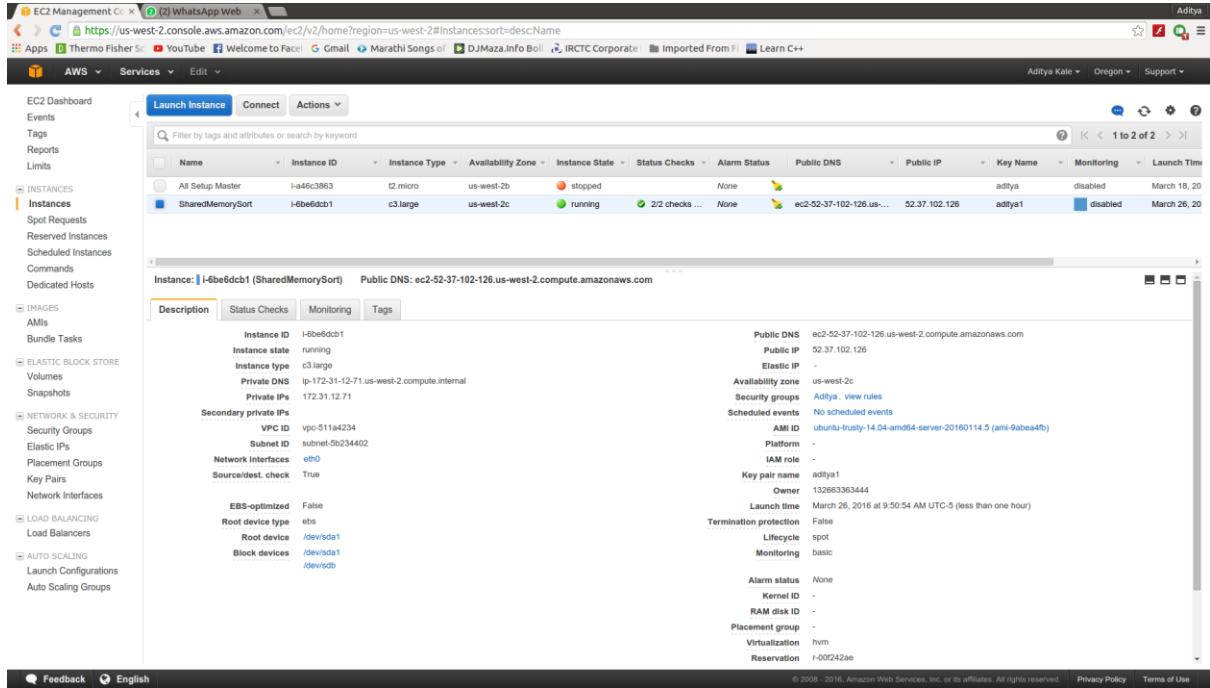


Cloud Computing

Programming Assignment – 2

- Execution screenshots for Shared Memory Sort 10GB File.



- No Of Parallel Thread – 1
 - Time Taken = 3746 seconds

- No Of Parallel Thread – 2
 - Time Taken = 2741 seconds

- No Of Parallel Thread – 4
 - Time Taken = 2323

```
ubuntu@ip-172-31-12-71:/aditya/SharedMemorySort/TeraSort$ ant
Buildfile: /aditya/SharedMemorySort/TeraSort/build.xml

init:
[delete] Deleting directory /aditya/SharedMemorySort/TeraSort/bin
[mkdir] Created dir: /aditya/SharedMemorySort/TeraSort/bin
[delete] Deleting directory /aditya/SharedMemorySort/TeraSort/destination
[mkdir] Created dir: /aditya/SharedMemorySort/TeraSort/destination

compileTeraSort:
[mkdir] Created dir: /aditya/SharedMemorySort/TeraSort/build/classes
[javac] /aditya/SharedMemorySort/TeraSort/build.xml:32: warning: 'includeant
runtime' was not set, defaulting to build.sysclasspath=last; set to false for re
peatable builds
[javac] Compiling 3 source files to /aditya/SharedMemorySort/TeraSort/build/
classes

RunSortjar:
[mkdir] Created dir: /aditya/SharedMemorySort/TeraSort/build/jar
[jar] Building jar: /aditya/SharedMemorySort/TeraSort/build/jar/SharedMem
orySort.jar

RunSorting:
[java] Mar 26, 2016 3:04:22 PM edu.terasort.TeraSort divideInChunks
[java] INFO: Chunks creation started.
[java] Mar 26, 2016 3:09:00 PM edu.terasort.TeraSort divideInChunks
[java] INFO: Total time taken divide and sort a file is : 277.924 Seconds
[java] Mar 26, 2016 3:09:00 PM edu.terasort.TeraSort divideInChunks
[java] INFO: Finished all creation of chunks
[java] Mar 26, 2016 3:09:00 PM edu.terasort.TeraSort mergeFromChunks
[java] INFO: Merging started.
[java] Mar 26, 2016 3:47:44 PM edu.terasort.TeraSort mergeFromChunks
[java] INFO: Total time taken to sort a /aditya/SharedMemorySort/gensort-linux-
1.5/64/10GBFile file is : 2323.751 Seconds

main:

BUILD SUCCESSFUL
Total time: 43 minutes 23 seconds
ubuntu@ip-172-31-12-71:/aditya/SharedMemorySort/TeraSort$ █
```

