

I declare that the assignment submitted on Elearning system is original except for source material explicitly acknowledged, and that the same or related material has not been previously submitted for another course. I also acknowledge that I am aware of University policy and regulations on honesty in academic work, and of the disciplinary guidelines and procedures applicable to breaches of such policy and regulations, as contained in the website

<http://www.cuhk.edu.hk/policy/academic/honesty/>

Signed (Student Wu Yan) Date: 2020.09.25
Name WU Yan SID 1155148594

a. Single-node Hadoop Setup

i. The namenode web page(attached in the end)

You can also check the print page at the end of this PDF file

HadoopOverviewDatanodesDatanode Volume FailuresSnapshotStartup ProgressUtilities

Overview'hadoop-master:9000'(active)

Started:	Sat Sep 19 21:08:47 +0800 2020
Version:	2.9.2, r826afbeae31ca687bc2f8471dc841b66ed2c6704
Compiled:	Tue Nov 13 20:42:00 +0800 2018 by ajsaka from branch-2.9.2
Cluster ID:	CID-11539d2d-c608-4b9e-9ed2-a6a890677514
Block Pool ID:	BP-945784968-172.31.35.148-1600520911160

Summary

Security is off.
Safemode is off.
66 files and directories, 32 blocks = 98 total filesystem object(s).
Heap Memory used 100.35 MB of 272 MB Heap Memory. Max Heap Memory is 889 MB.
Non Heap Memory used 60.71 MB of 61.81 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	290.63 GB
DFS Used:	10.93 MB (0%)

ii. The Terasort example running process


1. Run the commands in Putty

```
ubuntu@ec2-54-158-67-211: /usr/local/hadoop/share/hadoop/mapreduce$ cd /usr/local/hadoop/share/hadoop/mapreduce
ubuntu@ec2-54-158-67-211: /usr/local/hadoop/share/hadoop/mapreduce$ hadoop jar hadoop-mapreduce-examples-2.9.2.jar teragen 100000 terasort terasort/input
teragen <num rows> <output dir>
ubuntu@ec2-54-158-67-211: /usr/local/hadoop/share/hadoop/mapreduce$ hadoop jar hadoop-mapreduce-examples-2.9.2.jar teragen 100000 terasort/input
20/09/17 14:22:02 INFO mapreduce.Job: INFO client.RMProxy: Connecting to ResourceManager at hadoop-master/172.31.77.85:8032
20/09/17 14:22:03 INFO terasort.Terasort: Generating 100000 using 2
20/09/17 14:22:03 INFO mapreduce.JobSubmitter: number of splits:2
20/09/17 14:22:03 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
20/09/17 14:22:03 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1600351918250_0005
20/09/17 14:22:03 INFO impl.YarnClientImpl: Submitted application application_1600351918250_0005
20/09/17 14:22:03 INFO mapreduce.Job: The url to track the job: http://hadoop-master:8088/proxy/application_1600351918250_0005/
20/09/17 14:22:03 INFO mapreduce.Job: Running job: job_1600351918250_0005
20/09/17 14:22:09 INFO mapreduce.Job: Job job_1600351918250_0005 running in uber mode : false
20/09/17 14:22:09 INFO mapreduce.Job: map 0% reduce 0%
20/09/17 14:22:22 INFO mapreduce.Job: Task Id : attempt_1600351918250_0005_m_000001_0, Status : FAILED
20/09/17 14:22:22 INFO mapreduce.Job: Task Id : attempt_1600351918250_0005_m_000000_0, Status : FAILED
20/09/17 14:22:35 INFO mapreduce.Job: Task Id : attempt_1600351918250_0005_m_000000_1, Status : FAILED
20/09/17 14:22:36 INFO mapreduce.Job: Task Id : attempt_1600351918250_0005_m_000001_1, Status : FAILED
20/09/17 14:22:42 INFO mapreduce.Job: map 100% reduce 0%
20/09/17 14:22:44 INFO mapreduce.Job: Job job_1600351918250_0005 completed successfully
20/09/17 14:22:44 INFO mapreduce.Job: counters: 32
File System Counters
FILE: Number of bytes read=0
FILE: Number of bytes written=386112
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=164
HDFS: Number of bytes written=10000000
HDFS: Number of read operations=6
HDFS: Number of large read operations=0
HDFS: Number of write operations=4
Job Counters
Failed map tasks=4
Launched map tasks=6
Other local map tasks=6
Total time spent by all maps in occupied slots (ms)=54616
Total time spent by all reducers in occupied slots (ms)=0
Total time spent by all map tasks (ms)=54616
Total vcore-milliseconds taken by all map tasks=54616
Total megabyte-milliseconds taken by all map tasks=83890176
Map-Reduce Framework
Map input records=100000
Map output records=100000
```

```
MTPUTTY (Multi-Tabbed PuTTY)
Server View Tools Help
<Quick send keys> Send Send to all
Servers
Search
Start page ubuntu@ec2-54-158-67-211: /usr/local/hadoop/share/hadoop/mapreduce X ubuntu@ec2-54-187-143-91: ~ X ubuntu@ec2-107-22-52-231: ~ X ubuntu@ec2-3-223-185-27: ~ X
PUTTY sessions
AmazonEC2DataNode001 (54)
AmazonEC2DataNode002 (10)
AmazonEC2DataNode003 (3)
AmazonEC2DataNode004 (54)
WinSCP temporary session (3)
ubuntu@ec2-54-158-67-211: /usr/local/hadoop/share/hadoop/mapreduce$ hadoop jar hadoop-mapreduce-examples-2.9.2.jar terasort terasort/input terasort/output
BytesWritten:1000000
20/09/17 14:26:41 INFO terasort.TeraSort: starting
20/09/17 14:26:42 INFO input.FileInputFormat: Total input files to process : 2
Spent 59ms computing base-splits.
Spent 2ms computing TeraScheduler splits.
Computing input splits took 102ms
Sampling 2 splits of 2
Making 1 from 100000 sampled records
Computing partitions took 402ms
Spent 507ms computing partitions.
20/09/17 14:26:43 INFO client.RMProxy: Connecting to ResourceManager at hadoop-master/172.31.77.65:8032
20/09/17 14:26:43 INFO mapreduce.JobSubmitter: number of splits:2
20/09/17 14:26:43 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
20/09/17 14:26:43 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1600351918250_0007
20/09/17 14:26:43 INFO impl.YarnClientImpl: Submitted application application_1600351918250_0007
20/09/17 14:26:43 INFO mapreduce.Job: The url to track the job: http://hadoop-master:8088/proxy/application_1600351918250_0007/
20/09/17 14:26:43 INFO mapreduce.Job: Running job: job_1600351918250_0007
20/09/17 14:26:48 INFO mapreduce.Job: Job job_1600351918250_0007 running in uber mode : false
20/09/17 14:26:48 INFO mapreduce.Job: map 0% reduce 0%
20/09/17 14:27:02 INFO mapreduce.Job: Task Id : attempt_1600351918250_0007_m_000000_0, Status : FAILED
20/09/17 14:27:02 INFO mapreduce.Job: Task Id : attempt_1600351918250_0007_m_000001_0, Status : FAILED
20/09/17 14:27:08 INFO mapreduce.Job: map 50% reduce 0%
20/09/17 14:27:16 INFO mapreduce.Job: Task Id : attempt_1600351918250_0007_m_000001_1, Status : FAILED
20/09/17 14:27:24 INFO mapreduce.Job: map 50% reduce 17%
20/09/17 14:27:29 INFO mapreduce.Job: Task Id : attempt_1600351918250_0007_m_000001_2, Status : FAILED
20/09/17 14:27:35 INFO mapreduce.Job: map 100% reduce 17%
20/09/17 14:27:36 INFO mapreduce.Job: Job job_1600351918250_0007 completed successfully
20/09/17 14:27:36 INFO mapreduce.Job: Counters: 53
File System Counters
FILE: Number of bytes read=10400006
FILE: Number of bytes written=21401445
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=10000260
HDFS: Number of bytes written=10000000
HDFS: Number of read operations=9
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
Job Counters
Failed map tasks=4
Killed map tasks=1

MTPUTTY (Multi-Tabbed PuTTY)
Server View Tools Help
<Quick send keys> Send Send to all
Servers
Search
Start page ubuntu@ec2-54-158-67-211: /usr/local/hadoop/share/hadoop/mapreduce X ubuntu@ec2-54-187-143-91: ~ X ubuntu@ec2-107-22-52-231: ~ X ubuntu@ec2-3-223-185-27: ~ X
PUTTY sessions
AmazonEC2DataNode001 (54)
AmazonEC2DataNode002 (10)
AmazonEC2DataNode003 (3)
AmazonEC2DataNode004 (54)
WinSCP temporary session (3)
ubuntu@ec2-54-158-67-211: /usr/local/hadoop/share/hadoop/mapreduce$ hadoop jar hadoop-mapreduce-examples-2.9.2.jar teravalidate terasort/output terasort/check
20/09/17 14:32:16 INFO client.RMProxy: Connecting to ResourceManager at hadoop-master/172.31.77.65:8032
20/09/17 14:32:17 INFO input.FileInputFormat: Total input files to process : 1
Spent 16ms computing base-splits.
Spent 3ms computing TeraScheduler splits.
20/09/17 14:32:17 INFO mapreduce.JobSubmitter: number of splits:1
20/09/17 14:32:17 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
20/09/17 14:32:17 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1600351918250_0010
20/09/17 14:32:17 INFO impl.YarnClientImpl: Submitted application application_1600351918250_0010
20/09/17 14:32:17 INFO mapreduce.Job: The url to track the job: http://hadoop-master:8088/proxy/application_1600351918250_0010/
20/09/17 14:32:23 INFO mapreduce.Job: Job job_1600351918250_0010 running in uber mode : false
20/09/17 14:32:23 INFO mapreduce.Job: map 0% reduce 0%
20/09/17 14:32:28 INFO mapreduce.Job: map 100% reduce 0%
20/09/17 14:32:42 INFO mapreduce.Job: Task Id : attempt_1600351918250_0010_r_000000_0, Status : FAILED
20/09/17 14:32:56 INFO mapreduce.Job: Task Id : attempt_1600351918250_0010_r_000000_1, Status : FAILED
20/09/17 14:33:02 INFO mapreduce.Job: map 100% reduce 100%
20/09/17 14:33:02 INFO mapreduce.Job: Job job_1600351918250_0010 completed successfully
20/09/17 14:33:02 INFO mapreduce.Job: Counters: 50
File System Counters
FILE: Number of bytes read=92
FILE: Number of bytes written=399127
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=10000131
HDFS: Number of bytes written=22
HDFS: Number of read operations=6
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
Job Counters
Failed reduce tasks=2
Launched map tasks=1
Launched reduce tasks=3
Rack-local map tasks=1
Total time spent by all maps in occupied slots (ms)=2480
Total time spent by all reduces in occupied slots (ms)=25274
Total time spent by all map tasks (ms)=2480
Total time spent by all reduce tasks (ms)=25274
Total vcore-milliseconds taken by all map tasks=2480
Total vcore-milliseconds taken by all reduce tasks=25274
Total megabyte-milliseconds taken by all map tasks=389280
Total megabyte-milliseconds taken by all reduce tasks=38820864
Map-Reduce Framework
```

