Cloud Computing
Author: Kedar Kaushikkar
TeraSort Using Hadoop on Amazon EC2 Cluster, Java and Spark
Performance Evaluation Report

Introduction:

This report gives the performance evaluation of terasort application implemented in three different ways: Java (Shared Memory), Hadoop and Spark.
We have created datasets of different data sets and executed this application on Amazon EC2 instances along with attaching storage volumes to it.

System Configuration Used on Amazon:

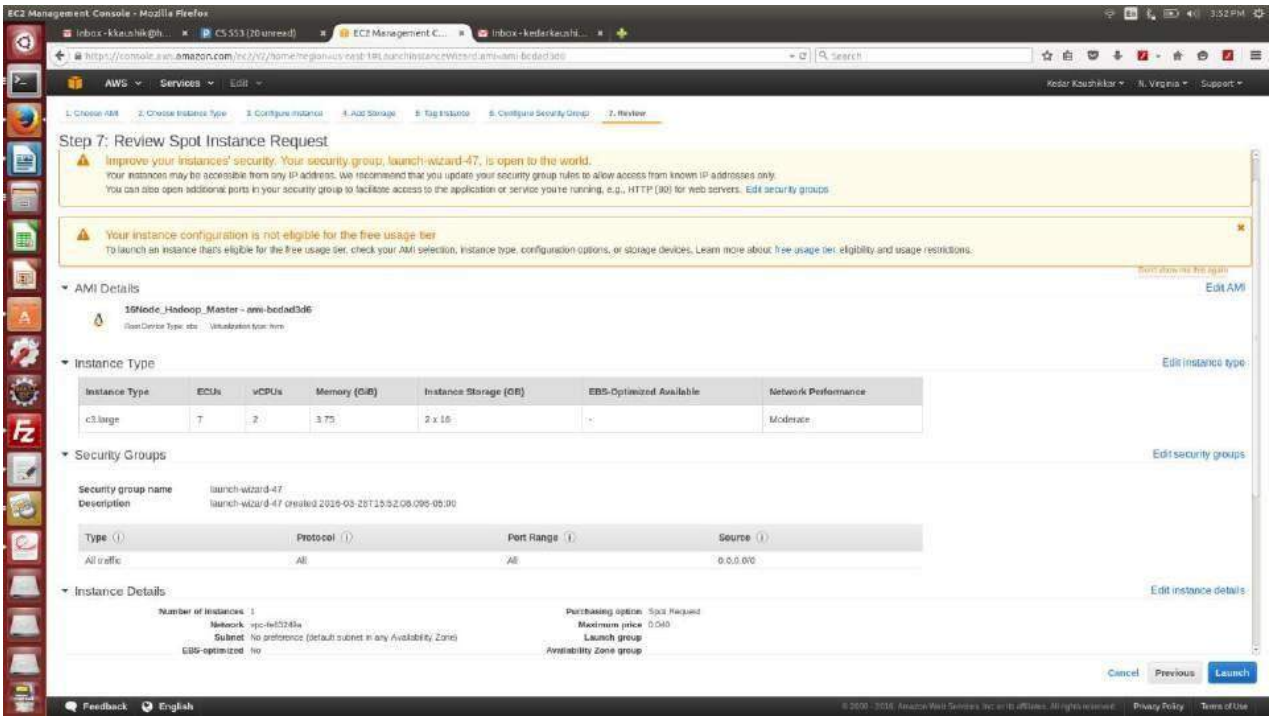
- AMI: Ubuntu Server 14.04 LTS (HVM), SSD Volume Type - ami-3d50120d
- Instance Type: c3.large. d2.xlarge
- Security Group: Created and used security group with SSH enabled for both inbound and outbound

Shared Memory Sort

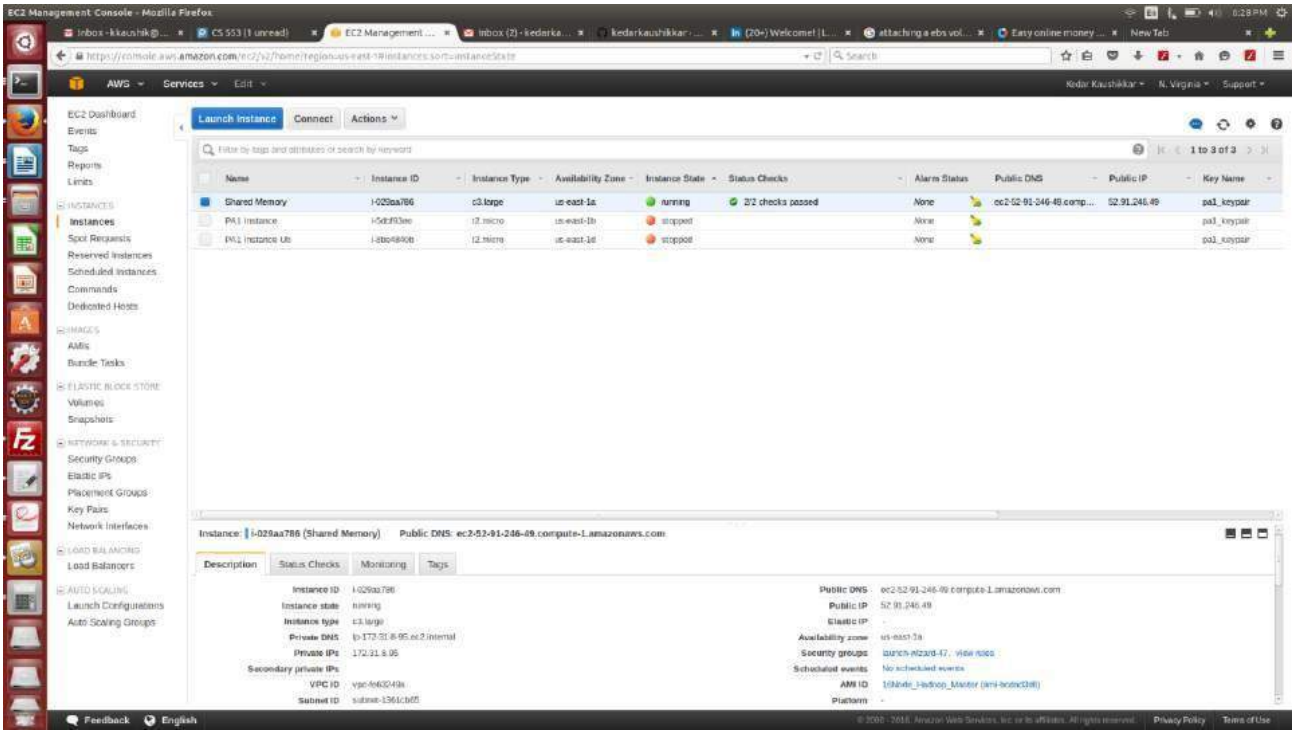
- Shared Memory sort is an application that will sort file-resident dataset and particularly those datasets that are larger than the allocated cluster memory.

- The application is implemented in Java with multithreading functionality. The 1GB and 10GB dataset is divided into chunks.
- For measuring performance, the dataset is stored in c3.large instance.
- To get efficient results, the test is performed on 1 thread, 2 thread, 4 thread and 8 thread.

The below screenshot shows the c3.large instance created on which the shared memory sort application can be executed.



Once we launch the instance, one active node will be actively running on the Amazon AWS Dashboard.



- To connect this above instance followed the following steps.

- We use chmod 400 My_Key.pem to make sure your private key file isn't publicly viewable.

1. Use the **ssh** command to connect to the instance.

2. After connection established we need to install Java on our virtual cluster
   sudo apt-get update
3. Install Default Java
   sudo apt-get install openjdk-7-jdk

- I have attached an EBS volume of 100 Gb to the instance to keep buffer storage and storing temporary files.

- To mount the EBS volume attached to the system, we follow the mentioned below steps.

Sudo yum install mdadm

- We use the **mdadm** command to create a logical RAID device from the newly attached Amazon EBS volumes.

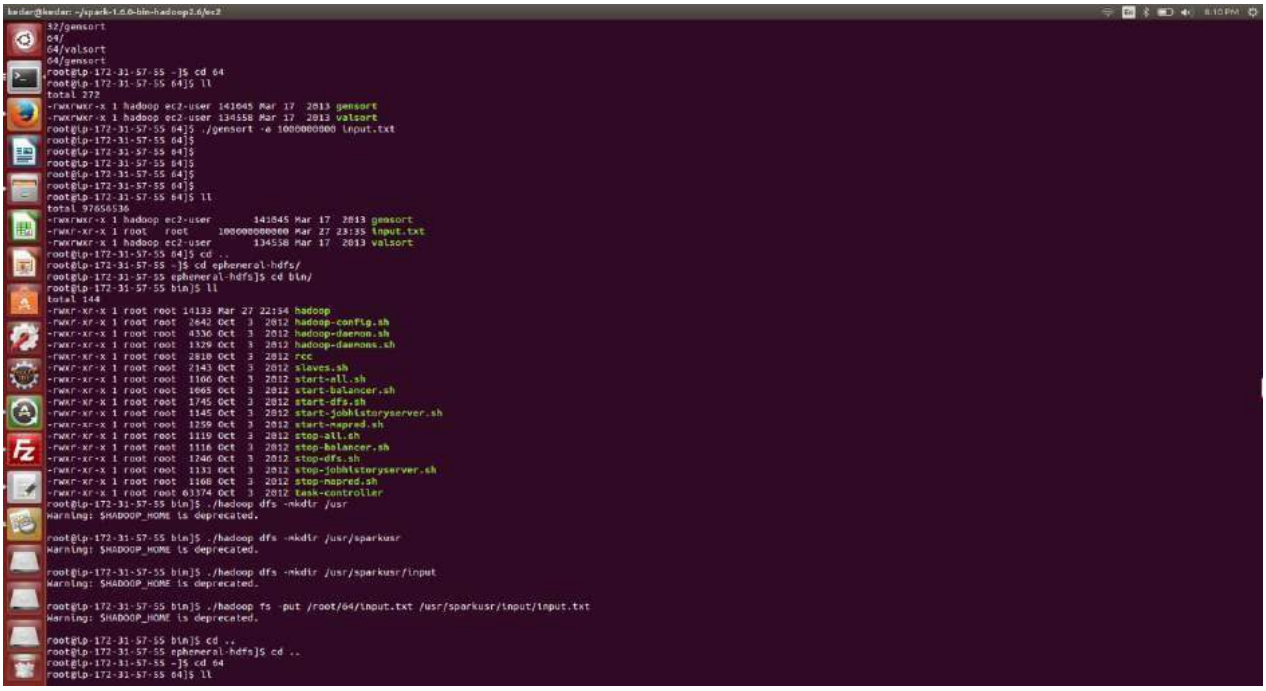
- We can list the devices on our instance using **lsblk** command

- To create a RAID 0 array, execute the following command
  sudo mdadm --create --verbose /dev/md0 --level=0 --name=RAID_NAME --raid-devices=No_Of_Devices Device1 Device2

- Create a file system on your RAID array, and give that file system a label to use when you mount it later.
  sudo mkfs.ext4 -L RAID_NAME /dev/md0

- Create a mount point for your RAID array.
  sudo mkdir -p /mnt/raid

- Finally, mount the RAID device on the mount point that you created:
  sudo mount LABEL=MAPREDUCE_RAID /mnt/raid



Shared Memory 10GB

- We need to create an input data of 10GB for which we use Gensort by using the following command.

  ./gensort –a 100000000 input.txt

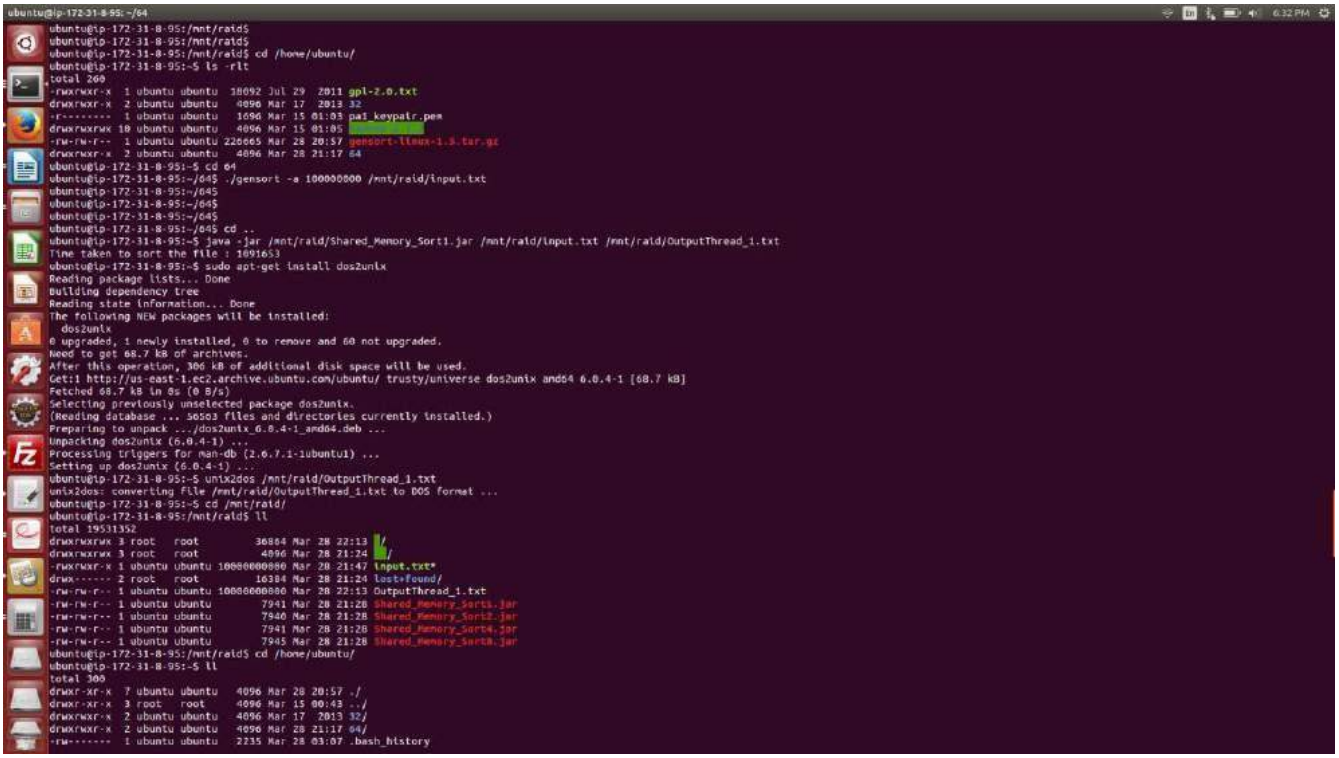
- The output of the above command will be a dataset of 10GB which will be stored in the mnt/raid directory.

The application jar files are present in the same directory where the input file is present.

The application JAR can be executed using the following command,

Java –jar /mnt/raid/Shared_Memory_Sort1.jar /mnt.raid/input.txt /mnt/raid/output1Thread.txt

Below screenshot shows the execution of Java Shared memory sort application on 1 thread.



```
ubuntu@ip-172-31-8-95: ~/64
ubuntu@ip-172-31-8-95:/mnt/raid$
ubuntu@ip-172-31-8-95:/mnt/raid$
ubuntu@ip-172-31-8-95:/mnt/raid$ cd /home/ubuntu/
ubuntu@ip-172-31-8-95:~$ ls -rlt
total 260
-rwxrwxr-x  1 ubuntu ubuntu  18092 Jul 29  2011 gpl-2.0.txt
drwxrwxr-x  2 ubuntu ubuntu   4096 Mar 17  2013 32
-r--------  1 ubuntu ubuntu   1696 Mar 15 01:03 pa1_keypair.pem
drwxrwxrwx 10 ubuntu ubuntu   4096 Mar 15 01:05 [redacted]
-rw-rw-r--  1 ubuntu ubuntu 226065 Mar 28 20:57 gensort-linux-1.3.tar.gz
drwxrwxr-x  2 ubuntu ubuntu   4096 Mar 28 21:17 64
ubuntu@ip-172-31-8-95:~$ cd 64
ubuntu@ip-172-31-8-95:~/64$ ./gensort -a 100000000 /mnt/raid/input.txt
ubuntu@ip-172-31-8-95:~/64$
ubuntu@ip-172-31-8-95:~/64$
ubuntu@ip-172-31-8-95:~/64$
ubuntu@ip-172-31-8-95:~/64$ cd ..
ubuntu@ip-172-31-8-95:~$ java -jar /mnt/raid/Shared_Memory_Sort1.jar /mnt/raid/input.txt /mnt/raid/OutputThread_1.txt
Time taken to sort the file : 1091653
ubuntu@ip-172-31-8-95:~$ sudo apt-get install dos2unix
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following NEW packages will be installed:
  dos2unix
0 upgraded, 1 newly installed, 0 to remove and 60 not upgraded.
Need to get 68.7 kB of archives.
After this operation, 306 kB of additional disk space will be used.
Get:1 http://us-east-1.ec2.archive.ubuntu.com/ubuntu/ trusty/universe dos2unix amd64 6.0.4-1 [68.7 kB]
Fetched 68.7 kB in 0s (0 B/s)
Selecting previously unselected package dos2unix.
(Reading database ... 56503 files and directories currently installed.)
Preparing to unpack .../dos2unix_6.0.4-1_amd64.deb ...
Unpacking dos2unix (6.0.4-1) ...
Processing triggers for man-db (2.6.7.1-1ubuntu1) ...
Setting up dos2unix (6.0.4-1) ...
ubuntu@ip-172-31-8-95:~$ unix2dos /mnt/raid/OutputThread_1.txt
unix2dos: converting file /mnt/raid/OutputThread_1.txt to DOS format ...
ubuntu@ip-172-31-8-95:~$ cd /mnt/raid/
ubuntu@ip-172-31-8-95:/mnt/raid$ ll
total 19531352
drwxrwxrwx 3 root   root        36864 Mar 28 22:13 [redacted]/
drwxrwxrwx 3 root   root         4096 Mar 28 21:24 [redacted]/
-rwxrwxr-x 1 ubuntu ubuntu 10000000000 Mar 28 21:47 input.txt*
drwx------ 2 root   root        16384 Mar 28 21:24 lost+found/
-rw-rw-r-- 1 ubuntu ubuntu 10000000000 Mar 28 22:13 OutputThread_1.txt
-rw-rw-r-- 1 ubuntu ubuntu        7941 Mar 28 21:28 Shared_Memory_Sort1.jar
-rw-rw-r-- 1 ubuntu ubuntu        7940 Mar 28 21:28 Shared_Memory_Sort2.jar
-rw-rw-r-- 1 ubuntu ubuntu        7941 Mar 28 21:28 Shared_Memory_Sort4.jar
-rw-rw-r-- 1 ubuntu ubuntu        7945 Mar 28 21:28 Shared_Memory_Sort8.jar
ubuntu@ip-172-31-8-95:/mnt/raid$ cd /home/ubuntu/
ubuntu@ip-172-31-8-95:~$ ll
total 300
drwxr-xr-x  7 ubuntu ubuntu   4096 Mar 28 20:57 ./
drwxr-xr-x  3 root   root     4096 Mar 15 00:43 ../
drwxrwxr-x  2 ubuntu ubuntu   4096 Mar 17  2013 32/
drwxrwxr-x  2 ubuntu ubuntu   4096 Mar 28 21:17 64/
-rw-------  1 ubuntu ubuntu   2235 Mar 28 03:07 .bash_history
```

- The output of the application will be stored in the /mnt/raid path as given in the execution command.

- The output file can be evaluated using valsort using the below command,

. /valsort output1Thread.txt

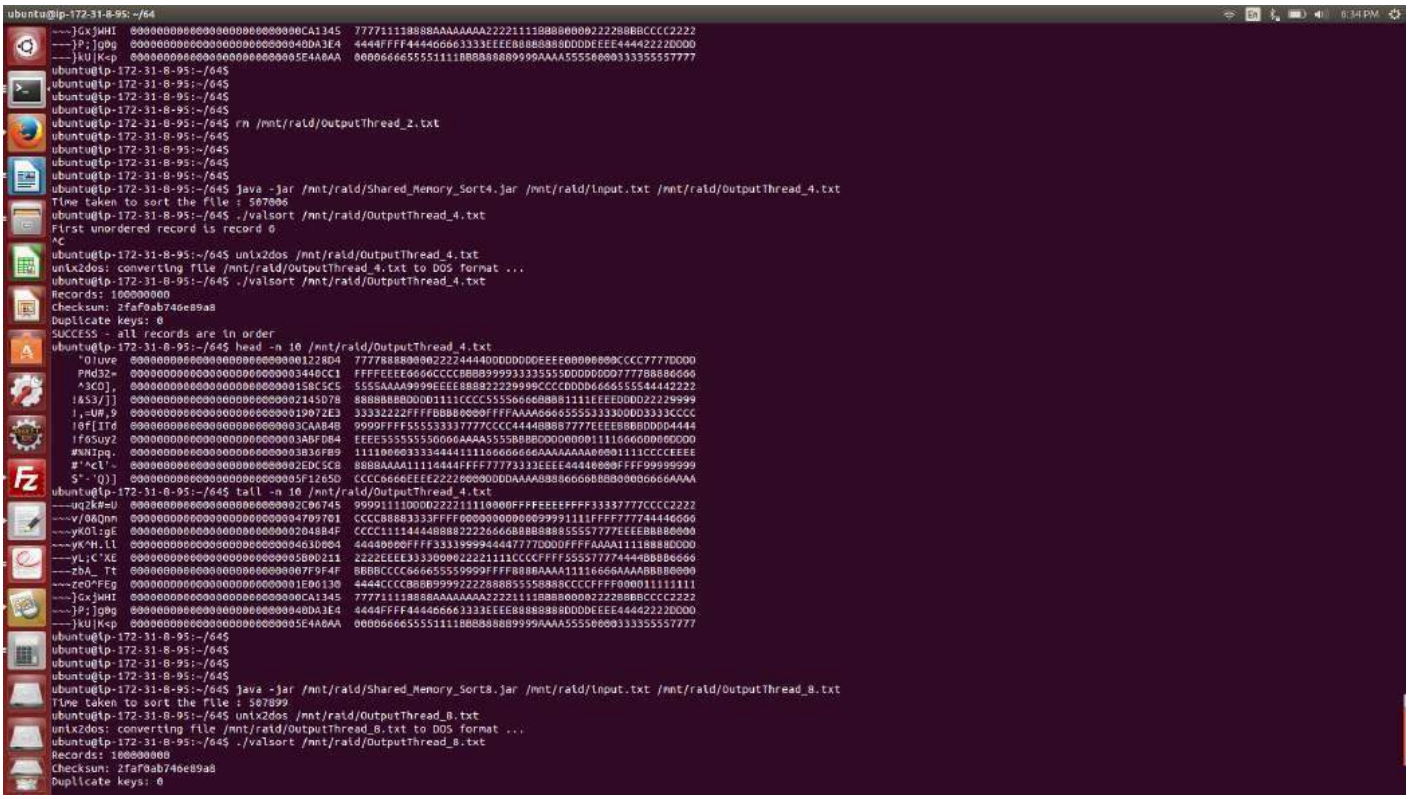
The output of the valsort command will be whether all the records in the output file are in order. The first 10 and last 10 lines of the output files are also shown.

```
ubuntu@ip-172-31-8-95: ~/64
drwxrwxrwx 3 root   root          4096 Mar 28 21:24 ./
-rwxrwxr-x 1 ubuntu ubuntu 10000000000 Mar 28 21:47 input.txt*
drwx------ 2 root   root         16384 Mar 28 21:24 lost+found/
-rw-rw-r-- 1 ubuntu ubuntu 10000000000 Mar 28 22:13 OutputThread_1.txt
-rw-rw-r-- 1 ubuntu ubuntu        7941 Mar 28 21:28 Shared_Memory_Sort1.jar
-rw-rw-r-- 1 ubuntu ubuntu        7940 Mar 28 21:28 Shared_Memory_Sort2.jar
-rw-rw-r-- 1 ubuntu ubuntu        7941 Mar 28 21:28 Shared_Memory_Sort4.jar
-rw-rw-r-- 1 ubuntu ubuntu        7945 Mar 28 21:28 Shared_Memory_Sort8.jar
ubuntu@ip-172-31-8-95:/mnt/raid$ cd /home/ubuntu/
ubuntu@ip-172-31-8-95:~$ ll
total 300
drwxr-xr-x  7 ubuntu ubuntu   4096 Mar 28 20:57 ./
drwxr-xr-x  3 root   root     4096 Mar 15 00:43 ../
drwxrwxr-x  2 ubuntu ubuntu   4096 Mar 17  2013 32/
drwxrwxr-x  2 ubuntu ubuntu   4096 Mar 28 21:17 64/
-rw-------  1 ubuntu ubuntu   2235 Mar 28 03:07 .bash_history
-rw-r--r--  1 ubuntu ubuntu    220 Apr  9  2014 .bash_logout
-rw-r--r--  1 ubuntu ubuntu   3789 Mar 15 00:56 .bashrc
drwx------  2 ubuntu ubuntu   4096 Mar 15 00:47 .cache/
-rw-rw-r--  1 ubuntu ubuntu 226665 Mar 28 20:57 gensort-linux-1.5.tar.gz
-rwxrwxr-x  1 ubuntu ubuntu  18092 Jul 29  2011 gpl-2.0.txt*
drwxrwxrwx 10 ubuntu ubuntu   4096 Mar 15 01:05 [illegible]/
-r--------  1 ubuntu ubuntu   1696 Mar 15 01:03 pai_keypair.pem
-rw-r--r--  1 ubuntu ubuntu    675 Apr  9  2014 .profile
drwx------  2 ubuntu ubuntu   4096 Mar 15 01:05 .ssh/
-rw-------  1 root   root     4746 Mar 28 03:02 .viminfo
ubuntu@ip-172-31-8-95:~$ cd 64
ubuntu@ip-172-31-8-95:~/64$ ./valsort /mnt/raid/OutputThread_1.txt
Records: 100000000
Checksum: 2faf0ab746e89a8
Duplicate keys: 0
SUCCESS - all records are in order
ubuntu@ip-172-31-8-95:~/64$ head -n 10 /mnt/raid/OutputThread_1.txt
    "O!uve  00000000000000000000000001228D4  77778888000022224444DDDDDDDDEEEE00000000CCCC7777DDDD
     PMd32=  00000000000000000000000003448CC1  FFFFEEEE6666CCCCBBBB99993333555SDDDDDDDD77778888666
     ^3CO],  00000000000000000000000015AC5C5  5555AAAA9999EEEE8888222299 99CCCCDDDD66665555444422 22
   !&S3/]]  00000000000000000000000002145D78  8888BBBBDDDD1111CCCC555566668888 1111EEEEDDDD22229999
   !,=U#,9  00000000000000000000000019072E3  33332222FFFFBBBB0000FFFFAAAA66665555333 3DDD3333CCCC
   !0f[ITd  00000000000000000000000003CAAB4B  9999FFFF5555333377 77CCCC4444BBBB7777EEEEBBBBDDDD4444
   !f6Suy2  00000000000000000000000003ABFDB4  EEEE5555555556666AAAA5555BBBBDDDD0000111166660000DDDD
   #%NIpq.  00000000000000000000000003B36FB9  1111000033334444111166666666AAAAAAAA00001111CCCCEEEE
   #'^cl'-  00000000000000000000000002EDC5C8  8888AAAA11114444FFFF77773333EEEE44440000FFFF99999999
   $"-'Q)]  00000000000000000000000005F1265D  CCCC6666EEEE22220000DDDDAAAA88886666BBBB0000666 6AAAA
ubuntu@ip-172-31-8-95:~/64$ tail -n 10 /mnt/raid/OutputThread_1.txt
~~~uq2k#=U  00000000000000000000000002C06745  99991111DDDD222211110000FFFFEEEEFFFF33337777CCCC2222
~~~v/0&Qnm  00000000000000000000000004709701  CCCC8888333 3FFFF0000000000000999911111FFFF777744446666
~~~yKOl:gE  00000000000000000000000002048B4F  CCCC11114444BBBB22226666BBBB8888555577 77EEEEBBBB0000
~~~yK^H.ll  00000000000000000000000004463D04  44440000FFFF333399994444777 7DDDDFFFFAAAA11118888DCDD
~~~yL;C'XE  00000000000000000000000005B0D211  2222EEEE333300002222111 1CCCCFFFF555577774444BBBB6666
~~~zbA_ Tt  00000000000000000000000006F9F4F  BBBBCCCC66665555999 9FFFF8888AAAA11116666AAAABBBB0000
~~~ze0^FEg  00000000000000000000000001E06130  4444CCCCBBBB999922228888555588888CCCCFFFF00001111111
~~~}GxjWHI  00000000000000000000000000CA1345  7777111188888AAAAAAAA22221111BBBB0000222 2BBBCCCC2222
~~~}P;]g0g  00000000000000000000000040DA3E4  4444FFFF444466663333EEEEBBBBBBBBDDDDEEEEE44442222DDDD
~~~}kU|K<p  00000000000000000000000005E4A0AA  00006666555511111BBBBBBBB9999AAAA555500003333555577 77
ubuntu@ip-172-31-8-95:~/64$
ubuntu@ip-172-31-8-95:~/64$
ubuntu@ip-172-31-8-95:~/64$
ubuntu@ip-172-31-8-95:~/64$
```

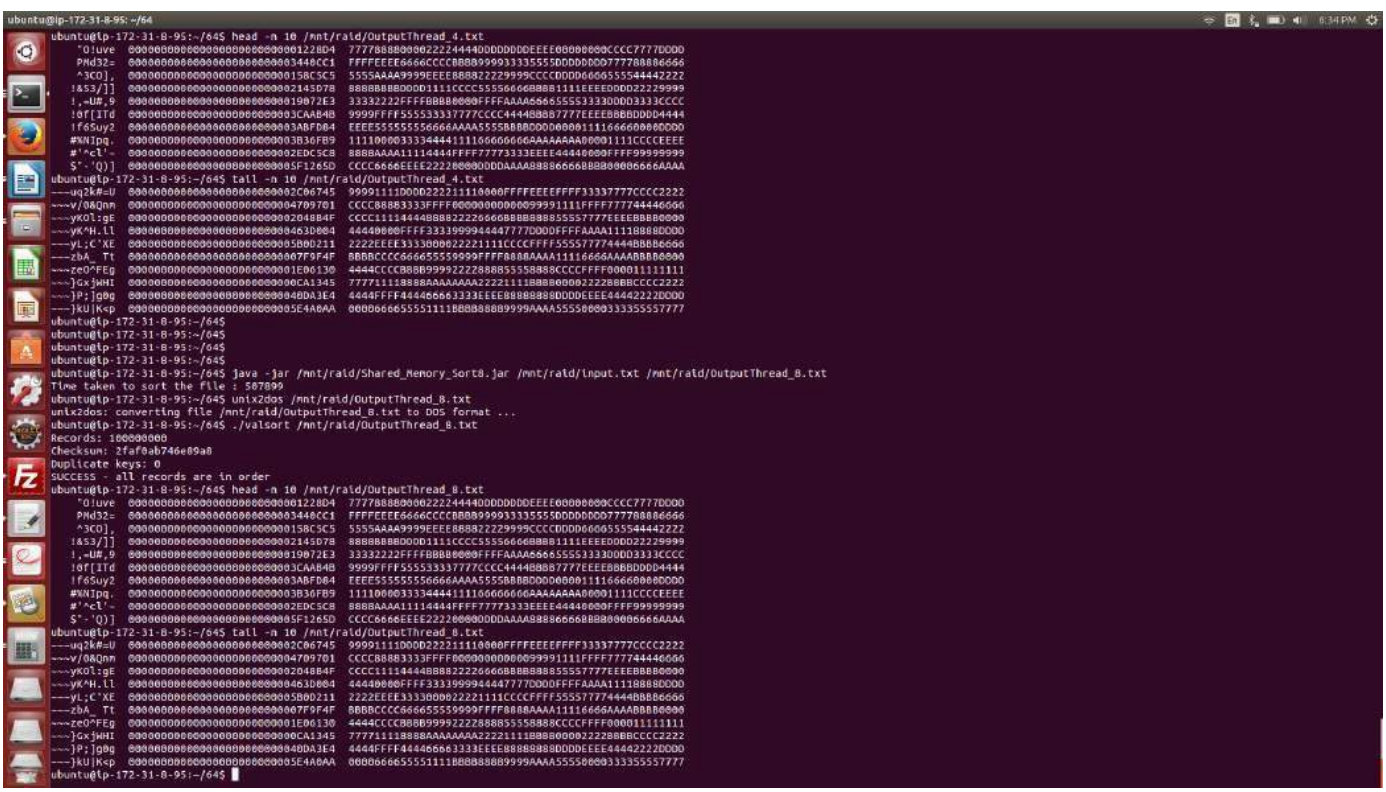
2 Threads

```
ubuntu@ip-172-31-8-95: ~/64
ubuntu@ip-172-31-8-95:~/64$
ubuntu@ip-172-31-8-95:~/64$
ubuntu@ip-172-31-8-95:~/64$
ubuntu@ip-172-31-8-95:~/64$
```

4 Threads.



8 Threads

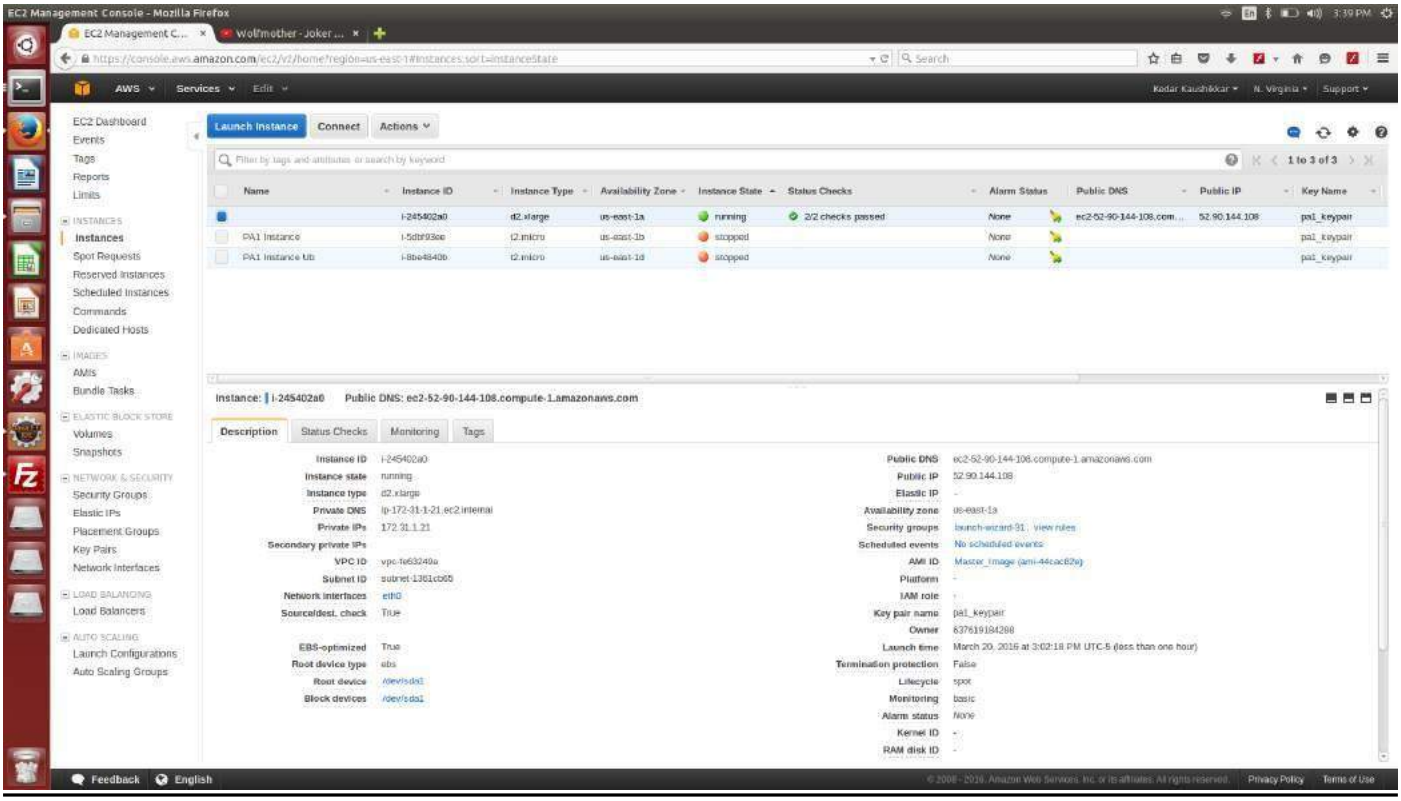


Performance Report:

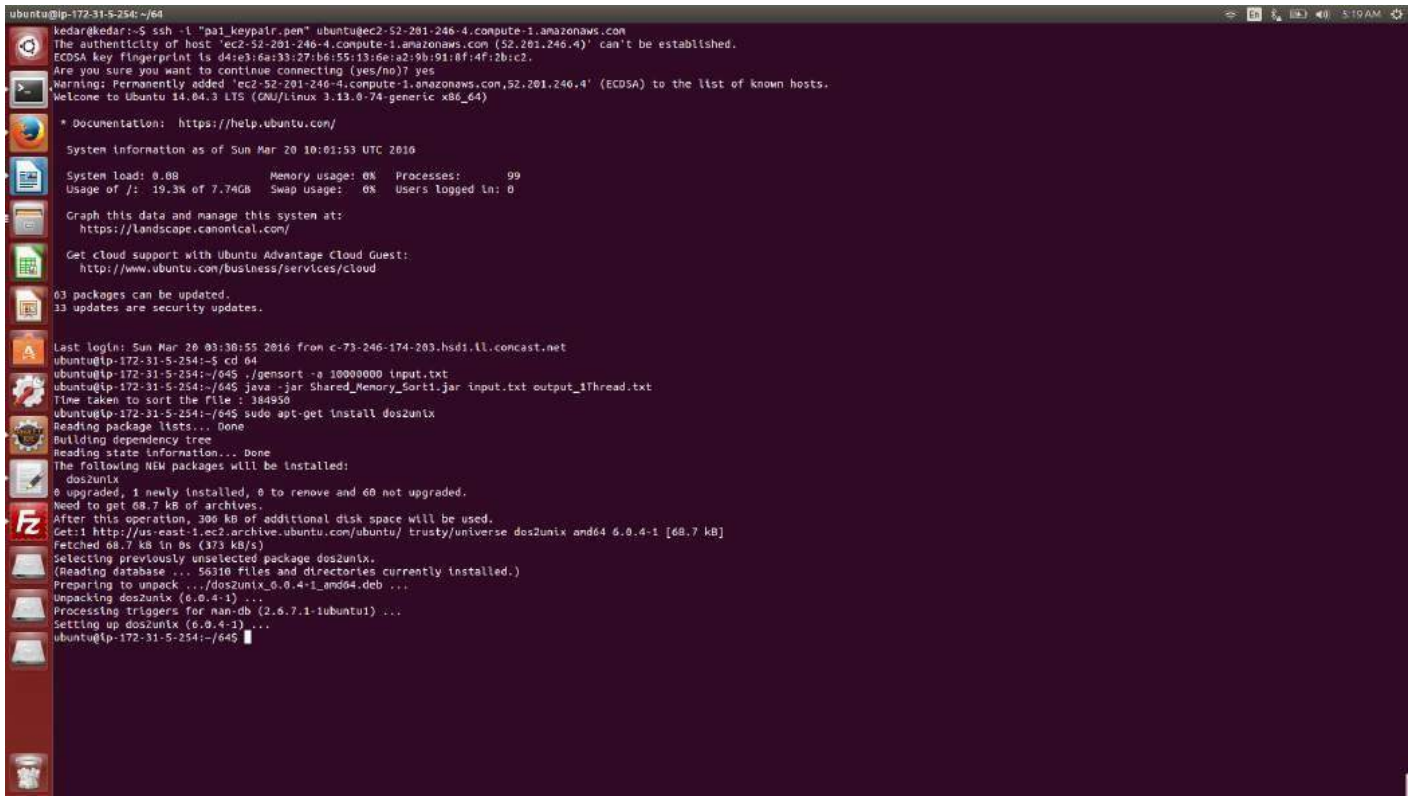
Time	1 Thread	2 Thread	4 Thread	8 Thread
Milliseconds	1091653	507420	507006	507899
Minutes	18.19	8.457	8.4501	8.4649

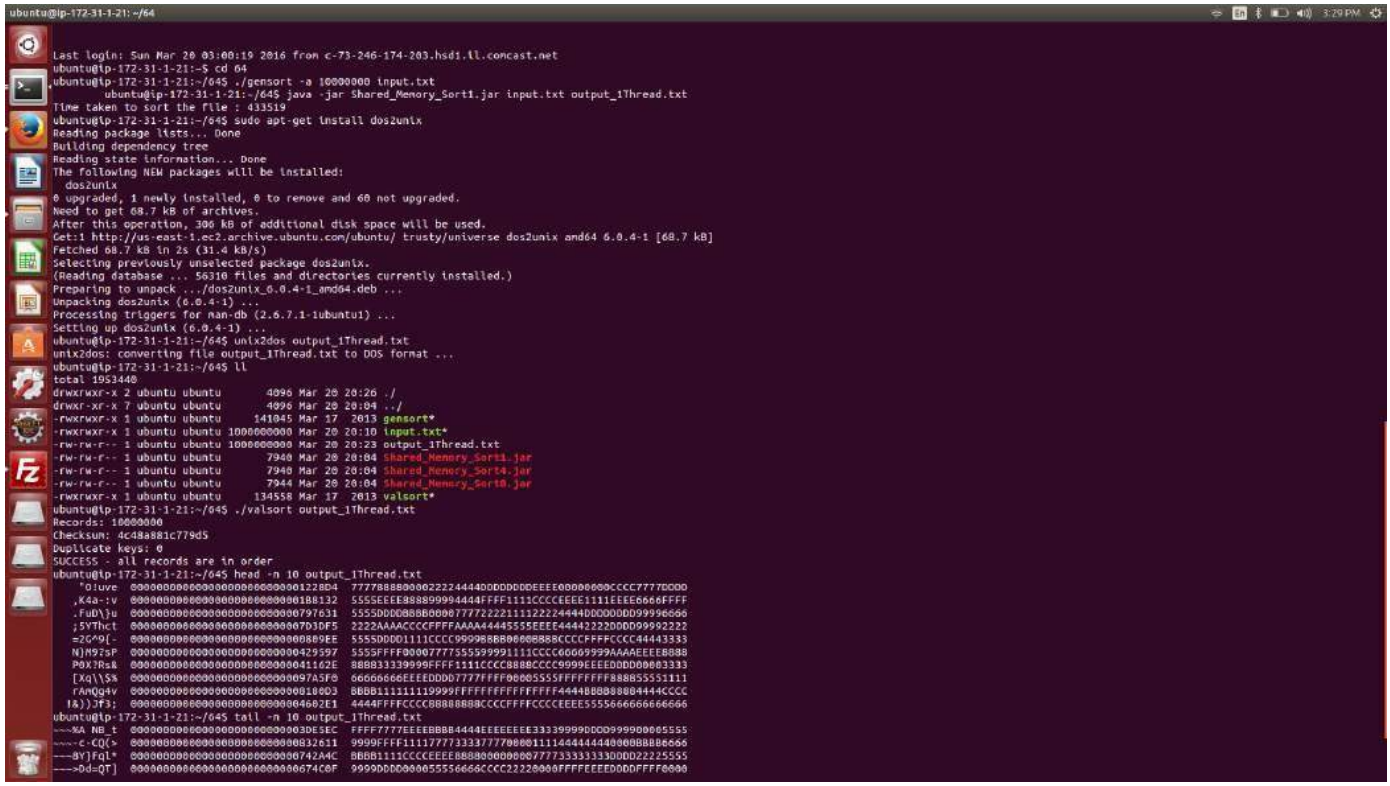
Shared Memory 1GB

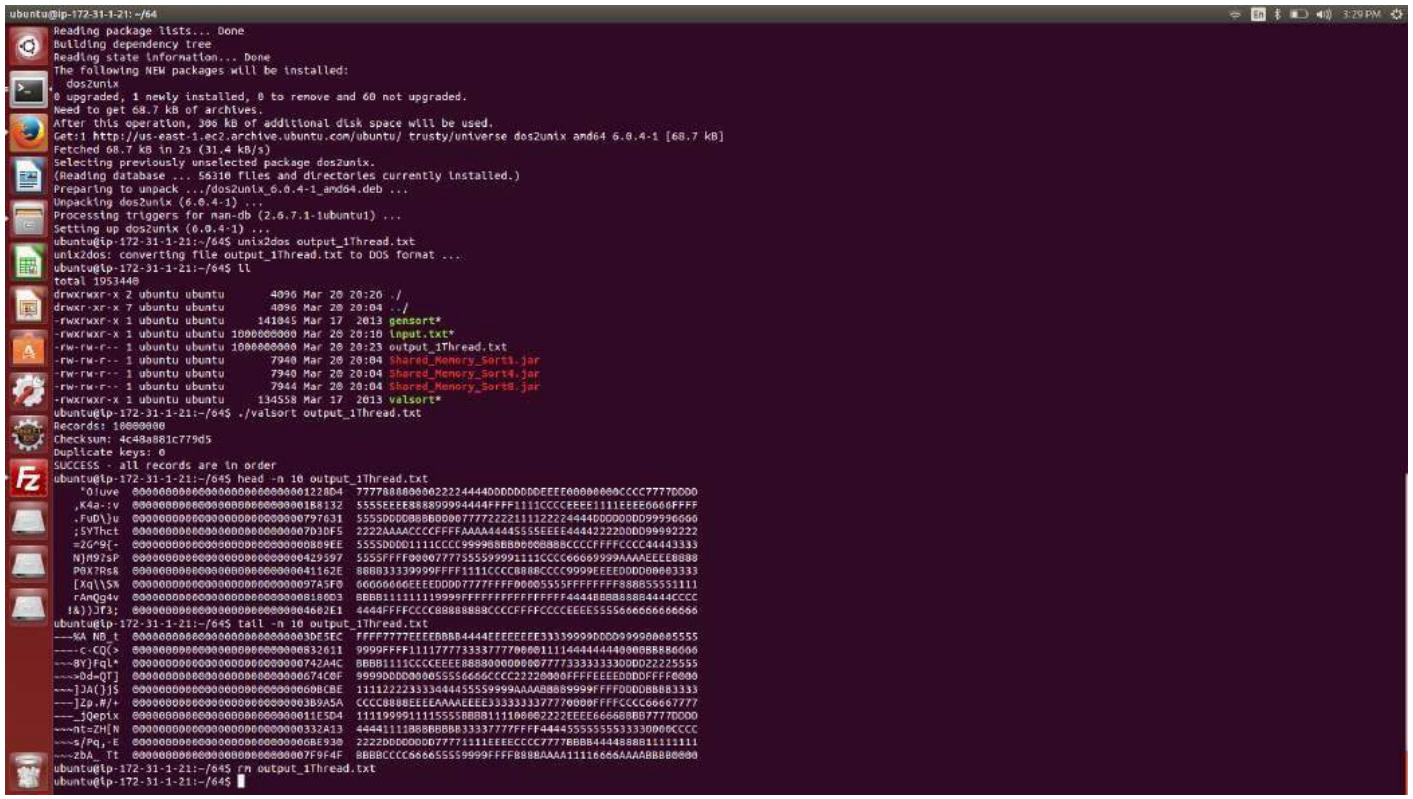
The dataset is changed from 10 GB to 1 GB with same setup of virtual cluster as for the 10GB dataset



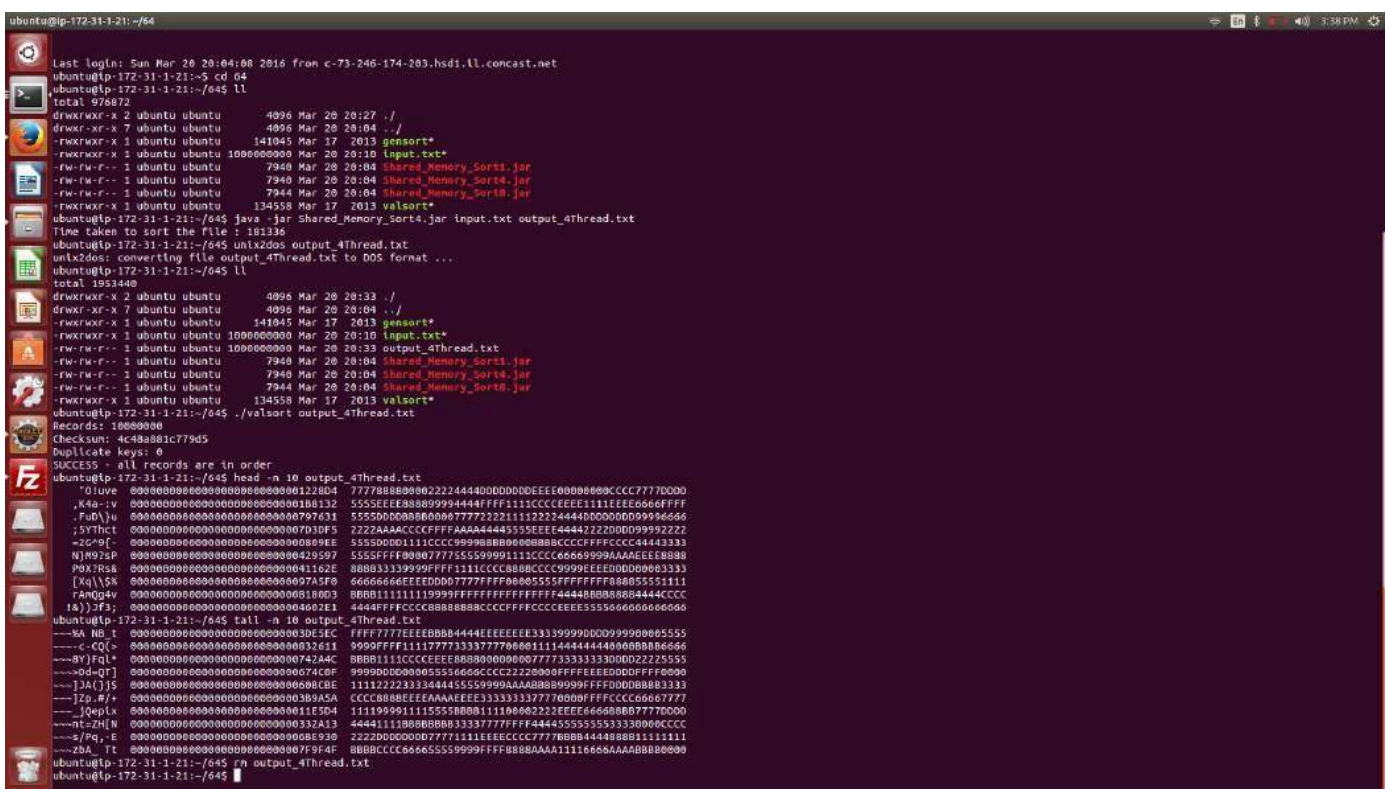
1 Thread.


ubuntu@ip-172-31-5-254: ~/64
kedar@kedar:~$ ssh -i "pa1_keypair.pem" ubuntu@ec2-52-201-246-4.compute-1.amazonaws.com
The authenticity of host 'ec2-52-201-246-4.compute-1.amazonaws.com (52.201.246.4)' can't be established.
ECDSA key fingerprint is d4:e3:6a:33:27:b6:55:13:6e:a2:9b:91:8f:4f:2b:c2.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'ec2-52-201-246-4.compute-1.amazonaws.com,52.201.246.4' (ECDSA) to the list of known hosts.
Welcome to Ubuntu 14.04.3 LTS (GNU/Linux 3.13.0-74-generic x86_64)
* Documentation: https://help.ubuntu.com/
System information as of Sun Mar 20 10:01:53 UTC 2016
System load: 0.08 Memory usage: 0% Processes: 99
Usage of /: 19.3% of 7.74GB Swap usage: 0% Users logged in: 0
Graph this data and manage this system at:
https://landscape.canonical.com/
Get cloud support with Ubuntu Advantage Cloud Guest:
http://www.ubuntu.com/business/services/cloud
63 packages can be updated.
33 updates are security updates.
Last login: Sun Mar 20 03:38:55 2016 from c-73-246-174-203.hsd1.ll.comcast.net
ubuntu@ip-172-31-5-254:~$ cd 64
ubuntu@ip-172-31-5-254:~/64$ ./gensort -a 10000000 input.txt
ubuntu@ip-172-31-5-254:~/64$ java -jar Shared_Memory_Sort1.jar input.txt output_1Thread.txt
Time taken to sort the file : 384950
ubuntu@ip-172-31-5-254:~/64$ sudo apt-get install dos2unix
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following NEW packages will be installed:
dos2unix
0 upgraded, 1 newly installed, 0 to remove and 60 not upgraded.
Need to get 68.7 kB of archives.
After this operation, 306 kB of additional disk space will be used.
Get:1 http://us-east-1.ec2.archive.ubuntu.com/ubuntu/ trusty/universe dos2unix amd64 6.0.4-1 [68.7 kB]
Fetched 68.7 kB in 0s (373 kB/s)
Selecting previously unselected package dos2unix.
(Reading database ... 56310 files and directories currently installed.)
Preparing to unpack .../dos2unix_6.0.4-1_amd64.deb ...
Unpacking dos2unix (6.0.4-1) ...
Processing triggers for man-db (2.6.7.1-1ubuntu1) ...
Setting up dos2unix (6.0.4-1) ...
ubuntu@ip-172-31-5-254:~/64$

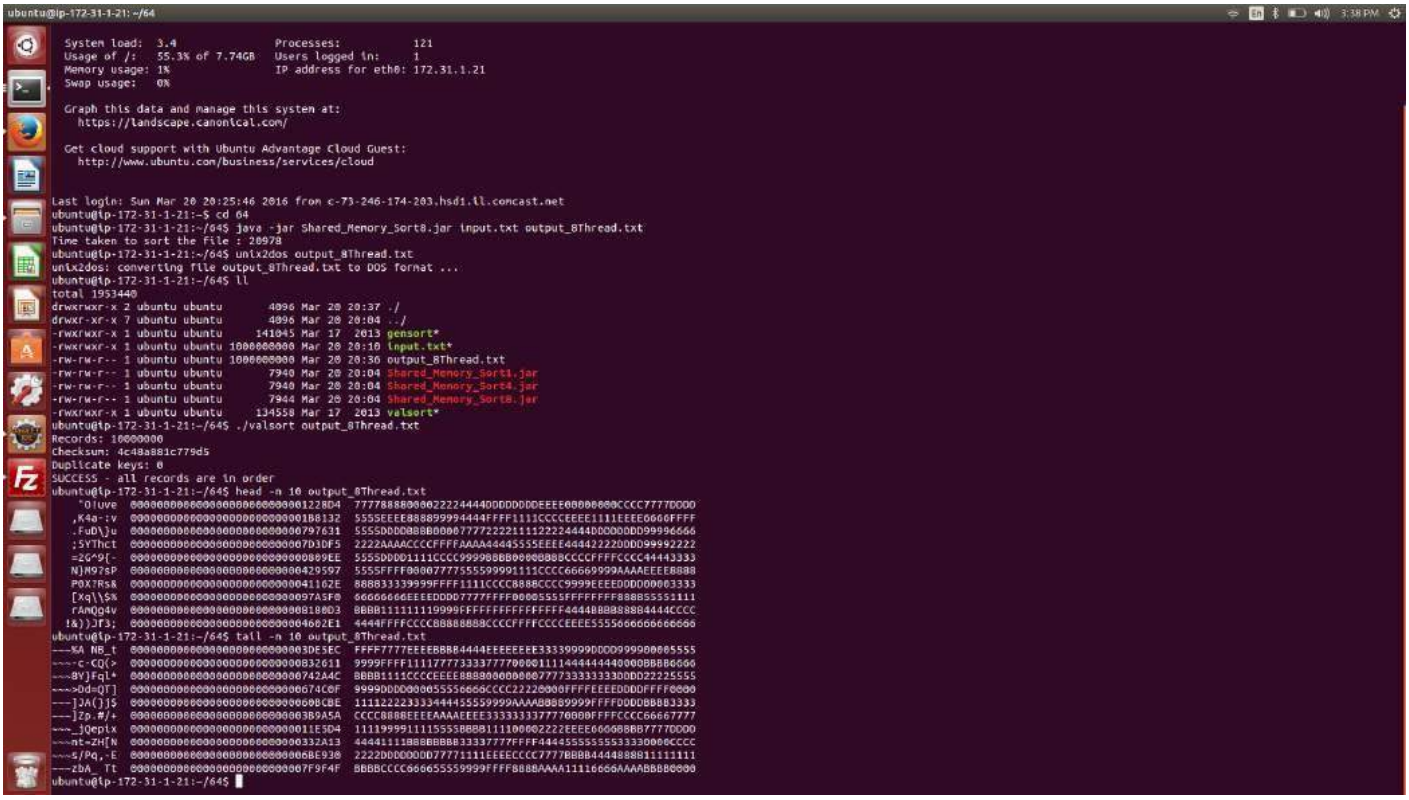

ubuntu@ip-172-31-1-21: ~/64
Last login: Sun Mar 20 03:00:19 2016 from c-73-246-174-203.hsd1.ll.comcast.net
ubuntu@ip-172-31-1-21:~$ cd 64
ubuntu@ip-172-31-1-21:~/64$ ./gensort -a 10000000 input.txt
ubuntu@ip-172-31-1-21:~/64$ java -jar Shared_Memory_Sort1.jar input.txt output_1Thread.txt
Time taken to sort the file : 433519
ubuntu@ip-172-31-1-21:~/64$ sudo apt-get install dos2unix
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following NEW packages will be installed:
dos2unix
0 upgraded, 1 newly installed, 0 to remove and 60 not upgraded.
Need to get 68.7 kB of archives.
After this operation, 306 kB of additional disk space will be used.
Get:1 http://us-east-1.ec2.archive.ubuntu.com/ubuntu/ trusty/universe dos2unix amd64 6.0.4-1 [68.7 kB]
Fetched 68.7 kB in 2s (31.4 kB/s)
Selecting previously unselected package dos2unix.
(Reading database ... 56310 files and directories currently installed.)
Preparing to unpack .../dos2unix_6.0.4-1_amd64.deb ...
Unpacking dos2unix (6.0.4-1) ...
Processing triggers for man-db (2.6.7.1-1ubuntu1) ...
Setting up dos2unix (6.0.4-1) ...
ubuntu@ip-172-31-1-21:~/64$ unix2dos output_1Thread.txt
unix2dos: converting file output_1Thread.txt to DOS format ...
ubuntu@ip-172-31-1-21:~/64$ ll
total 1953448
drwxrwxr-x 2 ubuntu ubuntu 4096 Mar 20 20:26 ./
drwxr-xr-x 7 ubuntu ubuntu 4096 Mar 20 20:04 ../
-rwxrwxr-x 1 ubuntu ubuntu 141045 Mar 17 2013 gensort*
-rwxrwxr-x 1 ubuntu ubuntu 1000000000 Mar 20 20:10 input.txt*
-rw-rw-r-- 1 ubuntu ubuntu 1000000000 Mar 20 20:23 output_1Thread.txt
-rw-rw-r-- 1 ubuntu ubuntu 7940 Mar 20 20:04 Shared_Memory_Sort1.jar
-rw-rw-r-- 1 ubuntu ubuntu 7940 Mar 20 20:04 Shared_Memory_Sort4.jar
-rw-rw-r-- 1 ubuntu ubuntu 7944 Mar 20 20:04 Shared_Memory_Sort8.jar
-rwxrwxr-x 1 ubuntu ubuntu 134558 Mar 17 2013 valsort*
ubuntu@ip-172-31-1-21:~/64$ ./valsort output_1Thread.txt
Records: 10000000
Checksum: 4c48a881c779d5
Duplicate keys: 0
SUCCESS - all records are in order
ubuntu@ip-172-31-1-21:~/64$ head -n 10 output_1Thread.txt
ubuntu@ip-172-31-1-21:~/64$ tail -n 10 output_1Thread.txt


ubuntu@ip-172-31-1-21: ~/64
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following NEW packages will be installed:
dos2unix
0 upgraded, 1 newly installed, 0 to remove and 60 not upgraded.
Need to get 68.7 kB of archives.
After this operation, 306 kB of additional disk space will be used.
Get:1 http://us-east-1.ec2.archive.ubuntu.com/ubuntu/ trusty/universe dos2unix amd64 6.0.4-1 [68.7 kB]
Fetched 68.7 kB in 2s (31.4 kB/s)
Selecting previously unselected package dos2unix.
(Reading database ... 56310 files and directories currently installed.)
Preparing to unpack .../dos2unix_6.0.4-1_amd64.deb ...
Unpacking dos2unix (6.0.4-1) ...
Processing triggers for man-db (2.6.7.1-1ubuntu1) ...
Setting up dos2unix (6.0.4-1) ...
ubuntu@ip-172-31-1-21:~/64$ unix2dos output_1Thread.txt
unix2dos: converting file output_1Thread.txt to DOS format ...
ubuntu@ip-172-31-1-21:~/64$ ll
total 1953440
drwxrwxr-x 2 ubuntu ubuntu 4096 Mar 20 20:20 ./
drwxr-xr-x 7 ubuntu ubuntu 4096 Mar 20 20:04 ../
-rwxrwxr-x 1 ubuntu ubuntu 141045 Mar 17 2013 gensort*
-rwxrwxr-x 1 ubuntu ubuntu 1000000000 Mar 20 20:10 input.txt*
-rw-rw-r-- 1 ubuntu ubuntu 1000000000 Mar 20 20:23 output_1Thread.txt
-rw-rw-r-- 1 ubuntu ubuntu 7940 Mar 20 20:04 Shared_Memory_Sort1.jar
-rw-rw-r-- 1 ubuntu ubuntu 7940 Mar 20 20:04 Shared_Memory_Sort4.jar
-rw-rw-r-- 1 ubuntu ubuntu 7944 Mar 20 20:04 Shared_Memory_Sort8.jar
-rwxrwxr-x 1 ubuntu ubuntu 134558 Mar 17 2013 valsort*
ubuntu@ip-172-31-1-21:~/64$ ./valsort output_1Thread.txt
Records: 10000000
Checksum: 4c48a881c779d5
Duplicate keys: 0
SUCCESS - all records are in order
ubuntu@ip-172-31-1-21:~/64$ head -n 10 output_1Thread.txt
ubuntu@ip-172-31-1-21:~/64$ tail -n 10 output_1Thread.txt
ubuntu@ip-172-31-1-21:~/64$ rm output_1Thread.txt
ubuntu@ip-172-31-1-21:~/64$

4 Threads


ubuntu@ip-172-31-1-21: ~/64
Last login: Sun Mar 20 20:04:08 2016 from c-73-246-174-203.hsd1.ll.comcast.net
ubuntu@ip-172-31-1-21:~$ cd 64
ubuntu@ip-172-31-1-21:~/64$ ll
total 976872
drwxrwxr-x 2 ubuntu ubuntu 4096 Mar 20 20:27 ./
drwxr-xr-x 7 ubuntu ubuntu 4096 Mar 20 20:04 ../
-rwxrwxr-x 1 ubuntu ubuntu 141045 Mar 17 2013 gensort*
-rwxrwxr-x 1 ubuntu ubuntu 1000000000 Mar 20 20:10 input.txt*
-rw-rw-r-- 1 ubuntu ubuntu 7940 Mar 20 20:04 Shared_Memory_Sort1.jar
-rw-rw-r-- 1 ubuntu ubuntu 7940 Mar 20 20:04 Shared_Memory_Sort4.jar
-rw-rw-r-- 1 ubuntu ubuntu 7944 Mar 20 20:04 Shared_Memory_Sort8.jar
-rwxrwxr-x 1 ubuntu ubuntu 134558 Mar 17 2013 valsort*
ubuntu@ip-172-31-1-21:~/64$ java -jar Shared_Memory_Sort4.jar input.txt output_4Thread.txt
Time taken to sort the file : 181336
ubuntu@ip-172-31-1-21:~/64$ unix2dos output_4Thread.txt
unix2dos: converting file output_4Thread.txt to DOS format ...
ubuntu@ip-172-31-1-21:~/64$ ll
total 1953440
drwxrwxr-x 2 ubuntu ubuntu 4096 Mar 20 20:33 ./
drwxr-xr-x 7 ubuntu ubuntu 4096 Mar 20 20:04 ../
-rwxrwxr-x 1 ubuntu ubuntu 141045 Mar 17 2013 gensort*
-rwxrwxr-x 1 ubuntu ubuntu 1000000000 Mar 20 20:10 input.txt*
-rw-rw-r-- 1 ubuntu ubuntu 1000000000 Mar 20 20:33 output_4Thread.txt
-rw-rw-r-- 1 ubuntu ubuntu 7940 Mar 20 20:04 Shared_Memory_Sort1.jar
-rw-rw-r-- 1 ubuntu ubuntu 7940 Mar 20 20:04 Shared_Memory_Sort4.jar
-rw-rw-r-- 1 ubuntu ubuntu 7944 Mar 20 20:04 Shared_Memory_Sort8.jar
-rwxrwxr-x 1 ubuntu ubuntu 134558 Mar 17 2013 valsort*
ubuntu@ip-172-31-1-21:~/64$ ./valsort output_4Thread.txt
Records: 10000000
Checksum: 4c48a881c779d5
Duplicate keys: 0
SUCCESS - all records are in order
ubuntu@ip-172-31-1-21:~/64$ head -n 10 output_4Thread.txt
ubuntu@ip-172-31-1-21:~/64$ tail -n 10 output_4Thread.txt
ubuntu@ip-172-31-1-21:~/64$ rm output_4Thread.txt
ubuntu@ip-172-31-1-21:~/64$

8 Threads.



Performance

Time	1 Thread	2 Thread	4 Thread	8 Thread
Milliseconds	1091653	507420	507006	507899
Minutes	18.19	8.457	8.4501	8.4649

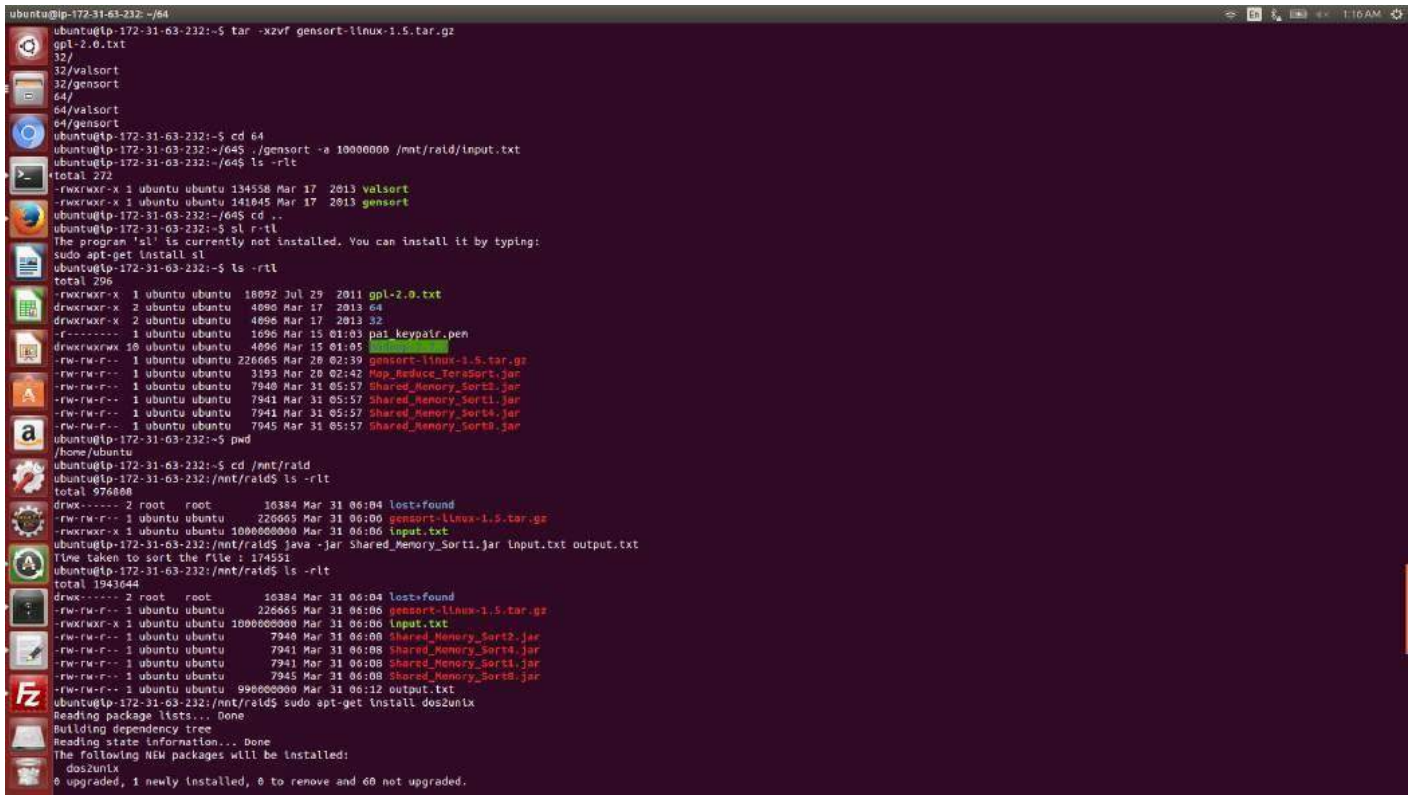
Throughput = (Datasize / Time taken)
= 10000 / 507.42
= 19.70 Mb/sec

Shared Memory 1GB

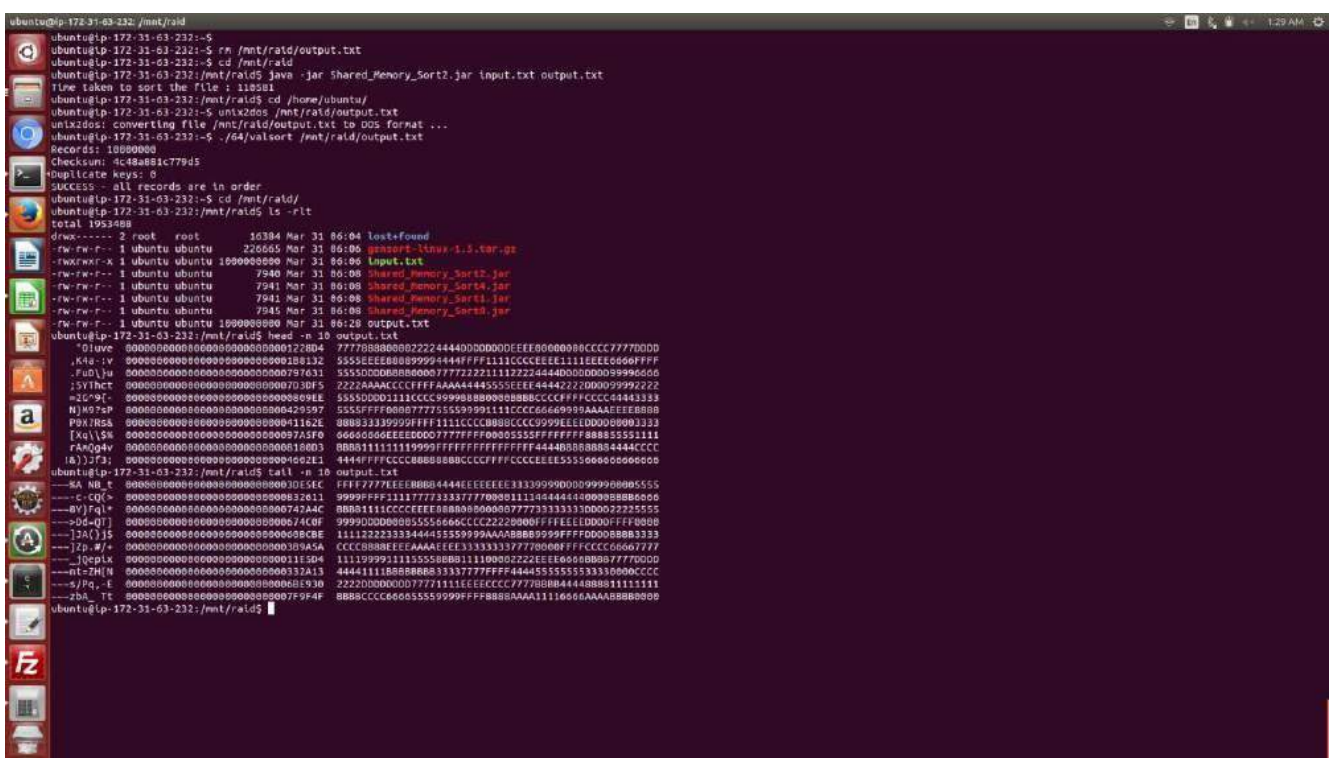
We follow the same steps as done for the shared memory 10 GB dataset but now we have done it for 1 GB dataset.

Below mentioned are the screenshot for the same.

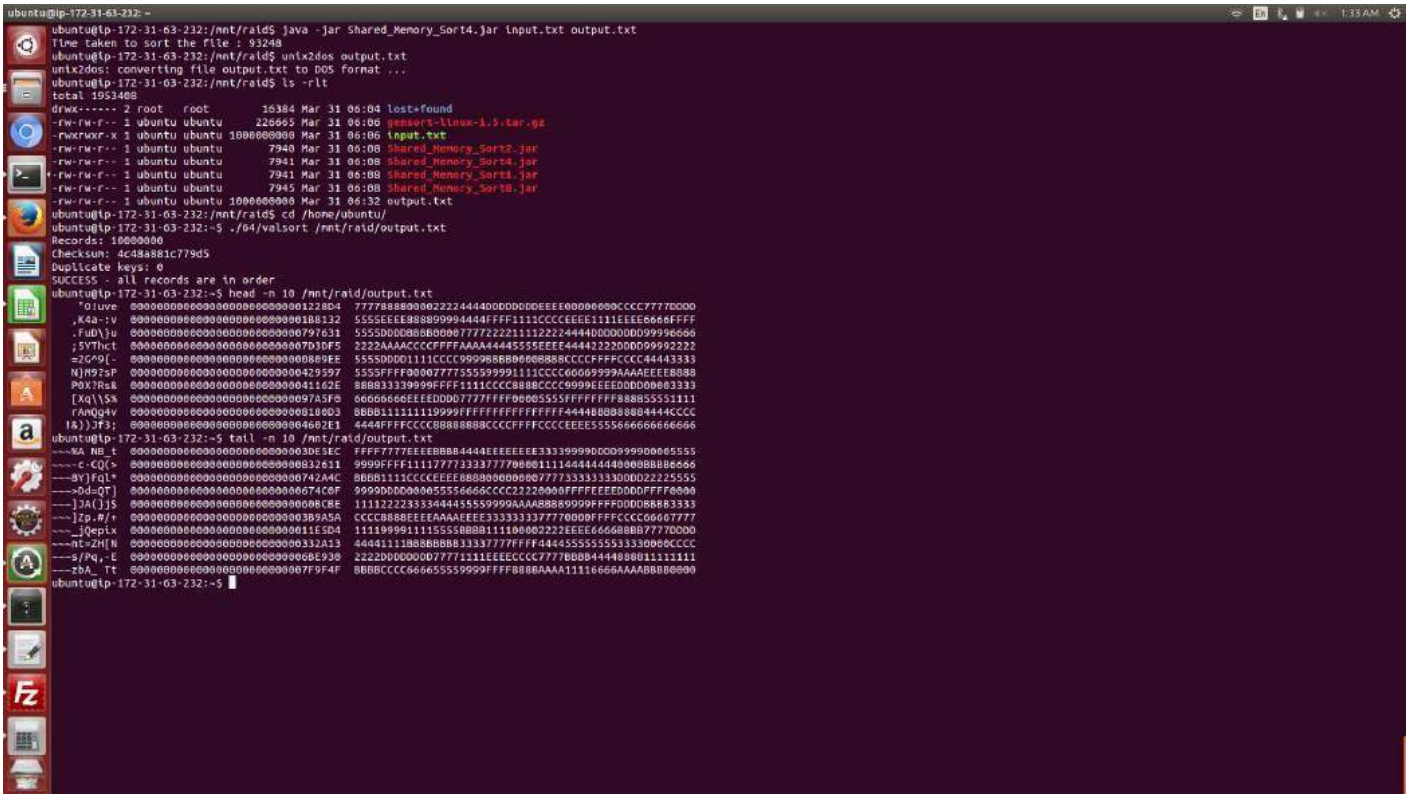
1 Thread.



