAL, 2170 bytes)
16/03/28 07:44:44 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 12.0 (TID 93) in 1786 ms on ip:172-31-6-35.us-west-2.compute.internal (5/75)
16/03/28 07:44:44 INFO scheduler.TaskSetManager: Starting task 7.0 in stage 12.0 (TID 96, ip:172-31-6-35.us-west-2.compute.internal, partition 7.NODE_LOCAL, 2170 bytes)
16/03/28 07:44:44 INFO scheduler.TaskSetManager: Finished task 5.0 in stage 12.0 (TID 94) in 1693 ms on ip:172-31-6-35.us-west-2.compute.internal (6/75)
16/03/28 07:44:45 INFO scheduler.TaskSetManager: Starting task 8.0 in stage 12.0 (TID 97, ip:172-31-6-35.us-west-2.compute.internal, partition 8.NODE_LOCAL, 2170 bytes)
16/03/28 07:44:45 INFO scheduler.TaskSetManager: Finished task 6.0 in stage 12.0 (TID 95) in 1688 ms on ip:172-31-6-35.us-west-2.compute.internal (7/75)
16/03/28 07:44:45 INFO scheduler.TaskSetManager: Starting task 9.0 in stage 12.0 (TID 98, ip:172-31-6-35.us-west-2.compute.internal, partition 9.NODE_LOCAL, 2170 bytes)
16/03/28 07:44:45 INFO scheduler.TaskSetManager: Finished task 7.0 in stage 12.0 (TID 96) in 1718 ms on ip:172-31-6-35.us-west-2.compute.internal (8/75)
16/03/28 07:44:45 INFO scheduler.TaskSetManager: Starting task 10.0 in stage 12.0 (TID 99, ip:172-31-6-35.us-west-2.compute.internal, partition 10.NODE_LOCAL, 2170 bytes)
16/03/28 07:44:45 INFO scheduler.TaskSetManager: Finished task 8.0 in stage 12.0 (TID 97) in 1747 ms on ip:172-31-6-35.us-west-2.compute.internal (9/75)
16/03/28 07:44:45 INFO scheduler.TaskSetManager: Starting task 11.0 in stage 12.0 (TID 100, ip:172-31-6-35.us-west-2.compute.internal, partition 11.NODE_LOCAL, 2170 bytes)
16/03/28 07:44:45 INFO scheduler.TaskSetManager: Finished task 9.0 in stage 12.0 (TID 98) in 1780 ms on ip:172-31-6-35.us-west-2.compute.internal (10/75)
16/03/28 07:44:45 INFO scheduler.TaskSetManager: Starting task 12.0 in stage 12.0 (TID 101, ip:172-31-6-35.us-west-2.compute.internal, partition 12.NODE_LOCAL, 2170 bytes)
16/03/28 07:44:45 INFO scheduler.TaskSetManager: Finished task 10.0 in stage 12.0 (TID 99) in 1784 ms on ip:172-31-6-35.us-west-2.compute.internal (11/75)
16/03/28 07:44:45 INFO scheduler.TaskSetManager: Starting task 13.0 in stage 12.0 (TID 102, ip:172-31-6-35.us-west-2.compute.internal, partition 13.NODE_LOCAL, 2170 bytes)
16/03/28 07:44:45 INFO scheduler.TaskSetManager: Finished task 11.0 in stage 12.0 (TID 100) in 1810 ms on ip:172-31-6-35.us-west-2.compute.internal (12/75)

```

```

Send keyboard input to current session only. [+] ON
16/03/28 07:45:40 INFO scheduler.TaskSetManager: Starting task 66.0 in stage 12.0 (TID 155, ip:172-31-6-35.us-west-2.compute.internal, partition 66.NODE_LOCAL, 2170 bytes)
16/03/28 07:45:40 INFO scheduler.TaskSetManager: Finished task 64.0 in stage 12.0 (TID 153) in 1668 ms on ip:172-31-6-35.us-west-2.compute.internal (65/75)
16/03/28 07:45:42 INFO scheduler.TaskSetManager: Starting task 67.0 in stage 12.0 (TID 156, ip:172-31-6-35.us-west-2.compute.internal, partition 67.NODE_LOCAL, 2170 bytes)
16/03/28 07:45:42 INFO scheduler.TaskSetManager: Finished task 65.0 in stage 12.0 (TID 154) in 1764 ms on ip:172-31-6-35.us-west-2.compute.internal (66/75)
16/03/28 07:45:42 INFO scheduler.TaskSetManager: Starting task 68.0 in stage 12.0 (TID 157, ip:172-31-6-35.us-west-2.compute.internal, partition 68.NODE_LOCAL, 2170 bytes)
16/03/28 07:45:42 INFO scheduler.TaskSetManager: Finished task 66.0 in stage 12.0 (TID 155) in 1686 ms on ip:172-31-6-35.us-west-2.compute.internal (67/75)
16/03/28 07:45:43 INFO scheduler.TaskSetManager: Starting task 69.0 in stage 12.0 (TID 158, ip:172-31-6-35.us-west-2.compute.internal, partition 69.NODE_LOCAL, 2170 bytes)
16/03/28 07:45:43 INFO scheduler.TaskSetManager: Finished task 67.0 in stage 12.0 (TID 156) in 1701 ms on ip:172-31-6-35.us-west-2.compute.internal (68/75)
16/03/28 07:45:44 INFO scheduler.TaskSetManager: Starting task 70.0 in stage 12.0 (TID 159, ip:172-31-6-35.us-west-2.compute.internal, partition 70.NODE_LOCAL, 2170 bytes)
16/03/28 07:45:44 INFO scheduler.TaskSetManager: Finished task 68.0 in stage 12.0 (TID 157) in 1682 ms on ip:172-31-6-35.us-west-2.compute.internal (69/75)
16/03/28 07:45:45 INFO scheduler.TaskSetManager: Starting task 71.0 in stage 12.0 (TID 160, ip:172-31-6-35.us-west-2.compute.internal, partition 71.NODE_LOCAL, 2170 bytes)
16/03/28 07:45:45 INFO scheduler.TaskSetManager: Finished task 69.0 in stage 12.0 (TID 158) in 1704 ms on ip:172-31-6-35.us-west-2.compute.internal (70/75)
16/03/28 07:45:46 INFO scheduler.TaskSetManager: Starting task 72.0 in stage 12.0 (TID 161, ip:172-31-6-35.us-west-2.compute.internal, partition 72.NODE_LOCAL, 2170 bytes)
16/03/28 07:45:46 INFO scheduler.TaskSetManager: Finished task 70.0 in stage 12.0 (TID 159) in 1741 ms on ip:172-31-6-35.us-west-2.compute.internal (71/75)
16/03/28 07:45:47 INFO scheduler.TaskSetManager: Starting task 73.0 in stage 12.0 (TID 162, ip:172-31-6-35.us-west-2.compute.internal, partition 73.NODE_LOCAL, 2170 bytes)
16/03/28 07:45:47 INFO scheduler.TaskSetManager: Finished task 71.0 in stage 12.0 (TID 160) in 1767 ms on ip:172-31-6-35.us-west-2.compute.internal (72/75)
16/03/28 07:45:47 INFO scheduler.TaskSetManager: Starting task 74.0 in stage 12.0 (TID 163, ip:172-31-6-35.us-west-2.compute.internal, partition 74.NODE_LOCAL, 2170 bytes)
16/03/28 07:45:47 INFO scheduler.TaskSetManager: Finished task 72.0 in stage 12.0 (TID 161) in 1749 ms on ip:172-31-6-35.us-west-2.compute.internal (73/75)
16/03/28 07:45:48 INFO scheduler.TaskSetManager: Finished task 74.0 in stage 12.0 (TID 163) in 1902 ms on ip:172-31-6-35.us-west-2.compute.internal (74/75)
16/03/28 07:45:48 INFO scheduler.TaskSetManager: Finished task 73.0 in stage 12.0 (TID 162) in 1574 ms on ip:172-31-6-35.us-west-2.compute.internal (75/75)
16/03/28 07:45:48 INFO DAGScheduler: ResultStage 12 (sortByKey at <console>:29) finished in 64.529 s
16/03/28 07:45:48 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 12.0, whose tasks have all completed, from pool
16/03/28 07:45:48 INFO scheduler.DAGScheduler: Job 8 finished: sortByKey at <console>:29, took 64.544514 s
data: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[38] at map at <console>:29

