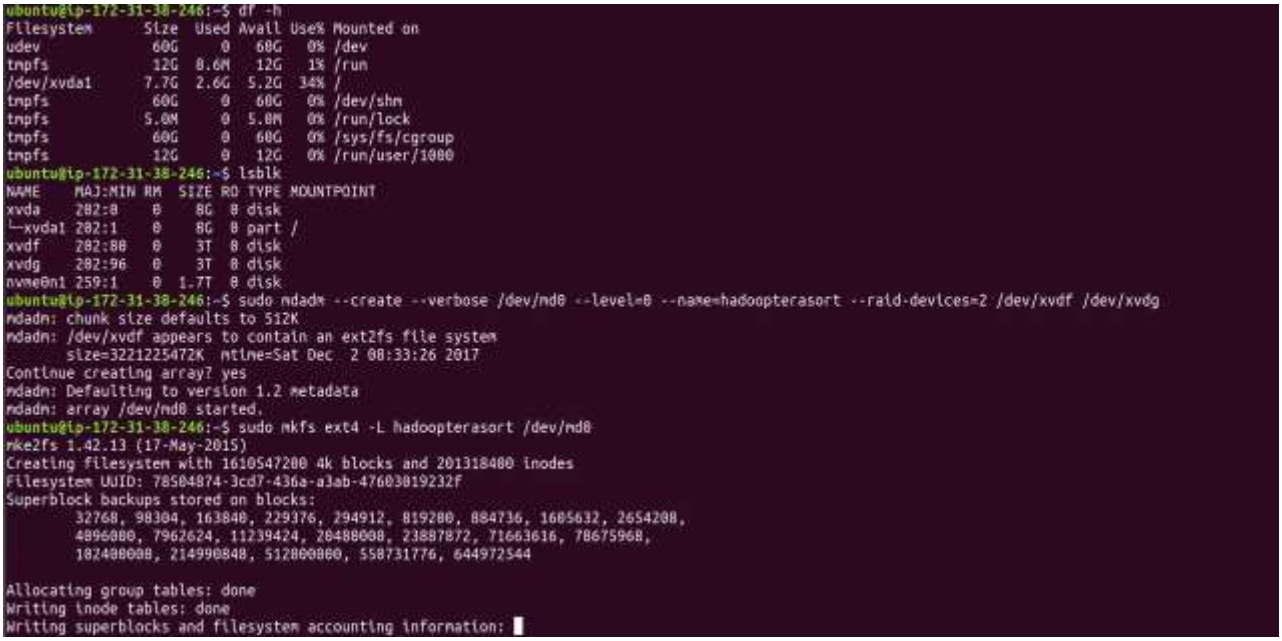
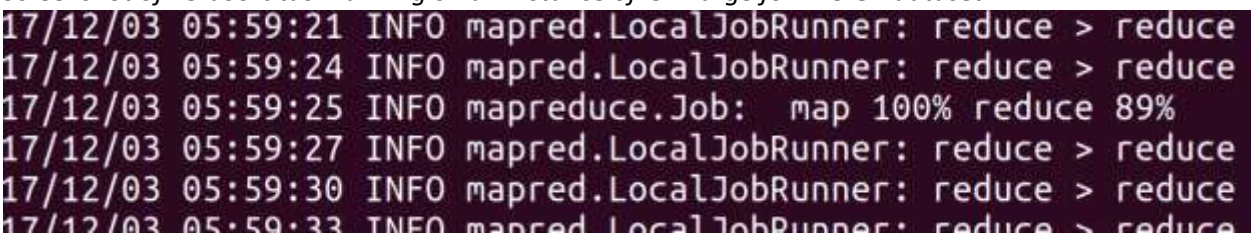
Sandeep Vuzzini [A20243379]
Abhinav Pimpalgaonkar[A20387324]
Shruthi Akkala [A20376505]

