

# CS 553 Cloud Computing

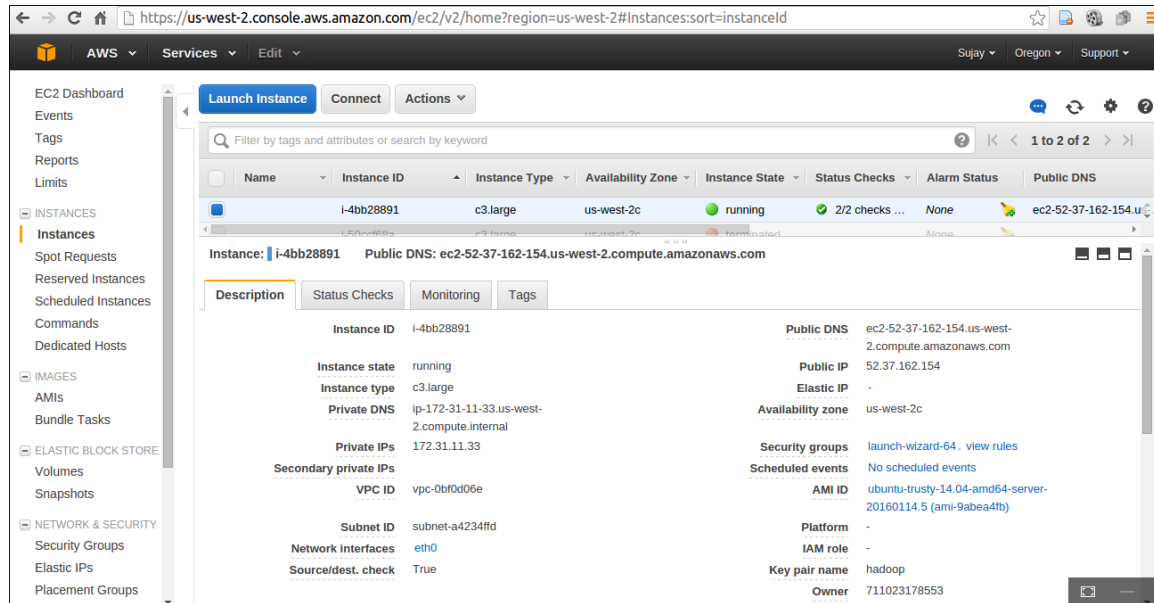
## Programming Assignment 2

Sujay Gunjal (CWID: A20351746)

### Screenshots

Screenshots of Shared-Memory, Hadoop and spark execution:-

Aws screenshot of Shared memory 10GB sort execution:-



4 threads Execution screenshot:-

```
ubuntu@ip-172-31-11-33: ~/code/Single_Node
f1_length 100000000
renamed
Execution complete !!!
Time required to sort dataset (Milliseconds):- 1914224
ubuntu@ip-172-31-11-33:~/code/Single_Node$
```

Time taken by 4 Thread to sort 10Gb file is: - 1914224 milliseconds

31.90375 minutes

### Valsort output of shared memory sort:-

```
ubuntu@ip-172-31-11-33: ~/64
ubuntu@ip-172-31-11-33:~/code/Single_Node$ pwd
/home/ubuntu/code/Single_Node
ubuntu@ip-172-31-11-33:~/code/Single_Node$ ls -lrt
total 9765696
-rw----- 1 ubuntu ubuntu      3085 Mar 26 18:53 Terasort.class
-rw----- 1 ubuntu ubuntu      3480 Mar 26 18:53 Terasort.java~
-rw----- 1 ubuntu ubuntu     8901 Mar 26 18:53 File_Sort.java
-rw----- 1 ubuntu ubuntu     1904 Mar 26 18:53 MergeSort.class
-rw----- 1 ubuntu ubuntu       724 Mar 26 18:53 File_Collector.java
-rw----- 1 ubuntu ubuntu       724 Mar 26 18:53 File_Collector.java~
-rw----- 1 ubuntu ubuntu     8901 Mar 26 18:53 File_Sort.java~
-rw----- 1 ubuntu ubuntu     6637 Mar 26 18:53 File_Sort.class
-rw----- 1 ubuntu ubuntu     2707 Mar 26 18:53 MergeSort.java
-rw----- 1 ubuntu ubuntu     3480 Mar 26 18:53 Terasort.java
-rw----- 1 ubuntu ubuntu       876 Mar 26 18:53 File_Collector.class
-rw-rw-r-- 1 ubuntu ubuntu 10000000000 Mar 26 19:42 final_sorted_dataset
ubuntu@ip-172-31-11-33:~/code/Single_Node$ cd /home/ubuntu/64
ubuntu@ip-172-31-11-33:~/64$ ./valsort /home/ubuntu/code/Single_Node/final_sorted_dataset
Records: 100000000
Checksum: 2faf0ab746e89a8
Duplicate keys: 0
SUCCESS - all records are in order
ubuntu@ip-172-31-11-33:~/64$
```

### Head output of shared memory sort:-

```
ubuntu@ip-172-31-11-33: ~/code/Single_Node
ubuntu@ip-172-31-11-33:~/code/Single_Node$ head -10 final_sorted_dataset
"O!uve 0000000000000000000000000000000000001228D4 77778888000022224444DDDDDDDEEEEE00000000CCCC7777DDDD
Pmd32= 0000000000000000000000000000000000003440CC1 FFFFEEEE6666CCCCBBBB999933335555DDDDDDDD777788886666
!&S3/] 0000000000000000000000000000000000002145D78 8888BBBBDDDD1111CCCC55556666BBBB1111EEEEDDDD22229999
!,=U#,9 00000000000000000000000000000000000019072E3 33332222FFFFBBBB0000FFFFAAAA666655553333DDDD3333CCCC
!0f[ITd 0000000000000000000000000000000000003CAAB4B 9999FFFF555533337777CCCC4444BBBB7777EEEEBBBBDDDD4444
!f6Suy2 0000000000000000000000000000000000003ABFD84 EEEE555555556666AAAA5555BBBBDDDD0000111166660000DDDD
#%NlPq. 1111000033334444111166666666AAAAAA00001111CCCCEEEE 8888AAAA11114444FFFF77773333EEEE44440000FFFF99999999
#'^cl'~ 0000000000000000000000000000000000002EDC5C8 8888AAAA11114444FFFF77773333EEEE44440000FFFF99999999
$"- 'Q)] 0000000000000000000000000000000000005F1265D CCCC6666EEEE22220000DDDDAAAA88886666BBBB00006666AAAA
ubuntu@ip-172-31-11-33:~/code/Single_Node$
```

### Tail output of shared memory sort:-

```
ubuntu@ip-172-31-11-33: ~/code/Single_Node
ubuntu@ip-172-31-11-33:~/code/Single_Node$ tail -10 final_sorted_dataset
~~~uq2k#=U 0000000000000000000000000000000000002C06745 99991111DDDD222211110000FFFFEEEEFFFF33337777CCCC2222
~~~v/0&Qnm 0000000000000000000000000000000000004709701 CCCC88883333FFFF0000000000000099991111FFFF777744446666
~~~yK0l:gE 0000000000000000000000000000000000002048B4F CCCC11114444888822226666BBBB888855557777EEEEBBBB0000
~~~yK^H.il 000000000000000000000000000000000000463D004 44440000FFFF3333999944447777DDDDFFFFAAAA11118888DDDD
~~~yL;c'XE 0000000000000000000000000000000000005B0D211 2222EEEE3333000022221111CCCCFFFF555577774444BBBB6666
~~~zbA_ Tt 0000000000000000000000000000000000007F9F4F BBBBCCCC666655559999FFFF8888AAAA11116666AAAA BBBB0000
~~~ze0^FEg 0000000000000000000000000000000000001E06130 4444CCCCBBBB99992222888855558888CCCCFFFF000011111111
~~~}GxjWHI 000000000000000000000000000000000000CA1345 777711118888AAAAAA22221111BBBB00002222BBBBCCCC2222
~~~}P;]g0g 00000000000000000000000000000000000040DA3E4 4444FFFF444466663333EEEE88888888DDDDDEEEE44442222DDDD
~~~}kU|K<p 0000000000000000000000000000000000005E4A0AA 0000666655551111BBBB88889999AAAA55550000333355557777
ubuntu@ip-172-31-11-33:~/code/Single_Node$
```

### 8 threads Execution screenshot:-

```
ubuntu@ip-172-31-1-2: ~/sujay/Single_Node
f2_length 25600000
Array List Size is:- 998
f1_length 48800000
f2_length 51200000
Array List Size is:- 999
f1_length 100000000
renamed
Execution complete !!!
Time required to sort dataset (Milliseconds):- 2075342
```

Time taken by 8 Thread to sort 10Gb file is :- 2075342 milliseconds

34.58903 minutes

### 2 threads Execution screenshot:-

```
ubuntu@ip-172-31-4-181: /sujay/sujay
f1_length 25600000
f2_length 25600000
Array List Size is:- 998
f1_length 48800000
f2_length 51200000
Array List Size is:- 999
f1_length 100000000
renamed
Execution complete !!!
Time required to sort dataset (Milliseconds):- 2109931
ubuntu@ip-172-31-4-181:/sujay/sujay$ java File_Collector
Enter the name of file to sort:-
```