scale> data.saveAsTextFile("Output")
16/03/28 07:45:47 INFO spark.SparkContext: Starting job: saveAsTextFile at <console>:32
16/03/28 07:45:47 INFO scheduler.DAGScheduler: Registering RDD 34 (map at <console>:32)
16/03/28 07:45:47 INFO scheduler.DAGScheduler: Got job 9 (saveAsTextFile at <console>:32) with 75 output partitions
16/03/28 07:45:47 INFO scheduler.DAGScheduler: Final stage: ResultStage 14 (saveAsTextFile at <console>:32)
16/03/28 07:45:47 INFO scheduler.DAGScheduler: Parents of final stage: List(ShuffleMapStage 13)
16/03/28 07:45:47 INFO scheduler.DAGScheduler: Missing parents: List()
16/03/28 07:45:47 INFO scheduler.DAGScheduler: Submitting ShuffleMapStage 13 (MapPartitionsRDD[34] at map at <console>:32), which has no missing parents
16/03/28 07:45:47 INFO storage.MemoryStore: Block broadcast_15 stored as values in memory (estimated size 11.7 KB, free 182.2 KB)
16/03/28 07:45:47 INFO storage.MemoryStore: Block broadcast_15_piece0 stored as bytes in memory (estimated size 6.1 KB, free 188.3 KB)
16/03/28 07:45:47 INFO storage.BlockManagerInfo: Added broadcast_15_piece0 in memory on 172.31.15.32:47604 (size: 6.1 KB, free: 511.5 MB)
16/03/28 07:45:47 INFO spark.SparkContext: Created broadcast 15 from broadcast at DAGScheduler.scala:1096
16/03/28 07:45:47 INFO scheduler.DAGScheduler: Submitting 75 missing tasks from ShuffleMapStage 13 (MapPartitionsRDD[34] at map at <console>:32)
16/03/28 07:45:47 INFO scheduler.TaskSchedulerImpl: Adding task set 13.0 with 75 tasks
16/03/28 07:45:47 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 13.0 (TID 164, ip:172-31-6-35.us-west-2.compute.internal, partition 0.NODE_LOCAL, 2159 bytes)
16/03/28 07:45:47 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 13.0 (TID 165, ip:172-31-6-35.us-west-2.compute.internal, partition 1.NODE_LOCAL, 2159 bytes)
16/03/28 07:45:47 INFO storage.BlockManagerInfo: Added broadcast_15_piece0 in memory on ip:172-31-6-35.us-west-2.compute.internal:38679 (size: 6.1 KB, free: 1539.0 MB)
16/03/28 07:45:47 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 13.0 (TID 166, ip:172-31-6-35.us-west-2.compute.internal, partition 2.NODE_LOCAL, 2159 bytes)
16/03/28 07:45:48 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 13.0 (TID 165) in 6051 ms on ip:172-31-6-35.us-west-2.compute.internal (1/75)
16/03/28 07:45:48 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 13.0 (TID 167, ip:172-31-6-35.us-west-2.compute.internal, partition 3.NODE_LOCAL, 2159 bytes)
16/03/28 07:45:48 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 13.0 (TID 164) in 6180 ms on ip:172-31-6-35.us-west-2.compute.internal (2/75)

```

```
+-----+  
Send keyboard input to current session only.  
16/03/28 07:56:00 INFO scheduler.TaskSetManager: Starting task 55.0 in stage 14.0 (TID 294, ip-172-31-6-35.us-west-2.compute.internal, partition 55,NODE_LOCAL, 1894 bytes)  
16/03/28 07:56:00 INFO scheduler.TaskSetManager: Finished task 53.0 in stage 14.0 (TID 292) in 17181 ms on ip-172-31-6-35.us-west-2.compute.internal (54/75)  
16/03/28 07:56:00 INFO scheduler.TaskSetManager: Starting task 56.0 in stage 14.0 (TID 295, ip-172-31-6-35.us-west-2.compute.internal, partition 56,NODE_LOCAL, 1894 bytes)  
16/03/28 07:56:00 INFO scheduler.TaskSetManager: Finished task 54.0 in stage 14.0 (TID 293) in 11882 ms on ip-172-31-6-35.us-west-2.compute.internal (55/75)  
16/03/28 07:56:11 INFO scheduler.TaskSetManager: Starting task 57.0 in stage 14.0 (TID 296, ip-172-31-6-35.us-west-2.compute.internal, partition 57,NODE_LOCAL, 1894 bytes)  
16/03/28 07:56:11 INFO scheduler.TaskSetManager: Finished task 55.0 in stage 14.0 (TID 294) in 10639 ms on ip-172-31-6-35.us-west-2.compute.internal (56/75)  
16/03/28 07:56:21 INFO scheduler.TaskSetManager: Starting task 58.0 in stage 14.0 (TID 297, ip-172-31-6-35.us-west-2.compute.internal, partition 58,NODE_LOCAL, 1894 bytes)  
16/03/28 07:56:21 INFO scheduler.TaskSetManager: Finished task 56.0 in stage 14.0 (TID 295) in 13297 ms on ip-172-31-6-35.us-west-2.compute.internal (57/75)  
16/03/28 07:56:23 INFO scheduler.TaskSetManager: Starting task 59.0 in stage 14.0 (TID 298, ip-172-31-6-35.us-west-2.compute.internal, partition 59,NODE_LOCAL, 1894 bytes)  
16/03/28 07:56:23 INFO scheduler.TaskSetManager: Finished task 57.0 in stage 14.0 (TID 296) in 12904 ms on ip-172-31-6-35.us-west-2.compute.internal (58/75)  
16/03/28 07:56:30 INFO scheduler.TaskSetManager: Starting task 60.0 in stage 14.0 (TID 299, ip-172-31-6-35.us-west-2.compute.internal, partition 60,NODE_LOCAL, 1894 bytes)  
16/03/28 07:56:30 INFO scheduler.TaskSetManager: Finished task 58.0 in stage 14.0 (TID 297) in 9907 ms on ip-172-31-6-35.us-west-2.compute.internal (59/75)  
16/03/28 07:56:30 INFO scheduler.TaskSetManager: Starting task 61.0 in stage 14.0 (TID 300, ip-172-31-6-35.us-west-2.compute.internal, partition 61,NODE_LOCAL, 1894 bytes)  
16/03/28 07:56:39 INFO scheduler.TaskSetManager: Finished task 59.0 in stage 14.0 (TID 298) in 12052 ms on ip-172-31-6-35.us-west-2.compute.internal (60/75)  
16/03/28 07:56:39 INFO scheduler.TaskSetManager: Starting task 62.0 in stage 14.0 (TID 301, ip-172-31-6-35.us-west-2.compute.internal, partition 62,NODE_LOCAL, 1894 bytes)  
16/03/28 07:56:39 INFO scheduler.TaskSetManager: Finished task 60.0 in stage 14.0 (TID 299) in 8957 ms on ip-172-31-6-35.us-west-2.compute.internal (61/75)  
16/03/28 07:56:44 INFO scheduler.TaskSetManager: Starting task 63.0 in stage 14.0 (TID 302, ip-172-31-6-35.us-west-2.compute.internal, partition 63,NODE_LOCAL, 1894 bytes)  
16/03/28 07:56:44 INFO scheduler.TaskSetManager: Finished task 61.0 in stage 14.0 (TID 300) in 10194 ms on ip-172-31-6-35.us-west-2.compute.internal (62/75)  
16/03/28 07:56:44 INFO scheduler.TaskSetManager: Starting task 64.0 in stage 14.0 (TID 303, ip-172-31-6-35.us-west-2.compute.internal, partition 64,NODE_LOCAL, 1894 bytes)  
16/03/28 07:56:44 INFO scheduler.TaskSetManager: Finished task 62.0 in stage 14.0 (TID 301) in 9771 ms on ip-172-31-6-35.us-west-2.compute.internal (63/75)  
16/03/28 07:56:57 INFO scheduler.TaskSetManager: Starting task 65.0 in stage 14.0 (TID 304, ip-172-31-6-35.us-west-2.compute.internal, partition 65,NODE_LOCAL, 1894 bytes)  
16/03/28 07:56:57 INFO scheduler.TaskSetManager: Finished task 63.0 in stage 14.0 (TID 302) in 11223 ms on ip-172-31-6-35.us-west-2.compute.internal (64/75)  
16/03/28 07:57:01 INFO scheduler.TaskSetManager: Starting task 66.0 in stage 14.0 (TID 305, ip-172-31-6-35.us-west-2.compute.internal, partition 66,NODE_LOCAL, 1894 bytes)  
16/03/28 07:57:01 INFO scheduler.TaskSetManager: Finished task 64.0 in stage 14.0 (TID 303) in 12886 ms on ip-172-31-6-35.us-west-2.compute.internal (65/75)  
16/03/28 07:57:09 INFO scheduler.TaskSetManager: Starting task 67.0 in stage 14.0 (TID 306, ip-172-31-6-35.us-west-2.compute.internal, partition 67,NODE_LOCAL, 1894 bytes)  
16/03/28 07:57:09 INFO scheduler.TaskSetManager: Finished task 65.0 in stage 14.0 (TID 304) in 11851 ms on ip-172-31-6-35.us-west-2.compute.internal (66/75)  
16/03/28 07:57:14 INFO scheduler.TaskSetManager: Starting task 68.0 in stage 14.0 (TID 307, ip-172-31-6-35.us-west-2.compute.internal, partition 68,NODE_LOCAL, 1894 bytes)  
16/03/28 07:57:14 INFO scheduler.TaskSetManager: Finished task 66.0 in stage 14.0 (TID 305) in 12495 ms on ip-172-31-6-35.us-west-2.compute.internal (67/75)  
16/03/28 07:57:16 INFO scheduler.TaskSetManager: Starting task 69.0 in stage 14.0 (TID 308, ip-172-31-6-35.us-west-2.compute.internal, partition 69,NODE_LOCAL, 1894 bytes)  
16/03/28 07:57:16 INFO scheduler.TaskSetManager: Finished task 67.0 in stage 14.0 (TID 306) in 7517 ms on ip-172-31-6-35.us-west-2.compute.internal (68/75)  
16/03/28 07:57:24 INFO scheduler.TaskSetManager: Starting task 70.0 in stage 14.0 (TID 309, ip-172-31-6-35.us-west-2.compute.internal, partition 70,NODE_LOCAL, 1894 bytes)  
16/03/28 07:57:24 INFO scheduler.TaskSetManager: Finished task 68.0 in stage 14.0 (TID 307) in 12153 ms on ip-172-31-6-35.us-west-2.compute.internal (69/75)  
16/03/28 07:57:28 INFO scheduler.TaskSetManager: Starting task 71.0 in stage 14.0 (TID 310, ip-172-31-6-35.us-west-2.compute.internal, partition 71,NODE_LOCAL, 1894 bytes)  
16/03/28 07:57:28 INFO scheduler.TaskSetManager: Finished task 69.0 in stage 14.0 (TID 308) in 11630 ms on ip-172-31-6-35.us-west-2.compute.internal (70/75)  
16/03/28 07:57:38 INFO scheduler.TaskSetManager: Starting task 72.0 in stage 14.0 (TID 311, ip-172-31-6-35.us-west-2.compute.internal, partition 72,NODE_LOCAL, 1894 bytes)  
16/03/28 07:57:38 INFO scheduler.TaskSetManager: Finished task 71.0 in stage 14.0 (TID 310) in 9988 ms on ip-172-31-6-35.us-west-2.compute.internal (71/75)  
16/03/28 07:57:44 INFO scheduler.TaskSetManager: Starting task 73.0 in stage 14.0 (TID 312, ip-172-31-6-35.us-west-2.compute.internal, partition 73,NODE_LOCAL, 1894 bytes)  
16/03/28 07:57:44 INFO scheduler.TaskSetManager: Finished task 70.0 in stage 14.0 (TID 309) in 13954 ms on ip-172-31-6-35.us-west-2.compute.internal (72/75)  
16/03/28 07:57:49 INFO scheduler.TaskSetManager: Starting task 74.0 in stage 14.0 (TID 313, ip-172-31-6-35.us-west-2.compute.internal, partition 74,NODE_LOCAL, 1894 bytes)  
16/03/28 07:57:49 INFO scheduler.TaskSetManager: Finished task 72.0 in stage 14.0 (TID 311) in 11477 ms on ip-172-31-6-35.us-west-2.compute.internal (73/75)  
16/03/28 07:57:49 INFO scheduler.TaskSetManager: Finished task 73.0 in stage 14.0 (TID 312) in 9245 ms on ip-172-31-6-35.us-west-2.compute.internal (74/75)  
16/03/28 07:57:55 INFO scheduler.TaskSetManager: Finished task 74.0 in stage 14.0 (TID 313) in 5467 ms on ip-172-31-6-35.us-west-2.compute.internal (75/75)  
16/03/28 07:57:55 INFO scheduler.TaskSchedulerImpl: ResultStage 14 (saveAsTextFile at <console>:32) finished in 492.148 s  
16/03/28 07:57:55 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 14.0, whose tasks have all completed, from pool  
16/03/28 07:57:55 INFO scheduler.DAGScheduler: Job 9 finished: saveAsTextFile at <console>:32, took 667.327069 s  
  