SCREENSHOTS

Raid Creation for Data Node

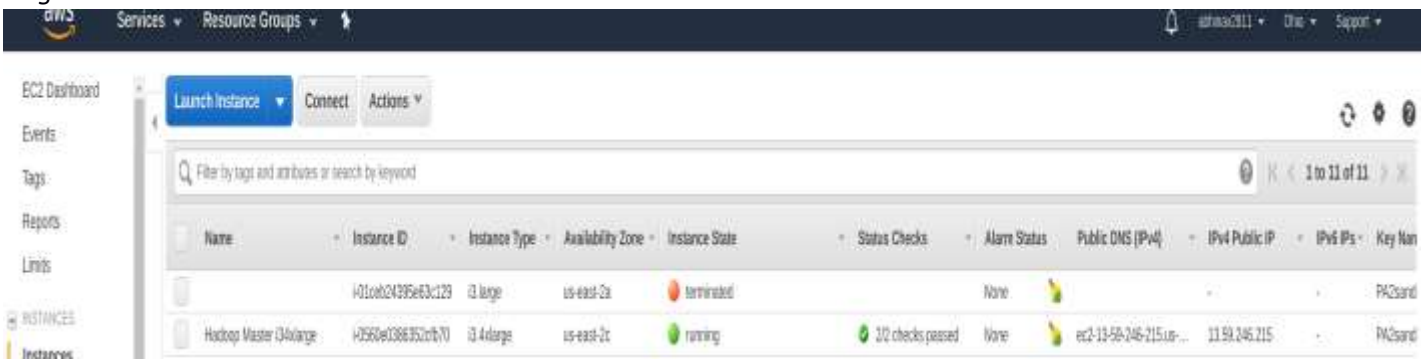


Screenshot of TeraSort task running on an instance of i34xlarge for 128 GB dataset



Sandeep Vuzzini [A20243379]
Abhinav Pimpalgaonkar[A20387324]
Shruthi Akkala [A20376505]

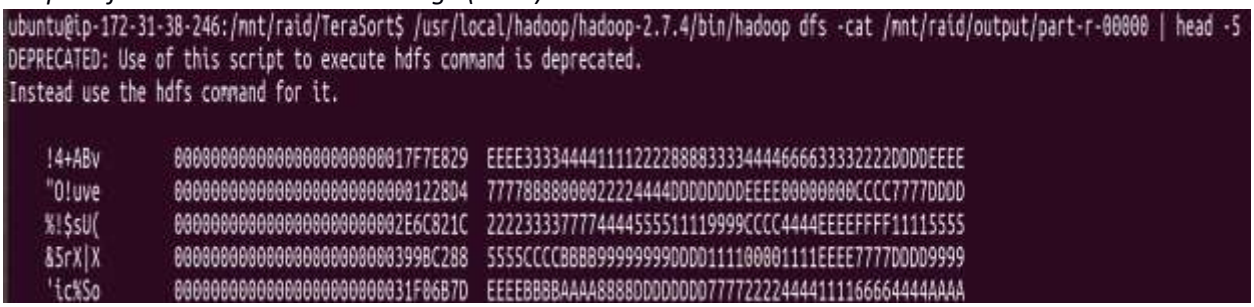
Single Node instance



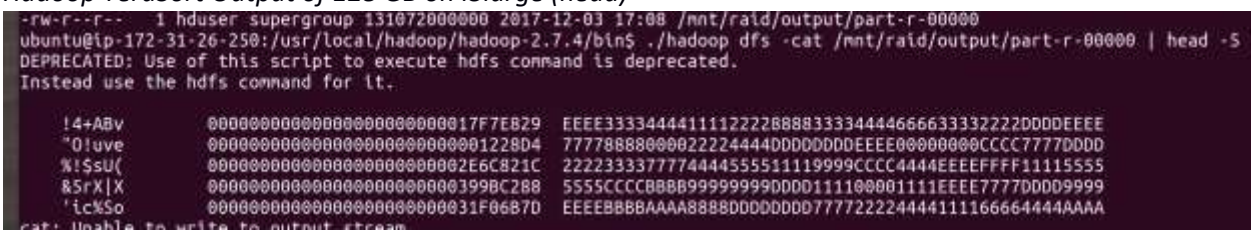
Cluster instance

List of nodes running with JPS

Output of 128 GB dataset on i34xlarge (head)



Hadoop TeraSort Output of 128 GB on i3large (head)