Time taken by 2 Thread to sort 10Gb file is :- 2109931 milliseconds

35.165516 minutes

1 thread Execution screenshot:-

```
ubuntu@ip-172-31-11-231: ~/Single_Node
Array List Size is:- 996
f1_length 23200000
f2_length 25600000
Array List Size is:- 997
f1_length 25600000
f2_length 25600000
Array List Size is:- 998
f1_length 48800000
f2_length 51200000
Array List Size is:- 999
f1_length 100000000
renamed
Execution complete !!!
Time required to sort dataset (Milliseconds):- 2951058
ubuntu@ip-172-31-11-231:~/Single_Node$
```

Time taken by 1 Thread to sort 10 Gb file is: - 2951058 milliseconds

49.1843 minutes

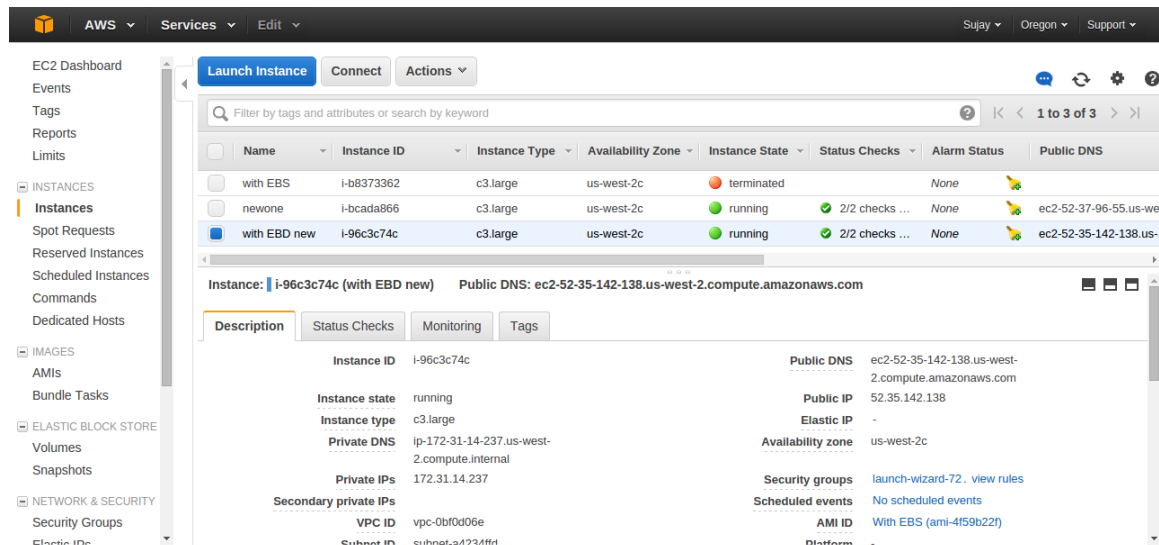
Valsort output of Shared memory application:-

```
ubuntu@ip-172-31-11-231: ~/64
ubuntu@ip-172-31-11-231:~/Single_Node$ ls
File_Collector.class  File_Collector.java~  File_Sort.java  final
File_Collector.java  File_Sort.class      File_Sort.java~  Merge
ubuntu@ip-172-31-11-231:~/Single_Node$ cd ..
ubuntu@ip-172-31-11-231:~$ ls
32  64  gensort-linux-1.5.tar.gz  gpl-2.0.txt  Single_Node
ubuntu@ip-172-31-11-231:~$ cd 64
ubuntu@ip-172-31-11-231:~/64$ pwd
/home/ubuntu/64
ubuntu@ip-172-31-11-231:~/64$ ./valsort /home/ubuntu/Single_Node/f
Records: 100000000
Checksum: 2faf0ab746e89a8
Duplicate keys: 0
SUCCESS - all records are in order
ubuntu@ip-172-31-11-231:~/64$
```

## Hadoop Execution:-

10 Gb Single Node Hadoop executions:-

Aws Screenshot:-



10Gb Hadoop output:-

```
ubuntu@ip-172-31-14-237:~/hadoop-2.7.2$ bin/hadoop jar TeraSort.jar
TeraSort /hdfs_sujay/10GB_input_file /hdfs_sujay/output
16/03/28 07:06:37 INFO client.RMProxy: Connecting to ResourceManager
at ec2-52-35-142-138.us-west-2.compute.amazonaws.com/172.31.14.237:9050
16/03/28 07:06:37 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 07:06:38 INFO input.FileInputFormat: Total input paths to
process : 1
16/03/28 07:06:38 INFO mapreduce.JobSubmitter: number of splits:75
16/03/28 07:06:38 INFO mapreduce.JobSubmitter: Submitting tokens for
job: job_1459147758523_0001
16/03/28 07:06:39 INFO impl.YarnClientImpl: Submitted application
application_1459147758523_0001
16/03/28 07:06:39 INFO mapreduce.Job: The url to track the job:
http://ec2-52-35-142-138.us-west-2.compute.amazonaws.com:9006/proxy/application_1459147758523_0001/
16/03/28 07:06:39 INFO mapreduce.Job: Running job:
job_1459147758523_0001
16/03/28 07:06:48 INFO mapreduce.Job: Job job_1459147758523_0001
running in uber mode : false
16/03/28 07:06:48 INFO mapreduce.Job: map 0% reduce 0%
16/03/28 07:07:12 INFO mapreduce.Job: map 1% reduce 0%
16/03/28 07:07:16 INFO mapreduce.Job: map 2% reduce 0%
```

```
16/03/28 07:07:20 INFO mapreduce.Job: map 3% reduce 0%
16/03/28 07:07:23 INFO mapreduce.Job: map 4% reduce 0%
16/03/28 07:07:36 INFO mapreduce.Job: map 5% reduce 0%
16/03/28 07:07:48 INFO mapreduce.Job: map 6% reduce 0%
16/03/28 07:07:51 INFO mapreduce.Job: map 7% reduce 0%
16/03/28 07:07:57 INFO mapreduce.Job: map 8% reduce 0%
16/03/28 07:08:19 INFO mapreduce.Job: map 9% reduce 0%
16/03/28 07:08:22 INFO mapreduce.Job: map 10% reduce 0%
16/03/28 07:08:24 INFO mapreduce.Job: map 11% reduce 0%
16/03/28 07:08:27 INFO mapreduce.Job: map 12% reduce 0%
16/03/28 07:08:38 INFO mapreduce.Job: map 13% reduce 0%
16/03/28 07:08:50 INFO mapreduce.Job: map 14% reduce 0%
16/03/28 07:08:56 INFO mapreduce.Job: map 15% reduce 0%
16/03/28 07:09:00 INFO mapreduce.Job: map 16% reduce 0%
16/03/28 07:09:25 INFO mapreduce.Job: map 17% reduce 0%
16/03/28 07:09:28 INFO mapreduce.Job: map 19% reduce 0%
16/03/28 07:09:32 INFO mapreduce.Job: map 20% reduce 0%
16/03/28 07:09:47 INFO mapreduce.Job: map 21% reduce 0%
16/03/28 07:09:59 INFO mapreduce.Job: map 22% reduce 0%
16/03/28 07:10:02 INFO mapreduce.Job: map 23% reduce 0%
16/03/28 07:10:08 INFO mapreduce.Job: map 24% reduce 0%
16/03/28 07:10:32 INFO mapreduce.Job: map 25% reduce 0%
16/03/28 07:10:35 INFO mapreduce.Job: map 26% reduce 0%
16/03/28 07:10:38 INFO mapreduce.Job: map 27% reduce 0%
16/03/28 07:10:41 INFO mapreduce.Job: map 27% reduce 1%
16/03/28 07:10:44 INFO mapreduce.Job: map 27% reduce 2%
16/03/28 07:10:51 INFO mapreduce.Job: map 28% reduce 2%
16/03/28 07:10:53 INFO mapreduce.Job: map 28% reduce 3%
16/03/28 07:11:00 INFO mapreduce.Job: map 29% reduce 3%
16/03/28 07:11:06 INFO mapreduce.Job: map 30% reduce 3%
16/03/28 07:11:11 INFO mapreduce.Job: map 31% reduce 3%
16/03/28 07:11:14 INFO mapreduce.Job: map 31% reduce 4%
16/03/28 07:11:20 INFO mapreduce.Job: map 31% reduce 5%
16/03/28 07:11:23 INFO mapreduce.Job: map 31% reduce 6%
16/03/28 07:11:29 INFO mapreduce.Job: map 31% reduce 7%
16/03/28 07:11:35 INFO mapreduce.Job: map 31% reduce 8%
16/03/28 07:11:42 INFO mapreduce.Job: map 32% reduce 8%
16/03/28 07:11:45 INFO mapreduce.Job: map 33% reduce 8%
16/03/28 07:11:51 INFO mapreduce.Job: map 34% reduce 8%
16/03/28 07:11:53 INFO mapreduce.Job: map 34% reduce 9%
16/03/28 07:12:04 INFO mapreduce.Job: map 35% reduce 9%
16/03/28 07:12:06 INFO mapreduce.Job: map 35% reduce 10%
16/03/28 07:12:16 INFO mapreduce.Job: map 36% reduce 10%
16/03/28 07:12:22 INFO mapreduce.Job: map 37% reduce 10%
16/03/28 07:12:40 INFO mapreduce.Job: map 37% reduce 11%
16/03/28 07:12:46 INFO mapreduce.Job: map 37% reduce 12%
16/03/28 07:13:01 INFO mapreduce.Job: map 39% reduce 12%
16/03/28 07:13:04 INFO mapreduce.Job: map 40% reduce 12%
16/03/28 07:13:19 INFO mapreduce.Job: map 41% reduce 12%
16/03/28 07:13:23 INFO mapreduce.Job: map 42% reduce 12%
16/03/28 07:13:35 INFO mapreduce.Job: map 43% reduce 12%
16/03/28 07:13:41 INFO mapreduce.Job: map 44% reduce 12%
16/03/28 07:13:56 INFO mapreduce.Job: map 44% reduce 13%
16/03/28 07:14:05 INFO mapreduce.Job: map 44% reduce 14%
16/03/28 07:14:11 INFO mapreduce.Job: map 44% reduce 15%
16/03/28 07:14:17 INFO mapreduce.Job: map 45% reduce 15%
```