- No Of Parallel Thread – 8
 - Time Taken = 2106 seconds

Validation Output:

```

ubuntu@ip-172-31-12-71:/aditya/SharedMemorySort/TeraSort/destination$ cd /aditya/SharedMemorySort/gensort-linux-1.5/64/
ubuntu@ip-172-31-12-71:/aditya/SharedMemorySort/gensort-linux-1.5/64$ ./valsor
tya/SharedMemorySort/TeraSort/destination/File_3.69E7111111111
rRecords: 10000000
Checksum: 2faf0ab746e89a8
Duplicate keys: 0
SUCCESS - all records are in order
ubuntu@ip-172-31-12-71:/aditya/SharedMemorySort/gensort-linux-1.5/64$ head /adit
haredMemorySort/TeraSort/destination/File_3.69E7111111111
    "0!uve 0000000000000000000000000000001228D4 7777888000022224444DDDDDDDEEEEEE00
00CCCC7777DDDD
    PMd32= 0000000000000000000000000000003440CC1 FFFFEEEE6666CCCCBBBB99993335555DD
DD77778886666
    ^3C0], 000000000000000000000000000000158C5C5 5555AAAA9999EEEE888822229999CCCCDD
66555544442222
    !&S3/] ] 0000000000000000000000000000002145D78 8888BBBBDDDD1111CCCC55556666BBBB11
EEDDDD22229999
    ! ,=U#,9 00000000000000000000000000000019072E3 33332222FFFFBBBB0000FFFFAAAA666655
33DDDD3333CCCC
    !0f[ITd 0000000000000000000000000000003CAAB4B 9999FFFF55553337777CCCC4444BBBB77
EBBBBBDDDD4444
    !f6Suy2 0000000000000000000000000000003ABFD84 EEEE5555556666AAAA5555BBBBDDDD00
1166660000DDDD
    #%"NIpq. 0000000000000000000000000000003B36FB9 1111000033334444111166666666AAAAAA
001111CCCCEEEE
    #'^cl`~ 0000000000000000000000000000002EDC5C8 8888AAAA11114444FFFF77773333EEEE44
00FFFF99999999
    $"-'Q)] 0000000000000000000000000000005F1265D CCCC6666EEEE22220000DDDAAAA888866
BB00006666AAAA
ubuntu@ip-172-31-12-71:/aditya/SharedMemorySort/gensort-linux-1.5/64$ tail /adit
haredMemorySort/TeraSort/destination/File_3.69E7111111111
~~~uq2k#=U 0000000000000000000000000000002C06745 99991111DDDD222211110000FFFFEEEEFF
337777CCCC2222
~~~v/0&Qnm 0000000000000000000000000000004709701 CCCC88883333FFFF00000000000999911
FF777744446666
~~~yK0l:gE 0000000000000000000000000000002048B4F CCCC11114444888822226666BBBB888855
77EEEEBBBB0000
~~~yK^H.il 000000000000000000000000000000463D004 44440000FFFF3333999944447777DDDDFF
AA11118888DDDD
~~~yL;C'XE 0000000000000000000000000000005B0D211 2222EEEE3333000022221111CCCCFFFF55
774444BBBB6666

```

- Execution screenshots for Hadoop Map Reduce Sort 10GB File.
 - Time Taken = 1140 seconds
= 19 minutes

```

ubuntu@ip-172-31-14-142:~/hadoop-2.7.2/sbin$ cd ..
ubuntu@ip-172-31-14-142:~/hadoop-2.7.2$ bin/hadoop fs -mkdir /hdfs aditya
ubuntu@ip-172-31-14-142:~/hadoop-2.7.2$ mv /home/ubuntu/MapReduceSort/gensort-linux-1.5/ /aditya/
ubuntu@ip-172-31-14-142:~/hadoop-2.7.2$ cd /aditya/gensort-linux-1.5/64/
ubuntu@ip-172-31-14-142:/aditya/gensort-linux-1.5/64$ ./gensort -a 100000000 10GBfile
ubuntu@ip-172-31-14-142:/aditya/gensort-linux-1.5/64$ cd /home/ubuntu/hadoop-2.7.2/
ubuntu@ip-172-31-14-142:~/hadoop-2.7.2$ bin/hadoop dfs -copyFromLocal /aditya/gensort-linux-1.5/64/10GBfile /hdfs_aditya
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

ubuntu@ip-172-31-14-142:~/hadoop-2.7.2$ bin/hadoop dfs -ls /hdfs aditya
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

Found 1 items
-rw-r--r-- 1 ubuntu supergroup 10000000000 2016-03-27 19:16 /hdfs/aditya/10GBfile
ubuntu@ip-172-31-14-142:~/hadoop-2.7.2$ bin/hadoop jar /home/ubuntu/MapReduceSort/MapReduceSort/build/jar/MapReduceTeraSort.jar /hdfs_aditya/10GBfile /hdfs_aditya/output
Mar 27, 2016 7:17:54 PM edu.sort.TeraSort main
INFO: Total Time for Execution of MapReduce is :1 Seconds
16/03/27 19:17:55 INFO client.RMProxy: Connecting to ResourceManager at ec2-52-26-45-247.us-west-2.compute.amazonaws.com/172.31.14.142:9050
16/03/27 19:17:55 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with T
oolRunner to remedy this.
16/03/27 19:17:56 INFO input.FileInputFormat: Total input paths to process : 1
16/03/27 19:17:56 INFO mapreduce.JobSubmitter: number of splits:5
16/03/27 19:17:56 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1459105690908_0001
16/03/27 19:17:57 INFO impl.YarnClientImpl: Submitted application application_1459105690908_0001
16/03/27 19:17:57 INFO mapreduce.Job: The url to track the job: http://ec2-52-26-45-247.us-west-2.compute.amazonaws.com:9006/proxy/application_1459105690908_0001/
16/03/27 19:17:57 INFO mapreduce.Job: Job: Running job: job_1459105690908_0001
16/03/27 19:18:08 INFO mapreduce.Job: Job job_1459105690908_0001 running in uber mode : false
16/03/27 19:18:08 INFO mapreduce.Job: map 0% reduce 0%

```

```

16/03/27 19:34:14 INFO mapreduce.Job: map 100% reduce 37%
16/03/27 19:34:17 INFO mapreduce.Job: map 100% reduce 64%
16/03/27 19:34:20 INFO mapreduce.Job: map 100% reduce 67%
16/03/27 19:34:26 INFO mapreduce.Job: map 100% reduce 68%
16/03/27 19:34:32 INFO mapreduce.Job: map 100% reduce 69%
16/03/27 19:34:35 INFO mapreduce.Job: map 100% reduce 70%
16/03/27 19:34:41 INFO mapreduce.Job: map 100% reduce 71%
16/03/27 19:34:41 INFO mapreduce.Job: map 100% reduce 72%
16/03/27 19:34:50 INFO mapreduce.Job: map 100% reduce 73%
16/03/27 19:34:50 INFO mapreduce.Job: map 100% reduce 74%
16/03/27 19:34:54 INFO mapreduce.Job: map 100% reduce 75%
16/03/27 19:35:05 INFO mapreduce.Job: map 100% reduce 76%
16/03/27 19:35:08 INFO mapreduce.Job: map 100% reduce 77%
16/03/27 19:35:14 INFO mapreduce.Job: map 100% reduce 78%
16/03/27 19:35:20 INFO mapreduce.Job: map 100% reduce 79%
16/03/27 19:35:23 INFO mapreduce.Job: map 100% reduce 80%
16/03/27 19:35:29 INFO mapreduce.Job: map 100% reduce 81%
16/03/27 19:35:35 INFO mapreduce.Job: map 100% reduce 82%
16/03/27 19:35:41 INFO mapreduce.Job: map 100% reduce 83%
16/03/27 19:35:41 INFO mapreduce.Job: map 100% reduce 84%
16/03/27 19:35:50 INFO mapreduce.Job: map 100% reduce 85%
16/03/27 19:35:50 INFO mapreduce.Job: map 100% reduce 86%
16/03/27 19:35:59 INFO mapreduce.Job: map 100% reduce 87%
16/03/27 19:36:05 INFO mapreduce.Job: map 100% reduce 88%
16/03/27 19:36:11 INFO mapreduce.Job: map 100% reduce 89%
16/03/27 19:36:14 INFO mapreduce.Job: map 100% reduce 90%
16/03/27 19:36:20 INFO mapreduce.Job: map 100% reduce 91%
16/03/27 19:36:26 INFO mapreduce.Job: map 100% reduce 92%
16/03/27 19:36:29 INFO mapreduce.Job: map 100% reduce 93%
16/03/27 19:36:35 INFO mapreduce.Job: map 100% reduce 94%
16/03/27 19:36:41 INFO mapreduce.Job: map 100% reduce 95%
16/03/27 19:36:44 INFO mapreduce.Job: map 100% reduce 96%
16/03/27 19:36:50 INFO mapreduce.Job: map 100% reduce 97%
16/03/27 19:36:53 INFO mapreduce.Job: map 100% reduce 98%
16/03/27 19:36:59 INFO mapreduce.Job: map 100% reduce 99%
16/03/27 19:37:04 INFO mapreduce.Job: map 100% reduce 100%
16/03/27 19:37:04 INFO mapreduce.Job: Job job_1459105690908_0001 completed successfully
16/03/27 19:37:07 INFO mapreduce.Job: Counters: 50
File System Counters
  FILE: Number of bytes read=30256951212
  FILE: Number of bytes written=40465906524
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=10000314054

```