Sandeep Vuzzini [A20243379]
Abhinav Pimpalgaonkar[A20387324]
Shruthi Akkala [A20376505]

SparkOutput on 128 GB dataset head

```
ubuntu@ip-172-31-38-246:/mnt/raid/SparkOutput$ ^C
ubuntu@ip-172-31-38-246:/mnt/raid/SparkOutput$ head -5 part-00000
    !4+ABv    0000000000000000000000017F7E829  EEEE333344441111222288883333444466663333222DDDDEEEE
    "O!uve    00000000000000000000000001228D4  77778888000022224444DDDDDDDDEEEE00000000CCCC7777DDDD
    %!$sU(    0000000000000000000000002E6C821C  22223333777744445555111199999CCCC4444EEEEFFFF11115555
    &5rX|X    00000000000000000000000399BC288  5555CCCCBBBB99999999DDDD11110001111EEEE7777DDDD9999
    'ic%So    000000000000000000000000031F06B7D  EEEEBBBBAAAA8888DDDDDDDD777722224444111166664444AAAA
```

SparkOutput on 128 GB tail

```
ubuntu@ip-172-31-38-246:/mnt/raid/SparkOutput$ tail -5 part-00488
~~~~UeTR]s    0000000000000000000000003E299FE5  DDDD0000CCCC22227777DDDD555588880000000222222222222
~~~~ZuHH~L    00000000000000000000000004320332A  111188882222CCCC6666CCCC8888CCCC88882222AAAADDDD7777
~~~~c+I&cP    0000000000000000000000000074BDF64  888800005555000DDDD22227777AAAA000033332222AAAADDDD
~~~~hb&5X*    000000000000000000000000032C0E06B  7777BBBBBBBB9999EEEEAAAAAAAA0000CCCCDDDD4444BBBB4444
~~~~lkLc*1    0000000000000000000000000045E4700F  9999777799991111AAAA22224444000011111CCCC9999FFFF0000
```

Input file generation

```
ubuntu@ip-172-31-38-246:/mnt/raid/SparkOutput$ ls -ltrh /mnt/raid/TeraSort/In
total 1.1T
-rwxr-xr-x 1 root root 954M Dec  2 20:59 a.txt
-rwxr-xr-x 1 root root 123G Dec  2 21:52 actualIn.txt
-rwxr-xr-x 1 root root 982G Dec  3 20:48 OneTBInput.txt
ubuntu@ip-172-31-38-246:/mnt/raid/SparkOutput$
```

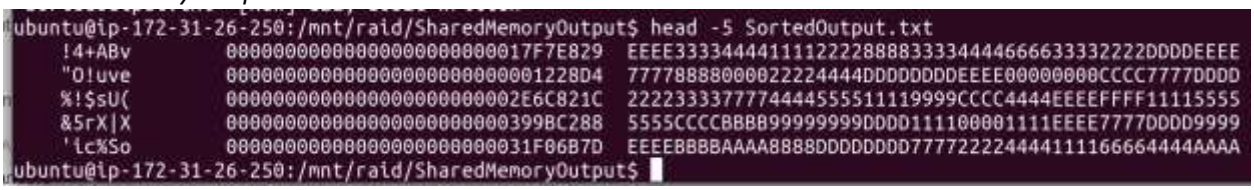
HDFS Input File

```
ubuntu@ip-172-31-38-246:/mnt/raid/SparkOutput$ /usr/local/hadoop/hadoop-2.7.4/bin/hadoop dfs -ls /mnt/raid/HDFSIn
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

Found 1 items
-rw-r--r--   1 hduser supergroup 131072000000 2017-12-03 03:26 /mnt/raid/HDFSIn/actualIn.txt
ubuntu@ip-172-31-38-246:/mnt/raid/SparkOutput$
```

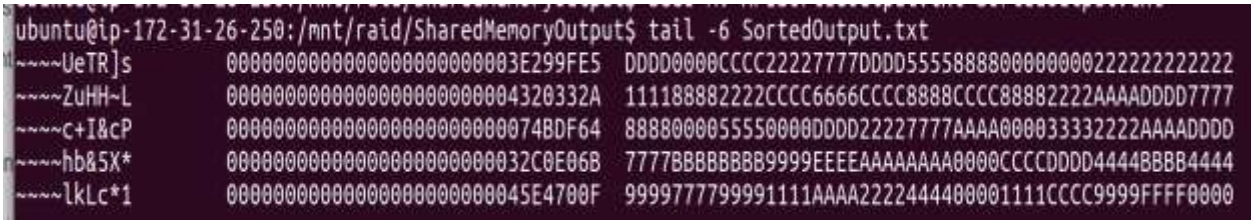
Terasort output tail 128 GB

