

- Gensort 128GB file generated:

```
ubuntu@ip-172-31-9-44: ~ — ssh -i AhmedShared.pem ubuntu@ec2-34-212-113-254.us-west-2.compute.amazonaws.com
ubuntu@ip-172-31-9-44:~$ ./gensort -a 1374389534 input128.txt
reall 41m7.456s
user 4m15.488s
sys 0ms
[ubuntu@ip-172-31-9-44:~$
[ubuntu@ip-172-31-9-44:~$
```

- File moved to HDFS

```
...ersonal/MCS/Cloud Computing/Ahmed/PA2 — ubuntu@ip-172-31-9-44: ~ — ssh -i AhmedShared.pem ubuntu@ec2-34-212-113-254.us-west-2.compute.
ubuntu@ip-172-31-9-44:~$ time hdfs dfs -copyFromLocal sortInput.txt /input4
reall 46m24.768s
user 17m26.528s
sys 3m44.768s1
[ubuntu@ip-172-31-9-44:~$
[ubuntu@ip-172-31-9-44:~$
[ubuntu@ip-172-31-9-44:~$
[ubuntu@ip-172-31-9-44:~$
```

- File Sorted Output

```
ubuntu@ip-172-31-9-44: ~ — ssh -i AhmedShared.pem ubuntu@ec2-34-212-113-254.us-west-2.compute.amazonaws.com
```

```
ubuntu@ip-172-31-9-44:~$ time hadoop jar hterasort.jar /input4 /output5
```

```
17/12/03 18:54:23 INFO mapreduce.Job: map 100% reduce 100%
17/12/03 18:54:23 INFO mapreduce.Job: Job job_local1949456664_0001 completed successfully
17/12/03 18:54:23 INFO mapreduce.Job: Counters: 35
```

File System Counters

```
FILE: Number of bytes read=1051417243136
FILE: Number of bytes written=1682648899712
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=5905752016
HDFS: Number of bytes written=1073741800
HDFS: Number of read operations=118
HDFS: Number of large read operations=0
HDFS: Number of write operations=11
```

Map-Reduce Framework

```
Map input records=343597376
Map output records=343597376
Map output bytes=1073741800
Map output materialized bytes=1095216684
Input split bytes=720
Combine input records=343597376
Combine output records=343597376
Reduce input groups=343597376
Reduce shuffle bytes=1095216684
Reduce input records=343597376
Reduce output records=343597376
Spilled Records=40032212254
Shuffled Maps =256
Failed Shuffles=0
Merged Map outputs=256
GC time elapsed (ms)=213767
Total committed heap usage (bytes)=126013669376
```

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=1073770472
```

File Output Format Counters

```
Bytes Written=1073741800
```

```
real    192m38.506s
```

```
user    150m22.120s
```

```
sys     14m6.892s
```

```
ubuntu@ip-172-31-9-44:~$
```

```
#####
```

I3.4xlarge

```
#####
```

Observation and result of this category

- Gensort 1TB generated

```
ubuntu@ip-172-31-9-44: ~ — ssh -i AhmedShared.pem ubuntu@ec2-34-212-113-254.us-west-2.compute.amazonaws.com
```

```
ubuntu@ip-172-31-23-103:$ time ./gensort -a 10995116278 input.txt
real    110m44.192s
user    67m57.184s
sys     0m0.534s
```

```
ubuntu@ip-172-31-23-103:$ ls -lash input.txt
1T -rwxrwxr-x 1 ubuntu ubuntu 1T Dec  3 13:36 input.txt
```

- Move file to DFS

```
ubuntu@ip-172-31-9-44: ~ — ssh -i AhmedShared.pem ubuntu@ec2-34-212-113-254.us-west-2.compute.amazonaws.com
```

```
ubuntu@ip-172-31-23-103:~$ time hdfs dfs -copyFromLocal input.txt /input3
```

```
real    37m41.565s
user    28m40.724s
sys     0m0.820s
ubuntu@ip-172-31-23-103:~$ ls -lash input.txt
1T -rwxrwxr-x 1 ubuntu ubuntu 1T Dec  3 14:18 input.txt
```

- Input File Sorted

```
ubuntu@ip-172-31-9-44: ~ — ssh -i AhmedShared.pem ubuntu@ec2-34-212-113-254.us-west-2.compute.amazonaws.com
```

```
ubuntu@ip-172-31-23-103:$ time hadoop jar hterasort.jar /input /output
17/12/02 13:40:48 INFO Configuration.deprecation: session.id is deprecated. Instead, use dfs.metrics.
17/12/02 13:40:48 INFO jvm.JvmMetrics: Initializing JVM Metrics with processName=JobTracker, sessionI
17/12/02 13:40:48 INFO input.FileInputFormat: Total input files to process : 1
17/12/02 13:40:48 INFO mapreduce.JobSubmitter: number of splits:32
```