```

Killed map tasks=1
Launched map tasks=76
Launched reduce tasks=1
Data-local map tasks=76
Total time spent by all maps in occupied slots (ms)=4606839
Total time spent by all reduces in occupied slots (ms)=940401
Total time spent by all map tasks (ms)=4606839
Total time spent by all reduce tasks (ms)=940401
Total vcore-milliseconds taken by all map tasks=4606839
Total vcore-milliseconds taken by all reduce tasks=940401
Total megabyte-milliseconds taken by all map tasks=4717403136
Total megabyte-milliseconds taken by all reduce tasks=962970624
Map-Reduce Framework
  Map input records=100000000
  Map output records=100000000
  Map output bytes=10000000000
  Map output materialized bytes=10200000450
  Input split bytes=10950
  Combine input records=100000000
  Combine output records=100000000
  Reduce input groups=100000000
  Reduce shuffle bytes=10200000450
  Reduce input records=100000000
  Reduce output records=100000000
  Spilled Records=396636763
  Shuffled Maps =75
  Failed Shuffles=0
  Merged Map outputs=75
  GC time elapsed (ms)=74122
  CPU time spent (ms)=1308480
  Physical memory (bytes) snapshot=19097853952
  Virtual memory (bytes) snapshot=144786001920
  Total committed heap usage (bytes)=15596519424
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=10000303104
File Output Format Counters
  Bytes Written=100000000000
ubuntu@ip-172-31-14-142:~/hadoop-2.7.2$ 

```

FINISHED Applications

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved	Active Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes
1	0	0	1	0	0 B	0 B	0 B	8	0	0	1	0	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation
Capacity Scheduler	[MEMORY]	<memory:1024, vCores:1>	<memory:8192, vCores:8>

Show 20 • entries

ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Progress	Tracking UI	Blacklisted Nodes
application_1459105690908_0001	ubuntu	Tera Sort	MAPREDUCE	default	Sun Mar 27, 14:17:56 -0500 2016	Sun Mar 27, 14:37:05 -0500 2016	FINISHED	SUCCEEDED	N/A	History	N/A

Showing 1 to 1 of 1 entries

First Previous 1 Next Last

hadoopxmlfiles.zip Show all downloads

Overview 'ec2-52-26-45-247.us-west-2.compute.amazonaws.com:9000' (active)

Started:	Sun Mar 27 19:07:48 UTC 2016
Version:	2.7.2, r1b16504fe8a74265c792ce23f546c64604acf0e41
Compiled:	2016-01-26T00:08Z by Jenkins from (detached from b165c4f)
Cluster ID:	CID-9f4ba6db-1bd8-4121-a7fb-d4a636709504
Block Pool ID:	BP-1290618776-172.31.14.142-1459105623691

Summary

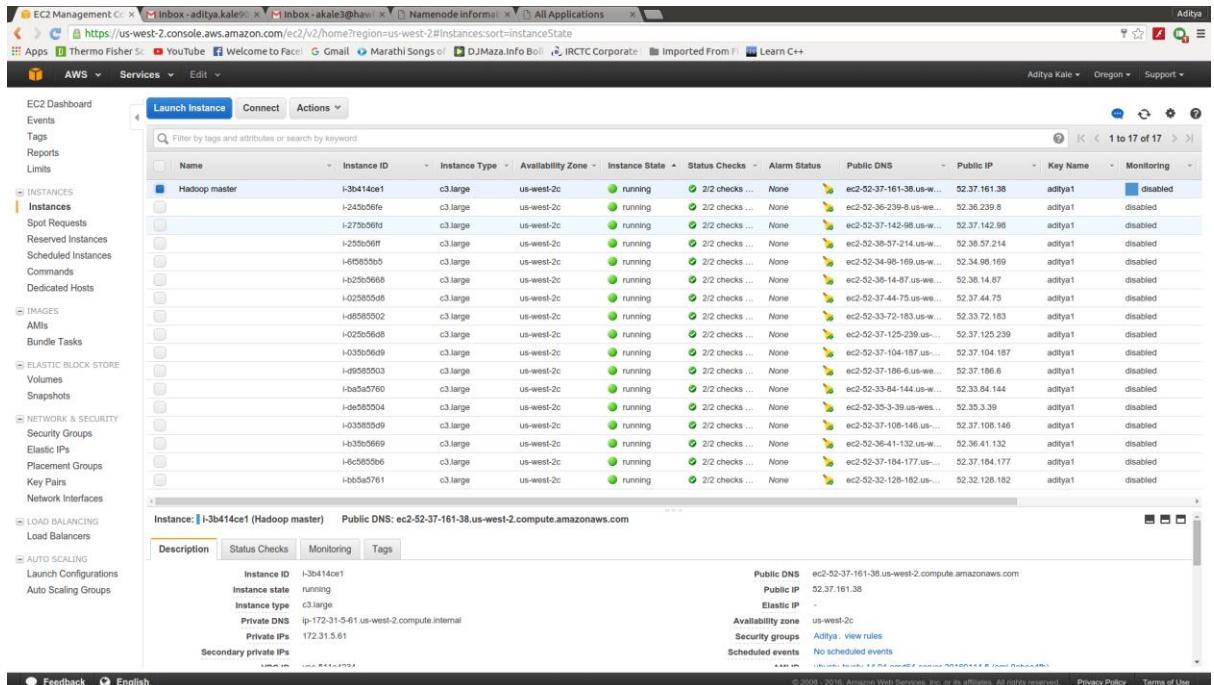
Security is off.
Safemode is off.