```
ubuntu@ip-172-31-26-250:/usr/local/hadoop/hadoop-2.7.4/bin$ ./hadoop dfs -cat /mnt/raid/output/part-r-00000 | tail -5
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

~~~~UeTR]s       0000000000000000000000003E299FE5  DDDD0000CCCC22227777DDDD555588880000000222222222222
~~~~ZuHH~L       00000000000000000000000004320332A  111188882222CCCC6666CCCC8888CCCC88882222AAAADDDD7777
~~~~c+I&cP       0000000000000000000000000074BDF64  888800005555000DDDD22227777AAAA000033332222AAAADDDD
~~~~hb&5X*       000000000000000000000000032C0E06B  7777BBBBBBBB9999EEEEAAAAAAAA0000CCCCDDDD4444BBBB4444
~~~~lkLc*1       0000000000000000000000000045E4700F  9999777799991111AAAA22224444000011111CCCC9999FFFF0000
```

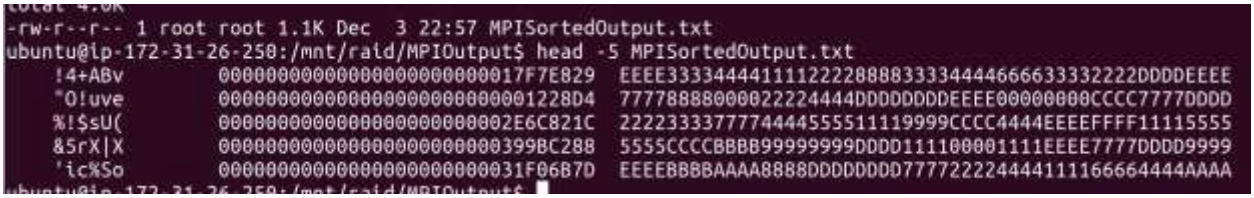
*Shared Memory Output **head***

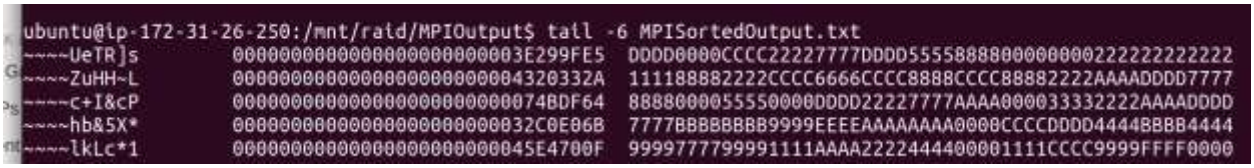


*Shared Memory Output **tail***



*MPI Sorted Output **head for 128 GB dataset***





ValSort Records FOR SPARK 1 file generated

ValSort Records FOR SPARK last file generated

Sandeep Vuzzini [A20243379]
Abhinav Pimpalgaonkar[A20387324]
Shruthi Akkala [A20376505]