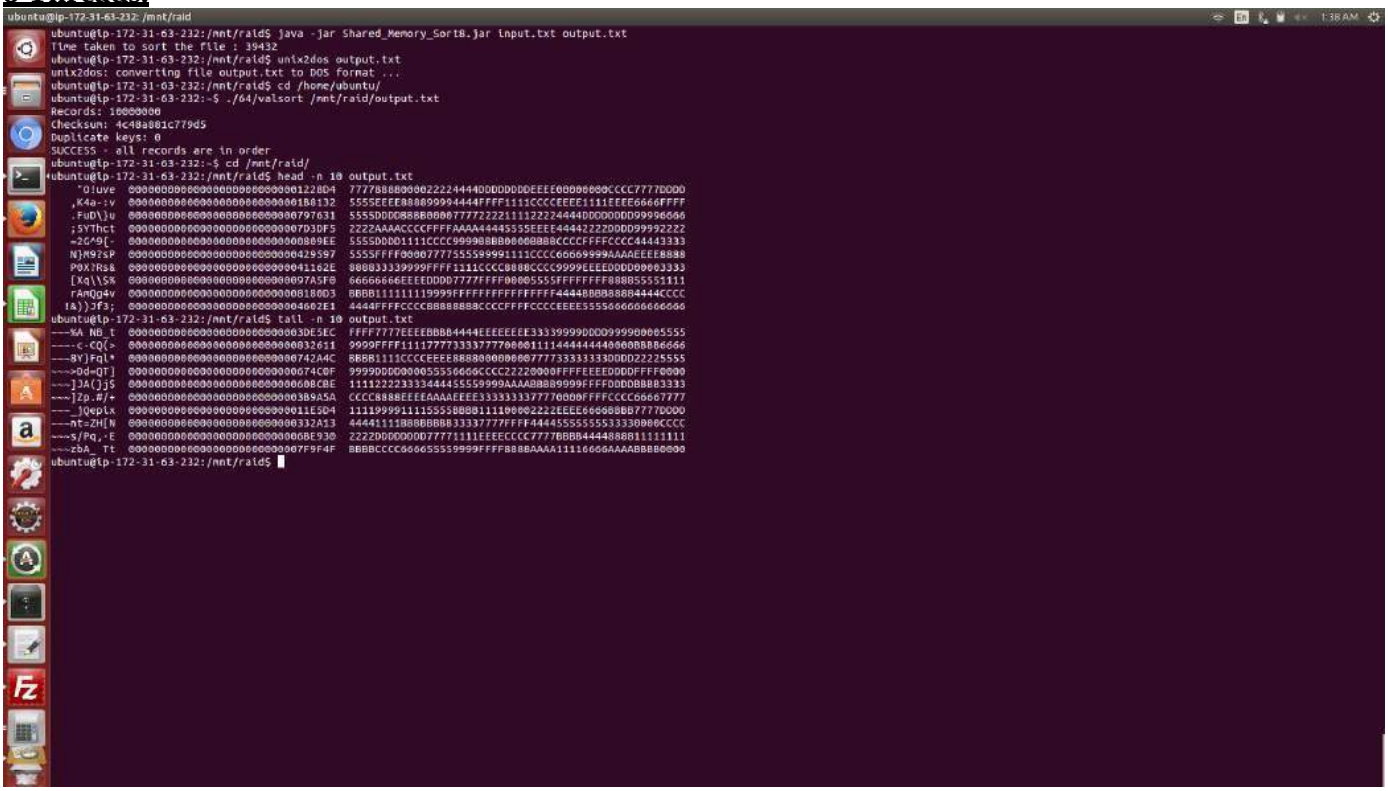
2 Threads.



4 Threads.



8 Threads.



Performance:

Time	1 Thread	2 Thread	4 Thread	8 Thread
Milliseconds	174551	110581	93248	39432
Minutes	2.909	1.8430	1.5541	0.32

Throughput = (Datasize / Time taken)
= 1000 / 93.24
= 10.70 Mb/sec

Conclusion:

- The worst efficiency is observed on single thread for both 1GB and 10 GB of dataset. Once multiple threads are used we get a better performance of approximately 50% improvement.
- For 10 GB, the experiment is performed on c3.large instance. The number of v-cores in c3.large instance is 2. Hence till parallel processing with number of threads 2 , we start getting better performance and thus after that , the timing does not change significantly.

- For 1 GB, the experiment is performed on d2.xlarge instance with number of vcores = 4. Hence we start getting good performance with 4 levels of multithreading. After increasing the number of threads more than 4 , the performance does not change significantly.

- Hence performance if observed better when the application is multithreaded rather than running the application on single thread.

- The Best performance is observed on 4 Thread with difference of hardly few milliseconds between 2 and 8 threads.

Hadoop

- What is Master Node? What is a Slave Node?
- Before Master and Slave node comes the HDFS which is one primary component of Hadoop cluster and it is designed to have the Master Slave Architecture.
- The Master node is also called the name node that is responsible for all the file system namespace operations.
- The Namespace operations include open, close rename files and directories.
- Master node is also responsible for determining the mapping of blocks to Slaves I.e. Data nodes and working on file permissions of accessing it by clients.
- Slaves are the data nodes responsible for file read/write requests from the FS along with block management (replication, creation, deletion) as per instruction from Master Nodes.
- Similar to HDFS, Map-Reduce is also another primary component of Hadoop having master slave architecture.
- In Map-Reduce Architecture Master node acts as a Job Tracker i.e. point of interaction between user and MR framework. The Jobtracker puts tasks in queue in FIFO format and manages the assignment of map-reduce tasks to the taskbar. Slaves execute the tasks upon instruction from jobtracker.

- Why do we need to set unique available ports to those configuration files on a shared environment ? What errors or side effects will show if we use the same port number for each user
- If we use the same port number for each resources of configuration file, some of the resources required for Hadoop cluster to run map-reduce will not work.
- For Ex. If the node manager is set to port 8080 and the same port number in assigned to data node in HDFS, then either of the two won't start and the jobtracker will not reflect one of those resources. Hence it is necessary to write unique ports for each resources and it should be unique per individual user as the resources for each user should be unique. It can be distinguished with different Private IP or public DNS for each user.

- How do we change the number of mappers or reducers from the configuration file?
- By default, the number of mappers and reducers are 2 and 1 respectively.
- This can be changed by adding a property in mapred-site.xml as follows,

```
<property>
  <name>mapreduce.job.reduces</name>
  <value>30</value>
</property>

<property>
  <name>mapreduce.job.maps</name>
  <value>4</value>
</property>
```

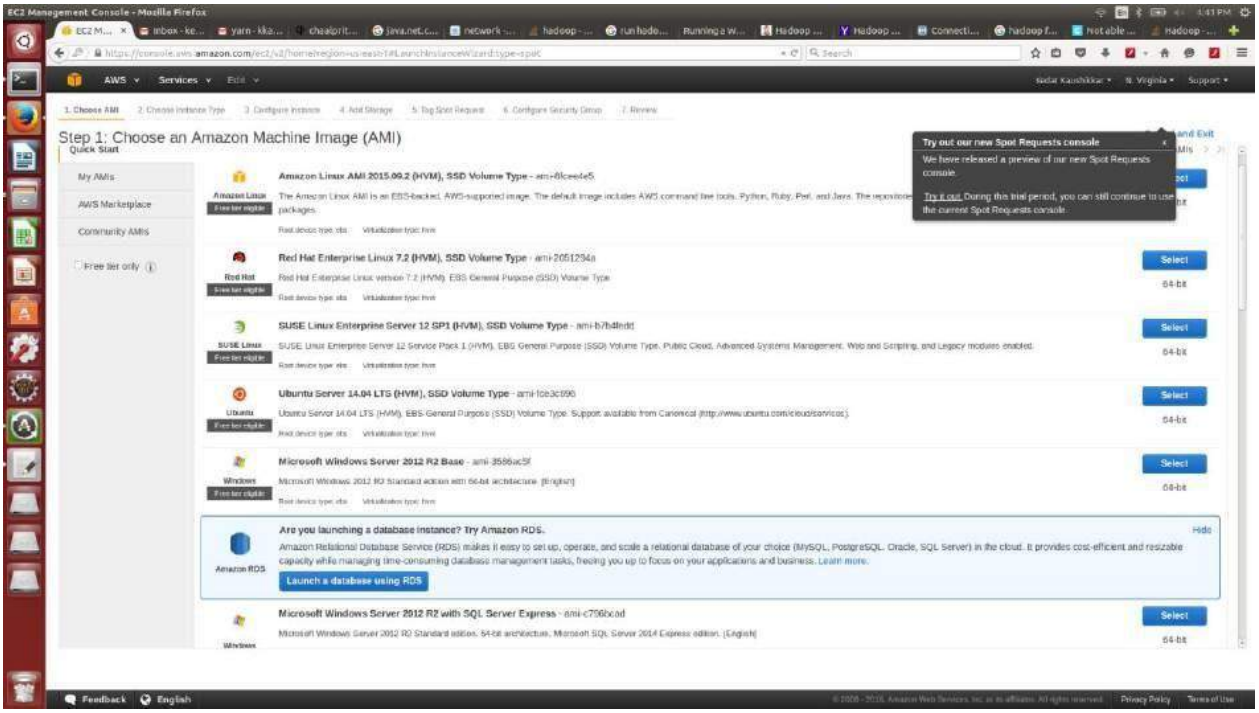
The above property will change the number of reduces to 30 and number of maps to 4.

Setting up Virtual Cluster 1 Node:

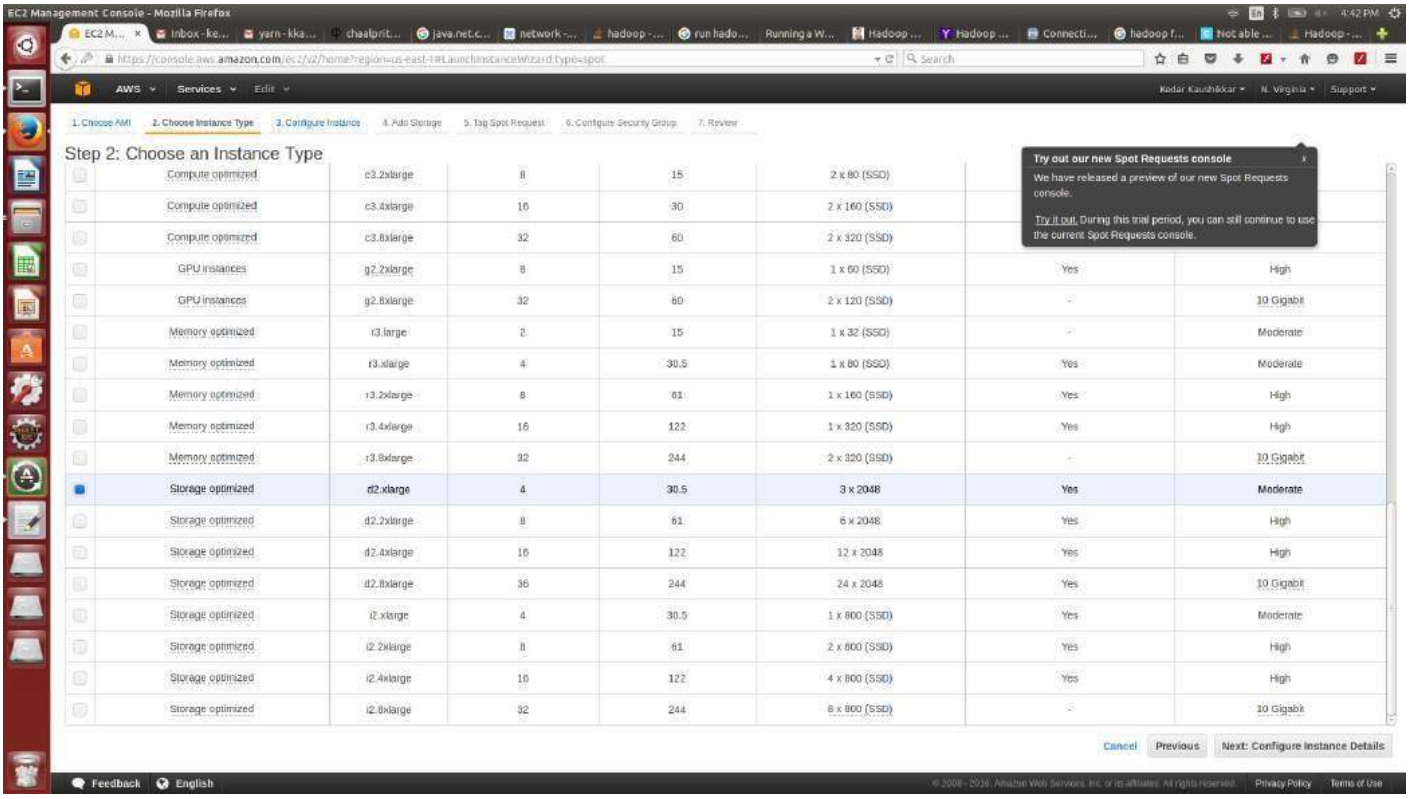
Installation Steps for Hadoop on Virtual Cluster of 1 Node on Amazon EC2 using d2.xlarge / c3.large instances.

- On the Launch Instance page, we need to choose the type of AMI

- For our installation, we choose the Ubuntu Server 14.04 LTS (HVM), SSD Volume Type

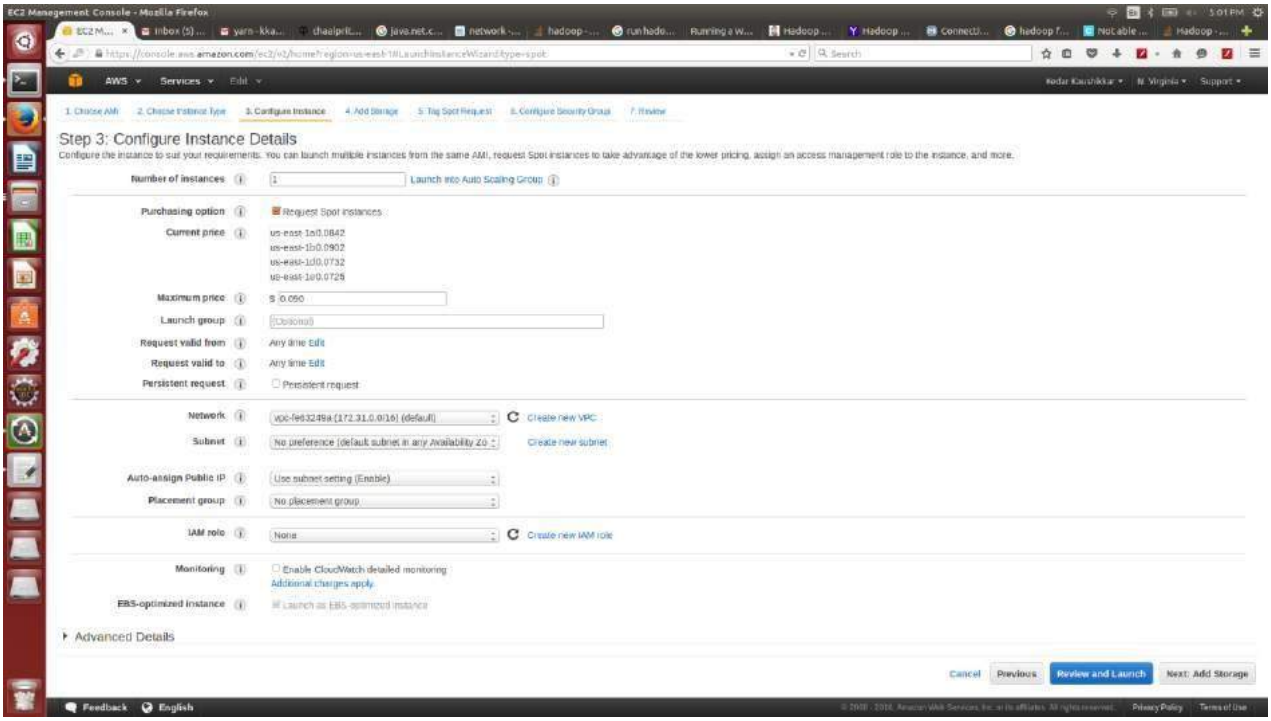


Once AMI is chosen we select the type of instance that we want for our Virtual Cluster. For Installation Steps we chose d2.xlarge instance that gives 3 *2048 Tb of disk Volume.

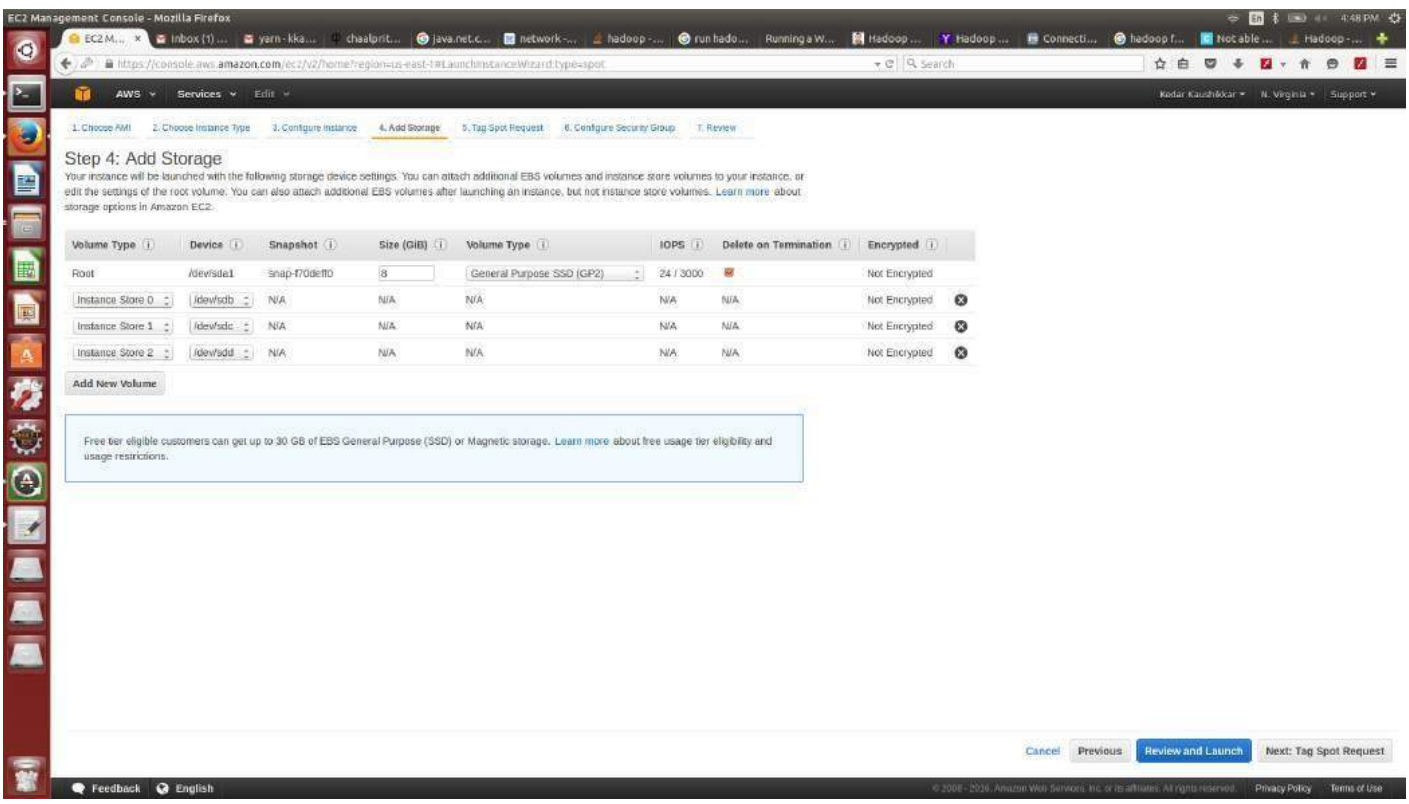


Compute optimized	c3.2xlarge	8	15	2 x 80 (SSD)		
Compute optimized	c3.4xlarge	16	30	2 x 160 (SSD)		
Compute optimized	c3.8xlarge	32	60	2 x 320 (SSD)		
GPU instances	g2.2xlarge	8	15	1 x 60 (SSD)	Yes	High
GPU instances	g2.8xlarge	32	60	2 x 120 (SSD)	-	10 Gigabit
Memory optimized	r3.large	2	15	1 x 32 (SSD)	-	Moderate
Memory optimized	r3.xlarge	4	30.5	1 x 80 (SSD)	Yes	Moderate
Memory optimized	r3.2xlarge	8	61	1 x 160 (SSD)	Yes	High
Memory optimized	r3.4xlarge	16	122	1 x 320 (SSD)	Yes	High
Memory optimized	r3.8xlarge	32	244	2 x 320 (SSD)	-	10 Gigabit
Storage optimized	d2.xlarge	4	30.5	3 x 2048	Yes	Moderate
Storage optimized	d2.2xlarge	8	61	6 x 2048	Yes	High
Storage optimized	d2.4xlarge	16	122	12 x 2048	Yes	High
Storage optimized	d2.8xlarge	36	244	24 x 2048	Yes	10 Gigabit
Storage optimized	i2.xlarge	4	30.5	1 x 800 (SSD)	Yes	Moderate
Storage optimized	i2.2xlarge	8	61	2 x 800 (SSD)	Yes	High
Storage optimized	i2.4xlarge	16	122	4 x 800 (SSD)	Yes	High
Storage optimized	i2.8xlarge	32	244	8 x 800 (SSD)	-	10 Gigabit

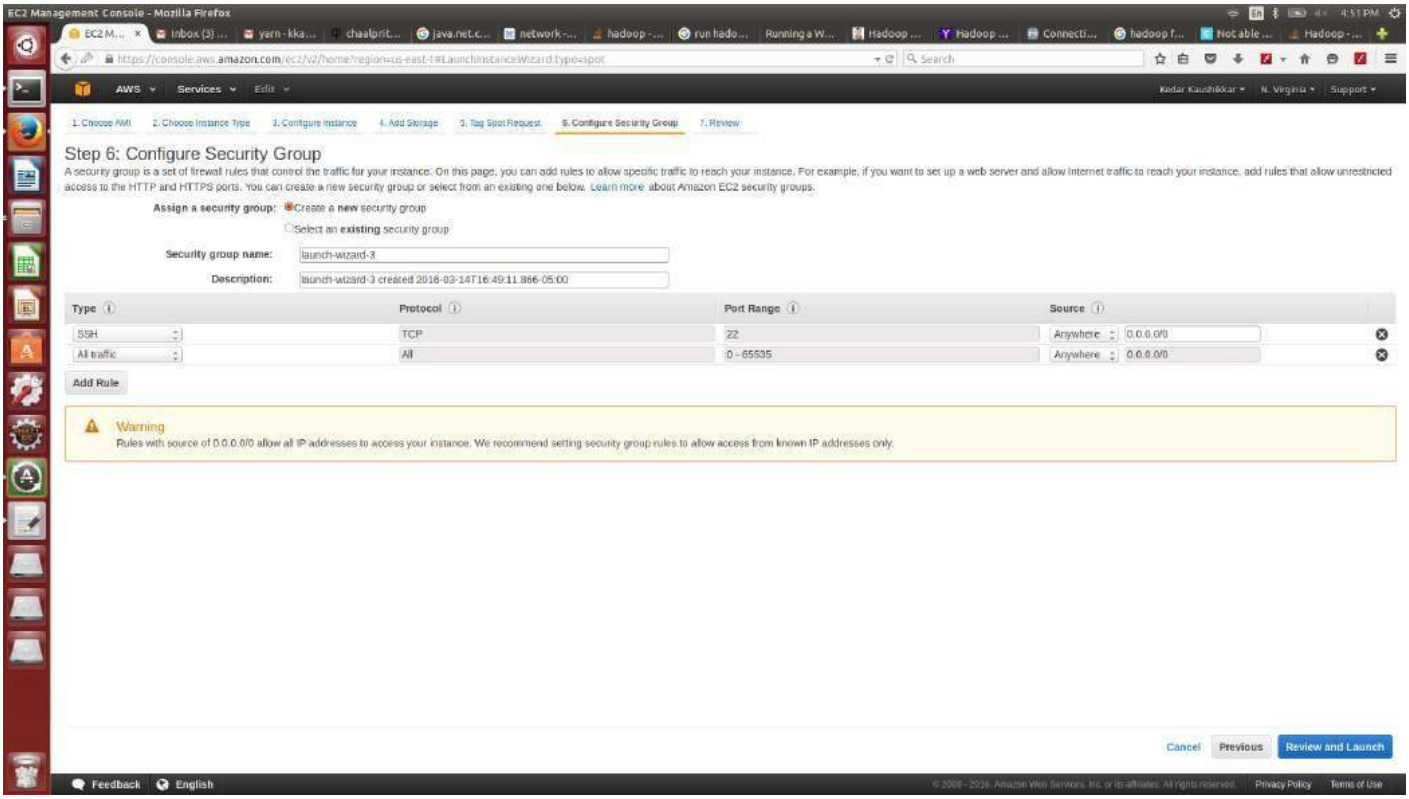
Next step is to the Configure the Instance Details where we type in the number of instances required and choosing the purchasing option i.e. Spot Instance / On-Demand



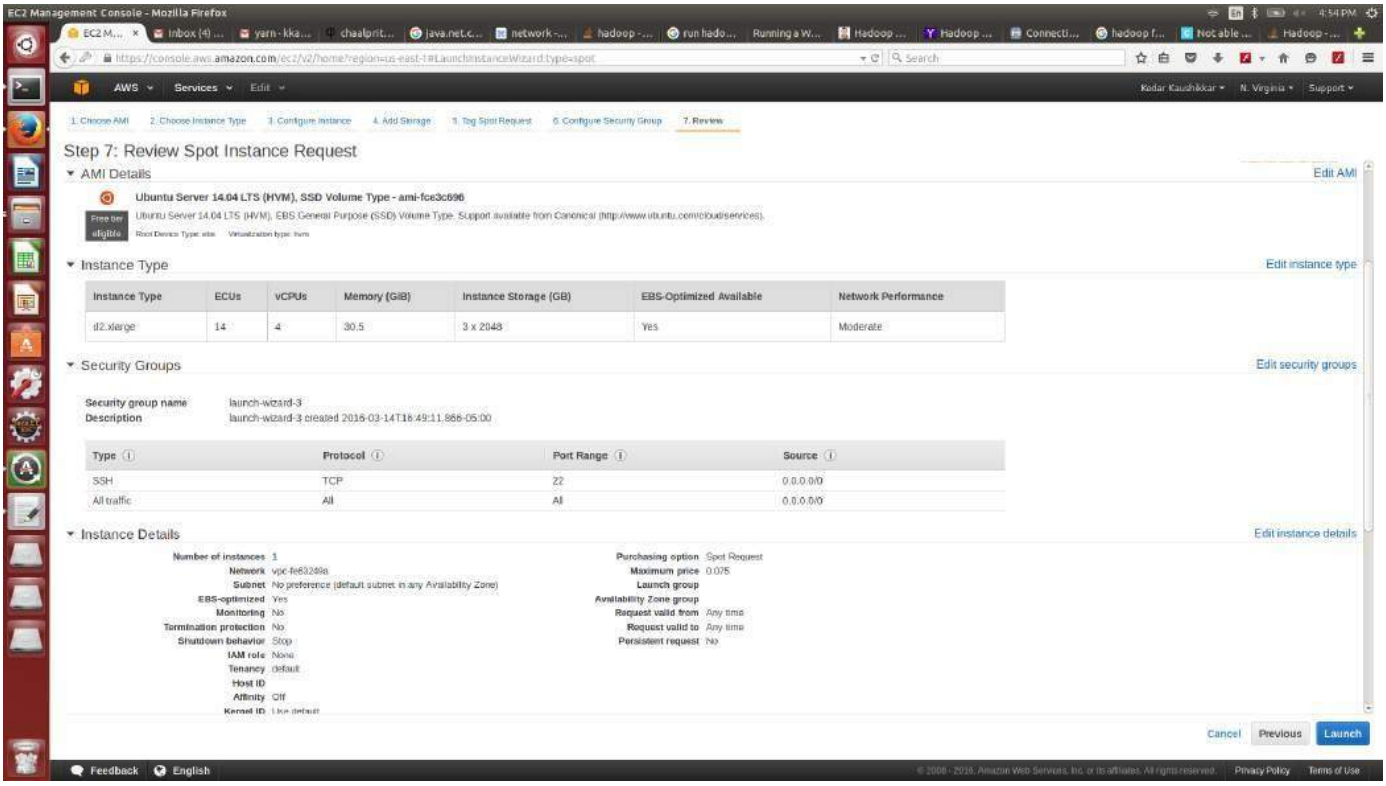
On configuring the instance , we can choose the Storage type and the size of the disk volume.

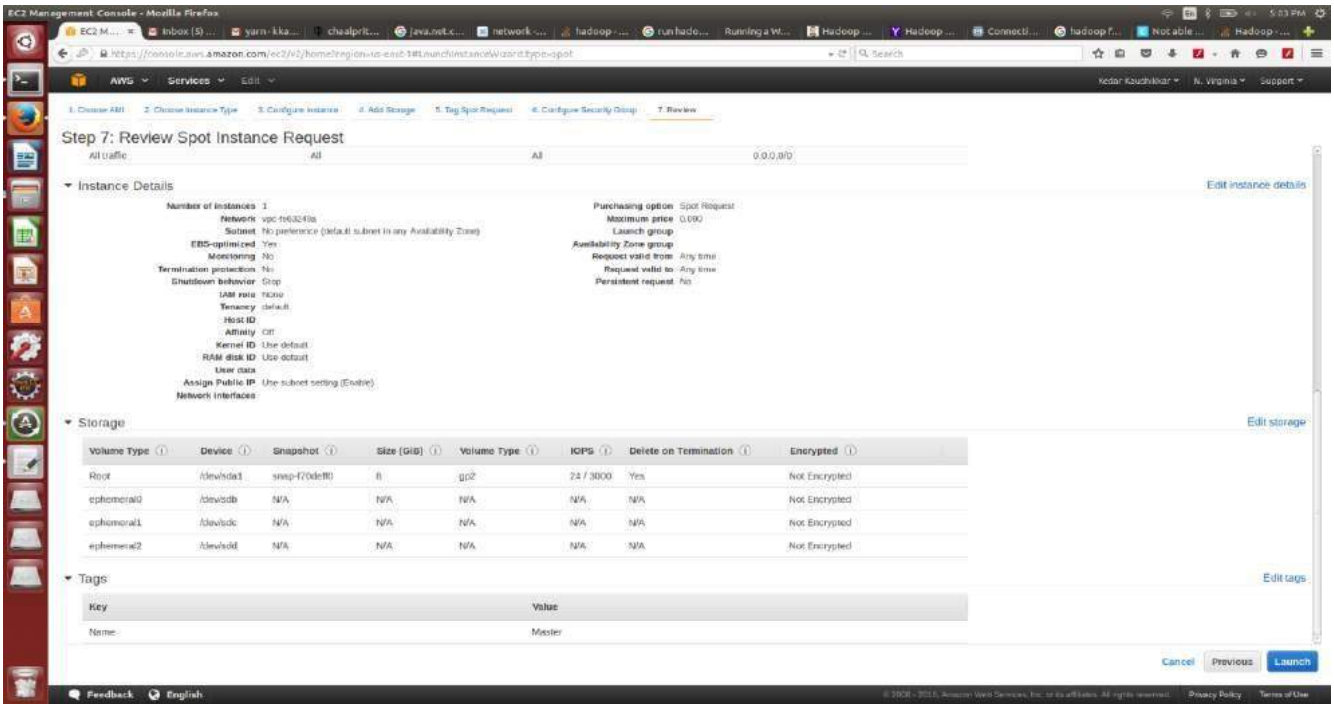


Step 6 is to configure the security group and provide the port range to connect our instances.



Once all the configurations are complete, we can review our configuration settings, storage type, security configurations.





- To connect this above instance followed the following steps.

1. We used chmod 400 My_Key.pem to make sure your private key file isn't publicly viewable.
2. Use the **ssh** command to connect to the instance.

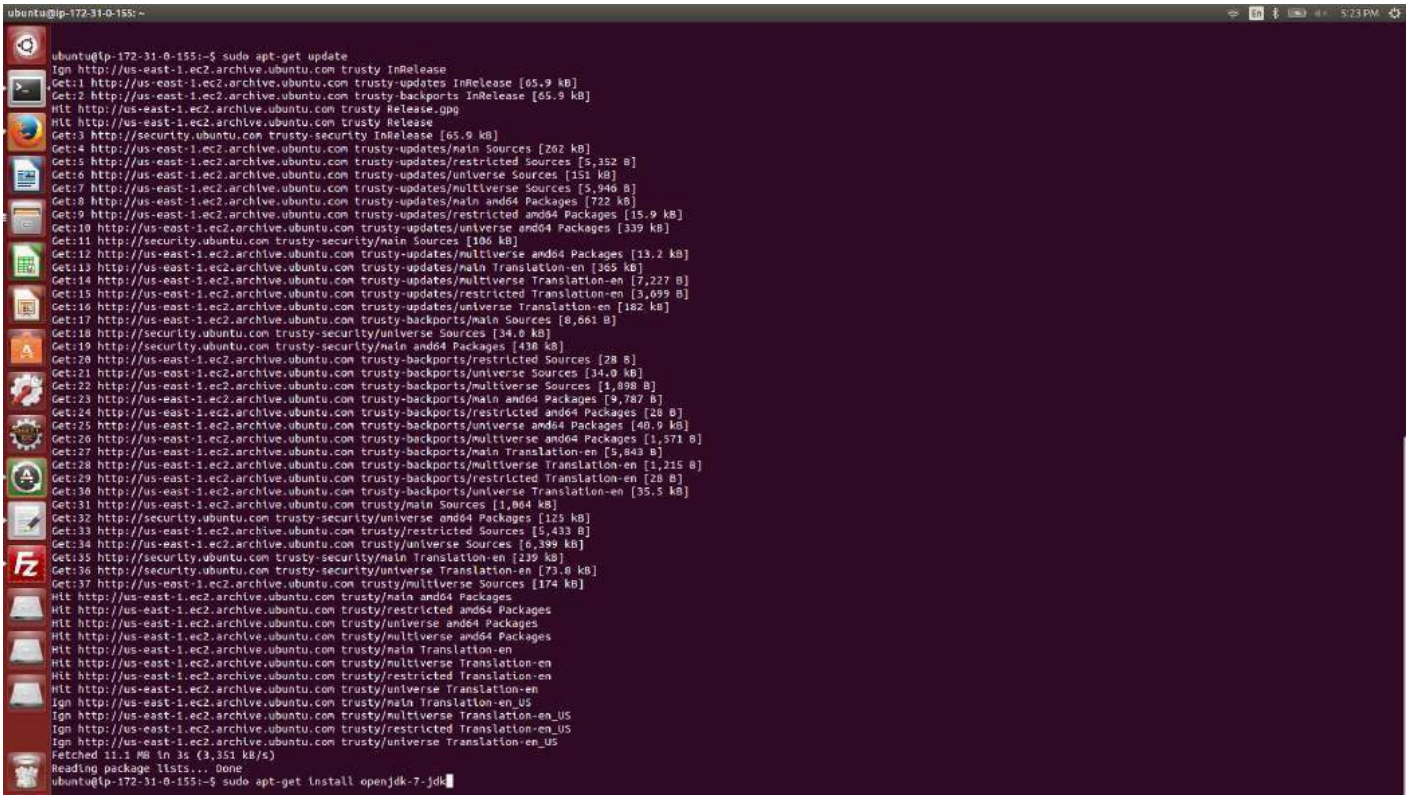
ssh –i 'pemFilePath' ubuntu@PublicDNS

- This command will log in to your instance where you will be installing your Hadoop Files.

- Once the instance is started, to run Hadoop Jobs we need to configure the amazon ec2 instance to be updated with latest Linux files and install Java Development Kit and Java Runtime Environment.

- After connection is established we need to install Java on our virtual cluster
  sudo apt-get update

- Install Default JDK and JRE
  sudo apt-get install openjdk-7-jdk

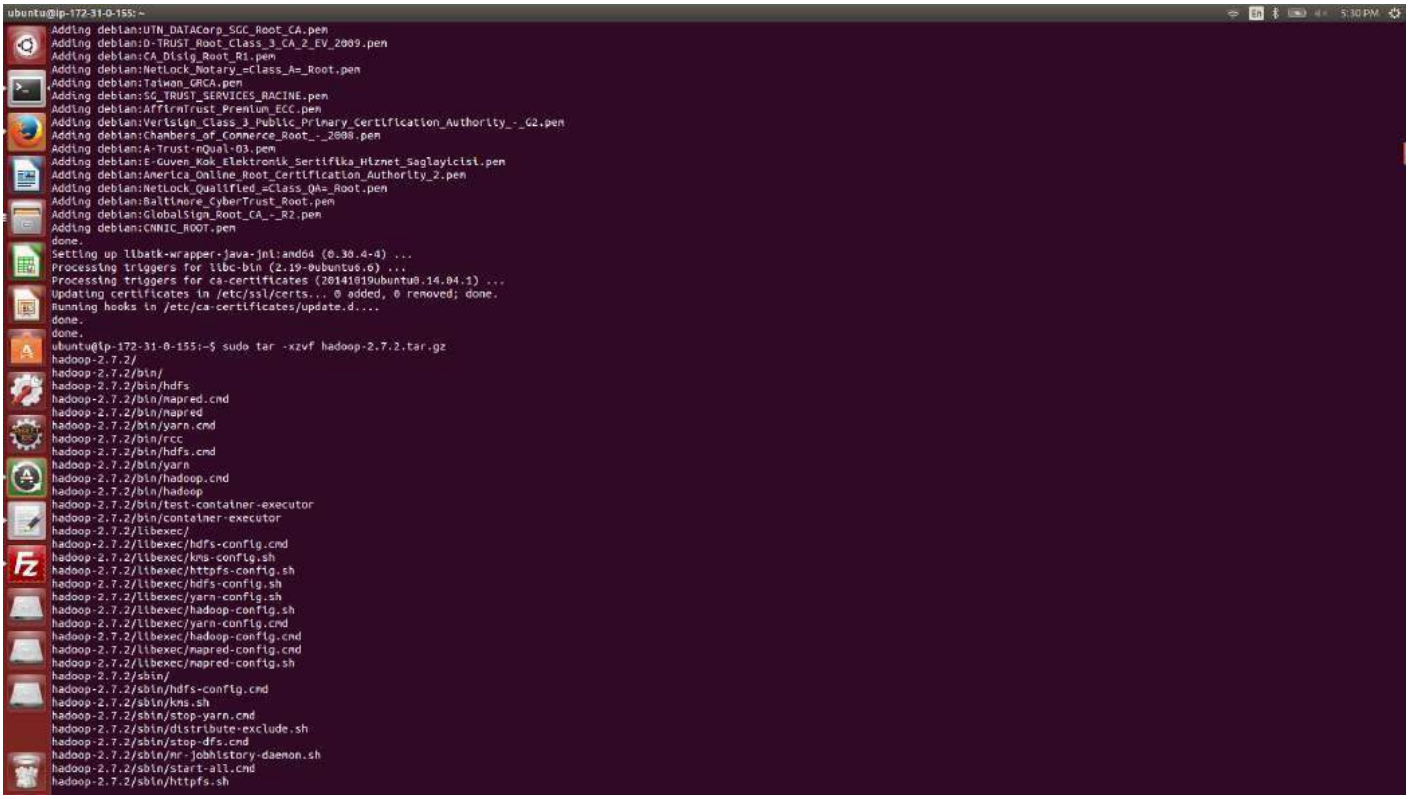


- Once installed, download the latest Hadoop 2.7.2 version on our amazon ec2 instance

  `wget` http://mirrors.sonic.net/apache/hadoop/common/hadoop-2.7.2/hadoop-2.7.2.tar.gz

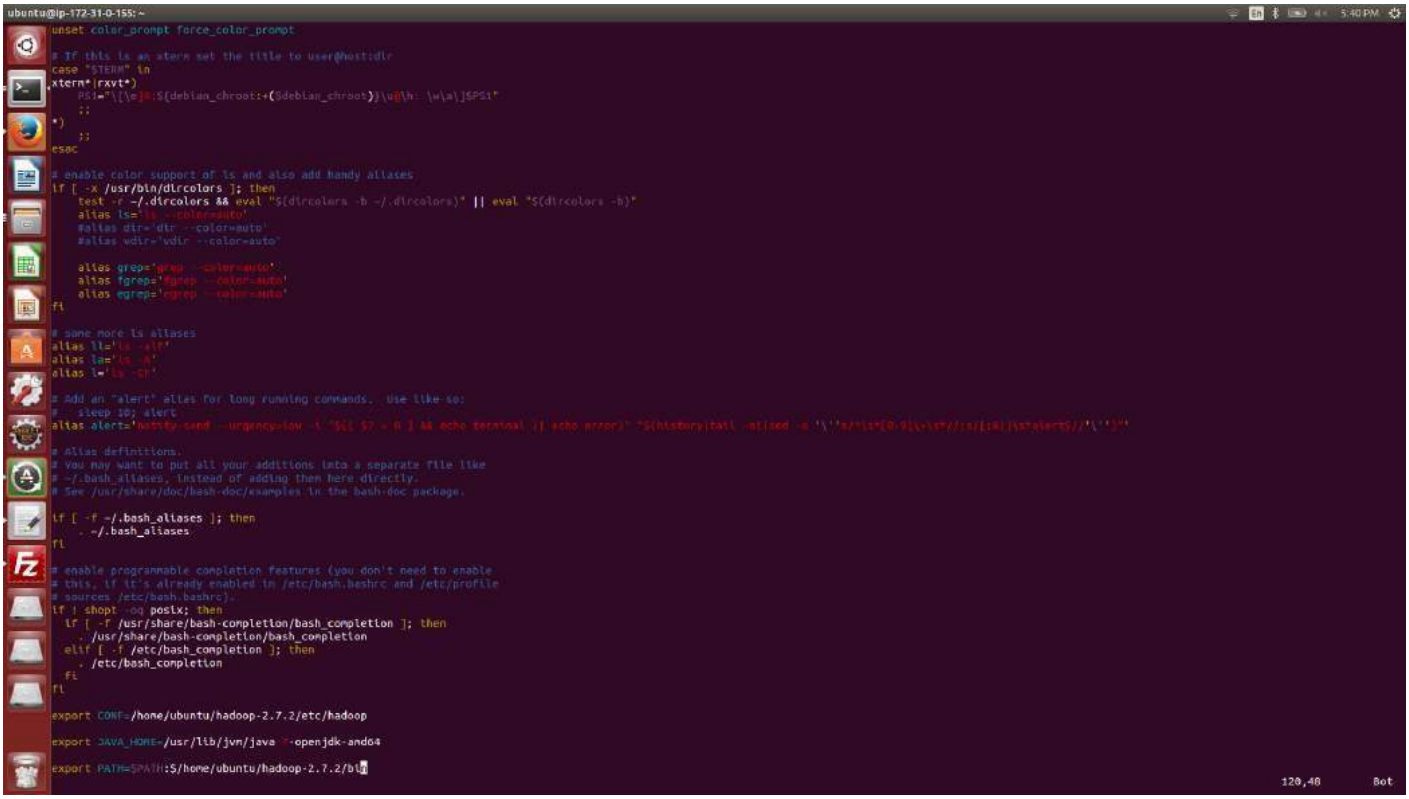
- Extract the downloaded Hadoop 2.7.2 Binary files to the system using command,

  tar –xzvf hadoop-2.7.2.tar.gz



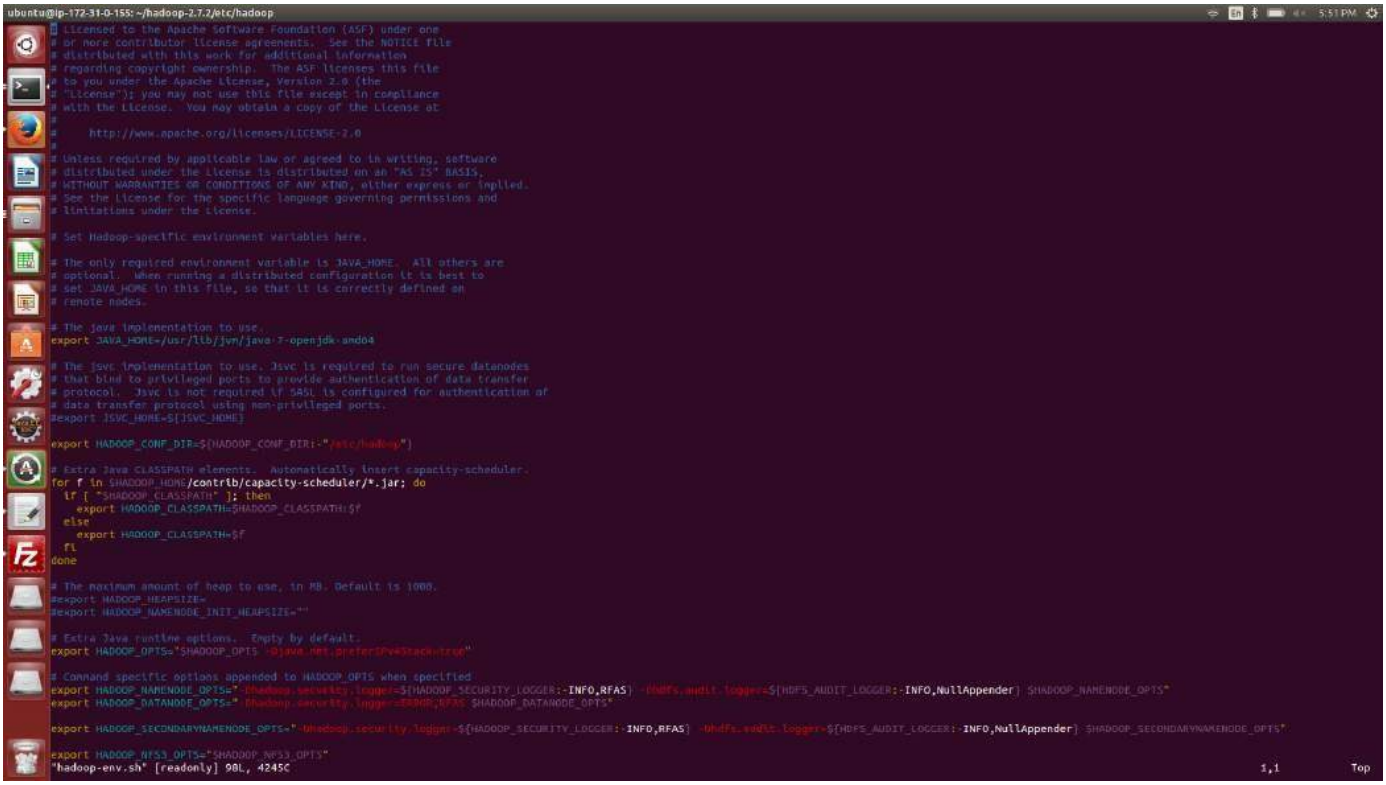
- Configure the .bashrc file to add Hadoop environment variables that is adding path of the Hadoop executable files.

export CONF=home/Ubuntu/Hadoop-2.7.2/etc/Hadoop
export JAVA_HOME = /usr/lib/jvm/java-7-openjdk-amd64
export PATH =$PATH:$/home/Ubuntu/Hadoop-2.7.2/bin



- Edit the Hadoop environment file and configure it to use the java installed in the previous step. The Hadoop environment script can be found on /Hadoop-2.7.2/etc/bin folder

Export JAVA_HOME = /usr/lib/jvm/java-7-openjdk-amd64

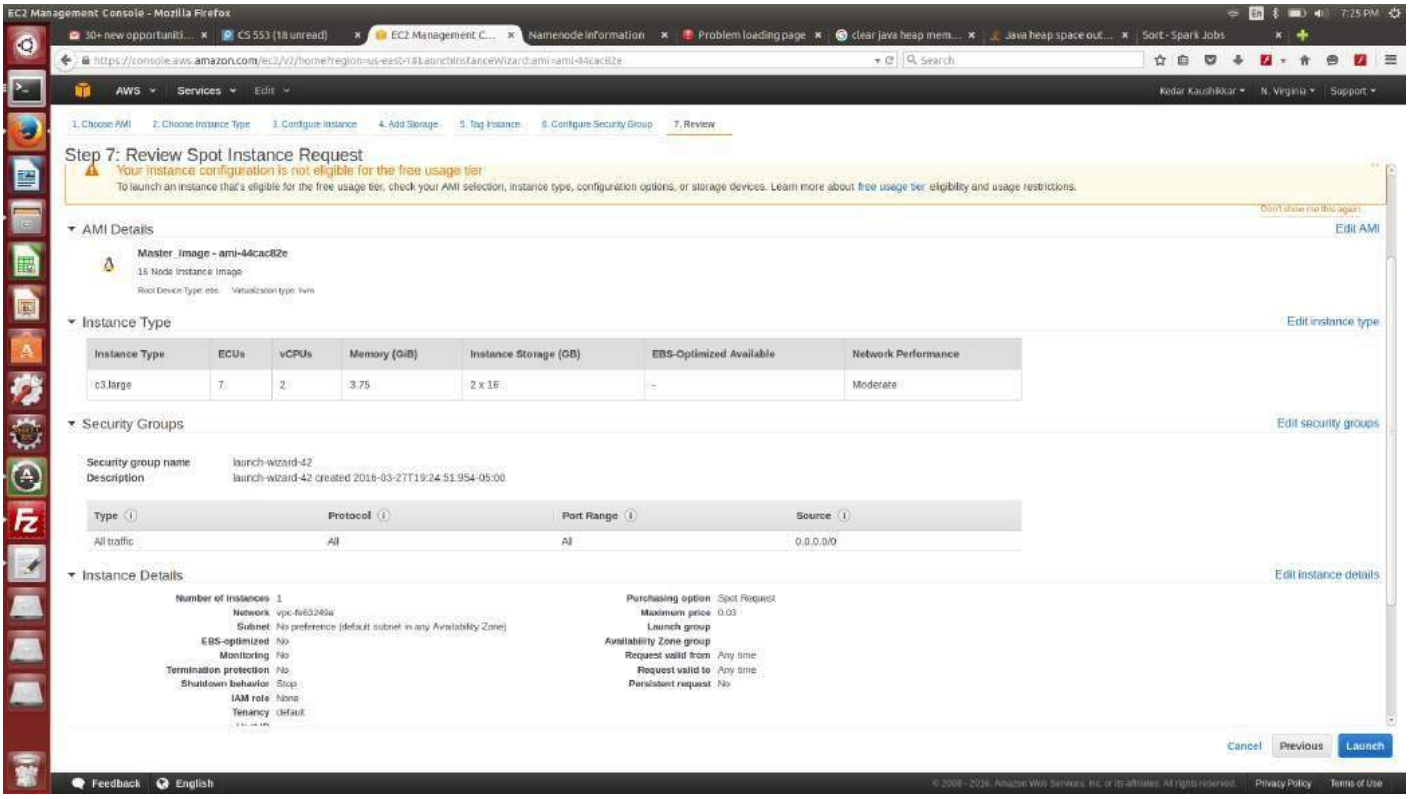


- This is the base installation that is needed in every virtual cluster that needs to run map-reduce jobs. More over depending upon the type of MapReduce jobs we change the Hadoop configuration files and start the Hadoop nodes on the instance.

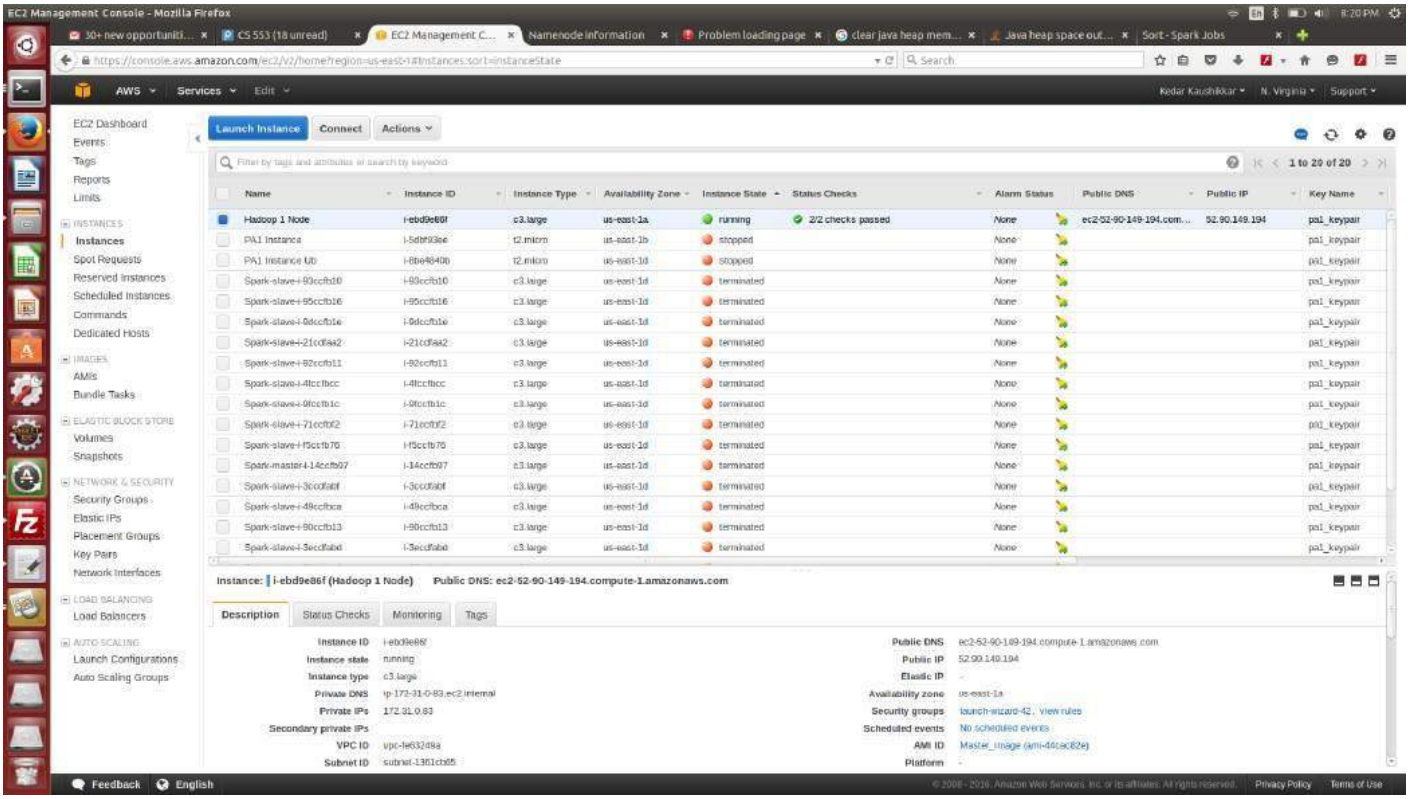
Hadoop 1 Node TeraSort (10GB)

- Following the above steps for installing Hadoop on the virtual cluster, we create another c3.large instance for our Hadoop 1 node mapreduce job for sorting a 10GB file.

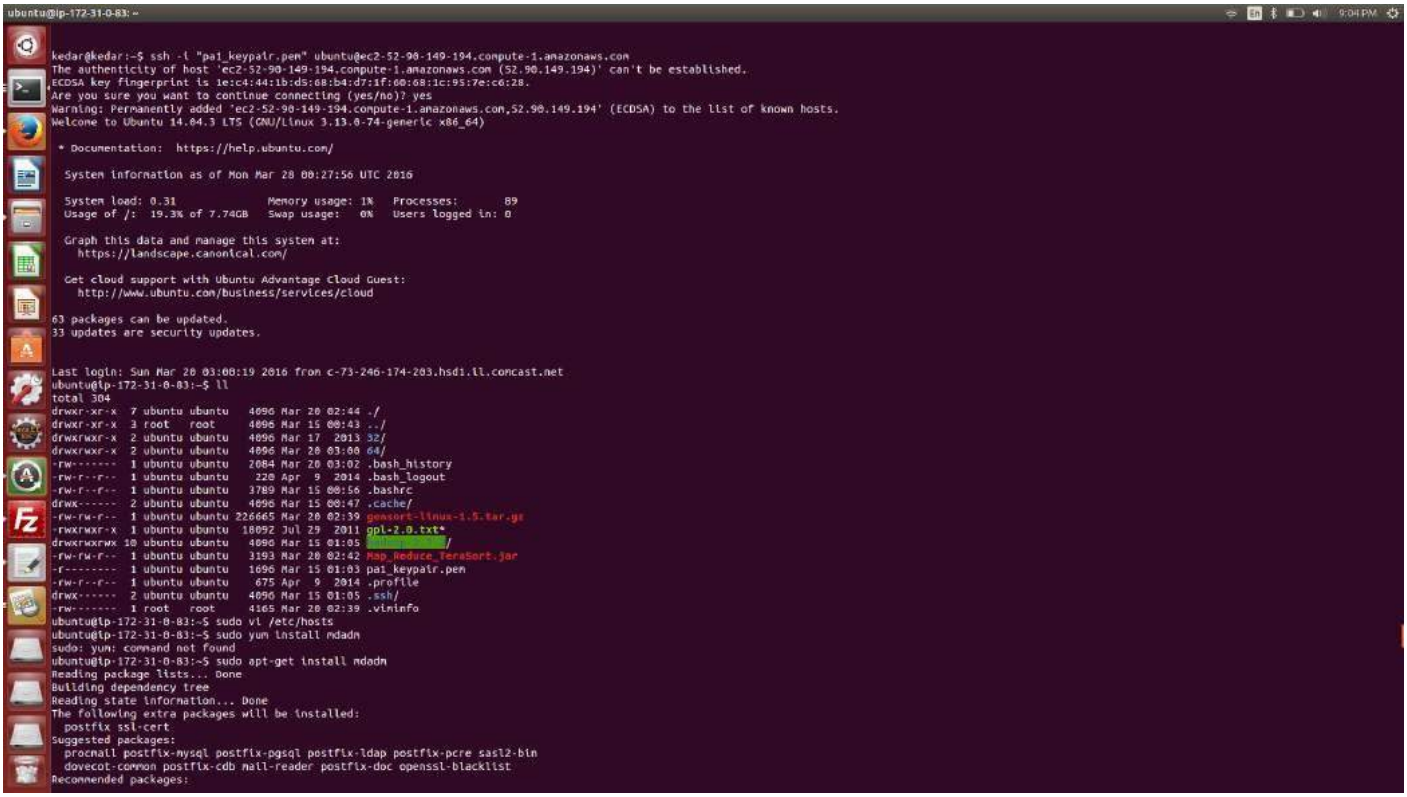
- Below given is the screenshot for review instance configuration for our c3.large instance.
  No of Instances = 1
  Spot Price = 0.030
  EBS Volume = 40GiB



- After launching the instance, the Hadoop 1 node virtual cluster will be started and Amazon AWS dashboard will show the active instance.



On the terminal, SSH the public DNS of the instance. Below screenshot shows up successful login.



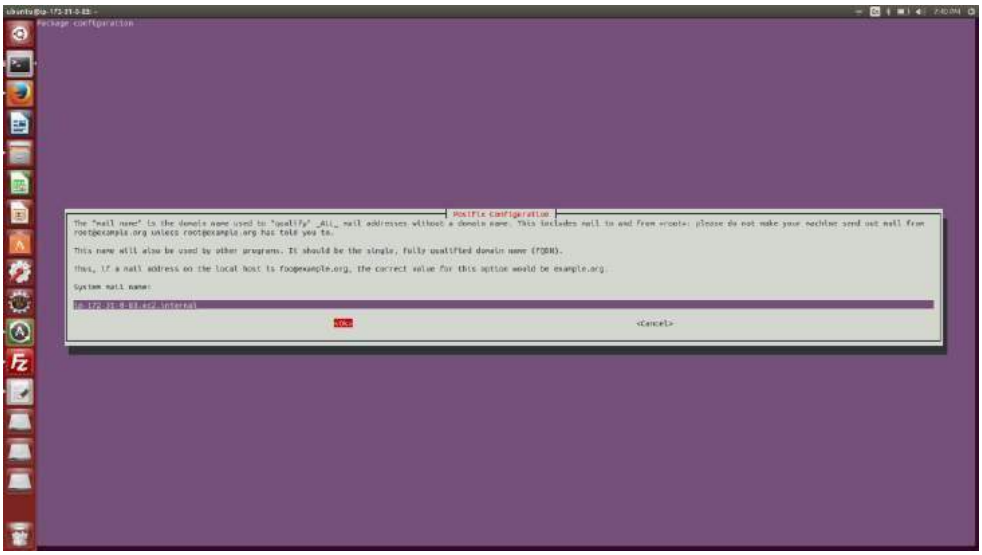
Once the amazon instance is started, follow the Hadoop installation procedure as explained in the earlier part. The Hadoop directory can be seen on the root directory.

- While creating an instance, we created an EBS volume of 40 GB which we will mount for our Hadoop to use it to store tmp files while running map-reduce jobs.

  Mentioned below are the steps to mount an EBS volume to our system.

  Sudo yum install mdadm

- We use the **mdadm** command to create a logical RAID device from the newly attached Amazon EBS volumes.



  - We can list the devices on our instance using **lsblk** command

  - To create a RAID 0 array, execute the following command
    sudo mdadm --create --verbose /dev/md0 --level=0 --name=RAID_NAME --raid-devices=No_Of_Devices Device1 Device2

  - Create a file system on your RAID array, and give that file system a label to use when you mount it later.
    sudo mkfs.ext4 -L RAID_NAME /dev/md0

  - Create a mount point for your RAID array.
    sudo mkdir -p /mnt/raid

  - Finally, mount the RAID device on the mount point that you created:
    sudo mount LABEL=MAPREDUCE_RAID /mnt/raid

The mentioned steps for mounting an EBS volume are shown in the below screenshot

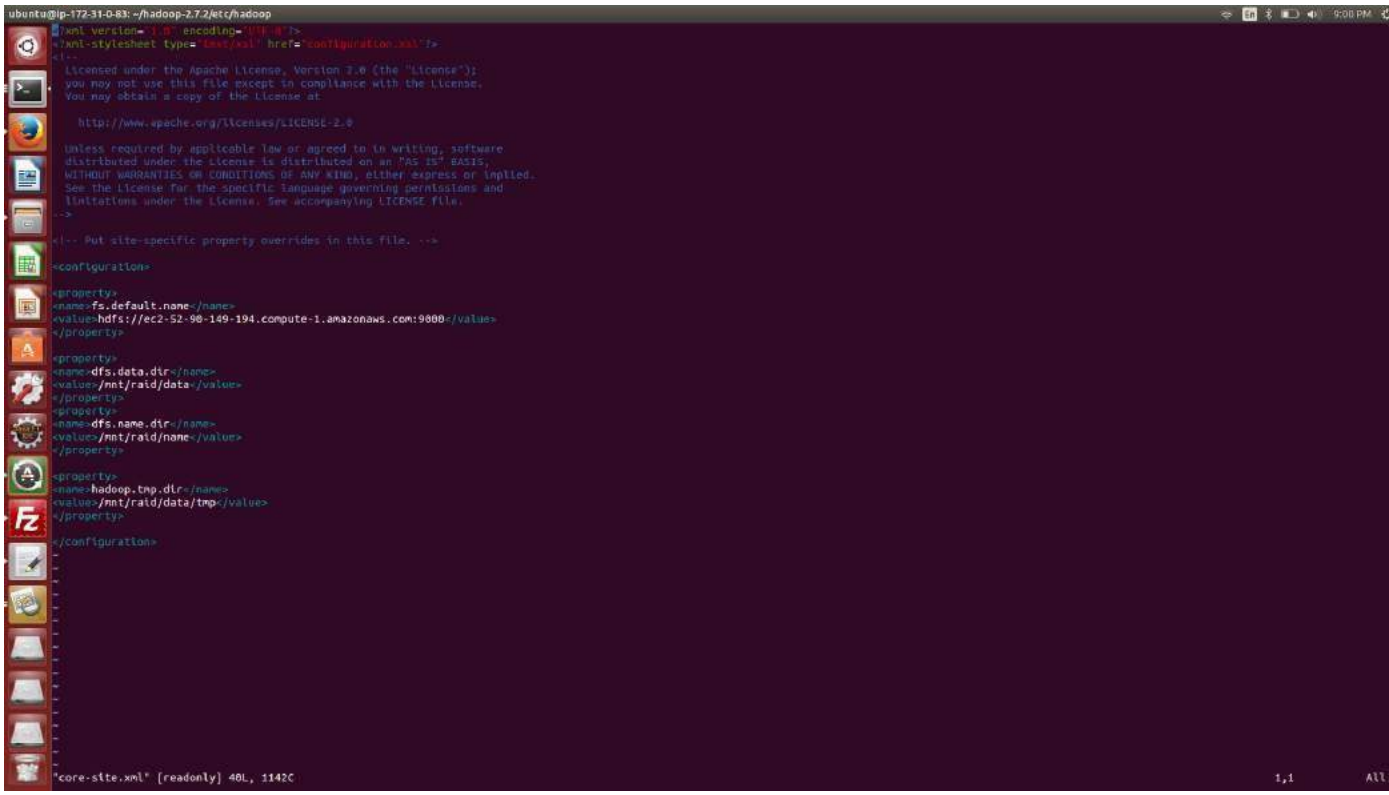
```
ubuntu@ip-172-31-0-83: ~
ubuntu@ip-172-31-0-83:~$ sudo mdadm --create --verbose /dev/md0 --level=0 --name=MAPREDUCE_RAID --raid-devices=3 /dev/xvdb /dev/xvdc /dev/xvdf
mdadm: super1.x cannot open /dev/xvdb: Device or resource busy
mdadm: ddf: Cannot use /dev/xvdb: Device or resource busy
mdadm: Cannot use /dev/xvdb: It is busy
mdadm: cannot open /dev/xvdb: Device or resource busy
ubuntu@ip-172-31-0-83:~$ sudo su
root@ip-172-31-0-83:/home/ubuntu# umount -l /dev/xvdb
root@ip-172-31-0-83:/home/ubuntu# exit
exit
ubuntu@ip-172-31-0-83:~$ sudo mdadm --create --verbose /dev/md0 --level=0 --name=MAPREDUCE_RAID --raid-devices=3 /dev/xvdb /dev/xvdc /dev/xvdf
mdadm: chunk size defaults to 512K
mdadm: /dev/xvdb appears to contain an ext2fs file system
    size=15996416K  mtime=Mon Mar 28 00:27:57 2016
mdadm: /dev/xvdc appears to contain an ext2fs file system
    size=15996416K  mtime=Thu Jan  1 00:00:00 1970
mdadm: /dev/xvdf appears to be part of a raid array:
    level=raid0 devices=0 ctime=Thu Jan  1 00:00:00 1970
mdadm: partition table exists on /dev/xvdf but will be lost or
       meaningless after creating array
Continue creating array? Y
mdadm: Defaulting to version 1.2 metadata
mdadm: array /dev/md0 started.
ubuntu@ip-172-31-0-83:~$ lsblk
NAME    MAJ:MIN RM   SIZE RO TYPE  MOUNTPOINT
xvda    202:0    0    20G  0 disk
└─xvda1 202:1    0    20G  0 part  /
xvdb    202:16   0  15.3G  0 disk
└─md0     9:0    0  60.5G  0 raid0
xvdc    202:32   0  15.3G  0 disk
└─md0     9:0    0  60.5G  0 raid0
xvdf    202:80   0    30G  0 disk
└─md0     9:0    0  60.5G  0 raid0
ubuntu@ip-172-31-0-83:~$ sudo su
root@ip-172-31-0-83:/home/ubuntu# umount -l /dev/md0
umount: /dev/md0: not mounted
root@ip-172-31-0-83:/home/ubuntu# exit
exit
ubuntu@ip-172-31-0-83:~$ mdadm --stop /dev/md0
mdadm: must be super-user to perform this action
ubuntu@ip-172-31-0-83:~$ sudo mdadm --stop /dev/md0
mdadm: stopped /dev/md0
ubuntu@ip-172-31-0-83:~$ lsblk
NAME    MAJ:MIN RM   SIZE RO TYPE MOUNTPOINT
xvda    202:0    0    20G  0 disk
└─xvda1 202:1    0    20G  0 part /
xvdb    202:16   0  15.3G  0 disk
xvdc    202:32   0  15.3G  0 disk
xvdf    202:80   0    30G  0 disk
ubuntu@ip-172-31-0-83:~$ sudo mkfs -t ext4 /dev/xvdf
mke2fs 1.42.9 (4-Feb-2014)
Filesystem label=
OS type: Linux
Block size=4096 (log=2)
Fragment size=4096 (log=2)
Stride=0 blocks, Stripe width=0 blocks
1966080 inodes, 7864320 blocks
393216 blocks (5.00%) reserved for the super user
First data block=0
```

```
ubuntu@ip-172-31-0-83: ~
32768 blocks per group, 32768 fragments per group
8192 inodes per group
Superblock backups stored on blocks:
        32768, 98304, 163840, 229376, 294912, 819200, 884736, 1605632, 2654208,
        4096000

Allocating group tables: done
Writing inode tables: done
Creating journal (32768 blocks): done
Writing superblocks and filesystem accounting information: done

ubuntu@ip-172-31-0-83:~$ sudo file -s /dev/xvdf
/dev/xvdf: Linux rev 1.0 ext4 filesystem data, UUID=746cc200-14a3-4643-9cc0-b1c3126773ff (extents) (large files) (huge files)
ubuntu@ip-172-31-0-83:~$ sudo mdadm --create --verbose /dev/md0 --level=0 --name=MAPREDUCE_RAID --raid-devices=3 /dev/xvdb /dev/xvdc /dev/xvdf
mdadm: chunk size defaults to 512K
mdadm: /dev/xvdb appears to contain an ext2fs file system
    size=15996416K  mtime=Mon Mar 28 00:27:57 2016
mdadm: /dev/xvdb appears to be part of a raid array:
    level=raid0 devices=3 ctime=Mon Mar 28 00:42:43 2016
mdadm: /dev/xvdc appears to contain an ext2fs file system
    size=15996416K  mtime=Thu Jan  1 00:00:00 1970
mdadm: /dev/xvdc appears to be part of a raid array:
    level=raid0 devices=3 ctime=Mon Mar 28 00:42:43 2016
mdadm: /dev/xvdf appears to contain an ext2fs file system
    size=31457280K  mtime=Thu Jan  1 00:00:00 1970
Continue creating array? Y
mdadm: Defaulting to version 1.2 metadata
mdadm: array /dev/md0 started.
ubuntu@ip-172-31-0-83:~$ lsblk
NAME    MAJ:MIN RM   SIZE RO TYPE  MOUNTPOINT
xvda    202:0    0    20G  0 disk
└─xvda1 202:1    0    20G  0 part  /
xvdb    202:16   0  15.3G  0 disk
└─md0     9:0    0  60.5G  0 raid0
xvdc    202:32   0  15.3G  0 disk
└─md0     9:0    0  60.5G  0 raid0
xvdf    202:80   0    30G  0 disk
└─md0     9:0    0  60.5G  0 raid0
ubuntu@ip-172-31-0-83:~$ sudo mkfs.ext4 -L MAPREDUCE_RAID /dev/md0
mke2fs 1.42.9 (4-Feb-2014)
Filesystem label=MAPREDUCE_RAID
OS type: Linux
Block size=4096 (log=2)
Fragment size=4096 (log=2)
Stride=128 blocks, Stripe width=384 blocks
3973120 inodes, 15862144 blocks
793107 blocks (5.00%) reserved for the super user
First data block=0
Maximum filesystem blocks=4294967296
485 block groups
32768 blocks per group, 32768 fragments per group
8192 inodes per group
Superblock backups stored on blocks:
        32768, 98304, 163840, 229376, 294912, 819200, 884736, 1605632, 2654208,
        4096000, 7962624, 11239424

Allocating group tables: done
Writing inode tables: done
```

Configuring Hadoop files for single node cluster.

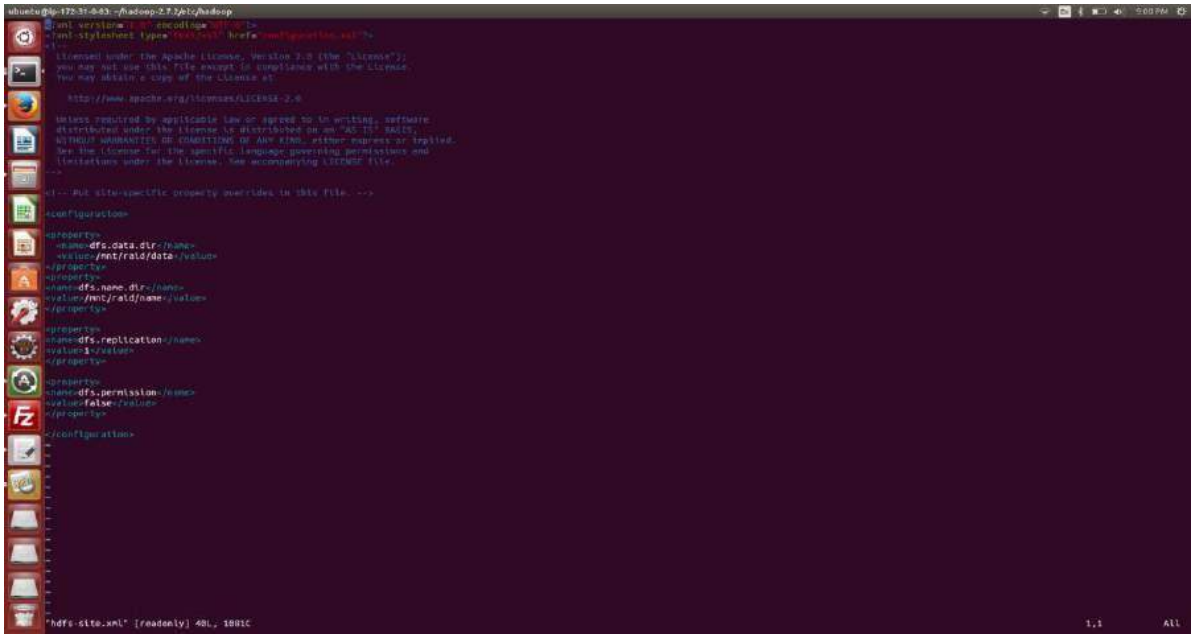
- Modify configuration files which are in /Hadoop-2.7.2/etc/Hadoop folder,

1. core-site.xml



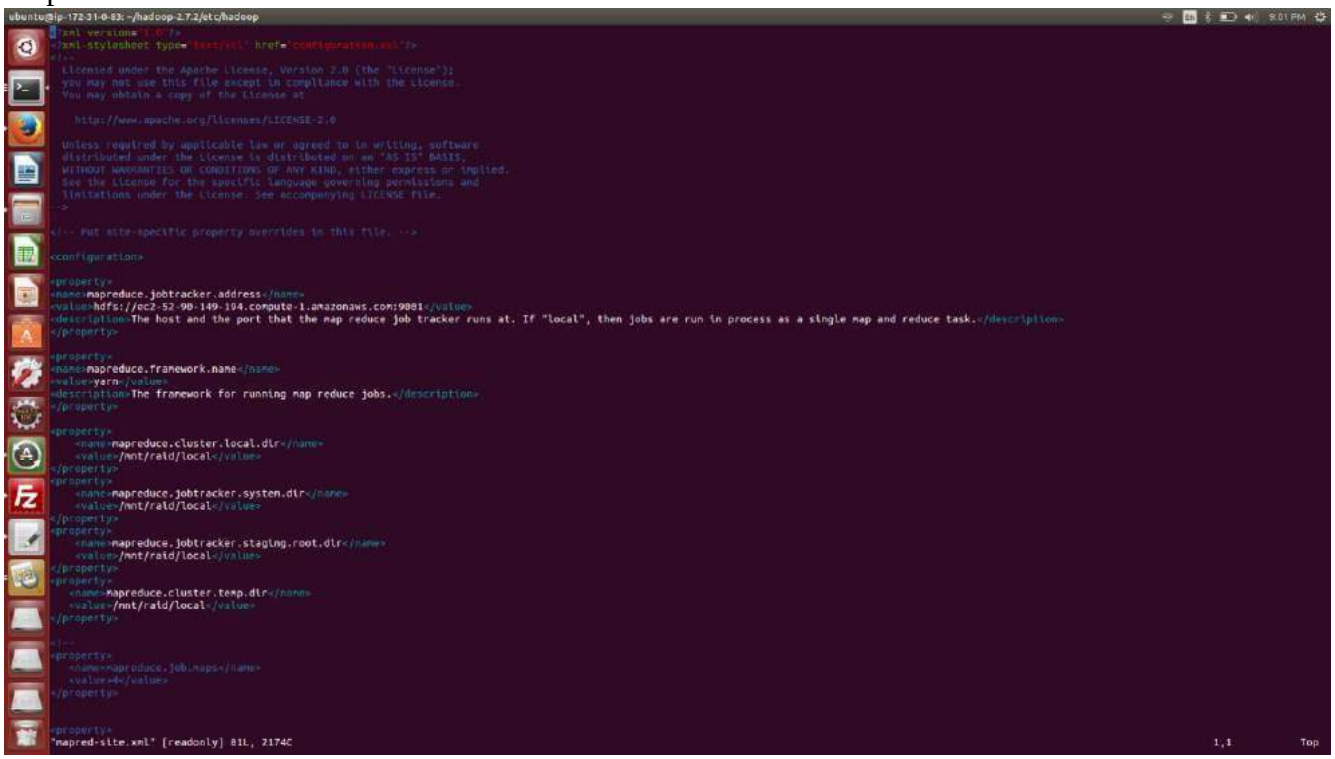
- The default name property will have the public DNS of our amazon instance.
- The Hadoop data directory, namenode directory and temporary directory will be now pointing to the new mounted directory that we created in the previous steps.