21 files and directories, 81 blocks = 102 total filesystem object(s).
Heap Memory used 156.2 MB of 224 MB Heap Memory. Max Heap Memory is 5.02 GB.
Non Heap Memory used 42.71 MB of 43.38 MB Committed Non Heap Memory. Max Non Heap Memory is -1 B.

Configured Capacity:	295.17 GB
DFS Used:	9.39 GB (3.18%)
Non DFS Used:	37.75 GB
DFS Remaining:	248.02 GB (84.03%)
Block Pool Used:	9.39 GB (3.18%)
DataNodes usages% (Min/Median/Max/stdDev):	3.18% / 3.18% / 3.18% / 0.00%
Live Nodes	1 (Decommissioned: 0)
Dead Nodes	0 (Decommissioned: 0)
Decommissioning Nodes	0
Total Datanode Volume Failures	0 (0 B)

hadoopxmlfiles.zip Show all downloads

Validation Output:

- Execution screenshots for Hadoop Map Reduce Sort 100GB File.
 - Time Taken = 7080 seconds
= 118 minutes



Overview 'ec2-52-37-161-38.us-west-2.compute.amazonaws.com:9000' (active)

Started:	Fri Apr 01 19:47:11 UTC 2016
Version:	2.7.2, r1650cf4eb874265792cc23f546d44604acf0e41
Compiled:	2016-01-26T00:08Z by Jenkins from (detached from b16504f)
Cluster ID:	CID-911c129-d515-4fb5-b4f4-e3cd67a9898
Block Pool ID:	BP-440267268-172.31.5.61-1459526750417

Summary

Security is off.
Safemode is off.
28 files and directories, 758 blocks = 786 total filesystem object(s).
Heap Memory used 163.47 MB of 269 MB Heap Memory. Max Heap Memory is 5.02 GB.
Non Heap Memory used 46 MB of 46.81 MB Committed Non Heap Memory. Max Non Heap Memory is -1 B.

Configured Capacity:	9.8 TB
DFS Used:	93.97 GB (0.94%)
Non DFS Used:	606.15 GB
DFS Remaining:	9.12 TB (93.03%)
Block Pool Used:	93.97 GB (0.94%)
DataNodes usages% (Min/Median/Max/StdDev):	0.00% / 0.00% / 15.90% / 3.74%
Live Nodes	17 (Decommissioned: 0)
Dead Nodes	0 (Decommissioned: 0)
Decommissioning Nodes	0
Total Datanode Volume Failures	0 (0 B)
Number of Under-Replicated Blocks	0
Number of Blocks Pending Deletion	0

All Applications

Scheduler Metrics											
Capacity Scheduler	Scheduling Resource Type	Minimum Allocation	Maximum Allocation								
Show 20 entries	[MEMORY]	<memory:1024, vCores:1>	<memory:8192, vCores:8>								
ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Progress	Tracking UI	Blacklisted Nodes
application_1459540056749_0003	ubuntu	Tera Sort	MAPREDUCE	default	Fri Apr 1 14:54:27 -0500 2016	Fri Apr 1 16:56:53 -0500 2016	FINISHED	SUCCEEDED	N/A	History	

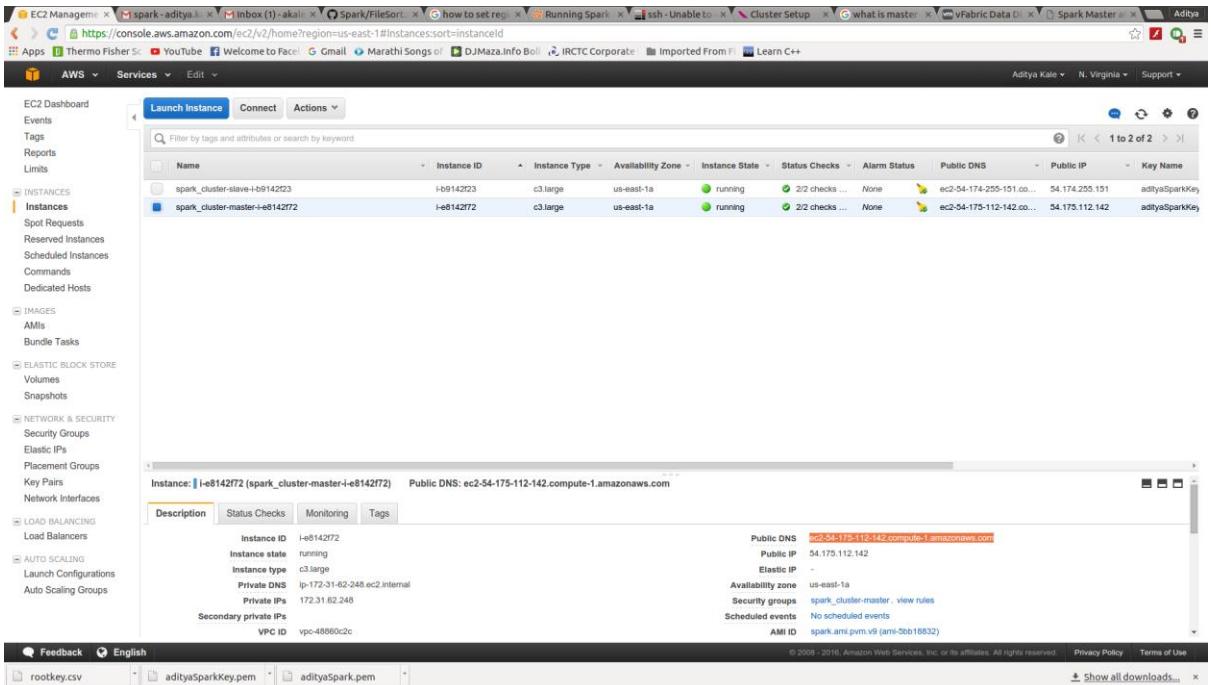
Showing 1 to 1 of 1 entries

First Previous 1 Next Last

Spark-master.zip PA2-graphs.ods Show all downloads

Validation Output:

- Execution screenshots for Spark Sort 10GB File.
 - Time Taken = 758 seconds
= 12.63 minutes