Hadoop Terasort Output on 1 TB dataset for a Single Node Instance & Multi Node Instance (Please note that the output was same for difference instance configurations, hence pasting a single screenshot for all instances to avoid confusion)

```
ubuntu@ip-172-31-26-250:/usr/local/hadoop/hadoop-2.7.4/bin$ ./hadoop dfs -cat /mnt/raid/output/part-r-00000 | head -5
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

    !"L5R;  0000000000000000000000001A68EC8A8  AAAA0000BBBBDDDD4444AAAAFFFF1111DDDD2222DDDD33339999
    !4+ABv  00000000000000000000000017F7E829  EEEE333344441111222288883333444466663333222DDDDEEEE
    "O!uve  000000000000000000000000001228D4  77778888000022224444DDDDDDDEEEE00000000CCCC7777DDDD
    #Eejl=  00000000000000000000000020A68BA17  22222222FFFF7777DDDDEEEE8888EEEEEEEE6666666666668888
    $h#_DN  000000000000000000000001BCB770C2  55554444CCCCCCCC4444222211119999555522227777 1111FFFF
```

```
ubuntu@ip-172-31-26-250:/usr/local/hadoop/hadoop-2.7.4/bin$ ./hadoop dfs -cat /mnt/raid/output/part-r-00000 | tail -5
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

~~~UeTR]s       0000000000000000000000003E299FE5  DDDD0000CCCC22227777DDDD5555888800000000222222222222
~~~ZuHH~L       00000000000000000000000004320332A  11118888222CCCC6666CCCC8888CCCC88882222AAAADDDD7777
~~~c+I&cP       000000000000000000000000074BDF64  888800005555000DDDD22227777AAAA000033332222AAAADDDD
~~~hb&5X*       00000000000000000000000032C0E06B  7777BBBBBBBB9999EEEEAAAAAAAA0000CCCCDDDD4444BBBB4444
~~~lkLc*1       000000000000000000000000045E4700F  999977779999111 1AAAA2222444400001111CCCC9999FFFF0000
```

Spark Terasort Output on 1 TB dataset for a Single Node Instance & Multi Node Instance (Please note that the output was same for difference instance configurations, hence pasting a single screenshot for all instances to avoid confusion)

```
ubuntu@ip-172-31-26-250:/usr/local/hadoop/hadoop-2.7.4/bin$ ./hadoop dfs -cat /mnt/raid/SparkOutput/part-r-00000 | head -5
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

    !"L5R;  0000000000000000000000001A68EC8A8  AAAA0000BBBBDDDD4444AAAAFFFF1111DDDD2222DDDD33339999
    !4+ABv  00000000000000000000000017F7E829  EEEE333344441111222288883333444466663333222DDDDEEEE
    "O!uve  000000000000000000000000001228D4  77778888000022224444DDDDDDDEEEE00000000CCCC7777DDDD
    #Eejl=  00000000000000000000000020A68BA17  22222222FFFF7777DDDDEEEE8888EEEEEEEE6666666666668888
    $h#_DN  000000000000000000000001BCB770C2  55554444CCCCCCCC4444222211119999555522227777 1111FFFF
    %q/-'7  000000000000000000000000A674E940  7777FFFF1111EEEE4444DDDDCCCC8888AAAACCCC1111CCCC1111
```

```
ubuntu@ip-172-31-26-250:/usr/local/hadoop/hadoop-2.7.4/bin$ ./hadoop dfs -cat /mnt/raid/SparkOutput/part-r-01952 | tail -5
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

~~~UeTR]s       0000000000000000000000003E299FE5  DDDD0000CCCC22227777DDDD5555888800000000222222222222
~~~ZuHH~L       00000000000000000000000004320332A  11118888222CCCC6666CCCC8888CCCC88882222AAAADDDD7777
~~~c+I&cP       000000000000000000000000074BDF64  888800005555000DDDD22227777AAAA000033332222AAAADDDD
~~~hb&5X*       00000000000000000000000032C0E06B  7777BBBBBBBB9999EEEEAAAAAAAA0000CCCCDDDD4444BBBB4444
~~~lkLc*1       000000000000000000000000045E4700F  999977779999111 1AAAA2222444400001111CCCC9999FFFF0000
```

Sandeep Vuzzini [A20243379]
Abhinav Pimpalgaonkar[A20387324]
Shruthi Akkala [A20376505]

Shared Memory TeraSort Output on 1 TB dataset for a Single Node Instance & Multi Node Instance (Please note that the output was same for difference instance configurations, hence pasting a single screenshot for all instances to avoid confusion)