scala: 4
```



## Spark 100GB on 16 Nodes

**NameNode 'ip-172-31-19-190.us-west-2.compute.internal:9000'**

Started: Mon Mar 28 23:08:26 UTC 2016  
 Version: 1.0.4, r1393290  
 Compiled: Wed Oct 3 05:13:58 UTC 2012 by hortonfo  
 Upgrades: There are no upgrades in progress.

[Browse the filesystem](#)  
[Namenode Logs](#)

**Cluster Summary**

2 files and directories, 746 blocks = 748 total. Heap Size is 443 MB / 889 MB (49%)

Configured Capacity	:	1.23 TB
DFS Used	:	282.09 GB
Non DFS Used	:	85.3 GB
DFS Remaining	:	892.27 GB
DFS Used%	:	22.39 %
DFS Remaining%	:	70.83 %
<a href="#">Live Nodes</a>	:	16
<a href="#">Dead Nodes</a>	:	0
<a href="#">Decommissioning Nodes</a>	:	0
Number of Under-Replicated Blocks	:	0

---

**NameNode Storage:**

[prog2-v5.pdf](#) Show all downloads... 7:36 PM 3/28/2016

**Live Datanodes : 16**

Node	Last Contact	Admin State	Configured Capacity (GB)	Used (GB)	Non DFS Used (GB)	Remaining (GB)	Used (%)	Used (%)	Remaining (%)	Blocks
ip-172-31-16-5	1	In Service	78.73	18.66	5.35	54.72	23.71	<div style="width: 23.71%;"></div>	69.5	148
ip-172-31-17-108	1	In Service	78.73	15.51	5.35	57.87	19.7	<div style="width: 19.7%;"></div>	73.51	123
ip-172-31-17-161	0	In Service	78.73	16.27	5.35	57.12	20.66	<div style="width: 20.66%;"></div>	72.55	129
ip-172-31-19-229	1	In Service	78.73	17.41	5.35	55.97	22.11	<div style="width: 22.11%;"></div>	71.1	139
ip-172-31-23-22	1	In Service	78.73	18.66	5.35	54.72	23.71	<div style="width: 23.71%;"></div>	69.5	148
ip-172-31-23-231	2	In Service	78.73	19.8	5.35	53.59	25.15	<div style="width: 25.15%;"></div>	68.06	157
ip-172-31-25-22	2	In Service	78.73	15.92	5.32	57.49	20.22	<div style="width: 20.22%;"></div>	73.02	127
ip-172-31-27-166	0	In Service	78.73	17.02	5.35	56.36	21.62	<div style="width: 21.62%;"></div>	71.59	135
ip-172-31-27-47	1	In Service	78.73	19.29	5.23	54.22	24.5	<div style="width: 24.5%;"></div>	68.86	152
ip-172-31-28-255	1	In Service	78.73	18.79	5.35	54.59	23.87	<div style="width: 23.87%;"></div>	69.34	149
ip-172-31-29-125	2	In Service	78.73	16.27	5.35	57.12	20.66	<div style="width: 20.66%;"></div>	72.55	129
ip-172-31-29-9	1	In Service	78.73	17.53	5.35	55.86	22.26	<div style="width: 22.26%;"></div>	70.95	139
ip-172-31-30-212	1	In Service	78.73	17.53	5.35	55.86	22.26	<div style="width: 22.26%;"></div>	70.95	139
ip-172-31-30-253	0	In Service	78.73	18.41	5.35	54.97	23.39	<div style="width: 23.39%;"></div>	69.82	146
ip-172-31-30-28	1	In Service	78.73	16.27	5.35	57.12	20.66	<div style="width: 20.66%;"></div>	72.55	129
ip-172-31-31-119	2	In Service	78.73	18.76	5.26	54.71	23.83	<div style="width: 23.83%;"></div>	69.49	149

This is Apache Hadoop release 1.0.4

[prog2-v5.pdf](#) Show all downloads... 7:36 PM 3/28/2016

Spark-Master - Xshell 5 (Free for Home/School)

File Edit View Tools Tab Window Help

ssh://root@ec2-52-37-249-240.us-west-2.compute.amazonaws.com:22

To add the current session, click on the left arrow button.

Cloud Spark-Master Spark-Master +