16/03/28	07:14:20	INFO	mapreduce.Job:	map	46%	reduce	15%
16/03/28	07:14:21	INFO	mapreduce.Job:	map	47%	reduce	15%
16/03/28	07:14:33	INFO	mapreduce.Job:	map	48%	reduce	15%
16/03/28	07:14:45	INFO	mapreduce.Job:	map	49%	reduce	15%
16/03/28	07:14:48	INFO	mapreduce.Job:	map	50%	reduce	15%
16/03/28	07:14:52	INFO	mapreduce.Job:	map	51%	reduce	15%
16/03/28	07:14:57	INFO	mapreduce.Job:	map	51%	reduce	16%
16/03/28	07:15:15	INFO	mapreduce.Job:	map	51%	reduce	17%
16/03/28	07:15:29	INFO	mapreduce.Job:	map	52%	reduce	17%
16/03/28	07:15:32	INFO	mapreduce.Job:	map	53%	reduce	17%
16/03/28	07:15:35	INFO	mapreduce.Job:	map	54%	reduce	17%
16/03/28	07:15:51	INFO	mapreduce.Job:	map	55%	reduce	17%
16/03/28	07:16:01	INFO	mapreduce.Job:	map	56%	reduce	17%
16/03/28	07:16:04	INFO	mapreduce.Job:	map	57%	reduce	17%
16/03/28	07:16:13	INFO	mapreduce.Job:	map	57%	reduce	18%
16/03/28	07:16:19	INFO	mapreduce.Job:	map	57%	reduce	19%
16/03/28	07:16:35	INFO	mapreduce.Job:	map	58%	reduce	19%
16/03/28	07:16:36	INFO	mapreduce.Job:	map	59%	reduce	19%
16/03/28	07:16:40	INFO	mapreduce.Job:	map	60%	reduce	19%
16/03/28	07:16:52	INFO	mapreduce.Job:	map	62%	reduce	19%
16/03/28	07:17:04	INFO	mapreduce.Job:	map	63%	reduce	19%
16/03/28	07:17:08	INFO	mapreduce.Job:	map	64%	reduce	19%
16/03/28	07:17:14	INFO	mapreduce.Job:	map	64%	reduce	20%
16/03/28	07:17:20	INFO	mapreduce.Job:	map	64%	reduce	21%
16/03/28	07:17:44	INFO	mapreduce.Job:	map	65%	reduce	21%
16/03/28	07:17:47	INFO	mapreduce.Job:	map	66%	reduce	21%
16/03/28	07:17:50	INFO	mapreduce.Job:	map	67%	reduce	21%
16/03/28	07:18:03	INFO	mapreduce.Job:	map	68%	reduce	21%
16/03/28	07:18:15	INFO	mapreduce.Job:	map	69%	reduce	21%
16/03/28	07:18:18	INFO	mapreduce.Job:	map	70%	reduce	21%
16/03/28	07:18:23	INFO	mapreduce.Job:	map	71%	reduce	21%
16/03/28	07:18:27	INFO	mapreduce.Job:	map	71%	reduce	22%
16/03/28	07:18:30	INFO	mapreduce.Job:	map	71%	reduce	23%
16/03/28	07:18:33	INFO	mapreduce.Job:	map	71%	reduce	24%
16/03/28	07:18:48	INFO	mapreduce.Job:	map	72%	reduce	24%
16/03/28	07:18:51	INFO	mapreduce.Job:	map	73%	reduce	24%
16/03/28	07:18:54	INFO	mapreduce.Job:	map	74%	reduce	24%
16/03/28	07:19:07	INFO	mapreduce.Job:	map	75%	reduce	24%
16/03/28	07:19:16	INFO	mapreduce.Job:	map	76%	reduce	24%
16/03/28	07:19:19	INFO	mapreduce.Job:	map	77%	reduce	24%
16/03/28	07:19:34	INFO	mapreduce.Job:	map	77%	reduce	25%
16/03/28	07:19:40	INFO	mapreduce.Job:	map	77%	reduce	26%
16/03/28	07:20:02	INFO	mapreduce.Job:	map	78%	reduce	26%
16/03/28	07:20:03	INFO	mapreduce.Job:	map	79%	reduce	26%
16/03/28	07:20:06	INFO	mapreduce.Job:	map	80%	reduce	26%
16/03/28	07:20:18	INFO	mapreduce.Job:	map	81%	reduce	26%
16/03/28	07:20:21	INFO	mapreduce.Job:	map	82%	reduce	26%
16/03/28	07:20:30	INFO	mapreduce.Job:	map	83%	reduce	26%
16/03/28	07:20:33	INFO	mapreduce.Job:	map	84%	reduce	26%
16/03/28	07:20:43	INFO	mapreduce.Job:	map	84%	reduce	27%
16/03/28	07:20:46	INFO	mapreduce.Job:	map	84%	reduce	28%
16/03/28	07:21:01	INFO	mapreduce.Job:	map	85%	reduce	28%
16/03/28	07:21:04	INFO	mapreduce.Job:	map	86%	reduce	28%
16/03/28	07:21:05	INFO	mapreduce.Job:	map	87%	reduce	28%
16/03/28	07:21:17	INFO	mapreduce.Job:	map	88%	reduce	28%
16/03/28	07:21:27	INFO	mapreduce.Job:	map	89%	reduce	28%



```
16/03/28 07:21:30 INFO mapreduce.Job: map 90% reduce 28%
16/03/28 07:21:36 INFO mapreduce.Job: map 91% reduce 28%
16/03/28 07:21:42 INFO mapreduce.Job: map 91% reduce 29%
16/03/28 07:21:48 INFO mapreduce.Job: map 91% reduce 30%
16/03/28 07:22:15 INFO mapreduce.Job: map 92% reduce 30%
16/03/28 07:22:18 INFO mapreduce.Job: map 93% reduce 30%
16/03/28 07:22:22 INFO mapreduce.Job: map 94% reduce 30%
16/03/28 07:22:31 INFO mapreduce.Job: map 95% reduce 30%
16/03/28 07:22:43 INFO mapreduce.Job: map 96% reduce 30%
16/03/28 07:22:46 INFO mapreduce.Job: map 97% reduce 30%
16/03/28 07:22:54 INFO mapreduce.Job: map 97% reduce 31%
16/03/28 07:22:57 INFO mapreduce.Job: map 97% reduce 32%
16/03/28 07:23:13 INFO mapreduce.Job: map 98% reduce 32%
16/03/28 07:23:14 INFO mapreduce.Job: map 99% reduce 32%
16/03/28 07:23:25 INFO mapreduce.Job: map 99% reduce 33%
16/03/28 07:23:28 INFO mapreduce.Job: map 100% reduce 33%
16/03/28 07:23:37 INFO mapreduce.Job: map 100% reduce 41%
16/03/28 07:23:40 INFO mapreduce.Job: map 100% reduce 61%
16/03/28 07:23:43 INFO mapreduce.Job: map 100% reduce 67%
16/03/28 07:23:49 INFO mapreduce.Job: map 100% reduce 68%
16/03/28 07:23:55 INFO mapreduce.Job: map 100% reduce 69%
16/03/28 07:23:58 INFO mapreduce.Job: map 100% reduce 70%
16/03/28 07:24:04 INFO mapreduce.Job: map 100% reduce 71%
16/03/28 07:24:07 INFO mapreduce.Job: map 100% reduce 72%
16/03/28 07:24:13 INFO mapreduce.Job: map 100% reduce 73%
16/03/28 07:24:19 INFO mapreduce.Job: map 100% reduce 74%
16/03/28 07:24:22 INFO mapreduce.Job: map 100% reduce 75%
16/03/28 07:24:28 INFO mapreduce.Job: map 100% reduce 76%
16/03/28 07:24:31 INFO mapreduce.Job: map 100% reduce 77%
16/03/28 07:24:37 INFO mapreduce.Job: map 100% reduce 78%
16/03/28 07:24:43 INFO mapreduce.Job: map 100% reduce 79%
16/03/28 07:24:46 INFO mapreduce.Job: map 100% reduce 80%
16/03/28 07:24:52 INFO mapreduce.Job: map 100% reduce 81%
16/03/28 07:24:58 INFO mapreduce.Job: map 100% reduce 82%
16/03/28 07:25:04 INFO mapreduce.Job: map 100% reduce 83%
16/03/28 07:25:07 INFO mapreduce.Job: map 100% reduce 84%
16/03/28 07:25:13 INFO mapreduce.Job: map 100% reduce 85%
16/03/28 07:25:19 INFO mapreduce.Job: map 100% reduce 86%
16/03/28 07:25:22 INFO mapreduce.Job: map 100% reduce 87%
16/03/28 07:25:28 INFO mapreduce.Job: map 100% reduce 88%
16/03/28 07:25:34 INFO mapreduce.Job: map 100% reduce 89%
16/03/28 07:25:37 INFO mapreduce.Job: map 100% reduce 90%
16/03/28 07:25:43 INFO mapreduce.Job: map 100% reduce 91%
16/03/28 07:25:49 INFO mapreduce.Job: map 100% reduce 92%
16/03/28 07:25:52 INFO mapreduce.Job: map 100% reduce 93%
16/03/28 07:25:58 INFO mapreduce.Job: map 100% reduce 94%
16/03/28 07:26:04 INFO mapreduce.Job: map 100% reduce 95%
16/03/28 07:26:07 INFO mapreduce.Job: map 100% reduce 96%
16/03/28 07:26:13 INFO mapreduce.Job: map 100% reduce 97%
16/03/28 07:26:19 INFO mapreduce.Job: map 100% reduce 98%
16/03/28 07:26:22 INFO mapreduce.Job: map 100% reduce 99%
16/03/28 07:26:28 INFO mapreduce.Job: map 100% reduce 100%
16/03/28 07:26:30 INFO mapreduce.Job: Job job_1459147758523_0001
completed successfully
16/03/28 07:26:31 INFO mapreduce.Job: Counters: 50
File System Counters
```



```

FILE: Number of bytes read=30256951212
FILE: Number of bytes written=40465905460
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=10000314579
HDFS: Number of bytes written=10000000000
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)=4980632
  Total time spent by all reduces in occupied slots
(ms)=983744
  Total time spent by all map tasks (ms)=4980632
  Total time spent by all reduce tasks (ms)=983744
  Total vcore-milliseconds taken by all map tasks=4980632
  Total vcore-milliseconds taken by all reduce tasks=983744
  Total megabyte-milliseconds taken by all map
tasks=5100167168
  Total megabyte-milliseconds taken by all reduce
tasks=1007353856
Map-Reduce Framework
  Map input records=100000000
  Map output records=100000000
  Map output bytes=10000000000
  Map output materialized bytes=10200000450
  Input split bytes=11475
  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)=85255
  CPU time spent (ms)=1333390
  Physical memory (bytes) snapshot=19123703808
  Virtual memory (bytes) snapshot=144771522560
  Total committed heap usage (bytes)=15593897984
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=10000000000

**Total Time taken for Execution: 20.18 minutes**

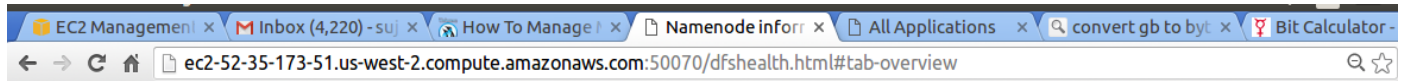
```
ubuntu@ip-172-31-14-237: ~  
Total time spent by all reduces in occupied slots (ms)=983744  
Total time spent by all map tasks (ms)=4980632  
Total time spent by all reduce tasks (ms)=983744  
Total vcore-milliseconds taken by all map tasks=4980632  
Total vcore-milliseconds taken by all reduce tasks=983744  
Total megabyte-milliseconds taken by all map tasks=5100167168  
Total megabyte-milliseconds taken by all reduce tasks=1007353856  
Map-Reduce Framework  
Map input records=100000000  
Map output records=100000000  
Map output bytes=1000000000  
Map output materialized bytes=10200000450  
Input split bytes=11475  
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)=85255  
CPU time spent (ms)=1333390  
Physical memory (bytes) snapshot=19123703808  
Virtual memory (bytes) snapshot=144771522560  
Total committed heap usage (bytes)=15593897984  
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=10000000000
```

**Final Output on terminal Hadoop single node:-**

Valsort output Hadoop single node:-