2. Check the logs in jobhistory web page



- Application
- Tools
- Configuration
- Local logs
- Server
- Stacks
- Server
- metrics

Log Type: syslog
Log Upload Time: Thu Sep 17 14:22:49 +0000 2020
Log Length: 72709

2020-09-17 14:22:04,787 INFO [main] org.apache.hadoop.mapreduce.v2.app.MRAppMaster: Created MRAppMaster for application appattemp_1600351918250_0005_

2020-09-17 14:22:04,953 INFO [main] org.apache.hadoop.mapreduce.v2.app.MRAppMaster: /*****

[system properties]

os.name: linux

os.version: 5.3.0-1035-aws

java.home: /usr/lib/jvm/java-8-openjdk-amd64/jre

java.runtime.version: 1.8.0_265-bu265-b01-0ubuntu2~18.04-b01

java.vendor: Private Build

java.version: 1.8.0_265

java.vm.name: OpenJDK 64-Bit Server VM

java.class.path: /home/ubuntu/home/hadoop.data/hdfs/tmp/nm-local-dir/usercache/ubuntu/appcache/application_1600351918250_0005/container_1600351918250_0005

java.io.tmpdir: /home/ubuntu/home/hadoop.data/hdfs/tmp/nm-local-dir/usercache/ubuntu/appcache/application_1600351918250_0005/container_1600351918250_0005

user.dir: /home/ubuntu/home/hadoop.data/hdfs/tmp/nm-local-dir/usercache/ubuntu/appcache/application_1600351918250_0005/container_1600351918250_0005

user.name: ubuntu

2020-09-17 14:22:05,004 INFO [main] org.apache.hadoop.security.SecurityUtil: Updating Configuration

2020-09-17 14:22:05,013 INFO [main] org.apache.hadoop.mapreduce.v2.app.MRAppMaster: Executing with tokens:

2020-09-17 14:22:05,013 INFO [main] org.apache.hadoop.mapreduce.v2.app.MRAppMaster: Kind: YARN_AM_RM_TOKEN, Service: , Ident: (appAttemptId {applicat

2020-09-17 14:22:05,121 INFO [main] org.apache.hadoop.mapreduce.v2.app.MRAppMaster: Using mapped newApiCommitter.

2020-09-17 14:22:05,125 INFO [main] org.apache.hadoop.mapreduce.v2.app.MRAppMaster: OutputCommiter set in config null

2020-09-17 14:22:05,172 INFO [main] org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter: File Output Committer Algorithm version is 1

2020-09-17 14:22:05,172 INFO [main] org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders ur

2020-09-17 14:22:05,802 INFO [main] org.apache.hadoop.mapreduce.v2.app.MRAppMaster: OutputCommiter is org.apache.hadoop.mapreduce.lib.output.FileOut

2020-09-17 14:22:05,948 INFO [main] org.apache.hadoop.yarn.event.AsyncDispatcher: Registering class org.apache.hadoop.mapreduce.jobhistory.EventType f

2020-09-17 14:22:05,949 INFO [main] org.apache.hadoop.yarn.event.AsyncDispatcher: Registering class org.apache.hadoop.mapreduce.v2.app.job.event.JobE

2020-09-17 14:22:05,950 INFO [main] org.apache.hadoop.yarn.event.AsyncDispatcher: Registering class org.apache.hadoop.mapreduce.v2.app.job.event.Task

2020-09-17 14:22:05,951 INFO [main] org.apache.hadoop.yarn.event.AsyncDispatcher: Registering class org.apache.hadoop.mapreduce.v2.app.job.event.Task/



Application

Tools

Configuration

Local logs

Server

stacks

Server

metrics

Log Type: syslog

Log Upload Time: Thu Sep 17 14:27:42 +0000 2020

Log Length: 95020

```
2020-09-17 14:26:44,746 INFO [main] org.apache.hadoop.mapreduce.v2.app.MRAppMaster: Created MRAppMaster for application appattempt_1600351918250_0007
2020-09-17 14:26:44,928 INFO [main] org.apache.hadoop.mapreduce.v2.app.MRAppMaster:
/*****
[system properties]
os.name: Linux
os.version: 5.3.0-1035-aws
java.home: /usr/lib/jvm/java-8-openjdk-amd64/jre
java.runtime.version: 1.8.0_265-8u265-b01-0ubuntu2~18.04-b01
java.vendor: Private Build
java.version: 1.8.0_265
java.vm.name: OpenJDK 64-Bit Server VM
java.class.path: /home/ubuntu/home/hadoop_data/hdfs/tmp/nm-local-dir/usercache/ubuntu/appcache/application_1600351918250_0007/container_1600351918250_0007
java.io.tmpdir: /home/ubuntu/home/hadoop_data/hdfs/tmp/nm-local-dir/usercache/ubuntu/appcache/application_1600351918250_0007/container_1600351918250_0007
user.dir: /home/ubuntu/home/hadoop_data/hdfs/tmp/nm-local-dir/usercache/ubuntu/appcache/application_1600351918250_0007/container_1600351918250_0007_01
user.name: ubuntu
*****/
2020-09-17 14:26:45,003 INFO [main] org.apache.hadoop.security.SecurityUtil: Updating Configuration
2020-09-17 14:26:45,015 INFO [main] org.apache.hadoop.mapreduce.v2.app.MRAppMaster: Executing with tokens:
2020-09-17 14:26:45,015 INFO [main] org.apache.hadoop.mapreduce.v2.app.MRAppMaster: Kind: YARN_AM_RM_TOKEN, Service: , Ident: (appAttemptId { applica
2020-09-17 14:26:45,124 INFO [main] org.apache.hadoop.mapreduce.v2.app.MRAppMaster: Using mapped newApiCommitter.
2020-09-17 14:26:45,126 INFO [main] org.apache.hadoop.mapreduce.v2.app.MRAppMaster: OutputCommitter set in config null
2020-09-17 14:26:45,164 INFO [main] org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter: File Output Committer Algorithm version is 1
2020-09-17 14:26:45,164 INFO [main] org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders ur
2020-09-17 14:26:45,678 INFO [main] org.apache.hadoop.mapreduce.v2.app.MRAppMaster: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutp
2020-09-17 14:26:45,819 INFO [main] org.apache.hadoop.mapreduce.v2.app.MRAppMaster: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutp
2020-09-17 14:26:45,819 INFO [main] org.apache.hadoop.yarn.event.AsyncDispatcher: Registering class org.apache.hadoop.mapreduce.jobhistory.EventType 1
2020-09-17 14:26:45,819 INFO [main] org.apache.hadoop.yarn.event.AsyncDispatcher: Registering class org.apache.hadoop.mapreduce.v2.app.job.event.JobFi
```



Application

Tools

Configuration

Local logs

Server

stacks

Server

metrics

Log Type: syslog

Log Upload Time: Thu Sep 17 14:33:08 +0000 2020

Log Length: 62278

```
2020-09-17 14:32:19,306 INFO [main] org.apache.hadoop.mapreduce.v2.app.MRAppMaster: Created MRAppMaster for application appattempt_1600351918250_0010
2020-09-17 14:32:19,459 INFO [main] org.apache.hadoop.mapreduce.v2.app.MRAppMaster:
/*****
[system properties]
os.name: Linux
os.version: 5.3.0-1035-aws
java.home: /usr/lib/jvm/java-8-openjdk-amd64/jre
java.runtime.version: 1.8.0_265-8u265-b01-0ubuntu2~18.04-b01
java.vendor: Private Build
java.version: 1.8.0_265
java.vm.name: OpenJDK 64-Bit Server VM
java.class.path: /home/ubuntu/home/hadoop_data/hdfs/tmp/nm-local-dir/usercache/ubuntu/appcache/application_1600351918250_0010/container_1600351918250_0010
java.io.tmpdir: /home/ubuntu/home/hadoop_data/hdfs/tmp/nm-local-dir/usercache/ubuntu/appcache/application_1600351918250_0010/container_1600351918250_0010_01
user.dir: /home/ubuntu/home/hadoop_data/hdfs/tmp/nm-local-dir/usercache/ubuntu/appcache/application_1600351918250_0010/container_1600351918250_0010_01
user.name: ubuntu
*****/
2020-09-17 14:32:19,526 INFO [main] org.apache.hadoop.security.SecurityUtil: Updating Configuration
2020-09-17 14:32:19,536 INFO [main] org.apache.hadoop.mapreduce.v2.app.MRAppMaster: Executing with tokens:
2020-09-17 14:32:19,536 INFO [main] org.apache.hadoop.mapreduce.v2.app.MRAppMaster: Kind: YARN_AM_RM_TOKEN, Service: , Ident: (appAttemptId { applica
2020-09-17 14:32:19,644 INFO [main] org.apache.hadoop mapreduce.v2.app.MRAppMaster: Using mapped newApiCommitter.
2020-09-17 14:32:19,645 INFO [main] org.apache.hadoop mapreduce.v2.app.MRAppMaster: OutputCommitter set in config null
2020-09-17 14:32:19,689 INFO [main] org.apache.hadoop mapreduce.lib.output.FileOutputCommitter: File Output Committer Algorithm version is 1
2020-09-17 14:32:19,689 INFO [main] org.apache.hadoop mapreduce.lib.output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders ur
2020-09-17 14:32:20,338 INFO [main] org.apache.hadoop mapreduce.v2.app.MRAppMaster: OutputCommitter is org.apache.hadoop mapreduce.lib.output.FileOutp
2020-09-17 14:32:20,482 INFO [main] org.apache.hadoop mapreduce.v2.app.MRAppMaster: OutputCommitter is org.apache.hadoop mapreduce.lib.output.FileOutp
2020-09-17 14:32:20,482 INFO [main] org.apache.hadoop.yarn.event.AsyncDispatcher: Registering class org.apache.hadoop mapreduce.jobhistory.EventType 1
2020-09-17 14:32:20,482 INFO [main] org.apache.hadoop.yarn.event.AsyncDispatcher: Registering class org.apache.hadoop mapreduce.v2.app.job.event.JobFi
```