```

root@ip-172-31-62-248 ~]$ cd spark/bin/
root@ip-172-31-62-248 bin]$ touch aditya-sort.scala
root@ip-172-31-62-248 bin]$ vi aditya-sort.scala
(spark-shell -i aditya-sort.scala
16/03/31 21:20:44 INFO spark.SecurityManager: Changing view acls to: root
16/03/31 21:20:44 INFO spark.SecurityManager: Changing modify acls to: root
16/03/31 21:20:44 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(root); users with modify permissions: Set(root)
16/03/31 21:20:45 INFO spark.HttpServer: Starting HTTP Server
16/03/31 21:20:45 INFO server.Server: jetty-8.y.z-SNAPSHOT
16/03/31 21:20:45 INFO server.AbstractConnector: Started SocketConnector@0.0.0.0:37066
16/03/31 21:20:45 INFO util.Utils: Successfully started service 'HTTP class server' on port 37066.
Welcome to
   __/\__\  / \ \ / \ \ / \ \
  / \ \ / \ . / \ \ / \ \ / \ \
   \ \ / \ / \ / \ / \ / \ / \ \
version 1.6.0

Using Scala version 2.10.5 (OpenJDK 64-Bit Server VM, Java 1.7.0_99)
Type in expressions to have them evaluated.
Type :help for more information.
16/03/31 21:20:50 INFO spark.SparkContext: Running Spark version 1.6.0
16/03/31 21:20:50 WARN spark.SparkConf:
SPARK WORKER INSTANCES was detected (set to '1').
This is deprecated in Spark 1.0+.

Please instead use:
- ./spark-submit with --num-executors to specify the number of executors
- Or set SPARK_EXECUTOR_INSTANCES
- spark.executor.instances to configure the number of instances in the spark config.

16/03/31 21:20:50 INFO spark.SecurityManager: Changing view acls to: root
16/03/31 21:20:50 INFO spark.SecurityManager: Changing modify acls to: root
16/03/31 21:20:50 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(root); users with modify permissions: Set(root)
16/03/31 21:20:50 INFO util.Utils: Successfully started service 'sparkDriver' on port 42927.
16/03/31 21:20:51 INFO slf4j.Slf4jLogger: Slf4jLogger started
16/03/31 21:20:51 INFO Remoting: Starting remoting

```

```

16/03/31 21:31:46 INFO scheduler.TaskSetManager: Finished task 56.0 in stage 2.0 (TID 206) in 11119 ms on ip-172-31-53-37.ec2.internal (56/75)
16/03/31 21:31:46 INFO scheduler.TaskSetManager: Starting task 58.0 in stage 2.0 (TID 208, ip-172-31-53-37.ec2.internal, partition 58, NODE LOCAL, 1894 bytes)
16/03/31 21:31:46 INFO scheduler.TaskSetManager: Finished task 55.0 in stage 2.0 (TID 205) in 12115 ms on ip-172-31-53-37.ec2.internal (57/75)
16/03/31 21:31:57 INFO scheduler.TaskSetManager: Starting task 59.0 in stage 2.0 (TID 209, ip-172-31-53-37.ec2.internal, partition 59, NODE LOCAL, 1894 bytes)
16/03/31 21:31:57 INFO scheduler.TaskSetManager: Finished task 57.0 in stage 2.0 (TID 207) in 10731 ms on ip-172-31-53-37.ec2.internal (58/75)
16/03/31 21:31:57 INFO scheduler.TaskSetManager: Starting task 60.0 in stage 2.0 (TID 210, ip-172-31-53-37.ec2.internal, partition 60, NODE LOCAL, 1894 bytes)
16/03/31 21:31:57 INFO scheduler.TaskSetManager: Finished task 58.0 in stage 2.0 (TID 208) in 11226 ms on ip-172-31-53-37.ec2.internal (59/75)
16/03/31 21:32:10 INFO scheduler.TaskSetManager: Starting task 61.0 in stage 2.0 (TID 211, ip-172-31-53-37.ec2.internal, partition 61, NODE LOCAL, 1894 bytes)
16/03/31 21:32:10 INFO scheduler.TaskSetManager: Finished task 59.0 in stage 2.0 (TID 209) in 13017 ms on ip-172-31-53-37.ec2.internal (60/75)
16/03/31 21:32:14 INFO scheduler.TaskSetManager: Starting task 62.0 in stage 2.0 (TID 212, ip-172-31-53-37.ec2.internal, partition 62, NODE LOCAL, 1894 bytes)
16/03/31 21:32:14 INFO scheduler.TaskSetManager: Finished task 60.0 in stage 2.0 (TID 210) in 17022 ms on ip-172-31-53-37.ec2.internal (61/75)
16/03/31 21:32:23 INFO scheduler.TaskSetManager: Starting task 63.0 in stage 2.0 (TID 213, ip-172-31-53-37.ec2.internal, partition 63, NODE LOCAL, 1894 bytes)
16/03/31 21:32:23 INFO scheduler.TaskSetManager: Finished task 61.0 in stage 2.0 (TID 211) in 13486 ms on ip-172-31-53-37.ec2.internal (62/75)
16/03/31 21:32:25 INFO scheduler.TaskSetManager: Starting task 64.0 in stage 2.0 (TID 214, ip-172-31-53-37.ec2.internal, partition 64, NODE LOCAL, 1894 bytes)
16/03/31 21:32:25 INFO scheduler.TaskSetManager: Finished task 62.0 in stage 2.0 (TID 212) in 10658 ms on ip-172-31-53-37.ec2.internal (63/75)
16/03/31 21:32:39 INFO scheduler.TaskSetManager: Starting task 65.0 in stage 2.0 (TID 215, ip-172-31-53-37.ec2.internal, partition 65, NODE LOCAL, 1894 bytes)
16/03/31 21:32:39 INFO scheduler.TaskSetManager: Finished task 64.0 in stage 2.0 (TID 214) in 13466 ms on ip-172-31-53-37.ec2.internal (64/75)
16/03/31 21:32:39 INFO scheduler.TaskSetManager: Starting task 66.0 in stage 2.0 (TID 216, ip-172-31-53-37.ec2.internal, partition 66, NODE LOCAL, 1894 bytes)
16/03/31 21:32:39 INFO scheduler.TaskSetManager: Finished task 63.0 in stage 2.0 (TID 213) in 15742 ms on ip-172-31-53-37.ec2.internal (65/75)
16/03/31 21:32:49 INFO scheduler.TaskSetManager: Starting task 67.0 in stage 2.0 (TID 217, ip-172-31-53-37.ec2.internal, partition 67, NODE LOCAL, 1894 bytes)
16/03/31 21:32:49 INFO scheduler.TaskSetManager: Finished task 65.0 in stage 2.0 (TID 215) in 10844 ms on ip-172-31-53-37.ec2.internal (66/75)
16/03/31 21:32:50 INFO scheduler.TaskSetManager: Starting task 68.0 in stage 2.0 (TID 218, ip-172-31-53-37.ec2.internal, partition 68, NODE LOCAL, 1894 bytes)
16/03/31 21:32:50 INFO scheduler.TaskSetManager: Finished task 66.0 in stage 2.0 (TID 216) in 10891 ms on ip-172-31-53-37.ec2.internal (67/75)
16/03/31 21:33:04 INFO scheduler.TaskSetManager: Starting task 69.0 in stage 2.0 (TID 219, ip-172-31-53-37.ec2.internal, partition 69, NODE LOCAL, 1894 bytes)
16/03/31 21:33:04 INFO scheduler.TaskSetManager: Finished task 68.0 in stage 2.0 (TID 218) in 13997 ms on ip-172-31-53-37.ec2.internal (68/75)
16/03/31 21:33:04 INFO scheduler.TaskSetManager: Starting task 70.0 in stage 2.0 (TID 220, ip-172-31-53-37.ec2.internal, partition 70, NODE LOCAL, 1894 bytes)
16/03/31 21:33:04 INFO scheduler.TaskSetManager: Finished task 67.0 in stage 2.0 (TID 217) in 14527 ms on ip-172-31-53-37.ec2.internal (69/75)
16/03/31 21:33:15 INFO scheduler.TaskSetManager: Starting task 71.0 in stage 2.0 (TID 221, ip-172-31-53-37.ec2.internal, partition 71, NODE LOCAL, 1894 bytes)
16/03/31 21:33:15 INFO scheduler.TaskSetManager: Finished task 69.0 in stage 2.0 (TID 219) in 11539 ms on ip-172-31-53-37.ec2.internal (70/75)
16/03/31 21:33:19 INFO scheduler.TaskSetManager: Starting task 72.0 in stage 2.0 (TID 222, ip-172-31-53-37.ec2.internal, partition 72, NODE LOCAL, 1894 bytes)
16/03/31 21:33:19 INFO scheduler.TaskSetManager: Finished task 70.0 in stage 2.0 (TID 220) in 15354 ms on ip-172-31-53-37.ec2.internal (71/75)
16/03/31 21:33:31 INFO scheduler.TaskSetManager: Starting task 73.0 in stage 2.0 (TID 223, ip-172-31-53-37.ec2.internal, partition 73, NODE LOCAL, 1894 bytes)
16/03/31 21:33:31 INFO scheduler.TaskSetManager: Finished task 71.0 in stage 2.0 (TID 221) in 15791 ms on ip-172-31-53-37.ec2.internal (72/75)
16/03/31 21:33:33 INFO scheduler.TaskSetManager: Starting task 74.0 in stage 2.0 (TID 224, ip-172-31-53-37.ec2.internal, partition 74, NODE LOCAL, 1894 bytes)
16/03/31 21:33:33 INFO scheduler.TaskSetManager: Finished task 72.0 in stage 2.0 (TID 222) in 13355 ms on ip-172-31-53-37.ec2.internal (73/75)
16/03/31 21:33:42 INFO scheduler.TaskSetManager: Finished task 73.0 in stage 2.0 (TID 223) in 11037 ms on ip-172-31-53-37.ec2.internal (74/75)
16/03/31 21:33:42 INFO scheduler.TaskSetManager: Starting task 72.0 in stage 2.0 (TID 224) in 9762 ms on ip-172-31-53-37.ec2.internal (75/75)
16/03/31 21:33:42 INFO scheduler.DAGScheduler: ResultStage 2 (saveAsTextFile at <console>:30) finished in 463.110 s
16/03/31 21:33:42 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 2.0, whose tasks have all completed, from pool
16/03/31 21:33:42 INFO scheduler.DAGScheduler: Job 1 finished: saveAsTextFile at <console>:30, took 697.452027 s
end_time: Long = 145946002971
Time for Completion :758382ms