```

16/03/29 00:37:31 INFO metastore.ObjectStore: Setting MetaStore object pin classes with hive.metastore.cache.pinobjtypes="Table,StorageDescriptor,SerDeInfo,Partition,Database,Type,FieldSchema,.Order"
16/03/29 00:37:32 INFO DataNucleus.Datastore: The class "org.apache.hadoop.hive.metastore.model.MFieldSchema" is tagged as "embedded-only" so does not have its own datastore table.
16/03/29 00:37:32 INFO DataNucleus.Datastore: The class "org.apache.hadoop.hive.metastore.model.MOrder" is tagged as "embedded-only" so does not have its own datastore table.
16/03/29 00:37:33 INFO DataNucleus.Datastore: The class "org.apache.hadoop.hive.metastore.model.MFieldSchema" is tagged as "embedded-only" so does not have its own datastore table.
16/03/29 00:37:33 INFO DataNucleus.Datastore: The class "org.apache.hadoop.hive.metastore.model.MOrder" is tagged as "embedded-only" so does not have its own datastore table.
16/03/29 00:37:34 INFO metastore.MetastoreDirectSql: Using direct SQL, underlying DB is Derby
16/03/29 00:37:34 INFO metastore.ObjectStore: Initialized objectStore
16/03/29 00:37:34 WARN metastore.ObjectStore: Version information not found in metastore. hive.metastore.schema.verification is not enabled so recording the schema version 1.2.0
16/03/29 00:37:34 WARN metastore.ObjectStore: Failed to get database default, returning NoSuchObjectException
16/03/29 00:37:34 INFO metastore.HiveMetastore: Added admin role in metastore
16/03/29 00:37:34 INFO metastore.HiveMetastore: Added public role in metastore
16/03/29 00:37:34 INFO metastore.HiveMetastore: No user is added in admin role, since config is empty
16/03/29 00:37:34 INFO metastore.HiveMetastore: 0: get_all_databases
16/03/29 00:37:34 INFO HiveMetaStore.audit: ugi=root ip=unknown-ip-addr cmd=get_all_databases
16/03/29 00:37:34 INFO metastore.HiveMetastore: 0: get_functions: db=default pat=*
16/03/29 00:37:34 INFO DataNucleus.Datastore: The class "org.apache.hadoop.hive.metastore.model.MResourceUri" is tagged as "embedded-only" so does not have its own datastore table.
16/03/29 00:37:34 INFO session.SessionState: Created local directory: /tmp/9822177a-bb97-43bc-b3a7-61af6a9bcebc_resources
16/03/29 00:37:34 INFO session.SessionState: Created HDFS directory: /tmp/hive/root/9822177a-bb97-43bc-b3a7-61af6a9bcebc
16/03/29 00:37:34 INFO session.SessionState: Created local directory: /tmp/root/9822177a-bb97-43bc-b3a7-61af6a9bcebc
16/03/29 00:37:34 INFO session.SessionState: Created HDFS directory: /tmp/hive/root/9822177a-bb97-43bc-b3a7-61af6a9bcebc/_tmp_space.db
16/03/29 00:37:34 INFO repl.SparkILoop: Created sql context (with Hive support)..SQL context available as sqlContext.

scala> val input = sc.textFile("Input")
16/03/29 00:38:35 INFO storage.MemoryStore: Block broadcast_0 stored as values in memory (estimated size 26.5 KB, free 26.5 KB)
16/03/29 00:38:35 INFO storage.MemoryStore: Block broadcast_0_piece0 stored as bytes in memory (estimated size 5.6 KB, free 26.5 KB)
16/03/29 00:38:35 INFO storage.BlockManagerInfo: Added broadcast_0_piece0 in memory on 172.31.19.190:38494 (size: 5.6 KB, free: 511.5 MB)
16/03/29 00:38:35 INFO spark.SparkContext: Created broadcast 0 from textFile at <console>:27
input: org.apache.spark.rdd.RDD[String] = /Input MapPartitionsRDD[1] at textFile at <console>:27

scala>
```

Send text to all sessions within current window

ssh://root@ec2-52-37-249-240.us-west-2.compute.amazonaws.com:22

SSH2 xterm I\* 191x33 33,8 3 sessions CAP NUM 7:38 PM 3/28/2016