b. Multi-node Hadoop Cluster Setup

i. "jps" command on 4 VMs


```
20/09/17 14:15:53 INFO blockmanagement.BlockManager: replicationRecheckInterval = 3000
20/09/17 14:15:53 INFO blockmanagement.BlockManager: encryptDataTransfer = false
20/09/17 14:15:53 INFO blockmanagement.BlockManager: maxNumBlocksToLog = 1000
20/09/17 14:15:53 INFO namenode.FSNamesystem: Append Enabled: true
20/09/17 14:15:53 INFO namenode.FSDirectory: GLOBAI serial map: bits=24 maxEntries=16777215
20/09/17 14:15:53 INFO util.GSet: Computing capacity for map INodeMap
20/09/17 14:15:53 INFO util.GSet: VM type = 64-bit
20/09/17 14:15:53 INFO util.GSet: 1.0% max memory 889 MB = 8.9 MB
20/09/17 14:15:53 INFO util.GSet: capacity = 2^20 = 1048576 entries
20/09/17 14:15:53 INFO namenode.FSDirectory: ACLs enabled? false
20/09/17 14:15:53 INFO namenode.FSDirectory: XAttrs enabled? true
20/09/17 14:15:53 INFO namenode.NameNode: Caching file names occurring more than 10 times
20/09/17 14:15:53 INFO snapshot.SnapshotManager: Loaded config captureOpenFiles: false skipCaptureAccessTimeOnlyChange: false
20/09/17 14:15:53 INFO util.GSet: Computing capacity for map cachedBlocks
20/09/17 14:15:53 INFO util.GSet: VM type = 64-bit
20/09/17 14:15:53 INFO util.GSet: 0.25% max memory 889 MB = 2.2 MB
20/09/17 14:15:53 INFO util.GSet: capacity = 2^18 = 262144 entries
20/09/17 14:15:53 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.num.buckets = 10
20/09/17 14:15:53 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10
20/09/17 14:15:53 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
20/09/17 14:15:53 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
20/09/17 14:15:53 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cache entry expiry time is 600000 millis
20/09/17 14:15:53 INFO util.GSet: Computing capacity for map NameNodeRetryCache
20/09/17 14:15:53 INFO util.GSet: VM type = 64-bit
20/09/17 14:15:53 INFO util.GSet: 0.029999999329447746% max memory 889 MB = 273.1 KB
20/09/17 14:15:53 INFO util.GSet: capacity = 2^15 = 32768 entries
20/09/17 14:15:53 INFO namenode.FSImage: Allocated new BlockPoolId: BP-617977013-127.0.1.1-1600352153175
20/09/17 14:15:53 INFO common.Storage: Storage directory /home/ubuntu/hadoop_data/hdfs/namenode has been successfully formatted.
20/09/17 14:15:53 INFO namenode.FSImageFormatProtobuf: Saving image file /home/ubuntu/hadoop_data/hdfs/namenode/current/fsimage.ckpt_000000000000000000 using no compression
20/09/17 14:15:53 INFO namenode.FSImageFormatProtobuf: Image file /home/ubuntu/hadoop_data/hdfs/namenode/current/fsimage.ckpt_000000000000000000 of size 324 bytes saved in 0 seconds
20/09/17 14:15:53 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
20/09/17 14:15:53 INFO namenode.NameNode: SHUTDOWN MSG:
/*****
SHUTDOWN MSG: Shutting down NameNode at ec2-54-197-143-91.compute-1.amazonaws.com/127.0.1.1
*****/
ubuntu@ec2-54-197-143-91:~$ jps
4098 NodeManager
4198 jps
ubuntu@ec2-54-197-143-91:~$ jps
4706 DataNode
4706 jps
4098 NodeManager
ubuntu@ec2-54-197-143-91:~$
```

```
20/09/17 14:11:26 INFO blockmanagement.BlockManager: maxReplication = 512
20/09/17 14:11:26 INFO blockmanagement.BlockManager: minReplication = 1
20/09/17 14:11:26 INFO blockmanagement.BlockManager: maxReplicationStreams = 2
20/09/17 14:11:26 INFO blockmanagement.BlockManager: replicationRecheckInterval = 3000
20/09/17 14:11:26 INFO blockmanagement.BlockManager: encryptDataTransfer = false
20/09/17 14:11:26 INFO blockmanagement.BlockManager: maxNumBlocksToLog = 1000
20/09/17 14:11:26 INFO namenode.FSNamesystem: Append Enabled: true
20/09/17 14:11:26 INFO namenode.FSDirectory: GLOBAI serial map: bits=24 maxEntries=16777215
20/09/17 14:11:26 INFO util.GSet: Computing capacity for map INodeMap
20/09/17 14:11:26 INFO util.GSet: VM type = 64-bit
20/09/17 14:11:26 INFO util.GSet: 1.0% max memory 889 MB = 8.9 MB
20/09/17 14:11:26 INFO util.GSet: capacity = 2^20 = 1048576 entries
20/09/17 14:11:26 INFO namenode.FSDirectory: ACLs enabled? false
20/09/17 14:11:26 INFO namenode.FSDirectory: XAttrs enabled? true
20/09/17 14:11:26 INFO namenode.NameNode: Caching file names occurring more than 10 times
20/09/17 14:11:26 INFO snapshot.SnapshotManager: Loaded config captureOpenFiles: false skipCaptureAccessTimeOnlyChange: false
20/09/17 14:11:26 INFO util.GSet: Computing capacity for map cachedBlocks
20/09/17 14:11:26 INFO util.GSet: VM type = 64-bit
20/09/17 14:11:26 INFO util.GSet: 0.25% max memory 889 MB = 2.2 MB
20/09/17 14:11:26 INFO util.GSet: capacity = 2^18 = 262144 entries
20/09/17 14:11:26 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.num.buckets = 10
20/09/17 14:11:26 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10
20/09/17 14:11:26 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
20/09/17 14:11:26 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
20/09/17 14:11:26 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cache entry expiry time is 600000 millis
20/09/17 14:11:26 INFO util.GSet: Computing capacity for map NameNodeRetryCache
20/09/17 14:11:26 INFO util.GSet: VM type = 64-bit
20/09/17 14:11:26 INFO util.GSet: 0.029999999329447746% max memory 889 MB = 273.1 KB
20/09/17 14:11:26 INFO util.GSet: capacity = 2^15 = 32768 entries
20/09/17 14:11:26 INFO namenode.FSImage: Allocated new BlockPoolId: BP-771571415-127.0.1.1-160035188646
20/09/17 14:11:26 INFO common.Storage: Storage directory /home/ubuntu/hadoop_data/hdfs/namenode has been successfully formatted.
20/09/17 14:11:26 INFO namenode.FSImageFormatProtobuf: Saving image file /home/ubuntu/hadoop_data/hdfs/namenode/current/fsimage.ckpt_000000000000000000 using no compression
20/09/17 14:11:26 INFO namenode.FSImageFormatProtobuf: Image file /home/ubuntu/hadoop_data/hdfs/namenode/current/fsimage.ckpt_000000000000000000 of size 325 bytes saved in 0 seconds
20/09/17 14:11:26 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
20/09/17 14:11:26 INFO namenode.NameNode: SHUTDOWN MSG:
/*****
SHUTDOWN MSG: Shutting down NameNode at ec2-107-22-52-231.compute-1.amazonaws.com/127.0.1.1
*****/
ubuntu@ec2-107-22-52-231:~$ jps
3840 NodeManager
3633 DataNode
3677 jps
ubuntu@ec2-107-22-52-231:~$
```

```
20/09/17 14:11:14 INFO blockmanagement.BlockManager: maxReplication = 512
20/09/17 14:11:14 INFO blockmanagement.BlockManager: minReplication = 1
20/09/17 14:11:14 INFO blockmanagement.BlockManager: maxReplicationStreams = 2
20/09/17 14:11:14 INFO blockmanagement.BlockManager: replicationRecheckInterval = 3000
20/09/17 14:11:14 INFO blockmanagement.BlockManager: encryptDataTransfer = false
20/09/17 14:11:14 INFO blockmanagement.BlockManager: maxNumBlocksToLog = 1000
20/09/17 14:11:14 INFO namenode.FSNamesystem: Append Enabled: true
20/09/17 14:11:14 INFO namenode.FSDirectory: GLOBAI serial map: bits=24 maxEntries=16777215
20/09/17 14:11:14 INFO util.GSet: Computing capacity for map INodeMap
20/09/17 14:11:14 INFO util.GSet: VM type = 64-bit
20/09/17 14:11:14 INFO util.GSet: 1.0% max memory 889 MB = 8.9 MB
20/09/17 14:11:14 INFO util.GSet: capacity = 2^20 = 1048576 entries
20/09/17 14:11:14 INFO namenode.FSDirectory: ACLs enabled? false
20/09/17 14:11:14 INFO namenode.FSDirectory: XAttrs enabled? true
20/09/17 14:11:14 INFO namenode.NameNode: Caching file names occurring more than 10 times
20/09/17 14:11:14 INFO snapshot.SnapshotManager: Loaded config captureOpenFiles: false skipCaptureAccessTimeOnlyChange: false
20/09/17 14:11:14 INFO util.GSet: Computing capacity for map cachedBlocks
20/09/17 14:11:14 INFO util.GSet: VM type = 64-bit
20/09/17 14:11:14 INFO util.GSet: 0.25% max memory 889 MB = 2.2 MB
20/09/17 14:11:14 INFO util.GSet: capacity = 2^18 = 262144 entries
20/09/17 14:11:14 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.num.buckets = 10
20/09/17 14:11:14 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10
20/09/17 14:11:14 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
20/09/17 14:11:14 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
20/09/17 14:11:14 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cache entry expiry time is 600000 millis
20/09/17 14:11:14 INFO util.GSet: Computing capacity for map NameNodeRetryCache
20/09/17 14:11:14 INFO util.GSet: VM type = 64-bit
20/09/17 14:11:14 INFO util.GSet: 0.029999999329447746% max memory 889 MB = 273.1 KB
20/09/17 14:11:14 INFO util.GSet: capacity = 2^15 = 32768 entries
20/09/17 14:11:14 INFO namenode.FSImage: Allocated new BlockPoolId: BP-17162080-127.0.1.1-160035187462
20/09/17 14:11:14 INFO common.Storage: Storage directory /home/ubuntu/hadoop_data/hdfs/namenode has been successfully formatted.
20/09/17 14:11:14 INFO namenode.FSImageFormatProtobuf: Saving image file /home/ubuntu/hadoop_data/hdfs/namenode/current/fsimage.ckpt_000000000000000000 using no compression
20/09/17 14:11:14 INFO namenode.FSImageFormatProtobuf: Image file /home/ubuntu/hadoop_data/hdfs/namenode/current/fsimage.ckpt_000000000000000000 of size 325 bytes saved in 0 seconds
20/09/17 14:11:14 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
20/09/17 14:11:14 INFO namenode.NameNode: SHUTDOWN MSG:
/*****
SHUTDOWN MSG: Shutting down NameNode at ec2-3-223-185-27.compute-1.amazonaws.com/127.0.1.1
*****/
ubuntu@ec2-3-223-185-27:~$ jps
4196 jps
4055 NodeManager
3668 DataNode
ubuntu@ec2-3-223-185-27:~$
```

ii. Run Terasort example with 2GB data

check the time consumed(about 27 minutes)