scala> ■

```

EC2 Manager x spark-aditya.x. x Inbox (1)-akali x Spark/FileSort x How to set reg x If running Spark x ssh - Unable to x Cluster Setup x what is master x vFabric Data x Spark Master x Aditya

Spark 1.6.0

Spark Master at spark://ec2-54-175-112-142.compute-1.amazonaws.com:7077

URL: spark://ec2-54-175-112-142.compute-1.amazonaws.com:7077
REST URL: spark://ec2-54-175-112-142.compute-1.amazonaws.com:6066 (cluster mode)

Alive Workers: 1

Cores in use: 2 Total: 0 Used
Memory in use: 2.7 GB Total: 0.0 B Used
Applications: 0 Running, 1 Completed
Drivers: 0 Running, 0 Completed
Status: ALIVE

Workers

Worker Id	Address	State	Cores	Memory
worker-20160331205426-172.31.53.37-34135	172.31.53.37-34135	ALIVE	2 (0 Used)	2.7 GB (0.0 B Used)

Running Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-2016033112052-0000	Spark shell	2	2.4 GB	2016/03/31 21:20:52	root	FINISHED	14 min

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-2016033112052-0000	Spark shell	2	2.4 GB	2016/03/31 21:20:52	root	FINISHED	14 min

rootkey.csv adityaSparkKey.pem adityaSpark.pem Show all downloads...

Validation Output:

- Execution screenshots for Spark Sort 100GB File.
 - Time Taken = 747 seconds
= 12.45 minutes

```

root@ip-172-31-63-28 ~$ ./spark/bin/spark-shell -i SparkSort.scala
16/03/31 23:58:24 INFO spark.SecurityManager: Changing view acls to: root
16/03/31 23:58:24 INFO spark.SecurityManager: Changing modify acls to: root
16/03/31 23:58:24 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(root); users with modify permissions: Set(root)
16/03/31 23:58:24 INFO spark.HttpServer: Starting HTTP Server
16/03/31 23:58:25 INFO server.Server: jetty-8.y.z-SNAPSHOT
16/03/31 23:58:25 INFO server.AbstractConnector: Started SocketConnector@0.0.0.0:48356
16/03/31 23:58:25 INFO util.Utils: Successfully started service 'HTTP class server' on port 48356.
Welcome to

          \   ^__^
           \  o.o
            __  .-
           )\/\  |
          _/ \_ \
         =  ^  =
         /   \
version 1.6.0

Using Scala version 2.10.5 (OpenJDK 64-Bit Server VM, Java 1.7.0_99)
Type in expressions to have them evaluated.
Type :help for more information.
16/03/31 23:58:30 INFO spark.SparkContext: Running Spark version 1.6.0
16/03/31 23:58:30 WARN spark.SparkConf:
SPARK_WORKER_INSTANCES was detected (set to '1').
This is deprecated in Spark 1.0+.