```

16/03/29 00:39:20 INFO scheduler.TaskSetManager: Finished task + compute.internal (702/745)
16/03/29 00:39:20 INFO scheduler.TaskSetManager: Finished task 690.0 in stage 0.0 (TID 697) in 1741 ms on ip-172-31-28-255.us-west-2.compute.internal (703/745)
16/03/29 00:39:20 INFO scheduler.TaskSetManager: Finished task 685.0 in stage 0.0 (TID 703) in 1591 ms on ip-172-31-27-47.us-west-2.compute.internal (704/745)
16/03/29 00:39:20 INFO scheduler.TaskSetManager: Finished task 706.0 in stage 0.0 (TID 705) in 1572 ms on ip-172-31-25-22.us-west-2.compute.internal (705/745)
16/03/29 00:39:20 INFO scheduler.TaskSetManager: Finished task 692.0 in stage 0.0 (TID 707) in 1591 ms on ip-172-31-31-119.us-west-2.compute.internal (706/745)
16/03/29 00:39:20 INFO scheduler.TaskSetManager: Finished task 698.0 in stage 0.0 (TID 708) in 1580 ms on ip-172-31-31-119.us-west-2.compute.internal (707/745)
16/03/29 00:39:20 INFO scheduler.TaskSetManager: Finished task 691.0 in stage 0.0 (TID 702) in 1709 ms on ip-172-31-28-255.us-west-2.compute.internal (708/745)
16/03/29 00:39:20 INFO scheduler.TaskSetManager: Finished task 695.0 in stage 0.0 (TID 706) in 1626 ms on ip-172-31-30-212.us-west-2.compute.internal (709/745)
16/03/29 00:39:20 INFO scheduler.TaskSetManager: Finished task 700.0 in stage 0.0 (TID 709) in 1632 ms on ip-172-31-23-22.us-west-2.compute.internal (710/745)
16/03/29 00:39:20 INFO scheduler.TaskSetManager: Finished task 710.0 in stage 0.0 (TID 714) in 1503 ms on ip-172-31-27-47.us-west-2.compute.internal (711/745)
16/03/29 00:39:20 INFO scheduler.TaskSetManager: Finished task 715.0 in stage 0.0 (TID 722) in 1311 ms on ip-172-31-30-212.us-west-2.compute.internal (712/745)
16/03/29 00:39:20 INFO scheduler.TaskSetManager: Finished task 708.0 in stage 0.0 (TID 710) in 1624 ms on ip-172-31-23-22.us-west-2.compute.internal (713/745)
16/03/29 00:39:20 INFO scheduler.TaskSetManager: Finished task 711.0 in stage 0.0 (TID 712) in 1577 ms on ip-172-31-28-255.us-west-2.compute.internal (714/745)
16/03/29 00:39:20 INFO scheduler.TaskSetManager: Finished task 713.0 in stage 0.0 (TID 711) in 1639 ms on ip-172-31-31-119.us-west-2.compute.internal (715/745)
16/03/29 00:39:20 INFO scheduler.TaskSetManager: Finished task 722.0 in stage 0.0 (TID 715) in 1580 ms on ip-172-31-30-28.us-west-2.compute.internal (716/745)
16/03/29 00:39:20 INFO scheduler.TaskSetManager: Finished task 714.0 in stage 0.0 (TID 713) in 1672 ms on ip-172-31-31-119.us-west-2.compute.internal (717/745)
16/03/29 00:39:20 INFO scheduler.TaskSetManager: Finished task 723.0 in stage 0.0 (TID 739) in 977 ms on ip-172-31-29-9.us-west-2.compute.internal (718/745)
16/03/29 00:39:20 INFO scheduler.TaskSetManager: Finished task 737.0 in stage 0.0 (TID 738) in 1003 ms on ip-172-31-30-253.us-west-2.compute.internal (719/745)
16/03/29 00:39:20 INFO scheduler.TaskSetManager: Finished task 718.0 in stage 0.0 (TID 718) in 1522 ms on ip-172-31-25-22.us-west-2.compute.internal (720/745)
16/03/29 00:39:20 INFO scheduler.TaskSetManager: Finished task 734.0 in stage 0.0 (TID 723) in 1501 ms on ip-172-31-23-22.us-west-2.compute.internal (721/745)
16/03/29 00:39:20 INFO scheduler.TaskSetManager: Finished task 709.0 in stage 0.0 (TID 716) in 1660 ms on ip-172-31-19-229.us-west-2.compute.internal (722/745)
16/03/29 00:39:20 INFO scheduler.TaskSetManager: Finished task 739.0 in stage 0.0 (TID 740) in 1634 ms on ip-172-31-30-253.us-west-2.compute.internal (723/745)
16/03/29 00:39:20 INFO scheduler.TaskSetManager: Finished task 716.0 in stage 0.0 (TID 721) in 1664 ms on ip-172-31-23-22.us-west-2.compute.internal (724/745)
16/03/29 00:39:20 INFO scheduler.TaskSetManager: Finished task 712.0 in stage 0.0 (TID 717) in 1624 ms on ip-172-31-19-229.us-west-2.compute.internal (725/745)
16/03/29 00:39:20 INFO scheduler.TaskSetManager: Finished task 735.0 in stage 0.0 (TID 731) in 1291 ms on ip-172-31-30-212.us-west-2.compute.internal (726/745)
16/03/29 00:39:20 INFO scheduler.TaskSetManager: Finished task 728.0 in stage 0.0 (TID 719) in 1614 ms on ip-172-31-17-161.us-west-2.compute.internal (727/745)
16/03/29 00:39:20 INFO scheduler.TaskSetManager: Finished task 726.0 in stage 0.0 (TID 727) in 1559 ms on ip-172-31-27-47.us-west-2.compute.internal (728/745)
16/03/29 00:39:20 INFO scheduler.TaskSetManager: Finished task 719.0 in stage 0.0 (TID 724) in 1593 ms on ip-172-31-27-47.us-west-2.compute.internal (729/745)
16/03/29 00:39:20 INFO scheduler.TaskSetManager: Finished task 725.0 in stage 0.0 (TID 730) in 1364 ms on ip-172-31-29-125.us-west-2.compute.internal (730/745)
16/03/29 00:39:20 INFO scheduler.TaskSetManager: Finished task 738.0 in stage 0.0 (TID 729) in 1444 ms on ip-172-31-25-22.us-west-2.compute.internal (731/745)
16/03/29 00:39:20 INFO scheduler.TaskSetManager: Finished task 724.0 in stage 0.0 (TID 726) in 1611 ms on ip-172-31-19-229.us-west-2.compute.internal (732/745)
16/03/29 00:39:20 INFO scheduler.TaskSetManager: Finished task 720.0 in stage 0.0 (TID 720) in 1696 ms on ip-172-31-17-108.us-west-2.compute.internal (733/745)
16/03/29 00:39:20 INFO scheduler.TaskSetManager: Finished task 725.0 in stage 0.0 (TID 725) in 1698 ms on ip-172-31-30-226.us-west-2.compute.internal (734/745)
16/03/29 00:39:20 INFO scheduler.TaskSetManager: Finished task 732.0 in stage 0.0 (TID 720) in 1331 ms on ip-172-31-19-229.us-west-2.compute.internal (735/745)
16/03/29 00:39:20 INFO scheduler.TaskSetManager: Finished task 730.0 in stage 0.0 (TID 743) in 1091 ms on ip-172-31-28-20.us-west-2.compute.internal (736/745)
16/03/29 00:39:21 INFO scheduler.TaskSetManager: Finished task 729.0 in stage 0.0 (TID 736) in 1402 ms on ip-172-31-29-125.us-west-2.compute.internal (738/745)
16/03/29 00:39:21 INFO scheduler.TaskSetManager: Finished task 741.0 in stage 0.0 (TID 742) in 1233 ms on ip-172-31-29-125.us-west-2.compute.internal (739/745)
16/03/29 00:39:21 INFO scheduler.TaskSetManager: Finished task 731.0 in stage 0.0 (TID 733) in 1501 ms on ip-172-31-17-108.us-west-2.compute.internal (740/745)
16/03/29 00:39:21 INFO scheduler.TaskSetManager: Finished task 740.0 in stage 0.0 (TID 735) in 1512 ms on ip-172-31-17-108.us-west-2.compute.internal (741/745)
16/03/29 00:39:21 INFO scheduler.TaskSetManager: Finished task 733.0 in stage 0.0 (TID 732) in 1573 ms on ip-172-31-17-161.us-west-2.compute.internal (742/745)
16/03/29 00:39:21 INFO scheduler.TaskSetManager: Finished task 736.0 in stage 0.0 (TID 734) in 1555 ms on ip-172-31-17-161.us-west-2.compute.internal (743/745)
16/03/29 00:39:21 INFO scheduler.TaskSetManager: Finished task 743.0 in stage 0.0 (TID 741) in 1425 ms on ip-172-31-17-161.us-west-2.compute.internal (744/745)
16/03/29 00:39:21 INFO scheduler.DAGScheduler: ResultStage 0 (sortByKey at <console>:29) finished in 23.200 s
16/03/29 00:39:21 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks have all completed, from pool
16/03/29 00:39:21 INFO scheduler.TaskSetManager: Job 0 finished: sortByKey at <console>:29, took 23.399483 s
data: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[7] at map at <console>:29

scala>
```

Spark-Master - Xshell 5 (Free for Home/School)

File Edit View Tools Tab Window Help

To add the current session, click on the left arrow button.

Cloud Spark-Master Spark-Master +