2. hdfs-site.xml



- The hdfs-site.xml will contain the data replication settings which we set to 1 in-order to avoid multiple replication of the same data. This will save the space in the memory and this configuration is used to have multiple back up of data in-case of any failure.

3. Mapred-site.xml

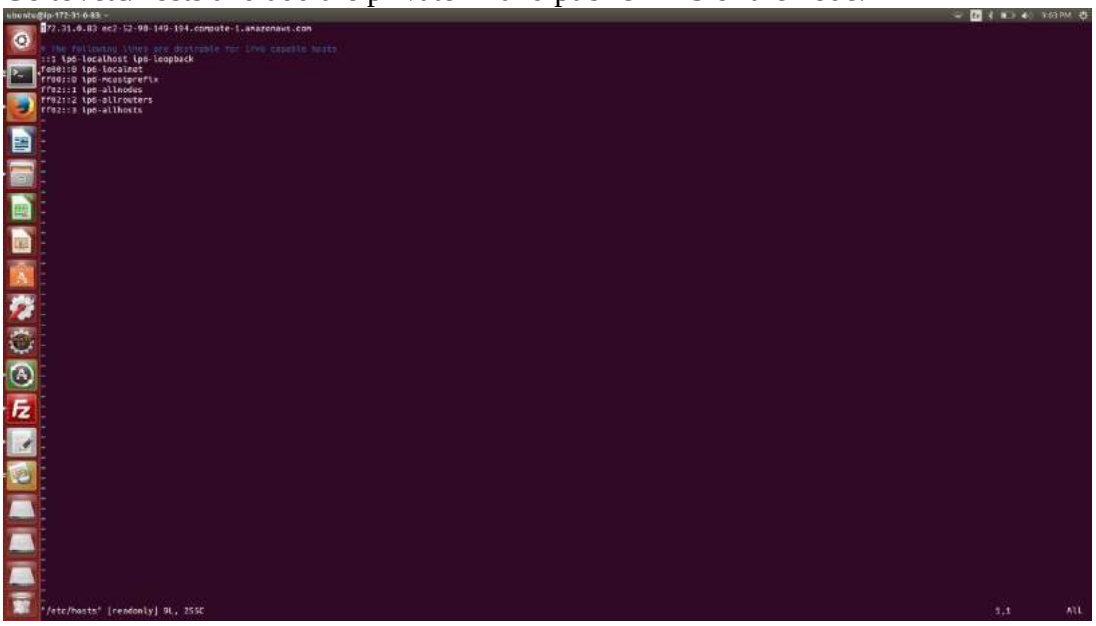


- The Hadoop mapred-site.xml contains configuration information that overrides the default values for MapReduce parameters.
- The framework name will be 'yarn' while the jobtracker address is the publicDNS of our amazon instance.
- The mapreduce cluster settings point to the new mounted raid device.
- Set up publicDNS address in the jobtracker.address values.

4. Slaves

The slaves file will have the public DNS of the amazon ec2 single node cluster. This being single node, the master acts as its own slave.

5. Finally, we need to change the host file of our instance to point it to its private IP and public DNS instead of localhost.
   Go to /etc/hosts and add the private IP and public DNS of the node.

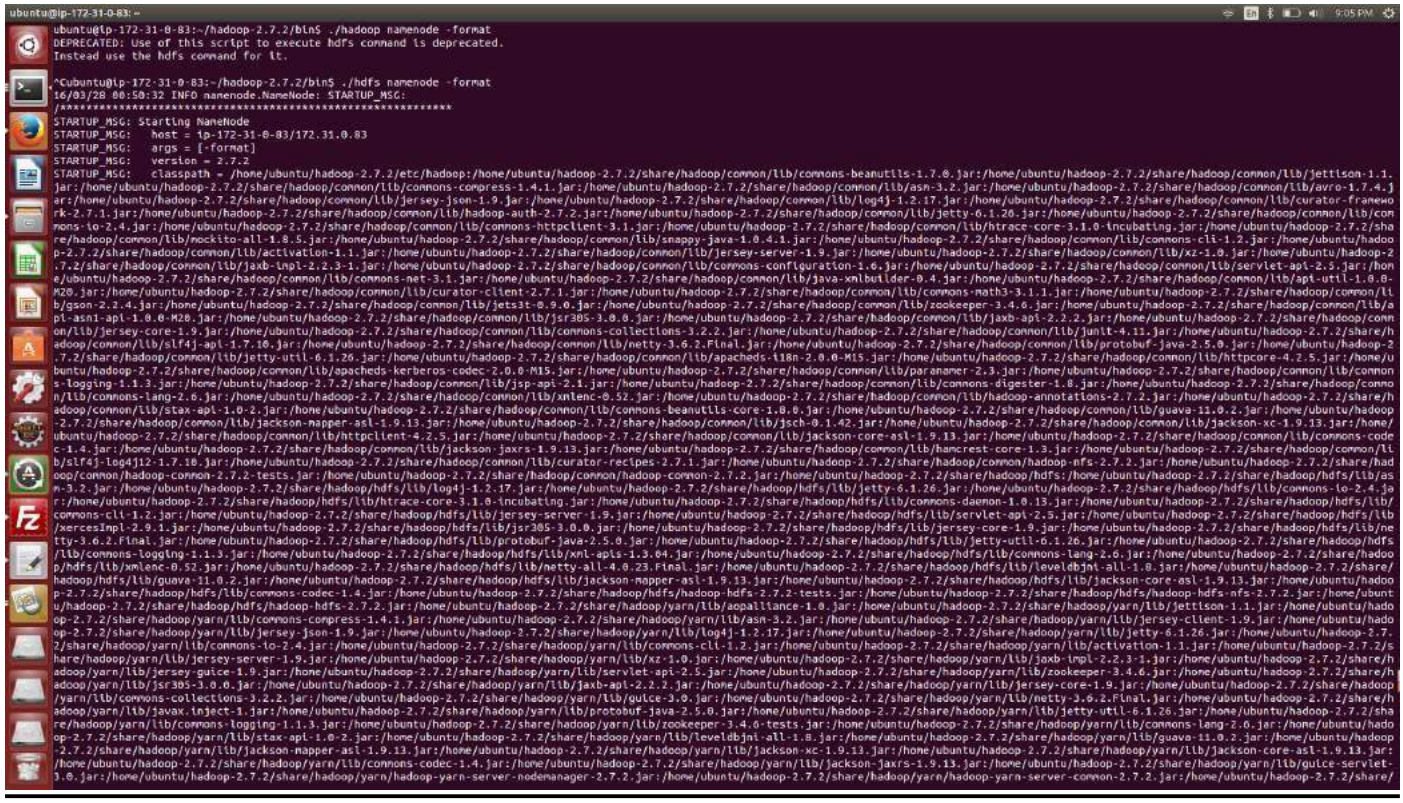


   Once we configure the Hadoop envoirnment with the above settings, we need to format the namenode that will create the data , name and temp directory as mentioned in the core-site.xml in the raid array .

6. Format the Namenode
   The namenode can be formatted using the below command on /Hadoop-2.7.2/bin directory

   ./hdfs namenode –format

   This command is used to format the file system at the location specified in the hdfs-site.xml file



- Once the name node is formatted, we can start Hadoop and all the configuration modified will be used for this Hadoop cluster.

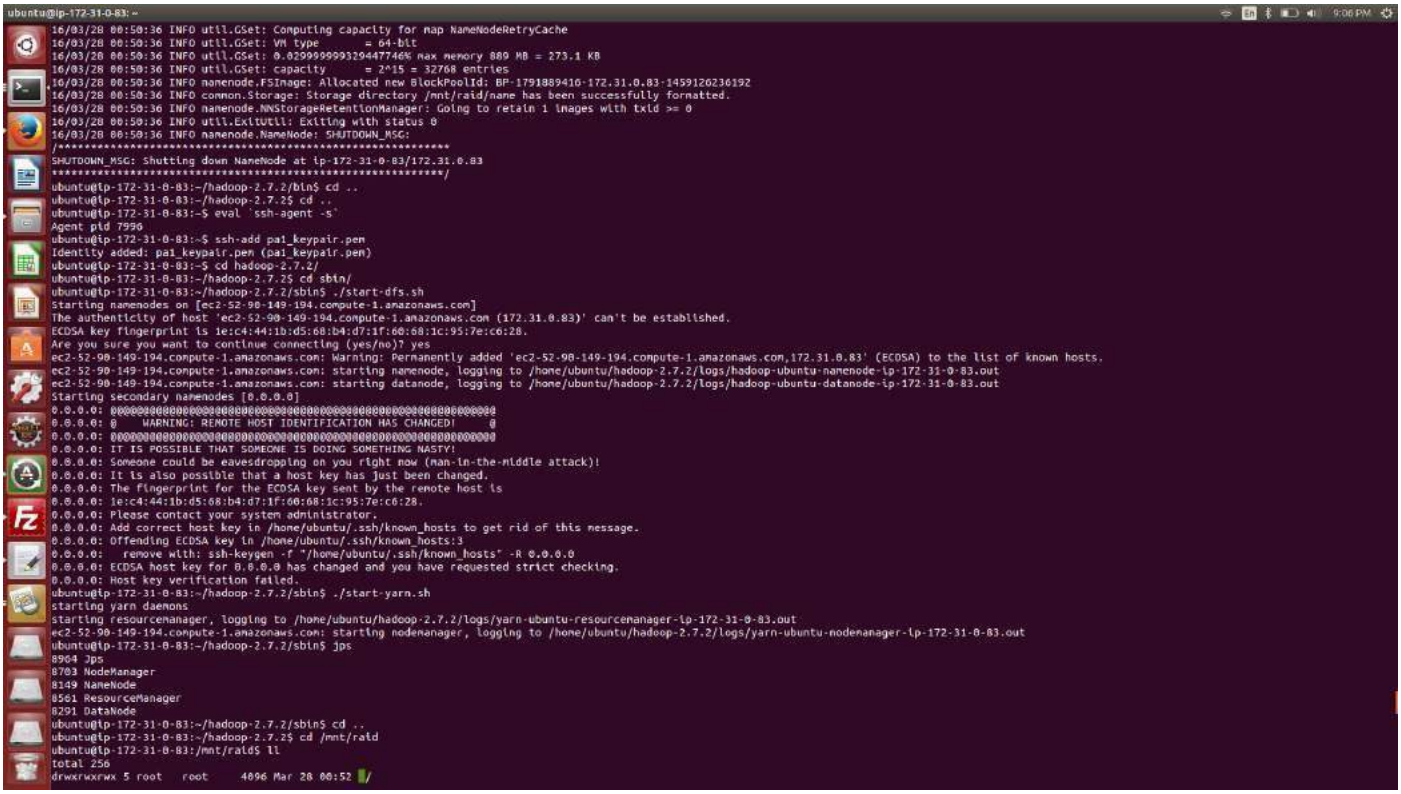
  To access permission write
  eval `ssh-agent –s`
  ssh-add keypair.pem

  If its restricted, then
  chmod 400 keypair.pem and then above type instructions

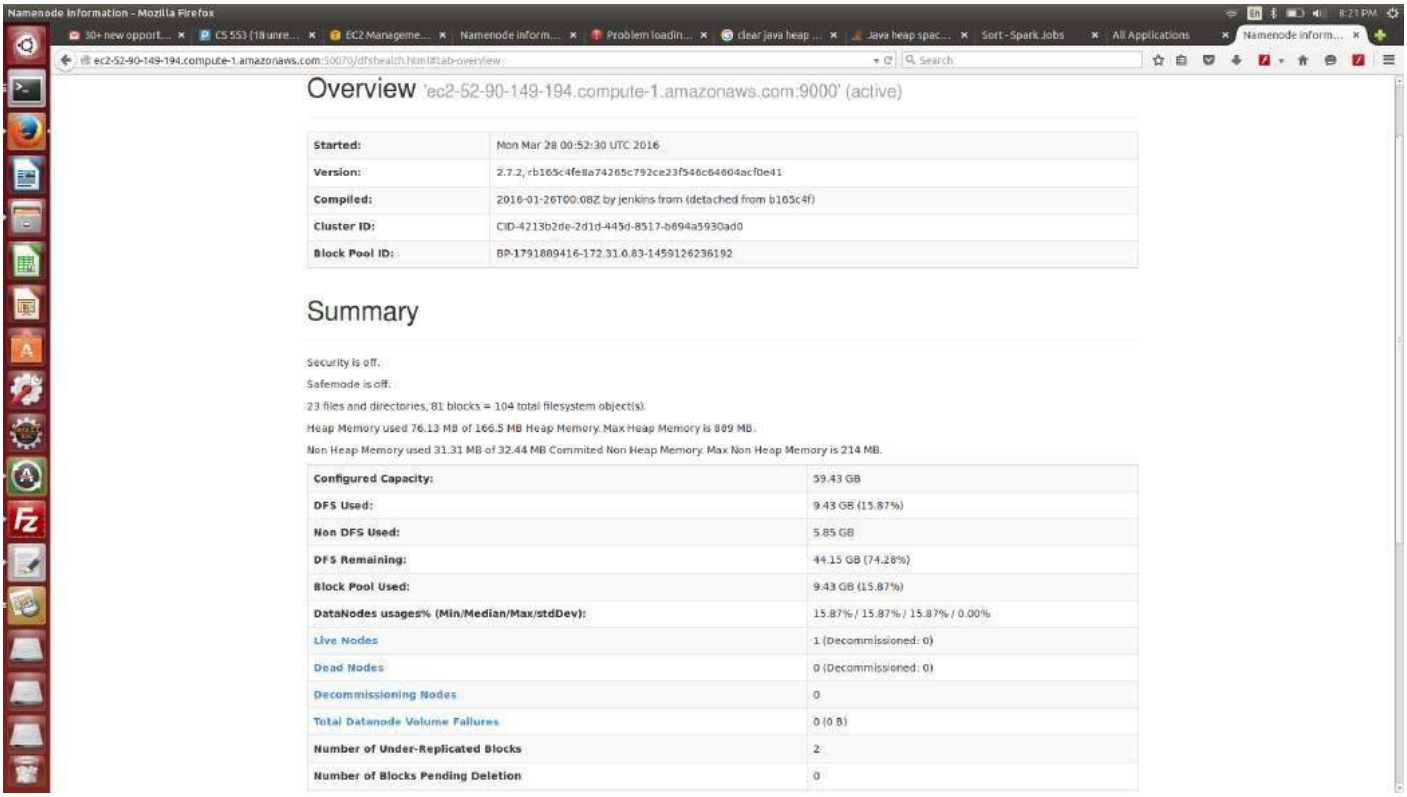
- Start the Hadoop cluster using

1. **./start-dfs.sh**
   This command will start all the Hadoop namenodes, datanodes, jobtrackers and tasktrackers
2. **./start-yarn.sh**
   This command will start the resourcemanagers and the node managers.
3. You can type in **jps** command to check whether the above nodes have been started.

All the above start up scripts have been executed and shown in the below screenshots



Once the job tracker has been started, we can verify whether our Hadoop cluster is setup up correctly by going to the public DNS address followed by port 50070



Here the Live Nodes define the number of Virtual Clusters actively running for the Hadoop setup.

- We will be running Hadoop map-reduce sort program which will take 10GB of input data taken by Map and convert it into another set of data where individual elements will be broken into tuples (key/value) pairs. The Reduce task will take the output from map as an input and combines those data tuples into smaller set of tuples.

  The program for the above example is present at the root directory.
  The input data of 10GB is generated through Gensort and will be stored in the Mounted disk.

  We then create a input directory at the HDFS and the input file will be moved from the mount directory to the HDFS and map-reduce sort job will be run.

  To create an input directory in HDFS, go to /Hadoop-2.7.2/bin folder and run the following commands,

  ./hdfs dfs -mkdir /user
  ./hdfs dfs -mkdir /user/hduser
  ./hdfs dfs -mkdir /user/hduser/input

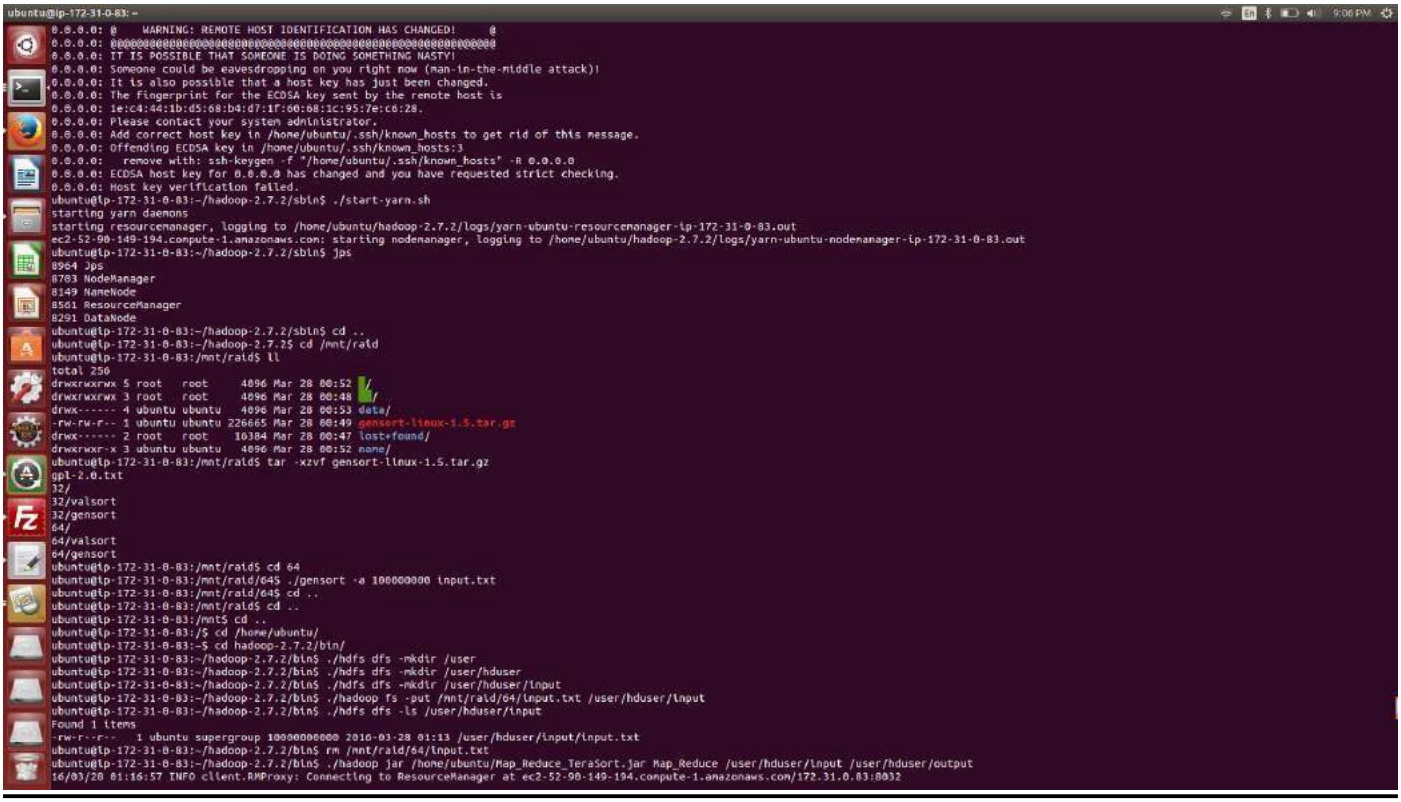
- Generate a 10GB input file on the mount disk using the below command,

  ./gensort –a 100000000 input.txt

- Transfer the input file into the HDFS
  ./hadoop fs -put /home/ubuntu/64/input.txt /user/hduser/input

  Verify if the file is successfully transferred to the HDFS,

  ./hdfs dfs -ls /user/hduser/input



```
ubuntu@ip-172-31-0-83: ~
0.0.0.0: @    WARNING: REMOTE HOST IDENTIFICATION HAS CHANGED!     @
0.0.0.0: @@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
0.0.0.0: IT IS POSSIBLE THAT SOMEONE IS DOING SOMETHING NASTY!
0.0.0.0: Someone could be eavesdropping on you right now (man-in-the-middle attack)!
0.0.0.0: It is also possible that a host key has just been changed.
0.0.0.0: The fingerprint for the ECDSA key sent by the remote host is
0.0.0.0: 1e:c4:44:1b:d5:68:b4:d7:1f:60:68:1c:95:7e:c6:28.
0.0.0.0: Please contact your system administrator.
0.0.0.0: Add correct host key in /home/ubuntu/.ssh/known_hosts to get rid of this message.
0.0.0.0: Offending ECDSA key in /home/ubuntu/.ssh/known_hosts:3
0.0.0.0:   remove with: ssh-keygen -f "/home/ubuntu/.ssh/known_hosts" -R 0.0.0.0
0.0.0.0: ECDSA host key for 0.0.0.0 has changed and you have requested strict checking.
0.0.0.0: Host key verification failed.
ubuntu@ip-172-31-0-83:~/hadoop-2.7.2/sbin$ ./start-yarn.sh
starting yarn daemons
starting resourcemanager, logging to /home/ubuntu/hadoop-2.7.2/logs/yarn-ubuntu-resourcemanager-ip-172-31-0-83.out
ec2-52-90-149-194.compute-1.amazonaws.com: starting nodemanager, logging to /home/ubuntu/hadoop-2.7.2/logs/yarn-ubuntu-nodemanager-ip-172-31-0-83.out
ubuntu@ip-172-31-0-83:~/hadoop-2.7.2/sbin$ jps
8964 Jps
8783 NodeManager
8149 NameNode
8561 ResourceManager
8291 DataNode
ubuntu@ip-172-31-0-83:~/hadoop-2.7.2/sbin$ cd ..
ubuntu@ip-172-31-0-83:~/hadoop-2.7.2$ cd /mnt/raid
ubuntu@ip-172-31-0-83:/mnt/raid$ ll
total 250
drwxrwxrwx 5 root   root     4096 Mar 28 00:52 ./
drwxrwxrwx 3 root   root     4096 Mar 28 00:48 ../
drwx------ 4 ubuntu ubuntu   4096 Mar 28 00:53 data/
-rw-rw-r-- 1 ubuntu ubuntu 226665 Mar 28 00:49 gensort-linux-1.5.tar.gz
drwx------ 2 root   root    16384 Mar 28 00:47 lost+found/
drwxrwxr-x 3 ubuntu ubuntu   4096 Mar 28 00:52 none/
ubuntu@ip-172-31-0-83:/mnt/raid$ tar -xzvf gensort-linux-1.5.tar.gz
gpl-2.0.txt
32/
32/valsort
32/gensort
64/
64/valsort
64/gensort
ubuntu@ip-172-31-0-83:/mnt/raid$ cd 64
ubuntu@ip-172-31-0-83:/mnt/raid/64$ ./gensort -a 100000000 input.txt
ubuntu@ip-172-31-0-83:/mnt/raid/64$ cd ..
ubuntu@ip-172-31-0-83:/mnt/raid$ cd ..
ubuntu@ip-172-31-0-83:/mnt$ cd ..
ubuntu@ip-172-31-0-83:/$ cd /home/ubuntu/
ubuntu@ip-172-31-0-83:~$ cd hadoop-2.7.2/bin/
ubuntu@ip-172-31-0-83:~/hadoop-2.7.2/bin$ ./hdfs dfs -mkdir /user
ubuntu@ip-172-31-0-83:~/hadoop-2.7.2/bin$ ./hdfs dfs -mkdir /user/hduser
ubuntu@ip-172-31-0-83:~/hadoop-2.7.2/bin$ ./hdfs dfs -mkdir /user/hduser/input
ubuntu@ip-172-31-0-83:~/hadoop-2.7.2/bin$ ./hadoop fs -put /mnt/raid/64/input.txt /user/hduser/input
ubuntu@ip-172-31-0-83:~/hadoop-2.7.2/bin$ ./hdfs dfs -ls /user/hduser/input
Found 1 items
-rw-r--r--   1 ubuntu supergroup 10000000000 2016-03-28 01:13 /user/hduser/input/input.txt
ubuntu@ip-172-31-0-83:~/hadoop-2.7.2/bin$ rm /mnt/raid/64/input.txt
ubuntu@ip-172-31-0-83:~/hadoop-2.7.2/bin$ ./hadoop jar /home/ubuntu/Map_Reduce_TeraSort.jar Map_Reduce /user/hduser/input /user/hduser/output
16/03/28 01:16:57 INFO client.RMProxy: Connecting to ResourceManager at ec2-52-90-149-194.compute-1.amazonaws.com/172.31.0.83:8032
```

- Once the input file is successfully transferred to the HDFS we can start the map-reduce sort example using the below command,

./hadoop jar /home/ubuntu/Map_Reduce_TeraSort.jar Map_Reduce /user/hduser/input /user/hduser/output

Map_Reduce_TeraSort is the name of the jar file that contains the map-reduce algorithm implemented.

Map_Reduce is the name of the class

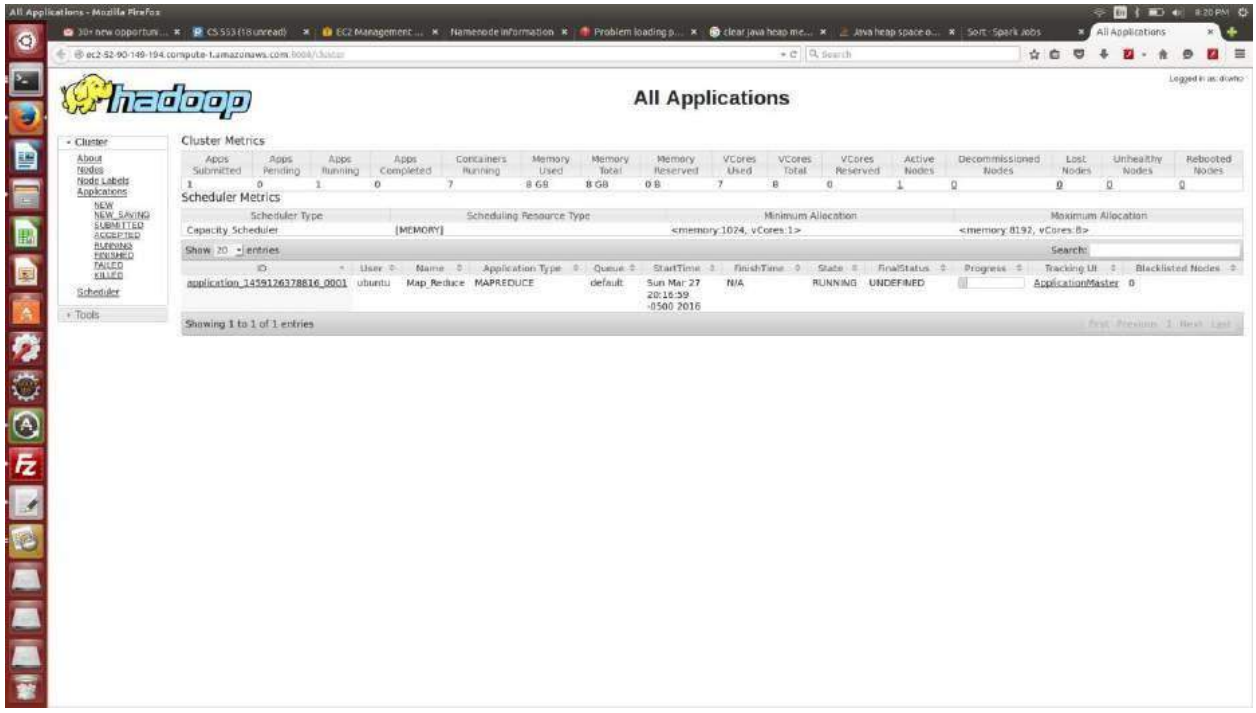
/user/hduser/input – Input directory of the text file

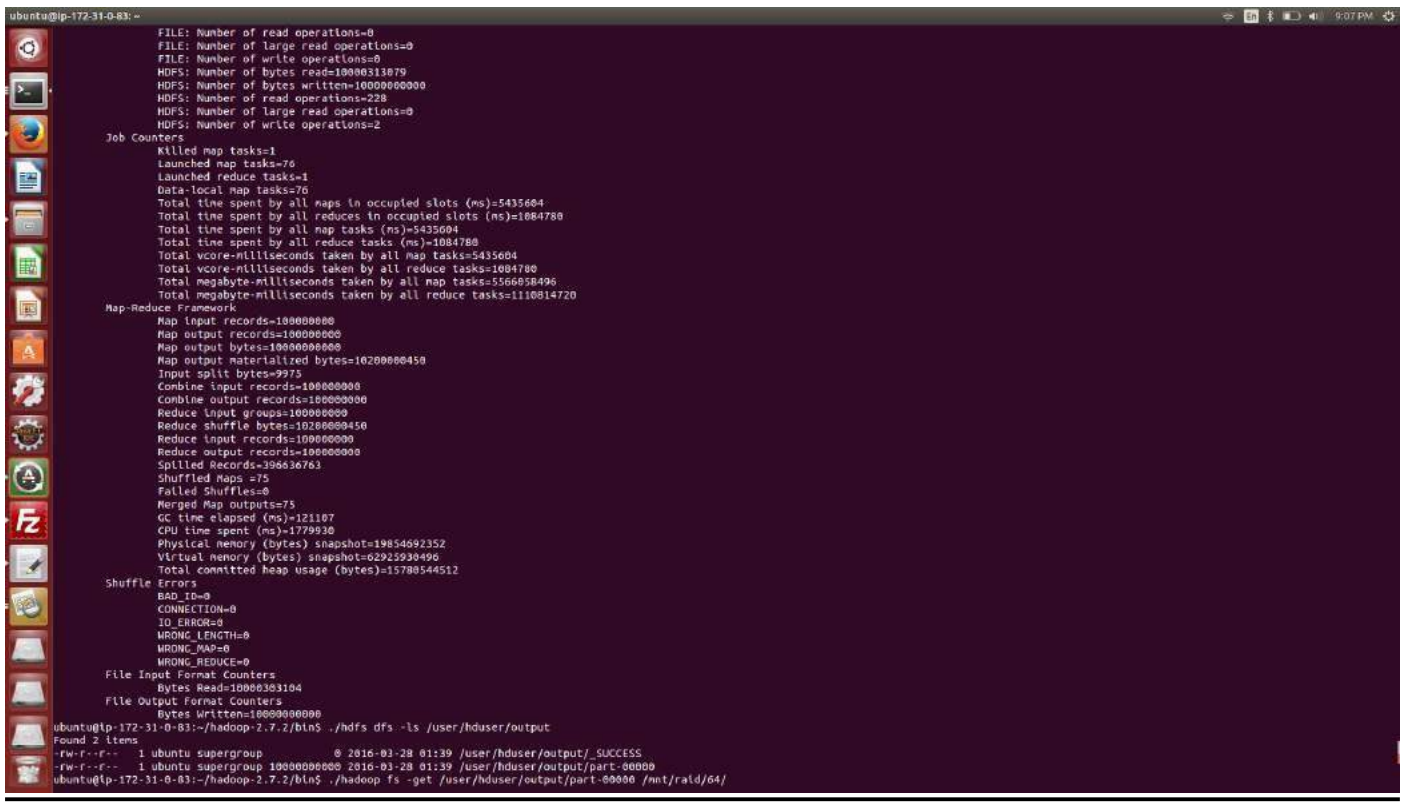
/user/hduser/output – Output Directory for the sorted file.

```
ubuntu@ip-172-31-0-83: ~
-rw-r--r--   1 ubuntu supergroup 10000000000 2016-03-28 01:13 /user/hduser/input/input.txt
ubuntu@ip-172-31-0-83:~/hadoop-2.7.2/bin$ rm /mnt/raid/64/input.txt
ubuntu@ip-172-31-0-83:~/hadoop-2.7.2/bin$ ./hadoop jar /home/ubuntu/Map_Reduce_TeraSort.jar Map_Reduce /user/hduser/input /user/hduser/output
16/03/28 01:16:57 INFO client.RMProxy: Connecting to ResourceManager at ec2-52-90-149-194.compute-1.amazonaws.com/172.31.0.83:8032
16/03/28 01:16:57 INFO client.RMProxy: Connecting to ResourceManager at ec2-52-90-149-194.compute-1.amazonaws.com/172.31.0.83:8032
16/03/28 01:16:57 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 01:16:58 INFO mapred.FileInputFormat: Total input paths to process : 1
16/03/28 01:16:58 INFO mapreduce.JobSubmitter: number of splits:75
16/03/28 01:16:59 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1459126378816_0001
16/03/28 01:16:59 INFO impl.YarnClientImpl: Submitted application application_1459126378810_0001
16/03/28 01:16:59 INFO mapreduce.Job: The url to track the job: http://ec2-52-90-149-194.compute-1.amazonaws.com:8088/proxy/application_1459126378816_0001/
16/03/28 01:16:59 INFO mapreduce.Job: Running job: job_1459126378816_0001
16/03/28 01:17:08 INFO mapreduce.Job: Job job_1459126378816_0001 running in uber mode : false
16/03/28 01:17:08 INFO mapreduce.Job:  map 0% reduce 0%
16/03/28 01:17:36 INFO mapreduce.Job:  map 1% reduce 0%
16/03/28 01:17:40 INFO mapreduce.Job:  map 2% reduce 0%
16/03/28 01:17:43 INFO mapreduce.Job:  map 3% reduce 0%
16/03/28 01:17:46 INFO mapreduce.Job:  map 4% reduce 0%
16/03/28 01:18:05 INFO mapreduce.Job:  map 5% reduce 0%
16/03/28 01:18:21 INFO mapreduce.Job:  map 6% reduce 0%
16/03/28 01:18:26 INFO mapreduce.Job:  map 7% reduce 0%
16/03/28 01:18:34 INFO mapreduce.Job:  map 8% reduce 0%
16/03/28 01:19:07 INFO mapreduce.Job:  map 9% reduce 0%
16/03/28 01:19:10 INFO mapreduce.Job:  map 10% reduce 0%
16/03/28 01:19:14 INFO mapreduce.Job:  map 11% reduce 0%
16/03/28 01:19:17 INFO mapreduce.Job:  map 12% reduce 0%
16/03/28 01:19:33 INFO mapreduce.Job:  map 13% reduce 0%
16/03/28 01:19:45 INFO mapreduce.Job:  map 14% reduce 0%
16/03/28 01:19:57 INFO mapreduce.Job:  map 15% reduce 0%
16/03/28 01:20:06 INFO mapreduce.Job:  map 16% reduce 0%
16/03/28 01:20:23 INFO mapreduce.Job:  map 17% reduce 0%
16/03/28 01:20:36 INFO mapreduce.Job:  map 18% reduce 0%
16/03/28 01:20:40 INFO mapreduce.Job:  map 19% reduce 0%
16/03/28 01:20:45 INFO mapreduce.Job:  map 20% reduce 0%
16/03/28 01:21:01 INFO mapreduce.Job:  map 21% reduce 0%
16/03/28 01:21:03 INFO mapreduce.Job:  map 22% reduce 0%
16/03/28 01:21:17 INFO mapreduce.Job:  map 23% reduce 0%
16/03/28 01:21:22 INFO mapreduce.Job:  map 24% reduce 0%
16/03/28 01:21:35 INFO mapreduce.Job:  map 24% reduce 1%
16/03/28 01:21:38 INFO mapreduce.Job:  map 24% reduce 2%
16/03/28 01:21:39 INFO mapreduce.Job:  map 25% reduce 2%
16/03/28 01:21:41 INFO mapreduce.Job:  map 25% reduce 3%
16/03/28 01:21:45 INFO mapreduce.Job:  map 25% reduce 4%
16/03/28 01:21:48 INFO mapreduce.Job:  map 25% reduce 5%
16/03/28 01:21:51 INFO mapreduce.Job:  map 25% reduce 6%
16/03/28 01:21:54 INFO mapreduce.Job:  map 25% reduce 8%
16/03/28 01:21:57 INFO mapreduce.Job:  map 26% reduce 8%
16/03/28 01:22:00 INFO mapreduce.Job:  map 27% reduce 8%
16/03/28 01:22:12 INFO mapreduce.Job:  map 28% reduce 8%
16/03/28 01:22:16 INFO mapreduce.Job:  map 29% reduce 8%
16/03/28 01:22:30 INFO mapreduce.Job:  map 30% reduce 8%
16/03/28 01:22:35 INFO mapreduce.Job:  map 31% reduce 8%
16/03/28 01:22:38 INFO mapreduce.Job:  map 31% reduce 9%
16/03/28 01:22:41 INFO mapreduce.Job:  map 31% reduce 10%
16/03/28 01:23:01 INFO mapreduce.Job:  map 32% reduce 10%
16/03/28 01:23:04 INFO mapreduce.Job:  map 33% reduce 10%
16/03/28 01:23:09 INFO mapreduce.Job:  map 34% reduce 10%
16/03/28 01:23:26 INFO mapreduce.Job:  map 35% reduce 10%
```

```
ubuntu@ip-172-31-0-83: ~
16/03/28 01:32:38 INFO mapreduce.Job:  map 86% reduce 28%
16/03/28 01:32:41 INFO mapreduce.Job:  map 87% reduce 28%
16/03/28 01:32:52 INFO mapreduce.Job:  map 88% reduce 28%
16/03/28 01:32:57 INFO mapreduce.Job:  map 89% reduce 28%
16/03/28 01:33:12 INFO mapreduce.Job:  map 90% reduce 28%
16/03/28 01:33:18 INFO mapreduce.Job:  map 90% reduce 29%
16/03/28 01:33:20 INFO mapreduce.Job:  map 91% reduce 29%
16/03/28 01:33:27 INFO mapreduce.Job:  map 91% reduce 30%
16/03/28 01:33:44 INFO mapreduce.Job:  map 92% reduce 30%
16/03/28 01:33:51 INFO mapreduce.Job:  map 93% reduce 30%
16/03/28 01:33:55 INFO mapreduce.Job:  map 94% reduce 30%
16/03/28 01:34:07 INFO mapreduce.Job:  map 94% reduce 31%
16/03/28 01:34:10 INFO mapreduce.Job:  map 95% reduce 31%
16/03/28 01:34:14 INFO mapreduce.Job:  map 96% reduce 31%
16/03/28 01:34:33 INFO mapreduce.Job:  map 97% reduce 31%
16/03/28 01:34:36 INFO mapreduce.Job:  map 98% reduce 31%
16/03/28 01:34:38 INFO mapreduce.Job:  map 98% reduce 32%
16/03/28 01:34:48 INFO mapreduce.Job:  map 99% reduce 32%
16/03/28 01:34:53 INFO mapreduce.Job:  map 99% reduce 33%
16/03/28 01:35:05 INFO mapreduce.Job:  map 100% reduce 33%
16/03/28 01:35:18 INFO mapreduce.Job:  map 100% reduce 44%
16/03/28 01:35:21 INFO mapreduce.Job:  map 100% reduce 67%
16/03/28 01:35:30 INFO mapreduce.Job:  map 100% reduce 68%
16/03/28 01:35:36 INFO mapreduce.Job:  map 100% reduce 69%
16/03/28 01:35:45 INFO mapreduce.Job:  map 100% reduce 70%
16/03/28 01:35:51 INFO mapreduce.Job:  map 100% reduce 71%
16/03/28 01:35:57 INFO mapreduce.Job:  map 100% reduce 72%
16/03/28 01:36:06 INFO mapreduce.Job:  map 100% reduce 73%
16/03/28 01:36:12 INFO mapreduce.Job:  map 100% reduce 74%
16/03/28 01:36:18 INFO mapreduce.Job:  map 100% reduce 75%
16/03/28 01:36:27 INFO mapreduce.Job:  map 100% reduce 76%
16/03/28 01:36:33 INFO mapreduce.Job:  map 100% reduce 77%
16/03/28 01:36:39 INFO mapreduce.Job:  map 100% reduce 78%
16/03/28 01:36:48 INFO mapreduce.Job:  map 100% reduce 79%
16/03/28 01:36:54 INFO mapreduce.Job:  map 100% reduce 80%
16/03/28 01:37:06 INFO mapreduce.Job:  map 100% reduce 81%
16/03/28 01:37:09 INFO mapreduce.Job:  map 100% reduce 82%
16/03/28 01:37:15 INFO mapreduce.Job:  map 100% reduce 83%
16/03/28 01:37:21 INFO mapreduce.Job:  map 100% reduce 84%
16/03/28 01:37:30 INFO mapreduce.Job:  map 100% reduce 85%
16/03/28 01:37:36 INFO mapreduce.Job:  map 100% reduce 86%
16/03/28 01:37:42 INFO mapreduce.Job:  map 100% reduce 87%
16/03/28 01:37:48 INFO mapreduce.Job:  map 100% reduce 88%
16/03/28 01:37:57 INFO mapreduce.Job:  map 100% reduce 89%
16/03/28 01:38:04 INFO mapreduce.Job:  map 100% reduce 90%
16/03/28 01:38:10 INFO mapreduce.Job:  map 100% reduce 91%
16/03/28 01:38:19 INFO mapreduce.Job:  map 100% reduce 92%
16/03/28 01:38:25 INFO mapreduce.Job:  map 100% reduce 93%
16/03/28 01:38:31 INFO mapreduce.Job:  map 100% reduce 94%
16/03/28 01:38:40 INFO mapreduce.Job:  map 100% reduce 95%
16/03/28 01:38:46 INFO mapreduce.Job:  map 100% reduce 96%
16/03/28 01:38:52 INFO mapreduce.Job:  map 100% reduce 97%
16/03/28 01:38:58 INFO mapreduce.Job:  map 100% reduce 98%
16/03/28 01:39:07 INFO mapreduce.Job:  map 100% reduce 99%
16/03/28 01:39:13 INFO mapreduce.Job:  map 100% reduce 100%
16/03/28 01:39:16 INFO mapreduce.Job: Job job_1459126378816_0001 completed successfully
16/03/28 01:39:16 INFO mapreduce.Job: Counters: 50
        File System Counters
```

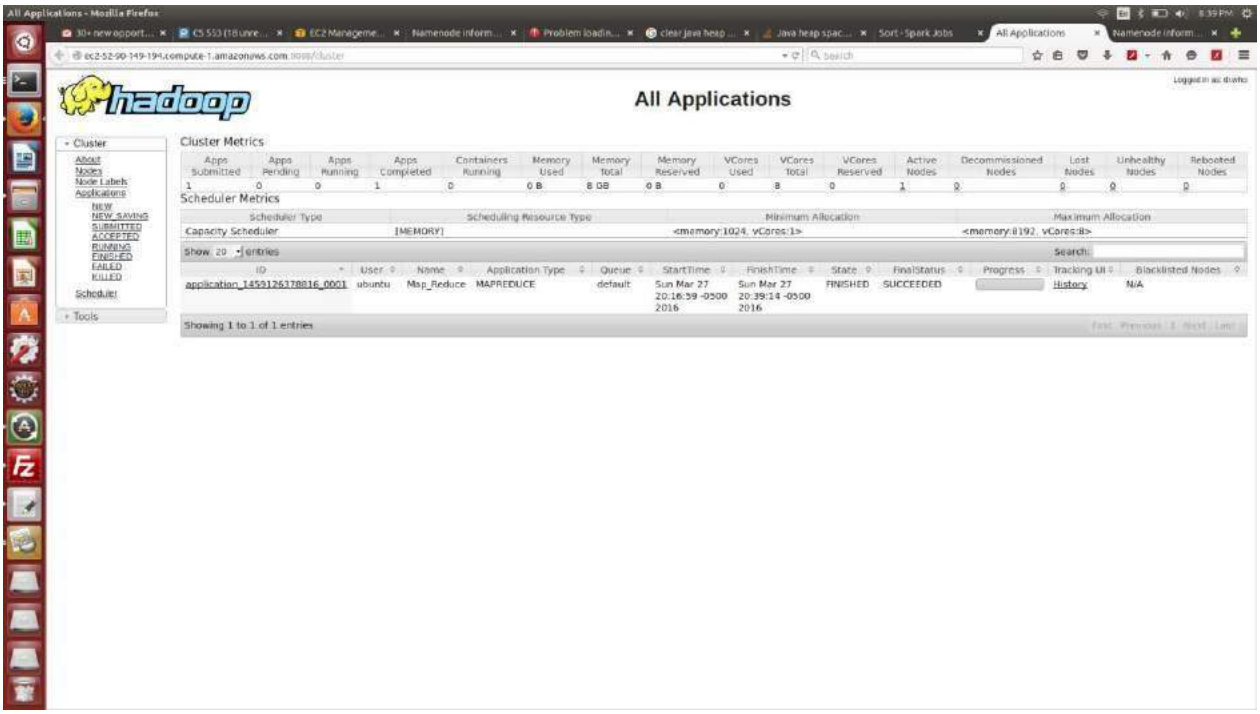
We can track the job that our application is running and also track its completion status.

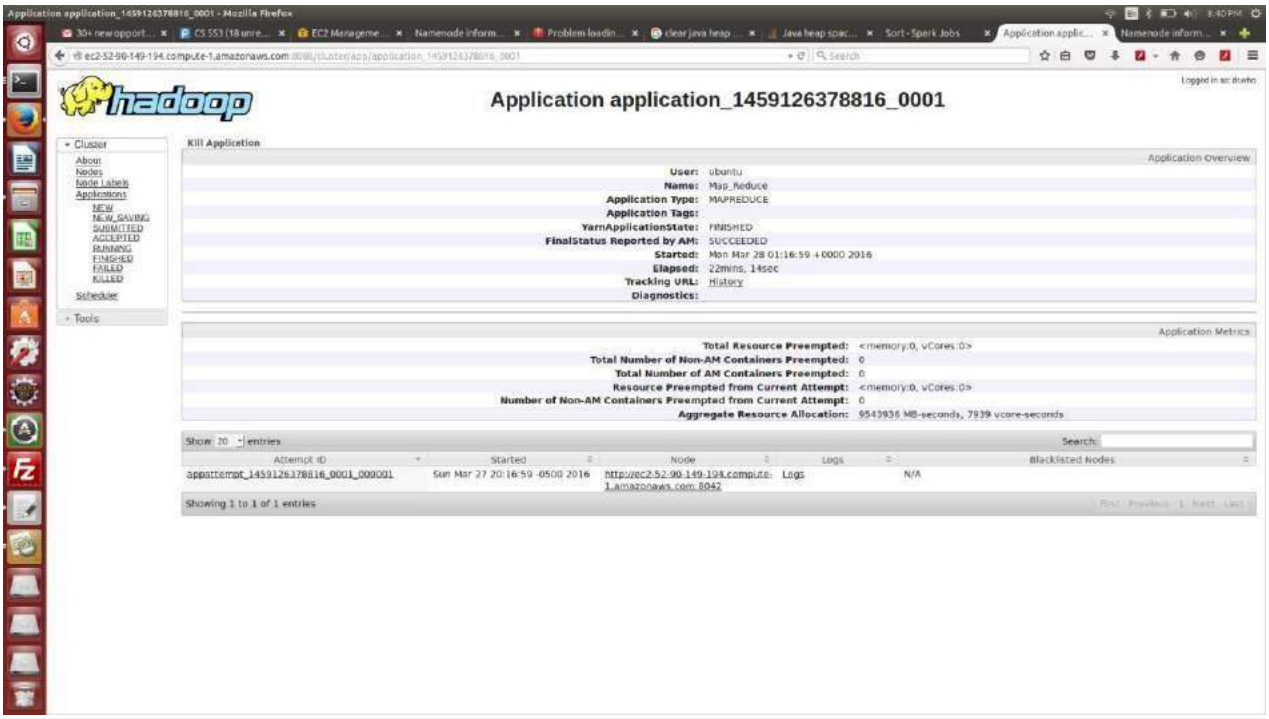


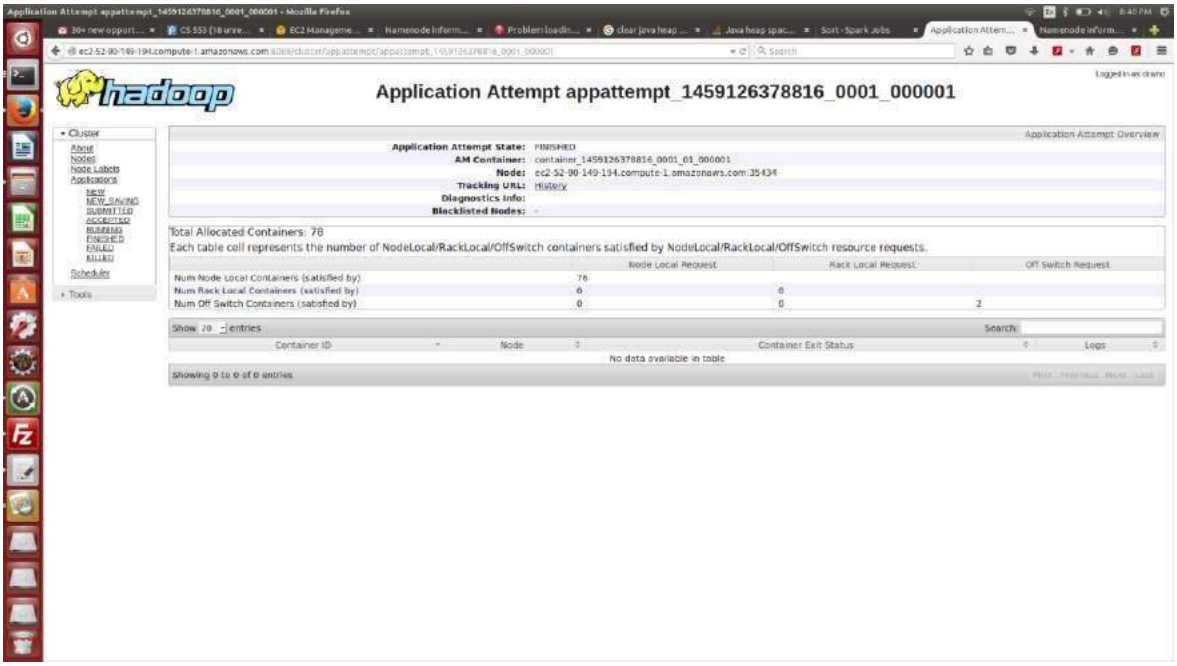


- Once the job is finished, the number of Bytes Written will be displayed on the terminal. In our case, the number of bytes written are 10GB.

- The details of the job completion are mentioned in the below screenshot.


All Applications - Mozilla Firefox
ec2-52-90-149-194.compute-1.amazonaws.com:8088/cluster
hadoop
All Applications
Logged in as: dr.who
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
1 | 0 | 0 | 1 | 0 | 0 B | 8 GB | 0 B | 0 | 8 | 0 | 1 | 0 | 0 | 0 | 0
Scheduler Metrics
Scheduler Type: Capacity Scheduler
Scheduling Resource Type: [MEMORY]
Minimum Allocation: <memory:1024, vCores:1>
Maximum Allocation: <memory:8192, vCores:8>
Show 20 entries
Search:
ID | User | Name | Application Type | Queue | StartTime | FinishTime | State | FinalStatus | Progress | Tracking UI | Blacklisted Nodes
application_1459126378816_0001 | ubuntu | Map_Reduce | MAPREDUCE | default | Sun Mar 27 20:16:59 -0500 2016 | Sun Mar 27 20:39:14 -0500 2016 | FINISHED | SUCCEEDED | | History | N/A
Showing 1 to 1 of 1 entries


Application application_1459126378816_0001 - Mozilla Firefox
ec2-52-90-149-194.compute-1.amazonaws.com:8088/cluster/app/application_1459126378816_0001
hadoop
Application application_1459126378816_0001
Logged in as: dr.who
Cluster
About
Nodes
Node Labels
Applications
NEW
NEW_SAVING
SUBMITTED
ACCEPTED
RUNNING
FINISHED
FAILED
KILLED
Scheduler
Tools
Kill Application
Application Overview
User: ubuntu
Name: Map_Reduce
Application Type: MAPREDUCE
Application Tags:
YarnApplicationState: FINISHED
FinalStatus Reported by AM: SUCCEEDED
Started: Mon Mar 28 01:16:59 +0000 2016
Elapsed: 22mins, 14sec
Tracking URL: History
Diagnostics:
Application Metrics
Total Resource Preempted: <memory:0, vCores:0>
Total Number of Non-AM Containers Preempted: 0
Total Number of AM Containers Preempted: 0
Resource Preempted from Current Attempt: <memory:0, vCores:0>
Number of Non-AM Containers Preempted from Current Attempt: 0
Aggregate Resource Allocation: 9543936 MB-seconds, 7939 vcore-seconds
Show 20 entries
Search:
Attempt ID | Started | Node | Logs | Blacklisted Nodes
appattempt_1459126378816_0001_000001 | Sun Mar 27 20:16:59 -0500 2016 | http://ec2-52-90-149-194.compute-1.amazonaws.com:8042 | Logs | N/A
Showing 1 to 1 of 1 entries



As stated in the above screenshot, the time taken to sort 10GB of data using Hadoop map-reduce with single node virtual cluster is 22mins 14 secs.

- Now the generated output file is transferred back to the mounted disk for evaluation. This can be done using the below commands,

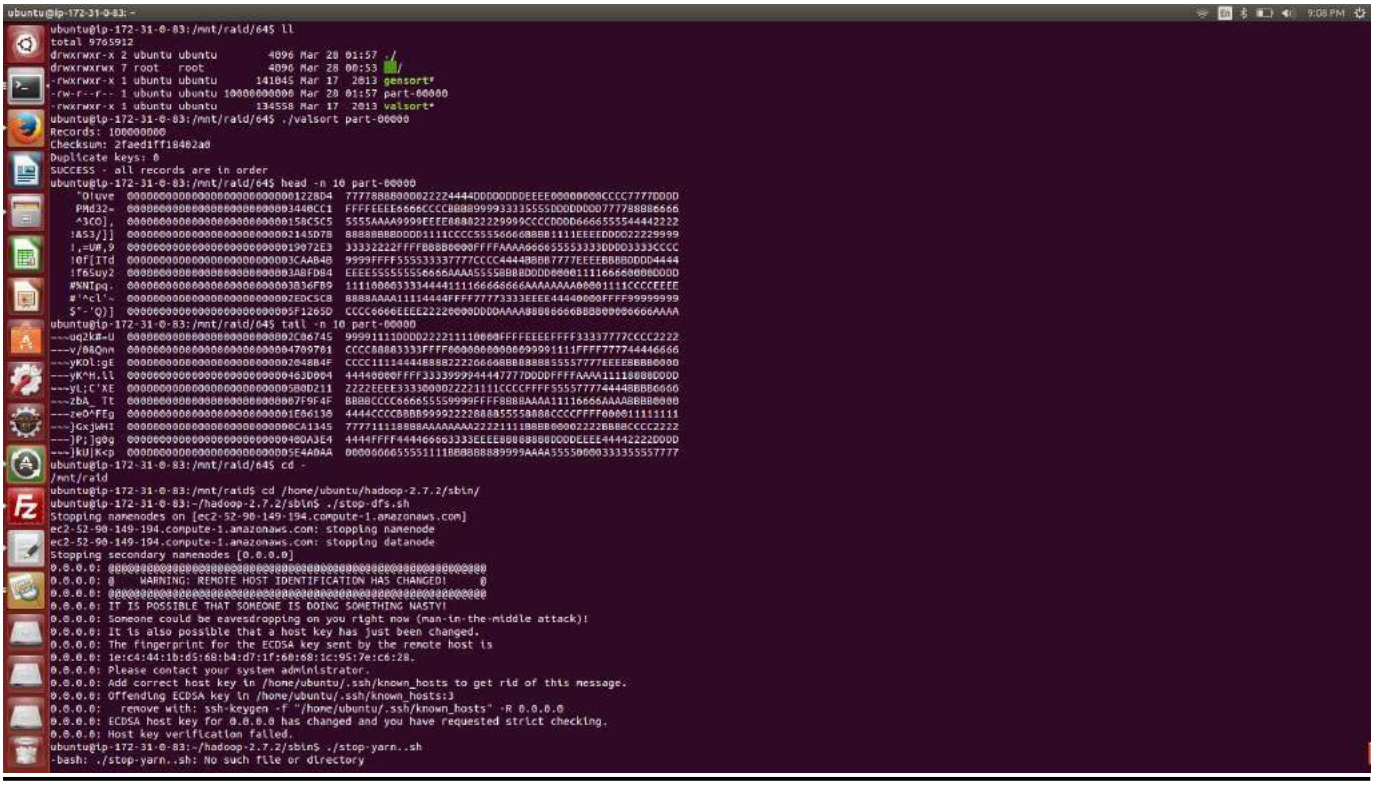
  ./hadoop fs -get /user/hduser/output/part-00000 /mnt/raid/64/

  The outfile is evaluated using Valsort by below command,

  ./valsort Output-file-Name

  The valsort command shows that all the records in the output file are in order.

  The first and last 10 lines of the output file are printed



- Finally, the Hadoop nodes can be stopped using the below commands,

  ./stop-dfs.sh

  ./stop-yarn.sh

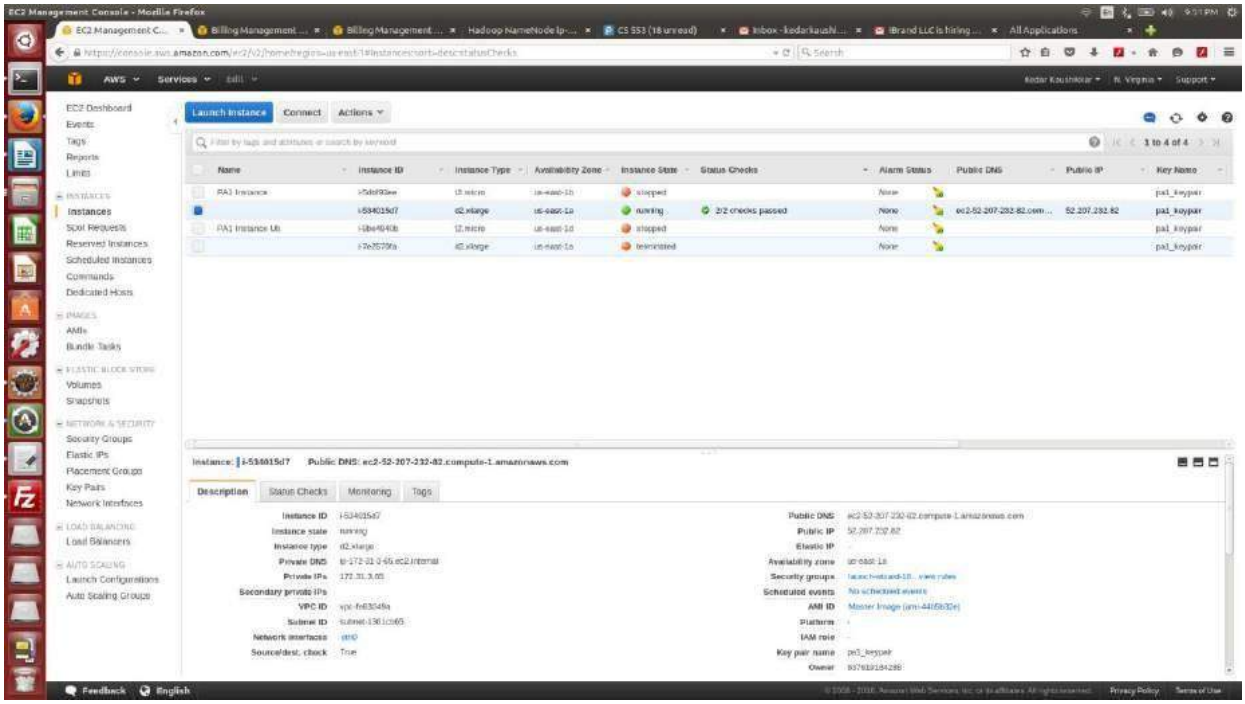
<u>Performance:</u>

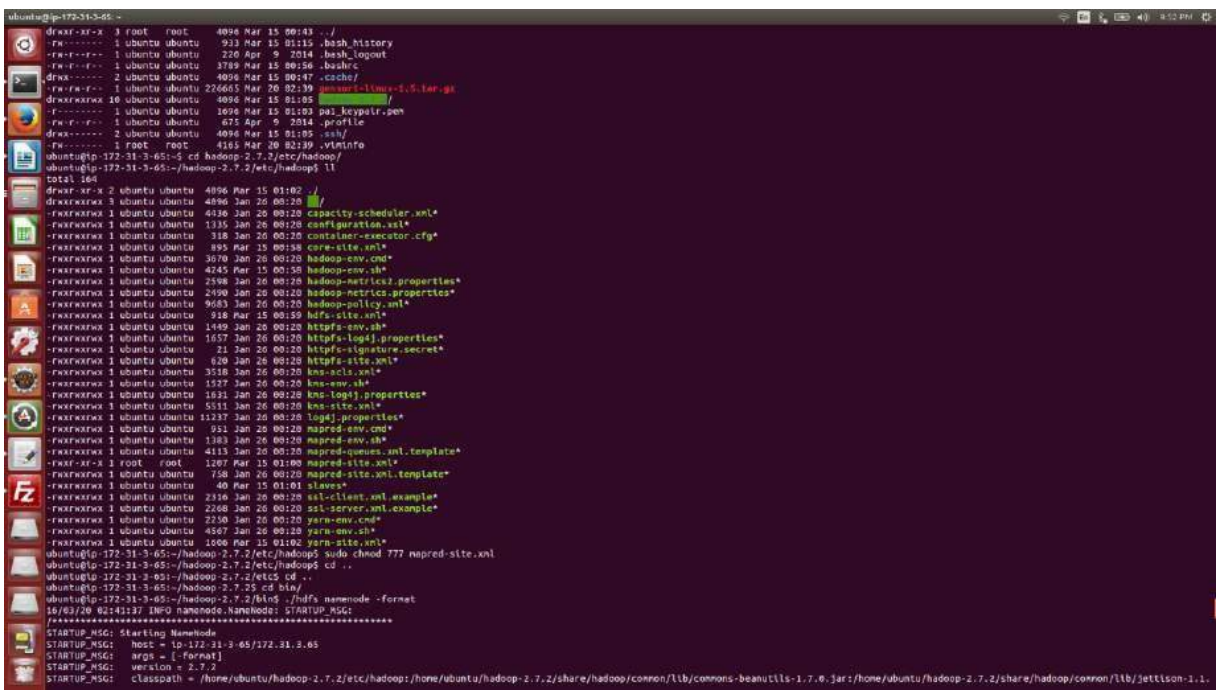
Throughput = (Datasize / Time taken)
= 10000 / 1334
= 7.49 Mb/sec

<u>Hadoop 1 Node TeraSort (1GB)</u>

- Following the above steps for installing Hadoop on the virtual cluster, we create another d2.xlarge instance for our Hadoop 1 node mapreduce job for sorting a 1GB file.

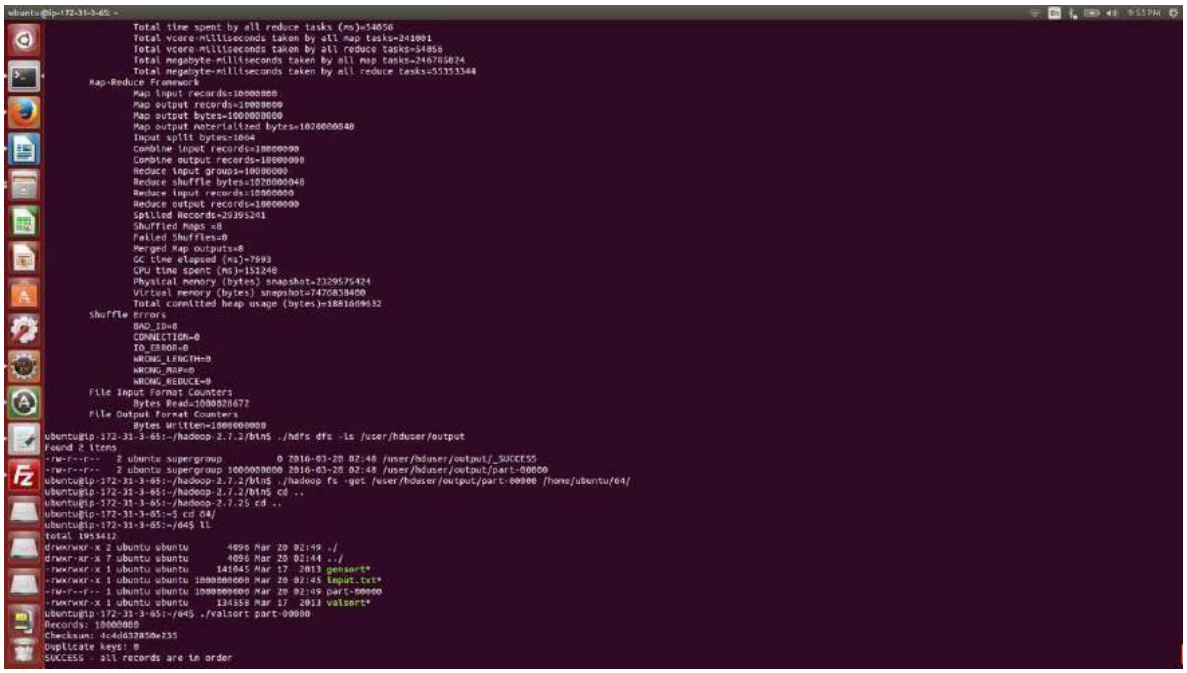
  All the steps are same as above configuration and below mentioned are the screenshots for 1GB dataset

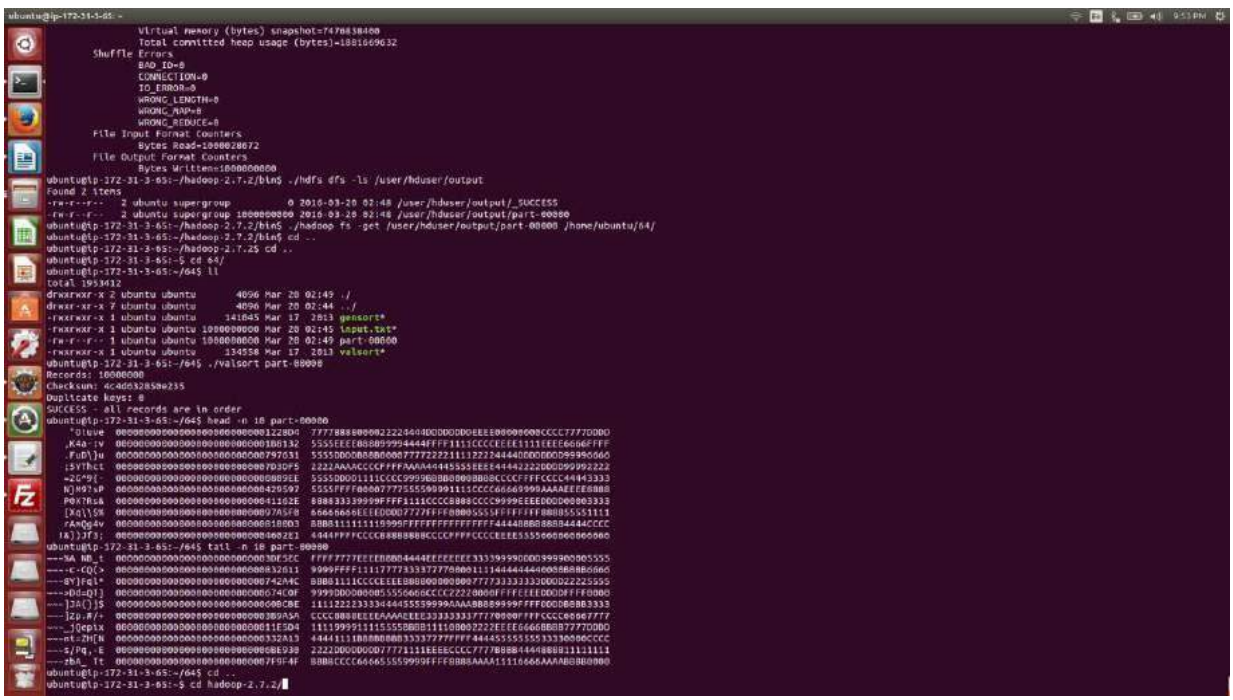

EC2 Management Console - Mozilla Firefox
AWS
Services
Launch Instance
Connect
Actions
EC2 Dashboard
Events
Tags
Reports
Limits
Instances
Spot Requests
Reserved Instances
Scheduled Instances
Commands
Dedicated Hosts
AMIs
Bundle Tasks
Volumes
Snapshots
Security Groups
Elastic IPs
Placement Groups
Key Pairs
Network Interfaces
Load Balancers
Launch Configurations
Auto Scaling Groups
Name
Instance ID
Instance Type
Availability Zone
Instance State
Status Checks
Alarm Status
Public DNS
Public IP
Key Name
running
2/2 checks passed
stopped
terminated
Description
Status Checks
Monitoring
Tags
Feedback
English


ubuntu@ip-172-31-3-65: ~
ubuntu@ip-172-31-3-65:~/hadoop-2.7.2/etc/hadoop$ sudo chmod 777 mapred-site.xml
ubuntu@ip-172-31-3-65:~/hadoop-2.7.2/etc/hadoop$ cd ..
ubuntu@ip-172-31-3-65:~/hadoop-2.7.2/etc$ cd ..
ubuntu@ip-172-31-3-65:~/hadoop-2.7.2$ cd bin/
ubuntu@ip-172-31-3-65:~/hadoop-2.7.2/bin$ ./hdfs namenode -format
16/03/20 02:41:37 INFO namenode.NameNode: STARTUP_MSG:
/**
STARTUP_MSG: Starting NameNode
STARTUP_MSG: host = ip-172-31-3-65/172.31.3.65
STARTUP_MSG: args = [-format]
STARTUP_MSG: version = 2.7.2

```
ubuntu@ip-172-31-3-65: ~
16/03/20 02:41:41 INFO namenode.FSImage: Allocated new BlockPoolId: BP-535733952-172.31.3.65-1458441701822
16/03/20 02:41:41 INFO common.Storage: Storage directory /tmp/hadoop-ubuntu/dfs/name has been successfully formatted.
16/03/20 02:41:41 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
16/03/20 02:41:41 INFO util.ExitUtil: Exiting with status 0
16/03/20 02:41:41 INFO namenode.NameNode: SHUTDOWN_MSG:
/************************************************************
SHUTDOWN_MSG: Shutting down NameNode at ip-172-31-3-65/172.31.3.65
************************************************************/
ubuntu@ip-172-31-3-65:~/hadoop-2.7.2/bin$ cd ..
ubuntu@ip-172-31-3-65:~/hadoop-2.7.2$ cd ..
ubuntu@ip-172-31-3-65:~$ eval `ssh-agent -s`
Agent pid 1691
ubuntu@ip-172-31-3-65:~$ ssh-add pa1_keypair.pem
Identity added: pa1_keypair.pem (pa1_keypair.pem)
ubuntu@ip-172-31-3-65:~$ cd hadoop-2.7.2/sbin/
ubuntu@ip-172-31-3-65:~/hadoop-2.7.2/sbin$ ./start-dfs.sh
Starting namenodes on [ec2-52-207-232-82.compute-1.amazonaws.com]
The authenticity of host 'ec2-52-207-232-82.compute-1.amazonaws.com (172.31.3.65)' can't be established.
ECDSA key fingerprint is 95:1b:c8:c0:38:5c:48:d0:63:07:4a:8c:f0:52:b1:a0.
Are you sure you want to continue connecting (yes/no)? yes
ec2-52-207-232-82.compute-1.amazonaws.com: Warning: Permanently added 'ec2-52-207-232-82.compute-1.amazonaws.com,172.31.3.65' (ECDSA) to the list of known hosts.
ec2-52-207-232-82.compute-1.amazonaws.com: starting namenode, logging to /home/ubuntu/hadoop-2.7.2/logs/hadoop-ubuntu-namenode-ip-172-31-3-65.out
ec2-52-207-232-82.compute-1.amazonaws.com: starting datanode, logging to /home/ubuntu/hadoop-2.7.2/logs/hadoop-ubuntu-datanode-ip-172-31-3-65.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: @@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
0.0.0.0: @    WARNING: REMOTE HOST IDENTIFICATION HAS CHANGED!     @
0.0.0.0: @@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
0.0.0.0: IT IS POSSIBLE THAT SOMEONE IS DOING SOMETHING NASTY!
0.0.0.0: Someone could be eavesdropping on you right now (man-in-the-middle attack)!
0.0.0.0: It is also possible that a host key has just been changed.
0.0.0.0: The fingerprint for the ECDSA key sent by the remote host is
0.0.0.0: 95:1b:c8:c0:38:5c:48:d0:63:07:4a:8c:f0:52:b1:a0.
0.0.0.0: Please contact your system administrator.
0.0.0.0: Add correct host key in /home/ubuntu/.ssh/known_hosts to get rid of this message.
0.0.0.0: Offending ECDSA key in /home/ubuntu/.ssh/known_hosts:3
0.0.0.0:  remove with: ssh-keygen -f "/home/ubuntu/.ssh/known_hosts" -R 0.0.0.0
0.0.0.0: ECDSA host key for 0.0.0.0 has changed and you have requested strict checking.
0.0.0.0: Host key verification failed.
ubuntu@ip-172-31-3-65:~/hadoop-2.7.2/sbin$ ./start-yarn.sh
starting yarn daemons
starting resourcemanager, logging to /home/ubuntu/hadoop-2.7.2/logs/yarn-ubuntu-resourcemanager-ip-172-31-3-65.out
ec2-52-207-232-82.compute-1.amazonaws.com: starting nodemanager, logging to /home/ubuntu/hadoop-2.7.2/logs/yarn-ubuntu-nodemanager-ip-172-31-3-65.out
ubuntu@ip-172-31-3-65:~/hadoop-2.7.2/sbin$ jps
7266 ResourceManager
1988 DataNode
1844 NameNode
7694 Jps
2410 NodeManager
ubuntu@ip-172-31-3-65:~/hadoop-2.7.2/sbin$ cd ..
ubuntu@ip-172-31-3-65:~/hadoop-2.7.2$ cd bin/
ubuntu@ip-172-31-3-65:~/hadoop-2.7.2/bin$ ./hdfs dfs -mkdir /user
ubuntu@ip-172-31-3-65:~/hadoop-2.7.2/bin$ ./hdfs dfs -mkdir /user/hduser
ubuntu@ip-172-31-3-65:~/hadoop-2.7.2/bin$ ./hdfs dfs -mkdir /user/hduser/input
ubuntu@ip-172-31-3-65:~/hadoop-2.7.2/bin$ cd ..
ubuntu@ip-172-31-3-65:~/hadoop-2.7.2$ cd ..
ubuntu@ip-172-31-3-65:~$ tar -xzvf gensort-linux-1.5.tar.gz
gpl-2.0.txt
32/
```

```
ubuntu@ip-172-31-3-65: ~
64/
64/valsort
64/gensort
ubuntu@ip-172-31-3-65:~$ cd 64
ubuntu@ip-172-31-3-65:~/64$ ./gensort -a 10000000 input.txt
ubuntu@ip-172-31-3-65:~/64$ cd ..
ubuntu@ip-172-31-3-65:~$ cd hadoop-2.7.2/bin/
ubuntu@ip-172-31-3-65:~/hadoop-2.7.2/bin$ ./hadoop fs -put /home/ubuntu/64/input.txt /user/hduser/input
ubuntu@ip-172-31-3-65:~/hadoop-2.7.2/bin$ ./hdfs dfs -ls /user/hduser/input
Found 1 items
-rw-r--r--   2 ubuntu supergroup 1000000000 2016-03-20 02:45 /user/hduser/input/input.txt
ubuntu@ip-172-31-3-65:~/hadoop-2.7.2/bin$ ./hadoop jar /home/ubuntu/Map_Reduce_TeraSort.jar Map_Reduce /user/hduser/input /user/hduser/output
16/03/20 02:46:23 INFO client.RMProxy: Connecting to ResourceManager at ec2-52-207-232-82.compute-1.amazonaws.com/172.31.3.65:8032
16/03/20 02:46:23 INFO client.RMProxy: Connecting to ResourceManager at ec2-52-207-232-82.compute-1.amazonaws.com/172.31.3.65:8032
16/03/20 02:46:25 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
16/03/20 02:46:23 INFO mapred.FileInputFormat: Total input paths to process : 1
16/03/20 02:46:24 INFO mapreduce.JobSubmitter: number of splits:8
16/03/20 02:46:25 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1458441797282_0001
16/03/20 02:46:26 INFO impl.YarnClientImpl: Submitted application application_1458441797282_0001
16/03/20 02:46:26 INFO mapreduce.Job: The url to track the job: http://ec2-52-207-232-82.compute-1.amazonaws.com:8088/proxy/application_1458441797282_0001/
16/03/20 02:46:26 INFO mapreduce.Job: Running job: job_1458441797282_0001
16/03/20 02:46:34 INFO mapreduce.Job: Job job_1458441797282_0001 running in uber mode : false
16/03/20 02:46:34 INFO mapreduce.Job:  map 0% reduce 0%
16/03/20 02:46:50 INFO mapreduce.Job:  map 4% reduce 0%
16/03/20 02:46:51 INFO mapreduce.Job:  map 28% reduce 0%
16/03/20 02:46:53 INFO mapreduce.Job:  map 29% reduce 0%
16/03/20 02:46:54 INFO mapreduce.Job:  map 34% reduce 0%
16/03/20 02:47:00 INFO mapreduce.Job:  map 41% reduce 0%
16/03/20 02:47:01 INFO mapreduce.Job:  map 44% reduce 0%
16/03/20 02:47:02 INFO mapreduce.Job:  map 46% reduce 0%
16/03/20 02:47:03 INFO mapreduce.Job:  map 50% reduce 0%
16/03/20 02:47:06 INFO mapreduce.Job:  map 51% reduce 0%
16/03/20 02:47:07 INFO mapreduce.Job:  map 52% reduce 0%
16/03/20 02:47:08 INFO mapreduce.Job:  map 53% reduce 0%
16/03/20 02:47:09 INFO mapreduce.Job:  map 59% reduce 0%
16/03/20 02:47:10 INFO mapreduce.Job:  map 65% reduce 0%
16/03/20 02:47:11 INFO mapreduce.Job:  map 71% reduce 0%
16/03/20 02:47:12 INFO mapreduce.Job:  map 75% reduce 0%
16/03/20 02:47:21 INFO mapreduce.Job:  map 81% reduce 0%
16/03/20 02:47:22 INFO mapreduce.Job:  map 93% reduce 0%
16/03/20 02:47:23 INFO mapreduce.Job:  map 93% reduce 25%
16/03/20 02:47:24 INFO mapreduce.Job:  map 96% reduce 25%
16/03/20 02:47:26 INFO mapreduce.Job:  map 100% reduce 29%
16/03/20 02:47:29 INFO mapreduce.Job:  map 100% reduce 68%
16/03/20 02:47:32 INFO mapreduce.Job:  map 100% reduce 73%
16/03/20 02:47:35 INFO mapreduce.Job:  map 100% reduce 78%
16/03/20 02:47:38 INFO mapreduce.Job:  map 100% reduce 83%
16/03/20 02:47:41 INFO mapreduce.Job:  map 100% reduce 88%
16/03/20 02:47:44 INFO mapreduce.Job:  map 100% reduce 93%
16/03/20 02:47:47 INFO mapreduce.Job:  map 100% reduce 94%
16/03/20 02:48:07 INFO mapreduce.Job:  map 100% reduce 100%
16/03/20 02:48:07 INFO mapreduce.Job: Job job_1458441797282_0001 completed successfully
16/03/20 02:48:07 INFO mapreduce.Job: Counters: 50
        File System Counters
                FILE: Number of bytes read=1978314714
                FILE: Number of bytes written=2999378912
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
```





As stated in the above screenshot, the time taken to sort 1GB of data using Hadoop map-reduce with single node virtual cluster is 2.5206 Minutes.