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
3	0	0	3	0	0 B	60 GB	0 B

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes
3	0	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation
Capacity Scheduler	[MEMORY]	<memory:1536, vCores:1>	<memory:20480, vCores:4>

Show 20 ▾ entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCoers	Allocated Memory MB	Reserved CPU VCoers
application_1600504455183_0003	ubuntu	TeraValidate	MAPREDUCE	default	0	Sat Sep 19 17:10:49 +0800 2020	Sat Sep 19 17:14:29 +0800 2020	FINISHED	SUCCEEDED	N/A	N/A	N/A	N/A
application_1600504455183_0002	ubuntu	TeraSort	MAPREDUCE	default	0	Sat Sep 19 17:09:11 +0800 2020	Sat Sep 19 16:48:51 +0800 2020	FINISHED	SUCCEEDED	N/A	N/A	N/A	N/A
application_1600504455183_0001	ubuntu	TeraGen	MAPREDUCE	default	0	Sat Sep 19 16:44:10 +0800 2020	Sat Sep 19 16:47:48 +0800 2020	FINISHED	SUCCEEDED	N/A	N/A	N/A	N/A

Showing 1 to 3 of 3 entries

iii. Run Terasort example with 20 GB data

1. Run commands in Putty(the same as above)


```
MTPUTTY (Multi-Tabbed PuTTY)
Server View Tools Help
ssh -o StrictHostKeyC Send Send to all
Servers
Start page x ubuntu@ec2-3-128-222-90: /usr/local/hadoop/share/hadoop/mapreduce x ubuntu@ip-172-31-36-91: x ubuntu@ip-172-31-43-64: x ubuntu@ip-172-31-39-134: x
Search
v PuTTY sessions
  DetailNode01 (ubuntu@ec2-3-128-222-90)
  DetailNode02 (ubuntu@ec2-3-128-222-90)
  DetailNode03 (ubuntu@ec2-3-128-222-90)
  NameNode (ubuntu@ec2-3-128-222-90)
2750 JobHistoryServer
ubuntu@ec2-3-128-222-90:~$ cd /usr/local/hadoop/share/hadoop/mapreduce
ubuntu@ec2-3-128-222-90:~$ hadoop jar hadoop-mapreduce-examples-2.9.2.jar teragen 214748365 terasort/input-20G
20/09/19 08:44:08 INFO client.RMProxy: Connecting to ResourceManager at hadoop-master/172.31.36.218:8032
20/09/19 08:44:09 INFO terasort.TeraGen: Generating 214748365 using 2
20/09/19 08:44:09 INFO mapreduce.JobSubmitter: number of splits=2
20/09/19 08:44:10 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
20/09/19 08:44:10 WARN hdfs.DataStreamer: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1252)
    at java.lang.Thread.join(Thread.java:1326)
    at org.apache.hadoop.hdfs.DataStreamer.closeResponder(DataStreamer.java:980)
    at org.apache.hadoop.hdfs.DataStreamer.endBlock(DataStreamer.java:630)
    at org.apache.hadoop.hdfs.DataStreamer.run(DataStreamer.java:807)
20/09/19 08:44:10 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1600504455183_0001
20/09/19 08:44:10 INFO impl.YarnClientImpl: Submitted application application_1600504455183_0001
20/09/19 08:44:10 INFO mapreduce.Job: The url to track the job: http://hadoop-master:8088/proxy/application_1600504455183_0001/
20/09/19 08:44:10 INFO mapreduce.Job: Running job: job_1600504455183_0001
20/09/19 08:44:18 INFO mapreduce.Job: Job job_1600504455183_0001 running in uber mode : false
20/09/19 08:44:18 INFO mapreduce.Job: map 0% reduce 0%
20/09/19 08:44:35 INFO mapreduce.Job: map 7% reduce 0%
20/09/19 08:44:41 INFO mapreduce.Job: map 10% reduce 0%
20/09/19 08:44:47 INFO mapreduce.Job: map 14% reduce 0%
20/09/19 08:44:53 INFO mapreduce.Job: map 17% reduce 0%
20/09/19 08:44:59 INFO mapreduce.Job: map 20% reduce 0%
20/09/19 08:45:05 INFO mapreduce.Job: map 24% reduce 0%
20/09/19 08:45:11 INFO mapreduce.Job: map 27% reduce 0%
20/09/19 08:45:17 INFO mapreduce.Job: map 30% reduce 0%
20/09/19 08:45:23 INFO mapreduce.Job: map 34% reduce 0%
20/09/19 08:45:29 INFO mapreduce.Job: map 37% reduce 0%
20/09/19 08:45:36 INFO mapreduce.Job: map 40% reduce 0%
20/09/19 08:45:42 INFO mapreduce.Job: map 43% reduce 0%
20/09/19 08:45:47 INFO mapreduce.Job: map 47% reduce 0%
20/09/19 08:45:53 INFO mapreduce.Job: map 50% reduce 0%
20/09/19 08:45:59 INFO mapreduce.Job: map 53% reduce 0%
20/09/19 08:46:05 INFO mapreduce.Job: map 56% reduce 0%
20/09/19 08:46:11 INFO mapreduce.Job: map 59% reduce 0%
20/09/19 08:46:17 INFO mapreduce.Job: map 62% reduce 0%
20/09/19 08:46:23 INFO mapreduce.Job: map 65% reduce 0%
20/09/19 08:46:29 INFO mapreduce.Job: map 68% reduce 0%
20/09/19 08:46:36 INFO mapreduce.Job: map 71% reduce 0%
```



```
ubuntu@ec2-3-128-222-90: /usr/local/hadoop/share/hadoop/mapreduce$ hadoop jar hadoop-mapreduce-examples-2.9.2.jar terasort terasort/input-20G terasort/output-20G
20/09/19 08:48:48 INFO terasort.TeraSort: starting
20/09/19 08:48:50 INFO input.FileInputFormat: Total input files to process : 2
Spent 157ms computing base-splits.
Spent 6ms computing TeraScheduler splits.
Computing input splits took 163ms
Sampling 10 splits of 160
Making 1 from 100000 sampled records
Computing partitions took 53ms
Spent 701ms computing partitions.
20/09/19 08:48:51 INFO client.RMProxy: Connecting to ResourceManager at hadoop-master/172.31.36.218:8032
20/09/19 08:48:51 INFO mapreduce.JobSubmitter: number of splits:160
20/09/19 08:48:51 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
20/09/19 08:48:51 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1600504455183_0002
20/09/19 08:48:51 INFO impl.YarnClientImpl: Submitted application application_1600504455183_0002
20/09/19 08:48:51 INFO mapreduce.Job: The url to track the job: http://hadoop-master:8088/proxy/application_1600504455183_0002/
20/09/19 08:48:57 INFO mapreduce.Job: Job job_1600504455183_0002 running in uber mode : false
20/09/19 08:48:57 INFO mapreduce.Job: map 0% reduce 0%
20/09/19 08:49:34 INFO mapreduce.Job: map 1% reduce 0%
20/09/19 08:49:35 INFO mapreduce.Job: map 2% reduce 0%
20/09/19 08:49:36 INFO mapreduce.Job: map 3% reduce 0%
20/09/19 08:49:37 INFO mapreduce.Job: map 4% reduce 0%
20/09/19 08:49:38 INFO mapreduce.Job: map 5% reduce 0%
20/09/19 08:49:42 INFO mapreduce.Job: map 10% reduce 0%
20/09/19 08:49:47 INFO mapreduce.Job: map 11% reduce 0%
21474836420/09/19 08:49:54 INFO mapreduce.Job: map 12% reduce 0%
20/09/19 08:49:55 INFO mapreduce.Job: map 13% reduce 0%
20/09/19 08:49:56 INFO mapreduce.Job: map 14% reduce 0%
20/09/19 08:50:01 INFO mapreduce.Job: map 15% reduce 0%
20/09/19 08:50:05 INFO mapreduce.Job: map 16% reduce 0%
20/09/19 08:50:11 INFO mapreduce.Job: map 17% reduce 0%
20/09/19 08:50:13 INFO mapreduce.Job: map 19% reduce 0%
20/09/19 08:50:14 INFO mapreduce.Job: map 20% reduce 0%
20/09/19 08:50:15 INFO mapreduce.Job: map 21% reduce 0%
20/09/19 08:50:19 INFO mapreduce.Job: map 22% reduce 0%
20/09/19 08:50:20 INFO mapreduce.Job: map 23% reduce 0%
20/09/19 08:50:37 INFO mapreduce.Job: map 24% reduce 0%
20/09/19 08:50:50 INFO mapreduce.Job: map 25% reduce 0%
20/09/19 08:50:51 INFO mapreduce.Job: map 27% reduce 1%
20/09/19 08:50:55 INFO mapreduce.Job: map 28% reduce 1%
20/09/19 08:50:56 INFO mapreduce.Job: map 29% reduce 1%
```

```
File Output Format: Counters
Bytes Written:21474836500
20/09/19 09:09:13 INFO terasort.TeraSort: done
ubuntu@ec2-3-128-222-90: /usr/local/hadoop/share/hadoop/mapreduce$ hadoop jar hadoop-mapreduce-examples-2.9.2.jar teravalidate terasort/output-20G terasort/check-20
20/09/19 09:10:48 INFO client.RMProxy: Connecting to ResourceManager at hadoop-master/172.31.36.218:8032
20/09/19 09:10:48 INFO input.FileInputFormat: Total input files to process : 1
Spent 29ms computing base-splits.
Spent 3ms computing TeraScheduler splits.
20/09/19 09:10:49 INFO mapreduce.JobSubmitter: number of splits:1
20/09/19 09:10:49 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
20/09/19 09:10:49 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1600504455183_0003
20/09/19 09:10:49 INFO impl.YarnClientImpl: Submitted application application_1600504455183_0003
20/09/19 09:10:49 INFO mapreduce.Job: The url to track the job: http://hadoop-master:8088/proxy/application_1600504455183_0003/
20/09/19 09:10:49 INFO mapreduce.Job: Running job: job_1600504455183_0003
20/09/19 09:10:56 INFO mapreduce.Job: Job job_1600504455183_0003 running in uber mode : false
20/09/19 09:10:56 INFO mapreduce.Job: map 0% reduce 0%
20/09/19 09:11:12 INFO mapreduce.Job: map 4% reduce 0%
20/09/19 09:11:18 INFO mapreduce.Job: map 7% reduce 0%
20/09/19 09:11:24 INFO mapreduce.Job: map 9% reduce 0%
20/09/19 09:11:30 INFO mapreduce.Job: map 11% reduce 0%
20/09/19 09:11:36 INFO mapreduce.Job: map 14% reduce 0%
20/09/19 09:11:42 INFO mapreduce.Job: map 16% reduce 0%
20/09/19 09:11:48 INFO mapreduce.Job: map 18% reduce 0%
20/09/19 09:11:54 INFO mapreduce.Job: map 20% reduce 0%
20/09/19 09:12:00 INFO mapreduce.Job: map 22% reduce 0%
20/09/19 09:12:06 INFO mapreduce.Job: map 25% reduce 0%
20/09/19 09:12:13 INFO mapreduce.Job: map 27% reduce 0%
20/09/19 09:12:19 INFO mapreduce.Job: map 29% reduce 0%
20/09/19 09:12:25 INFO mapreduce.Job: map 31% reduce 0%
20/09/19 09:12:31 INFO mapreduce.Job: map 33% reduce 0%
20/09/19 09:12:37 INFO mapreduce.Job: map 35% reduce 0%
20/09/19 09:12:42 INFO mapreduce.Job: map 37% reduce 0%
20/09/19 09:12:48 INFO mapreduce.Job: map 39% reduce 0%
20/09/19 09:12:54 INFO mapreduce.Job: map 41% reduce 0%
20/09/19 09:13:00 INFO mapreduce.Job: map 43% reduce 0%
20/09/19 09:13:06 INFO mapreduce.Job: map 45% reduce 0%
20/09/19 09:13:12 INFO mapreduce.Job: map 47% reduce 0%
20/09/19 09:13:18 INFO mapreduce.Job: map 49% reduce 0%
20/09/19 09:13:24 INFO mapreduce.Job: map 51% reduce 0%
20/09/19 09:13:30 INFO mapreduce.Job: map 53% reduce 0%
20/09/19 09:13:36 INFO mapreduce.Job: map 55% reduce 0%
20/09/19 09:13:42 INFO mapreduce.Job: map 56% reduce 0%
```