```
16/03/29 00:42:45 INFO scheduler.TaskSetManager: Finished task 739.0 in stage 1.0 (TID 1484) in 7609 ms on ip-172-31-30-253.us-west-2.compute.internal (739/745)
16/03/29 00:42:45 INFO scheduler.TaskSetManager: Finished task 739.0 in stage 1.0 (TID 1480) in 9016 ms on ip-172-31-25-32.us-west-2.compute.internal (740/745)
16/03/29 00:42:45 INFO scheduler.TaskSetManager: Finished task 724.0 in stage 1.0 (TID 1477) in 6229 ms on ip-172-31-25-22.us-west-2.compute.internal (741/745)
16/03/29 00:42:45 INFO scheduler.TaskSetManager: Finished task 737.0 in stage 1.0 (TID 1483) in 8657 ms on ip-172-31-30-253.us-west-2.compute.internal (742/745)
16/03/29 00:42:45 INFO scheduler.TaskSetManager: Finished task 748.0 in stage 1.0 (TID 1488) in 5307 ms on ip-172-31-31-119.us-west-2.compute.internal (743/745)
16/03/29 00:42:46 INFO scheduler.TaskSetManager: Finished task 742.0 in stage 1.0 (TID 1489) in 5546 ms on ip-172-31-27-47.us-west-2.compute.internal (744/745)
16/03/29 00:42:46 INFO scheduler.TaskSetManager: Finished task 744.0 in stage 1.0 (TID 1487) in 8107 ms on ip-172-31-19-229.us-west-2.compute.internal (745/745)
16/03/29 00:42:46 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 1.0, whose tasks have all completed, from pool
16/03/29 00:42:46 INFO scheduler.DAGScheduler: ShufflegateStage 1 (map at <console>:29) finished in 118.883 s
16/03/29 00:42:46 INFO scheduler.DAGScheduler: looking for newly runnable stages
16/03/29 00:42:46 INFO scheduler.DAGScheduler: running: Set()
16/03/29 00:42:46 INFO scheduler.DAGScheduler: waiting: Set(ResultStage 2)
16/03/29 00:42:46 INFO scheduler.DAGScheduler: failed: Set()
16/03/29 00:42:47 INFO scheduler.DAGScheduler: Submitting ResultStage 2 (MapPartitionsRDD[8] at saveAsTextFile at <console>:32), which has no missing parents
16/03/29 00:42:47 INFO storage.MemoryStore: Block broadcast_3 stored as values in memory (estimated size 24.8 KB, free 88.0 KB)
16/03/29 00:42:47 INFO storage.MemoryStore: Block broadcast_3_piece0 stored as bytes in memory (estimated size 16.9 KB, free 104.9 KB)
16/03/29 00:42:47 INFO storage.BlockManagerInfo: Added broadcast_3_piece0 in memory on 172.31.19.190:38494 (size: 16.9 KB, free: 511.5 MB)
16/03/29 00:42:47 INFO spark.SparkContext: Created broadcast 3 From broadcast at DAGScheduler.scala:1006
16/03/29 00:42:47 INFO scheduler.DAGScheduler: Submitting 745 missing tasks from ResultStage 2 (MapPartitionsRDD[8] at saveAsTextFile at <console>:32)
16/03/29 00:42:47 INFO scheduler.TaskSchedulerImpl: Adding task set 2.0 with 745 tasks
16/03/29 00:42:47 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 2.0 (TID 1490, ip-172-31-31-119.us-west-2.compute.internal, partition 0,PROCESS_LOCAL, 1894 bytes)
16/03/29 00:42:47 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 2.0 (TID 1491, ip-172-31-17-161.us-west-2.compute.internal, partition 1,PROCESS_LOCAL, 1894 bytes)
16/03/29 00:42:47 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 2.0 (TID 1492, ip-172-31-21-125.us-west-2.compute.internal, partition 2,PROCESS_LOCAL, 1894 bytes)
16/03/29 00:42:47 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 2.0 (TID 1493, ip-172-31-19-229.us-west-2.compute.internal, partition 3,PROCESS_LOCAL, 1894 bytes)
16/03/29 00:42:47 INFO scheduler.TaskSetManager: Starting task 4.0 in stage 2.0 (TID 1494, ip-172-31-16-5.us-west-2.compute.internal, partition 4,PROCESS_LOCAL, 1894 bytes)
16/03/29 00:42:47 INFO scheduler.TaskSetManager: Starting task 5.0 in stage 2.0 (TID 1495, ip-172-31-30-253.us-west-2.compute.internal, partition 5,PROCESS_LOCAL, 1894 bytes)
16/03/29 00:42:47 INFO scheduler.TaskSetManager: Starting task 6.0 in stage 2.0 (TID 1496, ip-172-31-28-255.us-west-2.compute.internal, partition 6,PROCESS_LOCAL, 1894 bytes)
16/03/29 00:42:47 INFO scheduler.TaskSetManager: Starting task 7.0 in stage 2.0 (TID 1497, ip-172-31-27-47.us-west-2.compute.internal, partition 7,PROCESS_LOCAL, 1894 bytes)
16/03/29 00:42:47 INFO scheduler.TaskSetManager: Starting task 8.0 in stage 2.0 (TID 1498, ip-172-31-30-28.us-west-2.compute.internal, partition 8,PROCESS_LOCAL, 1894 bytes)
16/03/29 00:42:47 INFO scheduler.TaskSetManager: Starting task 9.0 in stage 2.0 (TID 1499, ip-172-31-23.us-west-2.compute.internal, partition 9,PROCESS_LOCAL, 1894 bytes)
16/03/29 00:42:47 INFO scheduler.TaskSetManager: Starting task 10.0 in stage 2.0 (TID 1500, ip-172-31-30-212.us-west-2.compute.internal, partition 10,PROCESS_LOCAL, 1894 bytes)
16/03/29 00:42:47 INFO scheduler.TaskSetManager: Starting task 11.0 in stage 2.0 (TID 1501, ip-172-31-27-166.us-west-2.compute.internal, partition 11,PROCESS_LOCAL, 1894 bytes)
16/03/29 00:42:47 INFO scheduler.TaskSetManager: Starting task 12.0 in stage 2.0 (TID 1502, ip-172-31-11-100.us-west-2.compute.internal, partition 12,PROCESS_LOCAL, 1894 bytes)
```

Send text to all sessions within current window

ssh://root@ec2-52-37-249-240.us-west-2.compute.amazonaws.com:22

SSH2 xterm I\* 191x33 33,1 3 sessions ENG 7:44 PM 3/28/2016

EC2 Management Console prog2-v5.pdf Hadoop NameNode ip-17\* WhatsApp Web

Back Forward Home ec2-52-37-249-240.us-west-2.compute.amazonaws.com:50070/dfsnodelist.jsp?whatNodes=LIVE

Apps WhenToWork Facebook Blackboard Learn Illinois Institute of Tech WhatsApp Web Chicago South Shore Sign in Illinois Institut Imported From Firefox Go back to Dr. S homepage

### Live Datanodes : 16

Node	Last Contact	Admin State	Configured Capacity (GB)	Used (GB)	Non DFS Used (GB)	Remaining (GB)	Used (%)	Used (%)	Remaining (%)	Blocks
ip-172-31-16-5	0	In Service	78.73	24.23	8.11	46.38	30.78	30.78	58.91	213
ip-172-31-17-108	1	In Service	78.73	21.55	8.04	49.14	27.37	27.37	62.42	191
ip-172-31-17-161	2	In Service	78.73	21.58	8.04	49.11	27.41	27.41	62.38	195
ip-172-31-19-229	1	In Service	78.73	23.24	7.98	47.5	29.52	29.52	60.34	201
ip-172-31-23-22	1	In Service	78.73	24.77	7.97	45.99	31.46	31.46	58.41	221
ip-172-31-23-231	2	In Service	78.73	24.95	7.88	45.89	31.69	31.69	58.29	225
ip-172-31-25-22	2	In Service	78.73	20.99	7.95	49.79	26.66	26.66	63.24	189
ip-172-31-27-166	0	In Service	78.73	23.09	8.15	47.49	29.33	29.33	60.33	206
ip-172-31-27-47	1	In Service	78.73	25.69	8.19	44.85	32.63	32.63	56.96	226
ip-172-31-28-255	1	In Service	78.73	25.68	7.81	45.23	32.62	32.62	57.45	226
ip-172-31-29-125	2	In Service	78.73	22.35	7.7	48.68	28.39	28.39	61.83	203
ip-172-31-29-9	1	In Service	78.73	23.6	8.03	47.09	29.98	29.98	59.82	213
ip-172-31-30-212	0	In Service	78.73	23.17	8.11	47.45	29.43	29.43	60.27	204
ip-172-31-30-253	0	In Service	78.73	23.79	8.16	46.78	30.22	30.22	59.42	204
ip-172-31-30-28	1	In Service	78.73	20.71	7.87	50.15	26.31	26.31	63.7	184
ip-172-31-31-119	2	In Service	78.73	23.94	7.98	46.81	30.41	30.41	59.45	217

This is Apache Hadoop release 1.0.4

prog2-v5.pdf Show all downloads...

Windows File Explorer Google Chrome Microsoft Edge Microsoft Word Microsoft Excel

ENG 7:44 PM 3/28/2016