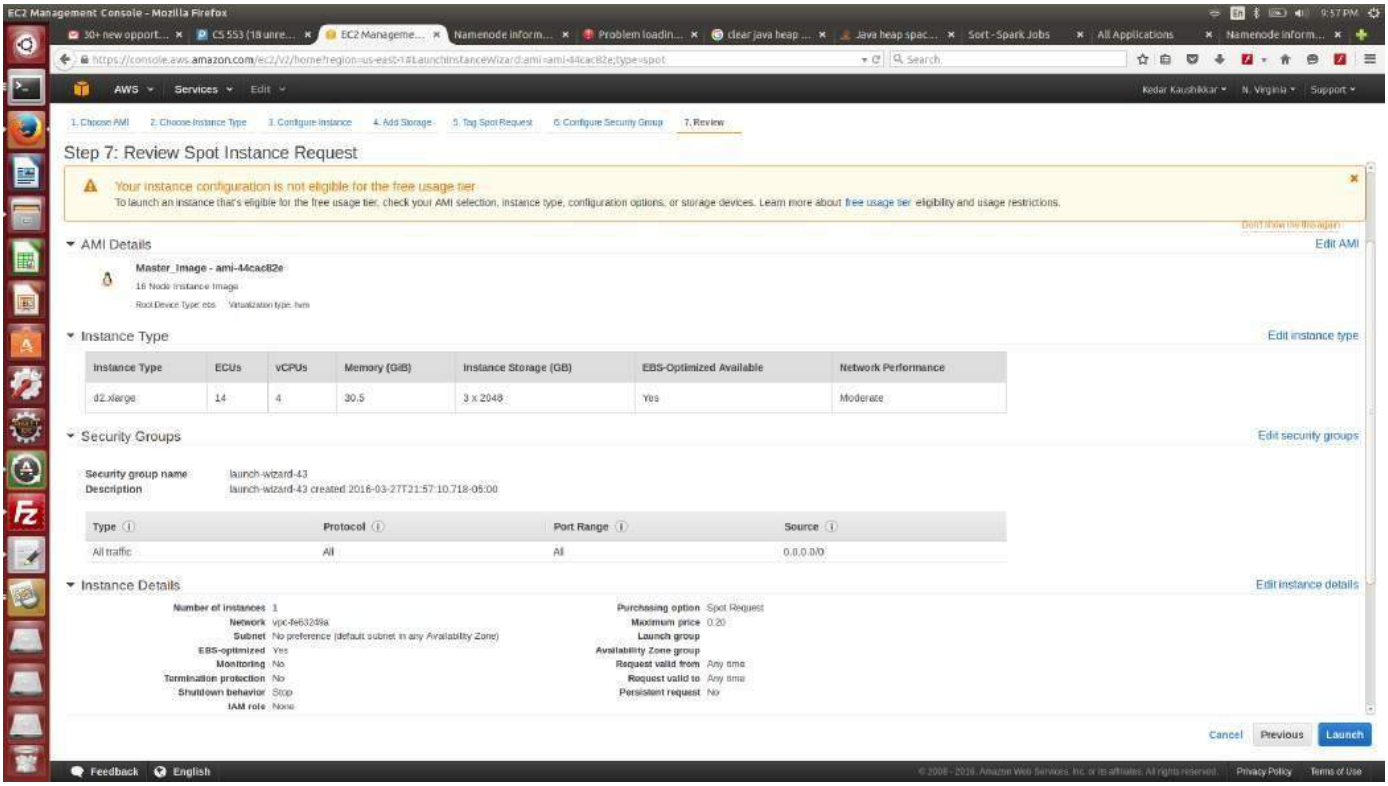
<u>Performance:</u>

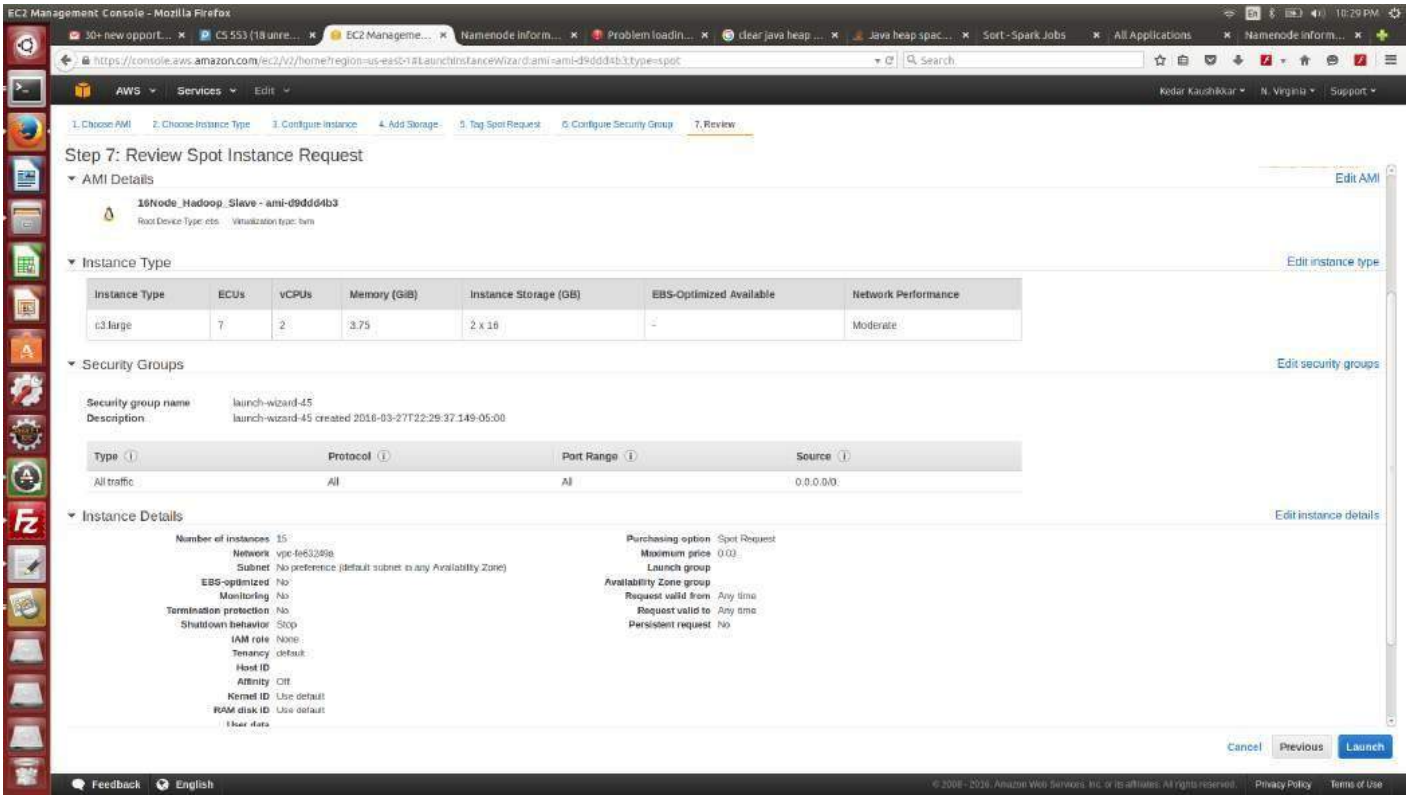
Throughput = (Datasize / Time taken)
= 1000 / 151.24
= 6.612 Mb/sec

Setting up 16 Node Virtual Cluster

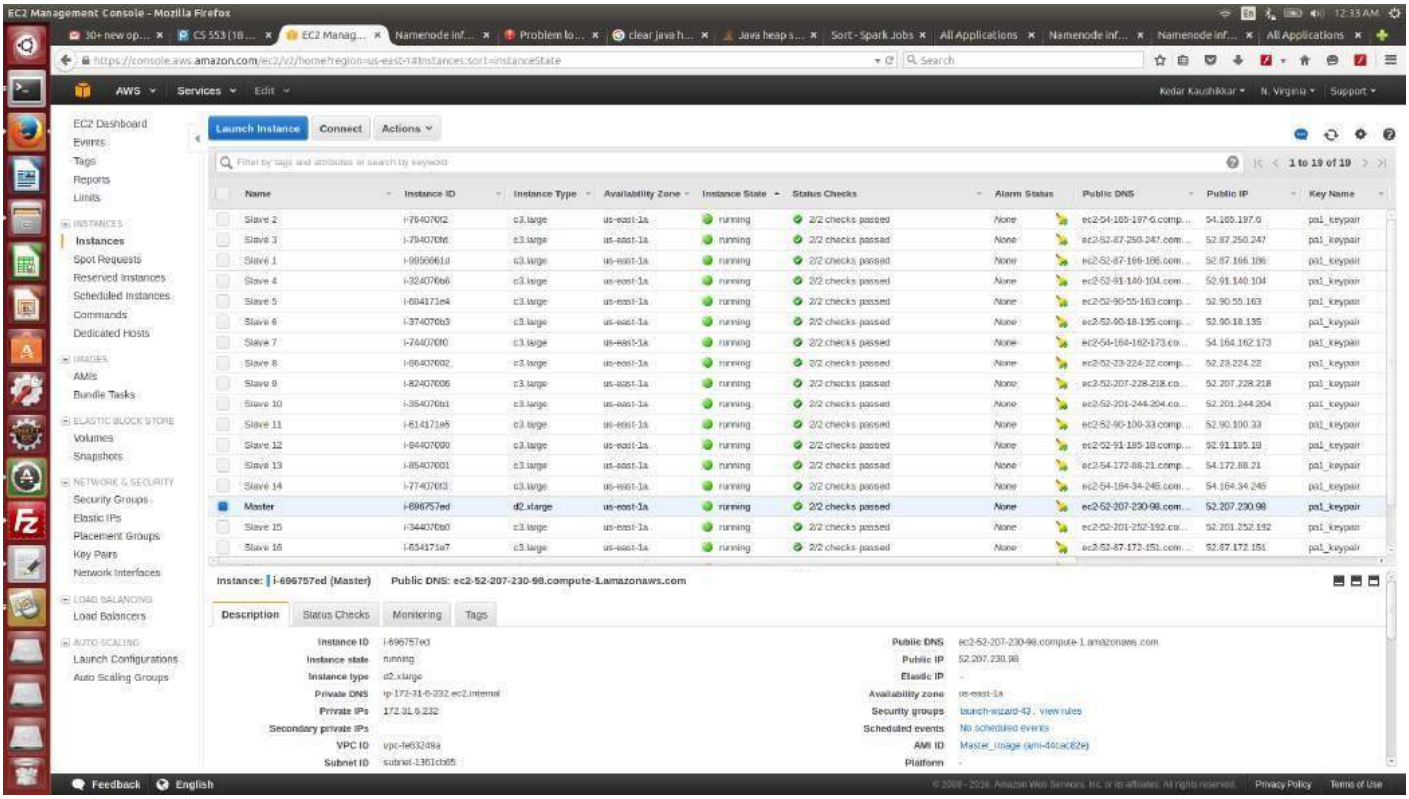
- The Hadoop 16 node cluster is same as the single node cluster when compared with respect to installation of Hadoop but the configuration files for 16 nodes are different for master node.

- For setting up Hadoop 16 node virtual cluster to run a map-reduce job of sorting 100GB of data, we created master node on d2.xlarge amazon ec2 instance and 16 slaves of c3.large .





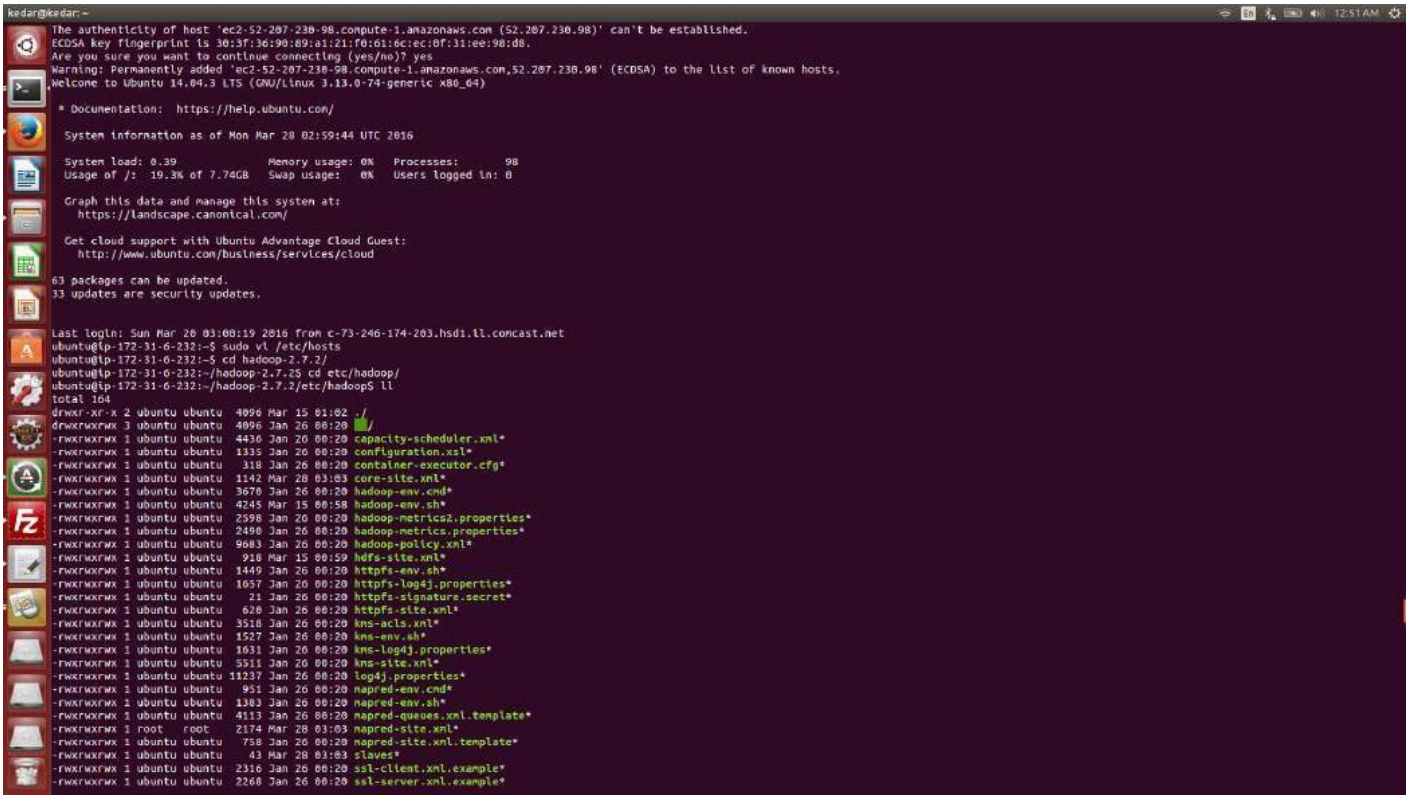
- While creating the instances I provided EBS volume of 150GiB to each slave node that will provide enough memory to store the temp files if the reducers fail and have to revert back to start.
- Once all the instances are granted , our Amazon AWS Dashboard will show up 17 active instances with 1 master and 16 slave nodes.



- As done on single node virtual cluster, we connect the master node on terminal using SSH and install and configure the Hadoop files to set up the 16 active slave nodes that will be linked with it.

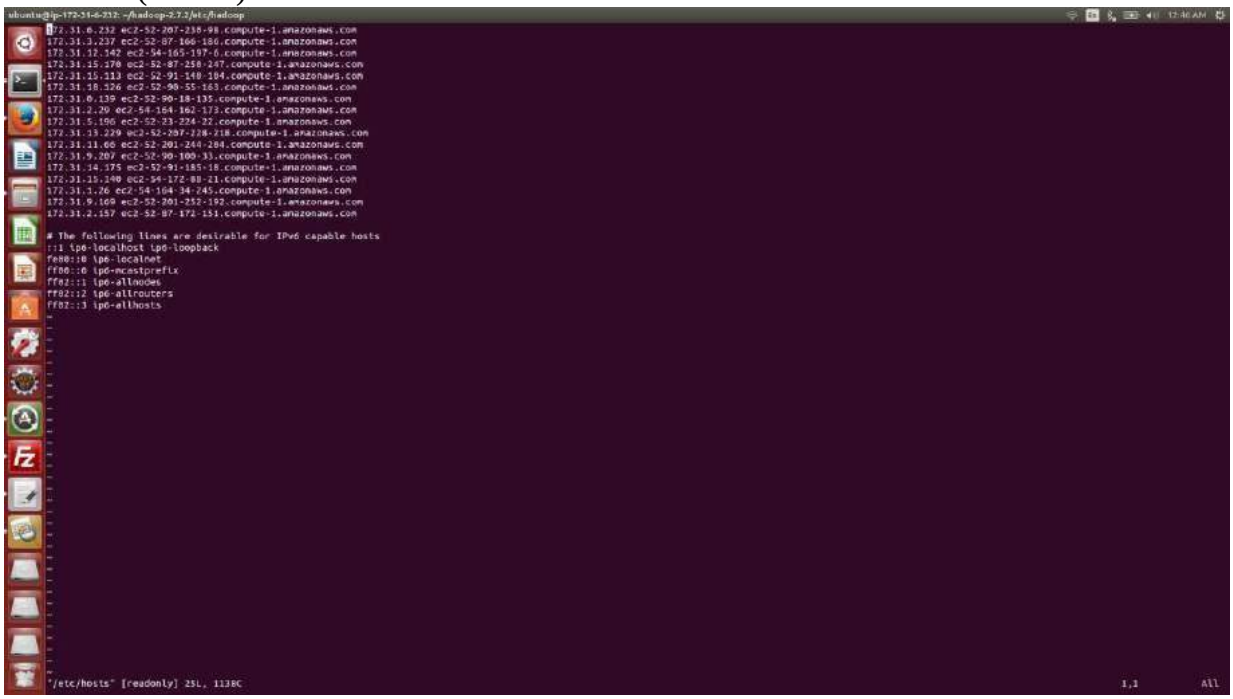
Configuration for Master Node

We connect to the master node using SSH on Public DNS of the Master Node Instance.



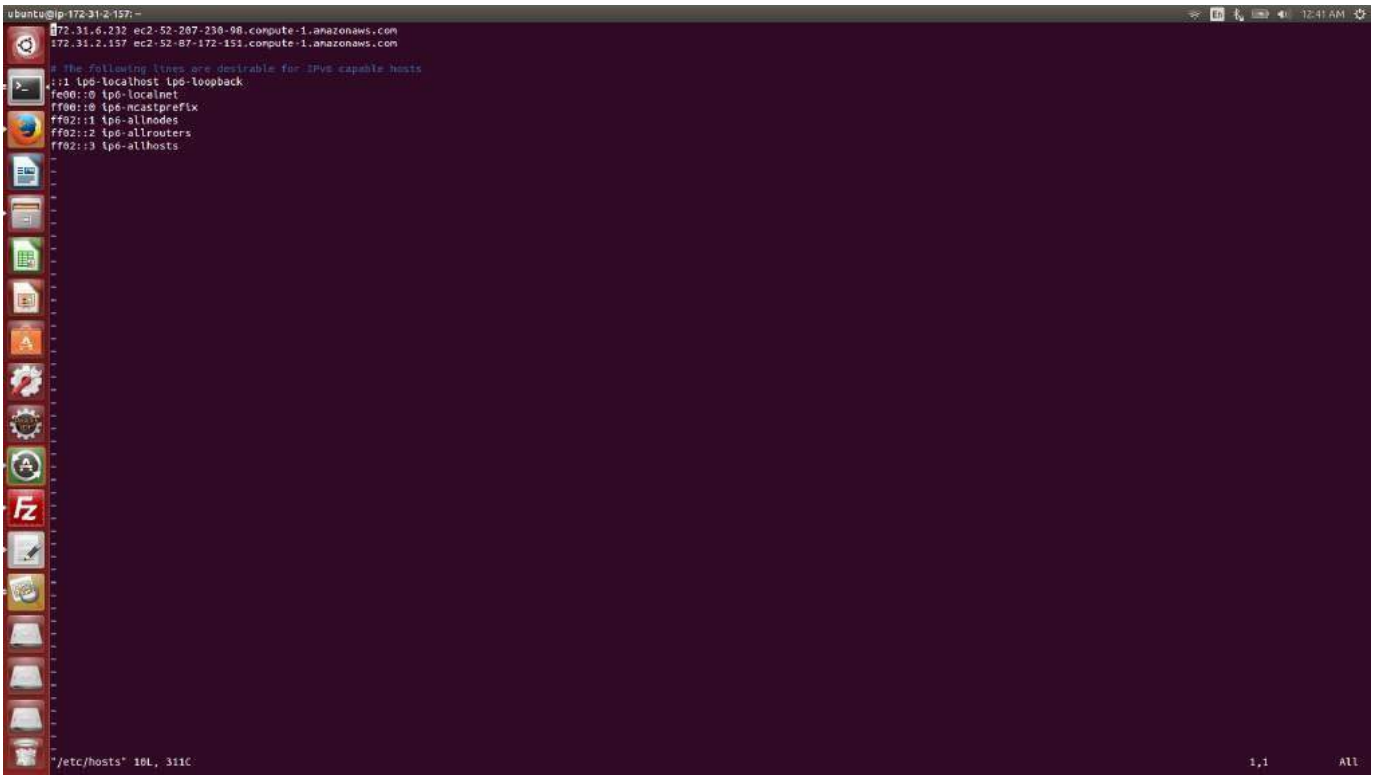
Below files changes are done to configure the Master Node and Slave-Node,

1. /hosts file (Master)



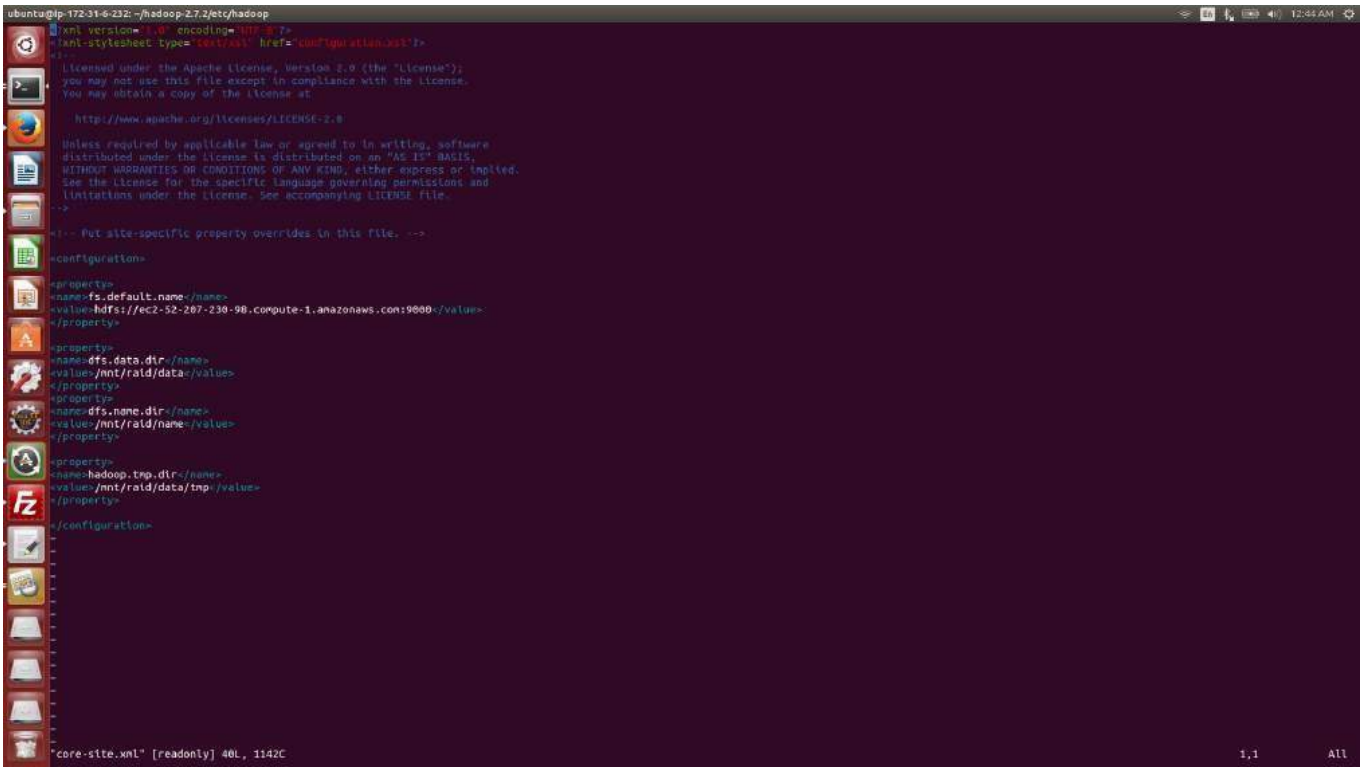
The host file for the master will have its private IP followed by its public DNS.
Similarly it will have the same information of all the slaves that it needs to connect with.

./hosts file (Slave)



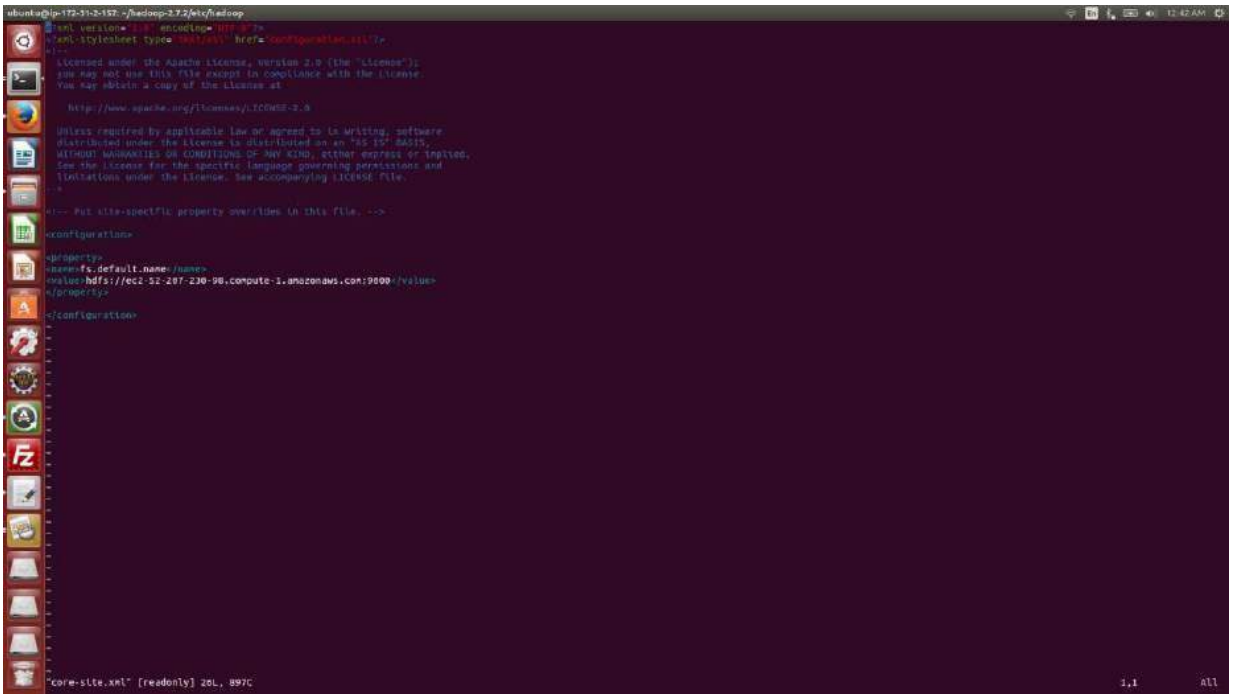
The Host file of Slaves will contain its own Private IP and Public DNS along with master's Private IP and Public DNS.

2. core-site.xml (Master)



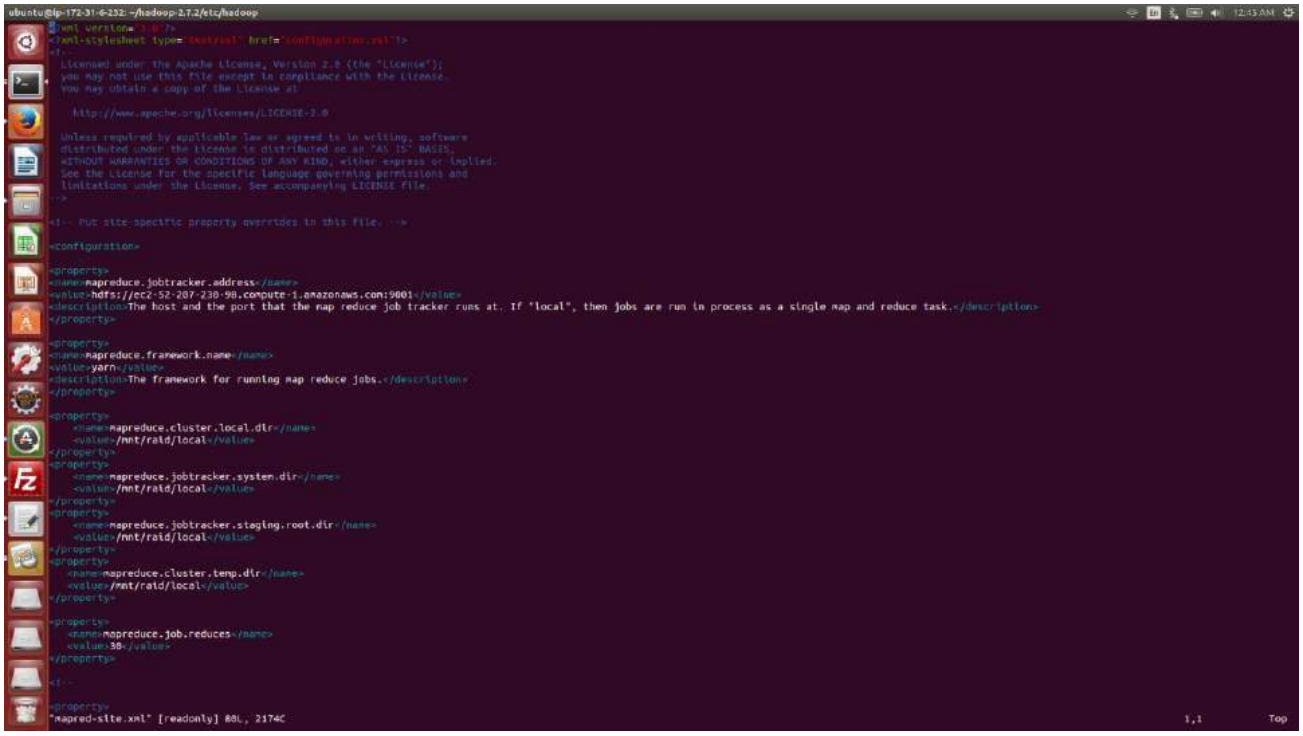
The core site for Master will have the same set of entries as done for single node virtual cluster.

core-site.xml (slave)



The core-site.xml for slave will contain only the default entries i.e. its own file system name having value as the Public DNS.

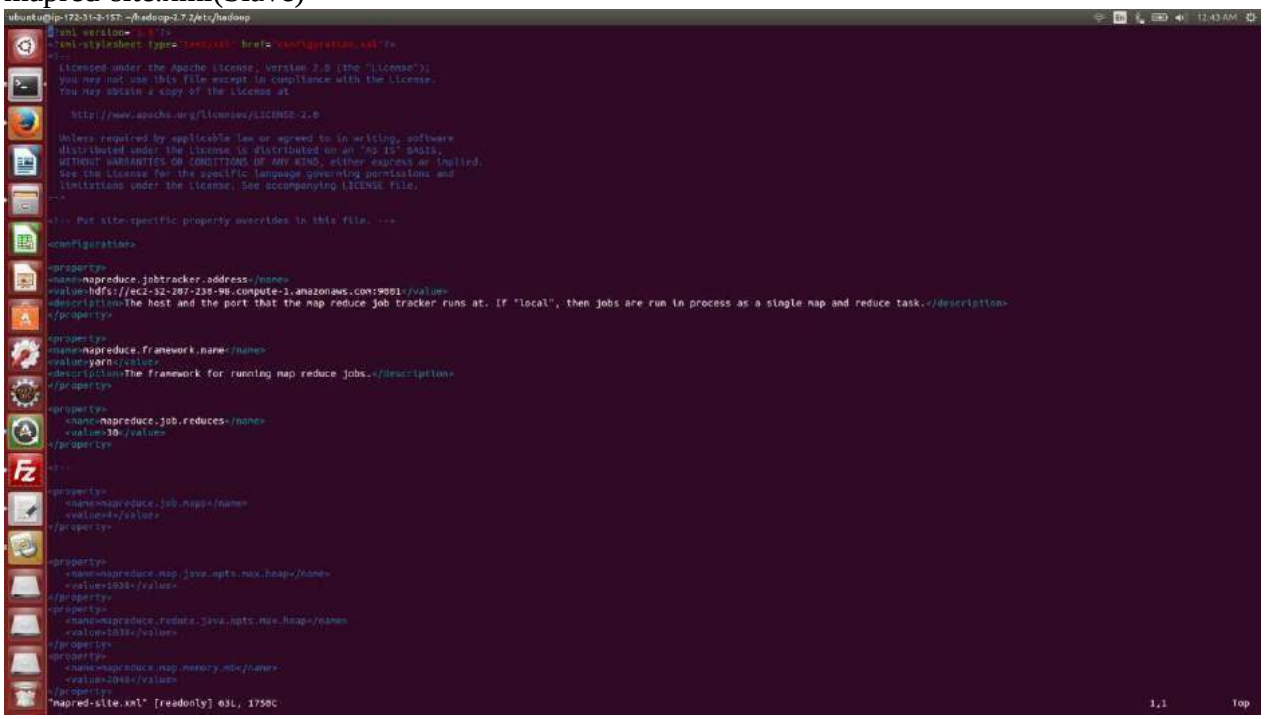
3. mapred-site.xml(master)



- mapred-site.xml for Master node will contain the jobtracker address and all the cluster temp address. Another important property that I defined in the mapred-site.xml is the number of reducers that the Hadoop map-reduce sort program will use while generating the output file.

- I have set the number of reducers to be 30 in order to tune mapreduce parallelism and even if the node fails the reduces can still be executed in the single wave.
  Note that this property is ignored when the job tracker address is local.

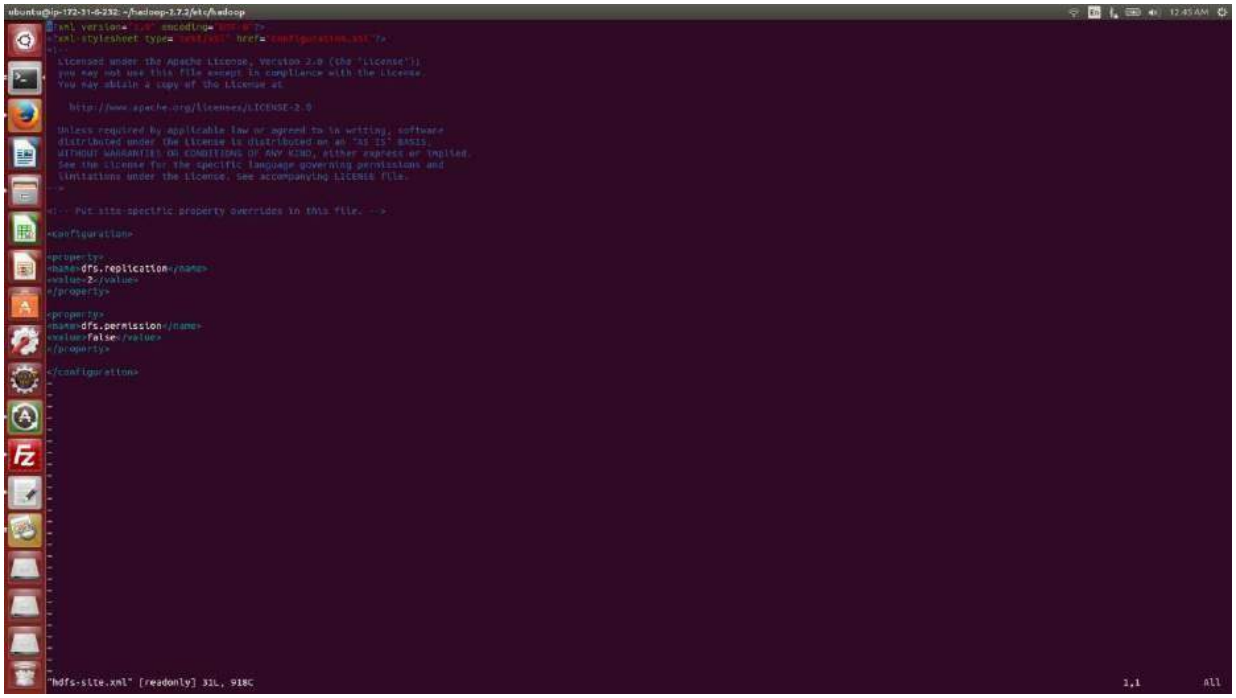
mapred-site.xml(Slave)



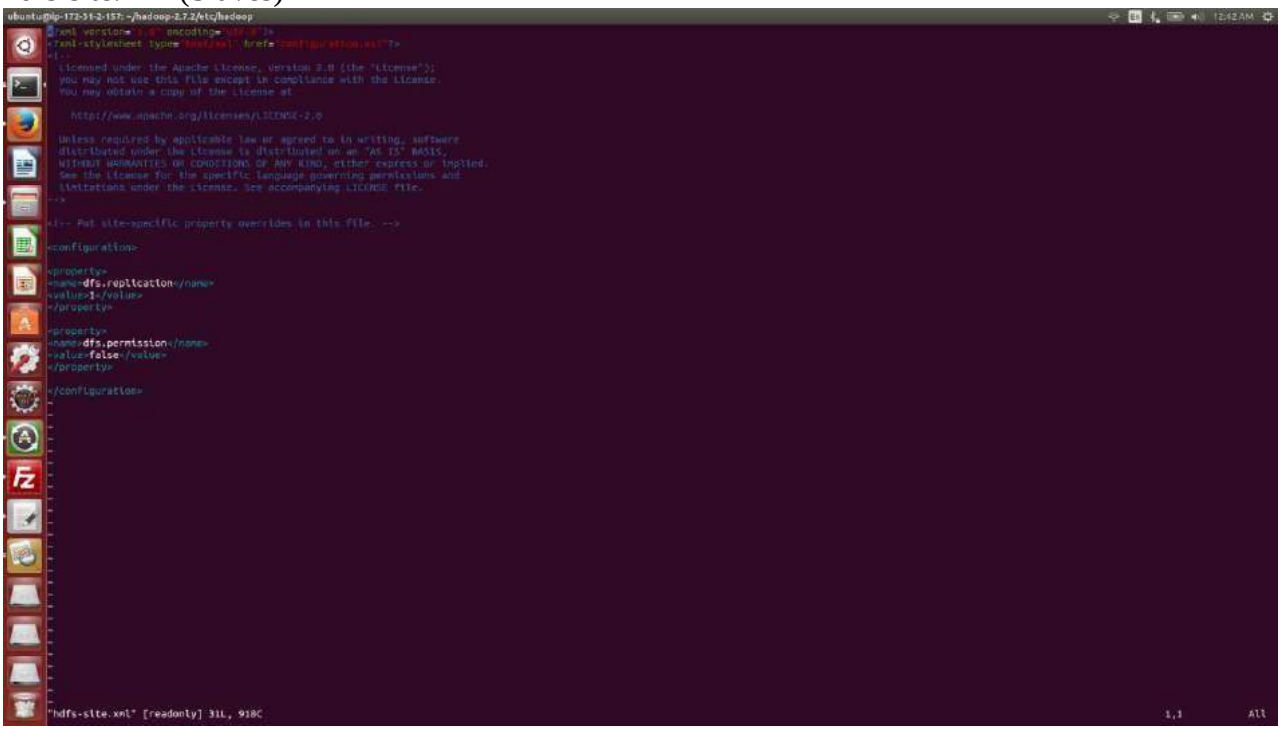
For Slaves, the mapred-site.xml will have all the default properties of the Hadoop setup done for single node virtual cluster.

The reducers defined in the master node are also defined in the slaves.

4. hdfs-site.xml(master)



hdfs-site.xml (Slaves)

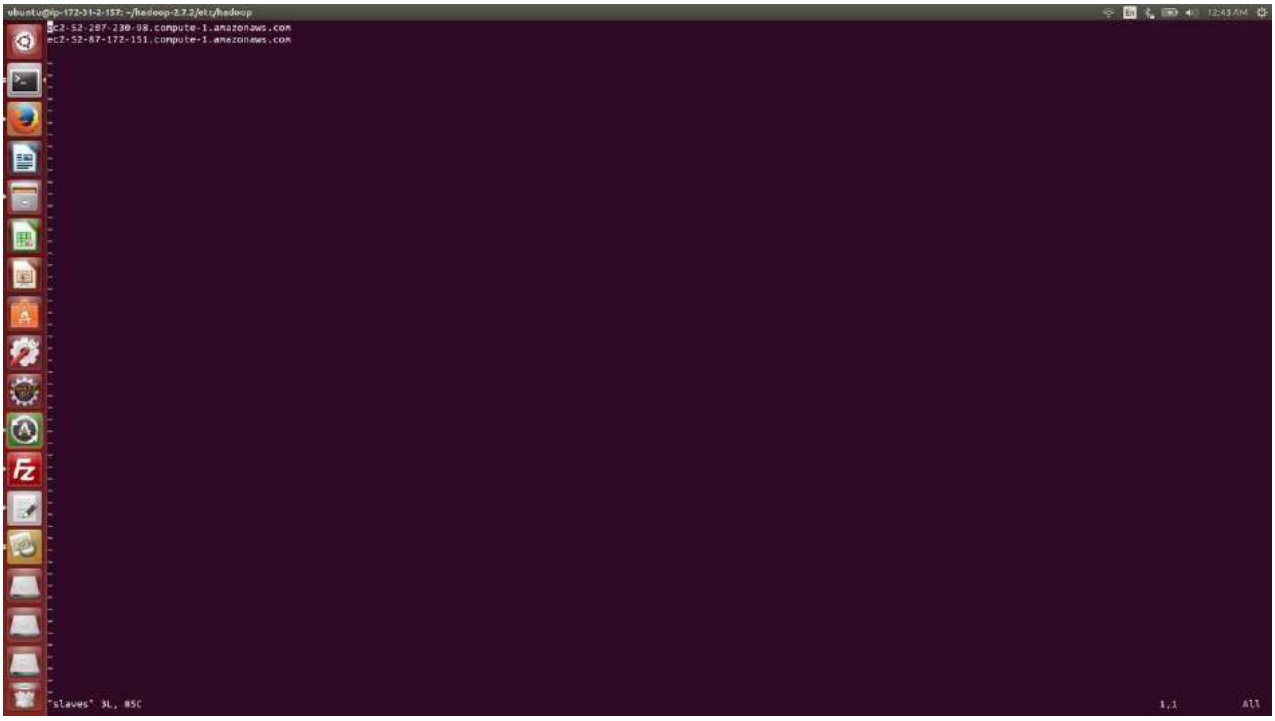


The hdfs-site.xml for master and slaves have default property of replication factor 1

5. slaves(Master)

The Slaves file of Master node will have all the Public DNS of the Slaves .

slaves(Slave)



The Slaves file of Slave node will have its own and master node's Public DNS address.

- Once the configuration is done on master and slave node, we need to generate a ssh private key that will authorize all the DNS of Master and Slaves.

  To access permission write

  eval `ssh-agent –s`
  ssh-add keypair.pem

  If its restricted, then

  chmod 400 keypair.pem and then above type instructions

  To authorize all the DNS we use the below command,

  ssh-keygen -t rsa

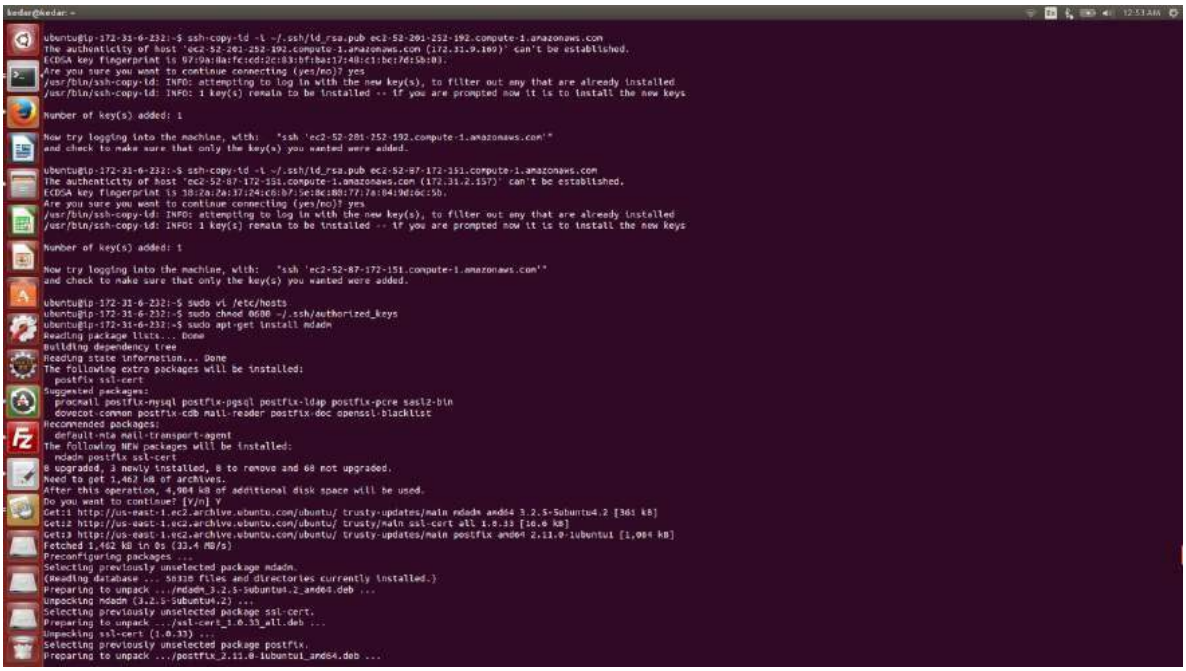
  ssh-copy-id -i ~/.ssh/id_rsa.pub PUBLIC_DNS

  chmod 0600 ~/.ssh/authorized_keys

  Below screenshot shows the output of the above scripts run on the terminal .

```
kader@kader: ~
/usr/bin/ssh-copy-id: ERROR: ssh: connect to host ec2-52-207-230-98.compute-1.amazonaws.com port 22: No route to host

ubuntu@ip-172-31-6-232:~$ sudo vi /etc/hosts
ubuntu@ip-172-31-6-232:~$ eval `ssh-agent -s`
Agent pid 1391
ubuntu@ip-172-31-6-232:~$ ssh-add pa1_keypair.pem
Identity added: pa1_keypair.pem (pa1_keypair.pem)
ubuntu@ip-172-31-6-232:~$ ssh-keygen -t rsa
Generating public/private rsa key pair.
Enter file in which to save the key (/home/ubuntu/.ssh/id_rsa):
/home/ubuntu/.ssh/id_rsa already exists.
Overwrite (y/n)? y
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /home/ubuntu/.ssh/id_rsa.
Your public key has been saved in /home/ubuntu/.ssh/id_rsa.pub.
The key fingerprint is:
9e:31:d5:76:8e:71:05:c6:ad:20:ae:86:05:05:b4:62 ubuntu@ip-172-31-6-232
The key's randomart image is:
+---[ RSA 2048]----+
|E  .o.    o.=+.  |
| .   .. . .X. .  |
|  . .o . .o.o.   |
|   .. . .. .     |
|    o .S         |
|     . o. +      |
|       . o       |
|                 |
|                 |
+----------------+
ubuntu@ip-172-31-6-232:~$ ssh-copy-id -i ~/.ssh/id_rsa.pub ec2-52-207-230-98.compute-1.amazonaws.com
The authenticity of host 'ec2-52-207-230-98.compute-1.amazonaws.com (172.31.6.232)' can't be established.
ECDSA key fingerprint is 38:3f:36:90:89:a1:21:f0:61:6c:ec:0f:31:ee:98:d8.
Are you sure you want to continue connecting (yes/no)? yes
/usr/bin/ssh-copy-id: INFO: attempting to log in with the new key(s), to filter out any that are already installed
/usr/bin/ssh-copy-id: INFO: 1 key(s) remain to be installed -- if you are prompted now it is to install the new keys

Number of key(s) added: 1

Now try logging into the machine, with:   "ssh 'ec2-52-207-230-98.compute-1.amazonaws.com'"
and check to make sure that only the key(s) you wanted were added.

ubuntu@ip-172-31-6-232:~$ ssh-copy-id -i ~/.ssh/id_rsa.pub ec2-52-87-166-186.compute-1.amazonaws.com
The authenticity of host 'ec2-52-87-166-186.compute-1.amazonaws.com (172.31.3.237)' can't be established.
ECDSA key fingerprint is 3a:4c:64:97:1d:cb:e4:ef:b8:59:e1:5a:8f:ef:e8:85.
Are you sure you want to continue connecting (yes/no)? yes
/usr/bin/ssh-copy-id: INFO: attempting to log in with the new key(s), to filter out any that are already installed
/usr/bin/ssh-copy-id: INFO: 1 key(s) remain to be installed -- if you are prompted now it is to install the new keys

Number of key(s) added: 1

Now try logging into the machine, with:   "ssh 'ec2-52-87-166-186.compute-1.amazonaws.com'"
and check to make sure that only the key(s) you wanted were added.

ubuntu@ip-172-31-6-232:~$ ssh-copy-id -i ~/.ssh/id_rsa.pub ec2-54-165-197-6.compute-1.amazonaws.com
The authenticity of host 'ec2-54-165-197-6.compute-1.amazonaws.com (172.31.12.142)' can't be established.
ECDSA key fingerprint is 44:f6:64:47:d0:ae:ca:84:90:e5:24:a8:a3:a5:53:ea.
```

```
kader@kader: ~
+----------------+
ubuntu@ip-172-31-6-232:~$ ssh-copy-id -i ~/.ssh/id_rsa.pub ec2-52-207-230-98.compute-1.amazonaws.com
The authenticity of host 'ec2-52-207-230-98.compute-1.amazonaws.com (172.31.6.232)' can't be established.
ECDSA key fingerprint is 38:3f:36:90:89:a1:21:f0:61:6c:ec:0f:31:ee:98:d8.
Are you sure you want to continue connecting (yes/no)? yes
/usr/bin/ssh-copy-id: INFO: attempting to log in with the new key(s), to filter out any that are already installed
/usr/bin/ssh-copy-id: INFO: 1 key(s) remain to be installed -- if you are prompted now it is to install the new keys

Number of key(s) added: 1

Now try logging into the machine, with:   "ssh 'ec2-52-207-230-98.compute-1.amazonaws.com'"
and check to make sure that only the key(s) you wanted were added.

ubuntu@ip-172-31-6-232:~$ ssh-copy-id -i ~/.ssh/id_rsa.pub ec2-52-87-166-186.compute-1.amazonaws.com
The authenticity of host 'ec2-52-87-166-186.compute-1.amazonaws.com (172.31.3.237)' can't be established.
ECDSA key fingerprint is 3a:4c:64:97:1d:cb:e4:ef:b8:59:e1:5a:8f:ef:e8:85.
Are you sure you want to continue connecting (yes/no)? yes
/usr/bin/ssh-copy-id: INFO: attempting to log in with the new key(s), to filter out any that are already installed
/usr/bin/ssh-copy-id: INFO: 1 key(s) remain to be installed -- if you are prompted now it is to install the new keys

Number of key(s) added: 1

Now try logging into the machine, with:   "ssh 'ec2-52-87-166-186.compute-1.amazonaws.com'"
and check to make sure that only the key(s) you wanted were added.

ubuntu@ip-172-31-6-232:~$ ssh-copy-id -i ~/.ssh/id_rsa.pub ec2-54-165-197-6.compute-1.amazonaws.com
The authenticity of host 'ec2-54-165-197-6.compute-1.amazonaws.com (172.31.12.142)' can't be established.
ECDSA key fingerprint is 44:f6:64:47:d0:ae:ca:84:90:e5:24:a8:a3:a5:53:ea.
Are you sure you want to continue connecting (yes/no)? yes
/usr/bin/ssh-copy-id: INFO: attempting to log in with the new key(s), to filter out any that are already installed
/usr/bin/ssh-copy-id: INFO: 1 key(s) remain to be installed -- if you are prompted now it is to install the new keys

Number of key(s) added: 1

Now try logging into the machine, with:   "ssh 'ec2-54-165-197-6.compute-1.amazonaws.com'"
and check to make sure that only the key(s) you wanted were added.

ubuntu@ip-172-31-6-232:~$ ssh-copy-id -i ~/.ssh/id_rsa.pub ec2-52-87-250-247.compute-1.amazonaws.com
The authenticity of host 'ec2-52-87-250-247.compute-1.amazonaws.com (172.31.15.170)' can't be established.
ECDSA key fingerprint is db:89:2b:c2:df:c8:d3:c4:d4:be:b1:ab:5d:75:ff:7c.
Are you sure you want to continue connecting (yes/no)? yes
/usr/bin/ssh-copy-id: INFO: attempting to log in with the new key(s), to filter out any that are already installed
/usr/bin/ssh-copy-id: INFO: 1 key(s) remain to be installed -- if you are prompted now it is to install the new keys

Number of key(s) added: 1

Now try logging into the machine, with:   "ssh 'ec2-52-87-250-247.compute-1.amazonaws.com'"
and check to make sure that only the key(s) you wanted were added.

ubuntu@ip-172-31-6-232:~$ ssh-copy-id -i ~/.ssh/id_rsa.pub ec2-52-91-148-104.compute-1.amazonaws.com
The authenticity of host 'ec2-52-91-148-104.compute-1.amazonaws.com (172.31.15.113)' can't be established.
ECDSA key fingerprint is b4:78:23:c3:94:15:e4:1e:f8:4e:b5:ab:5a:7d:8c:09.
Are you sure you want to continue connecting (yes/no)? yes
/usr/bin/ssh-copy-id: INFO: attempting to log in with the new key(s), to filter out any that are already installed
/usr/bin/ssh-copy-id: INFO: 1 key(s) remain to be installed -- if you are prompted now it is to install the new keys

Number of key(s) added: 1
```



Once all the keys are authorized, we mount a raid array of the available volume from d3.xlarge instance using the mount steps mentioned in the single node virtual cluster installation steps.

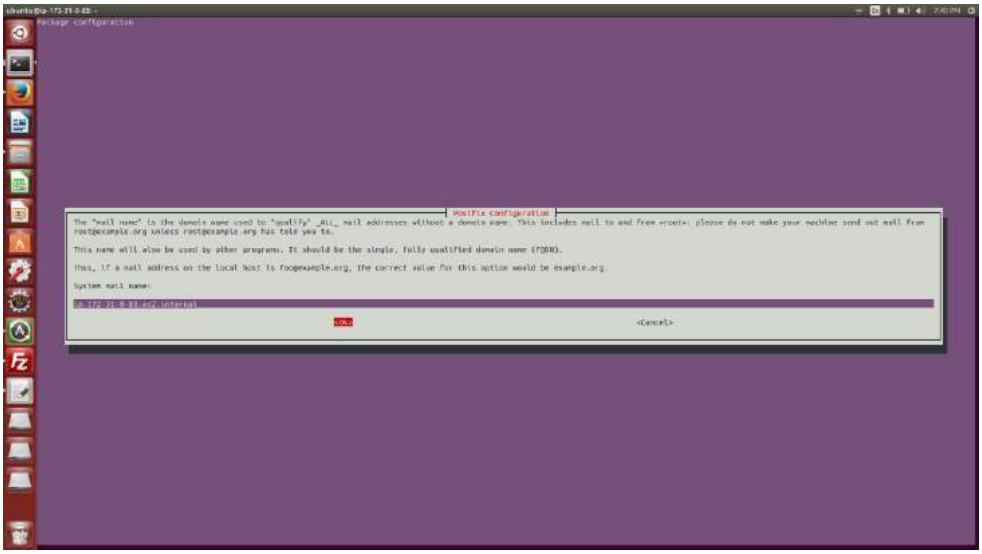
Hadoop TeraSort 16 Node (100GB)

- The d3.xlarge instance contains and available volume of 3*2048 TB size of an EBS volume which we will mount for our Hadoop to use it to store tmp files while running map-reduce jobs.

Mentioned below are the steps to mount an EBS volume to our system.

Sudo yum install mdadm

We use the **mdadm** command to create a logical RAID device from the newly attached Amazon EBS volumes.



- We can list the devices on our instance using **lsblk** command

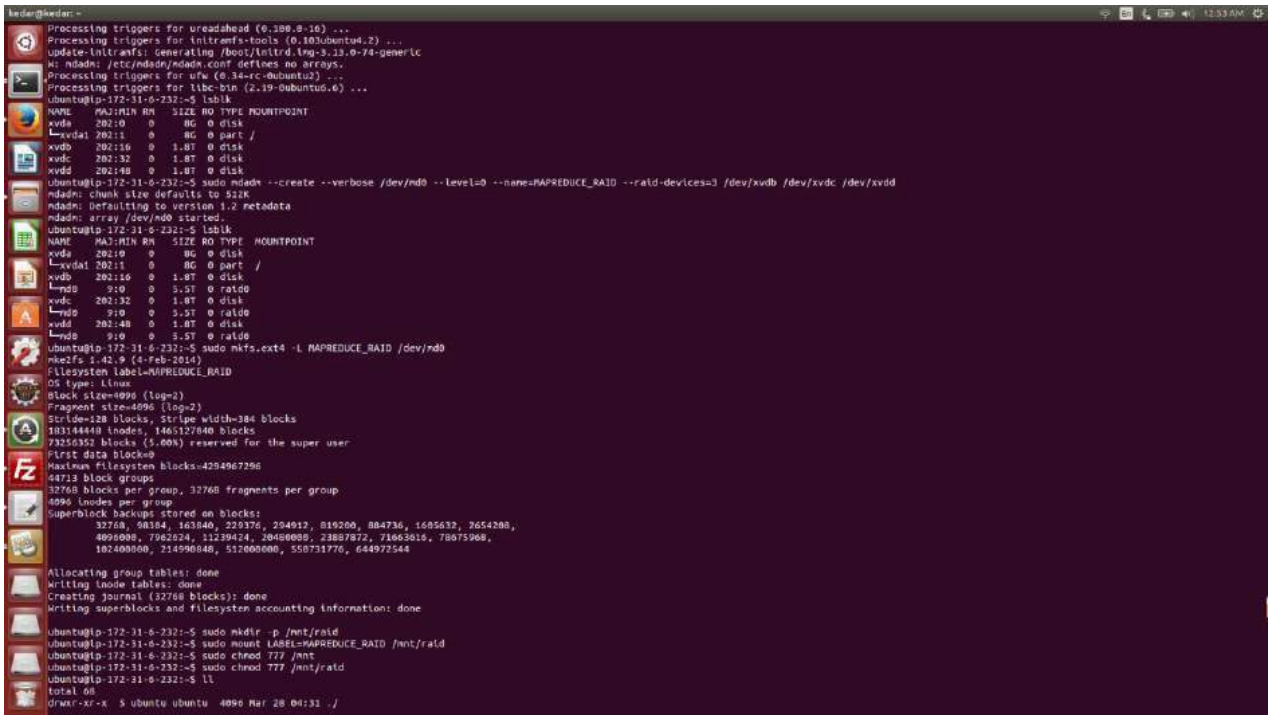
- To create a RAID 0 array, execute the following command
  sudo mdadm --create --verbose /dev/md0 --level=0 --name=RAID_NAME --raid-devices=No_Of_Devices Device1 Device2

- Create a file system on your RAID array, and give that file system a label to use when you mount it later.
  sudo mkfs.ext4 -L RAID_NAME /dev/md0

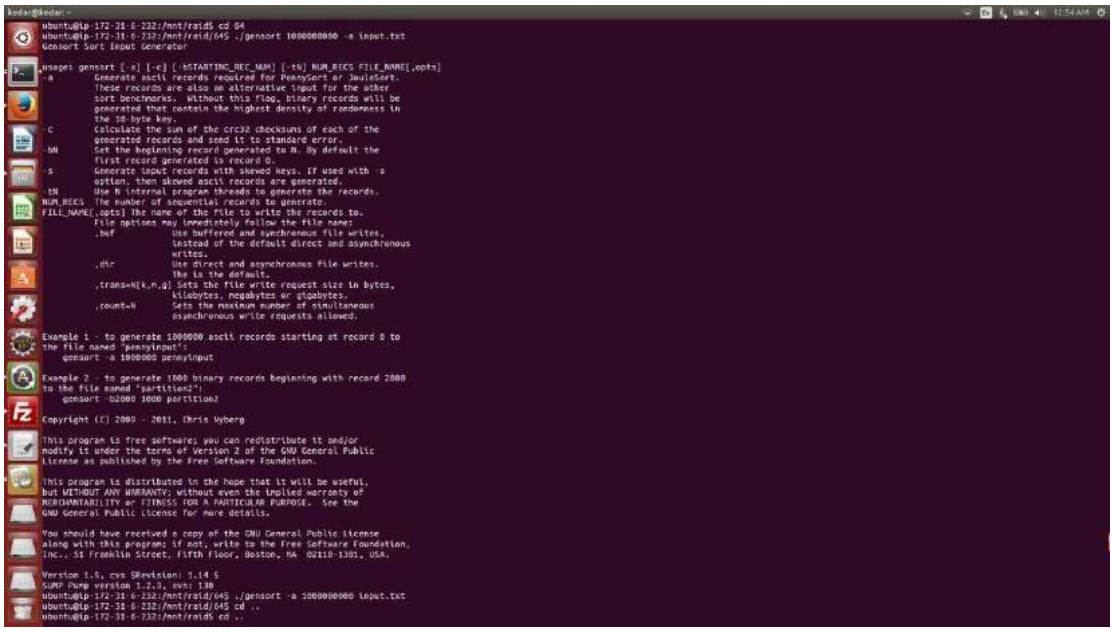
- Create a mount point for your RAID array.
  sudo mkdir -p /mnt/raid

- Finally, mount the RAID device on the mount point that you created:
  sudo mount LABEL=MAPREDUCE_RAID /mnt/raid

The mentioned steps for mounting an EBS volume are shown in the below screenshot



- We will be running an hadoop map-reduce algorithm of sorting 100GB of data .
  To generate 100 GB of data , we used gensort and stored all the data in the mounted raid disk .



- Once this data is generated, we will format the namenode as done previously for the single node cluster.

  Once the node is formatted, The Hadoop clusters are ready and can be started.

  Start the Hadoop cluster using

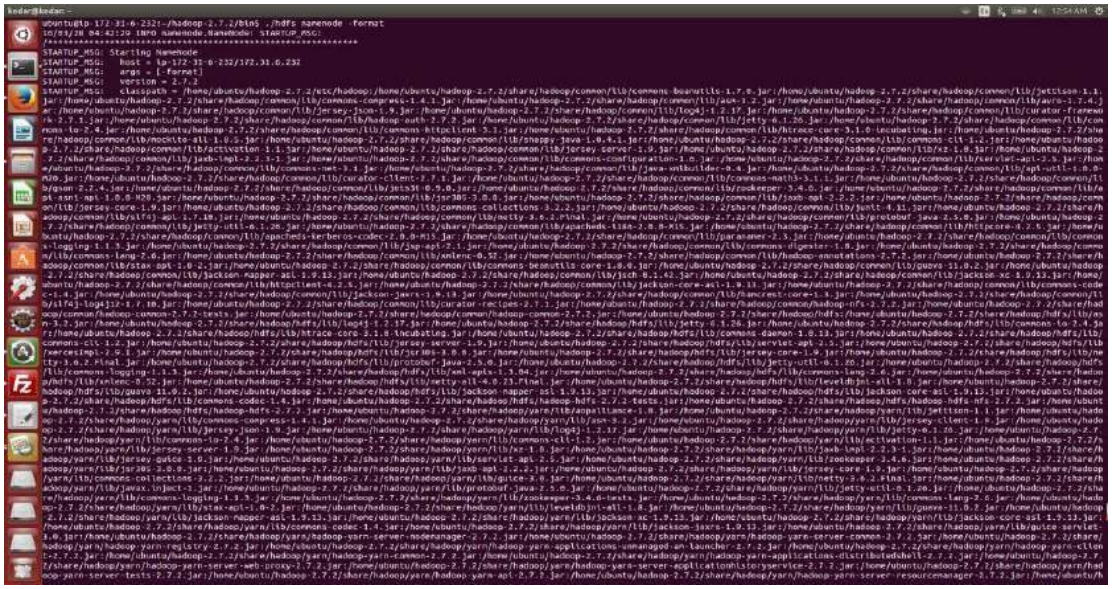
  ./start-dfs.sh

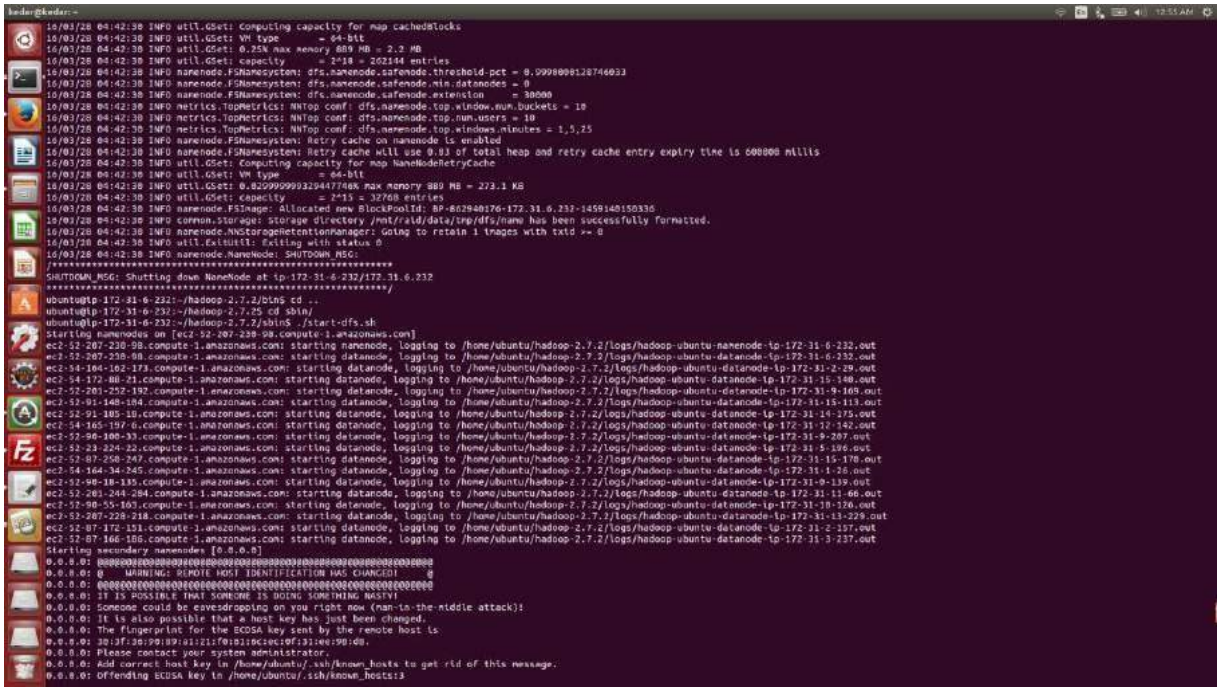
  This command will start all the Hadoop namenodes, datanodes, jobtrackers and tasktrackers

  ./start-yarn.sh

  This command will start the resourcemanagers and the node managers.

  You can type in **jps** command to check whether the above nodes have been started.





- The Data Nodes, Name Nodes will start up on all the Virtual Clusters defined in the localhost of the master nodes.

- Create an Input folder in the HDFS from where the Hadoop Map-Reduce job will gather the input.

  To create an input directory in the HDFS, we use the below commands,

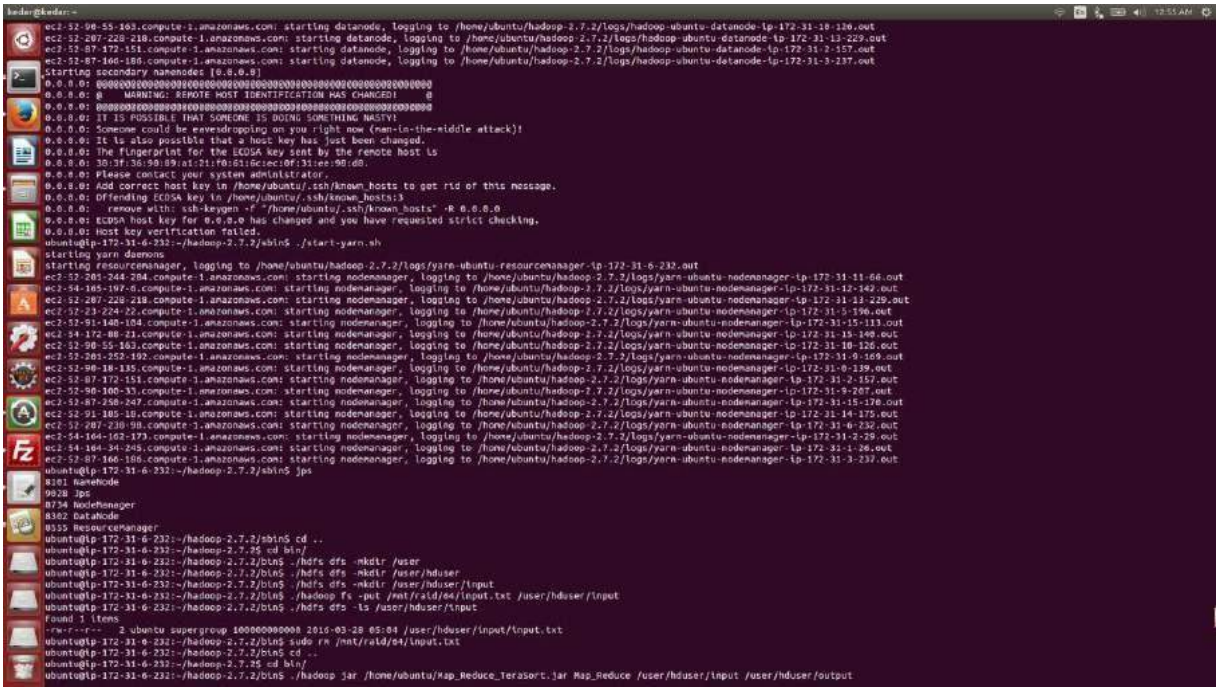
  ./hdfs dfs -mkdir /user

  ./hdfs dfs -mkdir /user/hduser

  ./hdfs dfs -mkdir /user/hduser/input

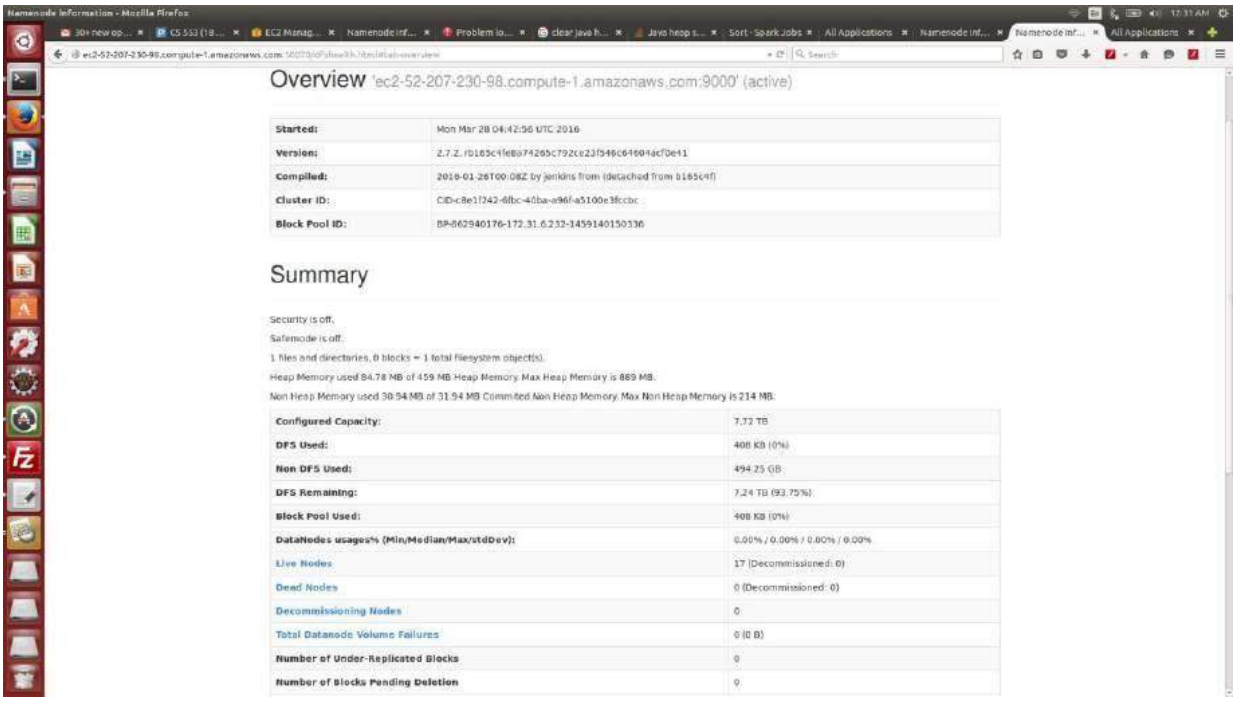
  Once the input directory is created, we transfer the 100GB of input stored in Mounted Disk to the HDFS using below command,

  ./hadoop fs -put /mnt/raid/64/input.txt /user/hduser/input

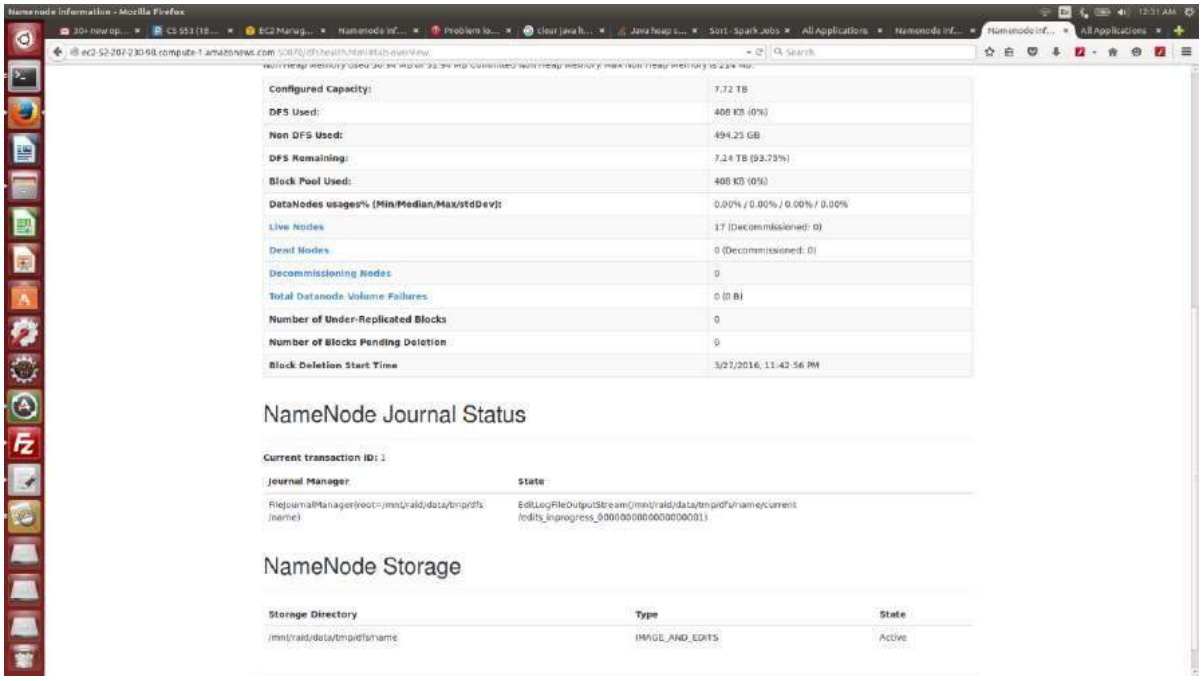


We can check if all the nodes are active and Hadoop setup is correctly configured on all the nodes.

This can be viewed using the port number 50070 of Master Node PublicDNS



As per the above screenshot, 17 Live nodes are running. This contains 1 Master and 16 Slave Nodes.



Once confirmed we can start our map-reduce jar file that we used for sorting on single node virtual cluster.

./hadoop jar /home/ubuntu/Map_Reduce_TeraSort.jar Map_Reduce /user/hduser/input /user/hduser/output

Map_Reduce_TeraSort is the name of the jar file that contains the map-reduce algorithm implemented.

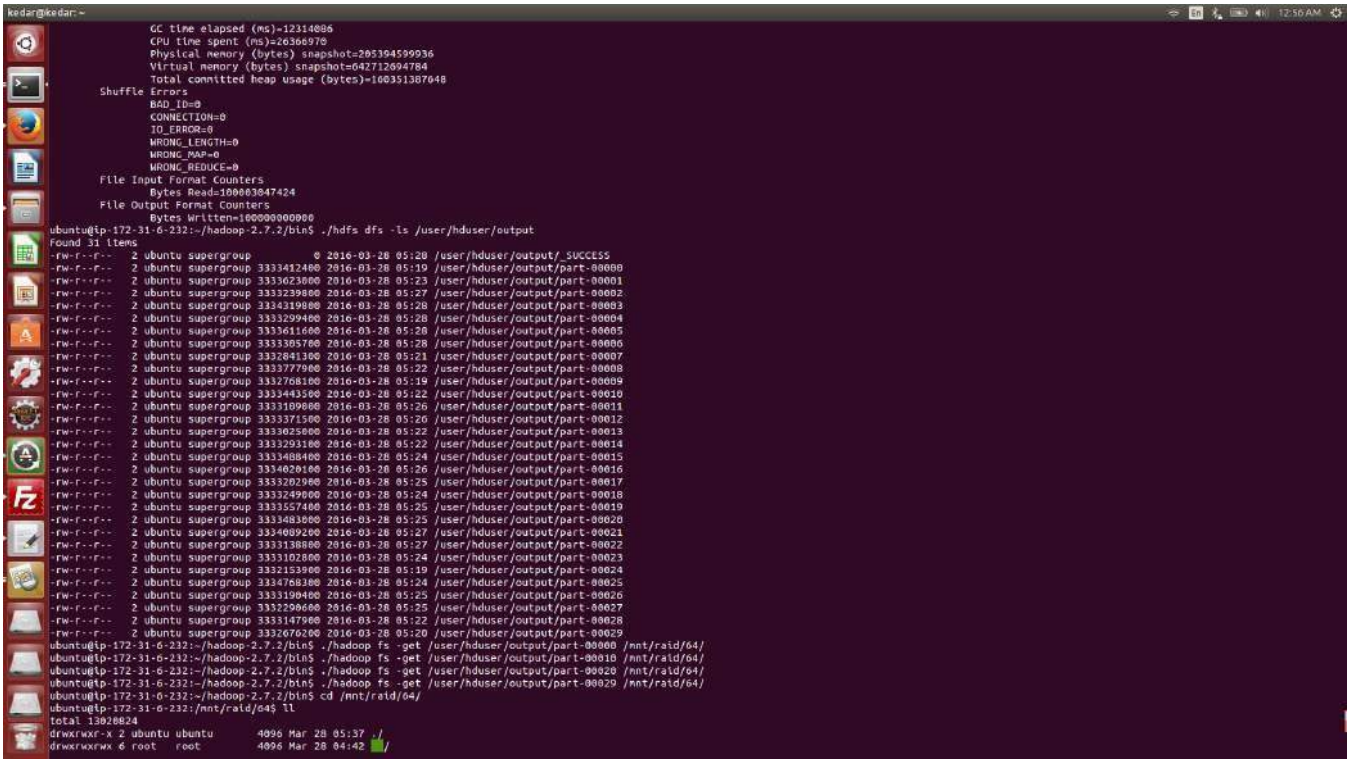
Map_Reduce is the name of the class

/user/hduser/input – Input directory of the text file

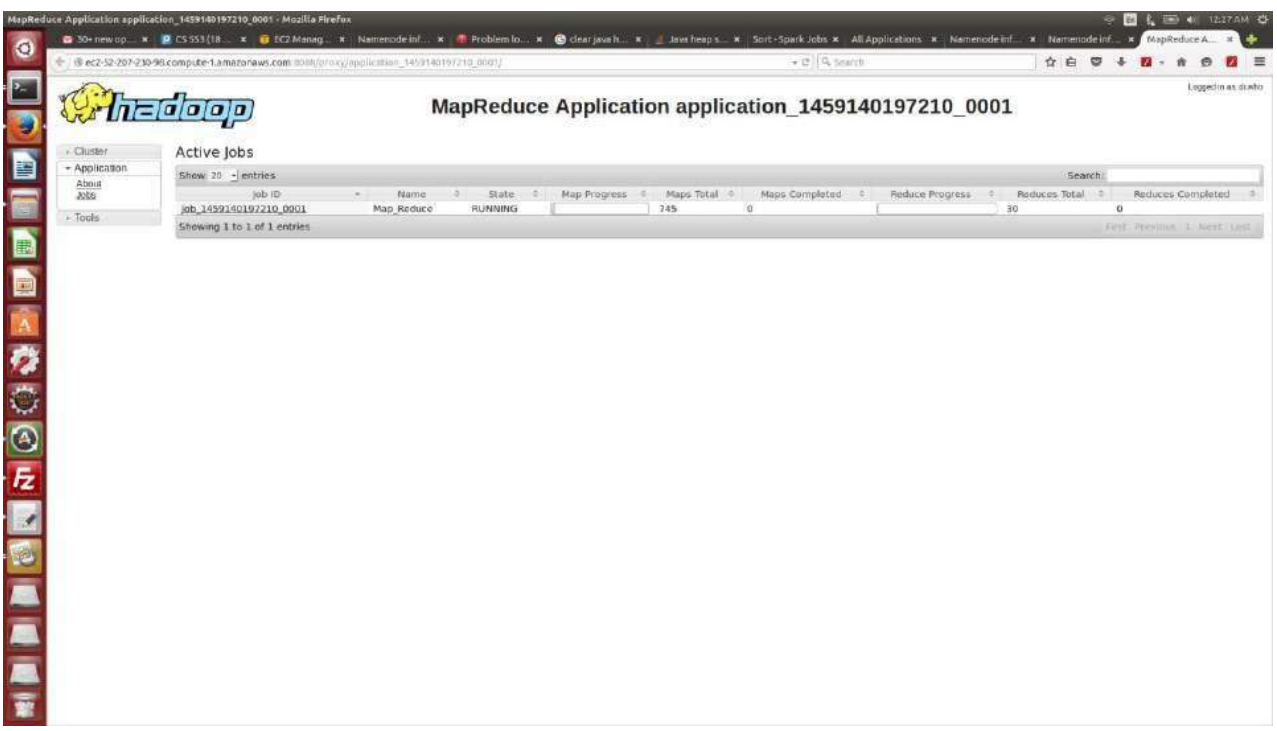
/user/hduser/output – Output Directory for the sorted file.

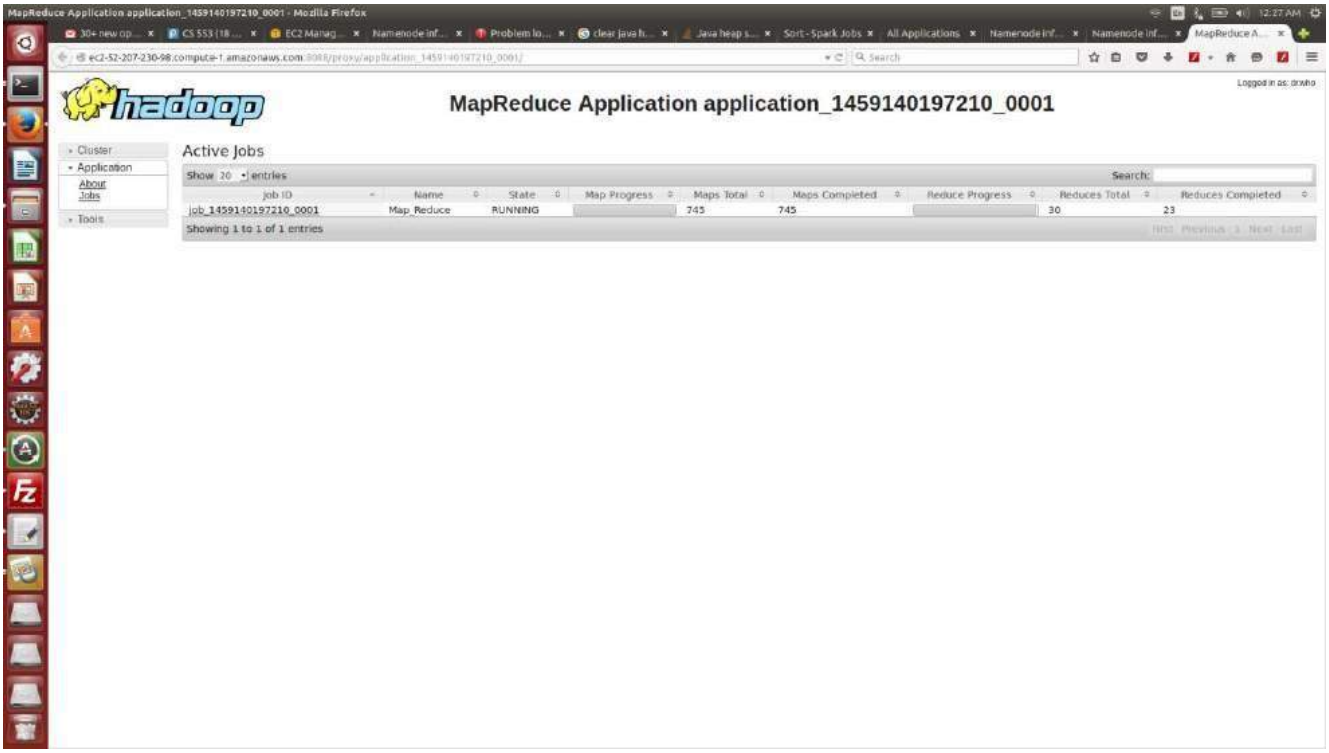
```
kedar@kedar: ~
ubuntu@ip-172-31-6-232:~/hadoop-2.7.2/bin$ sudo rm /mnt/raid/64/input.txt
ubuntu@ip-172-31-6-232:~/hadoop-2.7.2/bin$ cd ..
ubuntu@ip-172-31-6-232:~/hadoop-2.7.2$ cd bin/
ubuntu@ip-172-31-6-232:~/hadoop-2.7.2/bin$ ./hadoop jar /home/ubuntu/Map_Reduce_TeraSort.jar Map_Reduce /user/hduser/input /user/hduser/output
16/03/28 05:06:12 INFO client.RMProxy: Connecting to ResourceManager at ec2-52-207-238-98.compute-1.amazonaws.com/172.31.6.232:8032
16/03/28 05:06:12 INFO client.RMProxy: Connecting to ResourceManager at ec2-52-207-238-98.compute-1.amazonaws.com/172.31.6.232:8032
16/03/28 05:06:12 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 05:06:13 INFO mapred.FileInputFormat: Total input paths to process : 1
16/03/28 05:06:13 INFO net.NetworkTopology: Adding a new node: /default-rack/172.31.6.232:50010
16/03/28 05:06:13 INFO net.NetworkTopology: Adding a new node: /default-rack/172.31.2.29:50010
16/03/28 05:06:13 INFO mapreduce.JobSubmitter: number of splits:745
16/03/28 05:06:13 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1459140197210_0001
16/03/28 05:06:14 INFO impl.YarnClientImpl: Submitted application application_1459140197210_0001
16/03/28 05:06:14 INFO mapreduce.Job: The url to track the job: http://ec2-52-207-238-98.compute-1.amazonaws.com:8088/proxy/application_1459140197210_0001/
16/03/28 05:06:14 INFO mapreduce.Job: Running job: job_1459140197210_0001
16/03/28 05:06:22 INFO mapreduce.Job: Job job_1459140197210_0001 running in uber mode : false
16/03/28 05:06:22 INFO mapreduce.Job:  map 0% reduce 0%
16/03/28 05:06:51 INFO mapreduce.Job:  map 1% reduce 0%
16/03/28 05:07:01 INFO mapreduce.Job:  map 2% reduce 0%
16/03/28 05:07:02 INFO mapreduce.Job:  map 3% reduce 0%
16/03/28 05:07:05 INFO mapreduce.Job:  map 4% reduce 0%
16/03/28 05:07:07 INFO mapreduce.Job:  map 5% reduce 0%
16/03/28 05:07:09 INFO mapreduce.Job:  map 6% reduce 0%
16/03/28 05:07:10 INFO mapreduce.Job:  map 7% reduce 0%
16/03/28 05:07:13 INFO mapreduce.Job:  map 8% reduce 0%
16/03/28 05:07:16 INFO mapreduce.Job:  map 9% reduce 0%
16/03/28 05:07:33 INFO mapreduce.Job:  map 10% reduce 0%
16/03/28 05:07:35 INFO mapreduce.Job:  map 11% reduce 0%
16/03/28 05:07:38 INFO mapreduce.Job:  map 12% reduce 0%
16/03/28 05:07:44 INFO mapreduce.Job:  map 13% reduce 0%
16/03/28 05:07:54 INFO mapreduce.Job:  map 14% reduce 0%
16/03/28 05:07:58 INFO mapreduce.Job:  map 15% reduce 0%
16/03/28 05:08:03 INFO mapreduce.Job:  map 16% reduce 0%
16/03/28 05:08:06 INFO mapreduce.Job:  map 17% reduce 0%
16/03/28 05:08:10 INFO mapreduce.Job:  map 18% reduce 0%
16/03/28 05:08:19 INFO mapreduce.Job:  map 19% reduce 0%
16/03/28 05:08:30 INFO mapreduce.Job:  map 20% reduce 0%
16/03/28 05:08:42 INFO mapreduce.Job:  map 21% reduce 0%
16/03/28 05:08:45 INFO mapreduce.Job:  map 21% reduce 1%
16/03/28 05:08:46 INFO mapreduce.Job:  map 22% reduce 1%
16/03/28 05:08:49 INFO mapreduce.Job:  map 23% reduce 1%
16/03/28 05:08:52 INFO mapreduce.Job:  map 24% reduce 1%
16/03/28 05:08:56 INFO mapreduce.Job:  map 25% reduce 1%
16/03/28 05:09:03 INFO mapreduce.Job:  map 26% reduce 1%
16/03/28 05:09:13 INFO mapreduce.Job:  map 27% reduce 1%
16/03/28 05:09:16 INFO mapreduce.Job:  map 27% reduce 2%
16/03/28 05:09:18 INFO mapreduce.Job:  map 28% reduce 2%
16/03/28 05:09:23 INFO mapreduce.Job:  map 29% reduce 2%
16/03/28 05:09:31 INFO mapreduce.Job:  map 30% reduce 2%
16/03/28 05:09:36 INFO mapreduce.Job:  map 31% reduce 2%
16/03/28 05:09:42 INFO mapreduce.Job:  map 32% reduce 2%
16/03/28 05:09:46 INFO mapreduce.Job:  map 33% reduce 2%
16/03/28 05:09:48 INFO mapreduce.Job:  map 33% reduce 3%
16/03/28 05:09:52 INFO mapreduce.Job:  map 34% reduce 3%
16/03/28 05:10:00 INFO mapreduce.Job:  map 35% reduce 3%
16/03/28 05:10:09 INFO mapreduce.Job:  map 35% reduce 4%
16/03/28 05:10:20 INFO mapreduce.Job:  map 36% reduce 4%
16/03/28 05:10:25 INFO mapreduce.Job:  map 37% reduce 4%
```

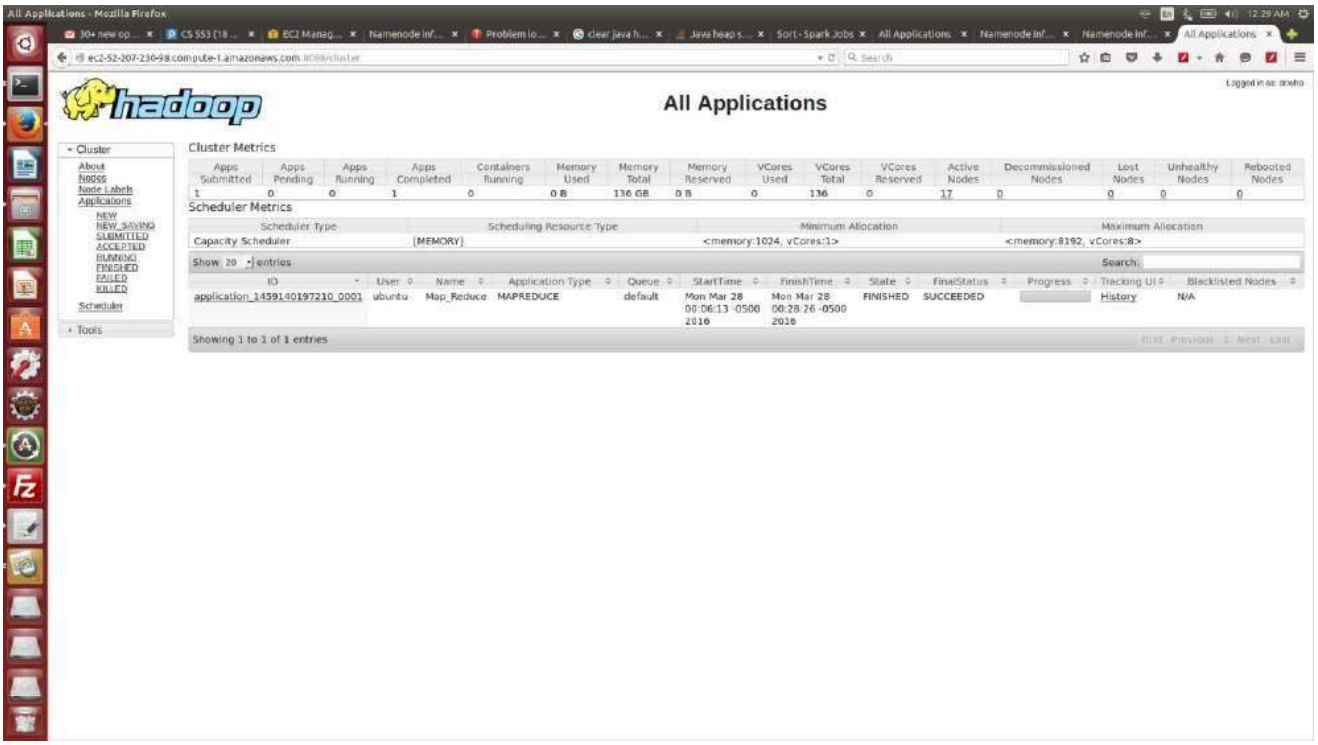
```
kedar@kedar: ~
16/03/28 05:20:42 INFO mapreduce.Job:  map 100% reduce 49%
16/03/28 05:20:48 INFO mapreduce.Job:  map 100% reduce 50%
16/03/28 05:20:51 INFO mapreduce.Job:  map 100% reduce 51%
16/03/28 05:20:54 INFO mapreduce.Job:  map 100% reduce 52%
16/03/28 05:21:01 INFO mapreduce.Job:  map 100% reduce 53%
16/03/28 05:21:04 INFO mapreduce.Job:  map 100% reduce 54%
16/03/28 05:21:09 INFO mapreduce.Job:  map 100% reduce 55%
16/03/28 05:21:10 INFO mapreduce.Job:  map 100% reduce 56%
16/03/28 05:21:13 INFO mapreduce.Job:  map 100% reduce 57%
16/03/28 05:21:15 INFO mapreduce.Job:  map 100% reduce 58%
16/03/28 05:21:18 INFO mapreduce.Job:  map 100% reduce 59%
16/03/28 05:21:24 INFO mapreduce.Job:  map 100% reduce 60%
16/03/28 05:21:36 INFO mapreduce.Job:  map 100% reduce 61%
16/03/28 05:21:43 INFO mapreduce.Job:  map 100% reduce 62%
16/03/28 05:21:49 INFO mapreduce.Job:  map 100% reduce 63%
16/03/28 05:21:57 INFO mapreduce.Job:  map 100% reduce 64%
16/03/28 05:22:06 INFO mapreduce.Job:  map 100% reduce 65%
16/03/28 05:22:14 INFO mapreduce.Job:  map 100% reduce 66%
16/03/28 05:22:20 INFO mapreduce.Job:  map 100% reduce 67%
16/03/28 05:22:24 INFO mapreduce.Job:  map 100% reduce 68%
16/03/28 05:22:29 INFO mapreduce.Job:  map 100% reduce 69%
16/03/28 05:22:32 INFO mapreduce.Job:  map 100% reduce 70%
16/03/28 05:22:34 INFO mapreduce.Job:  map 100% reduce 71%
16/03/28 05:22:37 INFO mapreduce.Job:  map 100% reduce 72%
16/03/28 05:22:40 INFO mapreduce.Job:  map 100% reduce 73%
16/03/28 05:22:42 INFO mapreduce.Job:  map 100% reduce 74%
16/03/28 05:22:46 INFO mapreduce.Job:  map 100% reduce 75%
16/03/28 05:22:52 INFO mapreduce.Job:  map 100% reduce 76%
16/03/28 05:23:00 INFO mapreduce.Job:  map 100% reduce 77%
16/03/28 05:23:10 INFO mapreduce.Job:  map 100% reduce 78%
16/03/28 05:23:22 INFO mapreduce.Job:  map 100% reduce 79%
16/03/28 05:23:35 INFO mapreduce.Job:  map 100% reduce 80%
16/03/28 05:23:40 INFO mapreduce.Job:  map 100% reduce 81%
16/03/28 05:23:47 INFO mapreduce.Job:  map 100% reduce 82%
16/03/28 05:24:01 INFO mapreduce.Job:  map 100% reduce 83%
16/03/28 05:24:10 INFO mapreduce.Job:  map 100% reduce 84%
16/03/28 05:24:14 INFO mapreduce.Job:  map 100% reduce 85%
16/03/28 05:24:19 INFO mapreduce.Job:  map 100% reduce 86%
16/03/28 05:24:22 INFO mapreduce.Job:  map 100% reduce 87%
16/03/28 05:24:25 INFO mapreduce.Job:  map 100% reduce 88%
16/03/28 05:24:29 INFO mapreduce.Job:  map 100% reduce 89%
16/03/28 05:24:36 INFO mapreduce.Job:  map 100% reduce 90%
16/03/28 05:24:48 INFO mapreduce.Job:  map 100% reduce 91%
16/03/28 05:25:00 INFO mapreduce.Job:  map 100% reduce 92%
16/03/28 05:25:14 INFO mapreduce.Job:  map 100% reduce 93%
16/03/28 05:25:31 INFO mapreduce.Job:  map 100% reduce 94%
16/03/28 05:25:46 INFO mapreduce.Job:  map 100% reduce 95%
16/03/28 05:26:05 INFO mapreduce.Job:  map 100% reduce 96%
16/03/28 05:26:27 INFO mapreduce.Job:  map 100% reduce 97%
16/03/28 05:26:54 INFO mapreduce.Job:  map 100% reduce 98%
16/03/28 05:27:25 INFO mapreduce.Job:  map 100% reduce 99%
16/03/28 05:27:55 INFO mapreduce.Job:  map 100% reduce 100%
16/03/28 05:28:27 INFO mapreduce.Job: Job job_1459140197210_0001 completed successfully
16/03/28 05:28:27 INFO mapreduce.Job: Counters: 52
        File System Counters
                FILE: Number of bytes read=209193669840
                FILE: Number of bytes written=391190780410
                FILE: Number of read operations=0
```

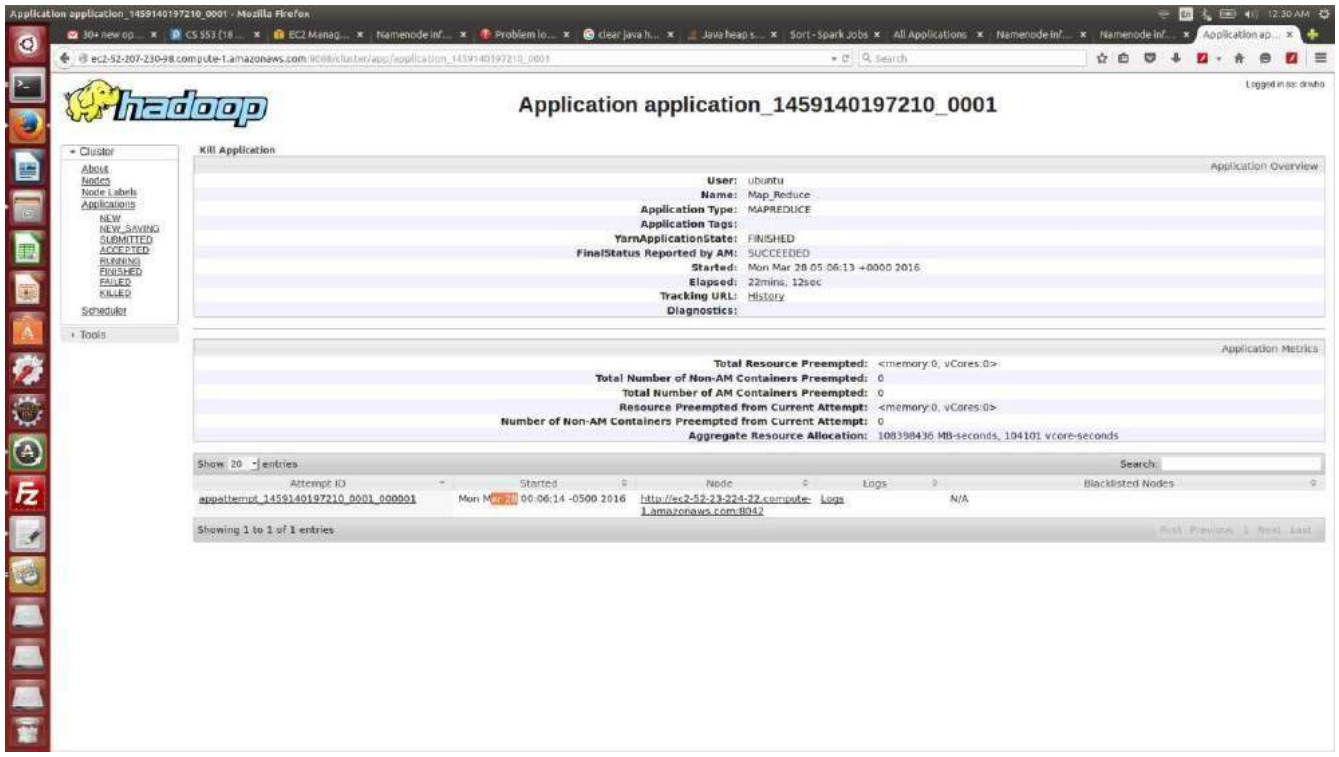


The Map-Reduce Job Run can be tracked and observed using the public DNS and port 8088 defined in the mapred-site file.




MapReduce Application application_1459140197210_0001 - Mozilla Firefox
ec2-52-207-230-98.compute-1.amazonaws.com:8088/proxy/application_1459140197210_0001/
hadoop
MapReduce Application application_1459140197210_0001
Cluster
Application
About
Jobs
Tools
Active Jobs
Show 20 entries
Search:
job ID | Name | State | Map Progress | Maps Total | Maps Completed | Reduce Progress | Reduces Total | Reduces Completed
job_1459140197210_0001 | Map_Reduce | RUNNING | 745 | 745 | 30 | 23
Showing 1 to 1 of 1 entries

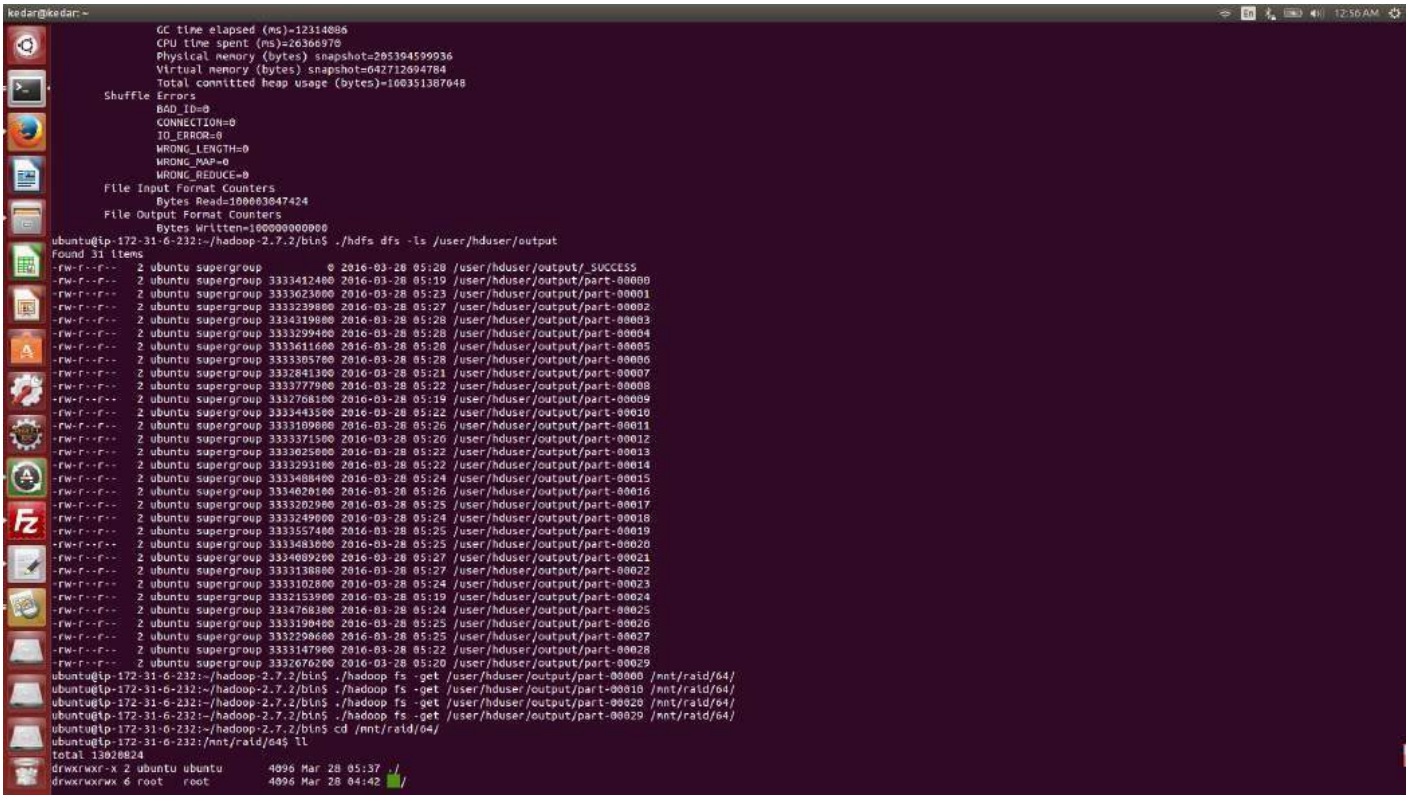

All Applications - Mozilla Firefox
ec2-52-207-230-98.compute-1.amazonaws.com:8088/cluster
hadoop
All Applications
Cluster
About
Nodes
Node Labels
Applications
NEW
NEW_SAVING
SUBMITTED
ACCEPTED
RUNNING
FINISHED
FAILED
KILLED
Scheduler
Tools
Cluster Metrics
Apps Submitted | Apps Pending | Apps Running | 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
1 | 0 | 0 | 1 | 0 | 0 B | 136 GB | 0 B | 0 | 136 | 0 | 17 | 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
Search:
ID | User | Name | Application Type | Queue | StartTime | FinishTime | State | FinalStatus | Progress | Tracking UI | Blacklisted Nodes
application_1459140197210_0001 | ubuntu | Map_Reduce | MAPREDUCE | default | Mon Mar 28 00:06:13 -0500 2016 | Mon Mar 28 00:28:26 -0500 2016 | FINISHED | SUCCEEDED | History | N/A
Showing 1 to 1 of 1 entries



<u>As mentioned in the screenshot, the total time required to execute map-reduce program to sort 100 GB of data on 16 node Virtual Cluster is 22min 12 sec.</u>

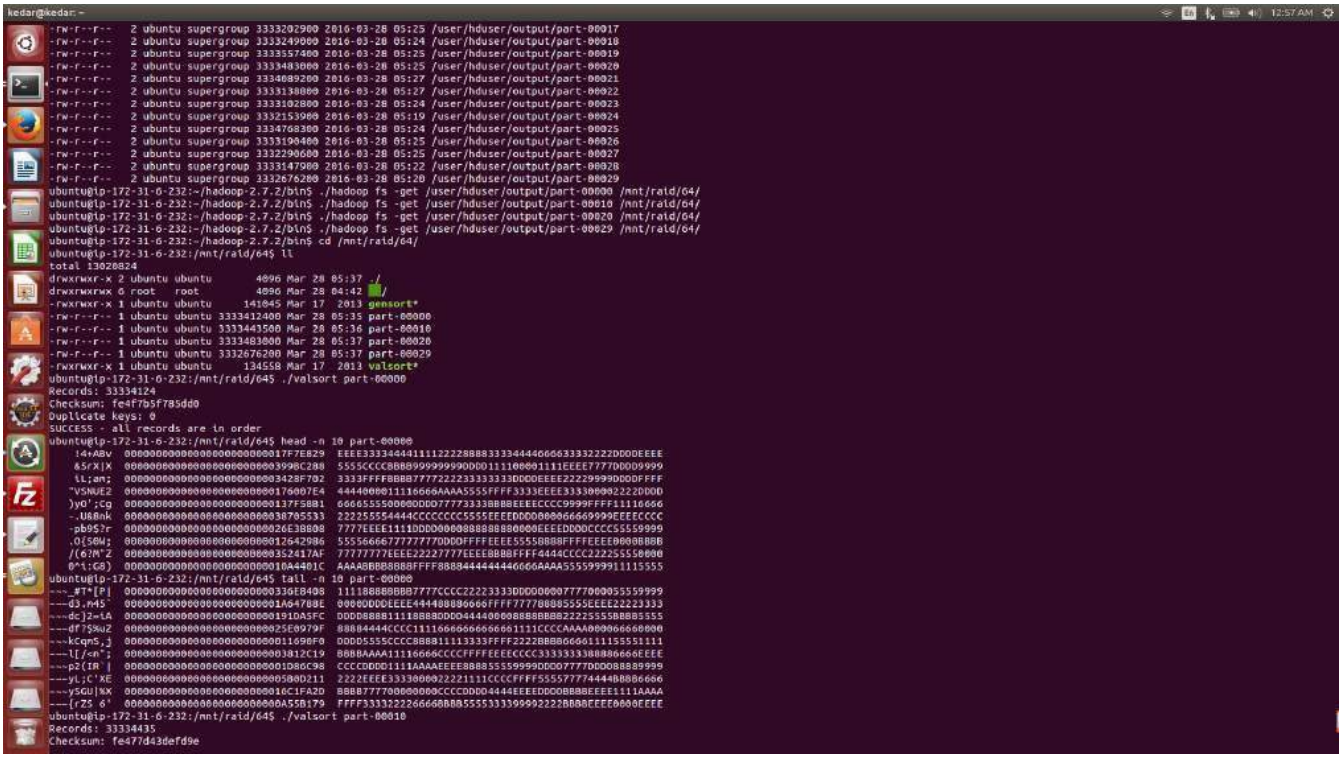
- As we have put up the number of reduces = 30, the output data will be generated with 30 parts.
- To evaluate the generated output , transfer the output parts back to the mount raid disk using the below command.

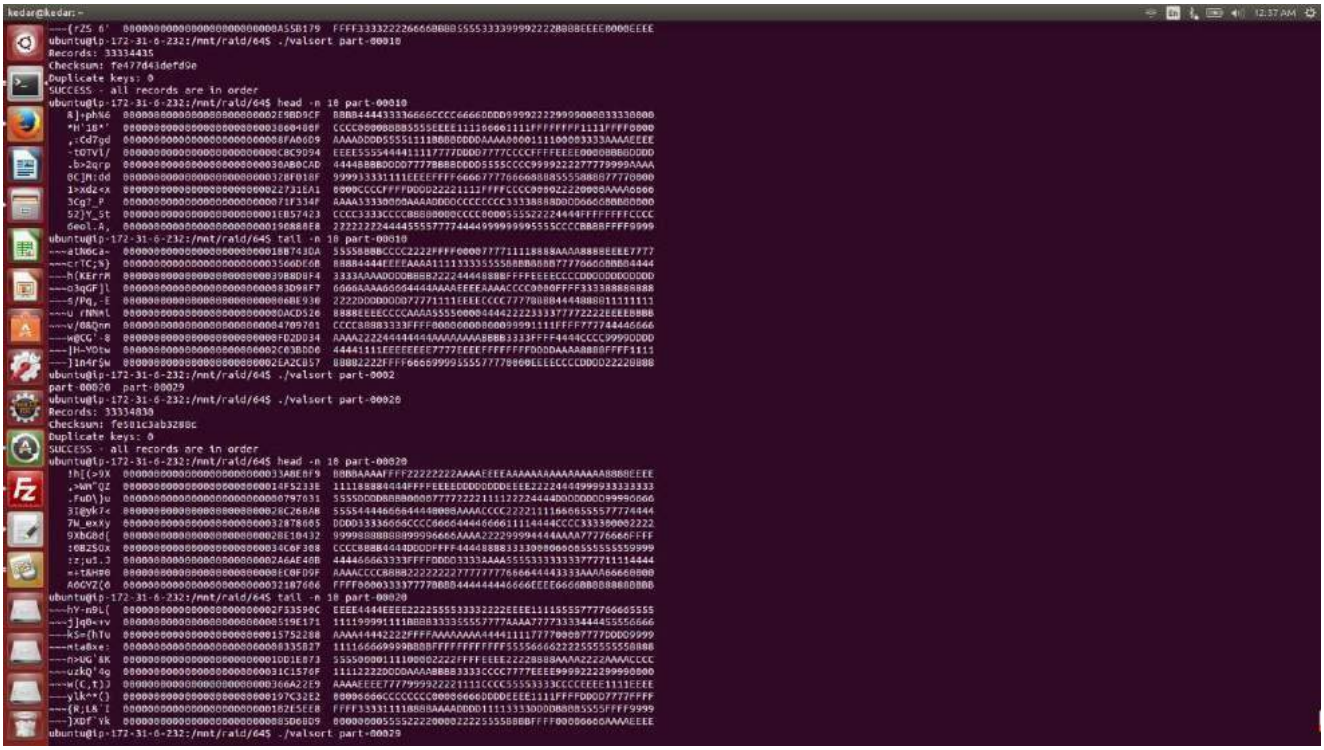
./hadoop fs -get /user/hduser/output/part* /mnt/raid/64/

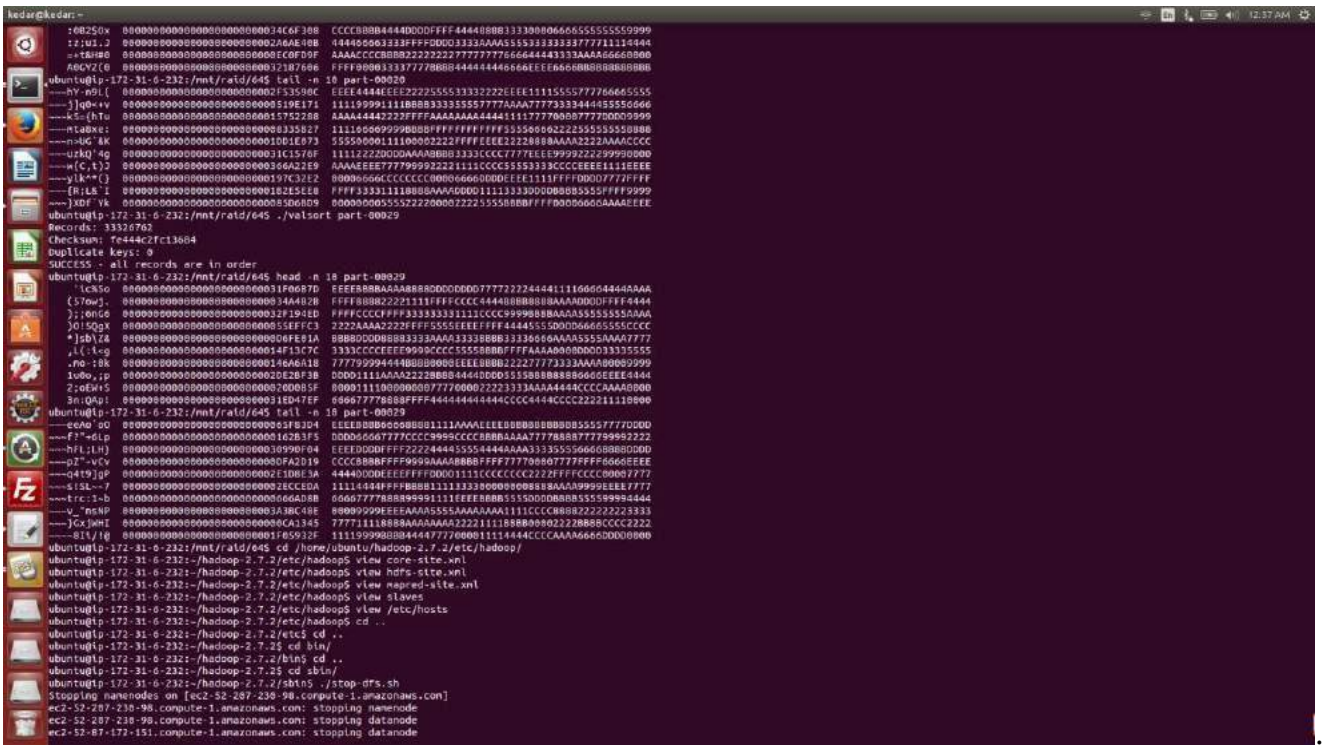


Each part can be evaluated using Valsort which will identify if all the records in the output part files are in order or not.

Also the first 10 and last 10 files of each part are shown in the screenshot.



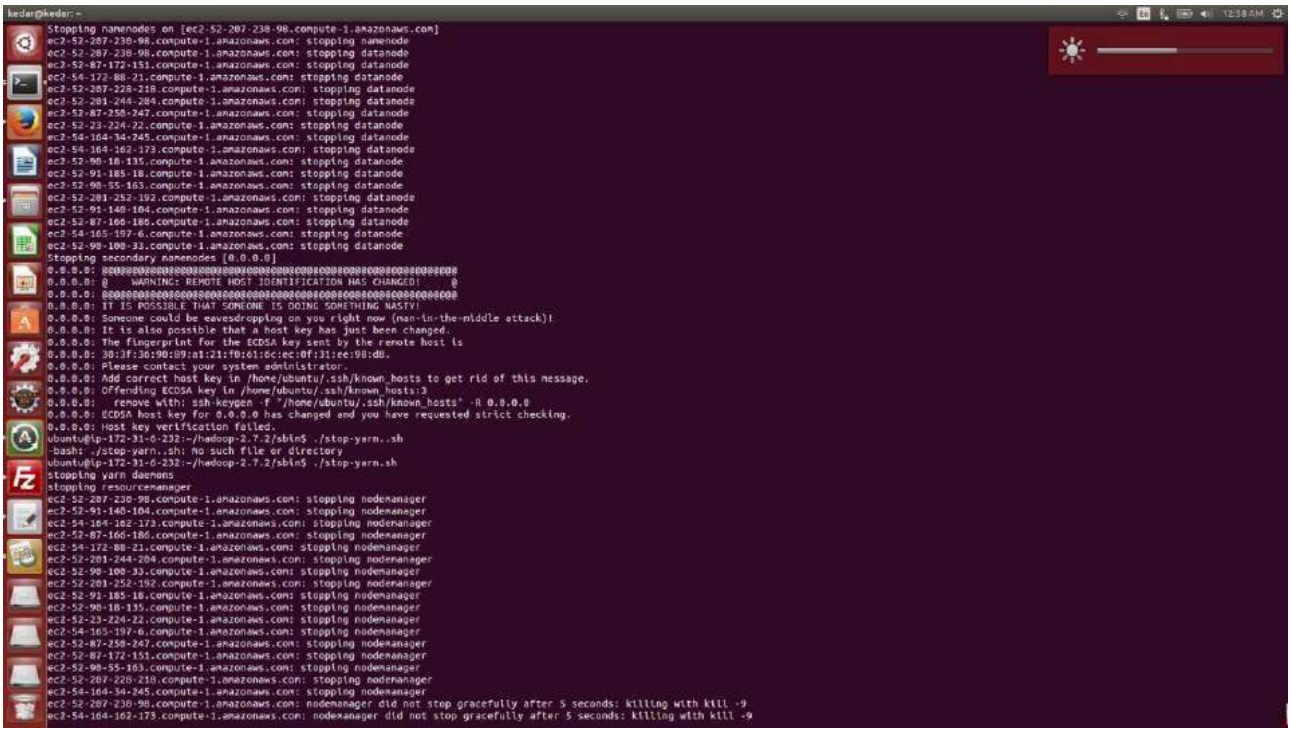




As shown in the above screenshots, all the output parts are in order. Hence the 100 GB file is sorted.

The Hadoop 17 Node Virtual Cluster can be stopped using the below commands,

./stop-dfs.sh
./stop-yarn.sh



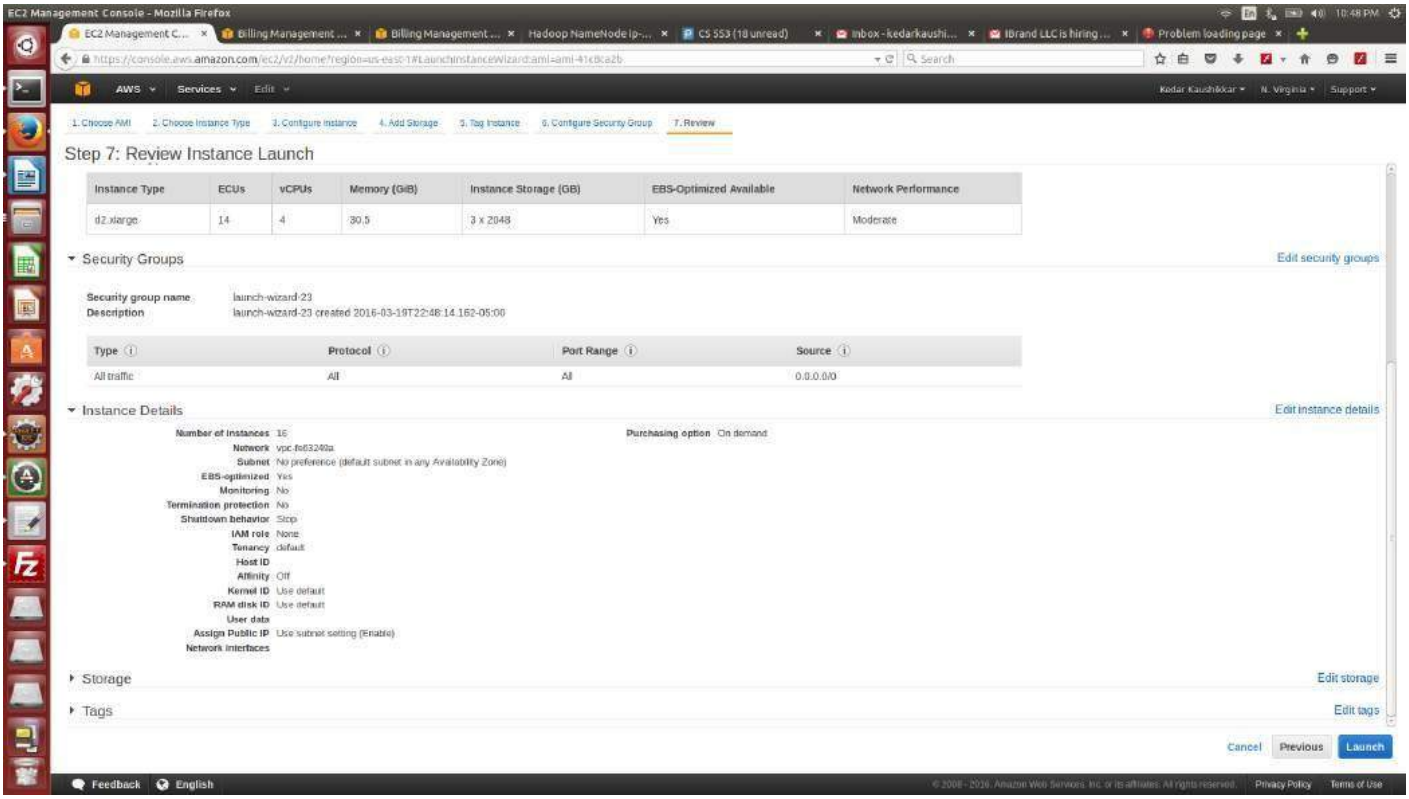
Performance:

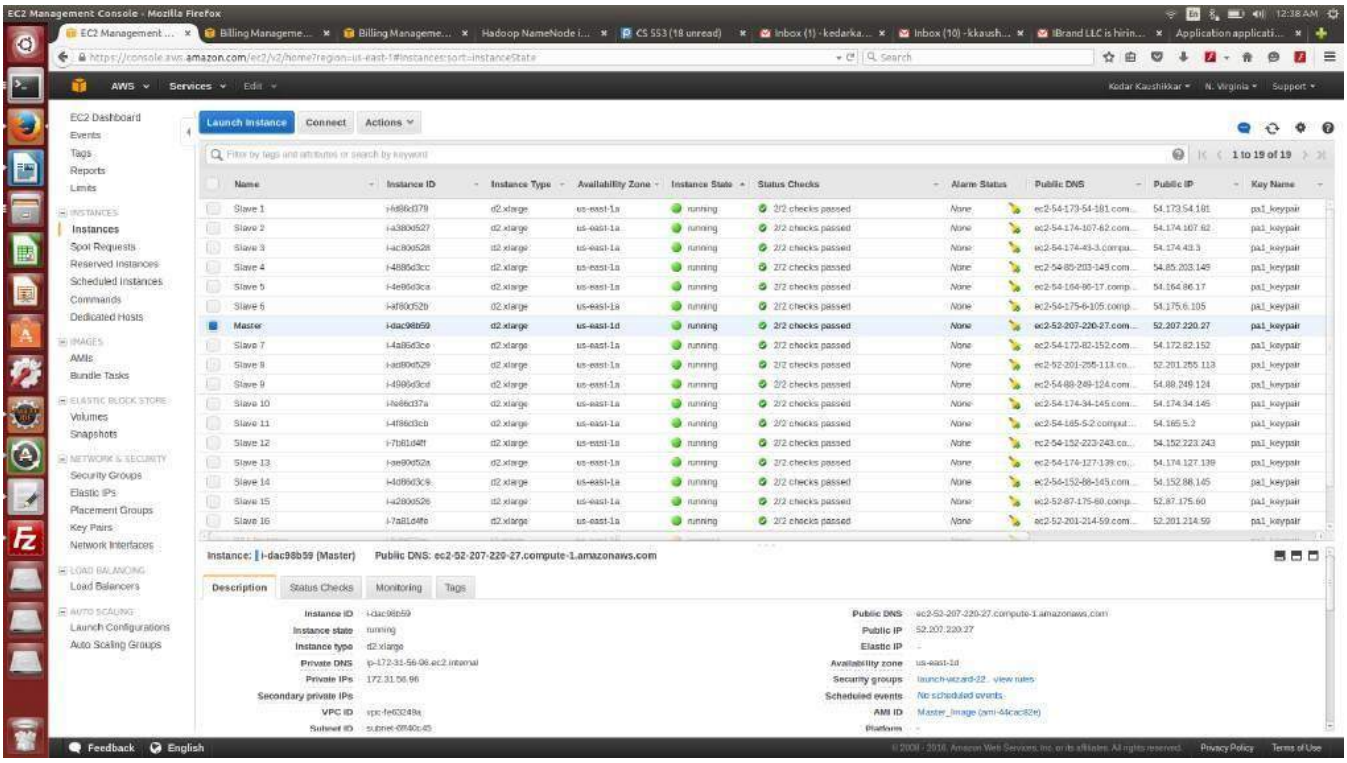
Throughput = (Datasize / Time taken)
= 100000 / 1332
= 75.07 Mb/sec

Hadoop TeraSort 16 Node (1GB)

The above steps have been followed for 1 GB dataset over d2.xlarge instance of Amazon ec2 . Here the Master and Slave nodes were both same d2.xlarge instances.

Below mentioned are the screenshots for the same.


EC2 Management Console - Mozilla Firefox
Step 7: Review Instance Launch
Instance Type | ECUs | vCPUs | Memory (GiB) | Instance Storage (GB) | EBS-Optimized Available | Network Performance
d2.xlarge | 14 | 4 | 30.5 | 3 x 2048 | Yes | Moderate
Security Groups
Edit security groups
Security group name launch-wizard-23
Description launch-wizard-23 created 2016-03-19T22:48:14.162-05:00
Type | Protocol | Port Range | Source
All traffic | All | All | 0.0.0.0/0
Instance Details
Edit instance details
Number of instances 16
Purchasing option On demand
Network vpc-fe63248a
Subnet No preference (default subnet in any Availability Zone)
EBS-optimized Yes
Monitoring No
Termination protection No
Shutdown behavior Stop
IAM role None
Tenancy default
Host ID
Affinity Off
Kernel ID Use default
RAM disk ID Use default
User data
Assign Public IP Use subnet setting (Enable)
Network interfaces
Storage
Edit storage
Tags
Edit tags
Cancel
Previous
Launch
Feedback
English


EC2 Management Console - Mozilla Firefox
EC2 Dashboard
Events
Tags
Reports
Limits
Instances
Spot Requests
Reserved Instances
Scheduled Instances
Commands
Dedicated Hosts
AMIs
Bundle Tasks
Volumes
Snapshots
Security Groups
Elastic IPs
Placement Groups
Key Pairs
Network Interfaces
Load Balancers
Launch Configurations
Auto Scaling Groups
Launch Instance
Connect
Actions
1 to 19 of 19
Name | Instance ID | Instance Type | Availability Zone | Instance State | Status Checks | Alarm Status | Public DNS | Public IP | Key Name
Master | i-dac98b59 | d2.xlarge | us-east-1d | running | 2/2 checks passed | None | ec2-52-207-220-27.com... | 52.207.220.27 | pa1_keypair
Instance: i-dac98b59 (Master) Public DNS: ec2-52-207-220-27.compute-1.amazonaws.com
Description
Status Checks
Monitoring
Tags
Instance ID i-dac98b59
Instance state running
Instance type d2.xlarge
Private DNS ip-172-31-56-96.ec2.internal
Private IPs 172.31.56.96
Secondary private IPs
VPC ID vpc-fe63248a
Subnet ID subnet-6f40c45
Public DNS ec2-52-207-220-27.compute-1.amazonaws.com
Public IP 52.207.220.27
Elastic IP
Availability zone us-east-1d
Security groups launch-wizard-22. view rules
Scheduled events No scheduled events
AMI ID Master_Image (ami-4cac82e)
Platform
Feedback
English

```
ubuntu@ip-172-31-56-96: ~/hadoop-2.7.2/bin
-rwxrwxrwx 1 ubuntu ubuntu  4436 Jan 26 00:20 capacity-scheduler.xml*
-rwxrwxrwx 1 ubuntu ubuntu  1335 Jan 26 00:20 configuration.xsl*
-rwxrwxrwx 1 ubuntu ubuntu   318 Jan 26 00:20 container-executor.cfg*
-rwxrwxrwx 1 ubuntu ubuntu   897 Mar 20 03:40 core-site.xml*
-rwxrwxrwx 1 ubuntu ubuntu  3670 Jan 26 00:20 hadoop-env.cmd*
-rwxrwxrwx 1 ubuntu ubuntu  4245 Mar 15 00:58 hadoop-env.sh*
-rwxrwxrwx 1 ubuntu ubuntu  2598 Jan 26 00:20 hadoop-metrics2.properties*
-rwxrwxrwx 1 ubuntu ubuntu  2490 Jan 26 00:20 hadoop-metrics.properties*
-rwxrwxrwx 1 ubuntu ubuntu  9683 Jan 26 00:20 hadoop-policy.xml*
-rwxrwxrwx 1 ubuntu ubuntu   918 Mar 15 00:59 hdfs-site.xml*
-rwxrwxrwx 1 ubuntu ubuntu  1449 Jan 26 00:20 httpfs-env.sh*
-rwxrwxrwx 1 ubuntu ubuntu  1657 Jan 26 00:20 httpfs-log4j.properties*
-rwxrwxrwx 1 ubuntu ubuntu    21 Jan 26 00:20 httpfs-signature.secret*
-rwxrwxrwx 1 ubuntu ubuntu   620 Jan 26 00:20 httpfs-site.xml*
-rwxrwxrwx 1 ubuntu ubuntu  3518 Jan 26 00:20 kms-acls.xml*
-rwxrwxrwx 1 ubuntu ubuntu  1527 Jan 26 00:20 kms-env.sh*
-rwxrwxrwx 1 ubuntu ubuntu  1631 Jan 26 00:20 kms-log4j.properties*
-rwxrwxrwx 1 ubuntu ubuntu  5511 Jan 26 00:20 kms-site.xml*
-rwxrwxrwx 1 ubuntu ubuntu 11237 Jan 26 00:20 log4j.properties*
-rwxrwxrwx 1 ubuntu ubuntu   951 Jan 26 00:20 mapred-env.cmd*
-rwxrwxrwx 1 ubuntu ubuntu  1383 Jan 26 00:20 mapred-env.sh*
-rwxrwxrwx 1 ubuntu ubuntu  4113 Jan 26 00:20 mapred-queues.xml.template*
-rwxrwxrwx 1 root   root    1209 Mar 20 03:40 mapred-site.xml*
-rwxrwxrwx 1 ubuntu ubuntu   758 Jan 26 00:20 mapred-site.xml.template*
-rwxrwxrwx 1 ubuntu ubuntu    42 Mar 20 03:40 slaves*
-rwxrwxrwx 1 ubuntu ubuntu  2316 Jan 26 00:20 ssl-client.xml.example*
-rwxrwxrwx 1 ubuntu ubuntu  2268 Jan 26 00:20 ssl-server.xml.example*
-rwxrwxrwx 1 ubuntu ubuntu  2250 Jan 26 00:20 yarn-env.cmd*
-rwxrwxrwx 1 ubuntu ubuntu  4567 Jan 26 00:20 yarn-env.sh*
-rwxrwxrwx 1 ubuntu ubuntu  1617 Mar 20 03:40 yarn-site.xml*
ubuntu@ip-172-31-56-96:~/hadoop-2.7.2/etc/hadoop$ vi slaves
ubuntu@ip-172-31-56-96:~/hadoop-2.7.2/etc/hadoop$ cd ..
ubuntu@ip-172-31-56-96:~/hadoop-2.7.2/etc$ cd ..
ubuntu@ip-172-31-56-96:~/hadoop-2.7.2$ cd ..
ubuntu@ip-172-31-56-96:~$ ssh-keygen -t rsa
Generating public/private rsa key pair.
Enter file in which to save the key (/home/ubuntu/.ssh/id_rsa): master_key
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in master_key.
Your public key has been saved in master_key.pub.
The key fingerprint is:
ad:8c:49:f6:1b:fa:0a:17:94:d6:3f:e3:48:9c:49:64 ubuntu@ip-172-31-56-96
The key's randomart image is:
+--[ RSA 2048]----+
|        E        |
|       =         |
|      + o        |
|     o o =       |
|      + S =      |
|     o B + o     |
|    . + * .      |
|     o . o       |
|      ooo        |
+-----------------+
ubuntu@ip-172-31-56-96:~$ ssh-copy-id -i ~/.ssh/master_key.pub ec2-52-207-220-27.compute-1.amazonaws.com

/usr/bin/ssh-copy-id: ERROR: failed to open ID file '/home/ubuntu/.ssh/master_key.pub': No such file
```

```
Terminal
ubuntu@ip-172-31-1-180: /etc
172.31.56.96 ec2-52-207-220-27.compute-1.amazonaws.com
172.31.1.180 ec2-54-172-82-152.compute-1.amazonaws.com
# The following lines are desirable for IPv6 capable hosts
::1 ip6-localhost ip6-loopback
fe00::0 ip6-localnet
ff00::0 ip6-mcastprefix
ff02::1 ip6-allnodes
ff02::2 ip6-allrouters
ff02::3 ip6-allhosts
~
~
~
~
~
~
~
~
~
~
~
~
~
~
"hosts" 9L, 318C                                  1,1           All
-rwxrwxrwx 1 ubuntu ubuntu  2268 Jan 26 00:20 ssl-server.xml.example*
-rwxrwxrwx 1 ubuntu ubuntu  2250 Jan 26 00:20 yarn-env.cmd*
-rwxrwxrwx 1 ubuntu ubuntu  4567 Jan 26 00:20 yarn-env.sh*
-rwxrwxrwx 1 ubuntu ubuntu  1617 Mar 20 03:40 yarn-site.xml*
ubuntu@ip-172-31-56-96:~/hadoop-2.7.2/etc/hadoop$ vi slaves
ubuntu@ip-172-31-56-96:~/hadoop-2.7.2/etc/hadoop$ cd ..
ubuntu@ip-172-31-56-96:~/hadoop-2.7.2/etc$ cd ..
ubuntu@ip-172-31-56-96:~/hadoop-2.7.2$ cd ..
ubuntu@ip-172-31-56-96:~$ ssh-keygen -t rsa
Generating public/private rsa key pair.
Enter file in which to save the key (/home/ubuntu/.ssh/id_rsa): master_key
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in master_key.
Your public key has been saved in master_key.pub.
The key fingerprint is:
ad:8c:49:f6:1b:fa:0a:17:94:d6:3f:e3:48:9c:49:64 ubuntu@ip-172-31-56-96
The key's randomart image is:
+--[ RSA 2048]----+
|        E        |
|       =         |
|      + o        |
|     o o =       |
|      + S =      |
|     o B + o     |
|    . + * .      |
|     o . o       |
|      ooo        |
+-----------------+
ubuntu@ip-172-31-56-96:~$ ssh-copy-id -i ~/.ssh/master_key.pub ec2-52-207-220-27.compute-1.amazonaws.com

/usr/bin/ssh-copy-id: ERROR: failed to open ID file '/home/ubuntu/.ssh/master_key.pub': No such file
```

```
ubuntu@ip-172-31-56-96: ~/hadoop-2.7.2/bin
Overwrite (y/n)? y
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /home/ubuntu/.ssh/id_rsa.
Your public key has been saved in /home/ubuntu/.ssh/id_rsa.pub.
The key fingerprint is:
d8:c2:b4:e5:3c:f4:fc:9b:c7:ca:78:2d:8d:04:dd:13 ubuntu@ip-172-31-56-96
The key's randomart image is:
+--[ RSA 2048]----+
|          . .E. |
|         . . o  |
|      . o  .  . |
|     o o o  .   |
|      = S o.    |
|       . . ..   |
|          .=    |
|         ooo=   |
|        ..=+    |
+-----------------+
ubuntu@ip-172-31-56-96:~$ ssh-copy-id -i ~/.ssh/id_rsa.pub ec2-52-207-220-27.compute-1.amazonaws.com
/usr/bin/ssh-copy-id: INFO: attempting to log in with the new key(s), to filter out any that are already installed
/usr/bin/ssh-copy-id: INFO: 1 key(s) remain to be installed -- if you are prompted now it is to install the new keys

Number of key(s) added: 1

Now try logging into the machine, with:   "ssh 'ec2-52-207-220-27.compute-1.amazonaws.com'"
and check to make sure that only the key(s) you wanted were added.

ubuntu@ip-172-31-56-96:~$ ssh-copy-id -i ~/.ssh/id_rsa.pub ec2-54-173-54-181.compute-1.amazonaws.com
The authenticity of host 'ec2-54-173-54-181.compute-1.amazonaws.com (172.31.7.5)' can't be established.
ECDSA key fingerprint is 17:54:91:23:d3:ef:9d:58:1e:64:7e:d5:6b:d9:0e:82.
Are you sure you want to continue connecting (yes/no)? yes
/usr/bin/ssh-copy-id: INFO: attempting to log in with the new key(s), to filter out any that are already installed
/usr/bin/ssh-copy-id: INFO: 1 key(s) remain to be installed -- if you are prompted now it is to install the new keys

Number of key(s) added: 1

Now try logging into the machine, with:   "ssh 'ec2-54-173-54-181.compute-1.amazonaws.com'"
and check to make sure that only the key(s) you wanted were added.

ubuntu@ip-172-31-56-96:~$ ssh-copy-id -i ~/.ssh/id_rsa.pub ec2-54-174-107-62.compute-1.amazonaws.com
The authenticity of host 'ec2-54-174-107-62.compute-1.amazonaws.com (172.31.10.28)' can't be established.
ECDSA key fingerprint is 98:00:dc:9d:ed:ea:89:cb:a4:a7:84:ba:a0:7c:1e:5d.
Are you sure you want to continue connecting (yes/no)? yes
/usr/bin/ssh-copy-id: INFO: attempting to log in with the new key(s), to filter out any that are already installed
/usr/bin/ssh-copy-id: INFO: 1 key(s) remain to be installed -- if you are prompted now it is to install the new keys

Number of key(s) added: 1

Now try logging into the machine, with:   "ssh 'ec2-54-174-107-62.compute-1.amazonaws.com'"
and check to make sure that only the key(s) you wanted were added.

ubuntu@ip-172-31-56-96:~$ ssh-copy-id -i ~/.ssh/id_rsa.pub ec2-54-164-86-17.compute-1.amazonaws.com
The authenticity of host 'ec2-54-164-86-17.compute-1.amazonaws.com (172.31.10.130)' can't be established.
ECDSA key fingerprint is 41:aa:56:cf:e3:ac:ff:b6:a5:eb:b3:d4:ea:c7:e6:02.
Are you sure you want to continue connecting (yes/no)? yes
/usr/bin/ssh-copy-id: INFO: attempting to log in with the new key(s), to filter out any that are already installed
/usr/bin/ssh-copy-id: INFO: 1 key(s) remain to be installed -- if you are prompted now it is to install the new keys
```

```
ubuntu@ip-172-31-56-96: ~/hadoop-2.7.2/bin
Now try logging into the machine, with:   "ssh 'ec2-52-201-214-59.compute-1.amazonaws.com'"
and check to make sure that only the key(s) you wanted were added.

ubuntu@ip-172-31-56-96:~$ sudo chmod 0600 ~/.ssh/authorized_keys
ubuntu@ip-172-31-56-96:~$ cd hadoop-2.7.2/
ubuntu@ip-172-31-56-96:~/hadoop-2.7.2$ cd bin/
ubuntu@ip-172-31-56-96:~/hadoop-2.7.2/bin$ cd ..
ubuntu@ip-172-31-56-96:~/hadoop-2.7.2$ cd sbin/
ubuntu@ip-172-31-56-96:~/hadoop-2.7.2/sbin$ ./start-dfs.sh
Starting namenodes on [ec2-52-207-220-27.compute-1.amazonaws.com]
ec2-52-207-220-27.compute-1.amazonaws.com: starting namenode, logging to /home/ubuntu/hadoop-2.7.2/logs/hadoop-ubuntu-namenode-ip-172-31-56-96.out
ec2-52-207-220-27.compute-1.amazonaws.com: starting datanode, logging to /home/ubuntu/hadoop-2.7.2/logs/hadoop-ubuntu-datanode-ip-172-31-56-96.out
ec2-54-174-34-145.compute-1.amazonaws.com: starting datanode, logging to /home/ubuntu/hadoop-2.7.2/logs/hadoop-ubuntu-datanode-ip-172-31-9-8.out
ec2-54-152-88-145.compute-1.amazonaws.com: starting datanode, logging to /home/ubuntu/hadoop-2.7.2/logs/hadoop-ubuntu-datanode-ip-172-31-11-94.out
ec2-52-201-214-59.compute-1.amazonaws.com: starting datanode, logging to /home/ubuntu/hadoop-2.7.2/logs/hadoop-ubuntu-datanode-ip-172-31-13-64.out
ec2-54-174-127-139.compute-1.amazonaws.com: starting datanode, logging to /home/ubuntu/hadoop-2.7.2/logs/hadoop-ubuntu-datanode-ip-172-31-1-209.out
ec2-52-201-255-113.compute-1.amazonaws.com: starting datanode, logging to /home/ubuntu/hadoop-2.7.2/logs/hadoop-ubuntu-datanode-ip-172-31-14-88.out
ec2-54-152-223-243.compute-1.amazonaws.com: starting datanode, logging to /home/ubuntu/hadoop-2.7.2/logs/hadoop-ubuntu-datanode-ip-172-31-5-92.out
ec2-54-88-249-124.compute-1.amazonaws.com: starting datanode, logging to /home/ubuntu/hadoop-2.7.2/logs/hadoop-ubuntu-datanode-ip-172-31-2-210.out
ec2-54-85-203-149.compute-1.amazonaws.com: starting datanode, logging to /home/ubuntu/hadoop-2.7.2/logs/hadoop-ubuntu-datanode-ip-172-31-6-170.out
ec2-54-174-43-3.compute-1.amazonaws.com: starting datanode, logging to /home/ubuntu/hadoop-2.7.2/logs/hadoop-ubuntu-datanode-ip-172-31-11-81.out
ec2-54-173-54-181.compute-1.amazonaws.com: starting datanode, logging to /home/ubuntu/hadoop-2.7.2/logs/hadoop-ubuntu-datanode-ip-172-31-7-5.out
ec2-54-174-107-62.compute-1.amazonaws.com: starting datanode, logging to /home/ubuntu/hadoop-2.7.2/logs/hadoop-ubuntu-datanode-ip-172-31-10-28.out
ec2-54-165-5-2.compute-1.amazonaws.com: starting datanode, logging to /home/ubuntu/hadoop-2.7.2/logs/hadoop-ubuntu-datanode-ip-172-31-10-21.out
ec2-52-87-175-60.compute-1.amazonaws.com: starting datanode, logging to /home/ubuntu/hadoop-2.7.2/logs/hadoop-ubuntu-datanode-ip-172-31-7-127.out
ec2-54-164-86-17.compute-1.amazonaws.com: starting datanode, logging to /home/ubuntu/hadoop-2.7.2/logs/hadoop-ubuntu-datanode-ip-172-31-10-130.out
ec2-54-172-82-152.compute-1.amazonaws.com: starting datanode, logging to /home/ubuntu/hadoop-2.7.2/logs/hadoop-ubuntu-datanode-ip-172-31-1-180.out
ec2-54-175-6-105.compute-1.amazonaws.com: starting datanode, logging to /home/ubuntu/hadoop-2.7.2/logs/hadoop-ubuntu-datanode-ip-172-31-5-3.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: @@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
0.0.0.0: @    WARNING: REMOTE HOST IDENTIFICATION HAS CHANGED!     @
0.0.0.0: @@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
0.0.0.0: IT IS POSSIBLE THAT SOMEONE IS DOING SOMETHING NASTY!
0.0.0.0: Someone could be eavesdropping on you right now (man-in-the-middle attack)!
0.0.0.0: It is also possible that a host key has just been changed.
0.0.0.0: The fingerprint for the ECDSA key sent by the remote host is
0.0.0.0: e3:df:38:a8:21:db:78:8f:e9:d9:1b:9b:b0:a2:7b:b0.
0.0.0.0: Please contact your system administrator.
0.0.0.0: Add correct host key in /home/ubuntu/.ssh/known_hosts to get rid of this message.
0.0.0.0: Offending ECDSA key in /home/ubuntu/.ssh/known_hosts:3
0.0.0.0:  remove with: ssh-keygen -f "/home/ubuntu/.ssh/known_hosts" -R 0.0.0.0
0.0.0.0: ECDSA host key for 0.0.0.0 has changed and you have requested strict checking.
0.0.0.0: Host key verification failed.
ubuntu@ip-172-31-56-96:~/hadoop-2.7.2/sbin$ ./start-yarn.sh
starting yarn daemons
starting resourcemanager, logging to /home/ubuntu/hadoop-2.7.2/logs/yarn-ubuntu-resourcemanager-ip-172-31-56-96.out
ec2-54-152-223-243.compute-1.amazonaws.com: starting nodemanager, logging to /home/ubuntu/hadoop-2.7.2/logs/yarn-ubuntu-nodemanager-ip-172-31-5-92.out
ec2-54-174-43-3.compute-1.amazonaws.com: starting nodemanager, logging to /home/ubuntu/hadoop-2.7.2/logs/yarn-ubuntu-nodemanager-ip-172-31-11-81.out
ec2-54-152-88-145.compute-1.amazonaws.com: starting nodemanager, logging to /home/ubuntu/hadoop-2.7.2/logs/yarn-ubuntu-nodemanager-ip-172-31-11-94.out
ec2-54-173-54-181.compute-1.amazonaws.com: starting nodemanager, logging to /home/ubuntu/hadoop-2.7.2/logs/yarn-ubuntu-nodemanager-ip-172-31-7-5.out
ec2-54-175-6-105.compute-1.amazonaws.com: starting nodemanager, logging to /home/ubuntu/hadoop-2.7.2/logs/yarn-ubuntu-nodemanager-ip-172-31-5-3.out
ec2-54-85-203-149.compute-1.amazonaws.com: starting nodemanager, logging to /home/ubuntu/hadoop-2.7.2/logs/yarn-ubuntu-nodemanager-ip-172-31-6-170.out
ec2-52-87-175-60.compute-1.amazonaws.com: starting nodemanager, logging to /home/ubuntu/hadoop-2.7.2/logs/yarn-ubuntu-nodemanager-ip-172-31-7-127.out
ec2-54-172-82-152.compute-1.amazonaws.com: starting nodemanager, logging to /home/ubuntu/hadoop-2.7.2/logs/yarn-ubuntu-nodemanager-ip-172-31-1-180.out
ec2-54-164-86-17.compute-1.amazonaws.com: starting nodemanager, logging to /home/ubuntu/hadoop-2.7.2/logs/yarn-ubuntu-nodemanager-ip-172-31-10-130.out
ec2-54-174-127-139.compute-1.amazonaws.com: starting nodemanager, logging to /home/ubuntu/hadoop-2.7.2/logs/yarn-ubuntu-nodemanager-ip-172-31-1-209.out
ec2-54-174-107-62.compute-1.amazonaws.com: starting nodemanager, logging to /home/ubuntu/hadoop-2.7.2/logs/yarn-ubuntu-nodemanager-ip-172-31-10-28.out
```

```
ubuntu@ip-172-31-56-96: ~/hadoop-2.7.2/bin
ec2-54-152-223-243.compute-1.amazonaws.com: starting nodemanager, logging to /home/ubuntu/hadoop-2.7.2/logs/yarn-ubuntu-nodemanager-ip-172-31-5-92.out
ec2-54-164-86-17.compute-1.amazonaws.com: starting nodemanager, logging to /home/ubuntu/hadoop-2.7.2/logs/yarn-ubuntu-nodemanager-ip-172-31-10-130.out
ec2-54-88-249-124.compute-1.amazonaws.com: starting nodemanager, logging to /home/ubuntu/hadoop-2.7.2/logs/yarn-ubuntu-nodemanager-ip-172-31-2-210.out
ec2-54-175-0-105.compute-1.amazonaws.com: starting nodemanager, logging to /home/ubuntu/hadoop-2.7.2/logs/yarn-ubuntu-nodemanager-ip-172-31-5-3.out
ec2-54-152-88-145.compute-1.amazonaws.com: starting nodemanager, logging to /home/ubuntu/hadoop-2.7.2/logs/yarn-ubuntu-nodemanager-ip-172-31-11-94.out
ec2-54-174-107-62.compute-1.amazonaws.com: starting nodemanager, logging to /home/ubuntu/hadoop-2.7.2/logs/yarn-ubuntu-nodemanager-ip-172-31-10-20.out
ec2-54-172-82-152.compute-1.amazonaws.com: starting nodemanager, logging to /home/ubuntu/hadoop-2.7.2/logs/yarn-ubuntu-nodemanager-ip-172-31-1-180.out
ec2-52-201-255-113.compute-1.amazonaws.com: starting nodemanager, logging to /home/ubuntu/hadoop-2.7.2/logs/yarn-ubuntu-nodemanager-ip-172-31-14-88.out
ec2-52-207-220-27.compute-1.amazonaws.com: starting nodemanager, logging to /home/ubuntu/hadoop-2.7.2/logs/yarn-ubuntu-nodemanager-ip-172-31-56-96.out
ubuntu@ip-172-31-56-96:~/hadoop-2.7.2/sbin$ cd ..
ubuntu@ip-172-31-56-96:~/hadoop-2.7.2$ cd bin/
ubuntu@ip-172-31-56-96:~/hadoop-2.7.2/bin$ ./hdfs dfs -mkdir /user
ubuntu@ip-172-31-56-96:~/hadoop-2.7.2/bin$ ./hdfs dfs -mkdir /user/hduser
ubuntu@ip-172-31-56-96:~/hadoop-2.7.2/bin$ ./hdfs dfs -mkdir /user/hduser/input
ubuntu@ip-172-31-56-96:~/hadoop-2.7.2/bin$ ./hadoop fs -put /home/ubuntu/64/input.txt /user/hduser/input
ubuntu@ip-172-31-56-96:~/hadoop-2.7.2/bin$ ./hdfs dfs -ls /user/hduser/input
Found 1 items
-rw-r--r--   2 ubuntu supergroup 1000000000 2016-03-20 05:14 /user/hduser/input/input.txt
ubuntu@ip-172-31-56-96:~/hadoop-2.7.2/bin$ ./hadoop jar /home/ubuntu/Map_Reduce_TeraSort.jar Map_Reduce /user/hduser/input /user/hduser/output
16/03/20 05:14:47 INFO client.RMProxy: Connecting to ResourceManager at ec2-52-207-220-27.compute-1.amazonaws.com/172.31.56.96:8032
16/03/20 05:14:51 INFO client.RMProxy: Connecting to ResourceManager at ec2-52-207-220-27.compute-1.amazonaws.com/172.31.56.96:8032
16/03/20 05:14:57 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
16/03/20 05:15:14 INFO mapred.FileInputFormat: Total input paths to process : 1
16/03/20 05:15:19 INFO mapreduce.JobSubmitter: number of splits:8
16/03/20 05:15:19 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1458450758145_0001
16/03/20 05:15:20 INFO impl.YarnClientImpl: Submitted application application_1458450758145_0001
16/03/20 05:15:20 INFO mapreduce.Job: The url to track the job: http://ec2-52-207-220-27.compute-1.amazonaws.com:8088/proxy/application_1458450758145_0001/
16/03/20 05:15:20 INFO mapreduce.Job: Running job: job_1458450758145_0001
16/03/20 05:15:27 INFO mapreduce.Job: Job job_1458450758145_0001 running in uber mode : false
16/03/20 05:15:27 INFO mapreduce.Job:  map 0% reduce 0%
16/03/20 05:15:36 INFO mapreduce.Job:  map 13% reduce 0%
16/03/20 05:15:38 INFO mapreduce.Job:  map 21% reduce 0%
16/03/20 05:15:39 INFO mapreduce.Job:  map 69% reduce 0%
16/03/20 05:15:41 INFO mapreduce.Job:  map 81% reduce 0%
16/03/20 05:15:42 INFO mapreduce.Job:  map 92% reduce 0%
16/03/20 05:15:43 INFO mapreduce.Job:  map 100% reduce 0%
16/03/20 05:15:47 INFO mapreduce.Job:  map 100% reduce 29%
16/03/20 05:15:50 INFO mapreduce.Job:  map 100% reduce 69%
16/03/20 05:15:53 INFO mapreduce.Job:  map 100% reduce 71%
16/03/20 05:16:02 INFO mapreduce.Job:  map 100% reduce 77%
16/03/20 05:16:05 INFO mapreduce.Job:  map 100% reduce 82%
16/03/20 05:16:08 INFO mapreduce.Job:  map 100% reduce 87%
16/03/20 05:16:11 INFO mapreduce.Job:  map 100% reduce 91%
16/03/20 05:16:14 INFO mapreduce.Job:  map 100% reduce 96%
16/03/20 05:16:17 INFO mapreduce.Job:  map 100% reduce 100%
16/03/20 05:16:23 INFO mapreduce.Job: Job job_1458450758145_0001 completed successfully
16/03/20 05:16:23 INFO mapreduce.Job: Counters: 51
        File System Counters
                FILE: Number of bytes read=1978314714
                FILE: Number of bytes written=2999378930
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=1000029730
                HDFS: Number of bytes written=1000000000
                HDFS: Number of read operations=27
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
```

```
ubuntu@ip-172-31-56-96: ~/hadoop-2.7.2/bin
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=1000028672
        File Output Format Counters
                Bytes Written=1000000000
ubuntu@ip-172-31-56-96:~/hadoop-2.7.2/bin$ ./hdfs dfs -ls /user/hduser/output
Found 2 items
-rw-r--r--   2 ubuntu supergroup          0 2016-03-20 05:16 /user/hduser/output/_SUCCESS
-rw-r--r--   2 ubuntu supergroup 1000000000 2016-03-20 05:16 /user/hduser/output/part-00000
ubuntu@ip-172-31-56-96:~/hadoop-2.7.2/bin$ ./hadoop fs -get /user/hduser/output/part-00000 /home/ubuntu/64/
ubuntu@ip-172-31-56-96:~/hadoop-2.7.2/bin$ cd ..
ubuntu@ip-172-31-56-96:~/hadoop-2.7.2$ cd ..
ubuntu@ip-172-31-56-96:~$ cd 64
ubuntu@ip-172-31-56-96:~/64$ ll
total 1953412
drwxrwxr-x 2 ubuntu ubuntu       4096 Mar 20 05:17 ./
drwxr-xr-x 7 ubuntu ubuntu       4096 Mar 20 04:48 ../
-rwxrwxr-x 1 ubuntu ubuntu     141045 Mar 17  2013 gensort*
-rwxrwxr-x 1 ubuntu ubuntu 1000000000 Mar 20 05:08 input.txt*
-rw-r--r-- 1 ubuntu ubuntu 1000000000 Mar 20 05:17 part-00000
-rwxrwxr-x 1 ubuntu ubuntu     134558 Mar 17  2013 valsort*
ubuntu@ip-172-31-56-96:~/64$
ubuntu@ip-172-31-56-96:~/64$ ./valsort part-00000
Records: 10000000
Checksum: 4c4d632850e235
Duplicate keys: 0
SUCCESS - all records are in order
ubuntu@ip-172-31-56-96:~/64$ head -n 10 part-00000
    "O!uve  0000000000000000000000000001228D4  77778888000022224444DDDDDDDDEEEEE00000000CCCC7777DDDD
   ,K4a-:v  0000000000000000000000000188132  5555EEEEE88889999444 4FFFF1111CCCCEEEEE1111EEEEE6666FFFF
   .FuD\)u  0000000000000000000000000797631  5555DDDDD888BB0000777722221111222244440DDDDDDDD99996666
   ;5Y?hct  0000000000000000000000000D7D1DF5  2222AAAACCCCFFFFAAAAAAAA5555EEEEE4444222220DDDD99992222
   =2G49(-  0000000000000000000000000889EE  5555DDDD1111CCCC9999BBBB0000BBBCCCCFFFFFCCCC44443333
   N]M97sP  0000000000000000000000000429597  5555FFFF0000777755555999911 11CCCC66669999AAAAEEEEBBBB
   P0X?Rs&  0000000000000000000000000041162E  8888333339999FFFF1111CCCCB888BCCCC9999EEEEDDDD00003333
   [Xq\\$%  0000000000000000000000000097A5F0  66666666EEEEDDDD7777FFFF0000555 5FFFFFFFF8888555 51111
   rAMQg4v  0000000000000000000000000B180D3  BBBB111111119999FFFFFFFFFFFFFFF4444BBBBBBBB4444CCCC
  I&})JT3;  0000000000000000000000000004602E1  4444FFFFCCCCB0000888BCCCCFFFFCCCCEEEE5555666666666666
ubuntu@ip-172-31-56-96:~/64$ tail -n 10 part-00000
~~~%A NB_t  0000000000000000000000000003DE5EC  FFFF7777EEEEBBBB4444EEEEEEEE33339999DDDD99990000555 5
~~~-c-CQ(>  0000000000000000000000000832611  9999FFFF111177773333777700001111444444440000BBBB6666
~~~8Y)FqL*  0000000000000000000000000742A4C  BBBB1111CCCCEEEEBBBB0000000077773333333 3DDDD22225555
~~~>0d=QT]  0000000000000000000000000674C0F  9999DDDD0000555556666CCCC22220000FFFFEEEEDDDDFFFF0000
~~~]JA(}]$  0000000000000000000000000060BCBE  1111222233334444555559999AAAABBBB9999FFFFDDDDBBBB3333
~~~]Zp.#/+  0000000000000000000000000389A5A  CCCC8888EEEEEAAAAEEEE3333333377770000FFFFCCCC66667777
~~~_jQepLx  0000000000000000000000001E5D4  1111999911555588BBB111100002222EEEE6666BBBB7777DDDD
~~~nt-2H[N  0000000000000000000000000332A13  4444111188BBBBBB333377777FFFF4444555555555333300000CCCC
~~~s/Pq,-E  0000000000000000000000000BE930  2222DDDDDDDD777711111EEEEEEEEC7777BBBB4444888811111111
~~~zbA_Tt  0000000000000000000000000007F9F4F  BBBBCCCC6666555559999FFFF8888AAAA11116666AAAABBBB0000
ubuntu@ip-172-31-56-96:~/64$ ll
total 1953416
drwxrwxr-x 2 ubuntu ubuntu       4096 Mar 20 05:17 ./
drwxr-xr-x 7 ubuntu ubuntu       4096 Mar 20 04:48 ../
-rwxrwxr-x 1 ubuntu ubuntu     141045 Mar 17  2013 gensort*
-rwxrwxr-x 1 ubuntu ubuntu 1000000000 Mar 20 05:08 input.txt*
```

```
ubuntu@ip-172-31-56-96: ~/hadoop-2.7.2/bin
                FILE: Number of bytes read=1978314714
                FILE: Number of bytes written=2999378930
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=1000029736
                HDFS: Number of bytes written=1000000000
                HDFS: Number of read operations=27
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Killed map tasks=1
                Launched map tasks=9
                Launched reduce tasks=1
                Data-local map tasks=5
                Rack-local map tasks=4
                Total time spent by all maps in occupied slots (ms)=100334
                Total time spent by all reduces in occupied slots (ms)=44175
                Total time spent by all map tasks (ms)=100334
                Total time spent by all reduce tasks (ms)=44175
                Total vcore-milliseconds taken by all map tasks=100334
                Total vcore-milliseconds taken by all reduce tasks=44175
                Total megabyte-milliseconds taken by all map tasks=102742016
                Total megabyte-milliseconds taken by all reduce tasks=45235200
        Map-Reduce Framework
                Map input records=10000000
                Map output records=10000000
                Map output bytes=1000000000
                Map output materialized bytes=1020000048
                Input split bytes=1064
                Combine input records=10000000
                Combine output records=10000000
                Reduce input groups=10000000
                Reduce shuffle bytes=1020000048
                Reduce input records=10000000
                Reduce output records=10000000
                Spilled Records=29395241
                Shuffled Maps =8
                Failed Shuffles=0
                Merged Map outputs=8
                GC time elapsed (ms)=2533
                CPU time spent (ms)=117330
                Physical memory (bytes) snapshot=2342481920
                Virtual memory (bytes) snapshot=7517401088
                Total committed heap usage (bytes)=1881669632
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=1000028672
        File Output Format Counters
                Bytes Written=1000000000
ubuntu@ip-172-31-56-96:~/hadoop-2.7.2/bin$ ./hdfs dfs -ls /user/hduser/output
Found 2 items
```

Namenode Information - Mozilla Firefox

ec2-52-207-220-27.compute-1.amazonaws.com:50070/dfshealth.html#tab-overview

Started:	Sun Mar 20 05:12:15 UTC 2016
Version:	2.7.2, rb165c4fe8a74265c792ce23f546c64604acf0e41
Compiled:	2016-01-26T00:08Z by jenkins from (detached from b165c4f)
Cluster ID:	CID-8c28f51d-15df-4120-8fd4-125bc0af5546
Block Pool ID:	BP-1893423351-172.31.56.96-1458450710916

Summary

Security is off.

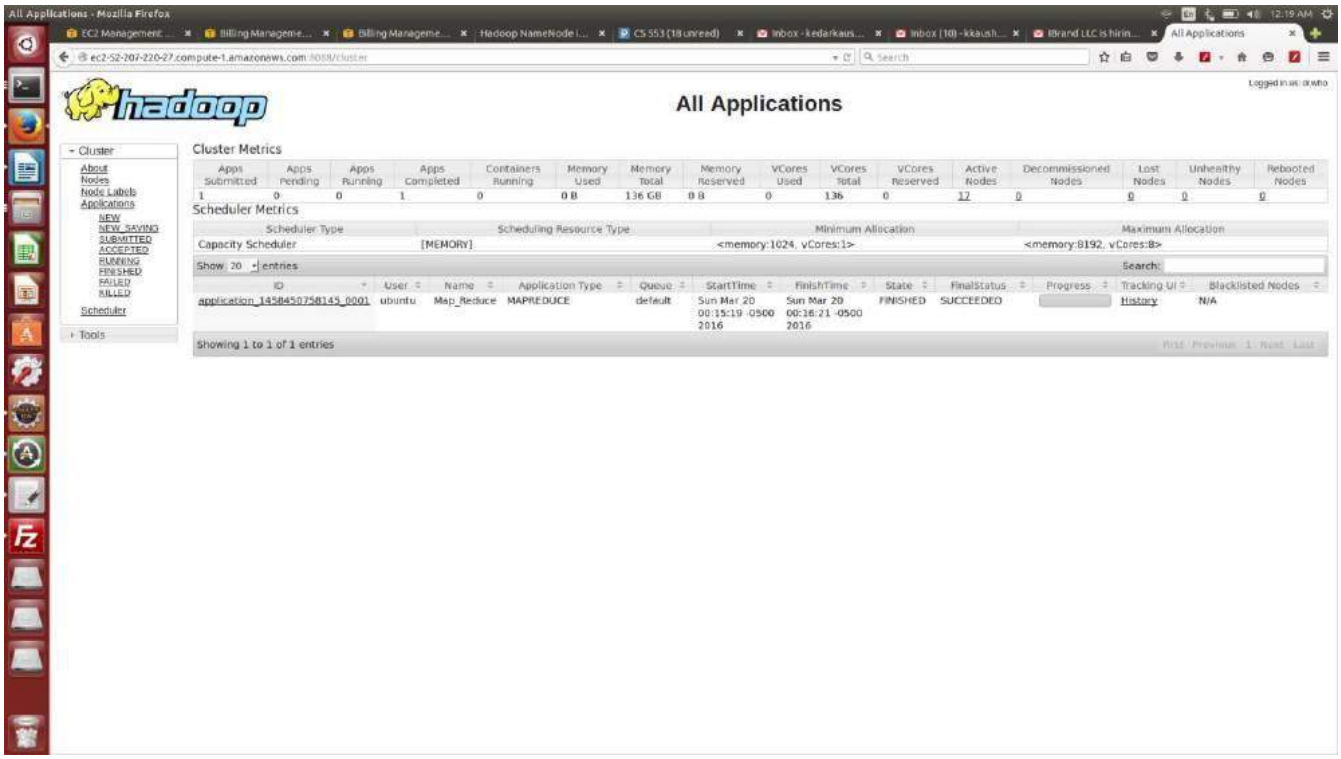
Safemode is off.

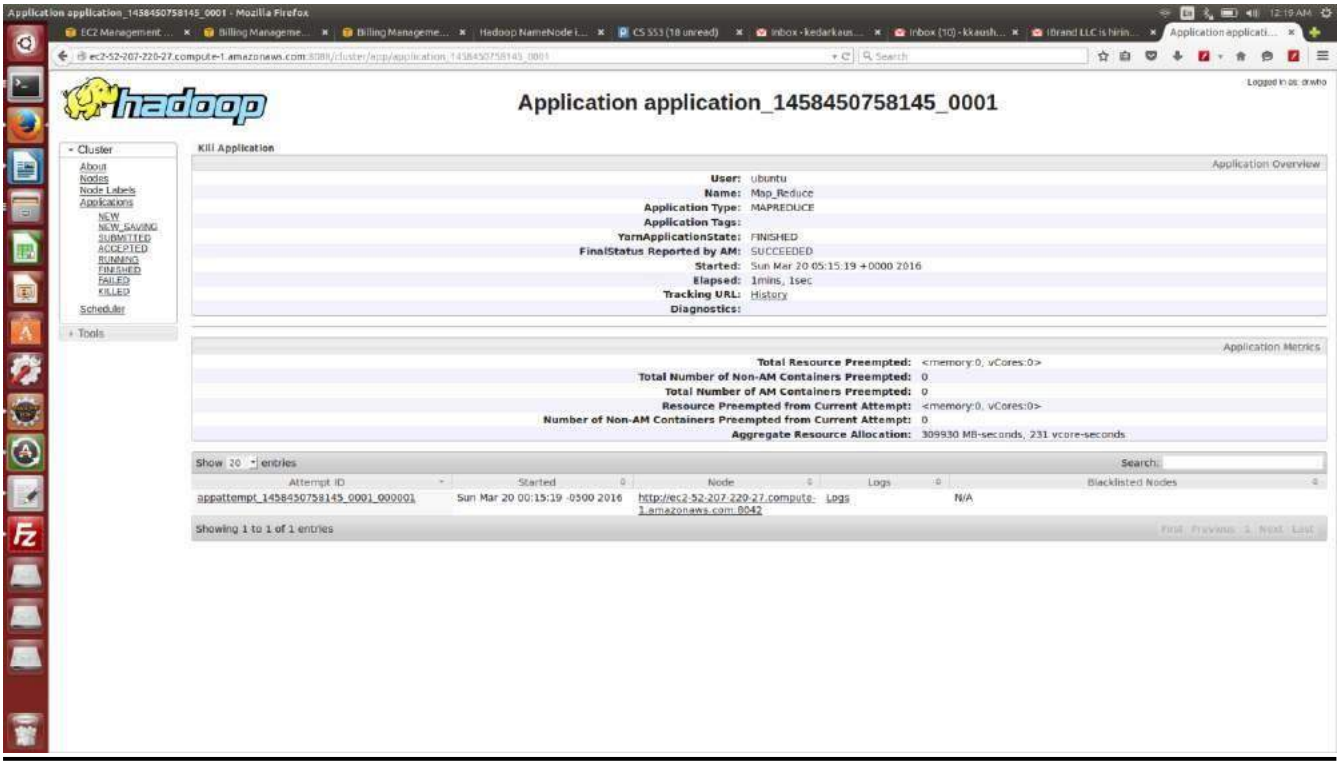
19 files and directories, 19 blocks = 38 total filesystem object(s).

Heap Memory used 96.93 MB of 459 MB Heap Memory. Max Heap Memory is 889 MB.

Non Heap Memory used 31.73 MB of 32.94 MB Commited Non Heap Memory. Max Non Heap Memory is 214 MB.

Configured Capacity:	131.57 GB
DFS Used:	3.76 GB (2.85%)
Non DFS Used:	34.34 GB
DFS Remaining:	93.48 GB (71.05%)
Block Pool Used:	3.76 GB (2.85%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 1.63% / 12.13% / 3.75%
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
Block Deletion Start Time	3/20/2016, 12:12:15 AM





<u>As mentioned in the screenshot, the total time required to execute map-reduce program to sort 1 GB of data on 16 node Virtual Cluster is 1mins 1 sec.</u>

Performance:

Throughput = (Datasize / Time taken)
= 1000 / 61
= 16.39 Mb/sec

Speedups:

As per the results that we get, we can generalize that when we increase the number of nodes from 1 Node to 16th the performance increases linearly stating that if the data size is increased 10 times for 16 nodes, the performance also increases by 10 times.

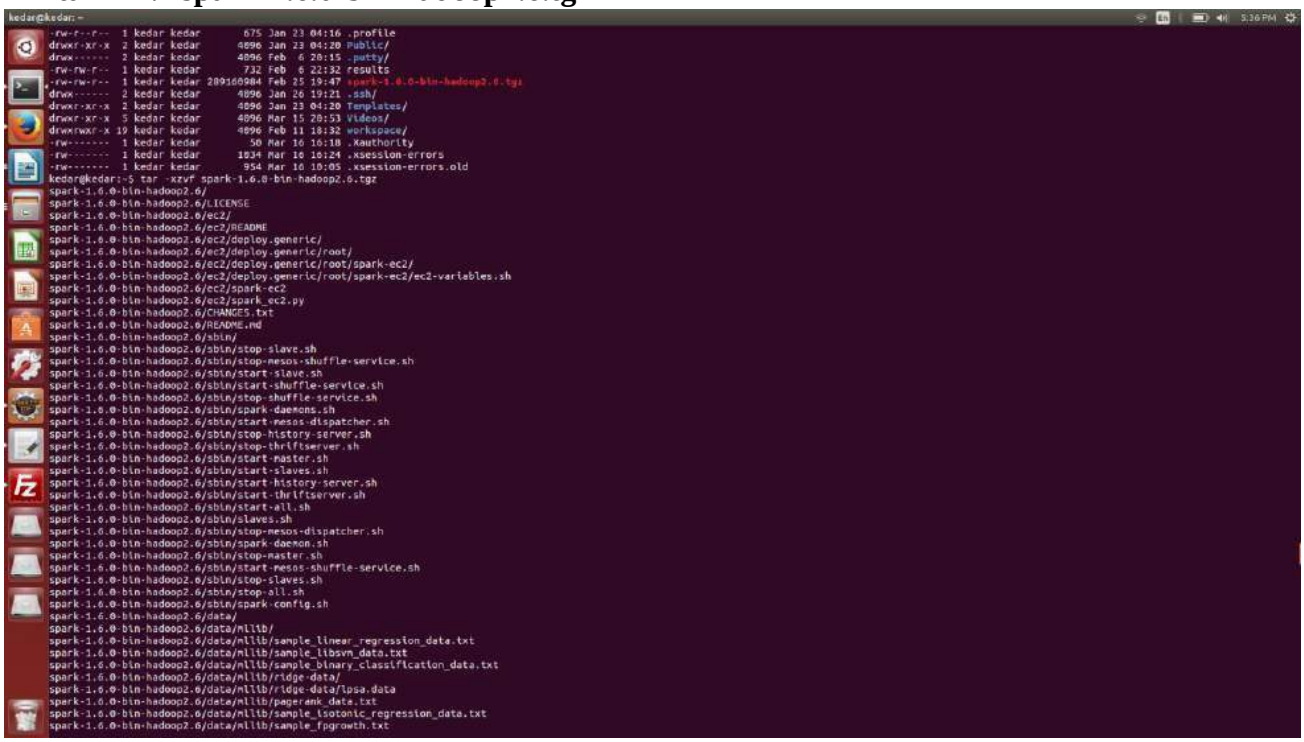
We can see that the throughput for 10GB 1 node is 7.4Mb/sec and for 100GB 16 Node is 75.07Mb/sec. Hence the throughput increases 10 times provided the data set is increased at the same rate and the number of nodes i.e. the parallelism and the virtual cores are increased.

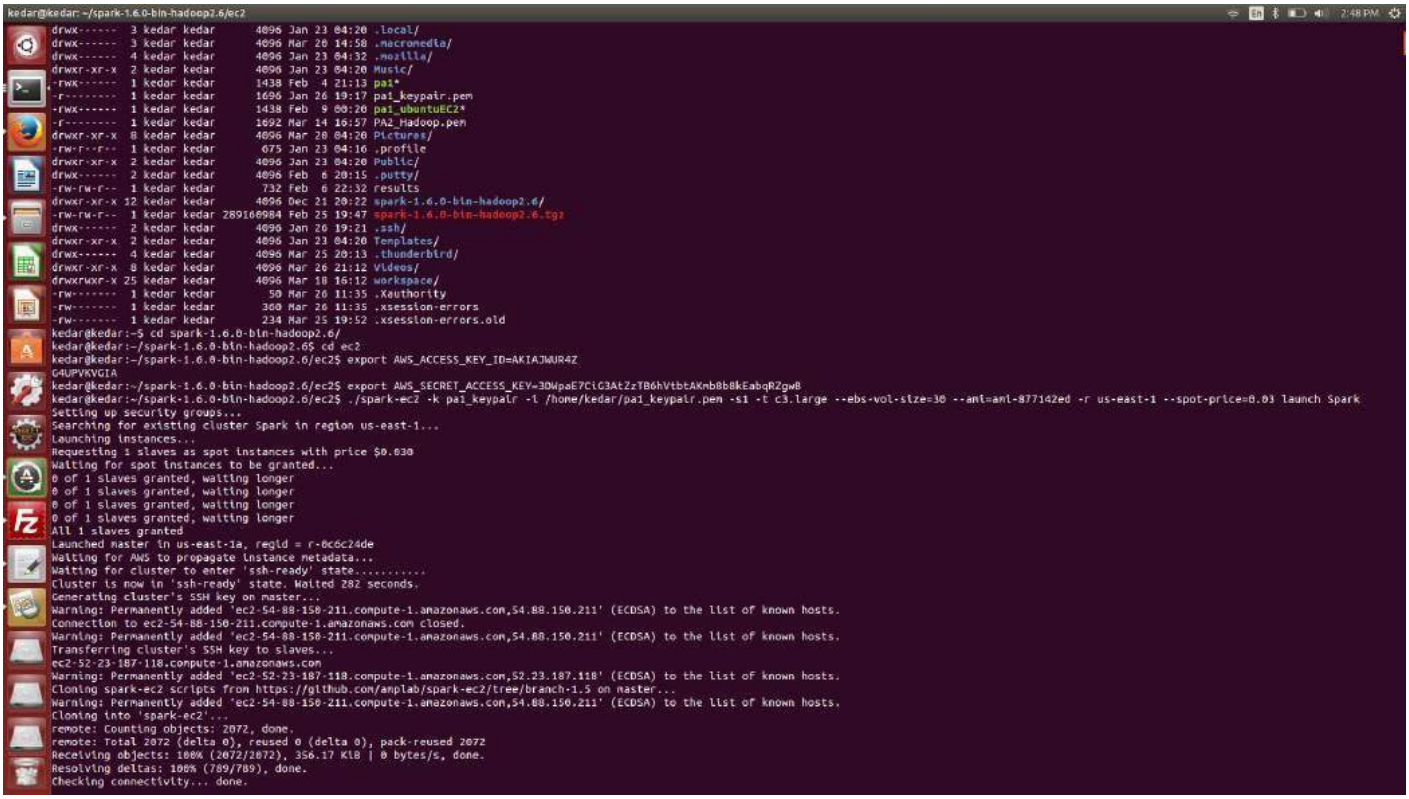
If the dataset size is kept constant and the number of virtual clusters are increased, then the performance will increase with an efficiency of 50% each time the nodes are doubled than the original.

Spark

Spark Virtual Cluster 1 Node:

- Download the Spark into your System and Extract it on the root directory
  tar -xzvf spark-1.6.0-bin-hadoop2.6.tgz





- Once the Spark Folder is extracted, you need to get the access key and the secret key that spark installer will use to initiate instance on your amazon account.

  Download your amazon account Access Key and Secret Key using below steps.

- Login to your amazon AWS account
- Go to Security Credentials
- Access Key (Access Key ID and Secret Access Key)
- Create New Access Key
- Download the file containing the keys

- Once we have the keys, provide your access keys to the Amazon EC2 CLI to set the AWS_ACCESS_KEY and AWS_SECRET_KEY environment variables using below commands.

- Go to EC2 folder of your spark installation directory,

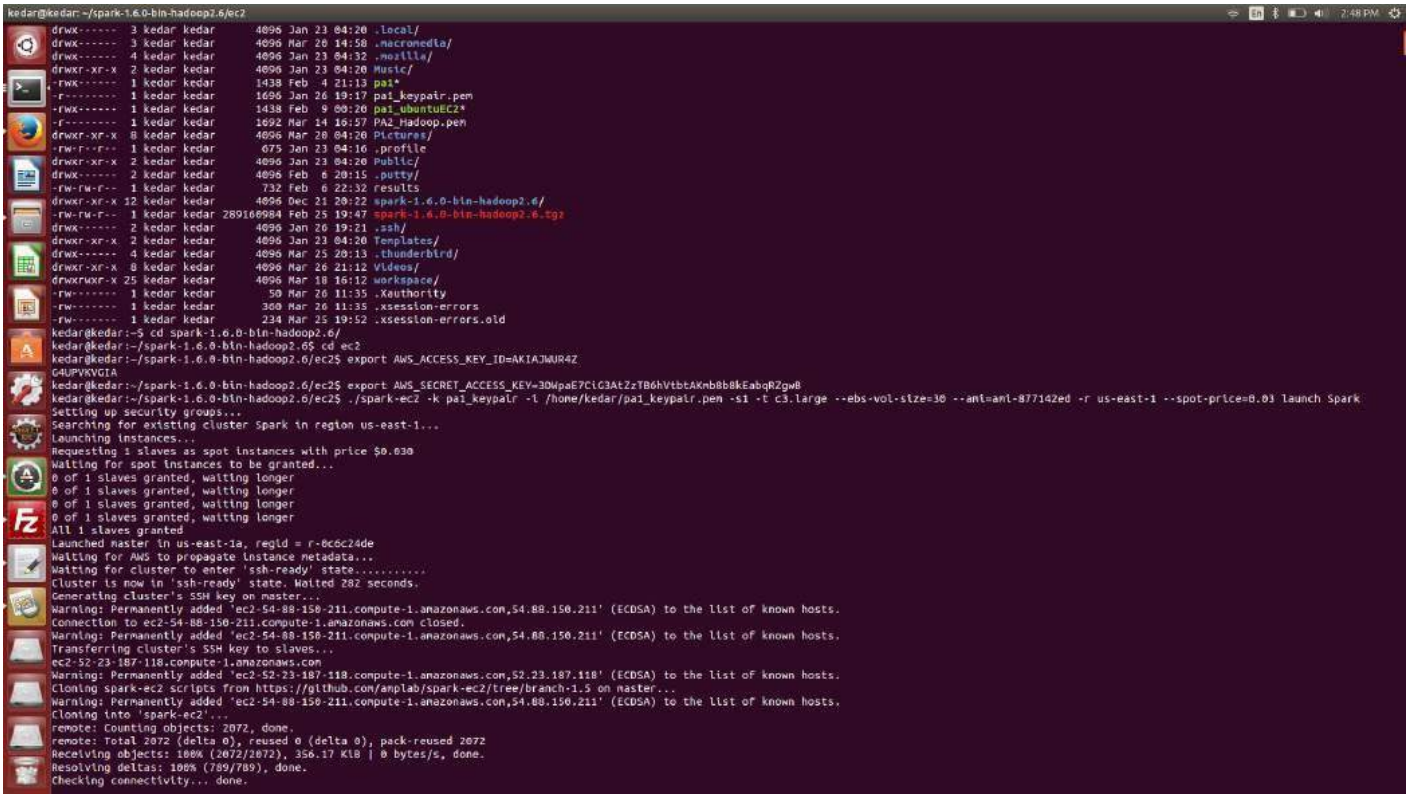
  export AWS_ACCESS_KEY_ID=AKIAJWUR4ZG4UPKKVGIA
  export AWS_SECRET_ACCESS_KEY=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

- Once the keys are exported, the spark instance can be started using the following command,

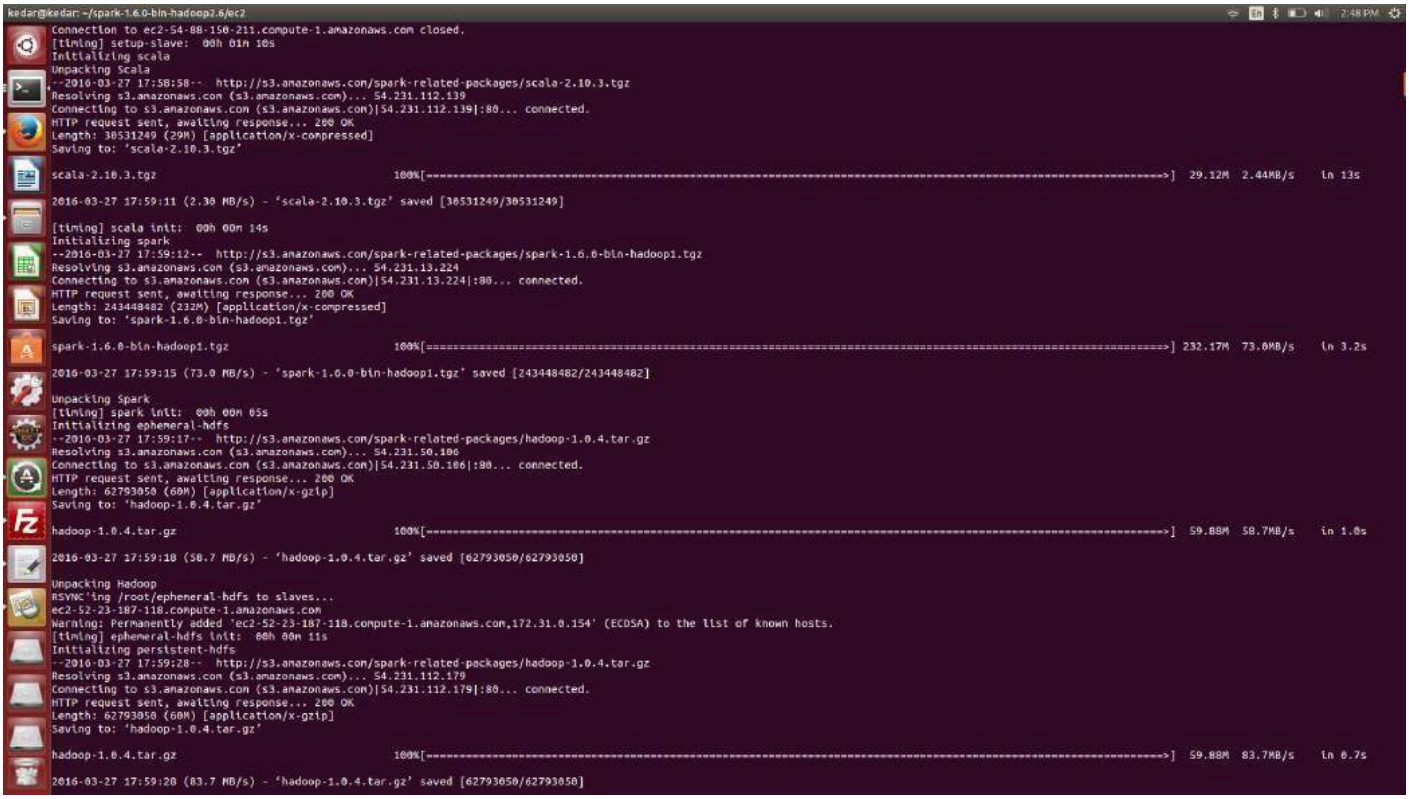
./spark-ec2 -k pa1_keypair -i /home/kedar/pa1_keypair.pem -s1 -t c3.large --ebs-vol-size=30 --ami=ami-877142ed -r us-east-1 --spot-price=0.03 launch Spark

-i **:** will represent the name and path of your .pem file
-sN: where N will represent the number of instances you need to create

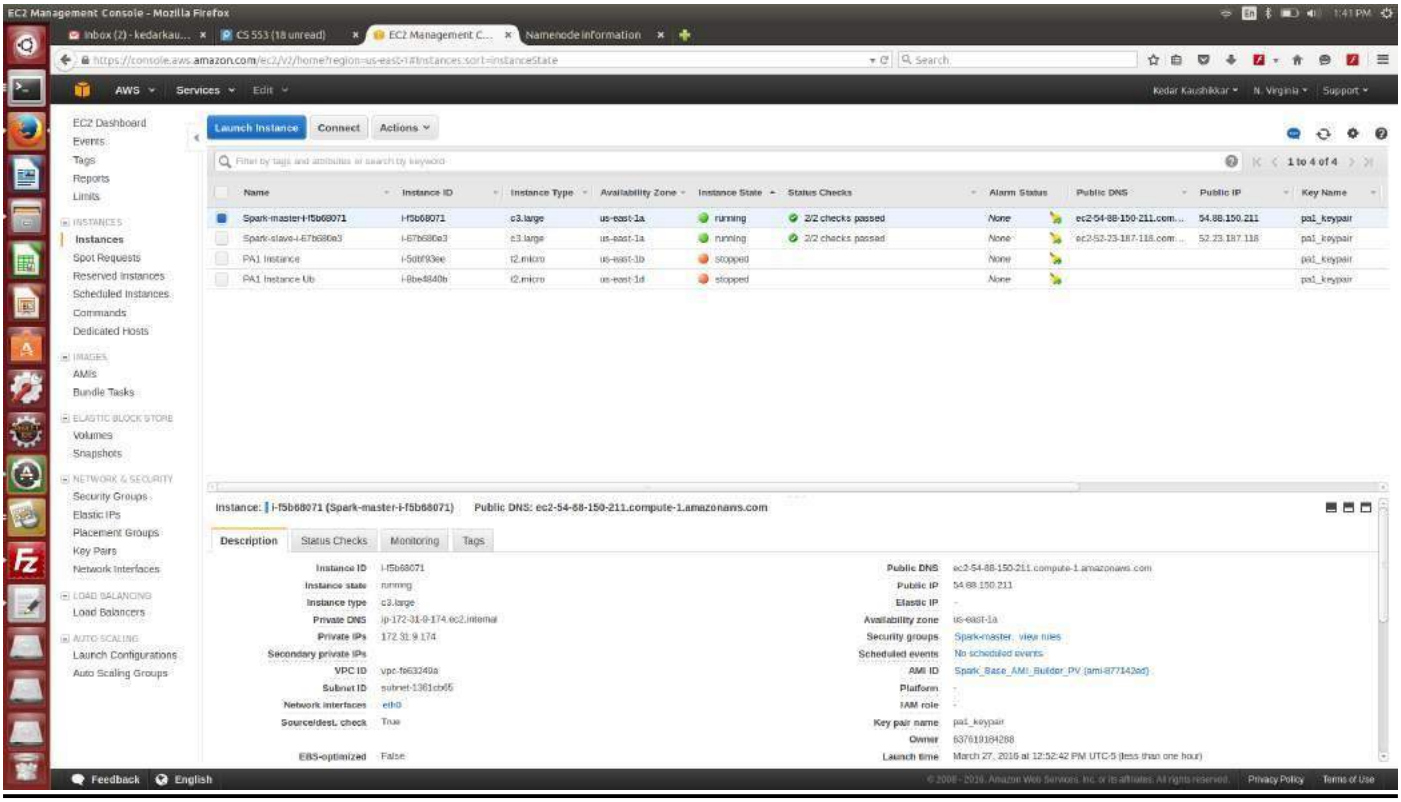
-t: the type of instance.
-ebs-vol-size : Size of Elastic Block Storage to be added to our instance.
-r: region from where once instance is requested.
-spot-price: the price you bid for your spot instance.
Launch NAME: Represents the name of your Spark instance.



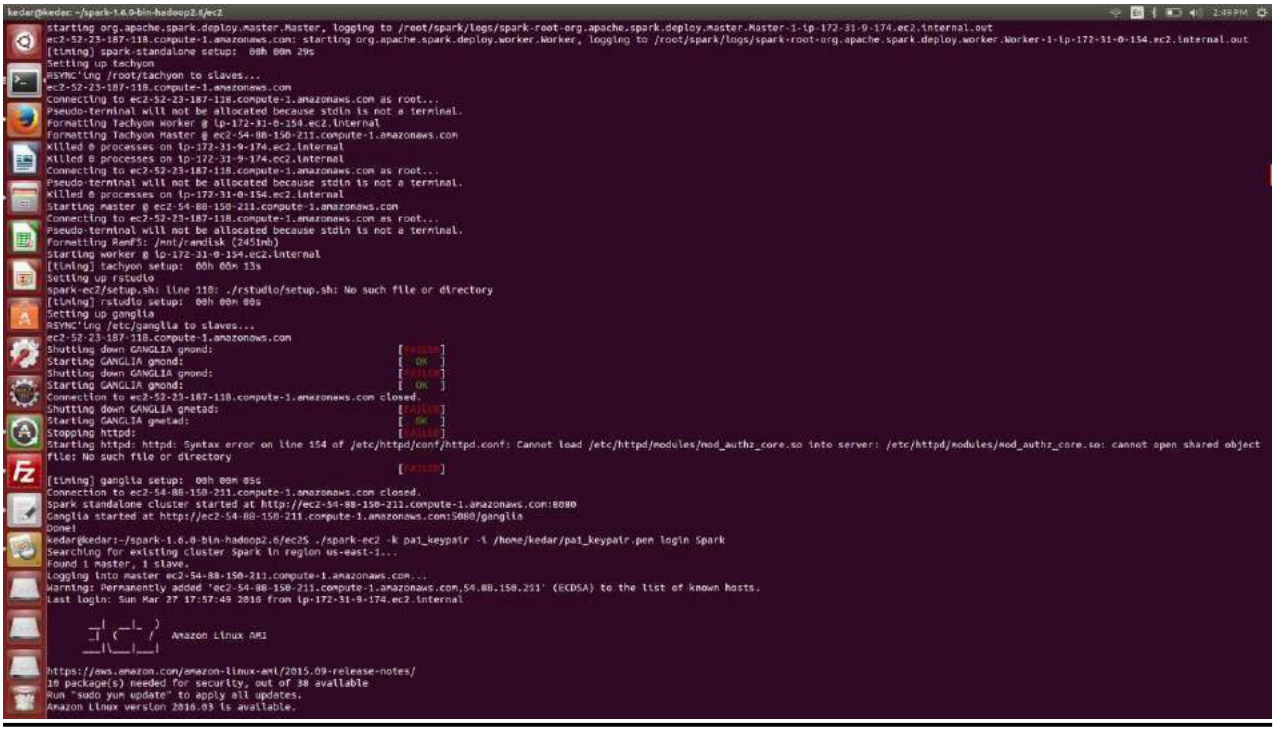
Once the command is run, the instance will be created and will be ssh ready with all Spark , Scala and Hadoop pre-installed in the cluster.



After successful installation of the Spark Virtual Cluster, our amazon AWS dashboard will show up the Spark Instances active.



The active instances can be started using the below command.
./spark-ec2 -k pa1_keypair -i /home/kedar/pa1_keypair.pem login Spark



Spark TeraSort 1 Node 10 GB

- After successful creation of spark cluster, we can run the sort program for 10GB input file. For this experiment, I created an instance with EBS Volume size of 30

  To create a 10GB data, we use GenSort

  ./gensort –a 100000000 input.txt

- Once data is created we create an input folder in the Hadoop File System HDFS where the data created by gensort is moved.

  As shown in the above screenshot, the input folder in the HDFS can be created by,

  On the ephemeral-hdfs/bin path, run the below commands

  ./hadoop dfs -mkdir /usr

  ./hadoop dfs -mkdir /usr/sparkusr

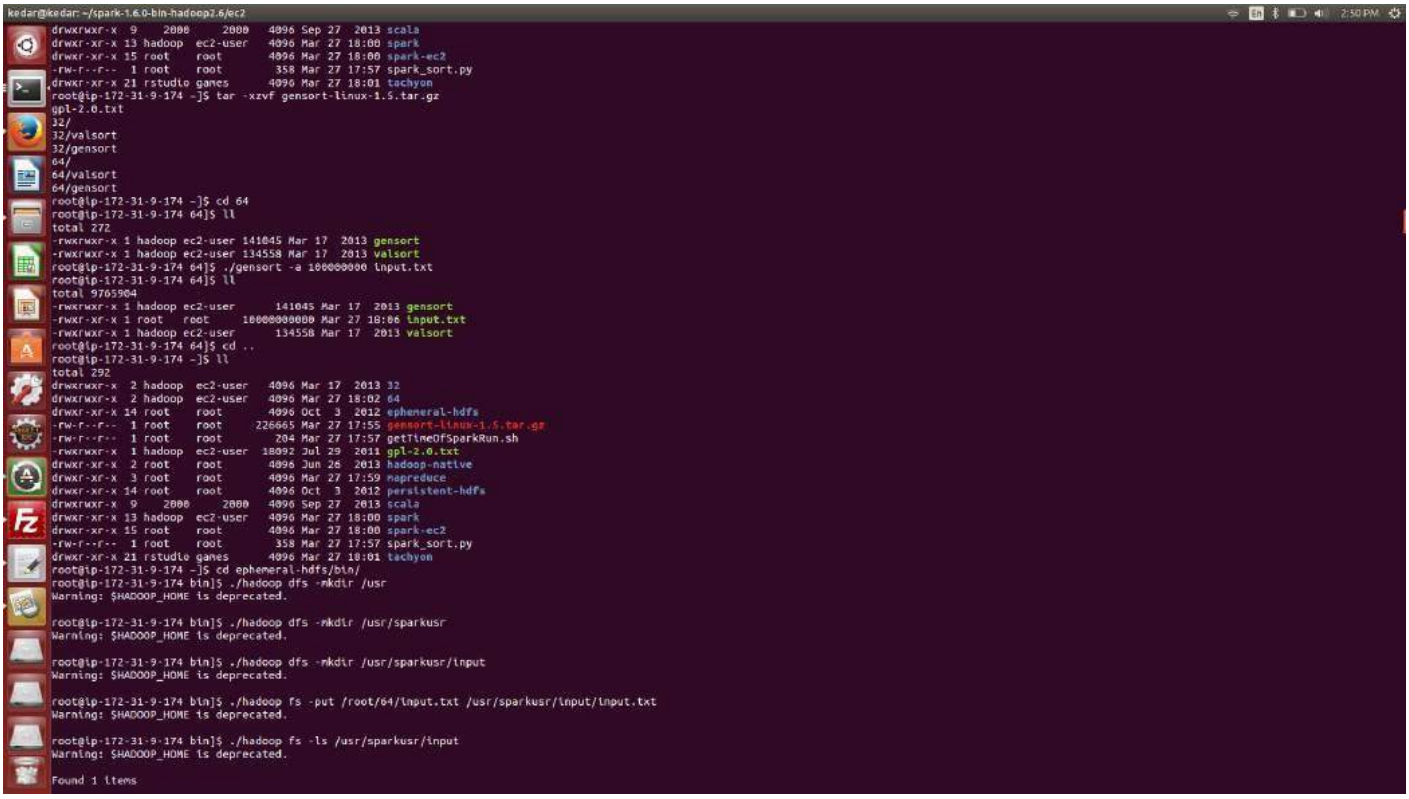
  ./hadoop dfs -mkdir /usr/sparkusr/input

  The data can be moved from the EBS volume to the gensort folder using command,

./hadoop fs -put /root/64/input.txt /usr/sparkusr/input/input.txt

Once moved the file can be listed on the HDFS using command,

./hadoop fs -ls /usr/sparkusr/input



- Once the unsorted 10GB of data is transferred to the HDFS file system, we can start the sort program.
  The sort commands on spark have been written in Python language.
  Spark has python compiler which it uses to execute python scripts.

- We use SortByKey () command for sorting the 10GB of data.
  To sort the data and calculate the time used in sorting the data I created a Shell Script that can be executed.
- The Shell script (getTimeOfSparkRun.sh) on executing will perform the following command,

- Display Start Time of the Sort Run
- Sort the Entire data using Python Commands
  The Python script spark.py contains the python commands to sort the data.
  The Python Script can be run using the command,

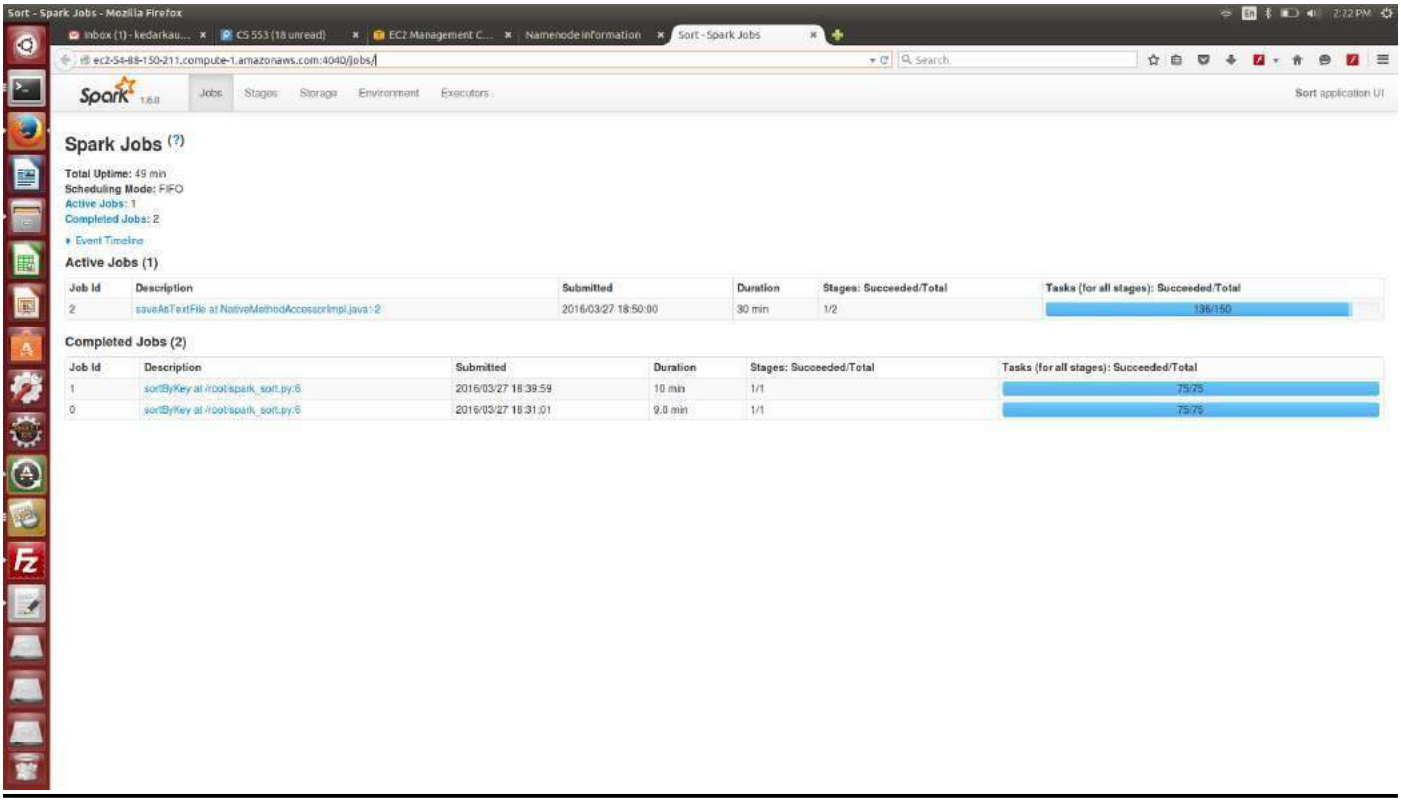
  ./spark-submit /root/sort_spark.py

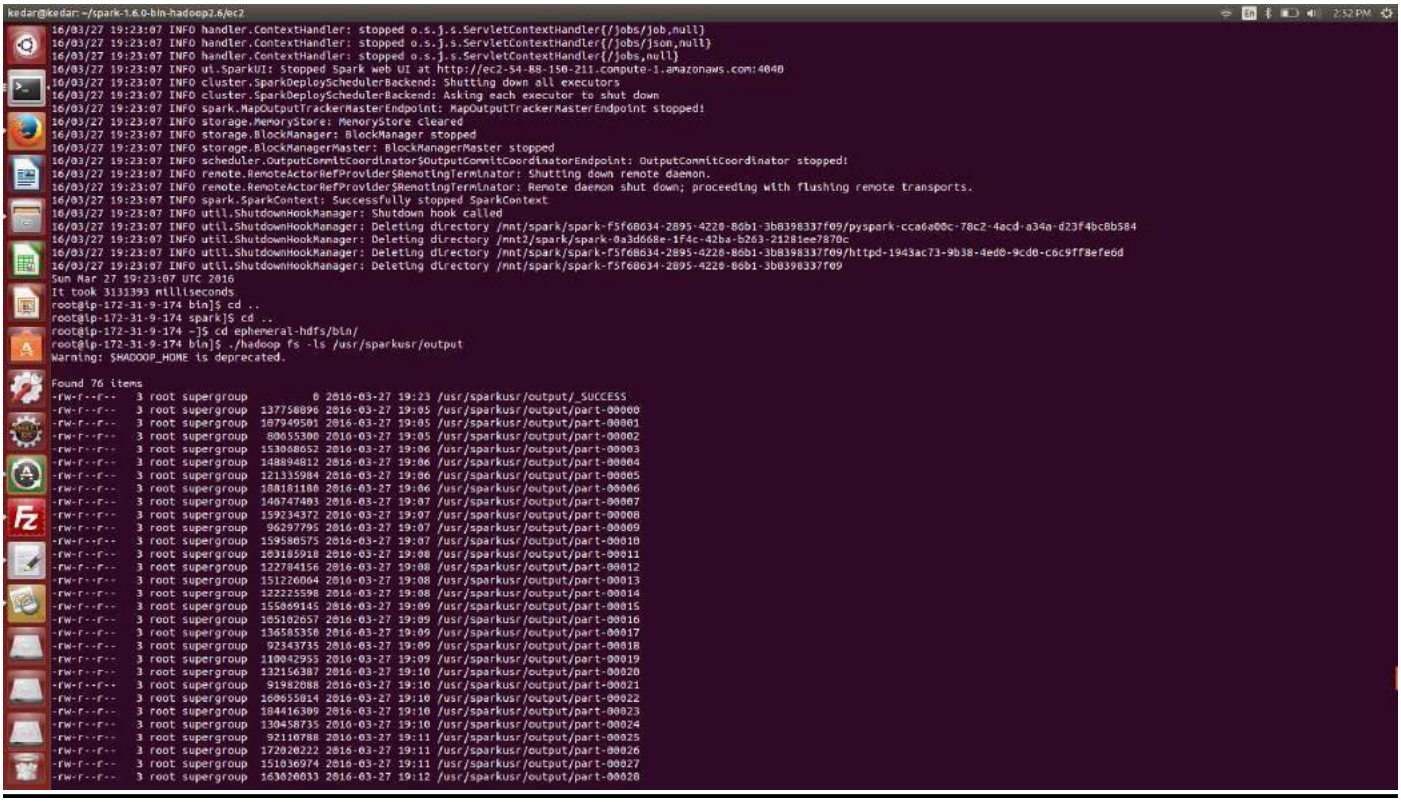
- Display the Actual time in Milliseconds took to sort the Data.

```
kedar@kedar: ~/spark-1.6.0-bin-hadoop2.6/ec2
total 288
drwxrwxr-x  2 hadoop  ec2-user   4096 Mar 17  2013 32
drwxrwxr-x  2 hadoop  ec2-user   4096 Mar 27 18:11 64
drwxr-xr-x 14 root    root       4096 Oct  3  2012 ephemeral-hdfs
-rw-r--r--  1 root    root     226665 Mar 27 17:55 gensort-linux-1.5.tar.gz
-rwxrwxr-x  1 hadoop  ec2-user  18092 Jul 29  2011 gpl-2.0.txt
drwxr-xr-x  2 root    root       4096 Jun 26  2013 hadoop-native
drwxr-xr-x  3 root    root       4096 Mar 27 17:59 mapreduce
drwxr-xr-x 14 root    root       4096 Oct  3  2012 persistent-hdfs
drwxrwxr-x  9    2006     2000  4096 Sep 27  2013 scala
drwxr-xr-x 13 hadoop  ec2-user   4096 Mar 27 18:00 spark
drwxr-xr-x 15 root    root       4096 Mar 27 18:00 spark-ec2
-rw-r--r--  1 root    root        358 Mar 27 18:18 spark_sort.py
drwxr-xr-x 21 rstudio games      4096 Mar 27 18:01 tachyon
root@ip-172-31-9-174 ~]$ vi spark_sort.py
root@ip-172-31-9-174 ~]$ cd spark/bin/
root@ip-172-31-9-174 bin]$ ./getTimeOfSparkRun.sh
Sun Mar 27 18:30:56 UTC 2016
16/03/27 18:30:57 INFO spark.SparkContext: Running Spark version 1.6.0
16/03/27 18:30:58 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.

16/03/27 18:30:58 INFO spark.SecurityManager: Changing view acls to: root
16/03/27 18:30:58 INFO spark.SecurityManager: Changing modify acls to: root
16/03/27 18:30:58 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(root); users with modify permissions: Set(root)
16/03/27 18:30:58 INFO util.Utils: Successfully started service 'sparkDriver' on port 36445.
16/03/27 18:30:58 INFO slf4j.Slf4jLogger: Slf4jLogger started
16/03/27 18:30:59 INFO Remoting: Starting remoting
16/03/27 18:30:59 INFO Remoting: Remoting started; listening on addresses :[akka.tcp://sparkDriverActorSystem@172.31.9.174:58491]
16/03/27 18:30:59 INFO util.Utils: Successfully started service 'sparkDriverActorSystem' on port 58491.
16/03/27 18:30:59 INFO spark.SparkEnv: Registering MapOutputTracker
16/03/27 18:30:59 INFO spark.SparkEnv: Registering BlockManagerMaster
16/03/27 18:30:59 INFO storage.DiskBlockManager: Created local directory at /mnt/spark/blockmgr-a94d884f-f03f-4f08-9bef-d4934ad9c9c6
16/03/27 18:30:59 INFO storage.DiskBlockManager: Created local directory at /mnt2/spark/blockmgr-4f513578-5c7a-454c-835f-4e4fa57b8821
16/03/27 18:30:59 INFO storage.MemoryStore: MemoryStore started with capacity 511.5 MB
16/03/27 18:30:59 INFO spark.SparkEnv: Registering OutputCommitCoordinator
16/03/27 18:30:59 INFO server.Server: jetty-8.y.z-SNAPSHOT
16/03/27 18:30:59 INFO server.AbstractConnector: Started SelectChannelConnector@0.0.0.0:4040
16/03/27 18:30:59 INFO util.Utils: Successfully started service 'SparkUI' on port 4040.
16/03/27 18:30:59 INFO ui.SparkUI: Started SparkUI at http://ec2-54-88-150-211.compute-1.amazonaws.com:4040
16/03/27 18:30:59 INFO spark.HttpFileServer: HTTP File server directory is /mnt/spark/spark-f5f68634-2895-4220-86b1-3b8398337f09/httpd-1943ac73-9b38-4ed0-9cd0-c6c9ff8efe6d
16/03/27 18:30:59 INFO spark.HttpServer: Starting HTTP Server
16/03/27 18:30:59 INFO server.Server: jetty-8.y.z-SNAPSHOT
16/03/27 18:30:59 INFO server.AbstractConnector: Started SocketConnector@0.0.0.0:33457
16/03/27 18:30:59 INFO util.Utils: Successfully started service 'HTTP file server' on port 33457.
16/03/27 18:30:59 INFO util.Utils: Copying /root/spark_sort.py to /mnt/spark/spark-f5f68634-2895-4220-86b1-3b8398337f09/userFiles-60dc02ef-737b-4c69-bba5-aa16d40d2378/spark_sort.py
16/03/27 18:30:59 INFO spark.SparkContext: Added file file:/root/spark_sort.py at http://172.31.9.174:33457/files/spark_sort.py with timestamp 1459103459638
16/03/27 18:30:59 INFO client.AppClient$ClientEndpoint: Connecting to master spark://ec2-54-88-150-211.compute-1.amazonaws.com:7077...
16/03/27 18:30:59 INFO cluster.SparkDeploySchedulerBackend: Connected to Spark cluster with app ID app-20160327183059-0001
16/03/27 18:30:59 INFO util.Utils: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port 49026.
16/03/27 18:30:59 INFO netty.NettyBlockTransferService: Server created on 49026
16/03/27 18:30:59 INFO storage.BlockManagerMaster: Trying to register BlockManager
```

```
kedar@kedar: ~/spark-1.6.0-bin-hadoop2.6/ec2
16/03/27 18:39:59 INFO scheduler.TaskSetManager: Starting task 8.0 in stage 1.0 (TID 75, ip-172-31-0-154.ec2.internal, partition 8,NODE_LOCAL, 2240 bytes)
16/03/27 18:39:59 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 1.0 (TID 76, ip-172-31-0-154.ec2.internal, partition 1,NODE_LOCAL, 2240 bytes)
16/03/27 18:39:59 INFO storage.BlockManagerInfo: Added broadcast_2_piece0 in memory on ip-172-31-0-154.ec2.internal:56666 (size: 3.7 KB, free: 1536.7 MB)
16/03/27 18:40:15 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 1.0 (TID 77, ip-172-31-0-154.ec2.internal, partition 2,NODE_LOCAL, 2240 bytes)
16/03/27 18:40:15 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 1.0 (TID 75) in 15712 ms on ip-172-31-0-154.ec2.internal (1/75)
16/03/27 18:40:15 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 1.0 (TID 78, ip-172-31-0-154.ec2.internal, partition 3,NODE_LOCAL, 2240 bytes)
16/03/27 18:40:15 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 1.0 (TID 76) in 16093 ms on ip-172-31-0-154.ec2.internal (2/75)
16/03/27 18:40:31 INFO scheduler.TaskSetManager: Starting task 4.0 in stage 1.0 (TID 79, ip-172-31-0-154.ec2.internal, partition 4,NODE_LOCAL, 2240 bytes)
16/03/27 18:40:31 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 1.0 (TID 77) in 15769 ms on ip-172-31-0-154.ec2.internal (3/75)
16/03/27 18:40:32 INFO scheduler.TaskSetManager: Starting task 5.0 in stage 1.0 (TID 80, ip-172-31-0-154.ec2.internal, partition 5,NODE_LOCAL, 2240 bytes)
16/03/27 18:40:32 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 1.0 (TID 78) in 16311 ms on ip-172-31-0-154.ec2.internal (4/75)
16/03/27 18:40:47 INFO scheduler.TaskSetManager: Starting task 6.0 in stage 1.0 (TID 81, ip-172-31-0-154.ec2.internal, partition 6,NODE_LOCAL, 2240 bytes)
16/03/27 18:40:47 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 1.0 (TID 79) in 16276 ms on ip-172-31-0-154.ec2.internal (5/75)
16/03/27 18:40:48 INFO scheduler.TaskSetManager: Starting task 7.0 in stage 1.0 (TID 82, ip-172-31-0-154.ec2.internal, partition 7,NODE_LOCAL, 2240 bytes)
16/03/27 18:40:48 INFO scheduler.TaskSetManager: Finished task 5.0 in stage 1.0 (TID 80) in 15872 ms on ip-172-31-0-154.ec2.internal (6/75)
16/03/27 18:41:03 INFO scheduler.TaskSetManager: Starting task 8.0 in stage 1.0 (TID 83, ip-172-31-0-154.ec2.internal, partition 8,NODE_LOCAL, 2240 bytes)
16/03/27 18:41:03 INFO scheduler.TaskSetManager: Finished task 6.0 in stage 1.0 (TID 81) in 15911 ms on ip-172-31-0-154.ec2.internal (7/75)
16/03/27 18:41:04 INFO scheduler.TaskSetManager: Starting task 9.0 in stage 1.0 (TID 84, ip-172-31-0-154.ec2.internal, partition 9,NODE_LOCAL, 2240 bytes)
16/03/27 18:41:04 INFO scheduler.TaskSetManager: Finished task 7.0 in stage 1.0 (TID 82) in 15927 ms on ip-172-31-0-154.ec2.internal (8/75)
16/03/27 18:41:19 INFO scheduler.TaskSetManager: Starting task 10.0 in stage 1.0 (TID 85, ip-172-31-0-154.ec2.internal, partition 10,NODE_LOCAL, 2240 bytes)
16/03/27 18:41:19 INFO scheduler.TaskSetManager: Finished task 8.0 in stage 1.0 (TID 83) in 16270 ms on ip-172-31-0-154.ec2.internal (9/75)
16/03/27 18:41:19 INFO scheduler.TaskSetManager: Starting task 11.0 in stage 1.0 (TID 86, ip-172-31-0-154.ec2.internal, partition 11,NODE_LOCAL, 2240 bytes)
16/03/27 18:41:19 INFO scheduler.TaskSetManager: Finished task 9.0 in stage 1.0 (TID 84) in 15739 ms on ip-172-31-0-154.ec2.internal (10/75)
16/03/27 18:41:35 INFO scheduler.TaskSetManager: Starting task 12.0 in stage 1.0 (TID 87, ip-172-31-0-154.ec2.internal, partition 12,NODE_LOCAL, 2240 bytes)
16/03/27 18:41:35 INFO scheduler.TaskSetManager: Finished task 11.0 in stage 1.0 (TID 86) in 15666 ms on ip-172-31-0-154.ec2.internal (11/75)
16/03/27 18:41:36 INFO scheduler.TaskSetManager: Starting task 13.0 in stage 1.0 (TID 88, ip-172-31-0-154.ec2.internal, partition 13,NODE_LOCAL, 2240 bytes)
16/03/27 18:41:36 INFO scheduler.TaskSetManager: Finished task 10.0 in stage 1.0 (TID 85) in 16418 ms on ip-172-31-0-154.ec2.internal (12/75)
16/03/27 18:41:51 INFO scheduler.TaskSetManager: Starting task 14.0 in stage 1.0 (TID 89, ip-172-31-0-154.ec2.internal, partition 14,NODE_LOCAL, 2240 bytes)
16/03/27 18:41:51 INFO scheduler.TaskSetManager: Finished task 12.0 in stage 1.0 (TID 87) in 16071 ms on ip-172-31-0-154.ec2.internal (13/75)
16/03/27 18:41:52 INFO scheduler.TaskSetManager: Starting task 15.0 in stage 1.0 (TID 90, ip-172-31-0-154.ec2.internal, partition 15,NODE_LOCAL, 2240 bytes)
16/03/27 18:41:52 INFO scheduler.TaskSetManager: Finished task 13.0 in stage 1.0 (TID 88) in 15947 ms on ip-172-31-0-154.ec2.internal (14/75)
16/03/27 18:42:07 INFO scheduler.TaskSetManager: Starting task 16.0 in stage 1.0 (TID 91, ip-172-31-0-154.ec2.internal, partition 16,NODE_LOCAL, 2240 bytes)
16/03/27 18:42:07 INFO scheduler.TaskSetManager: Finished task 14.0 in stage 1.0 (TID 89) in 16178 ms on ip-172-31-0-154.ec2.internal (15/75)
16/03/27 18:42:08 INFO scheduler.TaskSetManager: Starting task 17.0 in stage 1.0 (TID 92, ip-172-31-0-154.ec2.internal, partition 17,NODE_LOCAL, 2240 bytes)
16/03/27 18:42:08 INFO scheduler.TaskSetManager: Finished task 15.0 in stage 1.0 (TID 90) in 15962 ms on ip-172-31-0-154.ec2.internal (16/75)
16/03/27 18:42:23 INFO scheduler.TaskSetManager: Starting task 18.0 in stage 1.0 (TID 93, ip-172-31-0-154.ec2.internal, partition 18,NODE_LOCAL, 2240 bytes)
16/03/27 18:42:23 INFO scheduler.TaskSetManager: Finished task 17.0 in stage 1.0 (TID 92) in 15820 ms on ip-172-31-0-154.ec2.internal (17/75)
16/03/27 18:42:24 INFO scheduler.TaskSetManager: Starting task 19.0 in stage 1.0 (TID 94, ip-172-31-0-154.ec2.internal, partition 19,NODE_LOCAL, 2240 bytes)
16/03/27 18:42:24 INFO scheduler.TaskSetManager: Finished task 16.0 in stage 1.0 (TID 91) in 16320 ms on ip-172-31-0-154.ec2.internal (18/75)
16/03/27 18:42:39 INFO scheduler.TaskSetManager: Starting task 20.0 in stage 1.0 (TID 95, ip-172-31-0-154.ec2.internal, partition 20,NODE_LOCAL, 2240 bytes)
16/03/27 18:42:39 INFO scheduler.TaskSetManager: Finished task 19.0 in stage 1.0 (TID 94) in 15649 ms on ip-172-31-0-154.ec2.internal (19/75)
16/03/27 18:42:40 INFO scheduler.TaskSetManager: Starting task 21.0 in stage 1.0 (TID 96, ip-172-31-0-154.ec2.internal, partition 21,NODE_LOCAL, 2240 bytes)
16/03/27 18:42:40 INFO scheduler.TaskSetManager: Finished task 18.0 in stage 1.0 (TID 93) in 16545 ms on ip-172-31-0-154.ec2.internal (20/75)
16/03/27 18:42:55 INFO scheduler.TaskSetManager: Starting task 22.0 in stage 1.0 (TID 97, ip-172-31-0-154.ec2.internal, partition 22,NODE_LOCAL, 2240 bytes)
16/03/27 18:42:55 INFO scheduler.TaskSetManager: Finished task 20.0 in stage 1.0 (TID 95) in 15391 ms on ip-172-31-0-154.ec2.internal (21/75)
16/03/27 18:42:57 INFO scheduler.TaskSetManager: Starting task 23.0 in stage 1.0 (TID 98, ip-172-31-0-154.ec2.internal, partition 23,NODE_LOCAL, 2240 bytes)
16/03/27 18:42:57 INFO scheduler.TaskSetManager: Finished task 21.0 in stage 1.0 (TID 96) in 16927 ms on ip-172-31-0-154.ec2.internal (22/75)
16/03/27 18:43:10 INFO scheduler.TaskSetManager: Starting task 24.0 in stage 1.0 (TID 99, ip-172-31-0-154.ec2.internal, partition 24,NODE_LOCAL, 2240 bytes)
16/03/27 18:43:10 INFO scheduler.TaskSetManager: Finished task 22.0 in stage 1.0 (TID 97) in 15902 ms on ip-172-31-0-154.ec2.internal (23/75)
16/03/27 18:43:13 INFO scheduler.TaskSetManager: Starting task 25.0 in stage 1.0 (TID 100, ip-172-31-0-154.ec2.internal, partition 25,NODE_LOCAL, 2240 bytes)
16/03/27 18:43:13 INFO scheduler.TaskSetManager: Finished task 23.0 in stage 1.0 (TID 98) in 15972 ms on ip-172-31-0-154.ec2.internal (24/75)
16/03/27 18:43:26 INFO scheduler.TaskSetManager: Starting task 26.0 in stage 1.0 (TID 101, ip-172-31-0-154.ec2.internal, partition 26,NODE_LOCAL, 2240 bytes)
16/03/27 18:43:26 INFO scheduler.TaskSetManager: Finished task 24.0 in stage 1.0 (TID 99) in 15840 ms on ip-172-31-0-154.ec2.internal (25/75)
16/03/27 18:43:29 INFO scheduler.TaskSetManager: Starting task 27.0 in stage 1.0 (TID 102, ip-172-31-0-154.ec2.internal, partition 27,NODE_LOCAL, 2240 bytes)
16/03/27 18:43:29 INFO scheduler.TaskSetManager: Finished task 25.0 in stage 1.0 (TID 100) in 16299 ms on ip-172-31-0-154.ec2.internal (26/75)
16/03/27 18:43:43 INFO scheduler.TaskSetManager: Starting task 28.0 in stage 1.0 (TID 103, ip-172-31-0-154.ec2.internal, partition 28,NODE_LOCAL, 2240 bytes)
16/03/27 18:43:43 INFO scheduler.TaskSetManager: Finished task 26.0 in stage 1.0 (TID 101) in 16284 ms on ip-172-31-0-154.ec2.internal (27/75)
16/03/27 18:43:45 INFO scheduler.TaskSetManager: Starting task 29.0 in stage 1.0 (TID 104, ip-172-31-0-154.ec2.internal, partition 29,NODE_LOCAL, 2240 bytes)
```

The running Spark Job can be tracked using the instance Public DNS and 4040 port number as shown in the below screenshot,





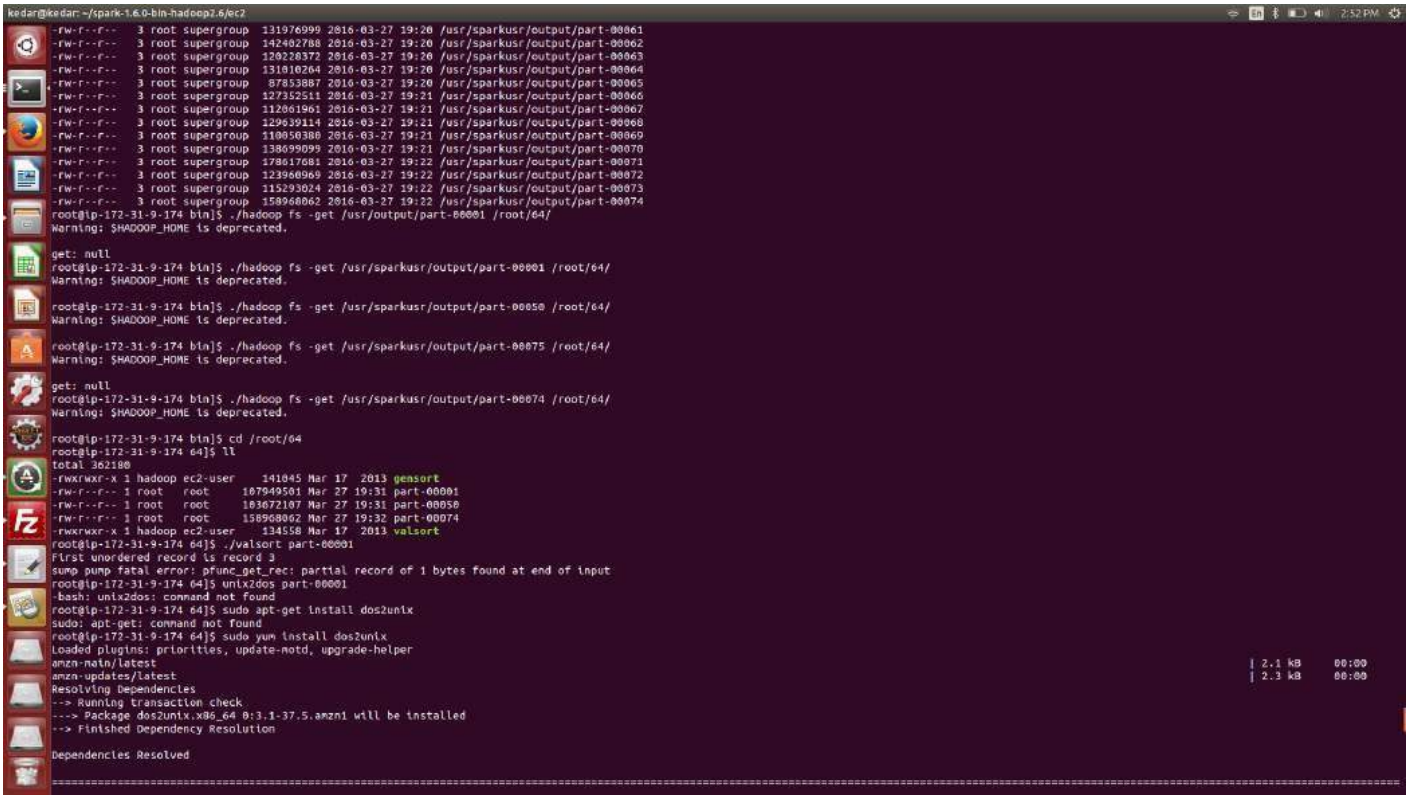
- As shown in the above figure ,

<u>The time taken to sort the 10GB of data on Spark 1 Node is 3131393 ms i.e. 52.18 Minutes.</u>

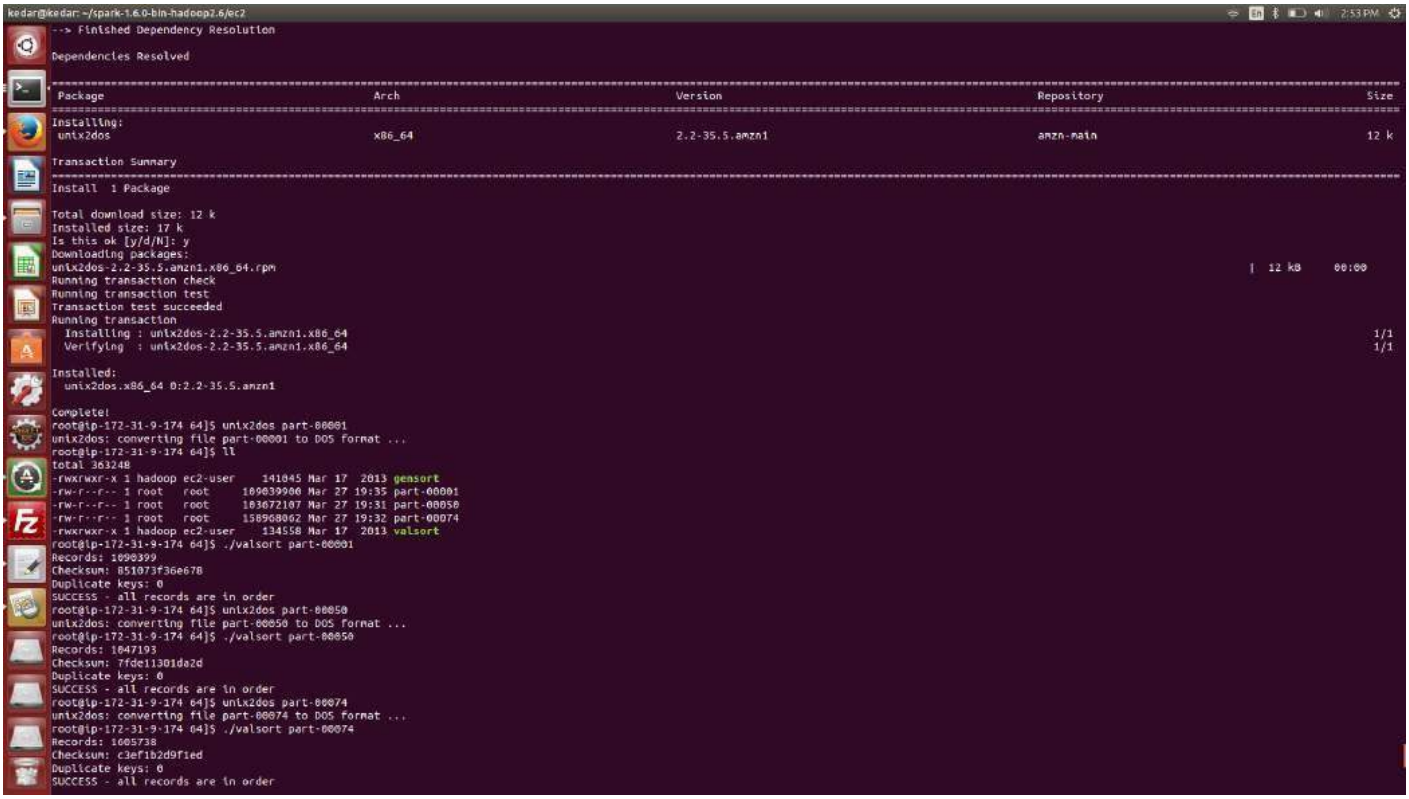
- Hadoop maps the data into 75 parts (sorted) which can be merged to get the entire sorted output file. The files Part 0 to Part 74 are present on hadoop file system which can be transferred back to the EBS Volume using the below commands,

  ./hadoop fs -get /usr/output/part* /root

- For the verification I have transferred back only three parts (randomly selected first, middle and last part) and transferred back to the EBS Volume.



- Once the data is transferred we need to convert all the output files from UNIX to DOS File Type. This can be done using unix2dos command for linux.
- Converting all the transferred parts into DOS file Type, we can evaluate the file whether it is sorted or not using valsort.



```
kedar@kedar: ~/spark-1.6.0-bin-hadoop2.6/ec2
--> Finished Dependency Resolution

Dependencies Resolved

================================================================================
 Package        Arch        Version                 Repository          Size
================================================================================
Installing:
 unix2dos       x86_64      2.2-35.5.amzn1          amzn-main           12 k

Transaction Summary
================================================================================
Install  1 Package

Total download size: 12 k
Installed size: 17 k
Is this ok [y/d/N]: y
Downloading packages:
unix2dos-2.2-35.5.amzn1.x86_64.rpm                                  |  12 kB     00:00
Running transaction check
Running transaction test
Transaction test succeeded
Running transaction
  Installing : unix2dos-2.2-35.5.amzn1.x86_64                                    1/1
  Verifying  : unix2dos-2.2-35.5.amzn1.x86_64                                    1/1

Installed:
  unix2dos.x86_64 0:2.2-35.5.amzn1

Complete!
root@ip-172-31-9-174 64]$ unix2dos part-00001
unix2dos: converting file part-00001 to DOS format ...
root@ip-172-31-9-174 64]$ ll
total 363248
-rwxrwxr-x 1 hadoop ec2-user    141045 Mar 17  2013 gensort
-rw-r--r-- 1 root   root     109039906 Mar 27 19:35 part-00001
-rw-r--r-- 1 root   root     103672107 Mar 27 19:31 part-00050
-rw-r--r-- 1 root   root     158968062 Mar 27 19:32 part-00074
-rwxrwxr-x 1 hadoop ec2-user    134558 Mar 17  2013 valsort
root@ip-172-31-9-174 64]$ ./valsort part-00001
Records: 1090399
Checksum: 851073f36e678
Duplicate keys: 0
SUCCESS - all records are in order
root@ip-172-31-9-174 64]$ unix2dos part-00050
unix2dos: converting file part-00050 to DOS format ...
root@ip-172-31-9-174 64]$ ./valsort part-00050
Records: 1047193
Checksum: 7fde11301da2d
Duplicate keys: 0
SUCCESS - all records are in order
root@ip-172-31-9-174 64]$ unix2dos part-00074
unix2dos: converting file part-00074 to DOS format ...
root@ip-172-31-9-174 64]$ ./valsort part-00074
Records: 1605738
Checksum: c3ef1b2d9f1ed
Duplicate keys: 0
SUCCESS - all records are in order
```

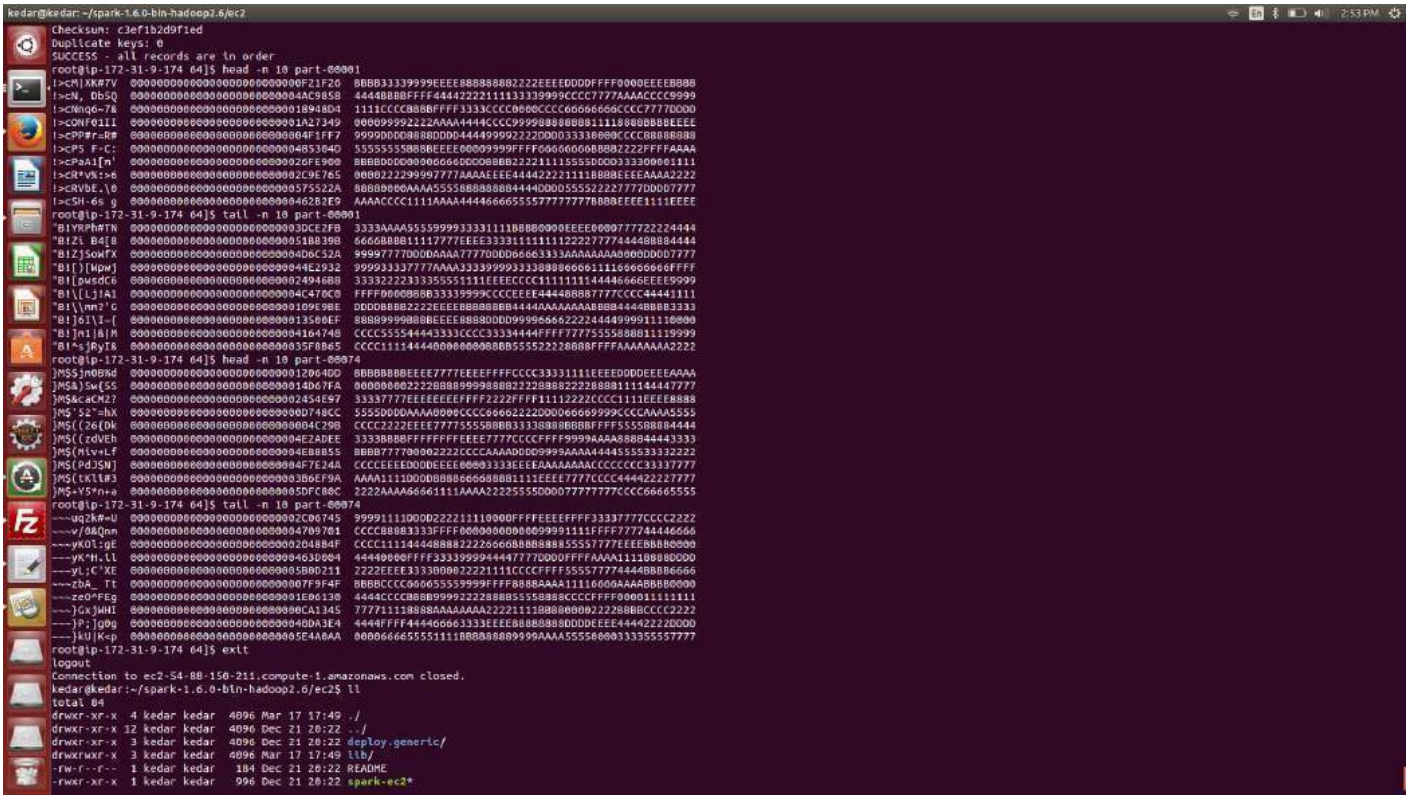
The above figure shows that the part-00001 ,part 00050 and part00074 are converted from Unix to Dos and then evaluated using valsort command as mentioned below,

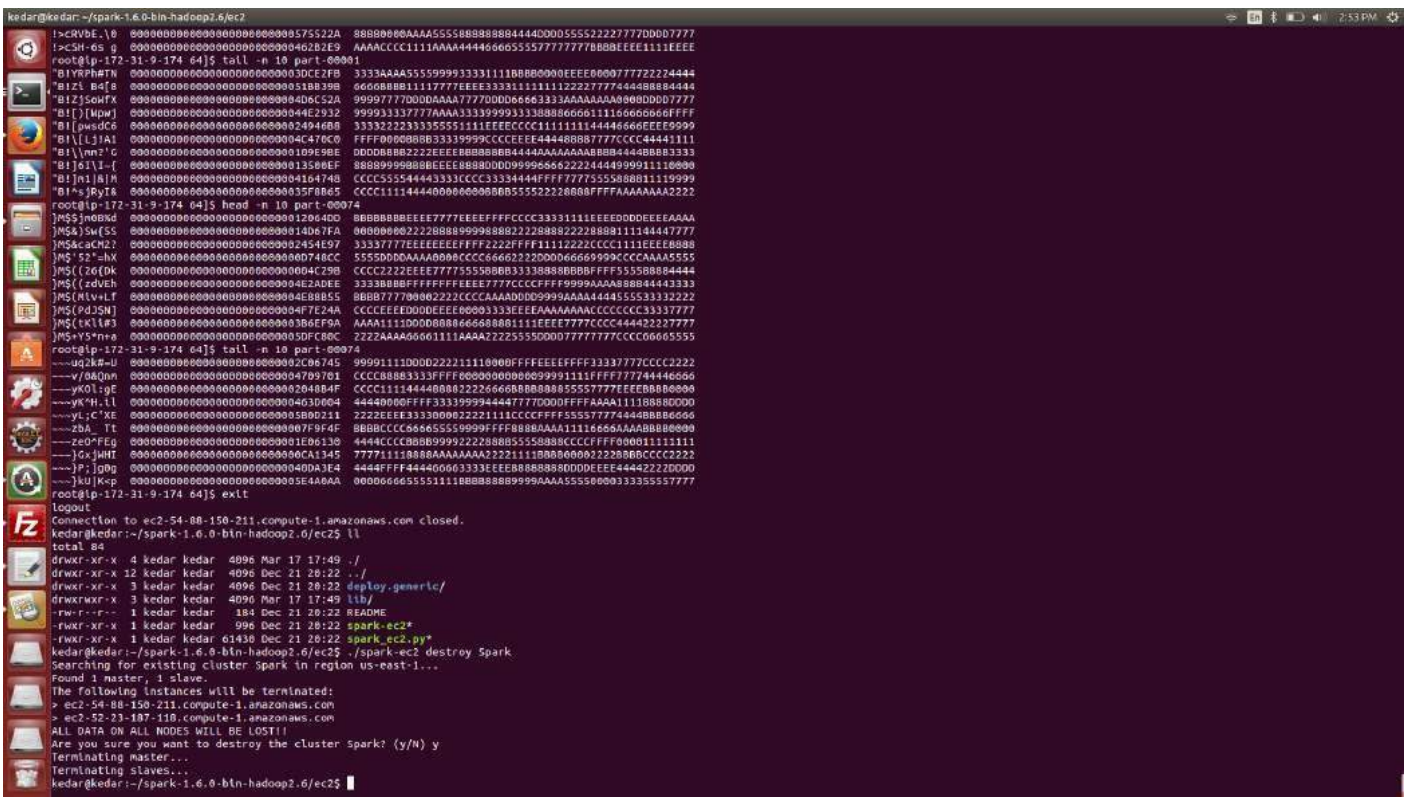
unix2dos part-00001

./valsort part-00001

The above command will check the file and output if the records present in the file are sorted or not.

As per the above screenshot, the records in all the parts are in order.





The above two screenshots represent the first 10 and last 10 lines of each part to show that they are sorted as per the keys.

Finally, the instance created can be terminated using the below command,

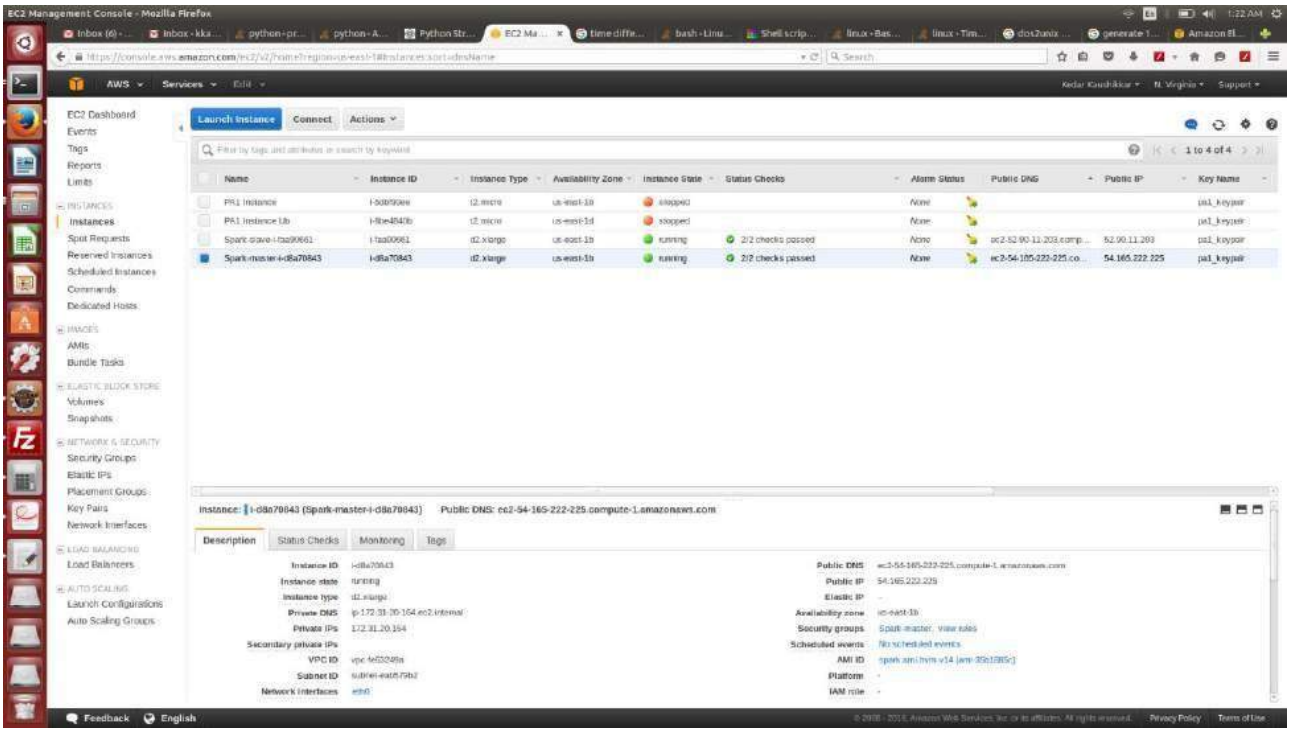
./spark-ec2 destroy [INSTANCE NAME]

Performance:

Throughput = (Datasize / Time taken)
= 10000 / 3130.8
= 3.19 Mb/sec

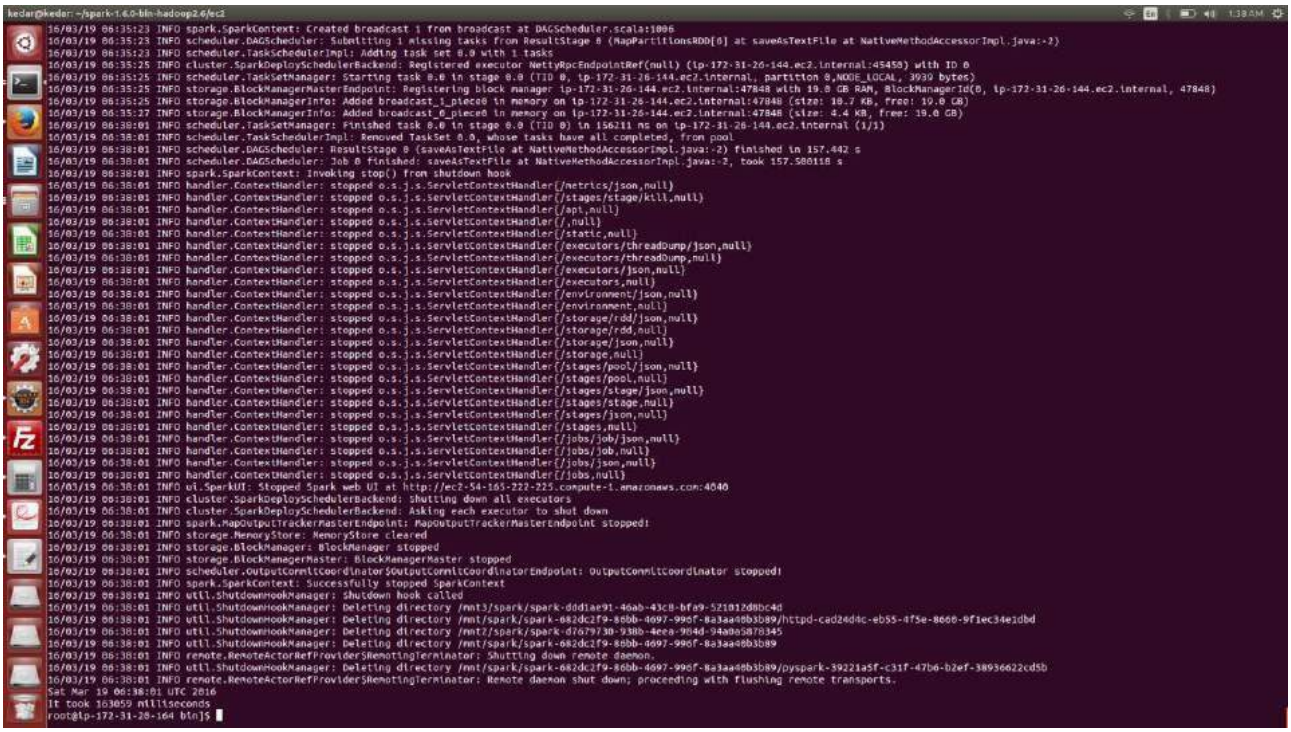
Spark TeraSort 1 Node 1 GB

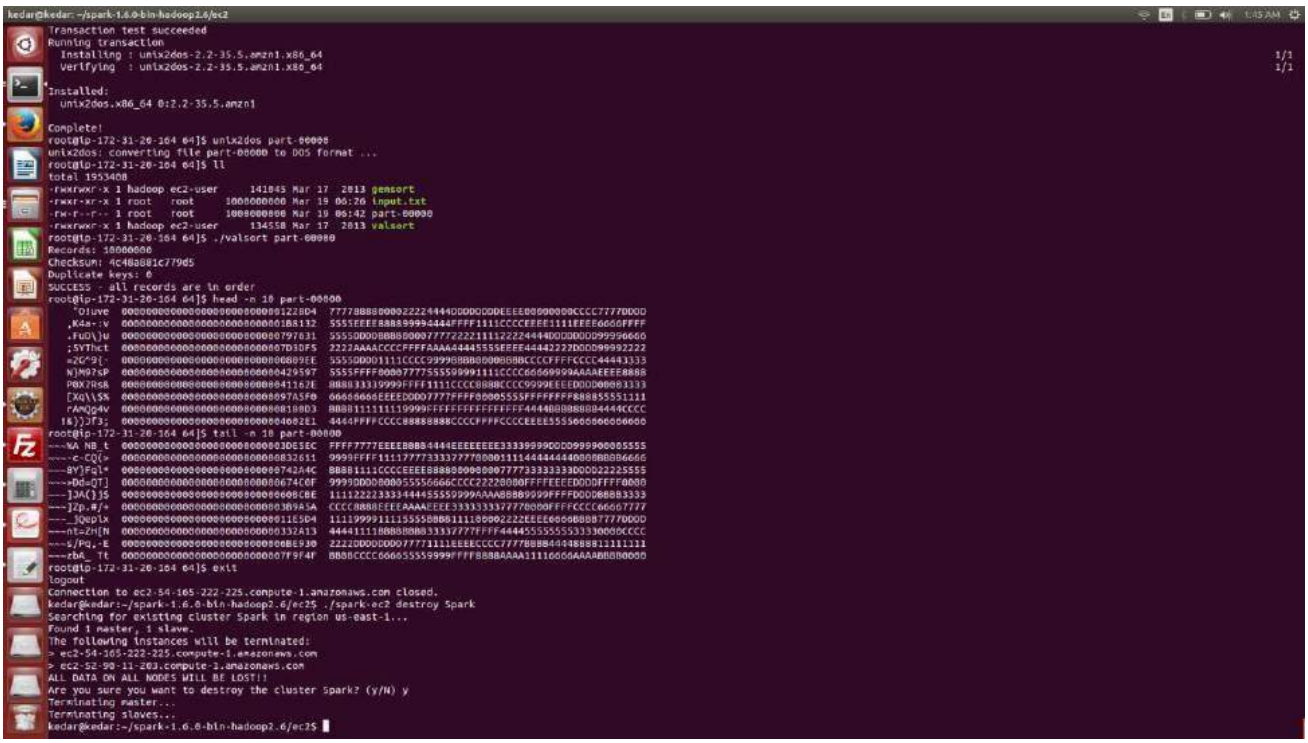
We followed the similar steps for 1GB dataset on d2.xlarge instance. Below are the screenshots for the same.



```
kedar@kedar: ~/spark-1.6.0-bin-hadoop2.6/ec2
ec2-52-90-11-203.compute-1.amazonaws.com
ec2-52-90-11-203.compute-1.amazonaws.com: no org.apache.spark.deploy.worker.Worker to stop
no org.apache.spark.deploy.master.Master to stop
starting org.apache.spark.deploy.master.Master, logging to /root/spark/logs/spark-root-org.apache.spark.deploy.master.Master-1-ip-172-31-20-164.ec2.internal.out
ec2-52-90-11-203.compute-1.amazonaws.com: starting org.apache.spark.deploy.worker.Worker, logging to /root/spark/logs/spark-root-org.apache.spark.deploy.worker.Worker-1-ip-172-31-26-144.ec2.internal.out
[timing] spark-standalone setup:  00h 00m 29s
Setting up tachyon
RSYNC'ing /root/tachyon to slaves...
ec2-52-90-11-203.compute-1.amazonaws.com
Connecting to ec2-52-90-11-203.compute-1.amazonaws.com as root...
Pseudo-terminal will not be allocated because stdin is not a terminal.
Formatting Tachyon Worker @ ip-172-31-26-144.ec2.internal
Formatting Tachyon Master @ ec2-54-165-222-225.compute-1.amazonaws.com
Killed 0 processes on ip-172-31-20-164.ec2.internal
Killed 0 processes on ip-172-31-20-164.ec2.internal
Connecting to ec2-52-90-11-203.compute-1.amazonaws.com as root...
Pseudo-terminal will not be allocated because stdin is not a terminal.
Killed 0 processes on ip-172-31-26-144.ec2.internal
Starting master @ ec2-54-165-222-225.compute-1.amazonaws.com
Connecting to ec2-52-90-11-203.compute-1.amazonaws.com as root...
Pseudo-terminal will not be allocated because stdin is not a terminal.
Formatting RamFS: /mnt/ramdisk (27341mb)
Starting worker @ ip-172-31-26-144.ec2.internal
[timing] tachyon setup:  00h 00m 14s
Setting up rstudio
spark-ec2/setup.sh: line 110: ./rstudio/setup.sh: No such file or directory
[timing] rstudio setup:  00h 00m 00s
Setting up ganglia
RSYNC'ing /etc/ganglia to slaves...
ec2-52-90-11-203.compute-1.amazonaws.com
Shutting down GANGLIA gmond:                               [FAILED]
Starting GANGLIA gmond:                                    [  OK  ]
Shutting down GANGLIA gmond:                               [FAILED]
Starting GANGLIA gmond:                                    [  OK  ]
Connection to ec2-52-90-11-203.compute-1.amazonaws.com closed.
Shutting down GANGLIA gmetad:                              [FAILED]
Starting GANGLIA gmetad:                                   [  OK  ]
Stopping httpd:                                            [FAILED]
Starting httpd:                                            [  OK  ]
[timing] ganglia setup:  00h 00m 02s
Connection to ec2-54-165-222-225.compute-1.amazonaws.com closed.
Spark standalone cluster started at http://ec2-54-165-222-225.compute-1.amazonaws.com:8080
Ganglia started at http://ec2-54-165-222-225.compute-1.amazonaws.com:5080/ganglia
Done!
kedar@kedar:~/spark-1.6.0-bin-hadoop2.6/ec2$ ./spark-ec2 -k pa1_keypair -i /home/kedar/pa1_keypair.pem login Spark
Searching for existing cluster Spark in region us-east-1...
Found 1 master, 1 slave.
Logging into master ec2-54-165-222-225.compute-1.amazonaws.com...
Warning: Permanently added 'ec2-54-165-222-225.compute-1.amazonaws.com,54.165.222.225' (ECDSA) to the list of known hosts.
Last login: Sat Mar 19 06:16:52 2016 from ip-172-31-20-164.ec2.internal

       __|  __|_  )
       _|  (     /   Amazon Linux AMI
      ___|\___|___|

https://aws.amazon.com/amazon-linux-ami/2013.03-release-notes/
Amazon Linux version 2015.09 is available.
root@ip-172-31-20-164 ~]$
```

```
kedar@kedar: ~/spark-1.6.0-bin-hadoop2.6/ec2
       __|  __|_  )
       _|  (     /   Amazon Linux AMI
      ___|\___|___|

https://aws.amazon.com/amazon-linux-ami/2013.03-release-notes/
Amazon Linux version 2015.09 is available.
root@ip-172-31-20-164 ~]$ tar -xzvf gensort-linux-1.5.tar.gz
gpl-2.0.txt
32/
32/valsort
32/gensort
64/
64/valsort
64/gensort
root@ip-172-31-20-164 ~]$ rm gensort-linux-1.5.tar.gz
root@ip-172-31-20-164 ~]$ cd 64/
root@ip-172-31-20-164 64]$ ll
total 272
-rwxrwxr-x 1 hadoop ec2-user 141045 Mar 17  2013 gensort
-rwxrwxr-x 1 hadoop ec2-user 134558 Mar 17  2013 valsort
root@ip-172-31-20-164 64]$ gensort 10000000 -a input.txt
-bash: gensort: command not found
root@ip-172-31-20-164 64]$ gensort -a 10000000 input.txt
-bash: gensort: command not found
root@ip-172-31-20-164 64]$ ./gensort -a 10000000 input.txt
root@ip-172-31-20-164 64]$ ll
total 976840
-rwxrwxr-x 1 hadoop ec2-user     141045 Mar 17  2013 gensort
-rwxr-xr-x 1 root   root      1000000000 Mar 19 06:26 input.txt
-rwxrwxr-x 1 hadoop ec2-user     134558 Mar 17  2013 valsort
root@ip-172-31-20-164 64]$ cd ..
root@ip-172-31-20-164 ~]$ cd ephemeral-hdfs/ec2
-bash: cd: ephemeral-hdfs/ec2: No such file or directory
root@ip-172-31-20-164 ~]$ cd ephemeral-hdfs/bin
root@ip-172-31-20-164 bin]$ ./hadoop dfs -mkdir /usr
Warning: $HADOOP_HOME is deprecated.

root@ip-172-31-20-164 bin]$ ./hadoop dfs -mkdir /usr/sparkusr
Warning: $HADOOP_HOME is deprecated.

root@ip-172-31-20-164 bin]$ ./hadoop dfs -mkdir /usr/sparkusr/input
Warning: $HADOOP_HOME is deprecated.

root@ip-172-31-20-164 bin]$ ./hadoop fs -put /root/64/input.txt /usr/sparkusr/input/input.txt
Warning: $HADOOP_HOME is deprecated.

root@ip-172-31-20-164 bin]$ ./hadoop fs -ls /usr/sparkusr/input
Warning: $HADOOP_HOME is deprecated.

Found 1 items
-rw-r--r--   3 root supergroup 1000000000 2016-03-19 06:28 /usr/sparkusr/input/input.txt
root@ip-172-31-20-164 bin]$ ^C
root@ip-172-31-20-164 bin]$ ^C
root@ip-172-31-20-164 bin]$ cd ..
root@ip-172-31-20-164 ephemeral-hdfs]$ cd ..
root@ip-172-31-20-164 ~]$ cd spark/bin
root@ip-172-31-20-164 bin]$
```





As shown in the above figure,

<u>The time taken to sort the 1GB of data on Spark 1 Node is 163059 ms i.e. 2.71 Minutes</u>.

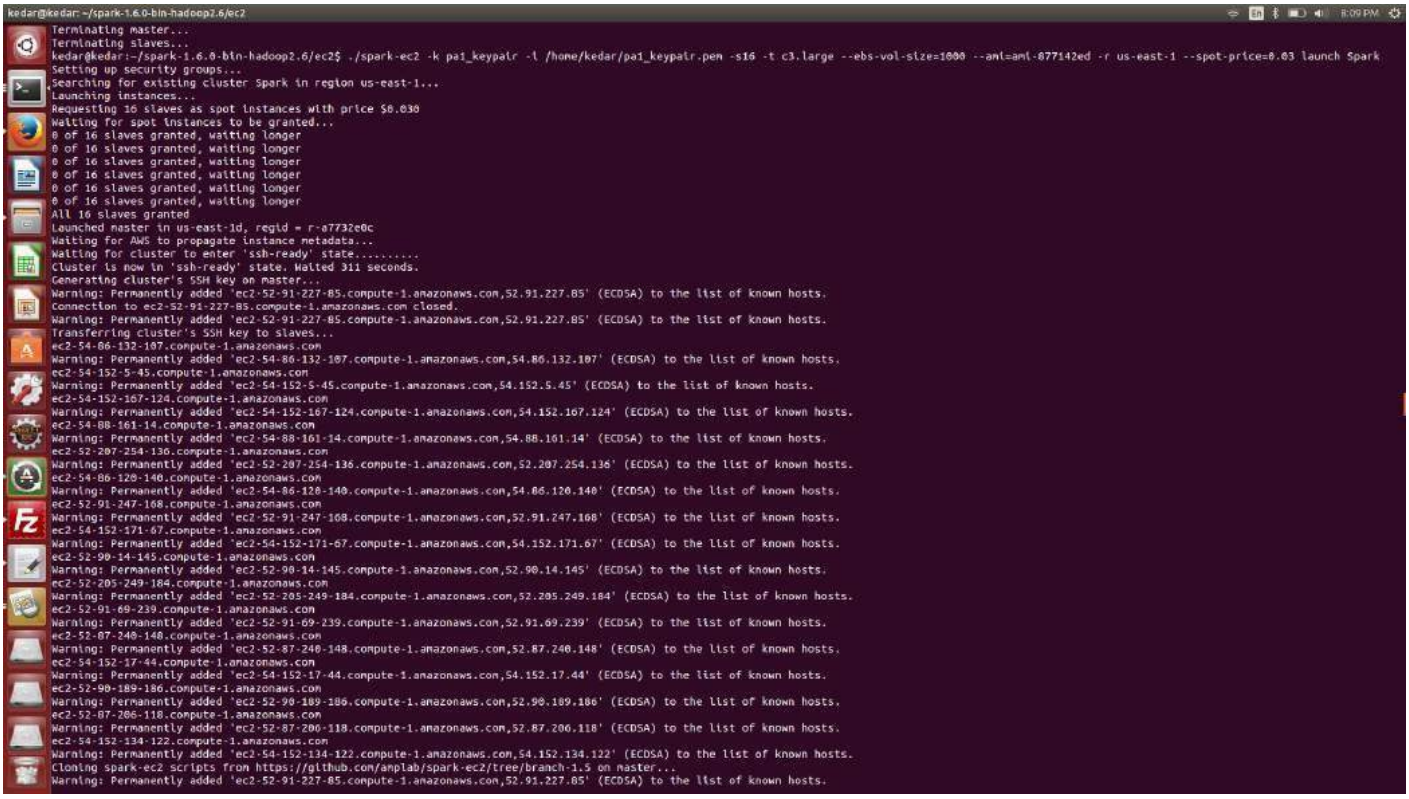
Performance:

Throughput = (Datasize / Time taken)
= 1000 / 162.6
= 6.15 Mb/sec

Spark 16 Node Virtual Cluster Setup

- The 16 node Virtual Cluster Setup can be done using the similar steps as done for 1 node but the Number of instances entered while creating the instance is 16. (s16).
  This will create a total of 17 nodes with 1 master and 16 slave node.

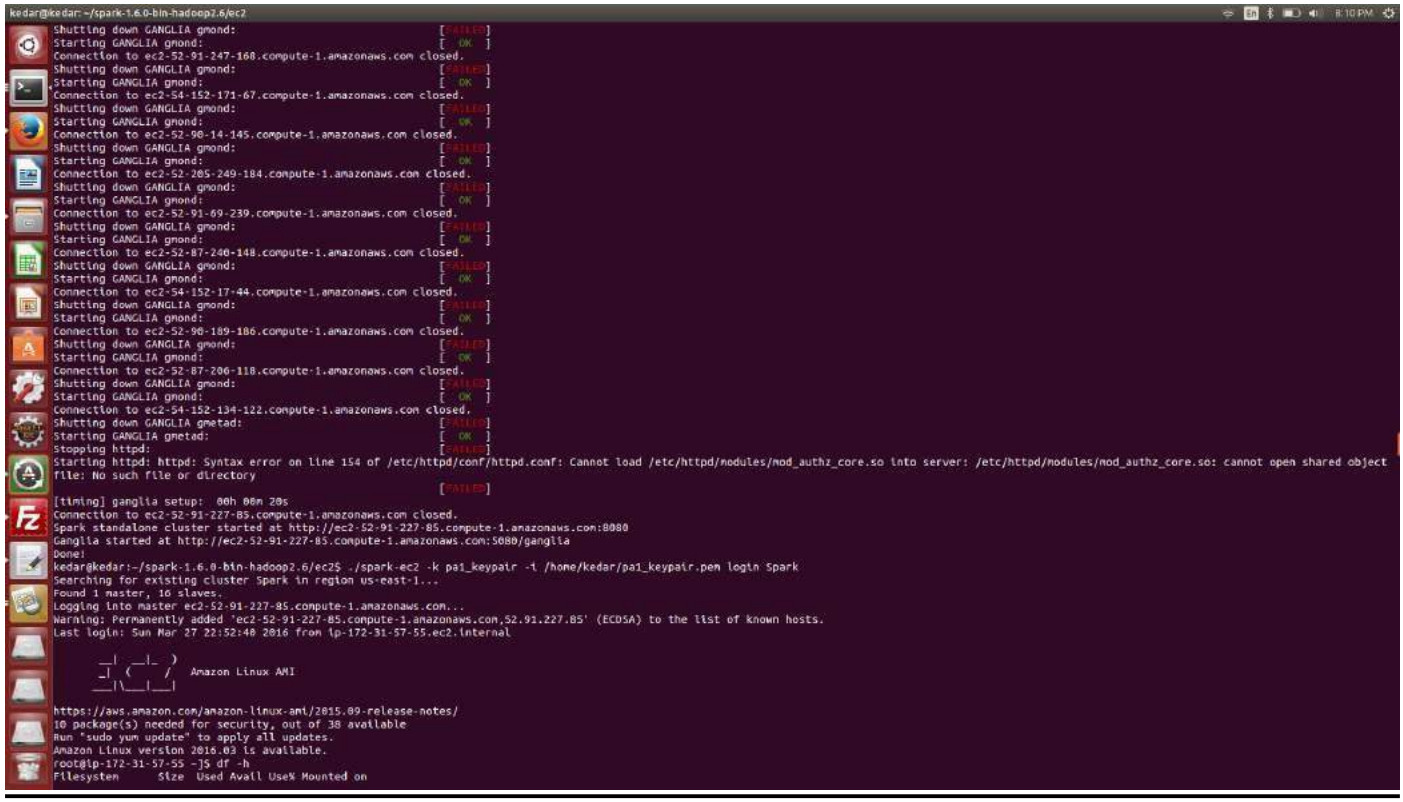
- For creating a 16 Node Virtual Cluster, we use a EBS volume of 1000 GB and spot price of 0.03.

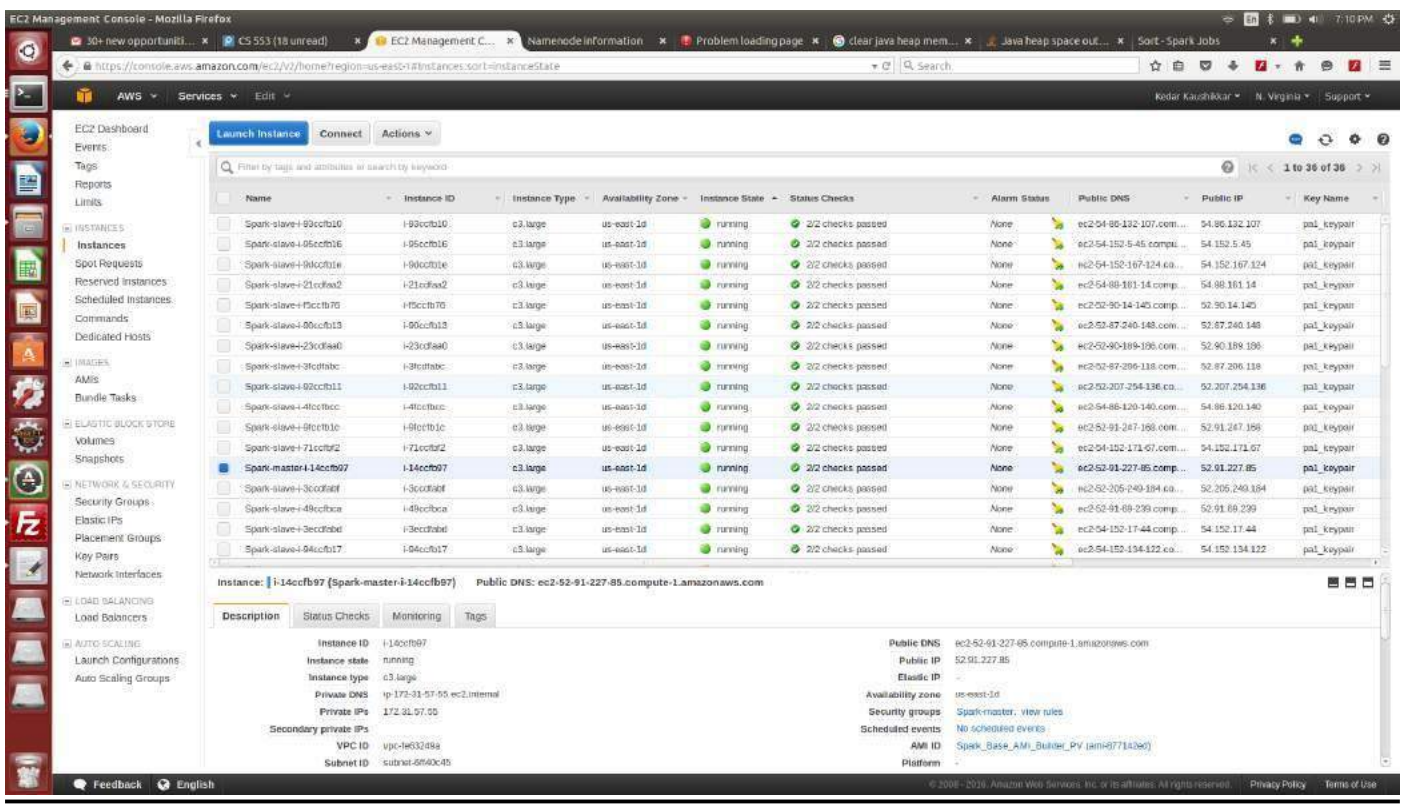


- The command to initiate creation of 16 nodes is mentioned below,

./spark-ec2 -k pa1_keypair -i /home/kedar/pa1_keypair.pem -s16 -t c3.large --ebs-vol-size=1000 --ami=ami-877142ed -r us-east-1 --spot-price=0.03 launch Spark

- Once all the instances are approved and are ssh-ready state , the amazon AWS dashboard will show 17 Virtual Clusters active with 1 Master and 16 Slave Nodes.


kedar@kedar: ~/spark-1.6.0-bin-hadoop2.6/ec2
Shutting down GANGLIA gmond: [FAILED]
Starting GANGLIA gmond: [OK]
Connection to ec2-52-91-247-168.compute-1.amazonaws.com closed.
Shutting down GANGLIA gmond: [FAILED]
Starting GANGLIA gmond: [OK]
Connection to ec2-54-152-171-67.compute-1.amazonaws.com closed.
Shutting down GANGLIA gmond: [FAILED]
Starting GANGLIA gmond: [OK]
Connection to ec2-52-90-14-145.compute-1.amazonaws.com closed.
Shutting down GANGLIA gmond: [FAILED]
Starting GANGLIA gmond: [OK]
Connection to ec2-52-205-249-184.compute-1.amazonaws.com closed.
Shutting down GANGLIA gmond: [FAILED]
Starting GANGLIA gmond: [OK]
Connection to ec2-52-91-69-239.compute-1.amazonaws.com closed.
Shutting down GANGLIA gmond: [FAILED]
Starting GANGLIA gmond: [OK]
Connection to ec2-52-87-240-148.compute-1.amazonaws.com closed.
Shutting down GANGLIA gmond: [FAILED]
Starting GANGLIA gmond: [OK]
Connection to ec2-54-152-17-44.compute-1.amazonaws.com closed.
Shutting down GANGLIA gmond: [FAILED]
Starting GANGLIA gmond: [OK]
Connection to ec2-52-90-189-186.compute-1.amazonaws.com closed.
Shutting down GANGLIA gmond: [FAILED]
Starting GANGLIA gmond: [OK]
Connection to ec2-52-87-206-118.compute-1.amazonaws.com closed.
Shutting down GANGLIA gmond: [FAILED]
Starting GANGLIA gmond: [OK]
Connection to ec2-54-152-134-122.compute-1.amazonaws.com closed.
Shutting down GANGLIA gmetad: [FAILED]
Starting GANGLIA gmetad: [OK]
Stopping httpd: [FAILED]
Starting httpd: httpd: Syntax error on line 154 of /etc/httpd/conf/httpd.conf: Cannot load /etc/httpd/modules/mod_authz_core.so into server: /etc/httpd/modules/mod_authz_core.so: cannot open shared object file: No such file or directory
[FAILED]
[timing] ganglia setup: 00h 08m 20s
Connection to ec2-52-91-227-85.compute-1.amazonaws.com closed.
Spark standalone cluster started at http://ec2-52-91-227-85.compute-1.amazonaws.com:8080
Ganglia started at http://ec2-52-91-227-85.compute-1.amazonaws.com:5080/ganglia
Done!
kedar@kedar:~/spark-1.6.0-bin-hadoop2.6/ec2$ ./spark-ec2 -k pa1_keypair -i /home/kedar/pa1_keypair.pem login Spark
Searching for existing cluster Spark in region us-east-1...
Found 1 master, 16 slaves.
Logging into master ec2-52-91-227-85.compute-1.amazonaws.com...
Warning: Permanently added 'ec2-52-91-227-85.compute-1.amazonaws.com,52.91.227.85' (ECDSA) to the list of known hosts.
Last login: Sun Mar 27 22:52:48 2016 from ip-172-31-57-55.ec2.internal
__| __|_)
_| (/ Amazon Linux AMI
___|___|___|
https://aws.amazon.com/amazon-linux-ami/2015.09-release-notes/
10 package(s) needed for security, out of 38 available
Run "sudo yum update" to apply all updates.
Amazon Linux version 2016.03 is available.
root@ip-172-31-57-55 ~]$ df -h
Filesystem Size Used Avail Use% Mounted on


EC2 Management Console - Mozilla Firefox
AWS
Services
Edit
Kedar Kaushikkar
N. Virginia
Support
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
Key Pairs
Network Interfaces
LOAD BALANCING
Load Balancers
AUTO SCALING
Launch Configurations
Auto Scaling Groups
Launch Instance
Connect
Actions
1 to 36 of 36
Name
Instance ID
Instance Type
Availability Zone
Instance State
Status Checks
Alarm Status
Public DNS
Public IP
Key Name
Spark-master-i-14ccfb97
i-14ccfb97
c3.large
us-east-1d
running
2/2 checks passed
None
ec2-52-91-227-85.comp...
52.91.227.85
pa1_keypair
Instance: i-14ccfb97 (Spark-master-i-14ccfb97) Public DNS: ec2-52-91-227-85.compute-1.amazonaws.com
Description
Status Checks
Monitoring
Tags
Instance ID i-14ccfb97
Instance state running
Instance type c3.large
Private DNS ip-172-31-57-55.ec2.internal
Private IPs 172.31.57.55
Secondary private IPs
VPC ID vpc-fe632d8a
Subnet ID subnet-6f640c45
Public DNS ec2-52-91-227-85.compute-1.amazonaws.com
Public IP 52.91.227.85
Elastic IP
Availability zone us-east-1d
Security groups Spark-master. view rules
Scheduled events No scheduled events
AMI ID Spark_Base_AMI_Builder_PV (ami-f6771a9e)
Platform
Feedback
English
Privacy Policy
Terms of Use

- Once the amazon instances are created we can connect to the master instance using the same command used previously for 1 node virtual cluster.

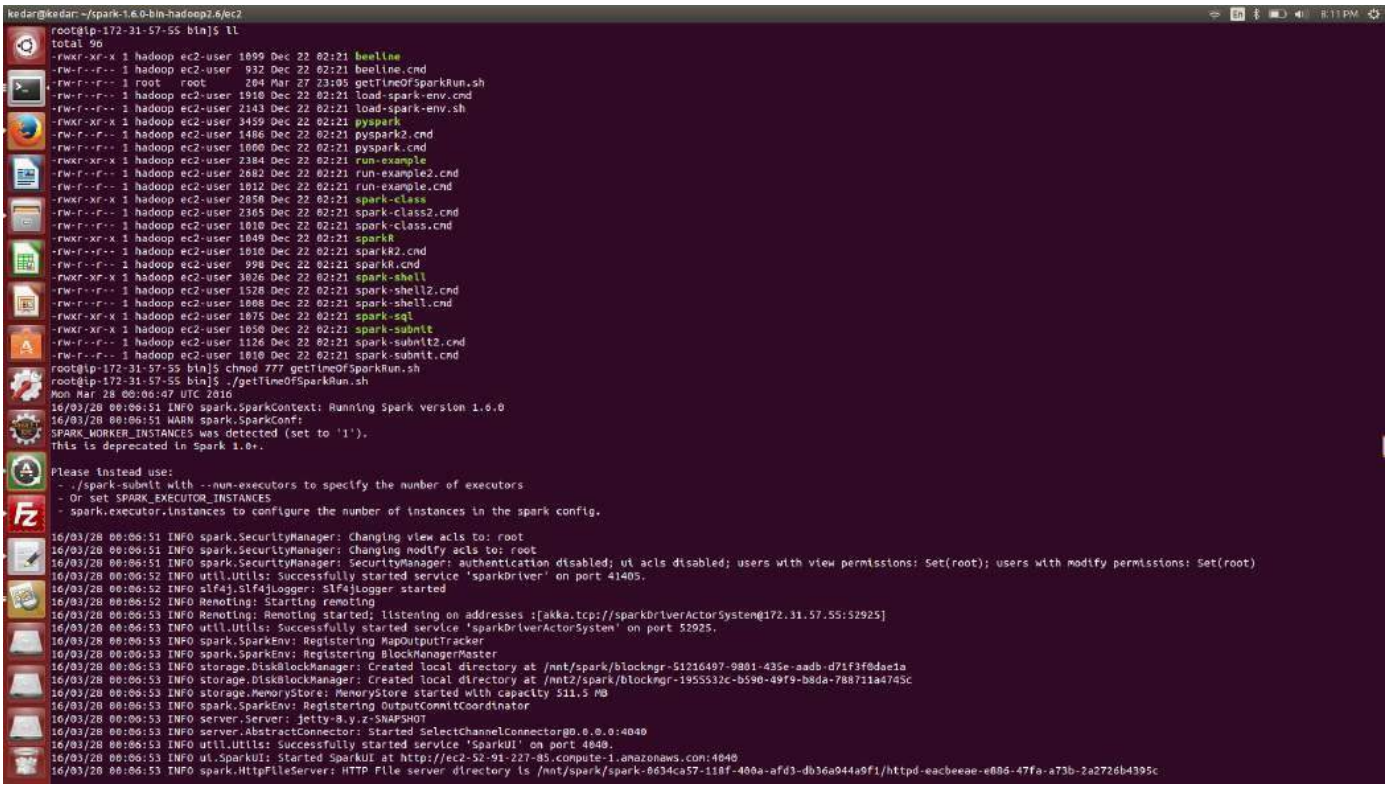
  ./spark-ec2 -k pa1_keypair -i /home/kedar/pa1_keypair.pem login Spark

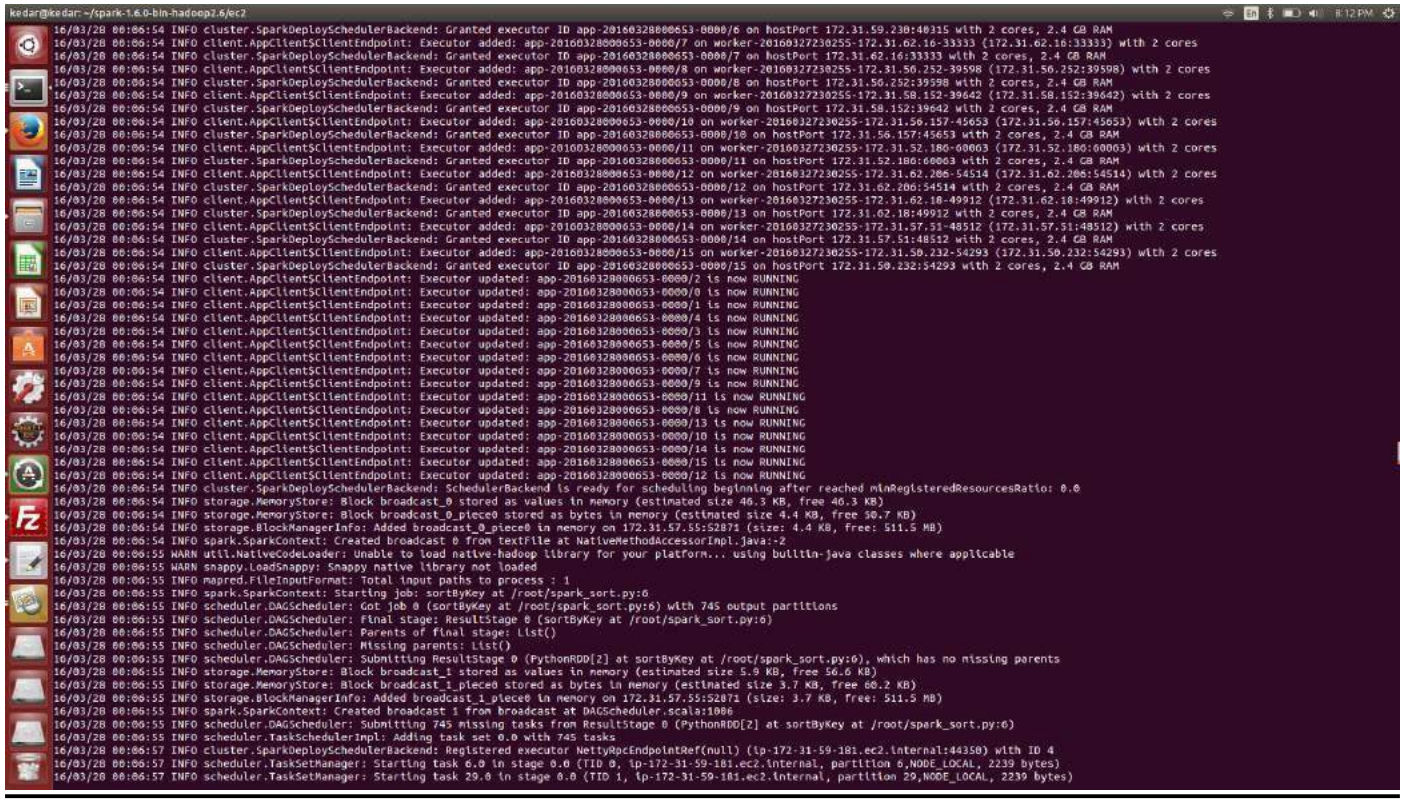
SPARK TeraSort 16 Node 100 GB

- Now follow the same steps to generate 100GB of data on the EBS volume using command,

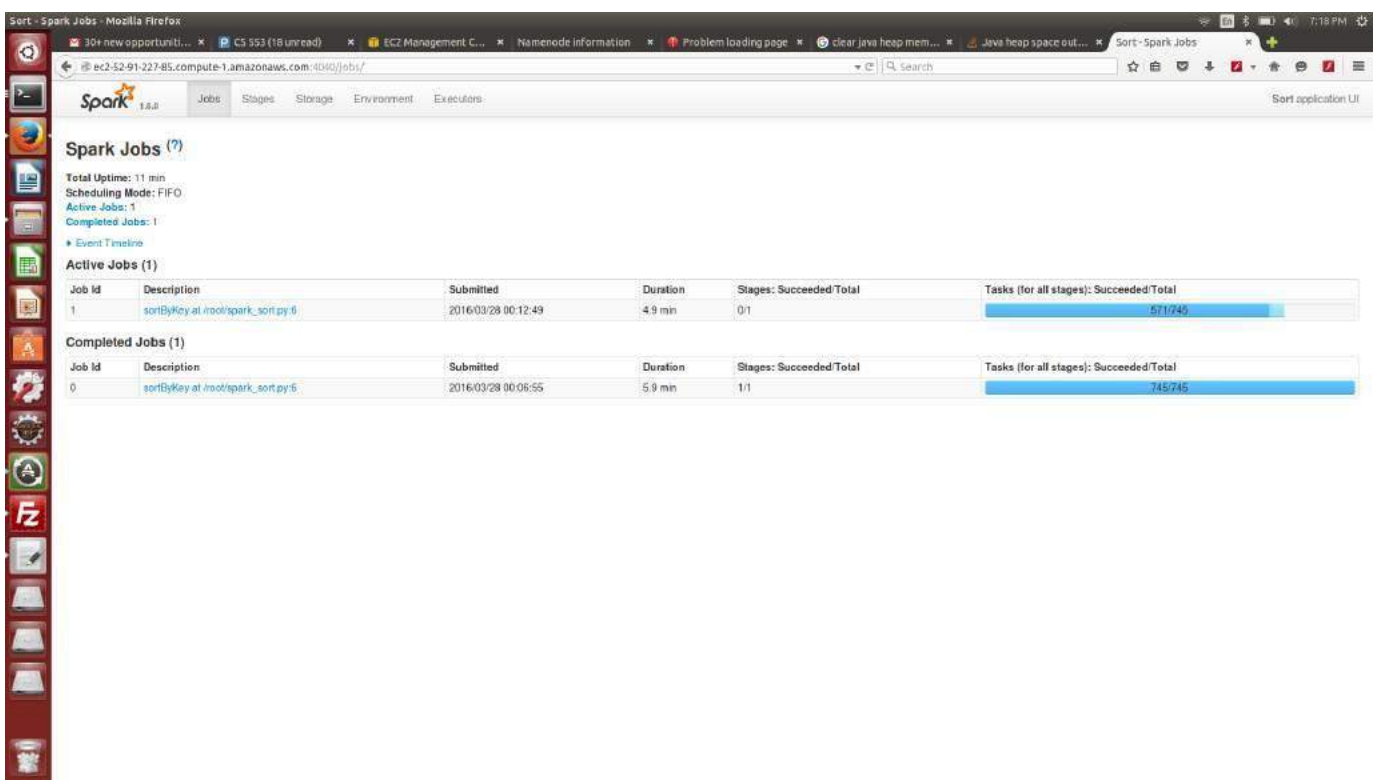
  ./gensort –a 1000000000 input.txt

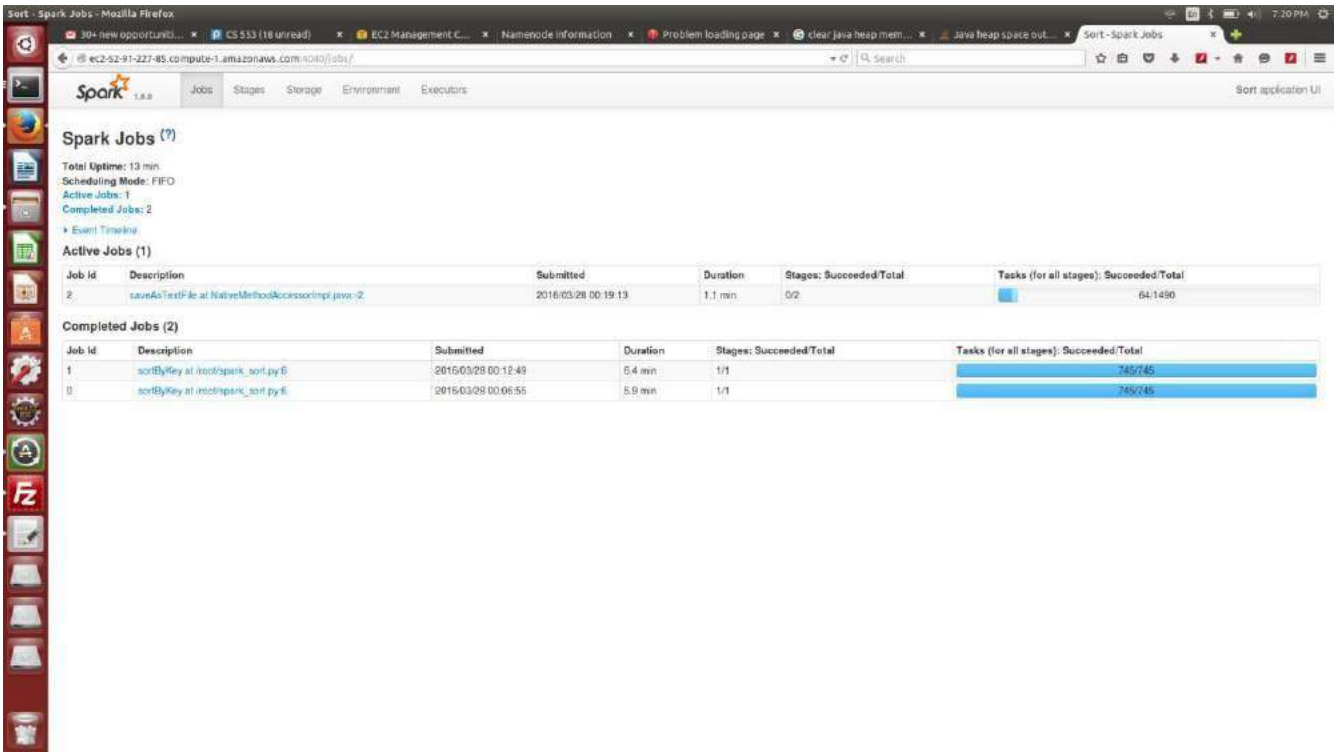
- Once the unsorted data is generated , the data can be moved at the Hadoop HDFS using the same command as of 1 Node VC. Follow the same execution steps and get the output sorted file parts back to the EBS volume.

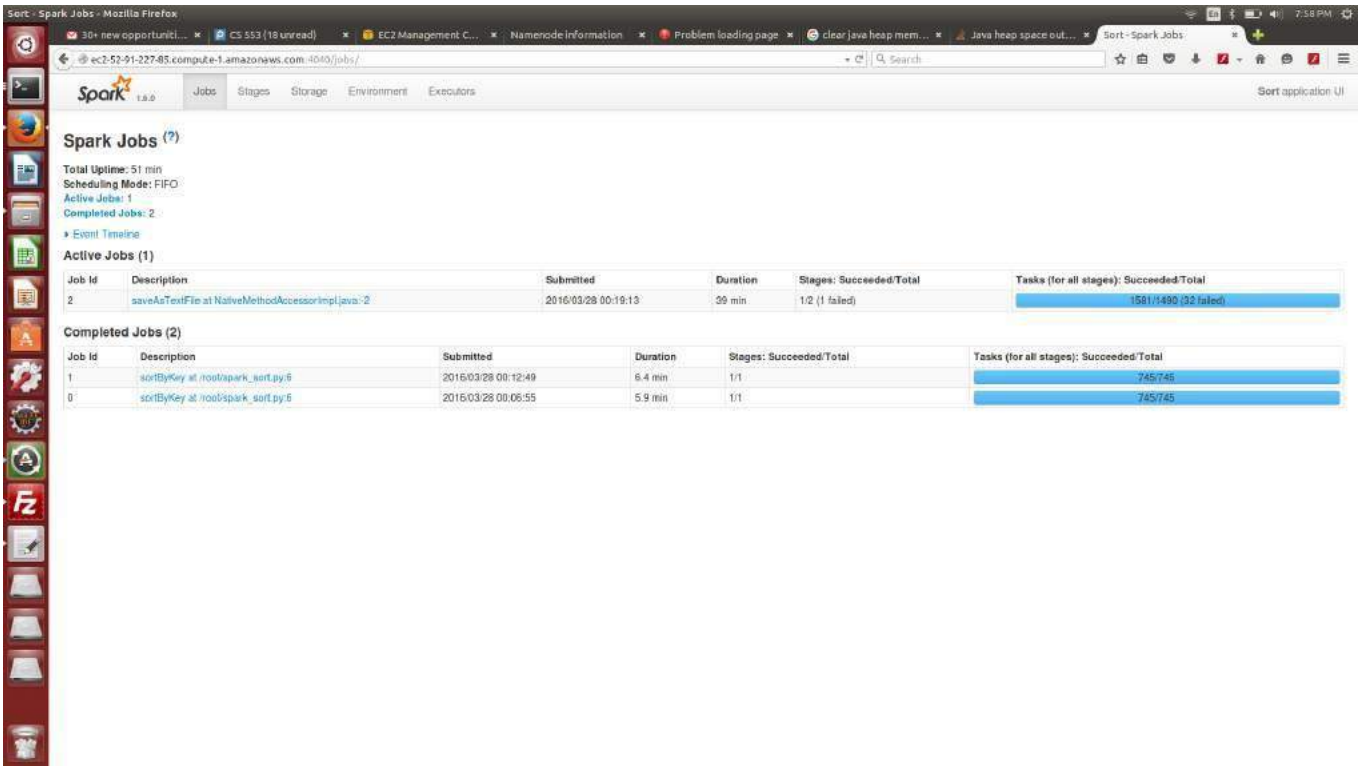


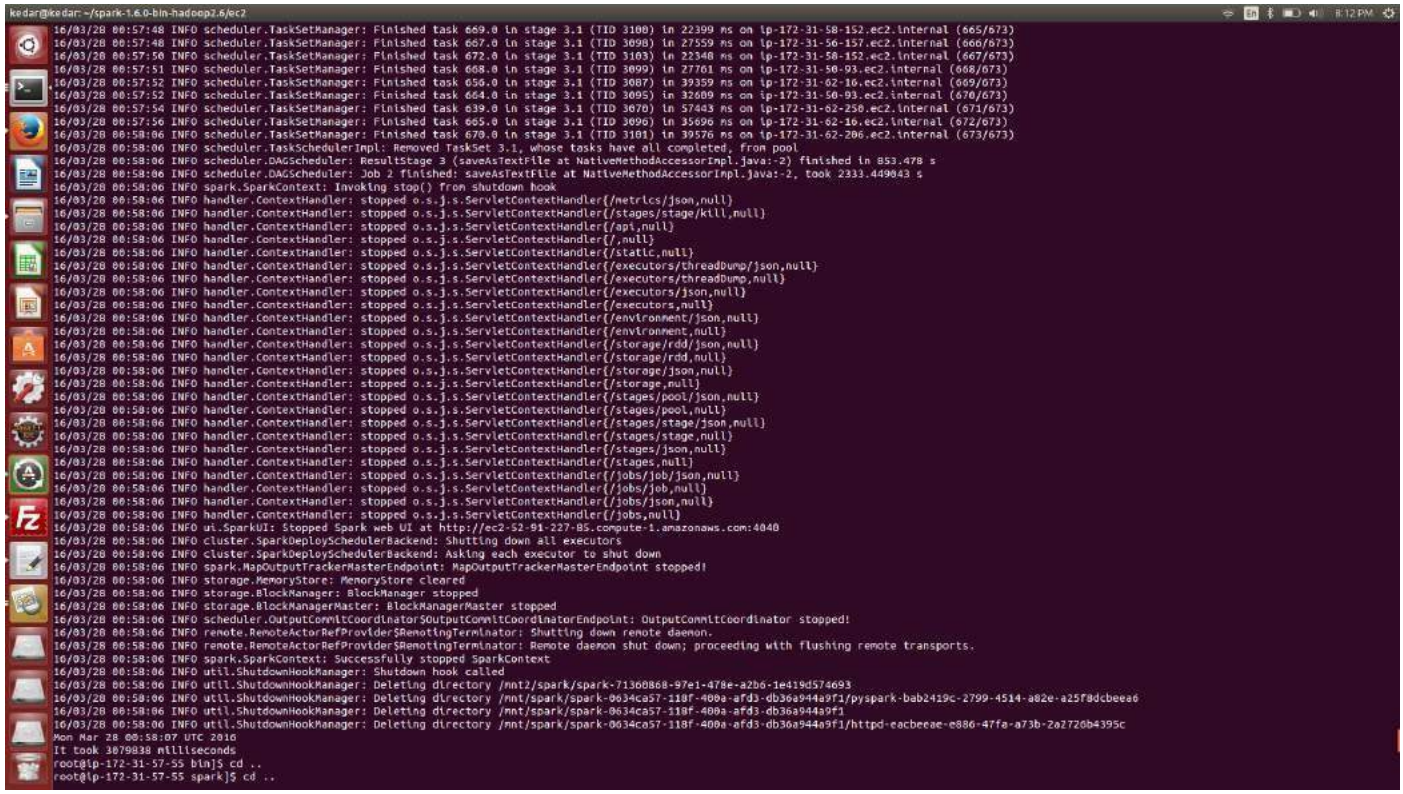


The sorting job running on the amazon aws can be tracked using the instance public DNS and 4040 port number.




Sort - Spark Jobs - Mozilla Firefox
ec2-52-91-227-85.compute-1.amazonaws.com:4040/jobs/
Spark 1.6.0
Jobs
Stages
Storage
Environment
Executors
Sort application UI
Spark Jobs (?)
Total Uptime: 13 min
Scheduling Mode: FIFO
Active Jobs: 1
Completed Jobs: 2
Event Timeline
Active Jobs (1)
Job Id | Description | Submitted | Duration | Stages: Succeeded/Total | Tasks (for all stages): Succeeded/Total
2 | saveAsTextFile at NativeMethodAccessorImpl.java:-2 | 2016/03/28 00:19:13 | 1.1 min | 0/2 | 64/1490
Completed Jobs (2)
Job Id | Description | Submitted | Duration | Stages: Succeeded/Total | Tasks (for all stages): Succeeded/Total
1 | sortByKey at /root/spark_sort.py:6 | 2016/03/28 00:12:49 | 6.4 min | 1/1 | 745/745
0 | sortByKey at /root/spark_sort.py:6 | 2016/03/28 00:06:55 | 5.9 min | 1/1 | 745/745


Sort - Spark Jobs - Mozilla Firefox
ec2-52-91-227-85.compute-1.amazonaws.com:4040/jobs/
Spark 1.6.0
Jobs
Stages
Storage
Environment
Executors
Sort application UI
Spark Jobs (?)
Total Uptime: 51 min
Scheduling Mode: FIFO
Active Jobs: 1
Completed Jobs: 2
Event Timeline
Active Jobs (1)
Job Id | Description | Submitted | Duration | Stages: Succeeded/Total | Tasks (for all stages): Succeeded/Total
2 | saveAsTextFile at NativeMethodAccessorImpl.java:-2 | 2016/03/28 00:19:13 | 39 min | 1/2 (1 failed) | 1581/1490 (32 failed)
Completed Jobs (2)
Job Id | Description | Submitted | Duration | Stages: Succeeded/Total | Tasks (for all stages): Succeeded/Total
1 | sortByKey at /root/spark_sort.py:6 | 2016/03/28 00:12:49 | 6.4 min | 1/1 | 745/745
0 | sortByKey at /root/spark_sort.py:6 | 2016/03/28 00:06:55 | 5.9 min | 1/1 | 745/745



<u>As per the above screenshot, the 16 node spark cluster sorting the data of 100 GB took 3079838 ms, i.e. 51.33 Minutes.</u>

- The output generated is in 745 parts as the hadoop map jobs are 745. To evaluate whether all those output parts are sorted we use valsort method.

  To evaluate, I exported few parts out of the 745 , and evaluated those parts using valsort

  ./hadoop fs -get /usr/output/part* /root/64

  This is mentioned in the below screenshot,

```
kedar@kedar: ~/spark-1.6.0-bin-hadoop2.6/ec2
-rw-r--r--   3 root supergroup  147575835 2016-03-28 00:56 /usr/sparkusr/output/part-00704
-rw-r--r--   3 root supergroup  121415580 2016-03-28 00:56 /usr/sparkusr/output/part-00705
-rw-r--r--   3 root supergroup  102222252 2016-03-28 00:56 /usr/sparkusr/output/part-00706
-rw-r--r--   3 root supergroup  101472327 2016-03-28 00:56 /usr/sparkusr/output/part-00707
-rw-r--r--   3 root supergroup  141059259 2016-03-28 00:56 /usr/sparkusr/output/part-00708
-rw-r--r--   3 root supergroup  110276892 2016-03-28 00:57 /usr/sparkusr/output/part-00709
-rw-r--r--   3 root supergroup   94730328 2016-03-28 00:57 /usr/sparkusr/output/part-00710
-rw-r--r--   3 root supergroup  190037449 2016-03-28 00:57 /usr/sparkusr/output/part-00711
-rw-r--r--   3 root supergroup  172395333 2016-03-28 00:57 /usr/sparkusr/output/part-00712
-rw-r--r--   3 root supergroup  124293708 2016-03-28 00:57 /usr/sparkusr/output/part-00713
-rw-r--r--   3 root supergroup  120305097 2016-03-28 00:57 /usr/sparkusr/output/part-00714
-rw-r--r--   3 root supergroup  107904258 2016-03-28 00:57 /usr/sparkusr/output/part-00715
-rw-r--r--   3 root supergroup   79425027 2016-03-28 00:57 /usr/sparkusr/output/part-00716
-rw-r--r--   3 root supergroup  127654461 2016-03-28 00:57 /usr/sparkusr/output/part-00717
-rw-r--r--   3 root supergroup   78449877 2016-03-28 00:57 /usr/sparkusr/output/part-00718
-rw-r--r--   3 root supergroup  128487051 2016-03-28 00:57 /usr/sparkusr/output/part-00719
-rw-r--r--   3 root supergroup  132851565 2016-03-28 00:57 /usr/sparkusr/output/part-00720
-rw-r--r--   3 root supergroup  116209764 2016-03-28 00:57 /usr/sparkusr/output/part-00721
-rw-r--r--   3 root supergroup  135995607 2016-03-28 00:57 /usr/sparkusr/output/part-00722
-rw-r--r--   3 root supergroup  150924114 2016-03-28 00:57 /usr/sparkusr/output/part-00723
-rw-r--r--   3 root supergroup  147515445 2016-03-28 00:57 /usr/sparkusr/output/part-00724
-rw-r--r--   3 root supergroup  143295075 2016-03-28 00:57 /usr/sparkusr/output/part-00725
-rw-r--r--   3 root supergroup   77077440 2016-03-28 00:57 /usr/sparkusr/output/part-00726
-rw-r--r--   3 root supergroup  124562493 2016-03-28 00:57 /usr/sparkusr/output/part-00727
-rw-r--r--   3 root supergroup  163677096 2016-03-28 00:57 /usr/sparkusr/output/part-00728
-rw-r--r--   3 root supergroup  147236958 2016-03-28 00:57 /usr/sparkusr/output/part-00729
-rw-r--r--   3 root supergroup   92219985 2016-03-28 00:57 /usr/sparkusr/output/part-00730
-rw-r--r--   3 root supergroup  143139942 2016-03-28 00:57 /usr/sparkusr/output/part-00731
-rw-r--r--   3 root supergroup  122086503 2016-03-28 00:57 /usr/sparkusr/output/part-00732
-rw-r--r--   3 root supergroup  117310545 2016-03-28 00:57 /usr/sparkusr/output/part-00733
-rw-r--r--   3 root supergroup  101297809 2016-03-28 00:57 /usr/sparkusr/output/part-00734
-rw-r--r--   3 root supergroup  111680316 2016-03-28 00:57 /usr/sparkusr/output/part-00735
-rw-r--r--   3 root supergroup  148061331 2016-03-28 00:57 /usr/sparkusr/output/part-00736
-rw-r--r--   3 root supergroup  166205358 2016-03-28 00:57 /usr/sparkusr/output/part-00737
-rw-r--r--   3 root supergroup  113725557 2016-03-28 00:57 /usr/sparkusr/output/part-00738
-rw-r--r--   3 root supergroup  176472549 2016-03-28 00:57 /usr/sparkusr/output/part-00739
-rw-r--r--   3 root supergroup  117385587 2016-03-28 00:57 /usr/sparkusr/output/part-00740
-rw-r--r--   3 root supergroup  108249669 2016-03-28 00:57 /usr/sparkusr/output/part-00741
-rw-r--r--   3 root supergroup  185911506 2016-03-28 00:57 /usr/sparkusr/output/part-00742
-rw-r--r--   3 root supergroup  113241645 2016-03-28 00:57 /usr/sparkusr/output/part-00743
-rw-r--r--   3 root supergroup  114428952 2016-03-28 00:57 /usr/sparkusr/output/part-00744
root@ip-172-31-57-55 bin]$ ./hadoop fs -get /usr/output/part-00001 /root/64/
Warning: $HADOOP_HOME is deprecated.

get: null
root@ip-172-31-57-55 bin]$ ./hadoop fs -get /usr/sparkusr/output/part-00001 /root/64/
Warning: $HADOOP_HOME is deprecated.

root@ip-172-31-57-55 bin]$ ./hadoop fs -get /usr/sparkusr/output/part-00250 /root/64/
Warning: $HADOOP_HOME is deprecated.

root@ip-172-31-57-55 bin]$ ./hadoop fs -get /usr/sparkusr/output/part-00500 /root/64/
Warning: $HADOOP_HOME is deprecated.

root@ip-172-31-57-55 bin]$ ./hadoop fs -get /usr/sparkusr/output/part-00744 /root/64/
Warning: $HADOOP_HOME is deprecated.

root@ip-172-31-57-55 bin]$ cd ..
```

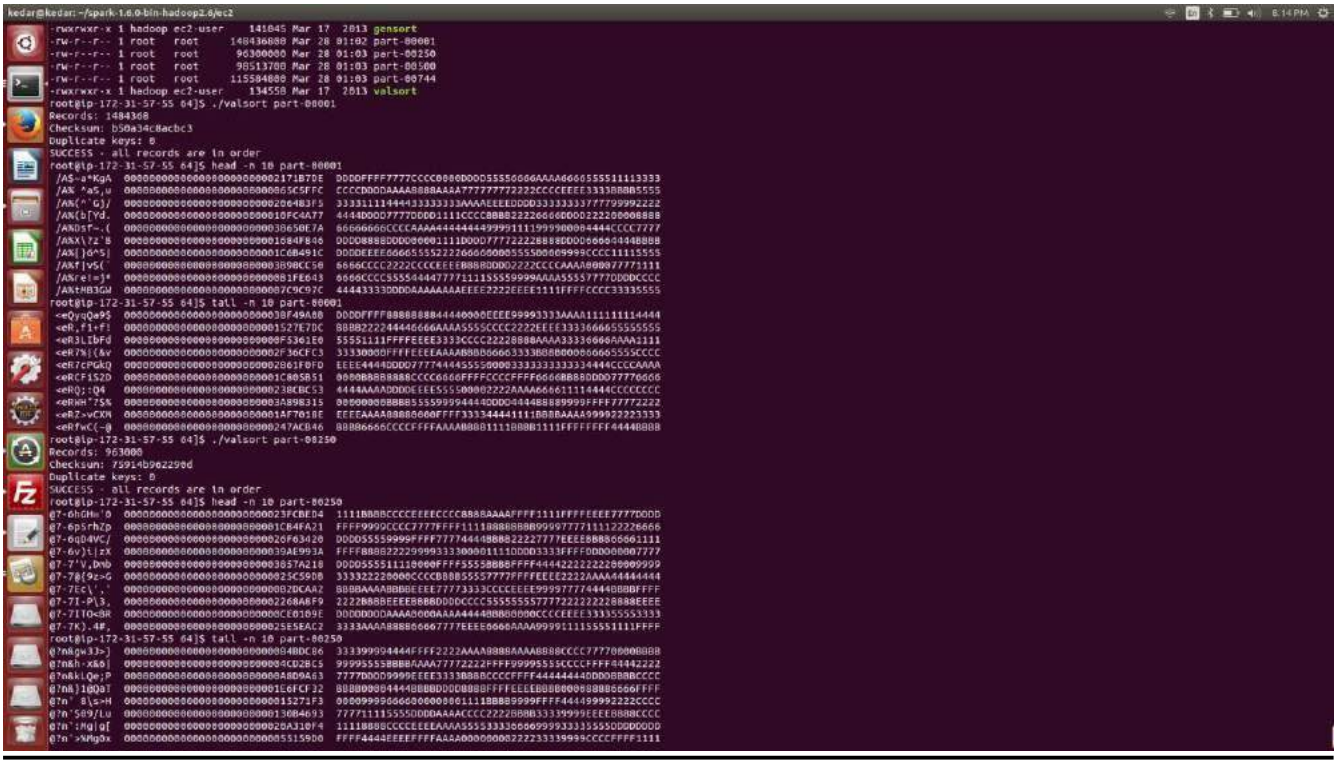
As per done in the 1 Node VC, the data needs to be converted from UNIX to DOS file system before evaluation by ValSort.

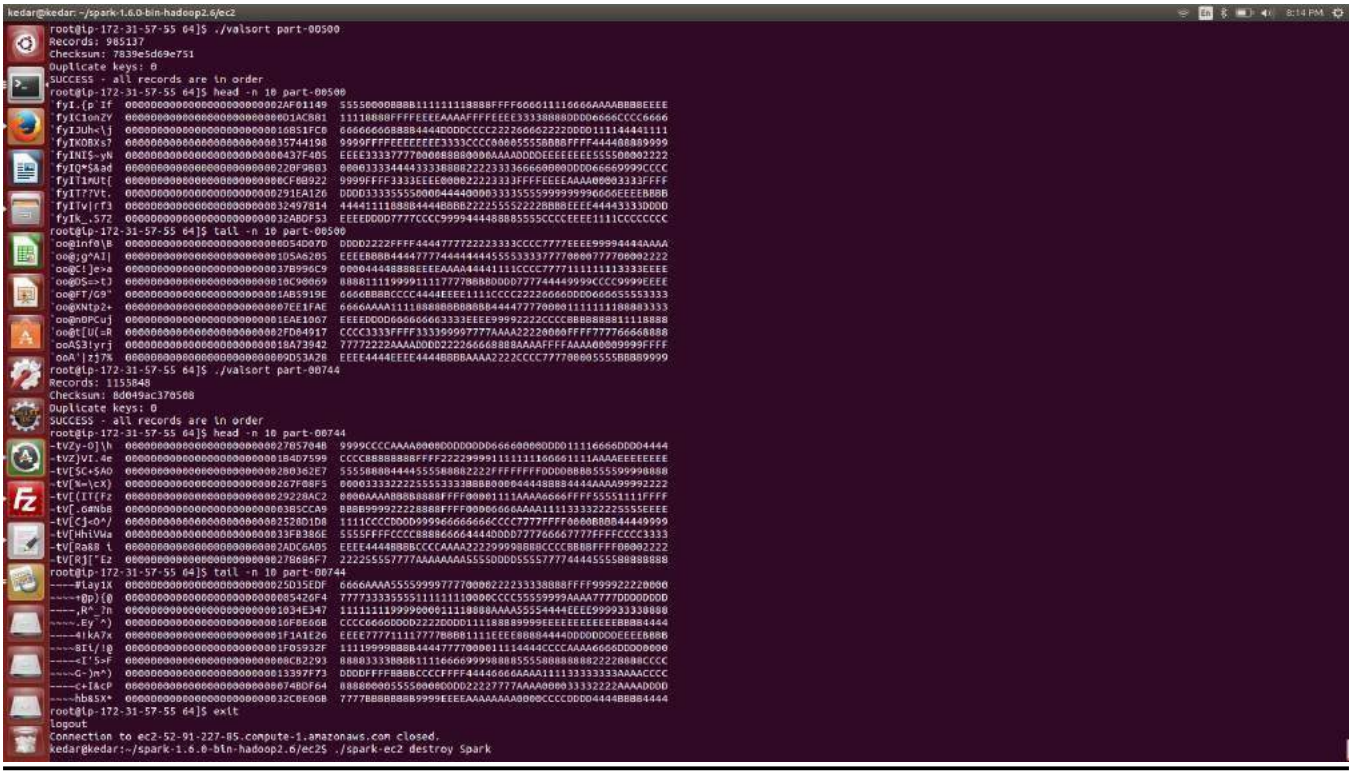
```
kedar@kedar: ~/spark-1.6.0-bin-hadoop2.6/ec2
-rw-r--r-- 1 root   root     114428952 Mar 28 01:01 part-00744
-rwxrwxr-x 1 hadoop ec2-user    134558 Mar 17  2013 valsort
root@ip-172-31-57-55 64]$ sudo yum install dos2unix
Loaded plugins: priorities, update-motd, upgrade-helper
amzn-main/latest                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | 2.1 kB     00:00
amzn-updates/latest                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | 2.3 kB     00:00
Resolving Dependencies
--> Running transaction check
---> Package dos2unix.x86_64 0:3.1-37.5.amzn1 will be installed
--> Finished Dependency Resolution

Dependencies Resolved

================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================
 Package                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                Arch                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                Version                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                Repository                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                Size
================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================
Installing:
 dos2unix                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                Arch                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                Version                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                Repository                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                Size
================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================
Installing:
 dos2unix                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                Arch                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                Version                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                Repository                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                Size
================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================
Installing:
 unix2dos                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                Arch                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                Version                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                Repository                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                Size
================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================
Transaction Summary
================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================
Install  1 Package

Total download size: 13 k
Installed size: 18 k
Is this ok [y/d/N]: y
Downloading packages:
dos2unix-3.1-37.5.amzn1.x86_64.rpm                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | 13 kB     00:00
Running transaction check
Running transaction test
Transaction test succeeded
Running transaction
  Installing : dos2unix-3.1-37.5.amzn1.x86_64                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 1/1
  Verifying  : dos2unix-3.1-37.5.amzn1.x86_64                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 1/1
================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================
Installed:
  dos2unix.x86_64 0:3.1-37.5.amzn1

Complete!
root@ip-172-31-57-55 64]$ sudo yum install unix2dos
Loaded plugins: priorities, update-motd, upgrade-helper
Resolving Dependencies
--> Running transaction check
---> Package unix2dos.x86_64 0:2.2-35.5.amzn1 will be installed
--> Finished Dependency Resolution

Dependencies Resolved

================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================
 Package                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                Arch                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                Version                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                Repository                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                Size
================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================
Installing:
 unix2dos                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                Arch                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                Version                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 2.2-35.5.amzn1                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                Repository                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                Size
================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================
Transaction Summary
================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================================
Install  1 Package
```

- Once the Output files are converted into DOS format, each output files can be evaluated using Valsort.

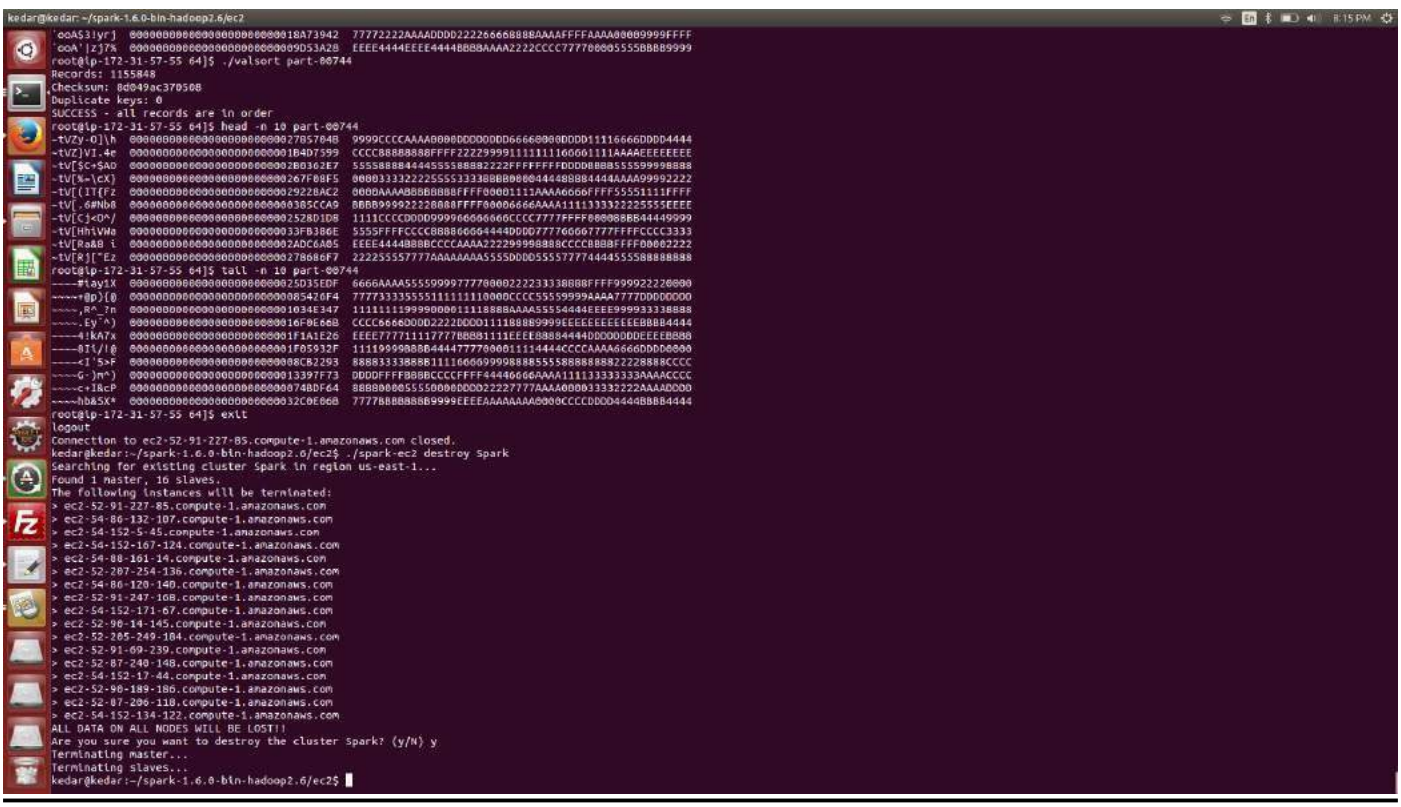
  ./valsort part*

- The output given by Valsort is whether the data generated by gensort is sorted and is valid. Hence on executing the Valsort command on any file will display if the all the records are in order.





As shown in the above screenshots, the records in part-00001, part-00250, part-00500, part-00744 are in order and the head i.e the first 10 lines and the tail i.e. the last 10 lines of the file are shown.



The Spark instance can be terminated by running the destroy command on master node which will end the instance on all of its 16 slaves as well.

./spark-ec2 destroy Spark.

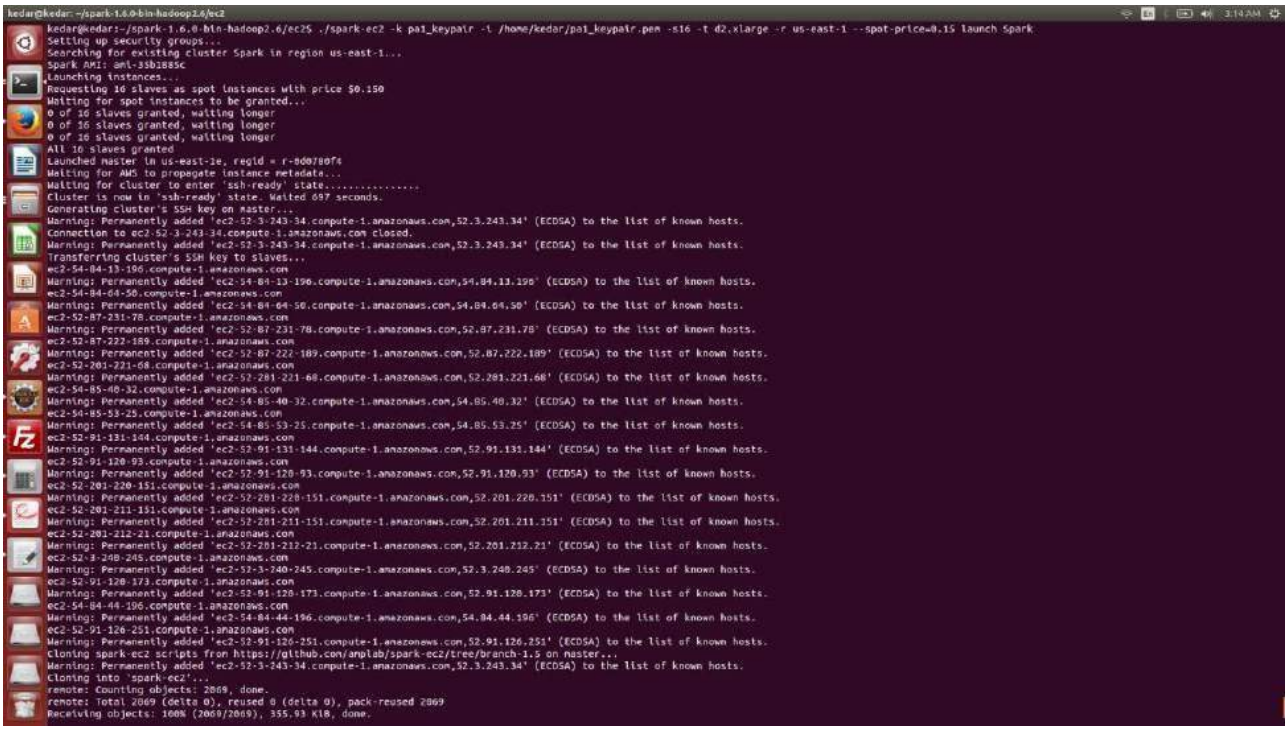
<u>Performance:</u>

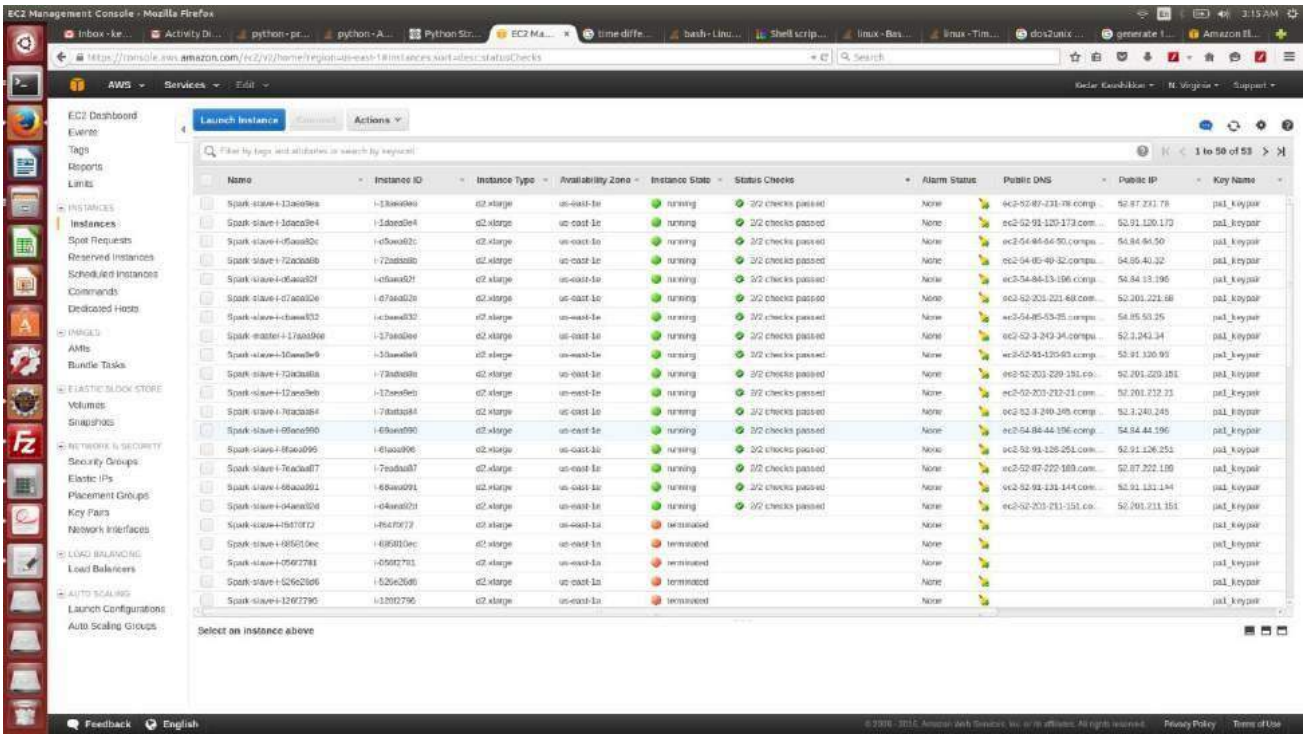
Throughput = (Datasize / Time taken)
= 100000 / 3079.8
= 32.46 Mb/sec

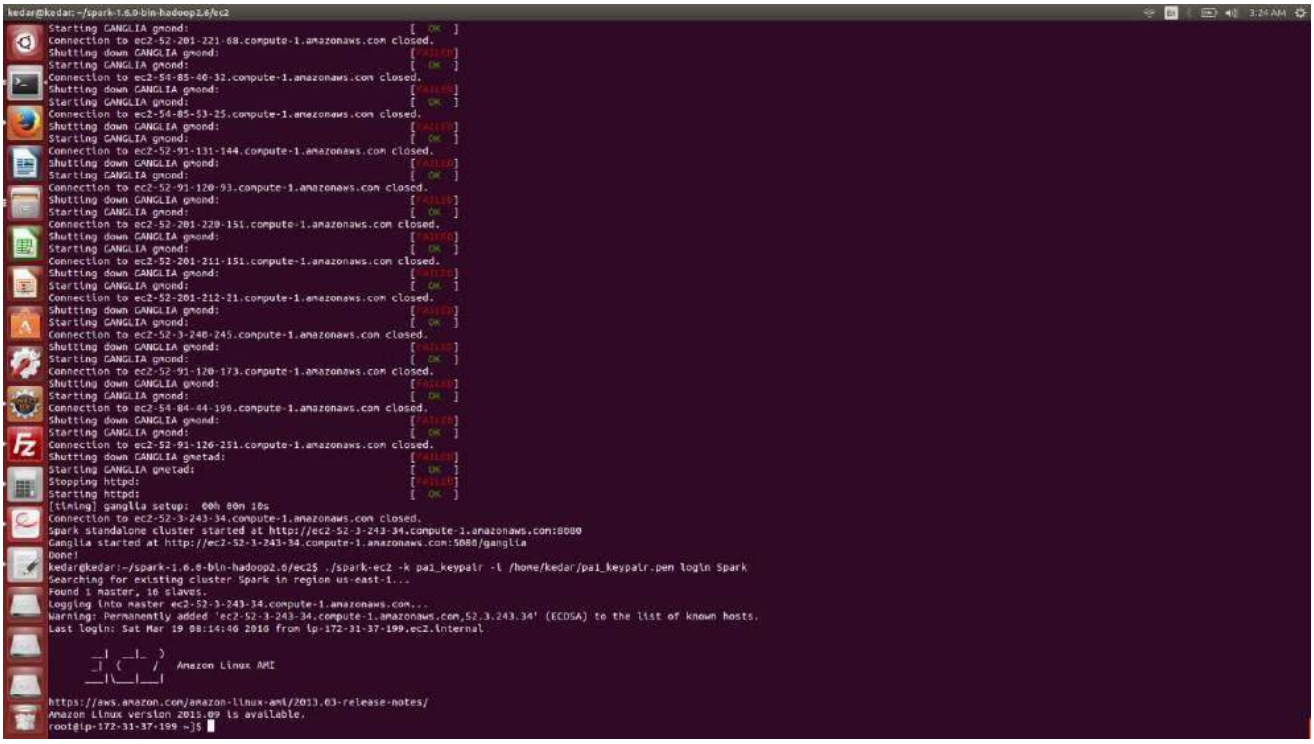
<u>SPARK TeraSort 16 Node 1 GB</u>

- We followed the similar steps for as done for 100 Gb except that for this experiment we used Amazon EC2 d2.xlarge instance.

  Below are the screenshots for the same.




EC2 Management Console - Mozilla Firefox
AWS
Services
Edit
Launch Instance
Connect
Actions
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
Key Pairs
Network Interfaces
LOAD BALANCING
Load Balancers
AUTO SCALING
Launch Configurations
Auto Scaling Groups
Name
Instance ID
Instance Type
Availability Zone
Instance State
Status Checks
Alarm Status
Public DNS
Public IP
Key Name
Select an instance above
Feedback
English

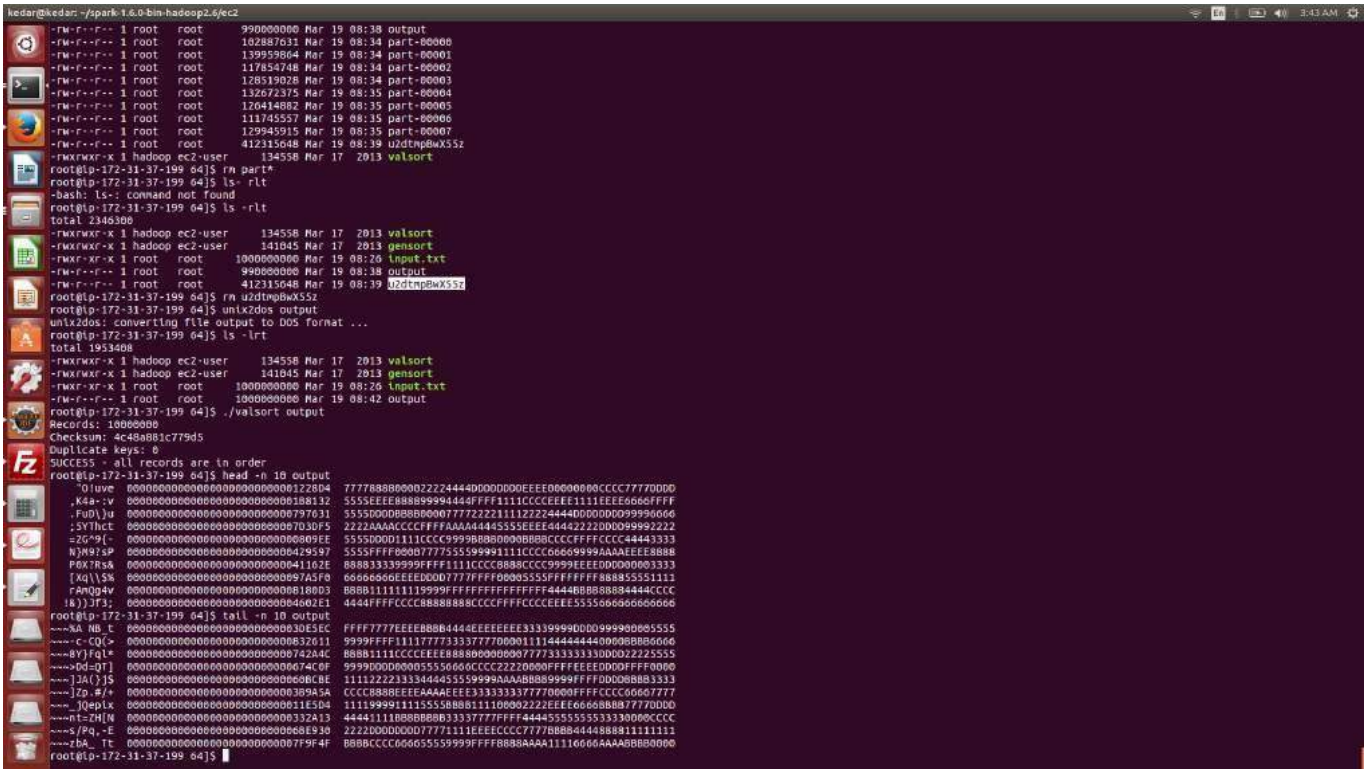

kedar@kedar: ~/spark-1.6.0-bin-hadoop2.6/ec2
Starting GANGLIA gmond: [OK]
Connection to ec2-52-201-221-68.compute-1.amazonaws.com closed.
Shutting down GANGLIA gmond: [FAILED]
Starting GANGLIA gmond: [OK]
Connection to ec2-54-85-40-32.compute-1.amazonaws.com closed.
Shutting down GANGLIA gmond: [FAILED]
Starting GANGLIA gmond: [OK]
Connection to ec2-54-85-53-25.compute-1.amazonaws.com closed.
Shutting down GANGLIA gmond: [FAILED]
Starting GANGLIA gmond: [OK]
Connection to ec2-52-91-131-144.compute-1.amazonaws.com closed.
Shutting down GANGLIA gmond: [FAILED]
Starting GANGLIA gmond: [OK]
Connection to ec2-52-91-120-93.compute-1.amazonaws.com closed.
Shutting down GANGLIA gmond: [FAILED]
Starting GANGLIA gmond: [OK]
Connection to ec2-52-201-220-151.compute-1.amazonaws.com closed.
Shutting down GANGLIA gmond: [FAILED]
Starting GANGLIA gmond: [OK]
Connection to ec2-52-201-211-151.compute-1.amazonaws.com closed.
Shutting down GANGLIA gmond: [FAILED]
Starting GANGLIA gmond: [OK]
Connection to ec2-52-201-212-21.compute-1.amazonaws.com closed.
Shutting down GANGLIA gmond: [FAILED]
Starting GANGLIA gmond: [OK]
Connection to ec2-52-3-240-245.compute-1.amazonaws.com closed.
Shutting down GANGLIA gmond: [FAILED]
Starting GANGLIA gmond: [OK]
Connection to ec2-52-91-120-173.compute-1.amazonaws.com closed.
Shutting down GANGLIA gmond: [FAILED]
Starting GANGLIA gmond: [OK]
Connection to ec2-54-84-44-196.compute-1.amazonaws.com closed.
Shutting down GANGLIA gmond: [FAILED]
Starting GANGLIA gmond: [OK]
Connection to ec2-52-91-126-251.compute-1.amazonaws.com closed.
Shutting down GANGLIA gmetad: [FAILED]
Starting GANGLIA gmetad: [OK]
Stopping httpd: [FAILED]
Starting httpd: [OK]
[timing] ganglia setup: 00h 00m 10s
Connection to ec2-52-3-243-34.compute-1.amazonaws.com closed.
Spark standalone cluster started at http://ec2-52-3-243-34.compute-1.amazonaws.com:8080
Ganglia started at http://ec2-52-3-243-34.compute-1.amazonaws.com:5080/ganglia
Done!
kedar@kedar:~/spark-1.6.0-bin-hadoop2.6/ec2$ ./spark-ec2 -k pai_keypair -i /home/kedar/pai_keypair.pem login Spark
Searching for existing cluster Spark in region us-east-1...
Found 1 master, 16 slaves.
Logging into master ec2-52-3-243-34.compute-1.amazonaws.com...
Warning: Permanently added 'ec2-52-3-243-34.compute-1.amazonaws.com,52.3.243.34' (ECDSA) to the list of known hosts.
Last login: Sat Mar 19 08:14:46 2016 from ip-172-31-37-199.ec2.internal
__| __|_)
_| (/ Amazon Linux AMI
___|___|___|
https://aws.amazon.com/amazon-linux-ami/2013.03-release-notes/
Amazon Linux version 2013.09 is available.
root@ip-172-31-37-199 ~]$

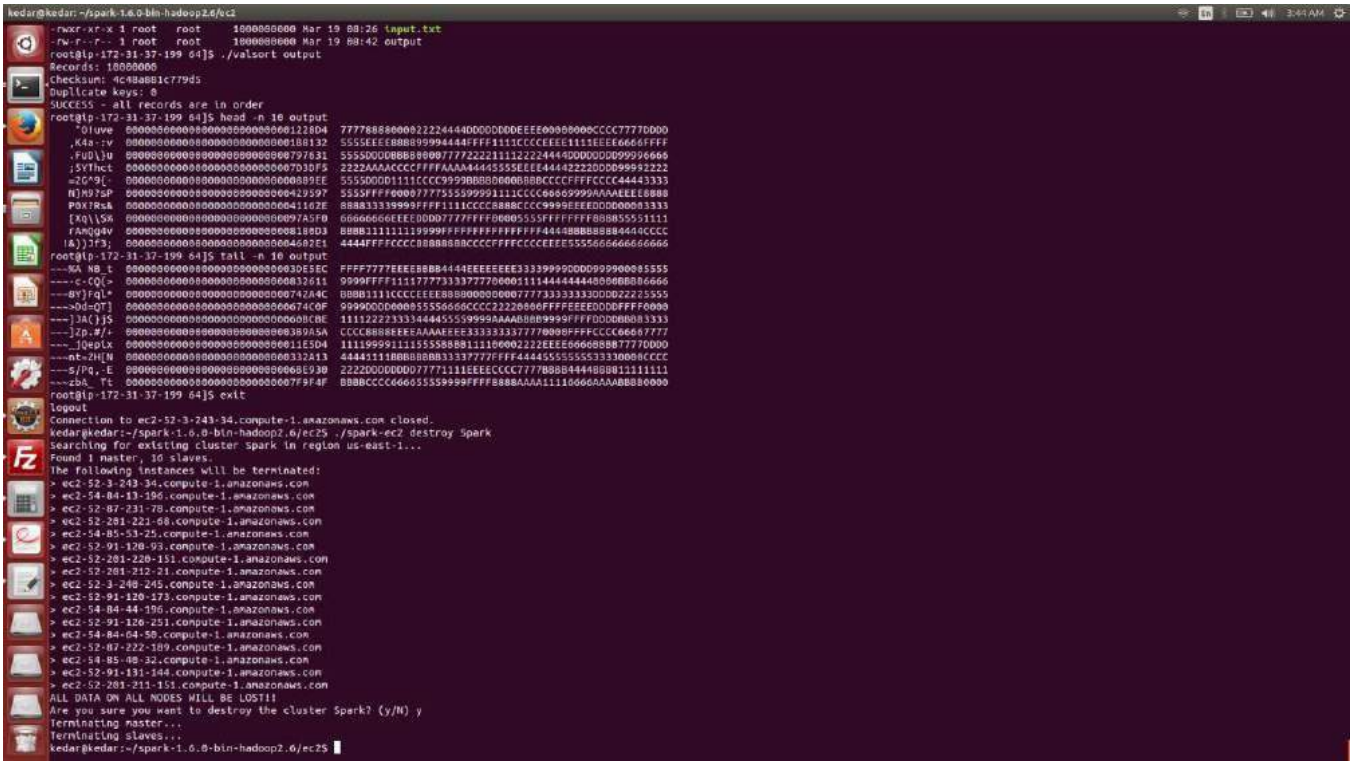
```
kedar@kedar: ~/spark-1.6.0-bin-hadoop2.6/ec2
drwxrwxr-x  9    2000     2000   4096 Sep 27  2013 scala
drwxr-xr-x 13 hadoop  ec2-user   4096 Mar 19 08:22 spark
drwxr-xr-x 15 root    root       4096 Mar 19 08:22 spark-ec2
-rw-r--r--  1 root    root        358 Mar 19 08:17 spark_sort.py
drwxr-xr-x 21 rstudio games      4096 Mar 19 08:23 tachyon
root@ip-172-31-37-199 ~]$ sudo vi spark_sort.py
root@ip-172-31-37-199 ~]$ tar -xzvf ge
gensort-linux-1.5.tar.gz  getTimeOfSparkRun.sh
root@ip-172-31-37-199 ~]$ tar -xzvf gensort-linux-1.5.tar.gz
gpl-2.0.txt
32/
32/valsort
32/gensort
64/
64/valsort
64/gensort
root@ip-172-31-37-199 ~]$ rm gensort-linux-1.5.tar.gz
root@ip-172-31-37-199 ~]$ cd 64
root@ip-172-31-37-199 64]$ ./gensort -a 10000000 input.txt
root@ip-172-31-37-199 64]$ cd ..
root@ip-172-31-37-199 ~]$ cd ephemeral-hdfs/bin/
root@ip-172-31-37-199 bin]$ ll
total 144
-rwxr-xr-x 1 root root 14133 Mar 19 08:17 hadoop
-rwxr-xr-x 1 root root  2642 Oct  3  2012 hadoop-config.sh
-rwxr-xr-x 1 root root  4330 Oct  3  2012 hadoop-daemon.sh
-rwxr-xr-x 1 root root  1329 Oct  3  2012 hadoop-daemons.sh
-rwxr-xr-x 1 root root  2810 Oct  3  2012 rcc
-rwxr-xr-x 1 root root  2143 Oct  3  2012 slaves.sh
-rwxr-xr-x 1 root root  1166 Oct  3  2012 start-all.sh
-rwxr-xr-x 1 root root  1065 Oct  3  2012 start-balancer.sh
-rwxr-xr-x 1 root root  1745 Oct  3  2012 start-dfs.sh
-rwxr-xr-x 1 root root  1145 Oct  3  2012 start-jobhistoryserver.sh
-rwxr-xr-x 1 root root  1259 Oct  3  2012 start-mapred.sh
-rwxr-xr-x 1 root root  1119 Oct  3  2012 stop-all.sh
-rwxr-xr-x 1 root root  1110 Oct  3  2012 stop-balancer.sh
-rwxr-xr-x 1 root root  1246 Oct  3  2012 stop-dfs.sh
-rwxr-xr-x 1 root root  1131 Oct  3  2012 stop-jobhistoryserver.sh
-rwxr-xr-x 1 root root  1168 Oct  3  2012 stop-mapred.sh
-rwxr-xr-x 1 root root 63374 Oct  3  2012 task-controller
root@ip-172-31-37-199 bin]$ ./hadoop dfs -mkdir /usr
Warning: $HADOOP_HOME is deprecated.

root@ip-172-31-37-199 bin]$ ./hadoop dfs -mkdir /usr/sparkusr
Warning: $HADOOP_HOME is deprecated.

root@ip-172-31-37-199 bin]$ ./hadoop dfs -mkdir /usr/sparkusr/input
Warning: $HADOOP_HOME is deprecated.

root@ip-172-31-37-199 bin]$ ./hadoop fs -put /root/64/input.txt /usr/sparkusr/input/input.txt
Warning: $HADOOP_HOME is deprecated.

root@ip-172-31-37-199 bin]$ ./hadoop fs -ls /usr/sparkusr/input
Warning: $HADOOP_HOME is deprecated.

Found 1 items
-rw-r--r--   3 root supergroup 1000000000 2016-03-19 08:27 /usr/sparkusr/input/input.txt
root@ip-172-31-37-199 bin]$
```

```
kedar@kedar: ~/spark-1.6.0-bin-hadoop2.6/ec2
-rwxr-xr-x 1 hadoop ec2-user 1050 Dec 22 02:21 spark-submit
-rw-r--r-- 1 hadoop ec2-user 1126 Dec 22 02:21 spark-submit2.cmd
-rw-r--r-- 1 hadoop ec2-user 1010 Dec 22 02:21 spark-submit.cmd
root@ip-172-31-37-199 bin]$ chmod 777 getTimeOfSparkRun.sh
root@ip-172-31-37-199 bin]$ ./getTimeOfSparkRun.sh
Sat Mar 19 08:30:48 UTC 2016
16/03/19 08:30:52 INFO spark.SparkContext: Running Spark version 1.6.0
16/03/19 08:30:52 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.

16/03/19 08:30:52 INFO spark.SecurityManager: Changing view acls to: root
16/03/19 08:30:52 INFO spark.SecurityManager: Changing modify acls to: root
16/03/19 08:30:52 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(root); users with modify permissions: Set(root)
16/03/19 08:30:52 INFO util.Utils: Successfully started service 'sparkDriver' on port 43823.
16/03/19 08:30:53 INFO slf4j.Slf4jLogger: Slf4jLogger started
16/03/19 08:30:53 INFO Remoting: Starting remoting
16/03/19 08:30:53 INFO Remoting: Remoting started; listening on addresses :[akka.tcp://sparkDriverActorSystem@172.31.37.199:36794]
16/03/19 08:30:53 INFO util.Utils: Successfully started service 'sparkDriverActorSystem' on port 36794.
16/03/19 08:30:53 INFO spark.SparkEnv: Registering MapOutputTracker
16/03/19 08:30:53 INFO spark.SparkEnv: Registering BlockManagerMaster
16/03/19 08:30:53 INFO storage.DiskBlockManager: Created local directory at /mnt/spark/blockmgr-8d241660-6638-4e5e-a391-4dc619fe9586
16/03/19 08:30:53 INFO storage.DiskBlockManager: Created local directory at /mnt2/spark/blockmgr-fe7aa591-5f1f-4417-955b-db8e3f35367a
16/03/19 08:30:53 INFO storage.DiskBlockManager: Created local directory at /mnt3/spark/blockmgr-928db6a9-c38d-4bbf-a225-d2b306efbbe7
16/03/19 08:30:53 INFO storage.MemoryStore: MemoryStore started with capacity 511.5 MB
16/03/19 08:30:53 INFO spark.SparkEnv: Registering OutputCommitCoordinator
16/03/19 08:30:53 INFO server.Server: jetty-8.y.z-SNAPSHOT
16/03/19 08:30:53 INFO server.AbstractConnector: Started SelectChannelConnector@0.0.0.0:4040
16/03/19 08:30:53 INFO util.Utils: Successfully started service 'sparkUI' on port 4040.
16/03/19 08:30:53 INFO ui.SparkUI: Started SparkUI at http://ec2-52-3-243-34.compute-1.amazonaws.com:4040
16/03/19 08:30:53 INFO spark.HttpFileServer: HTTP File server directory is /mnt/spark/spark-f45f038b-07da-4919-a579-53f7e32eca16/httpd-9cd2222d-5cee-4676-882e-ff137248d832
16/03/19 08:30:53 INFO spark.HttpServer: Starting HTTP Server
16/03/19 08:30:53 INFO server.Server: jetty-8.y.z-SNAPSHOT
16/03/19 08:30:53 INFO server.AbstractConnector: Started SocketConnector@0.0.0.0:39143
16/03/19 08:30:53 INFO util.Utils: Successfully started service 'HTTP file server' on port 39143.
16/03/19 08:30:53 INFO util.Utils: Copying /root/spark_sort.py to /mnt/spark/spark-f45f038b-07da-4919-a579-53f7e32eca16/userFiles-2028b13b-e062-435a-886f-a27898329c33/spark_sort.py
16/03/19 08:30:53 INFO spark.SparkContext: Added file file:/root/spark_sort.py at http://172.31.37.199:39143/files/spark_sort.py with timestamp 1458376253793
16/03/19 08:30:53 INFO client.AppClient$ClientEndpoint: Connecting to master spark://ec2-52-3-243-34.compute-1.amazonaws.com:7077...
16/03/19 08:30:54 INFO cluster.SparkDeploySchedulerBackend: Connected to Spark cluster with app ID app-20160319083054-0000
16/03/19 08:30:54 INFO util.Utils: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port 39367.
16/03/19 08:30:54 INFO netty.NettyBlockTransferService: Server created on 39367
16/03/19 08:30:54 INFO storage.BlockManagerMaster: Trying to register BlockManager
16/03/19 08:30:54 INFO storage.BlockManagerMasterEndpoint: Registering block manager 172.31.37.199:39367 with 511.5 MB RAM, BlockManagerId(driver, 172.31.37.199, 39367)
16/03/19 08:30:54 INFO storage.BlockManagerMaster: Registered BlockManager
16/03/19 08:30:54 INFO client.AppClient$ClientEndpoint: Executor added: app-20160319083054-0000/0 on worker-20160319082313-172.31.41.119-60205 (172.31.41.119:60205) with 4 cores
16/03/19 08:30:54 INFO cluster.SparkDeploySchedulerBackend: Granted executor ID app-20160319083054-0000/0 on hostPort 172.31.41.119:60205 with 4 cores, 26.7 GB RAM
16/03/19 08:30:54 INFO client.AppClient$ClientEndpoint: Executor added: app-20160319083054-0000/1 on worker-20160319082313-172.31.41.143-58754 (172.31.41.143:58754) with 4 cores
16/03/19 08:30:54 INFO cluster.SparkDeploySchedulerBackend: Granted executor ID app-20160319083054-0000/1 on hostPort 172.31.41.143:58754 with 4 cores, 26.7 GB RAM
16/03/19 08:30:54 INFO client.AppClient$ClientEndpoint: Executor added: app-20160319083054-0000/2 on worker-20160319082313-172.31.39.174-51615 (172.31.39.174:51615) with 4 cores
16/03/19 08:30:54 INFO cluster.SparkDeploySchedulerBackend: Granted executor ID app-20160319083054-0000/2 on hostPort 172.31.39.174:51615 with 4 cores, 26.7 GB RAM
16/03/19 08:30:54 INFO client.AppClient$ClientEndpoint: Executor added: app-20160319083054-0000/3 on worker-20160319082313-172.31.39.143-44543 (172.31.39.143:44543) with 4 cores
16/03/19 08:30:54 INFO cluster.SparkDeploySchedulerBackend: Granted executor ID app-20160319083054-0000/3 on hostPort 172.31.39.143:44543 with 4 cores, 26.7 GB RAM
16/03/19 08:30:54 INFO client.AppClient$ClientEndpoint: Executor added: app-20160319083054-0000/4 on worker-20160319082313-172.31.33.177-59175 (172.31.33.177:59175) with 4 cores
```

```
kedar@kedar: ~/spark-1.6.0-bin-hadoop2.6/ec2
16/03/19 08:31:31 INFO spark.MapOutputTrackerMasterEndpoint: Asked to send map output locations for shuffle 0 to ip-172-31-40-90.ec2.internal:43293
16/03/19 08:31:39 INFO scheduler.TaskSetManager: Finished task 8.0 in stage 3.0 (TID 24) in 9090 ms on ip-172-31-34-32.ec2.internal (1/8)
16/03/19 08:31:41 INFO scheduler.TaskSetManager: Finished task 5.0 in stage 3.0 (TID 29) in 10707 ms on ip-172-31-33-177.ec2.internal (2/8)
16/03/19 08:31:42 INFO scheduler.TaskSetManager: Finished task 7.0 in stage 3.0 (TID 31) in 11555 ms on ip-172-31-34-43.ec2.internal (3/8)
16/03/19 08:31:42 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 3.0 (TID 28) in 11884 ms on ip-172-31-39-174.ec2.internal (4/8)
16/03/19 08:31:42 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 3.0 (TID 27) in 11860 ms on ip-172-31-42-233.ec2.internal (5/8)
16/03/19 08:31:42 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 3.0 (TID 26) in 11971 ms on ip-172-31-41-143.ec2.internal (6/8)
16/03/19 08:31:43 INFO scheduler.TaskSetManager: Finished task 6.0 in stage 3.0 (TID 30) in 12958 ms on ip-172-31-40-90.ec2.internal (7/8)
16/03/19 08:31:44 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 3.0 (TID 25) in 13932 ms on ip-172-31-43-143.ec2.internal (8/8)
16/03/19 08:31:44 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 3.0, whose tasks have all completed, from pool
16/03/19 08:31:44 INFO scheduler.DAGScheduler: ResultStage 3 (saveAsTextFile at NativeMethodAccessorImpl.java:-2) finished in 13.934 s
16/03/19 08:31:44 INFO scheduler.DAGScheduler: Job 2 finished: saveAsTextFile at NativeMethodAccessorImpl.java:-2, took 24.772569 s
16/03/19 08:31:44 INFO spark.SparkContext: Invoking stop() from shutdown hook
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/metrics/json,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/stages/stage/kill,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/api,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/static,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/executors/threadDump/json,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/executors/threadDump,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/executors/json,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/executors,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/environment/json,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/environment,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/storage/rdd/json,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/storage/rdd,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/storage/json,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/storage,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/stages/pool/json,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/stages/pool,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/stages/stage/json,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/stages/stage,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/stages/json,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/stages,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/jobs/job/json,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/jobs/job,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/jobs/json,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/jobs,null}
16/03/19 08:31:44 INFO ui.SparkUI: Stopped Spark web UI at http://ec2-52-3-243-34.compute-1.amazonaws.com:4040
16/03/19 08:31:44 INFO cluster.SparkDeploySchedulerBackend: Shutting down all executors
16/03/19 08:31:44 INFO cluster.SparkDeploySchedulerBackend: Asking each executor to shut down
16/03/19 08:31:44 INFO spark.MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
16/03/19 08:31:44 INFO storage.MemoryStore: MemoryStore cleared
16/03/19 08:31:44 INFO storage.BlockManager: BlockManager stopped
16/03/19 08:31:44 INFO storage.BlockManagerMaster: BlockManagerMaster stopped
16/03/19 08:31:44 INFO scheduler.OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
16/03/19 08:31:44 INFO remote.RemoteActorRefProvider$RemotingTerminator: Shutting down remote daemon.
16/03/19 08:31:44 INFO spark.SparkContext: Successfully stopped SparkContext
16/03/19 08:31:44 INFO remote.RemoteActorRefProvider$RemotingTerminator: Remote daemon shut down; proceeding with flushing remote transports.
16/03/19 08:31:44 INFO util.ShutdownHookManager: Shutdown hook called
16/03/19 08:31:44 INFO util.ShutdownHookManager: Deleting directory /mnt/spark/spark-f45f038b-07da-4919-a579-53f7e32eca16
16/03/19 08:31:44 INFO util.ShutdownHookManager: Deleting directory /mnt3/spark/spark-a2bbac10-3ae4-4759-a719-00aa6508b6d6
16/03/19 08:31:44 INFO util.ShutdownHookManager: Deleting directory /mnt2/spark/spark-af9204fd-57c0-46cc-8a2d-c1163598087d
16/03/19 08:31:44 INFO util.ShutdownHookManager: Deleting directory /mnt/spark/spark-f45f038b-07da-4919-a579-53f7e32eca16/httpd-9cd2222d-5cee-4676-882e-ff137248d832
16/03/19 08:31:44 INFO util.ShutdownHookManager: Deleting directory /mnt/spark/spark-f45f038b-07da-4919-a579-53f7e32eca16/pyspark-01844546-09e1-463a-a4c6-1beeca1c6b40
Sat Mar 19 08:31:45 UTC 2016
It took 56122 milliseconds
root@ip-172-31-37-199 bin]$
```

```
kedar@kedar: ~/spark-1.6.0-bin-hadoop2.6/ec2
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/executors/json,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/executors,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/environment/json,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/environment,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/storage/rdd/json,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/storage/rdd,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/storage/json,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/storage,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/stages/pool/json,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/stages/pool,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/stages/stage/json,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/stages/stage,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/stages/json,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/stages,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/jobs/job/json,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/jobs/job,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/jobs/json,null}
16/03/19 08:31:44 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/jobs,null}
16/03/19 08:31:44 INFO ui.SparkUI: Stopped Spark web UI at http://ec2-52-3-243-34.compute-1.amazonaws.com:4040
16/03/19 08:31:44 INFO cluster.SparkDeploySchedulerBackend: Shutting down all executors
16/03/19 08:31:44 INFO cluster.SparkDeploySchedulerBackend: Asking each executor to shut down
16/03/19 08:31:44 INFO spark.MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
16/03/19 08:31:44 INFO storage.MemoryStore: MemoryStore cleared
16/03/19 08:31:44 INFO storage.BlockManager: BlockManager stopped
16/03/19 08:31:44 INFO storage.BlockManagerMaster: BlockManagerMaster stopped
16/03/19 08:31:44 INFO scheduler.OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
16/03/19 08:31:44 INFO remote.RemoteActorRefProvider$RemotingTerminator: Shutting down remote daemon.
16/03/19 08:31:44 INFO spark.SparkContext: Successfully stopped SparkContext
16/03/19 08:31:44 INFO remote.RemoteActorRefProvider$RemotingTerminator: Remote daemon shut down; proceeding with flushing remote transports.
16/03/19 08:31:44 INFO util.ShutdownHookManager: Shutdown hook called
16/03/19 08:31:44 INFO util.ShutdownHookManager: Deleting directory /mnt/spark/spark-f45f038b-07da-4919-a579-53f7e32eca16
16/03/19 08:31:44 INFO util.ShutdownHookManager: Deleting directory /mnt3/spark/spark-a2bbac10-3ae4-4759-a719-00aa6508b6d6
16/03/19 08:31:44 INFO util.ShutdownHookManager: Deleting directory /mnt2/spark/spark-af9204fd-57c0-46cc-8a2d-c1163598087d
16/03/19 08:31:44 INFO util.ShutdownHookManager: Deleting directory /mnt/spark/spark-f45f038b-07da-4919-a579-53f7e32eca16/httpd-9cd2222d-5cee-4676-882e-ff137248d832
16/03/19 08:31:44 INFO util.ShutdownHookManager: Deleting directory /mnt/spark/spark-f45f038b-07da-4919-a579-53f7e32eca16/pyspark-01844546-09e1-463a-a4c6-1beeca1c6b40
Sat Mar 19 08:31:45 UTC 2016
It took 56122 milliseconds
root@ip-172-31-37-199 bin]$ cd ..
root@ip-172-31-37-199 spark]$ cd ..
root@ip-172-31-37-199 ~]$ cd ephemeral-hdfs/bin/
root@ip-172-31-37-199 bin]$ ./hadoop fs -ls /usr/sparkusr/output
Warning: $HADOOP_HOME is deprecated.

Found 9 items
-rw-r--r--   3 root supergroup          0 2016-03-19 08:31 /usr/sparkusr/output/_SUCCESS
-rw-r--r--   3 root supergroup  102887631 2016-03-19 08:31 /usr/sparkusr/output/part-00000
-rw-r--r--   3 root supergroup  139959864 2016-03-19 08:31 /usr/sparkusr/output/part-00001
-rw-r--r--   3 root supergroup  117854748 2016-03-19 08:31 /usr/sparkusr/output/part-00002
-rw-r--r--   3 root supergroup  128519020 2016-03-19 08:31 /usr/sparkusr/output/part-00003
-rw-r--r--   3 root supergroup  132672375 2016-03-19 08:31 /usr/sparkusr/output/part-00004
-rw-r--r--   3 root supergroup  126414882 2016-03-19 08:31 /usr/sparkusr/output/part-00005
-rw-r--r--   3 root supergroup  111745557 2016-03-19 08:31 /usr/sparkusr/output/part-00006
-rw-r--r--   3 root supergroup  129945915 2016-03-19 08:31 /usr/sparkusr/output/part-00007
root@ip-172-31-37-199 bin]$ ^C
root@ip-172-31-37-199 bin]$ ./hadoop fs -get /usr/sparkusr/output/part* /root/64
Warning: $HADOOP_HOME is deprecated.

root@ip-172-31-37-199 bin]$
```





<u>As per the above screenshot, the 16 node spark cluster sorting the data of 1 GB took 56122 ms, i.e. 0.93 Minutes.</u>

Performance:

Throughput = (Data size / Time taken)
= 1000 / 55.8
= 17.92 Mb/sec

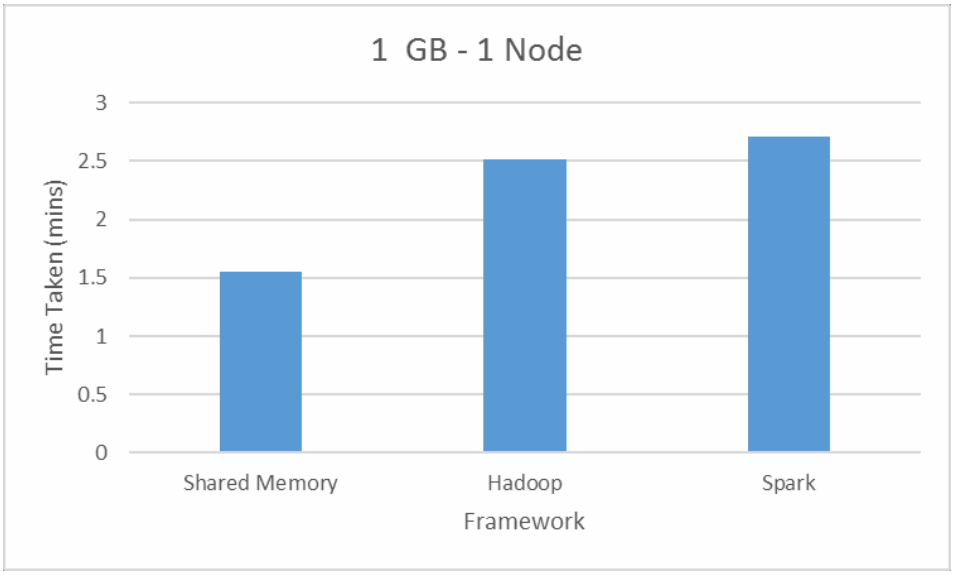
Speedups:

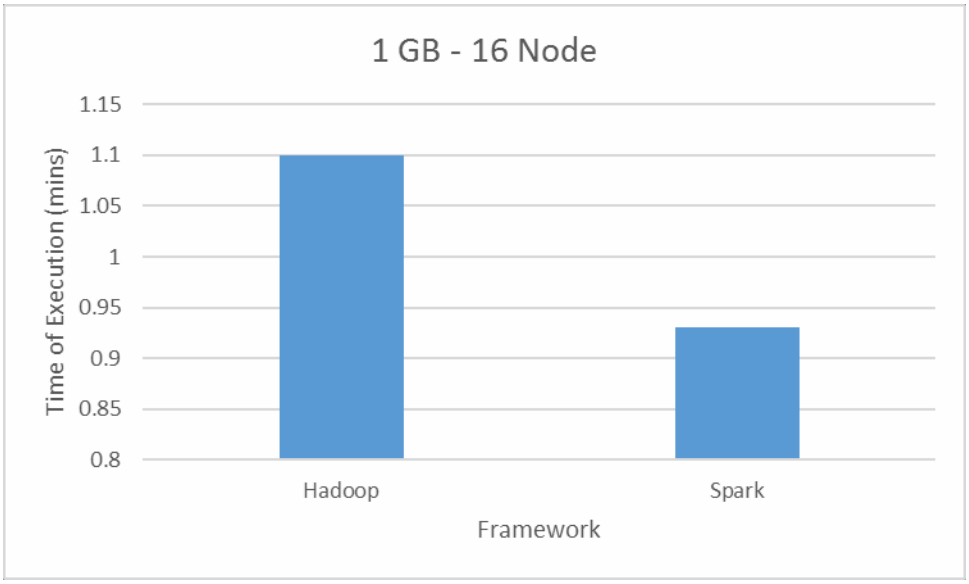
Similar to results for Hadoop map-reduce experiments, we can observe that Spark also had the same pattern. When we check the performance of Spark, if the number of nodes are increased for the increasing the dataset gradually , then the performance efficiency is increased. This can be seen as by comparing the result of 10GB on 1 Node and 100 GB on 16 Node .

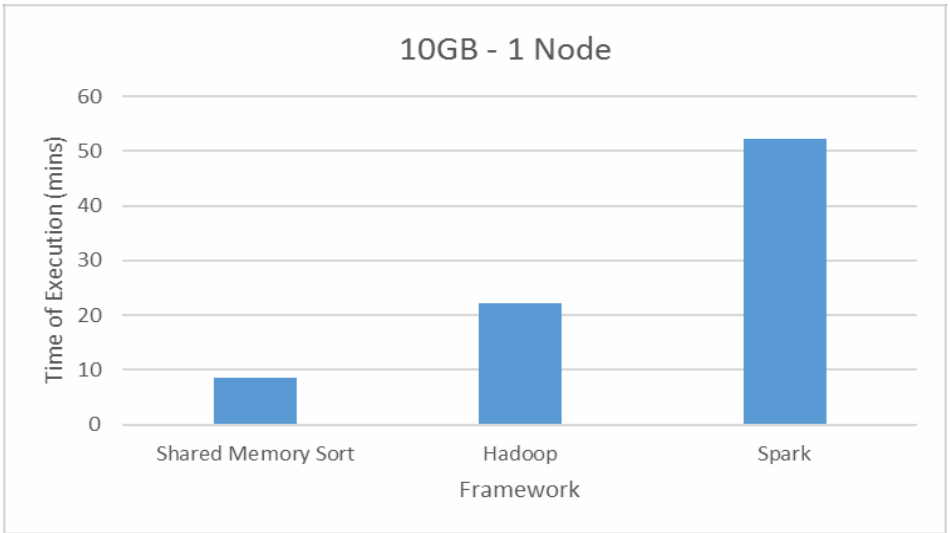
The performance kept constant and did not depreciate on increasing the data set as the number of virtual nodes keep increasing.

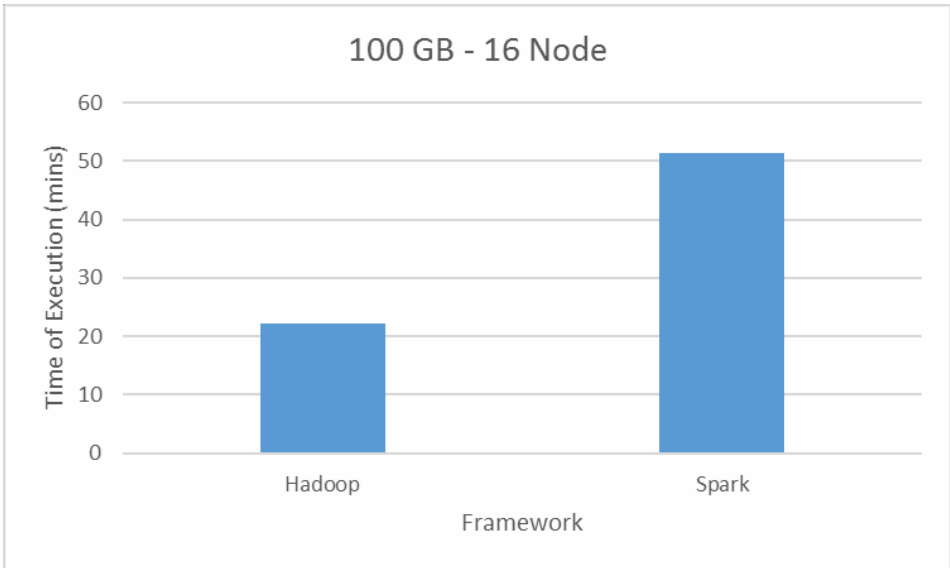
Hence it can be concluded that the performance of spark when dataset is increased will increase gradually if the number of virtual clusters are increased.

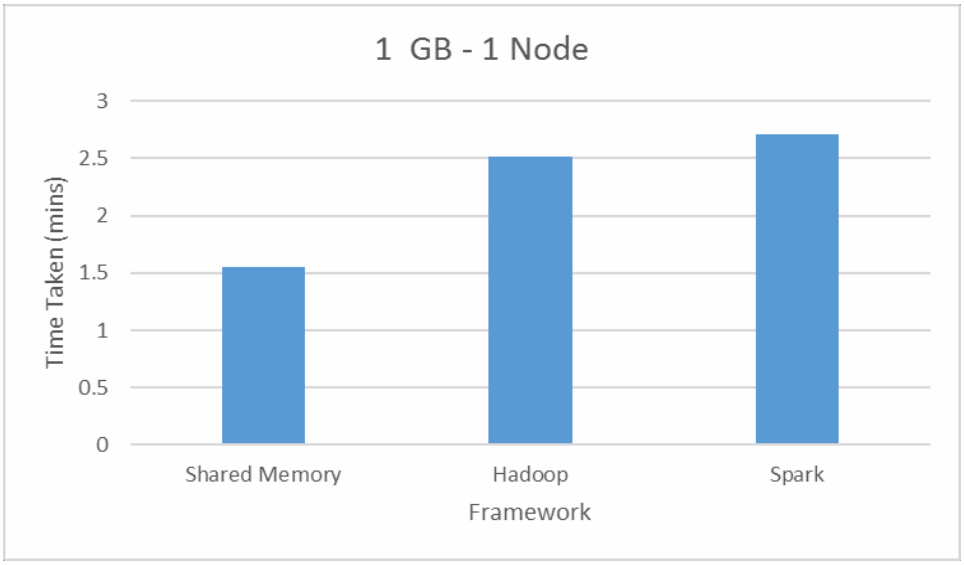
Graphs:

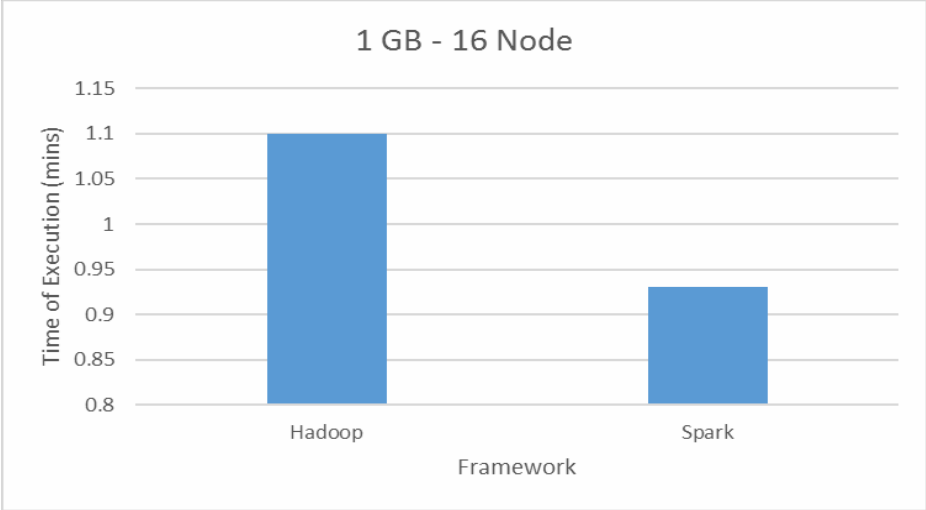


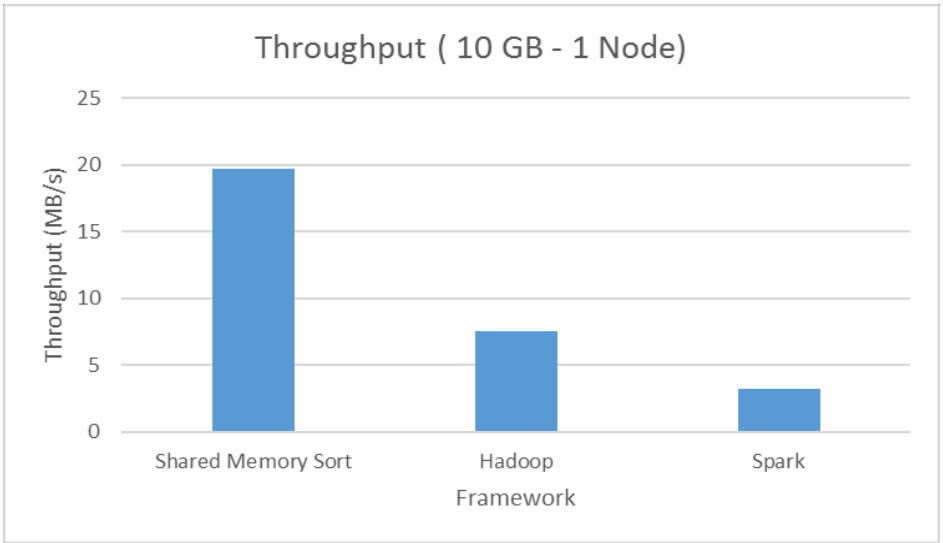

1 GB - 16 Node
Time of Execution (mins)
1.15
1.1
1.05
1
0.95
0.9
0.85
0.8
Hadoop
Spark
Framework

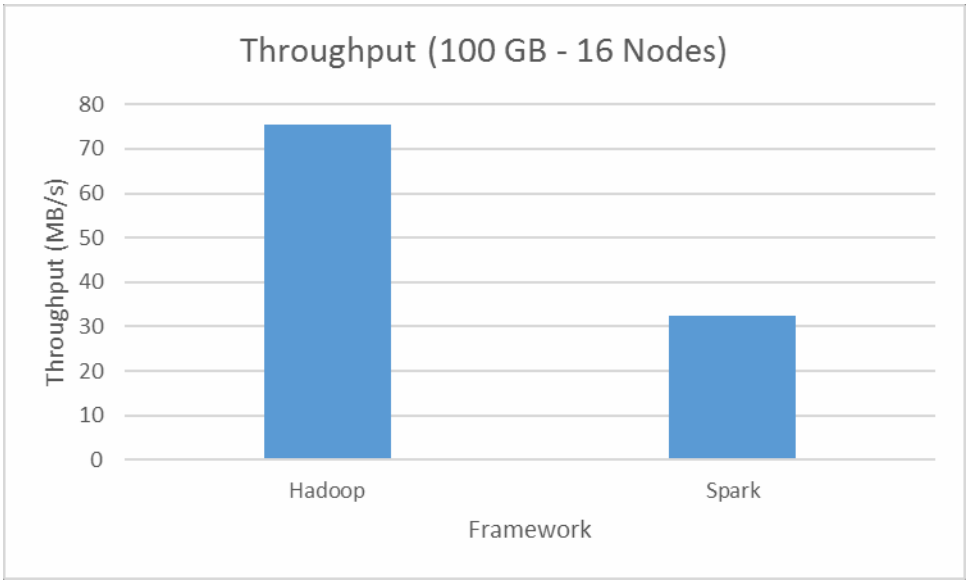

10GB - 1 Node
Time of Execution (mins)
60
50
40
30
20
10
0
Shared Memory Sort
Hadoop
Spark
Framework


100 GB - 16 Node
Time of Execution (mins)
60
50
40
30
20
10
0
Hadoop
Spark
Framework


1 GB - 1 Node
Time Taken (mins)
3
2.5
2
1.5
1
0.5
0
Shared Memory
Hadoop
Spark
Framework


1 GB - 16 Node
Time of Execution (mins)
1.15
1.1
1.05
1
0.95
0.9
0.85
0.8
Hadoop
Spark
Framework


Throughput (10 GB - 1 Node)
Throughput (MB/s)
25
20
15
10
5
0
Shared Memory Sort
Hadoop
Spark
Framework



Conclusion :

As we look up in the graphs, the performance for 1 node on 10GB of data set was better on Shared Memory while for the performance on 16 node was best in Hadoop.

For 16 Node, Hadoop performed better as the master node in our experiment was d2.xlarge while the master node in Spark was c3.large. Ideally Spark performs better when compared to Hadoop when there is multiple clusters but as the instance type used on Hadoop had better configuration hence it performed better on the experiment with same input.

For 100 nodes and 1000 nodes if made available on those experiments then the performance will increase gradually every time any time extra node is added. This is because the virtual cores that is the CPU performance will increase each time any extra node is added

Sort Benchmark

We check the results of the sort benchmark and for the year 2013 and 2014, the winners of the benchmarking are Hadoop and Apache Spark respectively.

Comparing the result of Hadoop system that won the 2013 award for sort, the configuration of the system is much better than the instances that we use for our experiments. Hence the performance is incomparable as the system specification is better for the Daytona Hadoop. The Hadoop system 2013 winner produced a throughput of 1.42TB/min which is very high as compared to the throughput that we receive with those experiments.

Similarly, Apache Spark winner of 2014 produces a throughput of 4.27Tb/min which has system of 32 Virtual cores and a memory of 244GB with 6400 GB SSD. This configuration is pretty less as compared to c3. large instances providing 2 vCPU cores on which this experiments have been tested.

Hence we get lesser throughput gradually with respect to the specifications on which both the experiments have been performed.

CloudSort:

This is benchmark based on external sorting basically to produce intended results on cloud envoirments with huge IO workload.
All the huge cloud platforms gradually write off the initial cost, still having large profitability by running the cloud services. Thus this paper defines the efficiency of external sort from total cost ownership i.e. asking itself the minimum cost for sorting a fixed number of records on any public cloud.

Thus it explains the advantages of public cloud that satisfy some issues found in the other sort benchmarks. Those advantages being,
Accessibility: Cloud sort levels the users. Even the groups with limited budget can develop their implementation and run their benchmarks.
Affordability: Comparison of AWS pricing for running sort benchmark is manageable for groups with modest budgets.
Auditability: Provides good detective work on the part of auditors.

The paper on Cloud Sort also states about the rules defined for the public cloud on the factors such as Pricing, Storage, Reliability and Reporting.

Hence it can be concluded that the cloud platforms are poorly provisioned for IO intensive workloads. The cloud sort benchmark will help to initiate innovation in the public cloud for supporting IO Intensive tasks.

As the cloud grows gradually, the benchmark will provide you best platforms for building your data fields but also find efficient sort implementation from Total cost of ownership perspective.