```
ubuntu@ip-172-31-14-237: /sujay/data/64
ubuntu@ip-172-31-14-237:~/hadoop-2.7.2$ cd /sujay
ubuntu@ip-172-31-14-237:/sujay$ ls
data dfs lost+found nm-local-dir part-r-00000
ubuntu@ip-172-31-14-237:/sujay$ pwd
/sujay
ubuntu@ip-172-31-14-237:/sujay$ cd data
ubuntu@ip-172-31-14-237:/sujay/data$ ls
32 64 gensort-linux-1.5.tar.gz gpl-2.0.txt
ubuntu@ip-172-31-14-237:/sujay/data$ cd 64
ubuntu@ip-172-31-14-237:/sujay/data/64$ ls
10GB_input_file gensort valsart
ubuntu@ip-172-31-14-237:/sujay/data/64$ ./valsart /sujay/part-r-00000
Records: 100000000
Checksum: 2fb0574596d67c8
Duplicate keys: 0
SUCCESS - all records are in order
ubuntu@ip-172-31-14-237:/sujay/data/64$
```

## Hadoop 100GB 16 Nodes Execution Name node:-



Started:	Tue Mar 29 17:47:20 UTC 2016
Version:	2.7.2, rb165c4fe8a74265c792ce23f546c54604ac0e41
Compiled:	2016-01-26T00:08Z by jenkins from (detached from b165c4f)
Cluster ID:	CID-caa2cba8-71a1-4d24-8935-7fbddcfe21c
Block Pool ID:	BP-450675934-172.31.6.180-1459260733017

### Summary

Security is off.

SafeMode is off.

21 files and directories, 752 blocks = 773 total filesystem object(s).

Heap Memory used 87.97 MB of 176.5 MB Heap Memory. Max Heap Memory is 889 MB.

Non Heap Memory used 51.53 MB of 52.4 MB Committed Non Heap Memory. Max Non Heap Memory is -1 B.

Configured Capacity:	4.9 TB
DFS Used:	93.91 GB (1.87%)
Non DFS Used:	335.72 GB
DFS Remaining:	4.48 TB (91.44%)
Block Pool Used:	93.91 GB (1.87%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 31.82% / 7.49%
<a href="#">Live Nodes</a>	17 (Decommissioned: 0)
<a href="#">Dead Nodes</a>	0 (Decommissioned: 0)
<a href="#">Decommissioning Nodes</a>	0
<a href="#">Total Datanode Volume Failures</a>	0 (0 B)
<a href="#">Number of Under-Replicated Blocks</a>	0
<a href="#">Number of Blocks Pending Deletion</a>	0
<a href="#">Block Deletion Start Time</a>	3/29/2016, 11:17:20 PM

## Hadoop 16 Nodes execution completion Screenshot:-

←

→

↺

🏠

ec2-52-35-173-51.us-west-2.compute.amazonaws.com:9006/cluster

🔍

☆


📄

👤

🖨

☰

Logged in as: dr.who



# All Applications

Cluster

About

Nodes

Node Labels

Applications

NEW

NEW SAVING

SUBMITTED

ACCEPTED

RUNNING

FINISHED

FAILED

KILLED

Scheduler

Tools

Cluster Metrics

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
2	0	0	2	0	0 B	128 GB	0 B	0	128	0	16	0	1	0	0

Scheduler Metrics

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

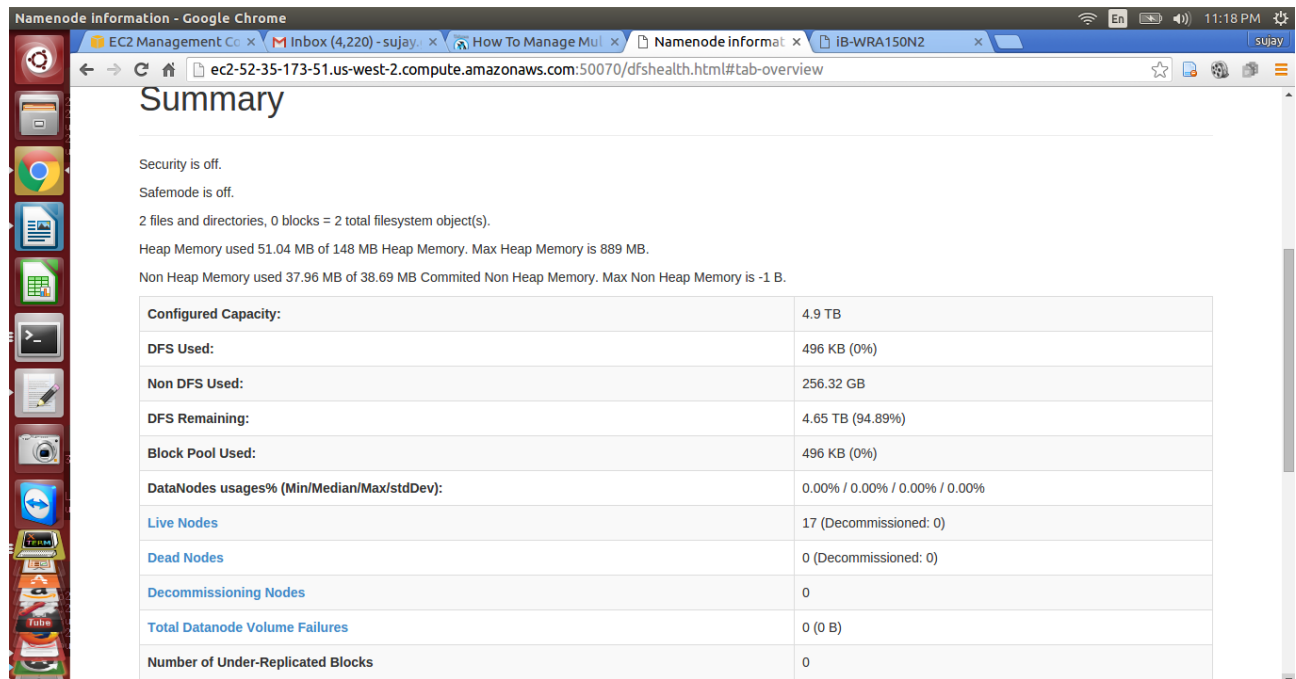
Show 20 entries

ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Progress	Tracking UI	Blacklisted Nodes
application_1459273697028_0001	ubuntu	Tera Sort	MAPREDUCE	default	Wed Mar 30 00:25:16 +0550 2016	Wed Mar 30 03:05:45 +0550 2016	FINISHED	SUCCEEDED	<div></div>	<a href="#">History</a>	N/A
application_1459273697028_0002	ubuntu	Tera Sort	MAPREDUCE	default	Wed Mar 30 01:33:58 +0550 2016	Wed Mar 30 04:06:06 +0550 2016	FINISHED	SUCCEEDED	<div></div>	<a href="#">History</a>	N/A

Showing 1 to 2 of 2 entries

First Previous 1 Next Last

Name node one more screenshot:-

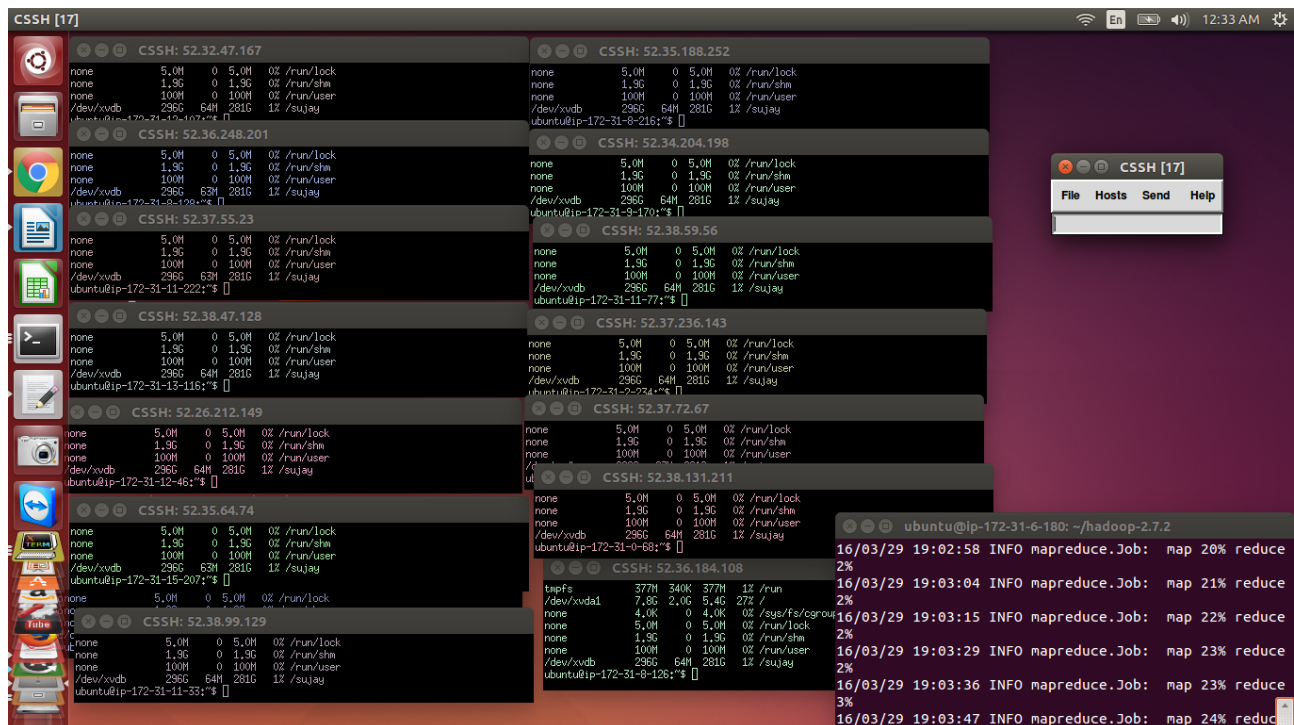


Summary

Security is off.  
Safemode is off.  
2 files and directories, 0 blocks = 2 total filesystem object(s).  
Heap Memory used 51.04 MB of 148 MB Heap Memory. Max Heap Memory is 889 MB.  
Non Heap Memory used 37.96 MB of 38.69 MB Committed Non Heap Memory. Max Non Heap Memory is -1 B.

Configured Capacity:	4.9 TB
DFS Used:	496 KB (0%)
Non DFS Used:	256.32 GB
DFS Remaining:	4.65 TB (94.89%)
Block Pool Used:	496 KB (0%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
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

Hadoop 16 Nodes execution on Ubuntu screenshot:-



CSSH [17]

CSSH: 52.32.47.167