2. Check the time consumed(about 10 minutes)



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
4	0	0	4	0	0 B	60 GB	0 B

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes
3	0	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation
Capacity Scheduler	[MEMORY]	<memory:1536, vCores:1>	<memory:20480, vCores:4>

Show 20 entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU V-Cores	Allocated Memory MB	Reserved CPU V-Cores
application_1600433360673_0005	ubuntu	TeraValidate	MAPREDUCE	default	0	Fri Sep 18 21:36:14 +0800 2020	Fri Sep 18 21:39:54 +0800 2020	FINISHED	SUCCEEDED	N/A	N/A	N/A	N/A
application_1600433360673_0004	ubuntu	TeraSort	MAPREDUCE	default	0	Fri Sep 18 21:27:32 +0800 2020	Fri Sep 18 21:33:54 +0800 2020	FINISHED	SUCCEEDED	N/A	N/A	N/A	N/A
application_1600433360673_0003	ubuntu	TeraGen	MAPREDUCE	default	0	Fri Sep 18 21:25:38 +0800 2020	Fri Sep 18 21:26:56 +0800 2020	FINISHED	SUCCEEDED	N/A	N/A	N/A	N/A
application_1600433360673_0002	ubuntu	TeraGen	MAPREDUCE	default	0	Fri Sep 18 21:13:45 +0800 2020	Fri Sep 18 21:15:49 +0800 2020	FINISHED	FAILED	N/A	N/A	N/A	N/A

Showing 1 to 4 of 4 entries

i. Necessary modifications

11655 Grind verdigris many times coloured with lemon juice and keep it
 11656 away from yellow (?).
 11657
 11658 Of preparing the panel.
 11659
 11660 629.
 11661
 11662 TO PREPARE A PANEL FOR PAINTING ON.
 11663
 11664 The panel should be cypress or pear or service-tree or walnut. You
 11665 must coat it over with mastic and turpentine twice distilled and
 11666 white or, if you like, lime, and put it in a frame so that it may
 11667 expand and shrink according to its moisture and dryness. Then give
 11668 it [a coat] of aqua vitae in which you have dissolved arsenic or
 11669 [corrosive] sublimate, 2 or 3 times. Then apply boiled linseed oil
 11670 in such a way as that it may penetrate every part, and before it is
 11671 cold rub it well with a cloth to dry it. Over this apply liquid
 11672 varnish with a stick; then wash it with urine when it is
 11673 dry, and dry it again. Then pounce and outline your drawing finely
 11674 and over it lay a priming of 30 parts of verdigris with one of
 11675 verdigris with two of yellow.
 11676
 11677 [Footnote: M. RAVAISSON'S reading varies from mine in the following
 11678 passages:
 11679
 11680 1. _opero allor [?] bo [alloro?]_ = "_ou bien de [laurier]_"
 11681
 11682 6. _fregalo bene con un panno_. He reads _pane_ for _panno_ and
 11683 renders it. "_Frotte le bien avec un pain de façon [jusqu'à ce]
 11684 qu'il_" etc.
 11685
 11686 7. _colla stecca po laua_. He reads "_polacca_" = "_avec le couteau
 11687 de Bois [?] polonais [?]"_.
 11688
 11689 The preparation of oils (629--634).
 11690
 11691 629.
 11692
 11693 OIL.
 11694
 11695 Make some oil of mustard seed; and if you wish to make it with
 11696 greater ease mix the ground seeds with linseed oil and put it all
 11697 under the press.
 11698