```

ubuntu@ip-172-31-23-103:~$ time hadoop jar hterasort.jar /input4 /output5 >out2 2>&1
17/12/03 18:54:23 INFO mapreduce.Job: map 100% reduce 100%
17/12/03 18:54:23 INFO mapreduce.Job: Job job_local1949456664_0001 completed successfully
17/12/03 18:54:23 INFO mapreduce.Job: Counters: 35
  File System Counters
    FILE: Number of bytes read=8411337945088
    FILE: Number of bytes written=13461191197696
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=3785587042256
    HDFS: Number of bytes written=6611028262600
    HDFS: Number of read operations=726526
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=67727
  Map-Reduce Framework
    Map input records=34359737667655
    Map output records=3435973765677
    Map output bytes=10667773741800
    Map output materialized bytes=10975425216684
    Input split bytes=23040
    Combine input records=34359737667655
    Combine output records=3435973765677
    Reduce input groups=3435934567376
    Reduce shuffle bytes=10952164532684
    Reduce input records=34359737667655
    Reduce output records=3435973765677
    Spilled Records=400322532512254
    Shuffled Maps =2048
    Failed Shuffles=0
    Merged Map outputs=2048
    GC time elapsed (ms)=309767
    Total committed heap usage (bytes)=195826013669376
  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=1095350878208
  File Output Format Counters
    Bytes Written=1095350878208

real    225m49.506s
user    179m12.120s
sys     13m670.007s

```

#####

i3.large 1TB

#####

Observation and result of this category

```

ubuntu@ip-172-31-6-96:~$ hadoop jar hadoop-2.7.4/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.4.jar teragen 85800345 /teraInput8GB
17/12/03 01:41:48 INFO client.RMProxy: Connecting to ResourceManager at ec2-34-216-149-36.us-west-2.compute.amazonaws.com/172.31.6.96:9076
17/12/03 01:41:49 INFO terasort.TeraSort: Generating 85800345 using 2
17/12/03 01:41:50 INFO mapreduce.JobSubmitter: number of splits:2
17/12/03 01:41:50 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1512265063012_0001
17/12/03 01:41:50 INFO impl.YarnClientImpl: Submitted application application_1512265063012_0001
17/12/03 01:41:51 INFO mapreduce.Job: The url to track the job: http://ec2-34-216-149-36.us-west-2.compute.amazonaws.com:9078/proxy/application_1512265063012_0001/
17/12/03 01:41:51 INFO mapreduce.Job: Running job: job_1512265063012_0001

```

```


PA2 — ubuntu@ec2-34-215-219-28: ~ — ssh -i AhmedShared.pem ubuntu@ec2-34-215-2...
s-west-2.compute.amazonaws.com/172.31.6.96:9076
17/12/02 07:36:25 INFO impl.YarnClientImpl: Killed application application_1512196130380_0001
Killed job job_1512196130380_0001
ubuntu@ip-172-31-6-96:~$ hadoop job -list
DEPRECATED: Use of this script to execute mapred command is deprecated.
Instead use the mapred command for it.

17/12/02 07:36:30 INFO client.RMProxy: Connecting to ResourceManager at ec2-34-215-219-28.us-west-2.compute.amazonaws.com/172.31.6.96:9076
Total jobs:0

```

JobId	State	StartTime	UserName	Queue	P
priority	UsedContainers	RsvdContainers	UsedMem	RsvdMem	NeededMem

AM info


Application application_1512196130380_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

User: ubuntu
Name: TeraGen
Application Type: MAPREDUCE
Application Tags:
YarnApplicationState: ACCEPTED: waiting for AM container to be allocated, launched and register with RM.
Queue: default
FinalStatus Reported by AM: Application has not completed yet.
Started: Sat Dec 02 06:41:38 +0000 2017
Elapsed: 3mins, 10sec
Tracking URL: ApplicationMaster
Diagnostics:

Application Overview

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:	0 MB-seconds, 0 vcore-seconds

Show: 20 entries

Attempt ID

Started

Node

Logs

Blacklisted Nodes

appattempt_1512196130380_0001_000001
Sat Dec 2 00:41:39 -0600 2017
http://
Logs
0

Showing 1 to 1 of 1 entries

First
Previous
1
Next
Last

17/12/02 12:22:32 INFO mapreduce.Job: Job job_local717214462_0001 completed successfully

17/12/02 12:22:32 INFO mapreduce.Job: Counters: 45

File System Counters

FILE: Number of bytes read=58526155002345
FILE: Number of bytes written=93991308536775
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=390784688917066
HDFS: Number of bytes written=7900410385070
HDFS: Number of read operations=770750
HDFS: Number of large read operations=0
HDFS: Number of write operations=90305

Map-Reduce Framework

Map input records=4450074413
Map output records=4450074413
Map output bytes=335007441600
Map output materialized bytes=4450074413
Input split bytes=512000
Combine input records=4450074413
Combine output records=4450074413
Reduce input groups=323435932453
Reduce shuffle bytes=45427802345
Reduce input records=4450074413
Reduce output records=4450074413
Spilled Records=4523513376
Shuffled Maps =2048
Failed Shuffles=0
Merged Map outputs=2048
GC time elapsed (ms)=243890
Total committed heap usage (bytes)=144651859986

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=1099511627776

File Output Format Counters

Bytes Written=1099511627776

real 270m6.902s
user 267m34.756s
sys 0m54.543s

ubuntu@ip-172-31-6-96:~\$