Please instead use:
- ./spark-submit with --num-executors to specify the number of executors
- Or set SPARK_EXECUTOR_INSTANCES
- spark.executor.instances to configure the number of instances in the spark config.

16/03/31 23:58:31 INFO spark.SecurityManager: Changing view acls to: root
16/03/31 23:58:31 INFO spark.SecurityManager: Changing modify acls to: root
16/03/31 23:58:31 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(root); users with modify permissions: Set(root)
16/03/31 23:58:31 INFO util.Utils: Successfully started service 'sparkDriver' on port 54835.
16/03/31 23:58:31 INFO slf4j.Slf4jLogger: Slf4jLogger started
16/03/31 23:58:31 INFO Remoting: Starting remoting
16/03/31 23:58:31 INFO Remoting: Remoting started, listening on addresses :[akka.tcp://sparkDriverActorSystem@172.31.63.28:48598]
16/03/31 23:58:31 INFO util.Utils: Successfully started service 'sparkDriverActorSystem' on port 48598.
16/03/31 23:58:31 INFO spark.SparkEnv: Registering MapOutputTracker
16/03/31 23:58:31 INFO spark.SparkEnv: Registering BlockManagerMaster
16/03/31 23:58:31 INFO storage.DiskBlockManager: Created local directory at /mnt/spark/blockmgr-71012a88-55b3-4c78-b3fa-16e1eb13a350
16/03/31 23:58:31 INFO storage.DiskBlockManager: Created local directory at /mnt2/spark/blockmgr-e4fa0146-5c69-4aac-af18-374d04298953
16/03/31 23:58:31 INFO storage.MemoryStore: MemoryStore started with capacity 511.5 MB
16/03/31 23:58:31 INFO spark.SparkEnv: Registering OutputCommitCoordinator
16/03/31 23:58:31 INFO server.Server: jetty-8.y.z-SNAPSHOT

```