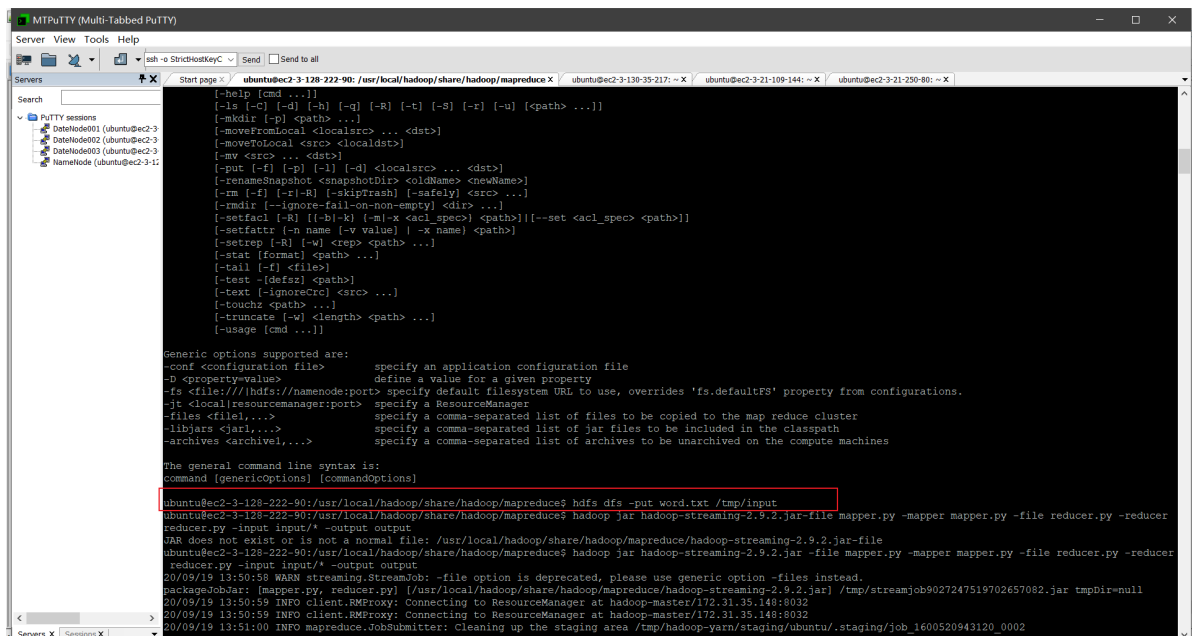
The screenshot shows a MobaXterm terminal window with the following content:

```
https://microk8s.io/ has docs and details.

00 updates can be installed immediately.
0 of these updates are security updates.
To see these additional updates run: apt list --upgradable

Last login: Sat Sep 19 12:42:31 2020 from 172.31.35.148
ubuntu@ec2-3-128-222-90:~$ hadoop version
Hadoop 2.9.2
Subversion https://git-wip-us.apache.org/repos/asf/hadoop.git -r 826afbae31ca6b7bc2f847d1481d4b66ed2c6704
Compiled by ajinks on 2018-11-13T12:42Z
Compiled with protoc 2.5.0
From source with checksum 3a939967262218aa556c684d107985
This command was run using /usr/local/hadoop/share/hadoop/common/hadoop-common-2.9.2.jar
ubuntu@ec2-3-128-222-90:~$ cd /usr/local/hadoop/share/hadoop/mapreduce
ubuntu@ec2-3-128-222-90:~/hadoop/mapreduce$ hadoop jar hadoop-streaming-2.9.2.jar -file mapper.py -mapper mapper.py -file reducer.py -reducer
reducer.py -input input/ -output output
20/09/19 13:45:44 WARN streaming.StreamJob: -file option is deprecated, please use generic option -files instead.
packageJobJar: [mapper.py, reducer.py] [/usr/local/hadoop/share/hadoop/mapreduce/hadoop-streaming-2.9.2.jar] /tmp/streamjob2544997733790292055.jar tmpDir=null
20/09/19 13:45:45 INFO client.RMProxy: Connecting to ResourceManager at hadoop-master/172.31.35.148:8032
20/09/19 13:45:45 INFO client.RMProxy: Connecting to ResourceManager at hadoop-master/172.31.35.148:8032
20/09/19 13:45:45 INFO mapreduce.JobSubmitter: Cleaning up the staging area /tmp/hadoop-yarn/staging/ubuntu/.staging/job_1600520943120_0001
20/09/19 13:45:45 ERROR streaming.StreamJob: Error Launching job : /user (is not a directory)
at org.apache.hadoop.hdfs.server.namenode.FSDirectory.resolvePath(FSDirectory.java:644)
at org.apache.hadoop.hdfs.server.namenode.FSDirectoryListingInfo.getListingInfo(FSDirectoryListingInfo.java:55)
at org.apache.hadoop.hdfs.server.namenode.FSNameSystem.getListListing(FSNameSystem.java:3630)
at org.apache.hadoop.hdfs.server.namenode.FSNameSystem.getServerListing(NameNodeRpcServer.java:1111)
at org.apache.hadoop.hdfs.protocolPB.ClientNameNodeProtocolServerSideTranslatorPB.getListListing(ClientNameNodeProtocolServerSideTranslatorPB.java:664)
at org.apache.hadoop.hdfs.protocol.proto.ClientNameNodeProtocolProtosClientNameNodeProtocol$2.callBlockingMethod(ClientNameNodeProtocolProtos.java)
at org.apache.hadoop.ipc.RPC$Server.call(RPC.java:969)
at org.apache.hadoop.ipc.Server$RpcCall.run(Server.java:871)
at org.apache.hadoop.ipc.Server$RpcCall.run(Server.java:871)
at java.security.AccessController.doPrivileged(Native Method)
at javax.security.auth.Subject.doAs(Subject.java:422)
at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1893)
at org.apache.hadoop.ipc.Server$Handler.run(Server.java:2606)

Streaming Command Failed!
ubuntu@ec2-3-128-222-90:~/hadoop/mapreduce$ hdfs dfs -mkdir input
mkdir: 'input': No such file or directory
ubuntu@ec2-3-128-222-90:~/hadoop/mapreduce$ hdfs dfs -mkdir /tmp/input
```

```
[~help [cmd ...]]
[-ls [-C] [-d] [-h] [-q] [-R] [-t] [-S] [-r] [-u] [-path] ...]]
[-mkdir [-p] [-path] ...]]
[-moveFromLocal <localsrc> ... <dst>]
[-moveToLocal <src> <localdst>]
[-mv <src> ... <dst>]
[-put [-f] [-p] [-l] [-d] <localsrc> ... <dst>]
[-renameSnapshot <snapshotDir> <oldName> <newName>]
[-rm [-f] [-r] [-R] [-skipTrash] [-safely] <src> ...]]
[-rmDir [-ignoreFailToNonEmpty] <dir> ...]]
[-setfacl [-R] [[-b|-k] [-n|-x <acl_spec>] <path>]] [--set <acl_spec> <path>]]
[-setfacl [-n name [-v value] | -x name] <path>]
[-setrep [-R] [-w] <rep> <path> ...]]
[-stat [format] <path> ...]]
[-tail [-f] <file>]
[-test [-defsz] <path>]
[-text [-ignoreCRC] <src> ...]]
[-touchz <path> ...]]
[-truncate [-w] <length> <path> ...]]
[-usage [cmd ...]]

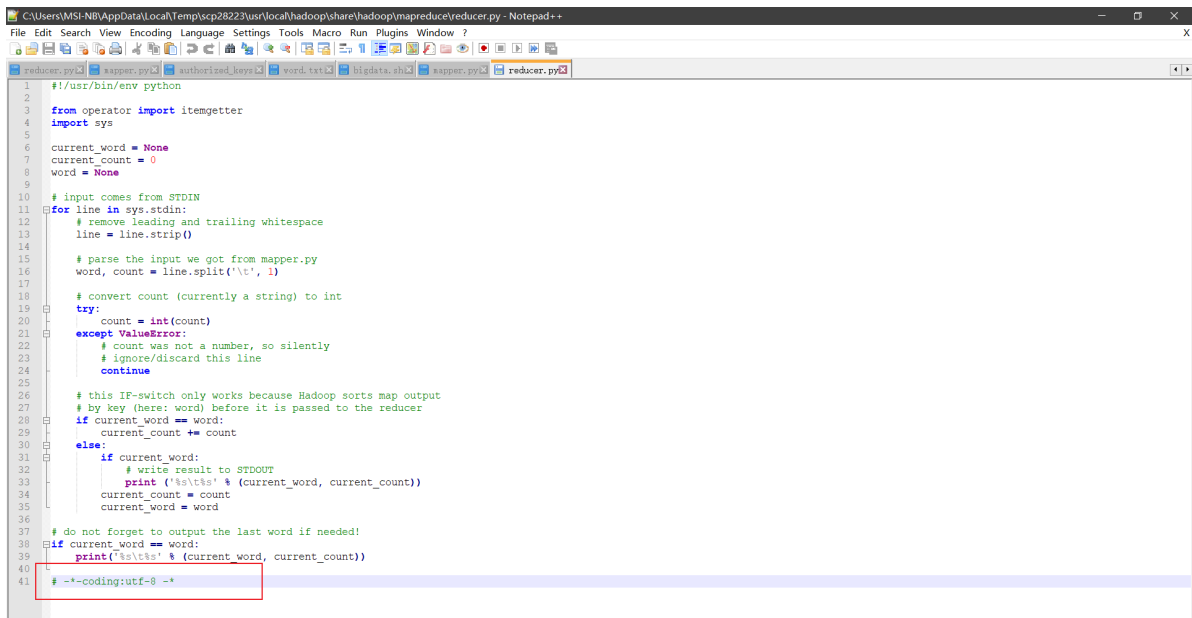
Generic options supported are:
-conf <configuration file>          specify an application configuration file
-D <property=value>                 define a value for a given property
-fs <file:///hdfs://namenode:port> specify default filesystem URL to use, overrides 'fs.defaultFS' property from configurations.
-jt <localresourcemanager:port>     specify a ResourceManager
-files <file1,...>                  specify a comma-separated list of files to be copied to the map reduce cluster
-libjars <jar1,...>                specify a comma-separated list of jar files to be included in the classpath
-archives <archive1,...>           specify a comma-separated list of archives to be unarchived on the compute machines

The general command line syntax is:
command [genericOptions] [commandOptions]
```

3. Modify the .py files' authority and format

- 1 `$ chmod +x mapper.py`
- 2
- 3 `$ chmod +x reducer.py`

add on the second line : `# --coding:utf-8``

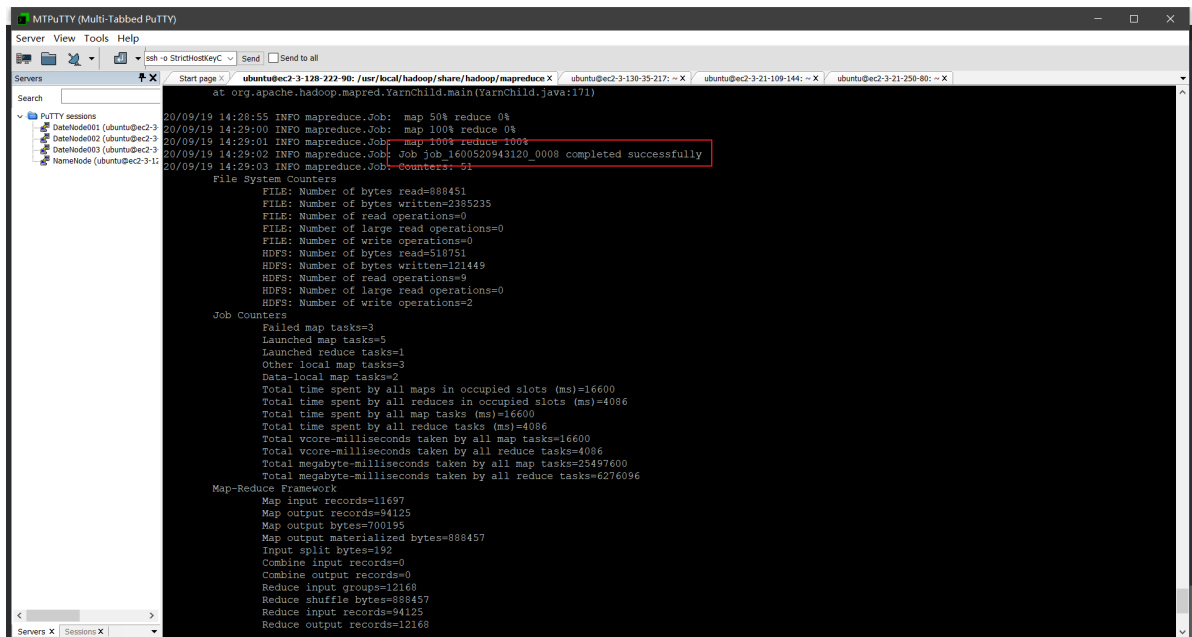


```
1 #!/usr/bin/env python
2
3 from operator import itemgetter
4 import sys
5
6 current_word = None
7 current_count = 0
8 word = None
9
10 # input comes from STDIN
11 for line in sys.stdin:
12     # remove leading and trailing whitespace
13     line = line.strip()
14
15     # parse the input we got from mapper.py
16     word, count = line.split('\t', 1)
17
18     # convert count (currently a string) to int
19     try:
20         count = int(count)
21     except ValueError:
22         # count was not a number, so silently
23         # ignore/discard this line
24         continue
25
26     # this IF-switch only works because Hadoop sorts map output
27     # by key (here: word) before it is passed to the reducer
28     if current_word == word:
29         current_count += count
30     else:
31         if current_word:
32             # write result to STDOUT
33             print('%s\t%s' % (current_word, current_count))
34         current_count = count
35         current_word = word
36
37 # do not forget to output the last word if needed!
38 if current_word == word:
39     print('%s\t%s' % (current_word, current_count))
40
41 # --coding:utf-8`
```

4. Run the following codes

- 1 `$ hadoop jar hadoop-streaming-2.9.2.jar -file`
`/usr/local/hadoop/share/hadoop/mapreduce/mapper.py -mapper`
`/usr/local/hadoop/share/hadoop/mapreduce/mapper.py -file`
`/usr/local/hadoop/share/hadoop/mapreduce/reducer.py -reducer`
`/usr/local/hadoop/share/hadoop/mapreduce/reducer.py -input /tmp/input/* -`
`output output4`