```

16/03/29 00:48:13 INFO scheduler.TaskSetManager: Finished task [+] - 2.compute.internal (701/745)
16/03/29 00:48:14 INFO scheduler.TaskSetManager: Finished task 699.0 in stage 2.0 (TID 2189) in 33555 ms on ip-172-31-23-231.us-west-2.compute.internal (702/745)
16/03/29 00:48:15 INFO scheduler.TaskSetManager: Finished task 721.0 in stage 2.0 (TID 2211) in 14433 ms on ip-172-31-17-161.us-west-2.compute.internal (704/745)
16/03/29 00:48:15 INFO scheduler.TaskSetManager: Finished task 703.0 in stage 2.0 (TID 2193) in 32829 ms on ip-172-31-15-5.us-west-2.compute.internal (705/745)
16/03/29 00:48:15 INFO scheduler.TaskSetManager: Finished task 719.0 in stage 2.0 (TID 2209) in 18656 ms on ip-172-31-27-47.us-west-2.compute.internal (706/745)
16/03/29 00:48:16 INFO scheduler.TaskSetManager: Finished task 687.0 in stage 2.0 (TID 2177) in 37135 ms on ip-172-31-31-119.us-west-2.compute.internal (707/745)
16/03/29 00:48:16 INFO scheduler.TaskSetManager: Finished task 696.0 in stage 2.0 (TID 2186) in 35340 ms on ip-172-31-25-22.us-west-2.compute.internal (708/745)
16/03/29 00:48:16 INFO scheduler.TaskSetManager: Finished task 698.0 in stage 2.0 (TID 2188) in 35229 ms on ip-172-31-25-22.us-west-2.compute.internal (709/745)
16/03/29 00:48:16 INFO scheduler.TaskSetManager: Finished task 706.0 in stage 2.0 (TID 2196) in 31291 ms on ip-172-31-27-166.us-west-2.compute.internal (710/745)
16/03/29 00:48:16 INFO scheduler.TaskSetManager: Finished task 655.0 in stage 2.0 (TID 2145) in 45296 ms on ip-172-31-29-9.us-west-2.compute.internal (711/745)
16/03/29 00:48:18 INFO scheduler.TaskSetManager: Finished task 673.0 in stage 2.0 (TID 2165) in 40996 ms on ip-172-31-17-166.us-west-2.compute.internal (712/745)
16/03/29 00:48:18 INFO scheduler.TaskSetManager: Finished task 720.0 in stage 2.0 (TID 2210) in 18188 ms on ip-172-31-20-20.us-west-2.compute.internal (714/745)
16/03/29 00:48:18 INFO scheduler.TaskSetManager: Finished task 718.0 in stage 2.0 (TID 2209) in 23900 ms on ip-172-31-19-229.us-west-2.compute.internal (715/745)
16/03/29 00:48:18 INFO scheduler.TaskSetManager: Finished task 712.0 in stage 2.0 (TID 2202) in 29016 ms on ip-172-31-30-28.us-west-2.compute.internal (716/745)
16/03/29 00:48:18 INFO scheduler.TaskSetManager: Finished task 674.0 in stage 2.0 (TID 2164) in 41090 ms on ip-172-31-31-119.us-west-2.compute.internal (717/745)
16/03/29 00:48:18 INFO scheduler.TaskSetManager: Finished task 717.0 in stage 2.0 (TID 2207) in 25828 ms on ip-172-31-30-28.us-west-2.compute.internal (718/745)
16/03/29 00:48:18 INFO scheduler.TaskSetManager: Finished task 664.0 in stage 2.0 (TID 2154) in 43312 ms on ip-172-31-31-119.us-west-2.compute.internal (719/745)
16/03/29 00:48:19 INFO scheduler.TaskSetManager: Finished task 684.0 in stage 2.0 (TID 2174) in 40476 ms on ip-172-31-17-108.us-west-2.compute.internal (720/745)
16/03/29 00:48:19 INFO scheduler.TaskSetManager: Finished task 693.0 in stage 2.0 (TID 2183) in 39509 ms on ip-172-31-25-22.us-west-2.compute.internal (721/745)
16/03/29 00:48:19 INFO scheduler.TaskSetManager: Finished task 715.0 in stage 2.0 (TID 2205) in 27532 ms on ip-172-31-30-28.us-west-2.compute.internal (722/745)
16/03/29 00:48:21 INFO scheduler.TaskSetManager: Finished task 709.0 in stage 2.0 (TID 2198) in 33607 ms on ip-172-31-30-28.us-west-2.compute.internal (723/745)
16/03/29 00:48:23 INFO scheduler.TaskSetManager: Finished task 728.0 in stage 2.0 (TID 2202) in 18660 ms on ip-172-31-17-161.us-west-2.compute.internal (724/745)
16/03/29 00:48:24 INFO scheduler.TaskSetManager: Finished task 724.0 in stage 2.0 (TID 2214) in 21273 ms on ip-172-31-19-229.us-west-2.compute.internal (725/745)
16/03/29 00:48:24 INFO scheduler.TaskSetManager: Finished task 727.0 in stage 2.0 (TID 2217) in 20073 ms on ip-172-31-30-212.us-west-2.compute.internal (726/745)
16/03/29 00:48:27 INFO scheduler.TaskSetManager: Finished task 733.0 in stage 2.0 (TID 2223) in 19847 ms on ip-172-31-17-161.us-west-2.compute.internal (727/745)
16/03/29 00:48:27 INFO scheduler.TaskSetManager: Finished task 734.0 in stage 2.0 (TID 2224) in 19905 ms on ip-172-31-17-108.us-west-2.compute.internal (728/745)
16/03/29 00:48:28 INFO scheduler.TaskSetManager: Finished task 731.0 in stage 2.0 (TID 2221) in 22983 ms on ip-172-31-27-166.us-west-2.compute.internal (729/745)
16/03/29 00:48:28 INFO scheduler.TaskSetManager: Finished task 738.0 in stage 2.0 (TID 2228) in 20504 ms on ip-172-31-28-255.us-west-2.compute.internal (730/745)
16/03/29 00:48:28 INFO scheduler.TaskSetManager: Finished task 739.0 in stage 2.0 (TID 2229) in 20803 ms on ip-172-31-16-5.us-west-2.compute.internal (731/745)
16/03/29 00:48:29 INFO scheduler.TaskSetManager: Finished task 746.0 in stage 2.0 (TID 2230) in 21172 ms on ip-172-31-29-9.us-west-2.compute.internal (732/745)
16/03/29 00:48:29 INFO scheduler.TaskSetManager: Finished task 729.0 in stage 2.0 (TID 2219) in 24777 ms on ip-172-31-30-212.us-west-2.compute.internal (733/745)
16/03/29 00:48:29 INFO scheduler.TaskSetManager: Finished task 722.0 in stage 2.0 (TID 2212) in 27701 ms on ip-172-31-27-47.us-west-2.compute.internal (734/745)
16/03/29 00:48:29 INFO scheduler.TaskSetManager: Finished task 730.0 in stage 2.0 (TID 2226) in 24636 ms on ip-172-31-30-253.us-west-2.compute.internal (735/745)
16/03/29 00:48:30 INFO scheduler.TaskSetManager: Finished task 743.0 in stage 2.0 (TID 2233) in 20694 ms on ip-172-31-30-253.us-west-2.compute.internal (736/745)
16/03/29 00:48:30 INFO scheduler.TaskSetManager: Finished task 735.0 in stage 2.0 (TID 2225) in 23207 ms on ip-172-31-23-22.us-west-2.compute.internal (737/745)
16/03/29 00:48:30 INFO scheduler.TaskSetManager: Finished task 744.0 in stage 2.0 (TID 2234) in 21008 ms on ip-172-31-29-9.us-west-2.compute.internal (738/745)
16/03/29 00:48:31 INFO scheduler.TaskSetManager: Finished task 737.0 in stage 2.0 (TID 2227) in 23404 ms on ip-172-31-23-22.us-west-2.compute.internal (739/745)
16/03/29 00:48:31 INFO scheduler.TaskSetManager: Finished task 736.0 in stage 2.0 (TID 2226) in 24234 ms on ip-172-31-19-229.us-west-2.compute.internal (740/745)
16/03/29 00:48:32 INFO scheduler.TaskSetManager: Finished task 725.0 in stage 2.0 (TID 2215) in 28824 ms on ip-172-31-19-229.us-west-2.compute.internal (741/745)
16/03/29 00:48:32 INFO scheduler.TaskSetManager: Finished task 732.0 in stage 2.0 (TID 2222) in 28134 ms on ip-172-31-29-125.us-west-2.compute.internal (742/745)
16/03/29 00:48:36 INFO scheduler.TaskSetManager: Finished task 726.0 in stage 2.0 (TID 2216) in 32913 ms on ip-172-31-29-125.us-west-2.compute.internal (743/745)
16/03/29 00:48:36 INFO scheduler.TaskSetManager: Finished task 741.0 in stage 2.0 (TID 2231) in 27867 ms on ip-172-31-29-125.us-west-2.compute.internal (744/745)
16/03/29 00:48:37 INFO scheduler.TaskSetManager: Finished task 742.0 in stage 2.0 (TID 2232) in 28043 ms on ip-172-31-29-125.us-west-2.compute.internal (745/745)
16/03/29 00:48:37 INFO scheduler.DAGScheduler: ResultStage 2 (saveAsTextFile at <console>:32) finished in 350.352 s
16/03/29 00:48:37 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 2.0, whose tasks have all completed, from pool
16/03/29 00:48:37 INFO scheduler.DAGScheduler: Job 1 finished: saveAsTextFile at <console>:32, took 469.542174 s

```

scala>

Live Datanodes : 16

Node	Last Contact	Admin State	Configured Capacity (GB)	Used (GB)	Non DFS Used (GB)	Remaining (GB)	Used (%)	Used (%)	Remaining (%)	Blocks
ip-172-31-16-5	2	In Service	78.73	34.5	8.09	36.14	43.82	<div style="width: 43.82%; background-color: #58D68D;"></div>	45.91	343
ip-172-31-17-108	2	In Service	78.73	33.81	7.85	37.07	42.95	<div style="width: 42.95%; background-color: #58D68D;"></div>	47.08	330
ip-172-31-17-161	1	In Service	78.73	31.32	8.02	39.39	39.79	<div style="width: 39.79%; background-color: #58D68D;"></div>	50.03	319
ip-172-31-19-229	0	In Service	78.73	36.41	7.93	34.39	46.25	<div style="width: 46.25%; background-color: #58D68D;"></div>	43.68	356
ip-172-31-23-22	0	In Service	78.73	38.77	7.86	32.1	49.25	<div style="width: 49.25%; background-color: #58D68D;"></div>	40.77	370
ip-172-31-23-231	0	In Service	78.73	38.72	7.75	32.25	49.18	<div style="width: 49.18%; background-color: #58D68D;"></div>	40.97	379
ip-172-31-25-22	0	In Service	78.73	32.69	7.86	38.18	41.52	<div style="width: 41.52%; background-color: #58D68D;"></div>	48.49	325
ip-172-31-27-166	2	In Service	78.73	33.32	8	37.4	42.33	<div style="width: 42.33%; background-color: #58D68D;"></div>	47.51	329
ip-172-31-27-247	0	In Service	78.73	38.58	7.95	32.2	49	<div style="width: 49%; background-color: #58D68D;"></div>	40.9	375
ip-172-31-28-255	1	In Service	78.73	37.55	7.73	33.45	47.69	<div style="width: 47.69%; background-color: #58D68D;"></div>	42.48	367
ip-172-31-29-125	1	In Service	78.73	35.2	7.58	35.95	44.71	<div style="width: 44.71%; background-color: #58D68D;"></div>	45.67	346
ip-172-31-29-9	2	In Service	78.73	35.76	7.82	35.15	45.42	<div style="width: 45.42%; background-color: #58D68D;"></div>	44.65	351
ip-172-31-30-212	2	In Service	78.73	35.07	8.03	35.63	44.55	<div style="width: 44.55%; background-color: #58D68D;"></div>	45.25	347
ip-172-31-30-253	2	In Service	78.73	34.83	8.03	35.87	44.23	<div style="width: 44.23%; background-color: #58D68D;"></div>	45.57	339
ip-172-31-30-28	0	In Service	78.73	31.33	7.79	39.62	39.79	<div style="width: 39.79%; background-color: #58D68D;"></div>	50.32	308
ip-172-31-31-119	0	In Service	78.73	36.98	7.85	33.9	46.97	<div style="width: 46.97%; background-color: #58D68D;"></div>	43.06	363

This is Apache Hadoop release 1.0.4



*Extra Credit:*

*Shared Memory Sort 1GB on d2.xlarge*

```
Connecting to www.ordinal.com (www.ordinal.com)|128.121.96.5| [+] - ⌂
HTTP request sent, awaiting response... 200 OK
Length: unspecified [application/octet-stream]
Saving to: "gensort-linux-1.5.tar.gz"

[  =>
2016-03-28 21:11:42 (437 KB/s) - 'gensort-linux-1.5.tar.gz' saved [226665]

ubuntu@ip-172-31-10-108:~$ sudo tar -xvzf gensort-linux-1.5.tar.gz
gpl-2.0.txt
32
32/gensort
32/gensort
64
64/gensort
64/gensort
ubuntu@ip-172-31-10-108:~$ sudo chown -R ubuntu 64
ubuntu@ip-172-31-10-108:~$ cd 64/
ubuntu@ip-172-31-10-108:~/64$ time ./gensort -a 10000000 Newl

real    1m34.336s
user    0m1.904s
sys     0m0.242s
ubuntu@ip-172-31-10-108:~/64$ ls -lh
total 954M
-rwxrwxr-x 1 ubuntu    500 138K Mar 17 2013 gensort
-rwxrwxr-x 1 ubuntu    954M Mar 28 21:13 Newl
-rwxrwxr-x 1 ubuntu    500 132K Mar 17 2013 valsrt
ubuntu@ip-172-31-10-108:~/64$ cd ..
ubuntu@ip-172-31-10-108:~/64$ java SharedMemorySort
Split Files: 19secs.
Quick Sort: 20 secs.
Merge Sort: 48 secs.
ubuntu@ip-172-31-10-108:~$ cd 64
ubuntu@ip-172-31-10-108:~/64$ time ./valsrt Output
Records: 10000000
Checksum: 4c499aae14ae2b
Duplicate keys: 0
SUCCESS - all records are in order

real    0m9.683s
user    0m1.573s
sys     0m0.014s
ubuntu@ip-172-31-10-108:~/64$ cd ..
ubuntu@ip-172-31-10-108:~$ vi SharedMemorySort.java
ubuntu@ip-172-31-10-108:~$ java SharedMemorySort.java
Error: Could not find or load main class SharedMemorySort.java
ubuntu@ip-172-31-10-108:~$ javac SharedMemorySort.java
ubuntu@ip-172-31-10-108:~$
```

## Hadoop Sort 1GB on d2.large

```

102400000, 214990848, 512000000, 550731776, 644972544 [+] - 8

Allocating group tables: done
Writing inode tables: done
Writing superblocks and filesystem accounting information: done

ubuntugip-172-31-10-108:~$ sudo mkdir -p /mnt/raid
ubuntugip-172-31-10-108:~$ sudo mount LABEL=MY_RAID /mnt/raid
ubuntugip-172-31-10-108:~$ df -h
Filesystem      Size  Used Avail Use% Mounted on
/dev           15G   12K  15G  1% /dev
tmpfs       3.0G  3.0G  0  100% /run
/dev/xvda1    7.0G  5.8G  1.0G  83% /
none          4.0K   0  4.0K  0% /sys/fs/cgroup
none          5.0M   0  5.0M  0% /dev/lock
none          15G   0  15G  0% /run/shm
none         100M   0  100M  0% /run/user
/dev/md0      5.5T  58M  5.2T  1% /mnt/raid
ubuntugip-172-31-10-108:~$ hadoop fs -put /home/ubuntu/64/New1 /Input
16/03/28 21:44:42 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
ubuntugip-172-31-10-108:~$ hadoop fs -ls /
16/03/28 21:44:59 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 1 items
-rw-r--r-- 1 ubuntu supergroup 1000000000 2016-03-28 21:44:42 /Input
ubuntugip-172-31-10-108:~$ hadoop jar HadoopSort.jar HadoopSort /Input /Output
16/03/28 21:45:15 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
16/03/28 21:45:15 INFO client.RMProxy: Connecting to ResourceManager at ec2-52-32-216-227.us-west-2.compute.amazonaws.com/172.31.10.108:8032
16/03/28 21:45:16 INFO client.RMProxy: Connecting to ResourceManager at ec2-52-32-216-227.us-west-2.compute.amazonaws.com/172.31.10.108:8032
16/03/28 21:45:16 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
16/03/28 21:45:17 INFO mapred.FileInputFormat: Total input paths to process : 1
16/03/28 21:45:18 INFO mapreduce.JobSubmitter: number of splits:8
16/03/28 21:45:18 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1459201455791_0001
16/03/28 21:45:18 INFO impl.YarnClientImpl: Submitted application application_1459201455791_0001
16/03/28 21:45:18 INFO mapreduce.Job: The url to track the job: http://ec2-52-32-216-227.us-west-2.compute.amazonaws.com:8088/proxy/application_1459201455791_0001/
16/03/28 21:45:18 INFO mapreduce.Job: Running job: job_1459201455791_0001
16/03/28 21:45:25 INFO mapreduce.Job: Job job_1459201455791_0001 running in uber mode : false
16/03/28 21:45:25 INFO mapreduce.Job: map 0% reduce 0%
16/03/28 21:45:41 INFO mapreduce.Job: map 17% reduce 0%
16/03/28 21:45:42 INFO mapreduce.Job: map 34% reduce 0%
16/03/28 21:45:45 INFO mapreduce.Job: map 35% reduce 0%
16/03/28 21:45:47 INFO mapreduce.Job: map 42% reduce 0%
16/03/28 21:45:48 INFO mapreduce.Job: map 47% reduce 0%
16/03/28 21:45:49 INFO mapreduce.Job: map 48% reduce 0%
16/03/28 21:45:50 INFO mapreduce.Job: map 50% reduce 0%
16/03/28 21:45:52 INFO mapreduce.Job: map 58% reduce 0%
16/03/28 21:45:53 INFO mapreduce.Job: map 64% reduce 0%
16/03/28 21:45:54 INFO mapreduce.Job: map 71% reduce 0%
16/03/28 21:45:55 INFO mapreduce.Job: map 75% reduce 0%
16/03/28 21:45:55 INFO mapreduce.Job: map 75% reduce 0%


[+] - 8

HDFS: Number of read operations=27
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
Job Counters
  Killed map tasks=1
  Launched map tasks=8
  Launched reduce tasks=1
  Data-local map tasks=8
  Total time spent by all maps in occupied slots (ms)=173484
  Total time spent by all reduces in occupied slots (ms)=20058
  Total time spent by all map tasks (ms)=173484
  Total time spent by all reduce tasks (ms)=20058
  Total vcore-milliseconds taken by all map tasks=173484
  Total vcore-milliseconds taken by all reduce tasks=20058
  Total megabyte-milliseconds taken by all map tasks=177647616
  Total megabyte-milliseconds taken by all reduce tasks=20539392
Map-Reduce Framework
  Map input records=100000000
  Map output records=100000000
  Map output bytes=1000000000
  Map output materialized bytes=1020000048
  Input split bytes=952
  Combine input records=0
  Combine output records=0
  Reduce input groups=100000000
  Reduce shuffle bytes=1020000048
  Reduce input records=100000000
  Reduce output records=100000000
  Spilled Records=29995241
  Shuffled Maps =8
  Failed Shuffles=0
  Merged Map outputs=8
  GC time elapsed (ms)=5302
  CPU time spent (ms)=96770
  Physical memory (bytes) snapshot=2362667008
  Virtual memory (bytes) snapshot=17202601984
  Total committed heap usage (bytes)=1800404992
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=1000028672
File Output Format Counters
  Bytes Written=1000000000
Time: 61 secs
ubuntugip-172-31-10-108:~$ 

```

## Spark Sort 1GB on d2.xlarge