```
none 5.0M 0 5.0M 0Z /run/lock
none 1.9G 0 1.9G 0Z /run/shm
none 100M 0 100M 0Z /run/user
/dev/xvdb 296G 64M 281G 1Z /sujoy
ubuntu@ip-172-31-15-102:~$
```

CSSH: 52.36.248.201

```
none 5.0M 0 5.0M 0Z /run/lock
none 1.9G 0 1.9G 0Z /run/shm
none 100M 0 100M 0Z /run/user
/dev/xvdb 296G 64M 281G 1Z /sujoy
ubuntu@ip-172-31-15-108:~$
```

CSSH: 52.37.55.23

```
none 5.0M 0 5.0M 0Z /run/lock
none 1.9G 0 1.9G 0Z /run/shm
none 100M 0 100M 0Z /run/user
/dev/xvdb 296G 64M 281G 1Z /sujoy
ubuntu@ip-172-31-11-222:~$
```

CSSH: 52.38.47.128

```
none 5.0M 0 5.0M 0Z /run/lock
none 1.9G 0 1.9G 0Z /run/shm
none 100M 0 100M 0Z /run/user
/dev/xvdb 296G 64M 281G 1Z /sujoy
ubuntu@ip-172-31-13-116:~$
```

CSSH: 52.26.212.149

```
none 5.0M 0 5.0M 0Z /run/lock
none 1.9G 0 1.9G 0Z /run/shm
none 100M 0 100M 0Z /run/user
/dev/xvdb 296G 64M 281G 1Z /sujoy
ubuntu@ip-172-31-12-48:~$
```

CSSH: 52.35.64.74

```
none 5.0M 0 5.0M 0Z /run/lock
none 1.9G 0 1.9G 0Z /run/shm
none 100M 0 100M 0Z /run/user
/dev/xvdb 296G 64M 281G 1Z /sujoy
ubuntu@ip-172-31-15-207:~$
```

CSSH: 52.38.99.129

```
none 5.0M 0 5.0M 0Z /run/lock
none 1.9G 0 1.9G 0Z /run/shm
none 100M 0 100M 0Z /run/user
/dev/xvdb 296G 64M 281G 1Z /sujoy
ubuntu@ip-172-31-11-33:~$
```

CSSH: 52.35.188.252

```
none 5.0M 0 5.0M 0Z /run/lock
none 1.9G 0 1.9G 0Z /run/shm
none 100M 0 100M 0Z /run/user
/dev/xvdb 296G 64M 281G 1Z /sujoy
ubuntu@ip-172-31-8-218:~$
```

CSSH: 52.34.204.198

```
none 5.0M 0 5.0M 0Z /run/lock
none 1.9G 0 1.9G 0Z /run/shm
none 100M 0 100M 0Z /run/user
/dev/xvdb 296G 64M 281G 1Z /sujoy
ubuntu@ip-172-31-9-170:~$
```

CSSH: 52.38.59.56

```
none 5.0M 0 5.0M 0Z /run/lock
none 1.9G 0 1.9G 0Z /run/shm
none 100M 0 100M 0Z /run/user
/dev/xvdb 296G 64M 281G 1Z /sujoy
ubuntu@ip-172-31-11-77:~$
```

CSSH: 52.37.236.143

```
none 5.0M 0 5.0M 0Z /run/lock
none 1.9G 0 1.9G 0Z /run/shm
none 100M 0 100M 0Z /run/user
/dev/xvdb 296G 64M 281G 1Z /sujoy
ubuntu@ip-172-31-2-234:~$
```

CSSH: 52.37.72.67

```
none 5.0M 0 5.0M 0Z /run/lock
none 1.9G 0 1.9G 0Z /run/shm
none 100M 0 100M 0Z /run/user
/dev/xvdb 296G 64M 281G 1Z /sujoy
ubuntu@ip-172-31-2-234:~$
```

CSSH: 52.38.131.211

```
none 5.0M 0 5.0M 0Z /run/lock
none 1.9G 0 1.9G 0Z /run/shm
none 100M 0 100M 0Z /run/user
/dev/xvdb 296G 64M 281G 1Z /sujoy
ubuntu@ip-172-31-0-68:~$
```

ubuntu@ip-172-31-6-180: ~/hadoop-2.7.2

```
16/03/29 19:02:58 INFO mapreduce.Job: map 20% reduce 2%
16/03/29 19:03:04 INFO mapreduce.Job: map 21% reduce 2%
16/03/29 19:03:15 INFO mapreduce.Job: map 22% reduce 2%
16/03/29 19:03:29 INFO mapreduce.Job: map 23% reduce 2%
16/03/29 19:03:36 INFO mapreduce.Job: map 23% reduce 3%
16/03/29 19:03:47 INFO mapreduce.Job: map 24% reduce 3%
```

## Hadoop 16 Nodes execution screen shot of AWS:-

EC2 Management Console - Google Chrome

EC2 Management | Inboxes (4,220) - Sujay | How To Manage | Namenode info | All Applications | convert gb to byt | Bit Calculator - Cr | Sujay

https://us-west-2.console.aws.amazon.com/ec2/v2/home?region=us-west-2#Instances:sort=securityGroupNames

AWS Services Edit

Launch Instance Connect Actions

Filter by tags and attributes or search by keyword

	Name	Instance ID	Instance Type	Availability Zone	Instance State	Status Checks	Alarm Status	Public DNS
<input type="checkbox"/>	slave4	i-c7101010	c3.large	us-west-2c	running	2/2 checks ...	None	ec2-52-30-227-157.us-west-2.amazonaws.com
<input type="checkbox"/>	slave5	i-d8383e02	c3.large	us-west-2c	running	2/2 checks ...	None	ec2-52-35-188-252.us-west-2.amazonaws.com
<input type="checkbox"/>	slave8	i-81c3c55b	c3.large	us-west-2c	running	2/2 checks ...	None	ec2-52-38-59-56.us-west-2.amazonaws.com
<input type="checkbox"/>	slave9	i-33c0c6e9	c3.large	us-west-2c	running	2/2 checks ...	None	ec2-52-37-55-23.us-west-2.amazonaws.com
<input type="checkbox"/>	slave10	i-b4c1c76e	c3.large	us-west-2c	running	2/2 checks ...	None	ec2-52-38-99-129.us-west-2.amazonaws.com
<input type="checkbox"/>	slave6	i-8ec3c554	c3.large	us-west-2c	running	2/2 checks ...	None	ec2-52-38-35-237.us-west-2.amazonaws.com
<input type="checkbox"/>	slave7	i-30c0c6ea	c3.large	us-west-2c	running	2/2 checks ...	None	ec2-52-38-47-128.us-west-2.amazonaws.com
<input type="checkbox"/>	slave11	i-2bde8f1	c3.large	us-west-2c	running	2/2 checks ...	None	ec2-52-26-212-149.us-west-2.amazonaws.com
<input type="checkbox"/>	slave12	i-50df98a	c3.large	us-west-2c	running	2/2 checks ...	None	ec2-52-37-72-67.us-west-2.amazonaws.com
<input type="checkbox"/>	slave13	i-e8ded832	c3.large	us-west-2c	running	2/2 checks ...	None	ec2-52-38-131-211.us-west-2.amazonaws.com
<input type="checkbox"/>	slave14	i-23e2e4f9	c3.large	us-west-2c	running	2/2 checks ...	None	ec2-52-34-204-198.us-west-2.amazonaws.com
<input type="checkbox"/>	slave15	i-cfe5e315	c3.large	us-west-2c	running	2/2 checks ...	None	ec2-52-37-236-143.us-west-2.amazonaws.com
<input type="checkbox"/>	slave16	i-ffe2e425	c3.large	us-west-2c	running	2/2 checks ...	None	ec2-52-36-184-108.us-west-2.amazonaws.com

Select an instance above

© 2008 - 2016, Amazon Internet Services Private Ltd. or its affiliates. All rights reserved. Privacy Policy Terms of Use



## Hadoop 100GB tail output:-

```
ubuntu@ip-172-31-6-180: /sujay
ubuntu@ip-172-31-6-180:/sujay$ tail -10 part-r-00000
~~~~~#1ay1X      000000000000000000000000000000000025D35EDF  6666AAAA5555999977770000222233338888FFFF999922220000
~~~~~+@p){@      0000000000000000000000000000000000085426F4  77773333555511111110000CCCC55559999AAAA7777DDDDDDDD
~~~~~,R^_?n      00000000000000000000000000000000000034E347  111111119999000011118888AAAA55554444EEEE999933338888
~~~~~,Ey`^      000000000000000000000000000000000016F0E66B  CCCC6666DDDD2222DDDD111188889999EEEEEEEEEEEEBBB84444
~~~~~4!kA7x      00000000000000000000000000000000001F1A1E26  EEEE777711117777BBB81111EEEE88884444DDDDDDDEEEEBBB8
~~~~~8Ii/!@      000000000000000000000000000000000001F05932F  11119999BBB844447777000011114444CCCCAAAA6666DDDD0000
~~~~~<I' 5>F      000000000000000000000000000000000008CB2293  88883333BBB8111166669999888855558888888822228888CCCC
~~~~~G-)m^      0000000000000000000000000000000000013397F73  DDDDDFFFB8B8CCCCFF44446666AAAA111133333333AAACCCC
~~~~~c+I&CP      0000000000000000000000000000000000074BD6F4  8888000055550000DDDD22227777AAAA000033332222AAAADDDD
~~~~~hb&5X*      0000000000000000000000000000000000032C0E06B  7777BBB8BBB89999EEEEAAAAAA0000CCCD0DD4444BBB84444
ubuntu@ip-172-31-6-180:/sujay$
```

## Hadoop 100GB head Output:-

```
ubuntu@ip-172-31-6-180: /sujay
ubuntu@ip-172-31-6-180:/sujay$ head -10 part-r-00000
!4+ABv      000000000000000000000000000000000017F7E829  EEEE3333444411112222888833334444666633332222DDDDDEEE
"!0!uve      000000000000000000000000000000000001228D4  77778888000022224444DDDDDDDEEE00000000CCCC7777DDDD
%!$sU(      000000000000000000000000000000000002E6C821C  2222333377774444555511119999CCCC4444EEEEFFFF11115555
&5rX|X      000000000000000000000000000000000000399BC288  5555CCCCBBB899999999DDDD111100001111EEEE7777DDDD9999
'!c%So      0000000000000000000000000000000000031F06B7D  EEEEBBB8AAAA8888DDDDDDDD777722224444111166664444AAAA
*0G1Io      0000000000000000000000000000000000003B5E85A1  1111AAAA9999CCCCBBB8111199991111333399991111AAAA6666
,(GhT_      0000000000000000000000000000000000002D0172DC  1111CCCC1111DDDDCCCCEEEE9999CCCC8888CCCCFFFF55555555
0*vYm3      00000000000000000000000000000000000026D61578  DDDD7777AAAAEEEEEEEE6666AAAA2222CCCC5555555522229999
2C>)8d      00000000000000000000000000000000000026C79E66  444400001111CCCC6666BBB8555577776666CCCC2222AAAABBB8
PMd32=      0000000000000000000000000000000000003440CC1  FFFFE6666CCCCBBB8999933335555DDDDDDDD777788886666
ubuntu@ip-172-31-6-180:/sujay$
```

## Spark execution screen-shots:-

### Spark 10Gb Execution report:-

Spark Master at spark://ec2-52-91-128-142.compute-1.amazonaws.com:7077 - Google Chrome

EC2 Management Co x Inbox (4,242) - sujay x Spark Master at spar x

ec2-52-91-128-142.compute-1.amazonaws.com:8080

### Spark 1.6.0 Spark Master at spark://ec2-52-91-128-142.compute-1.amazonaws.com:7077

URL: spark://ec2-52-91-128-142.compute-1.amazonaws.com:7077  
REST URL: spark://ec2-52-91-128-142.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, 2 Completed  
Drivers: 0 Running, 0 Completed  
Status: ALIVE

#### Workers

Worker Id	Address	State	Cores	Memory
worker-20160401143549-172.31.47.12-52483	172.31.47.12:52483	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-20160401151000-0001	Spark shell	2	2.4 GB	2016/04/01 15:10:00	root	FINISHED	16 min

## Spark Single node AWS screenshot:-

EC2 Dashboard  
Events  
Tags  
Reports  
Limits

INSTANCES

- Instances
- Spot Requests
- Reserved Instances
- Scheduled Instances
- Commands
- Dedicated Hosts

IMAGES

- AMIs
- Bundle Tasks

ELASTIC BLOCK STORE

- Volumes
- Snapshots

NETWORK & SECURITY

- Security Groups
- Elastic IPs
- Placement Groups

Launch Instance Connect Actions

Filter by tags and attributes or search by keyword

Name	Instance ID	Instance Type	Availability Zone	Instance State	Status Checks	Alarm Status	Public DNS
spark-master-i-1a3832e3	i-1a3832e3	c3.large	us-east-1e	running	2/2 checks ...	None	ec2-52-91-128-142.compute-1.amazonaws.com
spark-slave-i-1c3832e5	i-1c3832e5	c3.large	us-east-1e	running	2/2 checks ...	None	ec2-52-87-...

Instance: **i-1a3832e3 (spark-master-i-1a3832e3)** Public DNS: **ec2-52-91-128-142.compute-1.amazonaws.com**

Description Status Checks Monitoring Tags

Instance ID	Public DNS
i-1a3832e3	ec2-52-91-128-142.compute-1.amazonaws.com

Instance state	Public IP
running	52.91.128.142

Instance type	Elastic IP
c3.large	-

Private DNS	Availability zone
ip-172-31-47-234.ec2.internal	us-east-1e

Private IPs	Security groups
172.31.47.234	spark-master, view rules

Secondary private IPs	Scheduled events
-	No scheduled events

VPC ID	AMI ID
vpc-3836ba5c	spark.ami.pvm.v9 (ami-5bb18832)

Subnet ID	Platform
subnet-2160d01c	-

Network interfaces	IAM role
eth0	-

SourceDest. check	Key pair name
True	sparkeast

Owner
711023178553

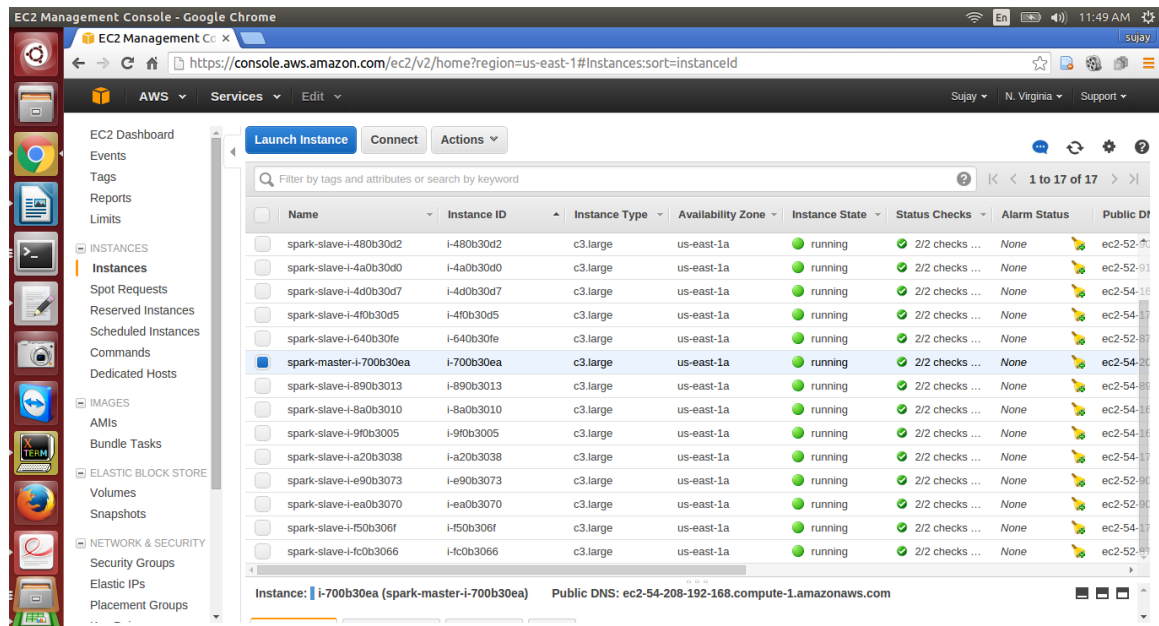
Feedback English © 2008 - 2016, Amazon Internet Services Private Ltd. or its affiliates. All rights reserved. Privacy Policy Terms of Use

## Spark 10GB Valsort output:-