5. Check the result



```
at org.apache.hadoop.mapred.YarnChild.main(YarnChild.java:171)

20/09/19 14:28:55 INFO mapreduce.Job: map 50% reduce 0%
20/09/19 14:29:00 INFO mapreduce.Job: map 100% reduce 0%
20/09/19 14:29:01 INFO mapreduce.Job: map 100% reduce 100%
20/09/19 14:29:02 INFO mapreduce.Job: Job Job_1600520943120_0000 completed successfully
20/09/19 14:29:03 INFO mapreduce.Job: Counters: 51

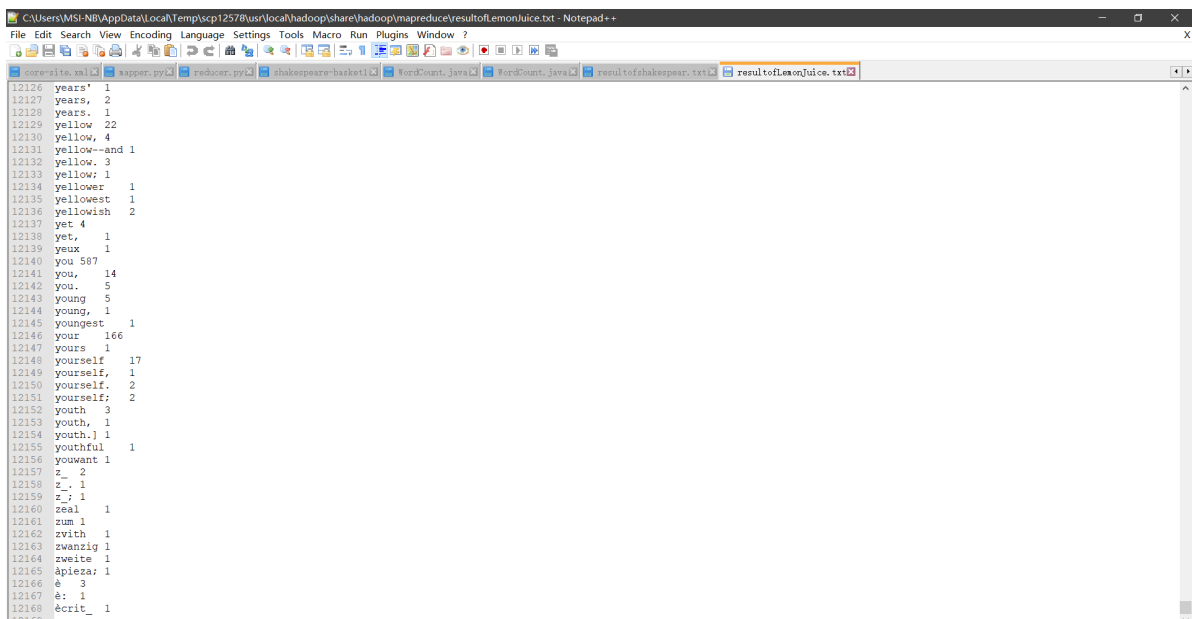
File System Counters
  FILE: Number of bytes read=888451
  FILE: Number of bytes written=2385235
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=518751
  HDFS: Number of bytes written=121449
  HDFS: Number of read operations=9
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2

Job Counters
  Failed map tasks=3
  Launched map tasks=5
  Launched reduce tasks=1
  Other local map tasks=3
  Data-local map tasks=2
  Total time spent by all maps in occupied slots (ms)=16600
  Total time spent by all reduces in occupied slots (ms)=4086
  Total time spent by all map tasks (ms)=16600
  Total time spent by all reduce tasks (ms)=4086
  Total vcore-milliseconds taken by all map tasks=16600
  Total vcore-milliseconds taken by all reduce tasks=4086
  Total megabyte-milliseconds taken by all map tasks=25497600
  Total megabyte-milliseconds taken by all reduce tasks=6276096

Map-Reduce Framework
  Map input records=11697
  Map output records=94125
  Map output bytes=700155
  Map output materialized bytes=888457
  Input split bytes=192
  Combine input records=0
  Combine output records=0
  Reduce input groups=12168
  Reduce shuffle bytes=888457
  Reduce input records=94125
  Reduce output records=12168
```

6. Cat the file to VM and check the result

```
1 $ hadoop fs -cat /user/ubuntu/output4/part-00000 > resultofLemonJuice.txt
```



```
12126 years: 1
12127 years: 2
12128 years: 1
12129 yellow: 22
12130 yellow: 4
12131 yellow-and: 1
12132 yellow: 3
12133 yellow: 1
12134 yellower: 1
12135 yellowest: 1
12136 yellowish: 2
12137 yet: 4
12138 yet: 1
12139 yeux: 1
12140 you: 587
12141 you: 14
12142 you: 5
12143 young: 5
12144 young: 1
12145 youngest: 1
12146 your: 166
12147 yours: 1
12148 yourself: 17
12149 yourself: 1
12150 yourself: 2
12151 yourself: 2
12152 youth: 3
12153 youth: 1
12154 youth: 1
12155 youthful: 1
12156 youwant: 1
12157 z: 2
12158 z: 1
12159 z: 1
12160 zeal: 1
12161 zum: 1
12162 zwitz: 1
12163 zwanzig: 1
12164 zweite: 1
12165 ¢pieza: 1
12166 ¢: 3
12167 ¢: 1
12168 ¢crit: 1
12169
```

d. Compiling the Java WordCount program for MapReduce

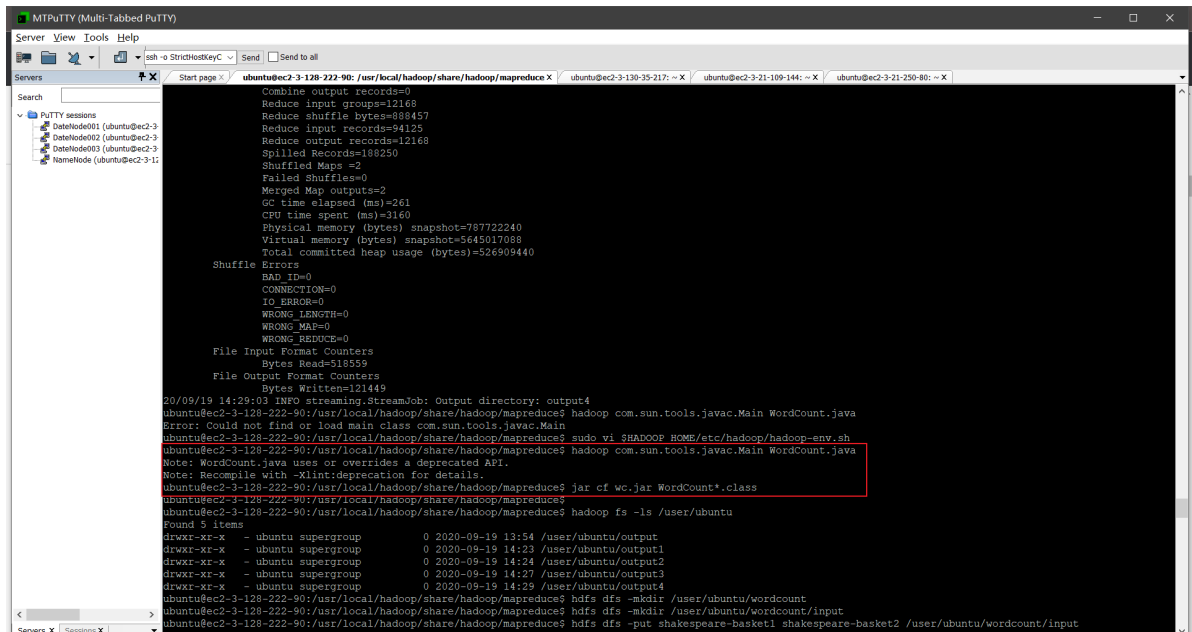
i. Make a preparation for running this job

1. Enter these path variables into \$HADOOP_HOME/etc/hadoop/hadoop-env.sh

- 1 `export PATH=${JAVA_HOME}/bin:${PATH}`
- 2 `export HADOOP_CLASSPATH=${JAVA_HOME}/lib/tools.jar`

2. Compile wordCount.java and create a jar

- 1 `$ hadoop com.sun.tools.javac.Main WordCount.java`
- 2 `$ jar cf wc.jar WordCount*.class`



```
Combine output records=0
Reduce input groups=12168
Reduce shuffle bytes=888457
Reduce input records=94125
Reduce output records=12168
Spilled Records=188250
Shuffled Maps =2
Failed Shuffles=0
Merged Map outputs=2
GC time elapsed (ms)=261
CPU time spent (ms)=3160
Physical memory (bytes) snapshot=787722240
Virtual memory (bytes) snapshot=5645017088
Total committed heap usage (bytes)=526909440

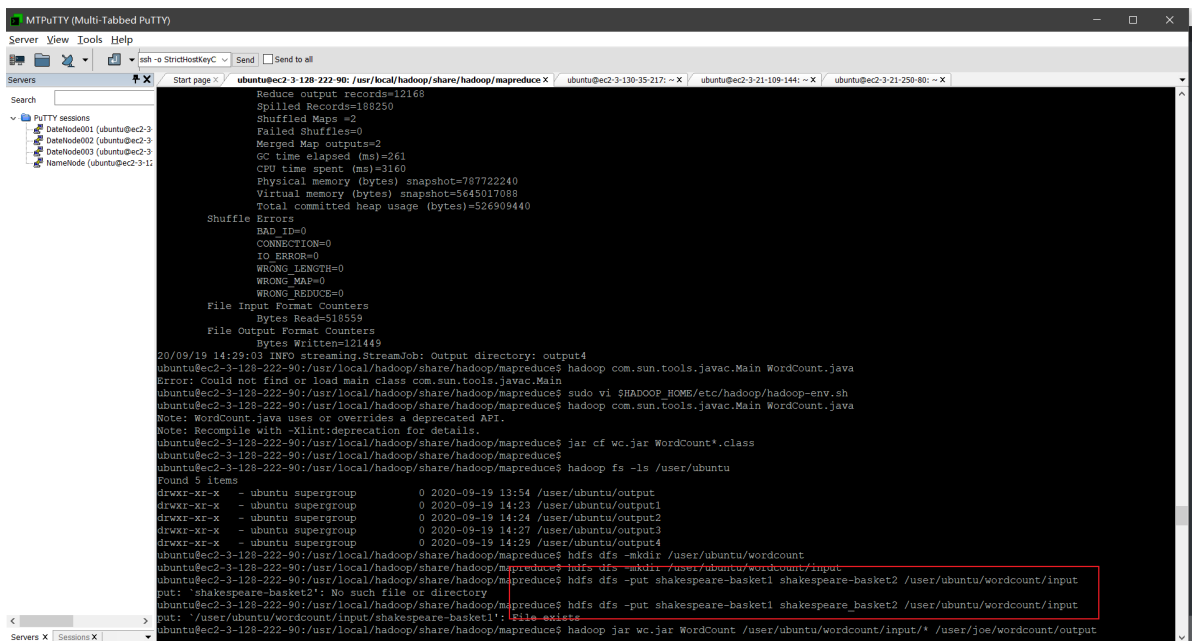
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0

File Input Format Counters
Bytes Read=518559
File Output Format Counters
Bytes Written=121449

20/09/19 14:29:03 INFO streaming.StreamJob: Output directory: output4
ubuntu@ec2-3-128-222-90:/usr/local/hadoop/share/hadoop/mapreduce$ hadoop com.sun.tools.javac.Main WordCount.java
Error: Could not find or load main class com.sun.tools.javac.Main
ubuntu@ec2-3-128-222-90:/usr/local/hadoop/share/hadoop/mapreduce$ sudo vi $HADOOP_HOME/etc/hadoop/hadoop-env.sh
ubuntu@ec2-3-128-222-90:/usr/local/hadoop/share/hadoop/mapreduce$ hadoop com.sun.tools.javac.Main WordCount.java
Note: WordCount.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
ubuntu@ec2-3-128-222-90:/usr/local/hadoop/share/hadoop/mapreduce$ jar cf wc.jar WordCount*.class
ubuntu@ec2-3-128-222-90:/usr/local/hadoop/share/hadoop/mapreduce$ hadoop fs -ls /user/ubunt
Found 5 items
drwxr-xr-x - ubuntu supergroup 0 2020-09-19 13:54 /user/ubunt/output
drwxr-xr-x - ubuntu supergroup 0 2020-09-19 14:23 /user/ubunt/output1
drwxr-xr-x - ubuntu supergroup 0 2020-09-19 14:24 /user/ubunt/output2
drwxr-xr-x - ubuntu supergroup 0 2020-09-19 14:27 /user/ubunt/output3
drwxr-xr-x - ubuntu supergroup 0 2020-09-19 14:29 /user/ubunt/output4
ubuntu@ec2-3-128-222-90:/usr/local/hadoop/share/hadoop/mapreduce$ hdfs dfs -mkdir /user/ubunt/wordcount
ubuntu@ec2-3-128-222-90:/usr/local/hadoop/share/hadoop/mapreduce$ hdfs dfs -mkdir /user/ubunt/wordcount/input
ubuntu@ec2-3-128-222-90:/usr/local/hadoop/share/hadoop/mapreduce$ hdfs dfs -put shakespeare-basket1 shakespeare-basket2 /user/ubunt/wordcount/input
```