```

16/04/01 00:14:04 INFO scheduler.TaskSetManager: Finished task 720.0 in stage 2.0 (TID 2210) in 16369 ms on ip-172-31-49-109.ec2.internal (711/745)
16/04/01 00:14:05 INFO scheduler.TaskSetManager: Starting task 743.0 in stage 2.0 (TID 2233, ip-172-31-55-178.ec2.internal, partition 743,PROCESS_LOCAL, 1894 bytes)
16/04/01 00:14:05 INFO scheduler.TaskSetManager: Finished task 707.0 in stage 2.0 (TID 2197) in 22647 ms on ip-172-31-55-178.ec2.internal (712/745)
16/04/01 00:14:05 INFO scheduler.TaskSetManager: Starting task 744.0 in stage 2.0 (TID 2234, ip-172-31-62-46.ec2.internal, partition 744,PROCESS_LOCAL, 1894 bytes)
16/04/01 00:14:05 INFO scheduler.TaskSetManager: Finished task 697.0 in stage 2.0 (TID 2187) in 29172 ms on ip-172-31-59-148.ec2.internal (714/745)
16/04/01 00:14:05 INFO scheduler.TaskSetManager: Finished task 704.0 in stage 2.0 (TID 2194) in 24519 ms on ip-172-31-52-236.ec2.internal (716/745)
16/04/01 00:14:05 INFO scheduler.TaskSetManager: Finished task 717.0 in stage 2.0 (TID 2207) in 19625 ms on ip-172-31-62-182.ec2.internal (717/745)
16/04/01 00:14:07 INFO scheduler.TaskSetManager: Finished task 728.0 in stage 2.0 (TID 2218) in 12323 ms on ip-172-31-49-229.ec2.internal (718/745)
16/04/01 00:14:07 INFO scheduler.TaskSetManager: Finished task 702.0 in stage 2.0 (TID 2192) in 25044 ms on ip-172-31-56-92.ec2.internal (715/745)
16/04/01 00:14:08 INFO scheduler.TaskSetManager: Finished task 729.0 in stage 2.0 (TID 2219) in 13587 ms on ip-172-31-50-145.ec2.internal (720/745)
16/04/01 00:14:10 INFO scheduler.TaskSetManager: Finished task 699.0 in stage 2.0 (TID 2189) in 32834 ms on ip-172-31-50-227.ec2.internal (721/745)
16/04/01 00:14:10 INFO scheduler.TaskSetManager: Finished task 714.0 in stage 2.0 (TID 2204) in 25067 ms on ip-172-31-55-149.ec2.internal (722/745)
16/04/01 00:14:10 INFO scheduler.TaskSetManager: Finished task 725.0 in stage 2.0 (TID 2215) in 20526 ms on ip-172-31-50-145.ec2.internal (723/745)
16/04/01 00:14:10 INFO scheduler.TaskSetManager: Finished task 724.0 in stage 2.0 (TID 2214) in 21073 ms on ip-172-31-54-214.ec2.internal (724/745)
16/04/01 00:14:11 INFO scheduler.TaskSetManager: Finished task 726.0 in stage 2.0 (TID 2216) in 20984 ms on ip-172-31-59-148.ec2.internal (725/745)
16/04/01 00:14:11 INFO scheduler.TaskSetManager: Finished task 715.0 in stage 2.0 (TID 2205) in 26431 ms on ip-172-31-63-87.ec2.internal (726/745)
16/04/01 00:14:12 INFO scheduler.TaskSetManager: Finished task 727.0 in stage 2.0 (TID 2217) in 20276 ms on ip-172-31-57-83.ec2.internal (727/745)
16/04/01 00:14:12 INFO scheduler.TaskSetManager: Finished task 716.0 in stage 2.0 (TID 2206) in 27026 ms on ip-172-31-49-109.ec2.internal (728/745)
16/04/01 00:14:13 INFO scheduler.TaskSetManager: Finished task 718.0 in stage 2.0 (TID 2208) in 25033 ms on ip-172-31-55-149.ec2.internal (729/745)
16/04/01 00:14:15 INFO scheduler.TaskSetManager: Finished task 719.0 in stage 2.0 (TID 2209) in 26864 ms on ip-172-31-63-87.ec2.internal (730/745)
16/04/01 00:14:15 INFO scheduler.TaskSetManager: Finished task 731.0 in stage 2.0 (TID 2221) in 15952 ms on ip-172-31-57-83.ec2.internal (731/745)
16/04/01 00:14:21 INFO scheduler.TaskSetManager: Finished task 730.0 in stage 2.0 (TID 2220) in 16984 ms on ip-172-31-54-91.ec2.internal (732/745)
16/04/01 00:14:22 INFO scheduler.TaskSetManager: Finished task 736.0 in stage 2.0 (TID 2226) in 18004 ms on ip-172-31-52-236.ec2.internal (733/745)
16/04/01 00:14:21 INFO scheduler.TaskSetManager: Finished task 733.0 in stage 2.0 (TID 2223) in 29117 ms on ip-172-31-55-178.ec2.internal (734/745)
16/04/01 00:14:22 INFO scheduler.TaskSetManager: Finished task 740.0 in stage 2.0 (TID 2230) in 18646 ms on ip-172-31-54-91.ec2.internal (735/745)
16/04/01 00:14:24 INFO scheduler.TaskSetManager: Finished task 739.0 in stage 2.0 (TID 2229) in 21722 ms on ip-172-31-49-229.ec2.internal (736/745)
16/04/01 00:14:25 INFO scheduler.TaskSetManager: Finished task 732.0 in stage 2.0 (TID 2222) in 24647 ms on ip-172-31-54-214.ec2.internal (737/745)
16/04/01 00:14:26 INFO scheduler.TaskSetManager: Finished task 741.0 in stage 2.0 (TID 2231) in 22635 ms on ip-172-31-62-46.ec2.internal (738/745)
16/04/01 00:14:26 INFO scheduler.TaskSetManager: Finished task 743.0 in stage 2.0 (TID 2233) in 21265 ms on ip-172-31-55-178.ec2.internal (739/745)
16/04/01 00:14:27 INFO scheduler.TaskSetManager: Finished task 742.0 in stage 2.0 (TID 2232) in 22875 ms on ip-172-31-49-109.ec2.internal (740/745)
16/04/01 00:14:28 INFO scheduler.TaskSetManager: Finished task 737.0 in stage 2.0 (TID 2227) in 26224 ms on ip-172-31-48-123.ec2.internal (741/745)
16/04/01 00:14:29 INFO scheduler.TaskSetManager: Finished task 734.0 in stage 2.0 (TID 2224) in 27043 ms on ip-172-31-62-182.ec2.internal (742/745)
16/04/01 00:14:29 INFO scheduler.TaskSetManager: Finished task 738.0 in stage 2.0 (TID 2228) in 26849 ms on ip-172-31-48-123.ec2.internal (743/745)
16/04/01 00:14:30 INFO scheduler.TaskSetManager: Finished task 735.0 in stage 2.0 (TID 2225) in 28243 ms on ip-172-31-56-92.ec2.internal (744/745)
16/04/01 00:14:30 INFO scheduler.TaskSetManager: Finished task 744.0 in stage 2.0 (TID 2234) in 25282 ms on ip-172-31-62-46.ec2.internal (745/745)
16/04/01 00:14:30 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 2.0, whose tasks have all completed, from pool
16/04/01 00:14:30 INFO scheduler.DAGScheduler: ResultStage 2 (saveAsTextFile at <console>:30) finished in 418.491 s
16/04/01 00:14:30 INFO scheduler.DAGScheduler: Job 1 finished: saveAsTextFile at <console>:30, took 693.664203 s
end time: Long = 1459469670809
Time Taken for Completion of Sorting is :746607Milliseconds
scala>

```

Spark Master at spark://ec2-54-164-238-161.compute-1.amazonaws.com:7077

URL: spark://ec2-54-164-238-161.compute-1.amazonaws.com:7077
REST URL: spark://ec2-54-164-238-161.compute-1.amazonaws.com:6066 (cluster mode)

Alive Workers: 16
Cores in use: 32 Total, 0 Used
Memory in use: 42.7 GB Total, 0.0 B Used
Applications: 0 Running, 3 Completed
Drivers: 0 Running, 0 Completed
Status: ALIVE

Workers

Worker Id	Address	State	Cores	Memory
worker_20160331223340-172.31.48.123-37487	172.31.48.123.37487	ALIVE	2 (0 Used)	2.7 GB (0.0 B Used)
worker_20160331223340-172.31.49.109-57900	172.31.49.109.57900	ALIVE	2 (0 Used)	2.7 GB (0.0 B Used)
worker_20160331223340-172.31.49.229-44374	172.31.49.229.44374	ALIVE	2 (0 Used)	2.7 GB (0.0 B Used)
worker_20160331223340-172.31.50.227-40937	172.31.50.227.40937	ALIVE	2 (0 Used)	2.7 GB (0.0 B Used)
worker_20160331223340-172.31.52.236-58597	172.31.52.236.58597	ALIVE	2 (0 Used)	2.7 GB (0.0 B Used)
worker_20160331223340-172.31.54.214-51450	172.31.54.214.51450	ALIVE	2 (0 Used)	2.7 GB (0.0 B Used)
worker_20160331223340-172.31.54.91-58137	172.31.54.91.58137	ALIVE	2 (0 Used)	2.7 GB (0.0 B Used)
worker_20160331223340-172.31.55.149-54128	172.31.55.149.54128	ALIVE	2 (0 Used)	2.7 GB (0.0 B Used)
worker_20160331223340-172.31.55.178-57765	172.31.55.178.57765	ALIVE	2 (0 Used)	2.7 GB (0.0 B Used)
worker_20160331223340-172.31.56.92-44461	172.31.56.92.44461	ALIVE	2 (0 Used)	2.7 GB (0.0 B Used)
worker_20160331223340-172.31.57.83-40064	172.31.57.83.40064	ALIVE	2 (0 Used)	2.7 GB (0.0 B Used)
worker_20160331223340-172.31.59.148-52363	172.31.59.148.52363	ALIVE	2 (0 Used)	2.7 GB (0.0 B Used)
worker_20160331223340-172.31.62.182-36193	172.31.62.182.36193	ALIVE	2 (0 Used)	2.7 GB (0.0 B Used)
worker_20160331223340-172.31.62.46-53486	172.31.62.46.53486	ALIVE	2 (0 Used)	2.7 GB (0.0 B Used)
worker_20160331223340-172.31.63.87-56941	172.31.63.87.56941	ALIVE	2 (0 Used)	2.7 GB (0.0 B Used)
worker_20160331223340-172.31.50.145-60284	172.31.50.145.60284	ALIVE	2 (0 Used)	2.7 GB (0.0 B Used)

Running Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
----------------	------	-------	-----------------	----------------	------	-------	----------

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-20160401000151-0002	Spark shell	32	2.4 GB	2016/04/01 00:01:51	root	FINISHED	15 min

Validation Output:

