# Assignment 2

**Problem:**

We are sorting large dataset using Shared Memory, Hadoop and Spark. We are sorting two files one of 128GB and one of 1TB so that we can perform benchmarking on the three methods.

**Methodology:**

1. **Shared Memory Sort**

- We are sorting 128 GB and 1 TB files using external sort. We are using merge sort algorithm for the same.
- Since the data is large we won't be able to sort this data in memory. So, we divide the data into small chunks that will fit into the memory and sort them using merge sort.
- We then write the sorted chunks into files. We have calculated the chunk size by calculating the free memory in MB and dividing it by 100. So, depending on the free memory of instance we create the number of chunk size.
- Sorting is performed based on the ascii value.
- So, take the file read it up to the chunk size. Store the data in to the array. Sort the array using merge sort and write these data into the file. Then read the next chunk sort and store. This will happen till we reach the length of the input file.
- After sorting all the files and writing into the chunks we merge these chunks into a single sorted output file.

2. **Hadoop**

- We have written a Java Map-Reduce program to take file as input, read the line and divide it into two parts key and value where key consist of first 10 characters of the line and rest are value.
- This is given to the mapper which counts the keys. Then is shuffled by shuffler so that same keys belong to the same bucket. The reducer then sorts the keys and returns the sorted result.
- We are using 128 GB and 1 TB data and running Hadoop on i3large and i3.4xlarge instances respectively and tracking the time to sort the data.

3. **Spark**

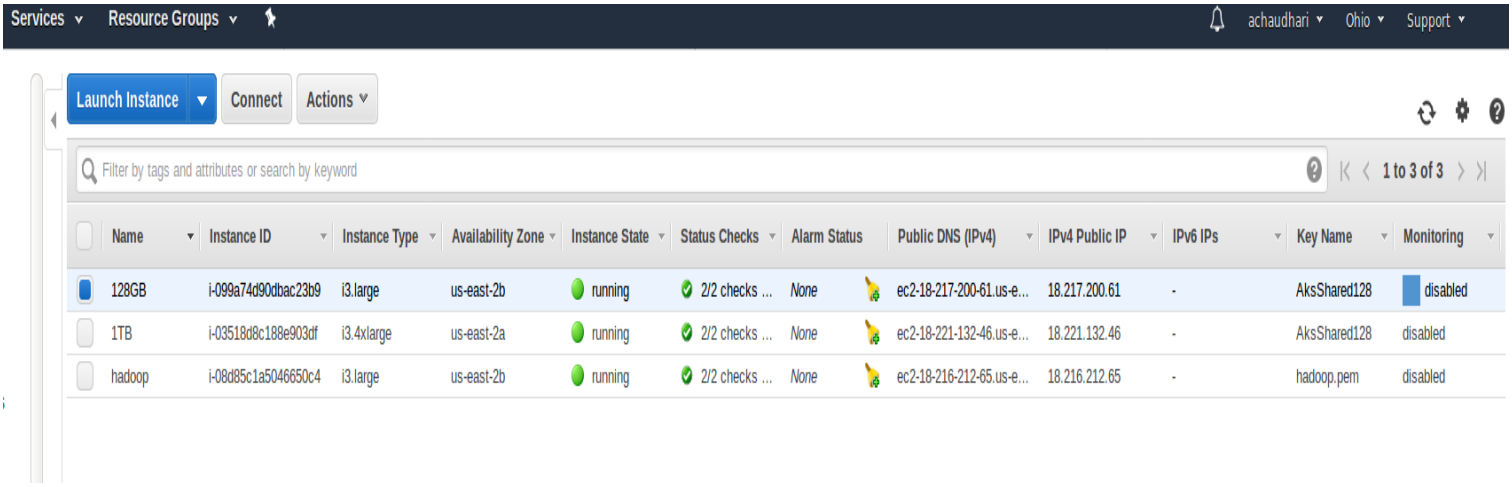
   Similar to Hadoop we configure the Spark and wrote the code in Scala. We take just the keys from the lines which consist of first 10 characters and drop the other part.
   We sort the file using sort_by_key() method and using the

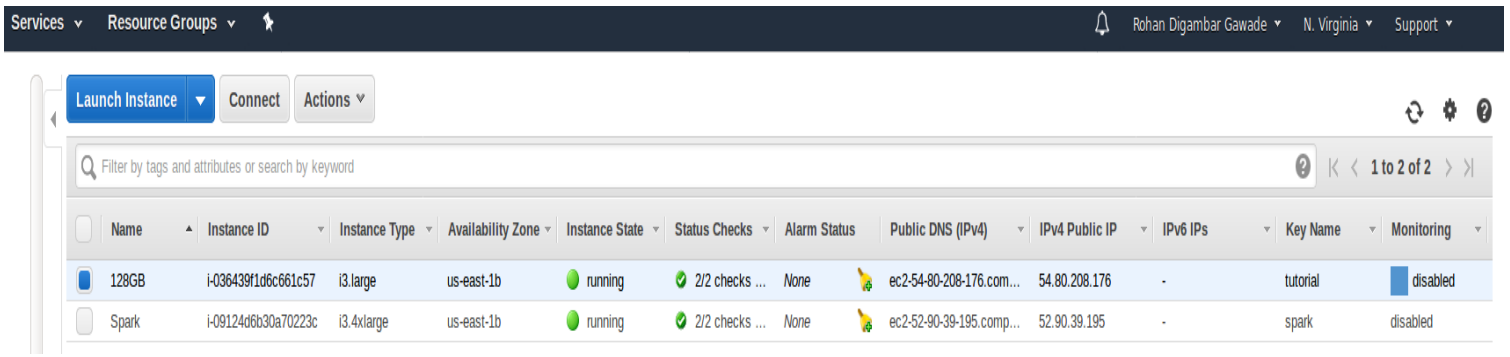
## Instance Amazon:

We are using i3.large and i3.4xlarge instance for 128GB and 1TB data respectively. The configuration for the above instances are as follows:

i3.large and i3.4xlarge configuration

| Family | Type | vCPUs | Memory (GiB) | Instance Storage (GB) |
|---|---|---|---|---|
| Storage optimized | i3.large | 2 | 15.25 | 1 x 475 (SSD) |
| Storage optimized | i3.4xlarge | 16 | 122 | 2 x 1900 (SSD) |



# Virtual Cluster (1-Node):

## Shared Memory:

The instance disk storage is just 8 GB so we need to mount the RAID to store the input file of 128GB.
I ran for 2 threads.
We created 128 GB file using gensort. Each line in gensort is of 100 bytes size.

## Commands

**The following commands are used to mount the external drive**

```
lsblk
sudo mke2fs -F -t ext4 /dev/nvme0n1
sudo mkdir /mount1
sudo mount /dev/nvme0n1 /mount1
```

**Copy the gensort file to /mount1 folder using**

```
cp ~/gensort /mount1
```

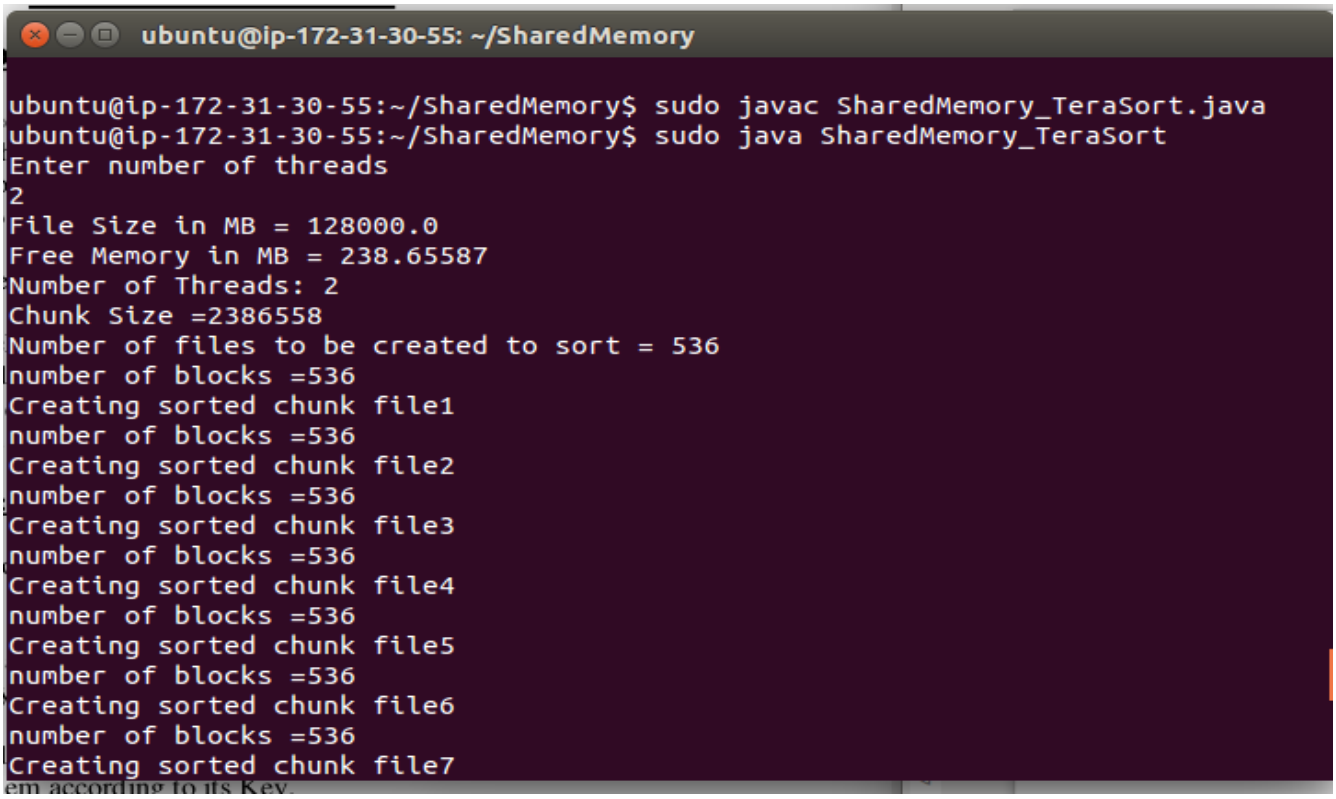
**Run gensort to generate 128GB data**

```
./gensort -a 1280000000 inputfile.txt
```

We created 1280000000 records as each record is for 100bytes
We ran our shared memory program SharedMemory_TeraSort on this 128GB file with 2 threads as there are 2 vCPUs.

The output is as follows:

**Output Screenshots**



```
ubuntu@ip-172-31-30-55: ~/SharedMemory
input_chunk213.txt  input_chunk336.txt  input_chunk459.txt  input_chunk97.txt
input_chunk214.txt  input_chunk337.txt  input_chunk45.txt   input_chunk98.txt
input_chunk215.txt  input_chunk338.txt  input_chunk460.txt  input_chunk99.txt
input_chunk216.txt  input_chunk339.txt  input_chunk461.txt  input_chunk9.txt
input_chunk217.txt  input_chunk33.txt   input_chunk462.txt  inputfile.txt
input_chunk218.txt  input_chunk340.txt  input_chunk463.txt  lost+found
input_chunk219.txt  input_chunk341.txt  input_chunk464.txt  output.txt
input_chunk21.txt   input_chunk342.txt  input_chunk465.txt
input_chunk220.txt  input_chunk343.txt  input_chunk466.txt
input_chunk221.txt  input_chunk344.txt  input_chunk467.txt
ubuntu@ip-172-31-30-55:/mount1$ sudo rm -r input_chunk*.txt
ubuntu@ip-172-31-30-55:/mount1$ ls -l
total 250000180
-rwxr-xr-x 1 root root       141045 Dec  3 02:57 gensort
-rwxr-xr-x 1 root root 128000000000 Dec  3 03:08 inputfile.txt
drwx------ 2 root root        16384 Dec  3 02:44 lost+found
-rw-r--r-- 1 root root 128000000000 Dec  3 06:28 output.txt
ubuntu@ip-172-31-30-55:/mount1$ sudo cp ~/SharedMemory/valsort /mount1/
ubuntu@ip-172-31-30-55:/mount1$ sudo ./valsort output.txt
Records: 1280000000
Checksum: 26258da068f32569
Duplicate keys: 0
SUCCESS - all records are in order
ubuntu@ip-172-31-30-55:/mount1$
```

```
ubuntu@ip-172-31-30-55: ~/SharedMemory
number of blocks =536
Creating sorted chunk file533
number of blocks =536
Creating sorted chunk file534
number of blocks =536
Creating sorted chunk file535
number of blocks =536
Creating sorted chunk file536
number of blocks =536
Creating sorted chunk file537
mergeLength = 1280000000
merge files
Merging: 10 percent
^[[B^[[B^[[BMerging: 20 percent
Merging: 30 percent
Merging: 40 percent
Merging: 50 percent
Merging: 60 percent
Merging: 70 percent
Merging: 80 percent
Merging: 90 percent
100% merged
Total Time Taken to sort = 10948.23 sec
ubuntu@ip-172-31-30-55:~/SharedMemory$
```

**Hadoop (1 node) i3.large:**

We ran Hadoop MapReduce for 128 GB data on i3.large node. Performed similar steps to mount the extra storage as in Shared Memory. Configured Hadoop for one node and run TeraSort_Hadoop and run it using following command.

**Hadoop configuration setting.**

We did the following configuration setting to setup the Hadoop system.

**core-site.xml**

```
<configuration>

<property>
      <name>fs.default.name</name>
      <value>hdfs://ec2-18-216-212-65.us-east-2.compute.amazonaws.com:9000</value>
</property>
<property>
      <name>hadoop.tmp.dir</name>
      <value>/mnt/temp/</value>
</property>
</configuration>
```

**yarn-site.xml**

```
<configuration>

<!-- Site specific YARN configuration properties -->

      <property>
            <name>yarn.nodemanager.aux-services</name>
            <value>mapreduce_shuffle</value>
      </property>
      <property>
      <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
      <value> org.apache.hadoop.mapred.ShuffleHandler</value>
   </property>
      <property>
            <name>yarn.resourcemanager.scheduler.address</name>
            <value>ec2-18-216-212-65.us-east-2.compute.amazonaws.com:8030</value>
      </property>

      </property>
      <property>
            <name>yarn.resourcemanager.address</name>
            <value>ec2-18-216-212-65.us-east-2.compute.amazonaws.com:8032</value>
      </property>
```

```
    </property>
    <property>
        <name>yarn.resourcemanager.address</name>
        <value>ec2-18-216-212-65.us-east-2.compute.amazonaws.com:8032</value>
    </property>
    <property>
        <name>yarn.resourcemanager.webapp.address</name>
        <value>ec2-18-216-212-65.us-east-2.compute.amazonaws.com:8088</value>
    </property>
    <property>
        <name>yarn.resourcemanager.resource-tracker.address</name>
        <value>ec2-18-216-212-65.us-east-2.compute.amazonaws.com:8031</value>
    </property>
    <property>
        <name>yarn.resourcemanager.admin.address</name>
        <value>ec2-18-216-212-65.us-east-2.compute.amazonaws.com:8033</value>
    </property>

</configuration>
```

**mapred-site.xml**

```
<configuration>

<property>
    <name>mapreduce.jobtracker.address</name>
    <value>ec2-18-216-212-65.us-east-2.compute.amazonaws.com:8021</value>
</property>
<property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
</property>
<property>
    <name>mapred.map.tasks</name>
    <value>2</value>
</property>
<property>
    <name>mapred.reduce.tasks</name>
    <value>2</value>
</property>

</configuration>
```

**hdfs-site.xml**

```
<configuration>
<property>
    <name>dfs.replication</name>
    <value>1</value>
```

```
</property>
<property>
        <name>dfs.permissions</name>
        <value>false</value>
</property>
<property>
   <name>dfs.namenode.name.dir</name>
   <value>file:///usr/local/hadoop/hadoop_data/hdfs/namenode</value>
</property>
<property>
   <name>dfs.datanode.data.dir</name>
   <value>file:///usr/local/hadoop/hadoop_store/hdfs/datanode</value>
</property>


</configuration>
```

**We ran our TeraSort_Hadoop.java file using the following commands in Hadoop.**

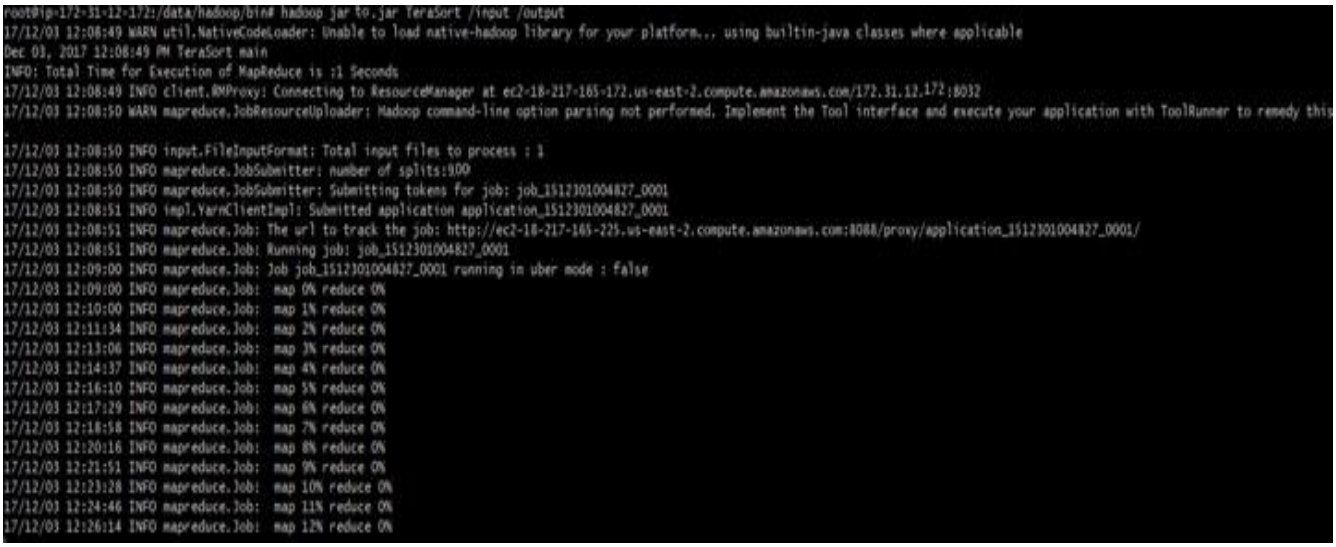
bin/hadoop com.sun.tools.javac.Main TeraSort_Hadoop.java

jar cf ts.jar  TeraSort_Hadoop*.class

hadoop jar ts.jar TeraSort_Hadoop /input /output

hdfs dfs –get /output

**Output:**



```
17/12/03 22:35:30 INFO mapreduce.Job:  map 100% reduce 97%
17/12/03 22:36:30 INFO mapreduce.Job:  map 100% reduce 98%
17/12/03 22:37:30 INFO mapreduce.Job:  map 100% reduce 99%
17/12/03 22:38:30 INFO mapreduce.Job:  map 100% reduce 100%
17/12/03 22:39:03 INFO mapreduce.Job: Job job_1512325297232_0001 completed successfully
17/12/03 22:39:03 INFO mapreduce.Job: Counters: 50
        File System Counters
                FILE: Number of bytes read=597946868118
                FILE: Number of bytes written=728634042772
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=128004059794
                HDFS: Number of bytes written=128000000000
                HDFS: Number of read operations=2803
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Killed map tasks=1
                Launched map tasks=901
                Launched reduce tasks=1
                Data-local map tasks=901
                Total time spent by all maps in occupied slots (ms)=50249794
                Total time spent by all reduces in occupied slots (ms)=12942772
                Total time spent by all map tasks (ms)=50249794
                Total time spent by all reduce tasks (ms)=12942772
                Total vcore-milliseconds taken by all map tasks=50249794
                Total vcore-milliseconds taken by all reduce tasks=12942772
                Total megabyte-milliseconds taken by all map tasks=51449489794
                Total megabyte-milliseconds taken by all reduce tasks=13252559794
        Map-Reduce Framework
                Map input records=1280000000
                Map output records=1280000000
                Map output bytes=128000000000
                Map output materialized bytes=130560059794
                Input split bytes=139284
                Combine input records=1280000000
                Combine output records=1280000000
                Reduce input groups=1280000000
                Reduce shuffle bytes=130560004027
                Reduce input records=1280000000
                Reduce output records=1280000000
                Spilled Records=7142221280
                Shuffled Maps =901
                Failed Shuffles=0
                Merged Map outputs=901
                GC time elapsed (ms)=447901
                CPU time spent (ms)=18790100
                Physical memory (bytes) snapshot=261063859794
                Virtual memory (bytes) snapshot=1875665178772
```

```
root@ip-172-31-12-172:/data/hadoop/bin# hdfs dfs -get /output /data
17/12/04 02:23:18 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
root@ip-172-31-12-172:/data/hadoop/bin# cd ../..
root@ip-172-31-12-172:/data# ls
hadoop  lost+found  output  tmp
root@ip-172-31-12-172:/data# cd output
root@ip-172-31-12-172:/data/output# ls
part-r-00000  _SUCCESS
root@ip-172-31-12-172:/data/output# cd ~
root@ip-172-31-12-172:~# ls
32  64  gensort-linux-1.5.tar.gz  gpl-2.0.txt  hadoop  hadoop-2.8.2.tar.gz  output  tmp
root@ip-172-31-12-172:~# cd 64
root@ip-172-31-12-172:~/64# ls
gensort  input.txt  smallFile.txt  valsort
root@ip-172-31-12-172:~/64# ./valsort /data/output/part-r-00000
Records: 1280000000
Checksum: 26258f2f171b7272
Duplicate keys: 0
SUCCESS - all records are in order
root@ip-172-31-12-224:~/64#
```

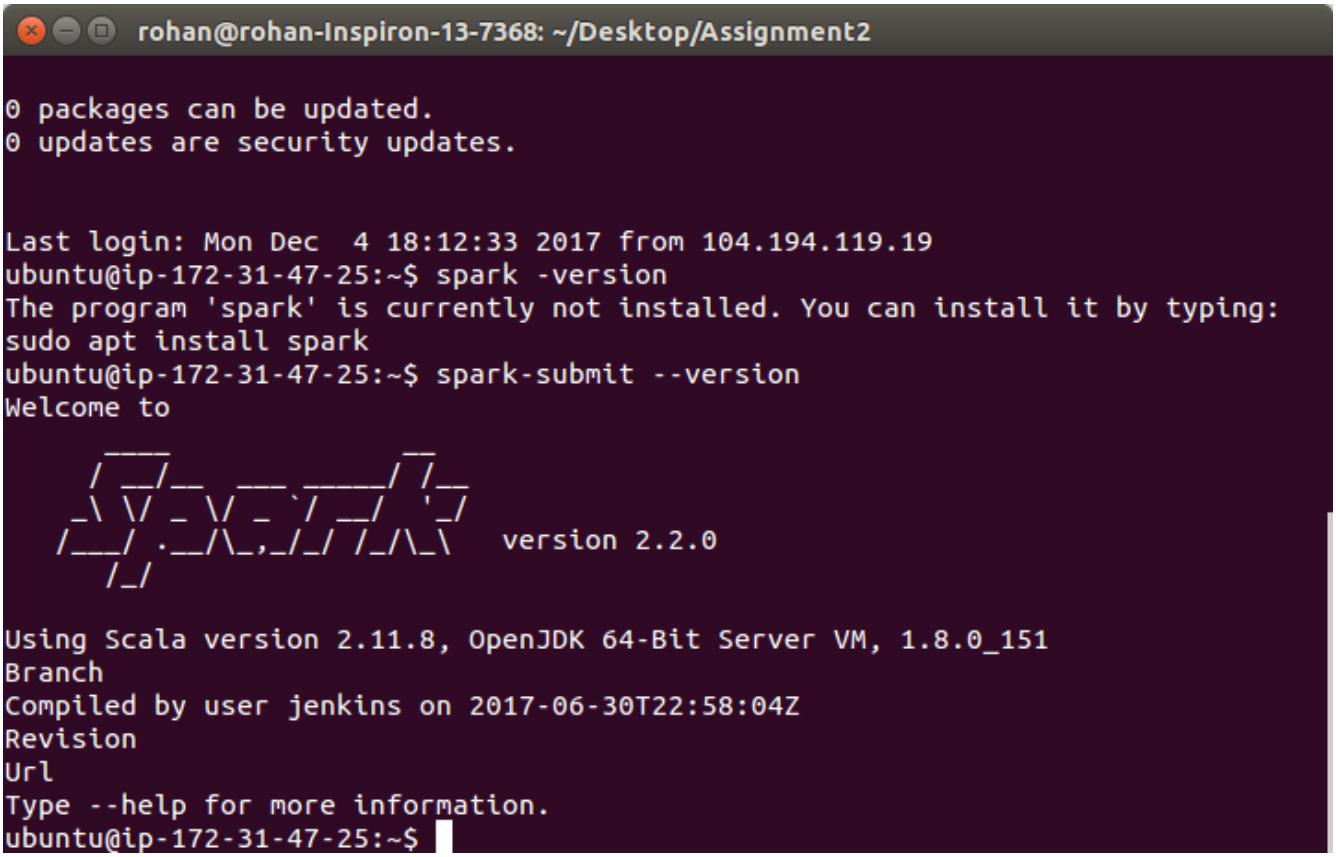
**Spark (1-node) i3.large:**

**We installed and configured spark using the following commands.**

```
wget https://d3kbcqa49mib13.cloudfront.net/spark-2.2.0-bin-hadoop2.7.tgz

sudo tar zxvf spark-2.2.0-bin-hadoop2.7.tgz -C /opt

sudo ln -fs spark-2.2.0-bin-hadoop2.7 /opt/spark
```

**Spark Home environment variables are set up using**

```
export SPARK_HOME=/opt/spark
PATH=$PATH:$SPARK_HOME/bin
export PATH

source ~/.bash_profile

# Confirm that spark-submit is now in the PATH.
spark-submit --version
```



```
17/12/04 01:32:21 INFO scheduler.TaskSetManager: Starting task 7.0 in stage 0.0 (TID 7, ip-172-31-78-246.ec2.internal, partition 7,ANY, 2162 bytes)
17/12/04 01:32:21 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 0.0 (TID 4) in 17282 ms on ip-172-31-78-246.ec2.internal (6/75)
17/12/04 01:32:32 INFO scheduler.TaskSetManager: Starting task 8.0 in stage 0.0 (TID 8, ip-172-31-78-246.ec2.internal, partition 8,ANY, 2162 bytes)
17/12/04 01:32:32 INFO scheduler.TaskSetManager: Finished task 7.0 in stage 0.0 (TID 7) in 10971 ms on ip-172-31-78-246.ec2.internal (7/75)
17/12/04 01:32:32 INFO scheduler.TaskSetManager: Starting task 9.0 in stage 0.0 (TID 9, ip-172-31-78-246.ec2.internal, partition 9,ANY, 2162 bytes)
17/12/04 01:32:32 INFO scheduler.TaskSetManager: Finished task 6.0 in stage 0.0 (TID 6) in 12845 ms on ip-172-31-78-246.ec2.internal (8/75)
17/12/04 01:32:42 INFO scheduler.TaskSetManager: Starting task 10.0 in stage 0.0 (TID 10, ip-172-31-78-246.ec2.internal, partition 10,ANY, 2162 bytes)
17/12/04 01:32:42 INFO scheduler.TaskSetManager: Finished task 8.0 in stage 0.0 (TID 8) in 9733 ms on ip-172-31-78-246.ec2.internal (9/75)
17/12/04 01:32:45 INFO scheduler.TaskSetManager: Starting task 11.0 in stage 0.0 (TID 11, ip-172-31-78-246.ec2.internal, partition 11,ANY, 2162 bytes)
17/12/04 01:32:45 INFO scheduler.TaskSetManager: Finished task 9.0 in stage 0.0 (TID 9) in 13102 ms on ip-172-31-78-246.ec2.internal (10/75)
17/12/04 01:32:54 INFO scheduler.TaskSetManager: Starting task 12.0 in stage 0.0 (TID 12, ip-172-31-78-246.ec2.internal, partition 12,ANY, 2162 bytes)
17/12/04 01:32:54 INFO scheduler.TaskSetManager: Finished task 10.0 in stage 0.0 (TID 10) in 11631 ms on ip-172-31-78-246.ec2.internal (11/75)
17/12/04 01:32:56 INFO scheduler.TaskSetManager: Starting task 13.0 in stage 0.0 (TID 13, ip-172-31-78-246.ec2.internal, partition 13,ANY, 2162 bytes)
17/12/04 01:32:56 INFO scheduler.TaskSetManager: Finished task 11.0 in stage 0.0 (TID 11) in 10222 ms on ip-172-31-78-246.ec2.internal (12/75)
17/12/04 01:33:03 INFO scheduler.TaskSetManager: Starting task 14.0 in stage 0.0 (TID 14, ip-172-31-78-246.ec2.internal, partition 14,ANY, 2162 bytes)
17/12/04 01:33:03 INFO scheduler.TaskSetManager: Finished task 12.0 in stage 0.0 (TID 12) in 9772 ms on ip-172-31-78-246.ec2.internal (13/75)
17/12/04 01:33:05 INFO scheduler.TaskSetManager: Starting task 15.0 in stage 0.0 (TID 15, ip-172-31-78-246.ec2.internal, partition 15,ANY, 2162 bytes)
17/12/04 01:33:05 INFO scheduler.TaskSetManager: Finished task 13.0 in stage 0.0 (TID 13) in 9884 ms on ip-172-31-78-246.ec2.internal (14/75)
17/12/04 01:33:13 INFO scheduler.TaskSetManager: Starting task 16.0 in stage 0.0 (TID 16, ip-172-31-78-246.ec2.internal, partition 16,ANY, 2162 bytes)
17/12/04 01:33:13 INFO scheduler.TaskSetManager: Finished task 14.0 in stage 0.0 (TID 14) in 9593 ms on ip-172-31-78-246.ec2.internal (15/75)
17/12/04 01:33:17 INFO scheduler.TaskSetManager: Starting task 17.0 in stage 0.0 (TID 17, ip-172-31-78-246.ec2.internal, partition 17,ANY, 2162 bytes)
17/12/04 01:33:17 INFO scheduler.TaskSetManager: Finished task 15.0 in stage 0.0 (TID 15) in 11696 ms on ip-172-31-78-246.ec2.internal (16/75)
17/12/04 01:33:27 INFO scheduler.TaskSetManager: Starting task 18.0 in stage 0.0 (TID 18, ip-172-31-78-246.ec2.internal, partition 18,ANY, 2162 bytes)
17/12/04 01:33:27 INFO scheduler.TaskSetManager: Finished task 16.0 in stage 0.0 (TID 16) in 14445 ms on ip-172-31-78-246.ec2.internal (17/75)
17/12/04 01:33:30 INFO scheduler.TaskSetManager: Starting task 19.0 in stage 0.0 (TID 19, ip-172-31-78-246.ec2.internal, partition 19,ANY, 2162 bytes)
17/12/04 01:33:30 INFO scheduler.TaskSetManager: Finished task 17.0 in stage 0.0 (TID 17) in 12820 ms on ip-172-31-78-246.ec2.internal (18/75)
17/12/04 01:33:40 INFO scheduler.TaskSetManager: Starting task 20.0 in stage 0.0 (TID 20, ip-172-31-78-246.ec2.internal, partition 20,ANY, 2162 bytes)
17/12/04 01:33:40 INFO scheduler.TaskSetManager: Finished task 18.0 in stage 0.0 (TID 18) in 13040 ms on ip-172-31-78-246.ec2.internal (19/75)
17/12/04 01:33:43 INFO scheduler.TaskSetManager: Starting task 21.0 in stage 0.0 (TID 21, ip-172-31-78-246.ec2.internal, partition 21,ANY, 2162 bytes)
17/12/04 01:33:43 INFO scheduler.TaskSetManager: Finished task 19.0 in stage 0.0 (TID 19) in 13463 ms on ip-172-31-78-246.ec2.internal (20/75)
17/12/04 01:33:56 INFO scheduler.TaskSetManager: Starting task 22.0 in stage 0.0 (TID 22, ip-172-31-78-246.ec2.internal, partition 22,ANY, 2162 bytes)
17/12/04 01:33:56 INFO scheduler.TaskSetManager: Finished task 20.0 in stage 0.0 (TID 20) in 15457 ms on ip-172-31-78-246.ec2.internal (21/75)
17/12/04 01:33:59 INFO scheduler.TaskSetManager: Starting task 23.0 in stage 0.0 (TID 23, ip-172-31-78-246.ec2.internal, partition 23,ANY, 2162 bytes)
```

```
scala> sortedOp.saveAsTextFile("/Output1")
17/12/04 01:39:53 INFO spark.SparkContext: Starting job: saveAsTextFile at <console>:32
17/12/04 01:39:53 INFO scheduler.DAGScheduler: Registering RDD 3 (map at <console>:29)
17/12/04 01:39:53 INFO scheduler.DAGScheduler: Got job 1 (saveAsTextFile at <console>:32) with 75 output partitions
17/12/04 01:39:53 INFO scheduler.DAGScheduler: Final stage: ResultStage 2 (saveAsTextFile at <console>:32)
17/12/04 01:39:53 INFO scheduler.DAGScheduler: Parents of final stage: List(ShuffleMapStage 1)
17/12/04 01:39:53 INFO scheduler.DAGScheduler: Missing parents: List(ShuffleMapStage 1)
17/12/04 01:39:53 INFO scheduler.DAGScheduler: Submitting ShuffleMapStage 1 (MapPartitionsRDD[3] at map at <console>:29), which has no missing parents
17/12/04 01:39:53 INFO storage.MemoryStore: Block broadcast_2 stored as values in memory (estimated size 5.4 KB, free 43.5 KB)
17/12/04 01:39:53 INFO storage.MemoryStore: Block broadcast_2_piece0 stored as bytes in memory (estimated size 3.4 KB, free 46.9 KB)
17/12/04 01:39:53 INFO storage.BlockManagerInfo: Added broadcast_2_piece0 in memory on 172.31.65.5:32918 (size: 3.4 KB, free: 511.5 MB)
17/12/04 01:39:53 INFO spark.SparkContext: Created broadcast 2 from broadcast at DAGScheduler.scala:1006
17/12/04 01:39:53 INFO scheduler.DAGScheduler: Submitting 75 missing tasks from ShuffleMapStage 1 (MapPartitionsRDD[3] at map at <console>:29)
17/12/04 01:39:53 INFO scheduler.TaskSchedulerImpl: Adding task set 1.0 with 75 tasks
17/12/04 01:39:53 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 1.0 (TID 75, ip-172-31-78-246.ec2.internal, partition 0,ANY, 2151 bytes)
17/12/04 01:39:53 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 1.0 (TID 76, ip-172-31-78-246.ec2.internal, partition 1,ANY, 2151 bytes)
17/12/04 01:39:54 INFO storage.BlockManagerInfo: Added broadcast_2_piece0 in memory on ip-172-31-78-246.ec2.internal:54393 (size: 3.4 KB, free: 7.6 GB)
17/12/04 01:40:06 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 1.0 (TID 77, ip-172-31-78-246.ec2.internal, partition 2,ANY, 2151 bytes)
17/12/04 01:40:06 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 1.0 (TID 75) in 12592 ms on ip-172-31-78-246.ec2.internal (1/75)
17/12/04 01:40:08 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 1.0 (TID 78, ip-172-31-78-246.ec2.internal, partition 3,ANY, 2151 bytes)
```

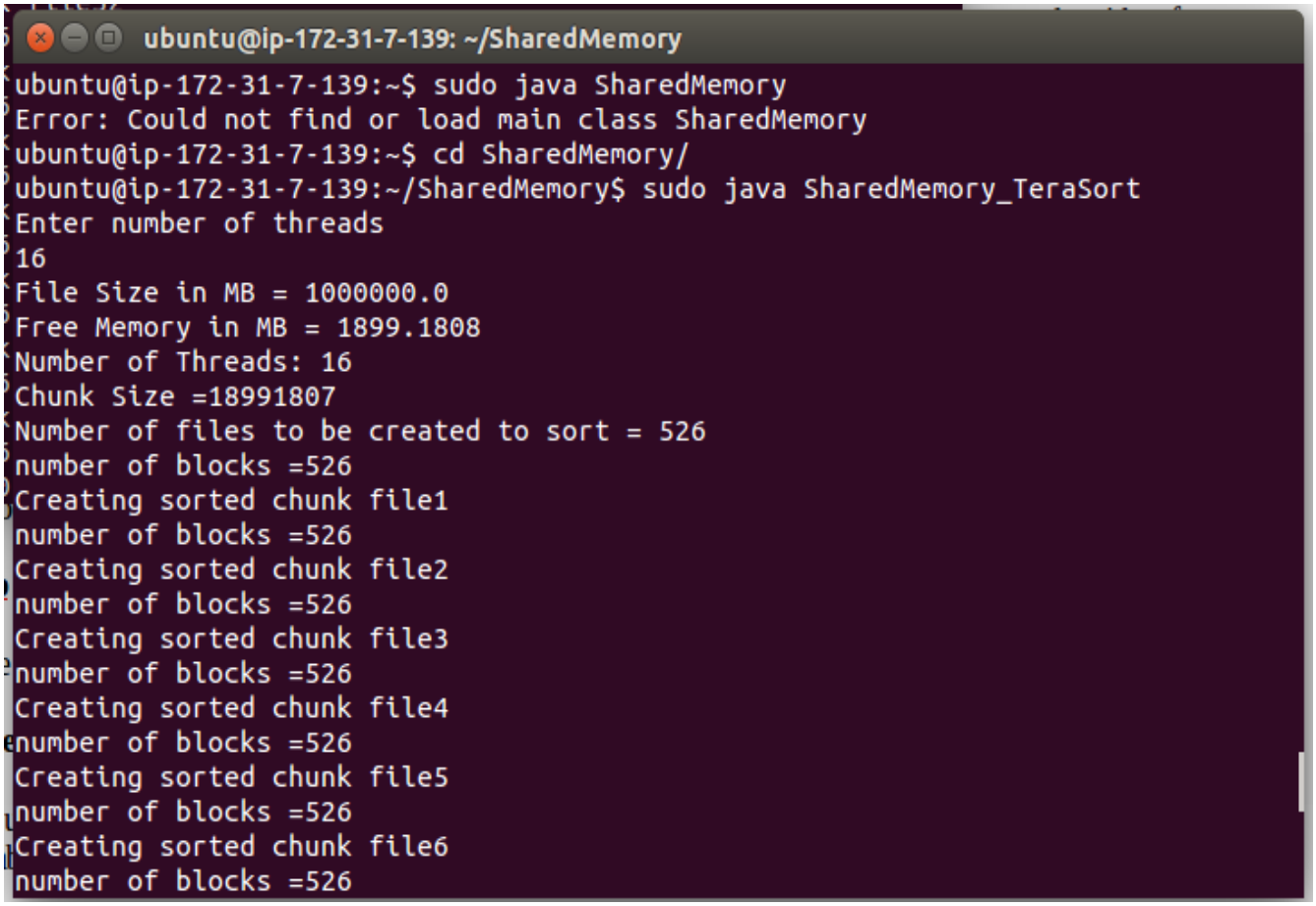
## Virtual Cluster (1-node) i3.4xlarge

We are using i3.4xlarge instance to store and sort 1 TB file. This instance has 1.7TB of two SSDs. To run our shared memory program we need atleast 3TB as we need space to store our 1 TB input file plus the chunk files which together will constitute to 1TB and also for storing the 1TB outfile file. So we merge the two RAID disk. We run the following script to merge the raid disk

```
#!/usr/bin/env bash
mkdir -p /data
mdadm --create --verbose --level=0 /dev/md0 --name=DATA --raid-devices=2 /dev/nvme0n1
/dev/nvme1n1
mdadm --wait /dev/md0
mkfs.ext4 /dev/md0
mdadm --detail --scan >> /etc/mdadm.conf
echo /dev/md0 /data ext4 defaults,nofail,noatime,discard 0 2 >> /etc/fstab
mount -a
```

**Shared Memory:**

We are running the program for 16 threads as this instance has 16 vCPUs. The configuration steps to run the java file is same as i3 instance.



```
ubuntu@ip-172-31-7-139: ~/SharedMemory
number of blocks =526
Creating sorted chunk file526
number of blocks =526
Creating sorted chunk file527
mergeLength = 10000000000
merge files
Merging: 10 percent
Merging: 20 percent
Merging: 30 percent
Merging: 40 percent
^[[B^[[B^[[B^[[B^[[B^[[B^[[B^[[B^[[B^[[B^[[B^[[B^[[B^[[B^[[B^[[B^[[B^[[A^[[A
^[[A^[[A^[[A^[[A^[[A^[[A^[[A^[[A^[[A^[[A^[[A^[[A^[[A^[[A^[[A^[[A^[[A^[[A^[[A
^[[A^[[A^[[B^[[B^[[B^[[B^[[B^[[B^[[B^[[B^[[B^[[B^[[B^[[B^[[B^[[B^[[B^[[B^[[B
^[[A^[[A^[[A^[[A^[[A^[[A^[[A^[[A^[[A^[[A^[[A^[[B^[[B^[[B^[[B^[[B^[[B^[[B^[[B
^[[B^[[B^[[B^[[B^[[B^[[B^[[B^[[B^[[B^[[B^[[A^[[A^[[A^[[A^[[A^[[A^[[A^[[A^[[A
^[[A^[[A^[[A^[[A^[[A^[[A^[[A^[[A^[[A^[[A^[[A^[[A^[[A^[[A^[[A^[[A^[[A^[[A^[[A
^[[A^[[A^[[A^[[A^[[A^[[AMerging: 50 percent
Merging: 60 percent
Merging: 70 percent
Merging: 80 percent
Merging: 90 percent
100% merged
Total Time Taken to sort = 32652.69 sec
ubuntu@ip-172-31-7-139:~/SharedMemory$
```

```
ubuntu@ip-172-31-7-139: ~/SharedMemory
input_chunk213.txt  input_chunk336.txt  input_chunk459.txt  input_chunk97.txt
input_chunk214.txt  input_chunk337.txt  input_chunk45.txt   input_chunk98.txt
input_chunk215.txt  input_chunk338.txt  input_chunk460.txt  input_chunk99.txt
input_chunk216.txt  input_chunk339.txt  input_chunk461.txt  input_chunk9.txt
input_chunk217.txt  input_chunk33.txt   input_chunk462.txt  inputfile.txt
input_chunk218.txt  input_chunk340.txt  input_chunk463.txt  lost+found
input_chunk219.txt  input_chunk341.txt  input_chunk464.txt  output.txt
input_chunk21.txt   input_chunk342.txt  input_chunk465.txt
input_chunk220.txt  input_chunk343.txt  input_chunk466.txt
input_chunk221.txt  input_chunk344.txt  input_chunk467.txt
ubuntu@ip-172-31-30-55:/mount1$ sudo rm -r input_chunk*.txt
ubuntu@ip-172-31-30-55:/mount1$ ls -l
total 250000180
-rwxr-xr-x 1 root root        141045 Dec  3 02:57 gensort
-rwxr-xr-x 1 root root 1000000000000 Dec  3 03:08 inputfile.txt
drwx------ 2 root root         16384 Dec  3 02:44 lost+found
-rw-r--r-- 1 root root 1000000000000 Dec  3 06:28 output.txt
ubuntu@ip-172-31-7-139:/mount1$ sudo cp ~/SharedMemory/valsort /mount1/
ubuntu@ip-172-31-7-139:/mount1$ sudo ./valsort output.txt
Records: 10000000000
Checksum: 4c26a01238df73d828b
Duplicate keys: 0
SUCCESS - all records are in order
ubuntu@ip-172-31-7-139:/mount1$
```

## Hadoop (1 Node) i3.4xlarge:

The Hadoop setup is same as that done in previous instance.
**Output.**

```
17/12/03 22:35:30 INFO mapreduce.Job:  map 100% reduce 97%
17/12/03 22:36:30 INFO mapreduce.Job:  map 100% reduce 98%
17/12/03 22:37:30 INFO mapreduce.Job:  map 100% reduce 99%
17/12/03 22:38:30 INFO mapreduce.Job:  map 100% reduce 100%
17/12/03 22:39:03 INFO mapreduce.Job: Job job_1512325297232_0001 completed successfully
17/12/03 22:39:03 INFO mapreduce.Job: Counters: 50
        File System Counters
                FILE: Number of bytes read=597946868118
                FILE: Number of bytes written=728634042772
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=1D00000000094
                HDFS: Number of bytes written=1D0000000000
                HDFS: Number of read operations=2803
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Killed map tasks=1
                Launched map tasks=901
                Launched reduce tasks=1
                Data-local map tasks=901
                Total time spent by all maps in occupied slots (ms)=50249794
                Total time spent by all reduces in occupied slots (ms)=12942772
                Total time spent by all map tasks (ms)=50249794
                Total time spent by all reduce tasks (ms)=12942772
                Total vcore-milliseconds taken by all map tasks=50249794
                Total vcore-milliseconds taken by all reduce tasks=12942772
                Total megabyte-milliseconds taken by all map tasks=51449489794
                Total megabyte-milliseconds taken by all reduce tasks=13252559794
        Map-Reduce Framework
                Map input records=1D00000000
                Map output records=1D00000000
                Map output bytes=1D0000000000
                Map output materialized bytes=130560059794
                Input split bytes=139284
                Combine input records=1D00000000
                Combine output records=1D00000000
                Reduce input groups=1D00000000
                Reduce shuffle bytes=130560004027
                Reduce input records=1D00000000
                Reduce output records=1D00000000
                Spilled Records=7142221280
                Shuffled Maps =901
                Failed Shuffles=0
                Merged Map outputs=901
                GC time elapsed (ms)=447901
                CPU time spent (ms)=48790100
                Physical memory (bytes) snapshot=261063859794
                Virtual memory (bytes) snapshot=1875665178772
```

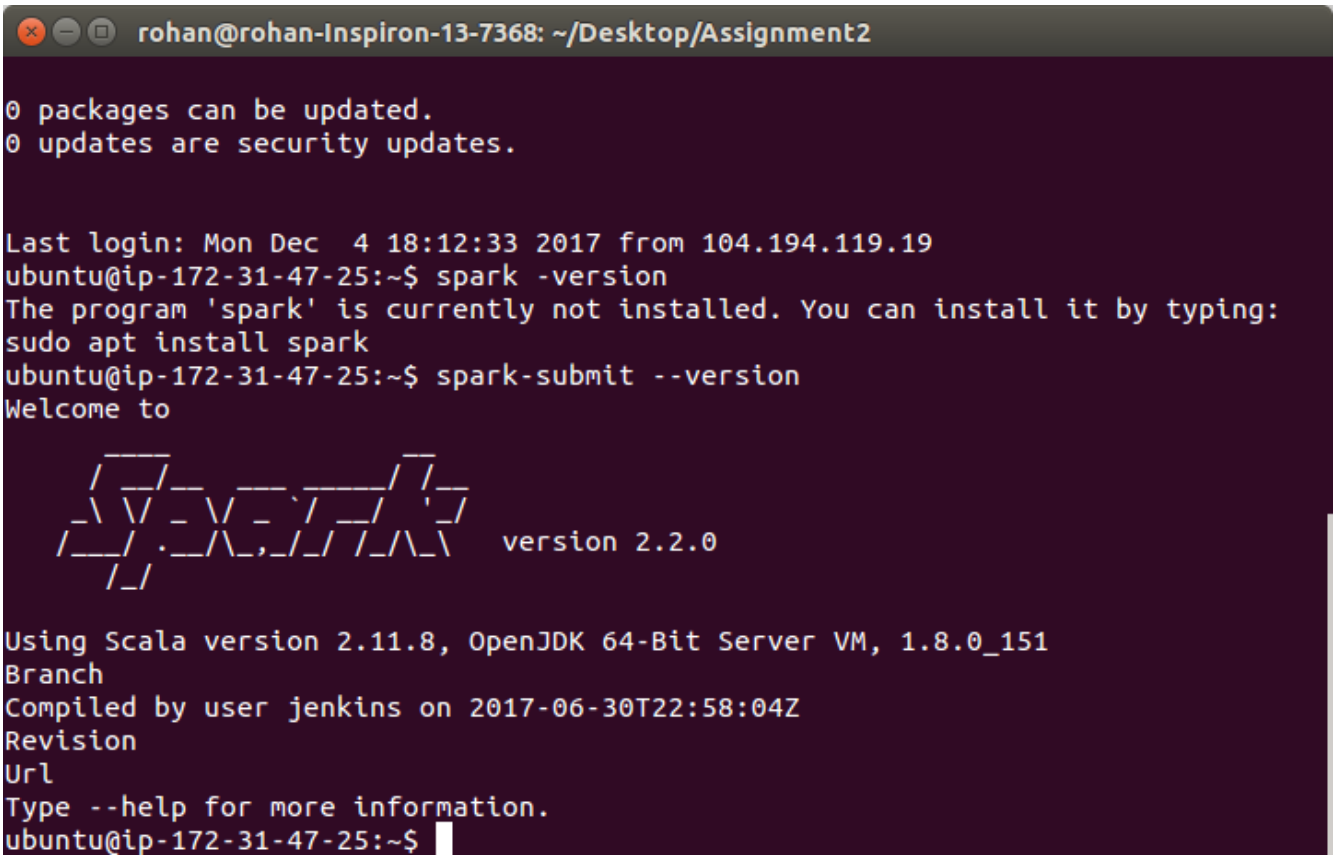
```
root@ip-172-31-12-172:/data/hadoop/bin# hdfs dfs -get /output /data
17/12/04 02:23:18 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
root@ip-172-31-12-172:/data/hadoop/bin# cd ../..
root@ip-172-31-12-172:/data# ls
hadoop  lost+found  output  tmp
root@ip-172-31-12-172:/data# cd output
root@ip-172-31-12-172:/data/output# ls
part-r-00000  _SUCCESS
root@ip-172-31-12-172:/data/output# cd ~
root@ip-172-31-12-172:~# ls
32  64  gensort-linux-1.5.tar.gz  gpl-2.0.txt  hadoop  hadoop-2.8.2.tar.gz  output  tmp
root@ip-172-31-12-172:~# cd 64
root@ip-172-31-12-172:~/64# ls
gensort  input.txt  smallFile.txt  valsort
root@ip-172-31-12-172:~/64# ./valsort /data/output/part-r-00000
Records: 1D00000000
Checksum: 26258f2f171b72721
Duplicate keys: 0
SUCCESS - all records are in order
root@ip-172-31-12-224:~/64#
```

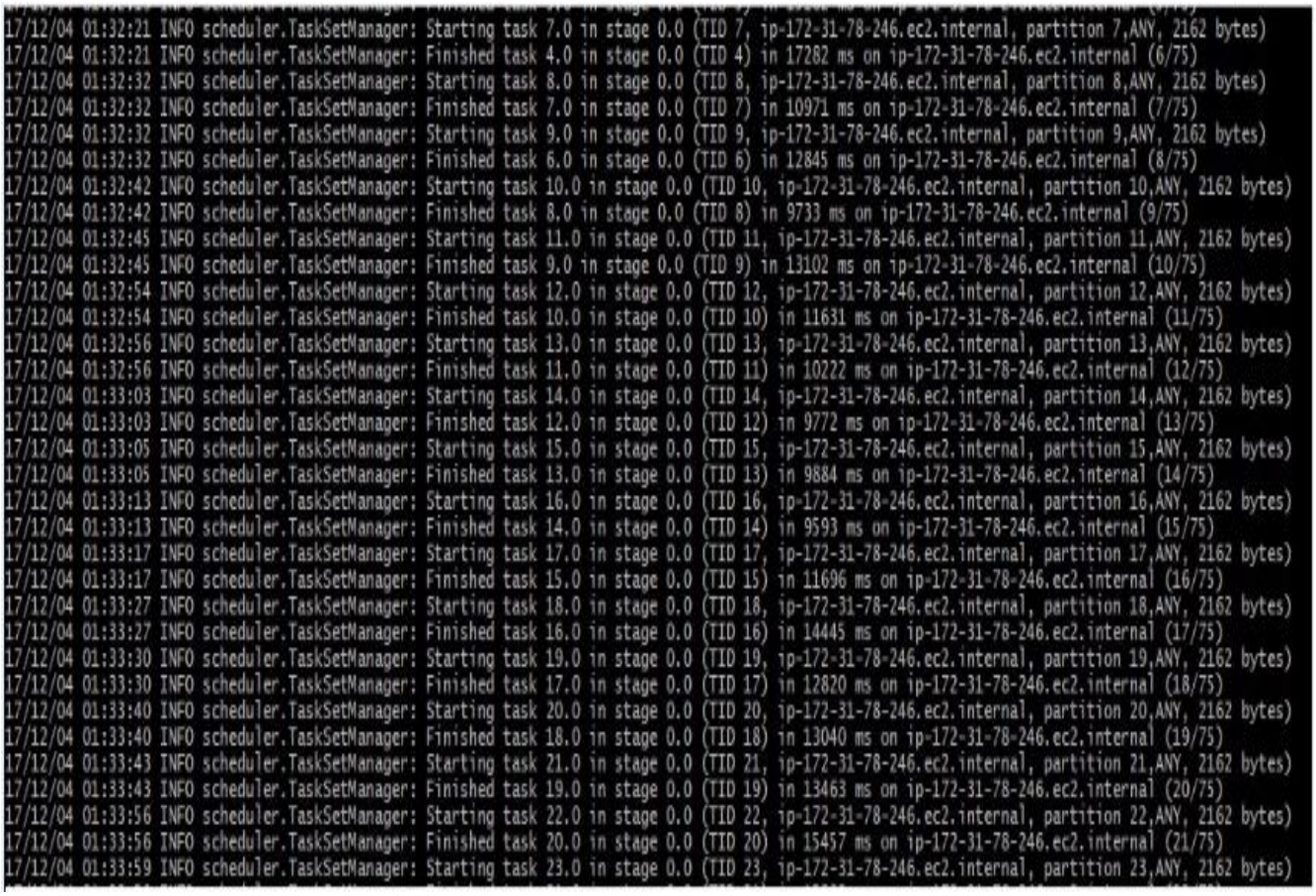
**Spark (1-node) i3.4xlarge:**

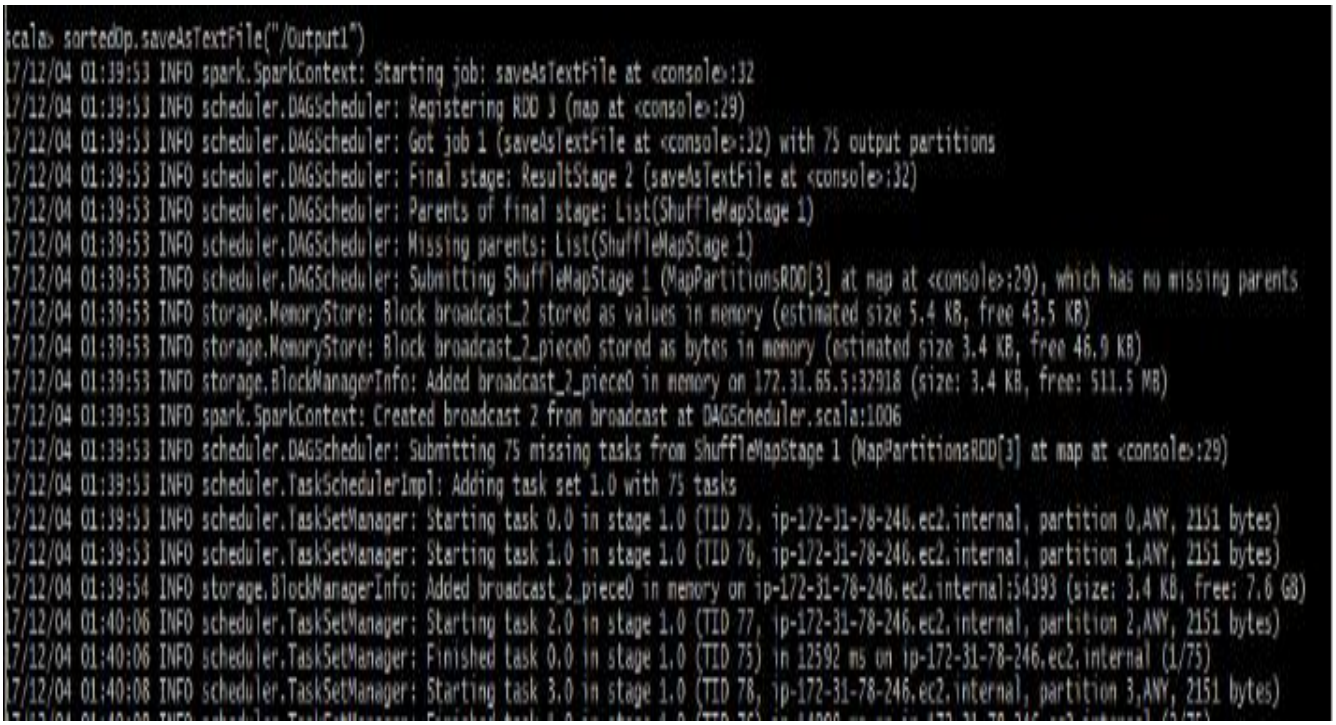
The steps to config and install spark are simlar to the above instance. We are using this instance to run 1TB data.

To run the scala file

```
./bin/spark-shell Spark_Sort.scala
```

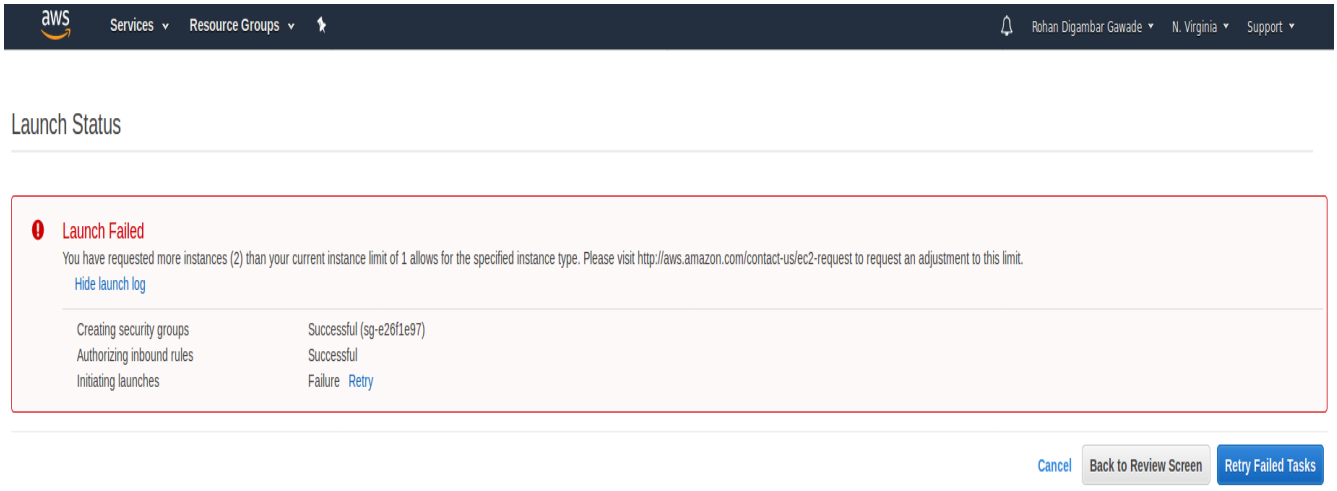



17/12/04 01:32:21 INFO scheduler.TaskSetManager: Starting task 7.0 in stage 0.0 (TID 7, ip-172-31-78-246.ec2.internal, partition 7,ANY, 2162 bytes)
17/12/04 01:32:21 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 0.0 (TID 4) in 17282 ms on ip-172-31-78-246.ec2.internal (6/75)
17/12/04 01:32:32 INFO scheduler.TaskSetManager: Starting task 8.0 in stage 0.0 (TID 8, ip-172-31-78-246.ec2.internal, partition 8,ANY, 2162 bytes)
17/12/04 01:32:32 INFO scheduler.TaskSetManager: Finished task 7.0 in stage 0.0 (TID 7) in 10971 ms on ip-172-31-78-246.ec2.internal (7/75)
17/12/04 01:32:32 INFO scheduler.TaskSetManager: Starting task 9.0 in stage 0.0 (TID 9, ip-172-31-78-246.ec2.internal, partition 9,ANY, 2162 bytes)
17/12/04 01:32:32 INFO scheduler.TaskSetManager: Finished task 6.0 in stage 0.0 (TID 6) in 12845 ms on ip-172-31-78-246.ec2.internal (8/75)
17/12/04 01:32:42 INFO scheduler.TaskSetManager: Starting task 10.0 in stage 0.0 (TID 10, ip-172-31-78-246.ec2.internal, partition 10,ANY, 2162 bytes)
17/12/04 01:32:42 INFO scheduler.TaskSetManager: Finished task 8.0 in stage 0.0 (TID 8) in 9733 ms on ip-172-31-78-246.ec2.internal (9/75)
17/12/04 01:32:45 INFO scheduler.TaskSetManager: Starting task 11.0 in stage 0.0 (TID 11, ip-172-31-78-246.ec2.internal, partition 11,ANY, 2162 bytes)
17/12/04 01:32:45 INFO scheduler.TaskSetManager: Finished task 9.0 in stage 0.0 (TID 9) in 13102 ms on ip-172-31-78-246.ec2.internal (10/75)
17/12/04 01:32:54 INFO scheduler.TaskSetManager: Starting task 12.0 in stage 0.0 (TID 12, ip-172-31-78-246.ec2.internal, partition 12,ANY, 2162 bytes)
17/12/04 01:32:54 INFO scheduler.TaskSetManager: Finished task 10.0 in stage 0.0 (TID 10) in 11631 ms on ip-172-31-78-246.ec2.internal (11/75)
17/12/04 01:32:56 INFO scheduler.TaskSetManager: Starting task 13.0 in stage 0.0 (TID 13, ip-172-31-78-246.ec2.internal, partition 13,ANY, 2162 bytes)
17/12/04 01:32:56 INFO scheduler.TaskSetManager: Finished task 11.0 in stage 0.0 (TID 11) in 10222 ms on ip-172-31-78-246.ec2.internal (12/75)
17/12/04 01:33:03 INFO scheduler.TaskSetManager: Starting task 14.0 in stage 0.0 (TID 14, ip-172-31-78-246.ec2.internal, partition 14,ANY, 2162 bytes)
17/12/04 01:33:03 INFO scheduler.TaskSetManager: Finished task 12.0 in stage 0.0 (TID 12) in 9772 ms on ip-172-31-78-246.ec2.internal (13/75)
17/12/04 01:33:05 INFO scheduler.TaskSetManager: Starting task 15.0 in stage 0.0 (TID 15, ip-172-31-78-246.ec2.internal, partition 15,ANY, 2162 bytes)
17/12/04 01:33:05 INFO scheduler.TaskSetManager: Finished task 13.0 in stage 0.0 (TID 13) in 9884 ms on ip-172-31-78-246.ec2.internal (14/75)
17/12/04 01:33:13 INFO scheduler.TaskSetManager: Starting task 16.0 in stage 0.0 (TID 16, ip-172-31-78-246.ec2.internal, partition 16,ANY, 2162 bytes)
17/12/04 01:33:13 INFO scheduler.TaskSetManager: Finished task 14.0 in stage 0.0 (TID 14) in 9593 ms on ip-172-31-78-246.ec2.internal (15/75)
17/12/04 01:33:17 INFO scheduler.TaskSetManager: Starting task 17.0 in stage 0.0 (TID 17, ip-172-31-78-246.ec2.internal, partition 17,ANY, 2162 bytes)
17/12/04 01:33:17 INFO scheduler.TaskSetManager: Finished task 15.0 in stage 0.0 (TID 15) in 11696 ms on ip-172-31-78-246.ec2.internal (16/75)
17/12/04 01:33:27 INFO scheduler.TaskSetManager: Starting task 18.0 in stage 0.0 (TID 18, ip-172-31-78-246.ec2.internal, partition 18,ANY, 2162 bytes)
17/12/04 01:33:27 INFO scheduler.TaskSetManager: Finished task 16.0 in stage 0.0 (TID 16) in 14445 ms on ip-172-31-78-246.ec2.internal (17/75)
17/12/04 01:33:30 INFO scheduler.TaskSetManager: Starting task 19.0 in stage 0.0 (TID 19, ip-172-31-78-246.ec2.internal, partition 19,ANY, 2162 bytes)
17/12/04 01:33:30 INFO scheduler.TaskSetManager: Finished task 17.0 in stage 0.0 (TID 17) in 12820 ms on ip-172-31-78-246.ec2.internal (18/75)
17/12/04 01:33:40 INFO scheduler.TaskSetManager: Starting task 20.0 in stage 0.0 (TID 20, ip-172-31-78-246.ec2.internal, partition 20,ANY, 2162 bytes)
17/12/04 01:33:40 INFO scheduler.TaskSetManager: Finished task 18.0 in stage 0.0 (TID 18) in 13040 ms on ip-172-31-78-246.ec2.internal (19/75)
17/12/04 01:33:43 INFO scheduler.TaskSetManager: Starting task 21.0 in stage 0.0 (TID 21, ip-172-31-78-246.ec2.internal, partition 21,ANY, 2162 bytes)
17/12/04 01:33:43 INFO scheduler.TaskSetManager: Finished task 19.0 in stage 0.0 (TID 19) in 13463 ms on ip-172-31-78-246.ec2.internal (20/75)
17/12/04 01:33:56 INFO scheduler.TaskSetManager: Starting task 22.0 in stage 0.0 (TID 22, ip-172-31-78-246.ec2.internal, partition 22,ANY, 2162 bytes)
17/12/04 01:33:56 INFO scheduler.TaskSetManager: Finished task 20.0 in stage 0.0 (TID 20) in 15457 ms on ip-172-31-78-246.ec2.internal (21/75)
17/12/04 01:33:59 INFO scheduler.TaskSetManager: Starting task 23.0 in stage 0.0 (TID 23, ip-172-31-78-246.ec2.internal, partition 23,ANY, 2162 bytes)


scala> sortedOp.saveAsTextFile("/Output1")
17/12/04 01:39:53 INFO spark.SparkContext: Starting job: saveAsTextFile at <console>:32
17/12/04 01:39:53 INFO scheduler.DAGScheduler: Registering RDD 3 (map at <console>:29)
17/12/04 01:39:53 INFO scheduler.DAGScheduler: Got job 1 (saveAsTextFile at <console>:32) with 75 output partitions
17/12/04 01:39:53 INFO scheduler.DAGScheduler: Final stage: ResultStage 2 (saveAsTextFile at <console>:32)
17/12/04 01:39:53 INFO scheduler.DAGScheduler: Parents of final stage: List(ShuffleMapStage 1)
17/12/04 01:39:53 INFO scheduler.DAGScheduler: Missing parents: List(ShuffleMapStage 1)
17/12/04 01:39:53 INFO scheduler.DAGScheduler: Submitting ShuffleMapStage 1 (MapPartitionsRDD[3] at map at <console>:29), which has no missing parents
17/12/04 01:39:53 INFO storage.MemoryStore: Block broadcast_2 stored as values in memory (estimated size 5.4 KB, free 43.5 KB)
17/12/04 01:39:53 INFO storage.MemoryStore: Block broadcast_2_piece0 stored as bytes in memory (estimated size 3.4 KB, free 46.9 KB)
17/12/04 01:39:53 INFO storage.BlockManagerInfo: Added broadcast_2_piece0 in memory on 172.31.65.5:32918 (size: 3.4 KB, free: 511.5 MB)
17/12/04 01:39:53 INFO spark.SparkContext: Created broadcast 2 from broadcast at DAGScheduler.scala:1006
17/12/04 01:39:53 INFO scheduler.DAGScheduler: Submitting 75 missing tasks from ShuffleMapStage 1 (MapPartitionsRDD[3] at map at <console>:29)
17/12/04 01:39:53 INFO scheduler.TaskSchedulerImpl: Adding task set 1.0 with 75 tasks
17/12/04 01:39:53 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 1.0 (TID 75, ip-172-31-78-246.ec2.internal, partition 0,ANY, 2151 bytes)
17/12/04 01:39:53 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 1.0 (TID 76, ip-172-31-78-246.ec2.internal, partition 1,ANY, 2151 bytes)
17/12/04 01:39:54 INFO storage.BlockManagerInfo: Added broadcast_2_piece0 in memory on ip-172-31-78-246.ec2.internal:54393 (size: 3.4 KB, free: 7.6 GB)
17/12/04 01:40:06 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 1.0 (TID 77, ip-172-31-78-246.ec2.internal, partition 2,ANY, 2151 bytes)
17/12/04 01:40:06 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 1.0 (TID 75) in 12592 ms on ip-172-31-78-246.ec2.internal (1/75)
17/12/04 01:40:08 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 1.0 (TID 78, ip-172-31-78-246.ec2.internal, partition 3,ANY, 2151 bytes)

## Multinode

We were not able to perform 8 node experiment due to following error



**Performance evaluation of Terasort:**

| Experiment (instance/dataset) | Shared Memory TeraSort | Hadoop TeraSort | Spark TeraSort |
|---|---|---|---|
| **Compute Time (sec) [1x i3.large 128GB]** | 10948.23 | 18790.21 | 9300 |
| **Data Read (GB) [1x i3.large 128GB]** | 128 | 128 | 128 |
| **Data Write (GB) [1x i3.large 128GB]** | 128 | 128 | 128 |
| **I/O Throughput (MB/sec) [1x i3.large 128GB]** | 23.382 MB/sec | 13.624 | 27.526 |
| **Compute Time (sec) [1x i3.4xlarge 1TB]** | 32652.69 | 48790.10 | 28398 |
| **Data Read (GB) [1x i3.4xlarge1TB]** | 1000 | 1000 | 1000 |
| **Data Write (GB) [1x i3.4xlarge 1TB]** | 1000 | 1000 | 1000 |
| **I/O Throughput (MB/sec) [1x i3.4xlarge 1TB]** | 61.251 MB/sec | 40.991 MB/sec | 70.427MB |
| **Speedup (weak scale)** | 2.682 | 3.08 | 2.62 |
| **Efficiency (weak scale)** | 33.529 | 38.512 | 32.74 |

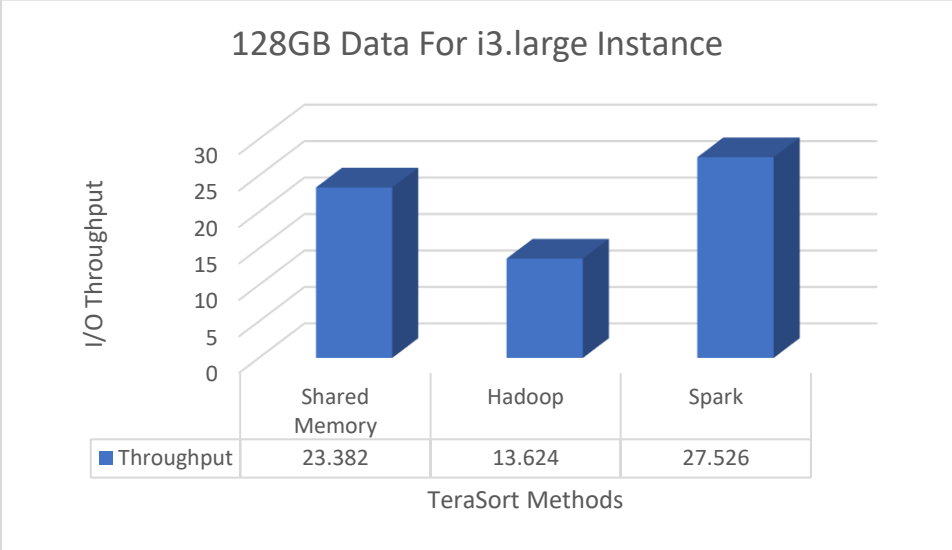
**Explanation**

- The Shared Memory Terasort for 128 GB takes around 3 hrs while Hadoop takes more than 5 hrs to sort the data even with 1 TB data we have the same case.
- This is because we have not increased the number of mapper and reducer. Since there are less mapper and reducer the throughput is also low for hadoop system.
- But with spark we increased the mapper and reducer to 8 so we have better results.
- The speed of Shared Memory for weak scale is 2.6 times as we go from 128 GB to 1 TB data. And 3 times and 2.6 times for Hadoop and Spark.
- The throughput is more as the number of core also increases.
- The multinode system should perform more faster as we have dedicated cores available for sorting. Here we can have one instance as master I.e namenode and other instances as slaves.(Data node).
- This data nodes are individual i3 instances or i3.4x instances so with multi node setting the hadoop and spark methods will perform sorting faster and efficiently.

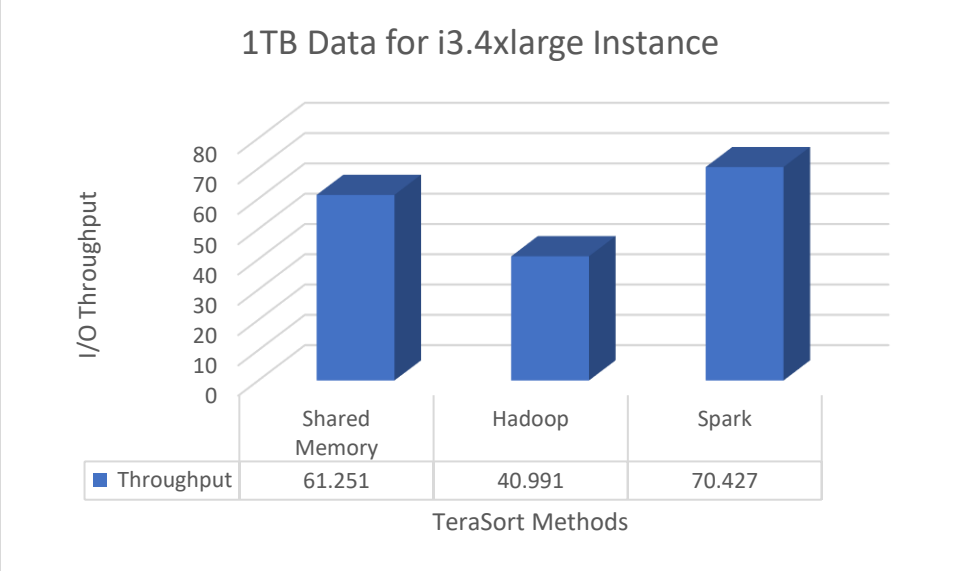
**Calculations**
Speed up = (8* compute time of i3large)/i3.4xlarge compute time

Efficiency = (100* compute time of i3large)/i3.4xlarge compute time

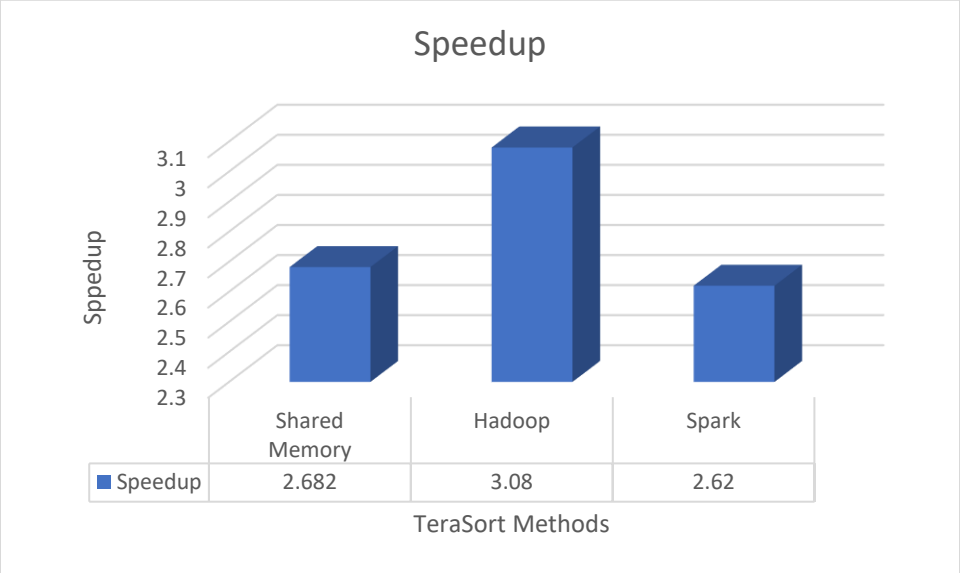
- We observe this using the following graphs.



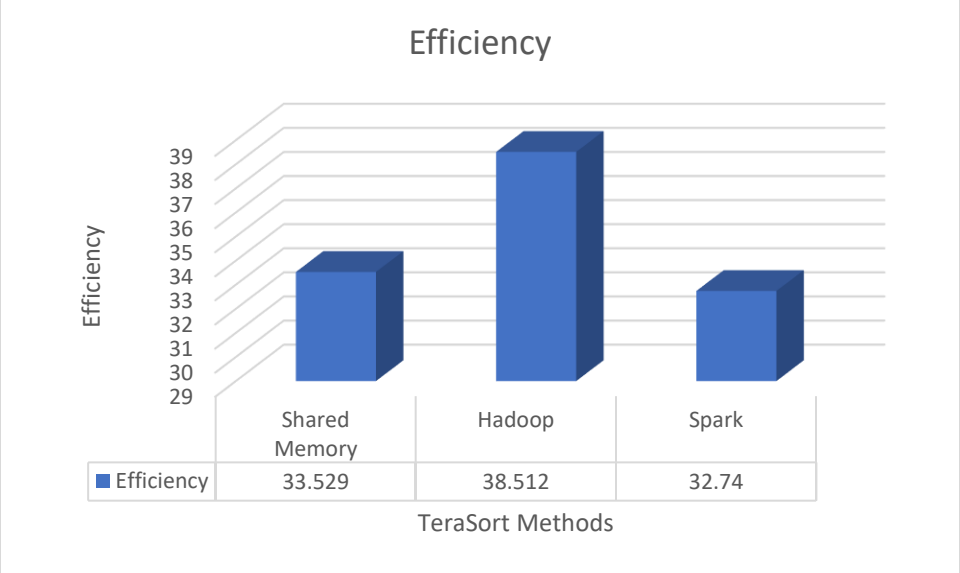

**Analysis**
The spark throughput is better among shared memory and hadoop. This may be because we have not increased the number of mapper and reducer for hadoop map-reduce




**Analysis:**
The results of this graph are similar to i3large instance. The throughput for 1TB is increased as this is i3.4xlarge instance and provides better performance




**Analysis**
This graph indicates that the hadoop speeds up more as compared to shared memory and spark.




**Analysis**:
Compared to Spark and Shared Memory Hadoop system is more efficient.

## Conclusion

- The i3.4xlarge system for 1 node will provide better results than 8 nodes of i3large systems.
- Spark seems to perform best at 1 node scale. But the speed up and efficiency of the Hadoop system is better for 1 node scale.
- For 8 nodes the Hadoop and spark system will perform well in 8 node settings. As per the speedup and efficiency, hadoop might performed better than spark for 100 nodes as well.
- Since spark is build on top of hadoop, for 1000 nodes the spark will perform better.
- As per the SORTING BENCHMARK, the Spark has sorted 100 TB of data in 23 minutes. https://opensource.com/business/15/1/apache-spark-new-world-record
- The hadoop uses 2100 nodes in this benchmark.
- The cloudsort measures the how efficient the external sort can be with respect to the cost of ownership. It provides a platform where we can derive more innovations on the clouds support to IO intensive tasks

## Installation steps you took to setup your virtual cluster:

- sudo apt-get update
- sudo apt-get upgrade
- sudo apt-get install default-jdk

This are the basic required packages to install other packages like hadoop and spark.

**For Hadoop we need to set up virtual cluster with the following environment variables.**
Adding environment variables in ~/.bashrc

```
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
export HADOOP_HOME=/usr/local/hadoop
export PATH=$PATH:$HADOOP_HOME/bin
export PATH=$PATH:$HADOOP_HOME/sbin
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib"
export HADOOP_CLASSPATH=$JAVA_HOME/lib/tools.jar
```

**For Hadoop we need to set up virtual cluster with the following environment variables.**
Add spark related environment variables to ~/.bash_profile:

```
export SPARK_HOME=/opt/spark
PATH=$PATH:$SPARK_HOME/bin
export PATH
```

## Difficulties that we faced during setuping virtual cluster:

1. The connection pipeline used to get broken.
2. For the 1TB setup, we had to mount the two raid disk into one and used for the experiment.
3. Manually configuring hadoop cluster was difficult task.
4. Spark Installation was more easy compared to Hadoop as hadoop took lot of time to configure. It had lot of configuration files to deal with and we were not able to set up for multinode.
5. Amazon didnt allow us to use more than one instances due to credit limits.

## Versions:

- OS - Ubuntu Server 16.04 LTS (HVM) - Linux 3.13.0-24-generic x86_64 kernel
- Ant version – Ant 1.9.9
- Java version – java-8-openjdk-amd64
- Hadoop version – Hadoop 2.7.4
- Spark version - Spark 2.2.0.