```
sujay@sujay-Studio-1558: ~/Desktop/PA2/spark-1.6.0-bin-hadoop2.6/ec2
root@ip-172-31-47-234 sujay]$ cd 64
root@ip-172-31-47-234 64]$ ./valsort /sujay/10GB_final_spark_sorted_file
Records: 100000000
Checksum: 2fae59101920038
Duplicate keys: 0
SUCCESS - all records are in order
root@ip-172-31-47-234 64]$
```

[illegible]

## Spark 100 Gb 16 nodes AWS screen-shot:-



EC2 Management Console - Google Chrome

https://console.aws.amazon.com/ec2/v2/home?region=us-east-1#Instances:sort=instancetype

AWS Services Edit

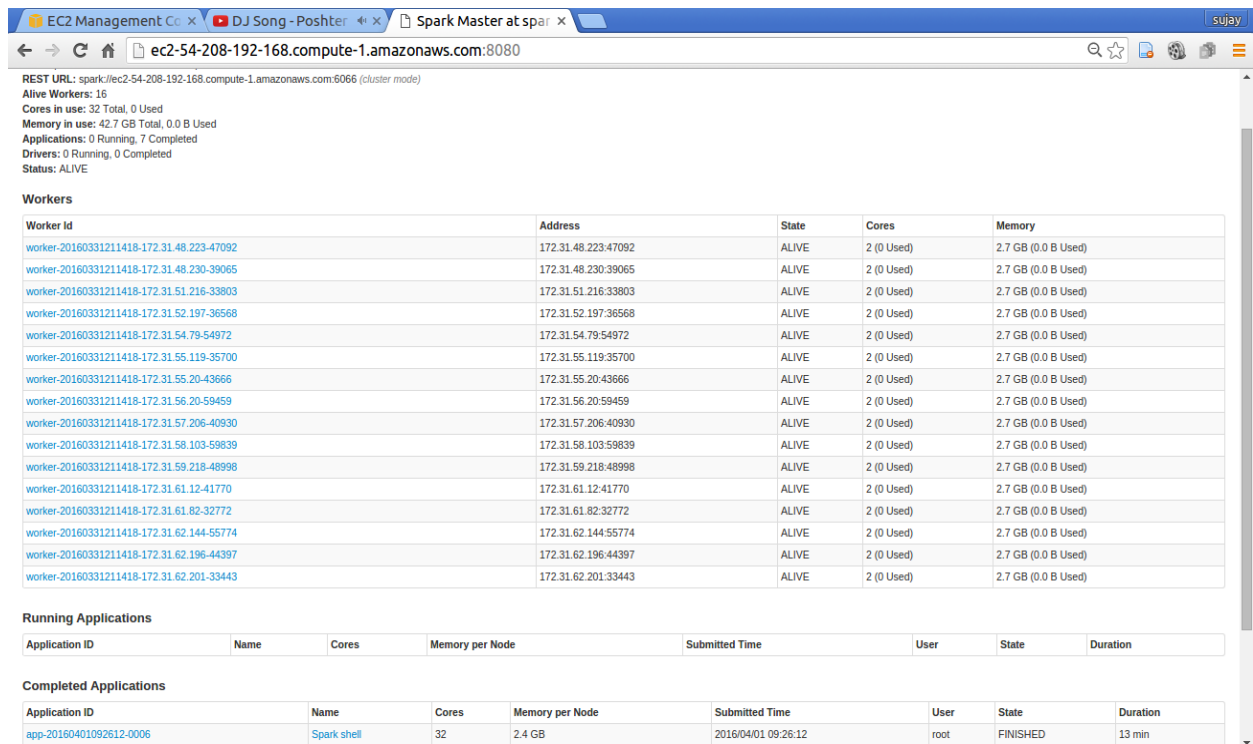
Launch Instance Connect Actions

Filter by tags and attributes or search by keyword

Name	Instance ID	Instance Type	Availability Zone	Instance State	Status Checks	Alarm Status	Public DNS
spark-slave-i-480b30d2	i-480b30d2	c3.large	us-east-1a	running	2/2 checks ...	None	ec2-52-208-192-168.compute-1.amazonaws.com
spark-slave-i-4a0b30d0	i-4a0b30d0	c3.large	us-east-1a	running	2/2 checks ...	None	ec2-52-208-192-168.compute-1.amazonaws.com
spark-slave-i-4d0b30d7	i-4d0b30d7	c3.large	us-east-1a	running	2/2 checks ...	None	ec2-54-187-192-168.compute-1.amazonaws.com
spark-slave-i-4f0b30d5	i-4f0b30d5	c3.large	us-east-1a	running	2/2 checks ...	None	ec2-54-187-192-168.compute-1.amazonaws.com
spark-slave-i-640b30fe	i-640b30fe	c3.large	us-east-1a	running	2/2 checks ...	None	ec2-52-208-192-168.compute-1.amazonaws.com
spark-master-i-700b30ea	i-700b30ea	c3.large	us-east-1a	running	2/2 checks ...	None	ec2-54-208-192-168.compute-1.amazonaws.com
spark-slave-i-890b3013	i-890b3013	c3.large	us-east-1a	running	2/2 checks ...	None	ec2-54-187-192-168.compute-1.amazonaws.com
spark-slave-i-8a0b3010	i-8a0b3010	c3.large	us-east-1a	running	2/2 checks ...	None	ec2-54-187-192-168.compute-1.amazonaws.com
spark-slave-i-9f0b3005	i-9f0b3005	c3.large	us-east-1a	running	2/2 checks ...	None	ec2-54-187-192-168.compute-1.amazonaws.com
spark-slave-i-a20b3038	i-a20b3038	c3.large	us-east-1a	running	2/2 checks ...	None	ec2-54-187-192-168.compute-1.amazonaws.com
spark-slave-i-e90b3073	i-e90b3073	c3.large	us-east-1a	running	2/2 checks ...	None	ec2-52-208-192-168.compute-1.amazonaws.com
spark-slave-i-ea0b3070	i-ea0b3070	c3.large	us-east-1a	running	2/2 checks ...	None	ec2-52-208-192-168.compute-1.amazonaws.com
spark-slave-i-f50b306f	i-f50b306f	c3.large	us-east-1a	running	2/2 checks ...	None	ec2-54-187-192-168.compute-1.amazonaws.com
spark-slave-i-fc0b3066	i-fc0b3066	c3.large	us-east-1a	running	2/2 checks ...	None	ec2-52-208-192-168.compute-1.amazonaws.com

Instance: i-700b30ea (spark-master-i-700b30ea) Public DNS: ec2-54-208-192-168.compute-1.amazonaws.com

## Spark 100 Gb 16 nodes completion screen-shot:-



EC2 Management Console - Google Chrome

Spark Master at spark-master-i-700b30ea

ec2-54-208-192-168.compute-1.amazonaws.com:8080

REST URL: spark://ec2-54-208-192-168.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, 7 Completed

Drivers: 0 Running, 0 Completed

Status: ALIVE

### Workers

Worker Id	Address	State	Cores	Memory
worker-20160331211418-172.31.48.223-47092	172.31.48.223-47092	ALIVE	2 (0 Used)	2.7 GB (0.0 B Used)
worker-20160331211418-172.31.48.230-39065	172.31.48.230-39065	ALIVE	2 (0 Used)	2.7 GB (0.0 B Used)
worker-20160331211418-172.31.51.216-33803	172.31.51.216-33803	ALIVE	2 (0 Used)	2.7 GB (0.0 B Used)
worker-20160331211418-172.31.52.197-36568	172.31.52.197-36568	ALIVE	2 (0 Used)	2.7 GB (0.0 B Used)
worker-20160331211418-172.31.54.79-54972	172.31.54.79-54972	ALIVE	2 (0 Used)	2.7 GB (0.0 B Used)
worker-20160331211418-172.31.55.119-35700	172.31.55.119-35700	ALIVE	2 (0 Used)	2.7 GB (0.0 B Used)
worker-20160331211418-172.31.55.20-43666	172.31.55.20-43666	ALIVE	2 (0 Used)	2.7 GB (0.0 B Used)
worker-20160331211418-172.31.56.20-59459	172.31.56.20-59459	ALIVE	2 (0 Used)	2.7 GB (0.0 B Used)
worker-20160331211418-172.31.57.206-40930	172.31.57.206-40930	ALIVE	2 (0 Used)	2.7 GB (0.0 B Used)
worker-20160331211418-172.31.58.103-59839	172.31.58.103-59839	ALIVE	2 (0 Used)	2.7 GB (0.0 B Used)
worker-20160331211418-172.31.59.218-48998	172.31.59.218-48998	ALIVE	2 (0 Used)	2.7 GB (0.0 B Used)
worker-20160331211418-172.31.61.12-41770	172.31.61.12-41770	ALIVE	2 (0 Used)	2.7 GB (0.0 B Used)
worker-20160331211418-172.31.61.82-32772	172.31.61.82-32772	ALIVE	2 (0 Used)	2.7 GB (0.0 B Used)
worker-20160331211418-172.31.62.144-55774	172.31.62.144-55774	ALIVE	2 (0 Used)	2.7 GB (0.0 B Used)
worker-20160331211418-172.31.62.196-44397	172.31.62.196-44397	ALIVE	2 (0 Used)	2.7 GB (0.0 B Used)
worker-20160331211418-172.31.62.201-33443	172.31.62.201-33443	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-20160401092612-0006	Spark shell	32	2.4 GB	2016/04/01 09:26:12	root	FINISHED	13 min

```
sujay@sujay-Studio-1558: ~/Desktop/PA2/spark-1.6.0-bin-hadoop2.6/ec2
root@ip-172-31-59-170 sujay$ head -10 100GB_final_spark_sorted_file
!4+ABv      000000000000000000000000000000000017F7E829    EEEE33334444111222288883333444466663333222DDDEEEE
!0luve      00000000000000000000000000000000001228D4    77778888000002224444DDDDDDDEEE00000000CCCC7777DDDD
%!$sU(      00000000000000000000000000000000002E6C821C    2222333377774444555511119999CCCC4444EEEEFFFF1115555
&5rX|X      000000000000000000000000000000000039B8C288    5555CCCCBB889999999DDDD111100001111EEEE7777DDDD9999
'ic$So      000000000000000000000000000000000031F06B7D    EEEEBBBBAAA8888DDDDDDDD777722224444111166664444AAAA
*0GIo       00000000000000000000000000000000003BS85A1    1111AAAA9999CCCCBB88111199991111333399991111AAAA6666
,(GHT_      00000000000000000000000000000000002D0172DC    1111CCCC1111DDDDCCCEEE9999CCC8888CCCCFF55555555
0*vYm3      000000000000000000000000000000000026D61578    DDDD7777AAAAEEEEEEEE6666AAAA2222CCCC555555552229999
2C>J8d      000000000000000000000000000000000026C79E66    444400001111CCCC6666BBB555577776666CCCC2222AAABBBB
PMD3z=      00000000000000000000000000000000003440CC1    FFFFEEEE6666CCCCBB8899993335555DDDDDDDD777788886666

root@ip-172-31-59-170 sujay$ tail -10 100GB_final_spark_sorted_file
~~~#layIX     000000000000000000000000000000000025D3SEDF    6666AAAA5555999977770000222233338888FFFF999922220000
~~~~+@p_|@   0000000000000000000000000000000000085426F4    77773333555511111110000CCCC55559999AAAA7777DDDDDDDD
~~~~,R^_?n    00000000000000000000000000000000001034E347    1111111199999000011118888AAAA55554444EEEE999933338888
~~~~,Ey_^)    000000000000000000000000000000000016F0E66B    CCCC6666DDDD2222DDDD111188889999EEEEEEEEEEEEEBBBB4444
~~~~4lKaTx    00000000000000000000000000000000001F1A1E26    EEEE777711117777BBBB1111EEEE88884444DDDDDDDDDEEEEBBBB
~~~~8I!/@     00000000000000000000000000000000000F05932F    11119999BBBB44447777000011114444CCCCAAAA6666DDDD0000
~~~~<-I>5>F    000000000000000000000000000000000008CB293    88883333BBB11116666999988885555888888822228888CCCC
~~~~-G-)m^    000000000000000000000000000000000013397F73    DDDDDFFFB BBBCCCCFFFF44446666AAAA111133333333AAAACCCC
~~~~c+T&cP     0000000000000000000000000000000000074BDf64    8888000055550000DDDD22227777AAAA000033332222AAADDDD
~~~~hb&5X*    000000000000000000000000000000000032C0E06B    7777BBBBBBBB9999EEEEAAAAAAAA0000CCCCDDDD4444BBBB4444

root@ip-172-31-59-170 sujay$
```

```

root@sujoy@sujoy-Studio-1558: ~/Desktop/PA2
https://aws.amazon.com/amazon-linux-ami/2013.03-release-notes/
Amazon Linux version 2016.03 is available.
root@ip-172-31-59-170 ~]$ cd /sujoy/64
root@ip-172-31-59-170 64]$ ;s
-bash: syntax error near unexpected token `;'
root@ip-172-31-59-170 64]$ ;s
-bash: syntax error near unexpected token `;'
root@ip-172-31-59-170 64]$ ls
100GB_input_file  gensort  valsort
root@ip-172-31-59-170 64]$ ./valsort /sujoy/100GB_final_spark_sorted_file

Records: 1000000000
Checksum: 1dcd7de9014883a9
Duplicate keys: 0
SUCCESS - all records are in order
root@ip-172-31-59-170 64]$
root@ip-172-31-59-170 64]$

```