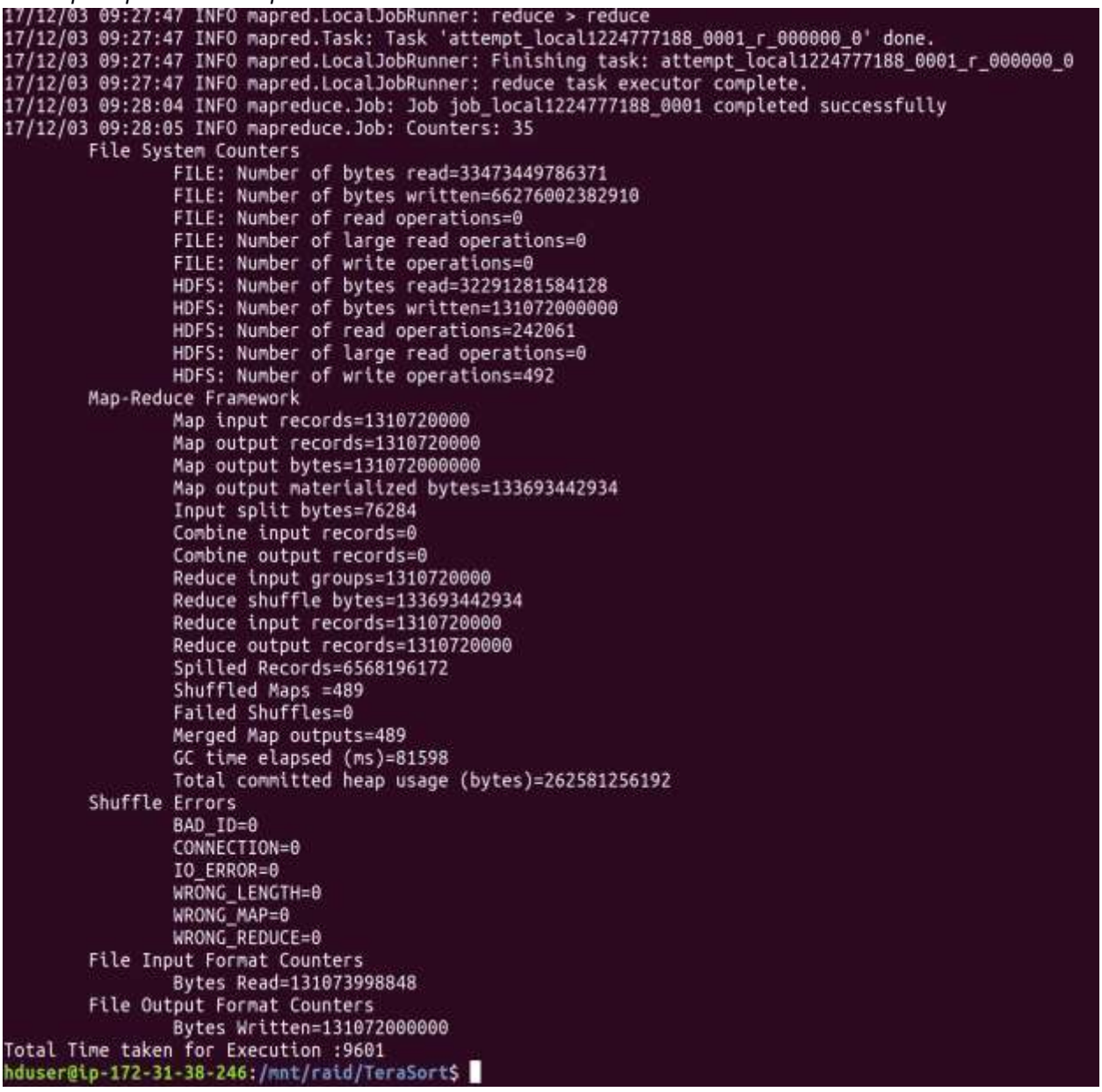
```
ubuntu@ip-172-31-26-250:cat /mnt/raid/SharedMemoryOutputfor1TB/SortedOutput.txt | head -5

    !"L5R;  0000000000000000000001A68EC8A8  AAAA0000BBBBDDDD4444AAAAFFFF1111DDDD2222DDDD33339999
    !4+ABv  00000000000000000000017F7E829  EEEE333344441111222288883333444466663333222DDDDEEEE
    "O!uve  000000000000000000000000001228D4  77778888000022224444DDDDDDDDEEEE00000000CCCC7777DDDD
    $h#_DN  0000000000000000000001BCB770C2  55554444CCCCCCCC444422221111999955552222777711111FFFF
    %q/-'7  00000000000000000000000A674E940  7777FFFF1111EEEE4444DDDDCCCC8888AAAACCCC1111CCCC1111

ubuntu@ip-172-31-26-250:cat /mnt/raid/SharedMemoryOutputfor1TB/SortedOutput.txt | tail -5

~~~UeTR]s       0000000000000000000003E299FE5  DDDD0000CCCC22227777DDDD555588880000000222222222222
~~~ZuHH~L       000000000000000000004320332A  111188882222CCCC6666CCCC8888CCCC88882222AAAADDDD7777
~~~c+I&cP       00000000000000000000074BDF64  888800005555000DDDD22227777AAAA000033332222AAAADDDD
~~~hb&5X*       0000000000000000000032C0E06B  7777BBBBBBBB9999EEEEAAAAAAAA0000CCCCDDDD4444BBBB4444
~~~lkLc*1       00000000000000000000045E4700F  9999777799991111AAAA2222444400001111CCCC9999FFFF0000
```

MPI TeraSort Output on 1 TB dataset for a Single Node Instance & Multi Node Instance (Please note that the output was same for difference instance configurations, hence pasting a single screenshot for all instances to avoid confusion)

```
ubuntu@ip-172-31-26-250:cat /mnt/raid/MPIOutputfor1TB/MPISortedOutput.txt | head -6

    !"L5R;  0000000000000000000001A68EC8A8  AAAA0000BBBBDDDD4444AAAAFFFF1111DDDD2222DDDD33339999
    !4+ABv  00000000000000000000017F7E829  EEEE333344441111222288883333444466663333222DDDDEEEE
    "O!uve  000000000000000000000000001228D4  77778888000022224444DDDDDDDDEEEE00000000CCCC7777DDDD
    #Eejl=  0000000000000000000000020A68BA17  22222222FFFF7777DDDDEEEE8888EEEEEEEE6666666666668888
    $h#_DN  0000000000000000000001BCB770C2  55554444CCCCCCCC444422221111999955552222777711111FFFF
    %q/-'7  00000000000000000000000A674E940  7777FFFF1111EEEE4444DDDDCCCC8888AAAACCCC1111CCCC1111

ubuntu@ip-172-31-26-250:cat /mnt/raid/MPIOutputfor1TB/MPISortedOutput.txt | tail -5

~~~UeTR]s       0000000000000000000003E299FE5  DDDD0000CCCC22227777DDDD555588880000000222222222222
~~~ZuHH~L       000000000000000000004320332A  111188882222CCCC6666CCCC8888CCCC88882222AAAADDDD7777
~~~c+I&cP       00000000000000000000074BDF64  888800005555000DDDD22227777AAAA000033332222AAAADDDD
~~~hb&5X*       0000000000000000000032C0E06B  7777BBBBBBBB9999EEEEAAAAAAAA0000CCCCDDDD4444BBBB4444
~~~lkLc*1       00000000000000000000045E4700F  9999777799991111AAAA2222444400001111CCCC9999FFFF0000
```

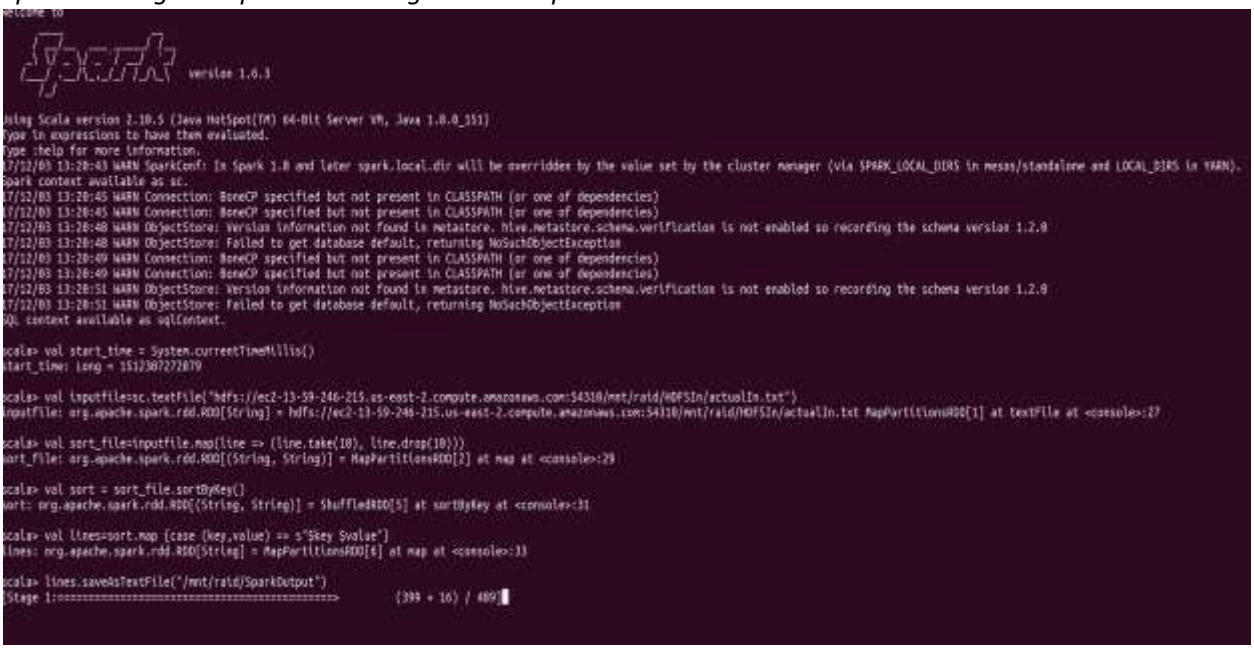
Sandeep Vuzzini [A20243379]
Abhinav Pimpalgaonkar[A20387324]
Shruthi Akkala [A20376505]

Hadoop Map Reduce sample screenshot



Sandeep Vuzzini [A20243379]
Abhinav Pimpalgaonkar[A20387324]
Shruthi Akkala [A20376505]

Spark running example on i3 4xlarge with sample dataset:



Sandeep Vuzzini [A20243379]
Abhinav Pimpalgaonkar[A20387324]
Shruthi Akkala [A20376505]

Hadoop TeraSort on Single Node i3large Instance for 128 GB dataset

```
17/12/03 17:08:04 INFO mapred.LocalJobRunner: reduce > reduce
17/12/03 17:08:04 INFO mapred.Task: Task 'attempt_local372026054_0001_r_000000_0' done.
17/12/03 17:08:04 INFO mapred.LocalJobRunner: Finishing task: attempt_local372026054_0001_r_000000_0
17/12/03 17:08:04 INFO mapred.LocalJobRunner: reduce task executor complete.
17/12/03 17:08:32 INFO mapreduce.Job: Job job_local372026054_0001 completed successfully
17/12/03 17:08:32 INFO mapreduce.Job: Counters: 35
        File System Counters
                FILE: Number of bytes read=33495978735672
                FILE: Number of bytes written=66298530699298
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=32291281584128
                HDFS: Number of bytes written=131072000000
                HDFS: Number of read operations=242061
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=492
        Map-Reduce Framework
                Map input records=1310720000
                Map output records=1310720000
                Map output bytes=131072000000
                Map output materialized bytes=133693442934
                Input split bytes=76773
                Combine input records=0
                Combine output records=0
                Reduce input groups=1310720000
                Reduce shuffle bytes=133693442934
                Reduce input records=1310720000
                Reduce output records=1310720000
                Spilled Records=6789068216
                Shuffled Maps =489
                Failed Shuffles=0
                Merged Map outputs=489
                GC time elapsed (ms)=100121
                Total committed heap usage (bytes)=262128795648
        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=131073998848
        File Output Format Counters
                Bytes Written=131072000000
Total Time taken for Execution :16809
```

RAID configuration

----------END of DOCUMENT----------