3. Create input directory in HDFS and upload the shakespeare Basket1/2 onto it

- 1 `$ hdfs dfs -put shakespeare-basket1 shakespeare-basket2 /user/ubunt/wordcount/input`



```
Reduce output records=12168
Spilled Records=188250
Shuffled Maps =2
Failed Shuffles=0
Merged Map outputs=2
GC time elapsed (ms)=261
CPU time spent (ms)=3160
Physical memory (bytes) snapshot=787722240
Virtual memory (bytes) snapshot=5645017088
Total committed heap usage (bytes)=526909440

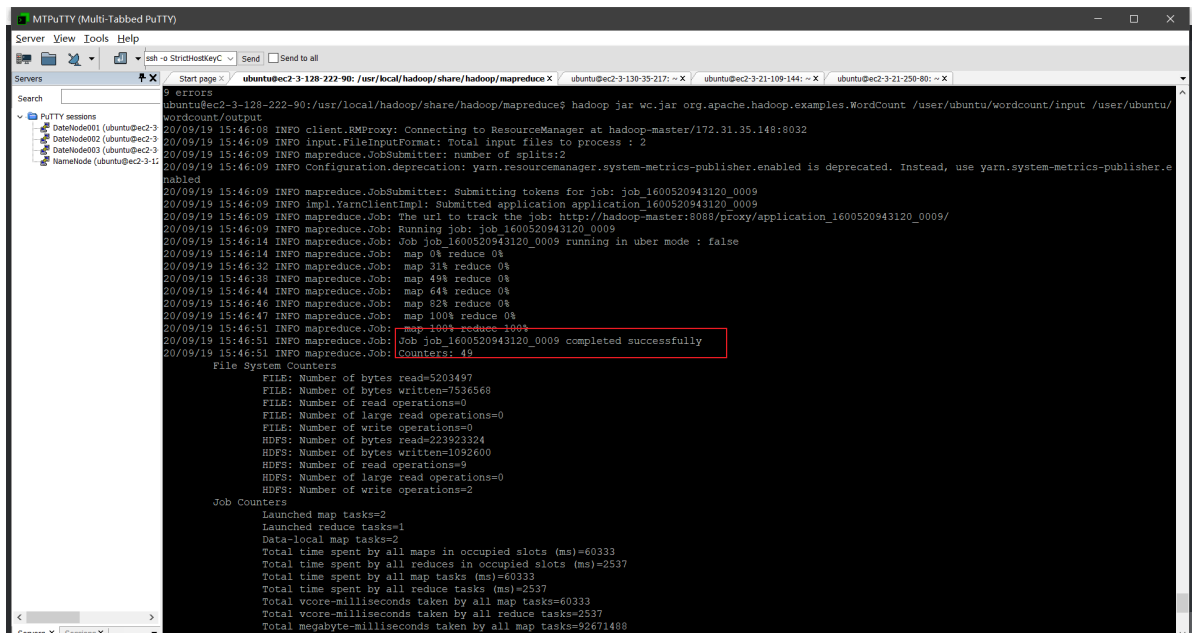
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0

File Input Format Counters
Bytes Read=518559
File Output Format Counters
Bytes Written=121449

20/09/19 14:29:03 INFO streaming.StreamJob: Output directory: output4
ubuntu@ec2-3-128-222-90:/usr/local/hadoop/share/hadoop/mapreduce$ hadoop com.sun.tools.javac.Main WordCount.java
Error: Could not find or load main class com.sun.tools.javac.Main
ubuntu@ec2-3-128-222-90:/usr/local/hadoop/share/hadoop/mapreduce$ sudo vi $HADOOP_HOME/etc/hadoop/hadoop-env.sh
ubuntu@ec2-3-128-222-90:/usr/local/hadoop/share/hadoop/mapreduce$ hadoop com.sun.tools.javac.Main WordCount.java
Note: WordCount.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
ubuntu@ec2-3-128-222-90:/usr/local/hadoop/share/hadoop/mapreduce$ jar cf wc.jar WordCount*.class
ubuntu@ec2-3-128-222-90:/usr/local/hadoop/share/hadoop/mapreduce$ hadoop fs -ls /user/ubunt
Found 5 items
drwxr-xr-x - ubuntu supergroup 0 2020-09-19 13:54 /user/ubunt/output
drwxr-xr-x - ubuntu supergroup 0 2020-09-19 14:23 /user/ubunt/output1
drwxr-xr-x - ubuntu supergroup 0 2020-09-19 14:24 /user/ubunt/output2
drwxr-xr-x - ubuntu supergroup 0 2020-09-19 14:27 /user/ubunt/output3
drwxr-xr-x - ubuntu supergroup 0 2020-09-19 14:29 /user/ubunt/output4
ubuntu@ec2-3-128-222-90:/usr/local/hadoop/share/hadoop/mapreduce$ hdfs dfs -mkdir /user/ubunt/wordcount
ubuntu@ec2-3-128-222-90:/usr/local/hadoop/share/hadoop/mapreduce$ hdfs dfs -mkdir /user/ubunt/wordcount/input
ubuntu@ec2-3-128-222-90:/usr/local/hadoop/share/hadoop/mapreduce$ hdfs dfs -put shakespeare-basket1 shakespeare-basket2 /user/ubunt/wordcount/input
ubuntu@ec2-3-128-222-90:/usr/local/hadoop/share/hadoop/mapreduce$ hdfs dfs -put shakespeare-basket1 shakespeare-basket2 /user/ubunt/wordcount/input
put: '/user/ubunt/wordcount/input/shakespeare-basket1': File exists
ubuntu@ec2-3-128-222-90:/usr/local/hadoop/share/hadoop/mapreduce$ hadoop jar wc.jar WordCount /user/ubunt/wordcount/input/* /user/joe/wordcount/output
```

4. Run the application

```
1 $ hadoop jar wc.jar org.apache.hadoop.examples.WordCount  
/user/ubuntu/wordcount/input /user/ubuntu/wordcount/output
```



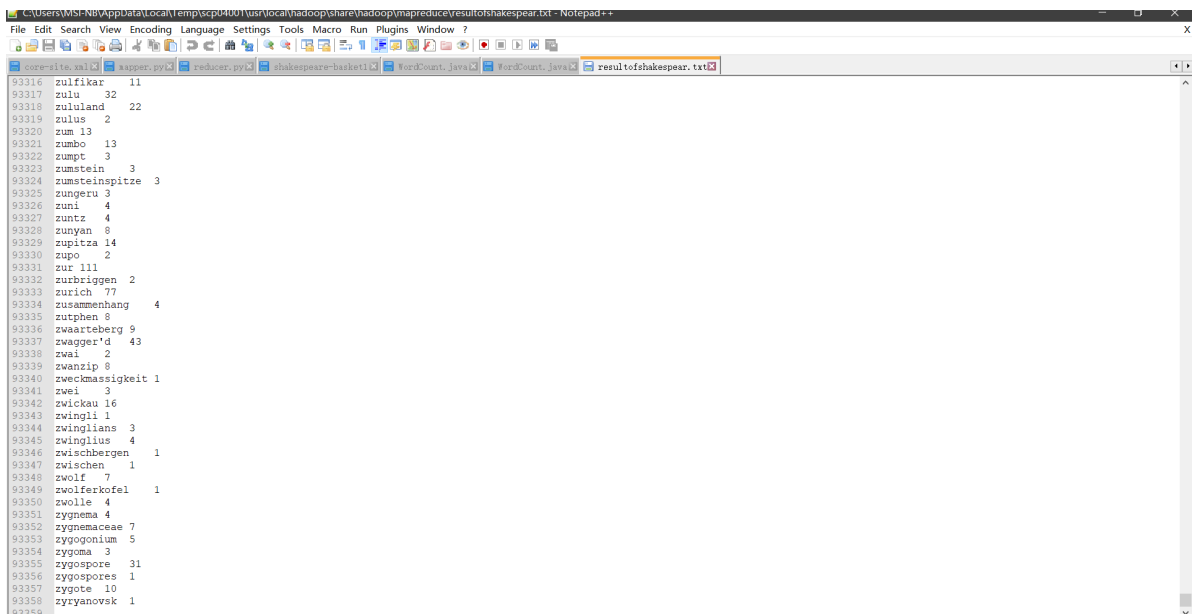
```
ubuntu@ec2-3-128-222-90:/usr/local/hadoop/share/hadoop/mapreduce$ hadoop jar wc.jar org.apache.hadoop.examples.WordCount /user/ubuntu/wordcount/input /user/ubuntu/wordcount/output
20/09/19 15:46:09 INFO client.RMProxy: Connecting to ResourceManager at hadoop-master/172.31.35.148:8032
20/09/19 15:46:09 INFO input.FileInputFormat: Total input files to process : 2
20/09/19 15:46:09 INFO mapreduce.JobSubmitter: number of splits:2
20/09/19 15:46:09 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
20/09/19 15:46:09 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1600520943120_0009
20/09/19 15:46:09 INFO impl.YarnClientImpl: Submitted application application_1600520943120_0009
20/09/19 15:46:09 INFO mapreduce.Job: The url to track the job: http://hadoop-master:8088/proxy/application_1600520943120_0009/
20/09/19 15:46:09 INFO mapreduce.Job: Running job: job_1600520943120_0009
20/09/19 15:46:14 INFO mapreduce.Job: Job Job_1600520943120_0009 running in uber mode : false
20/09/19 15:46:14 INFO mapreduce.Job: map 0% reduce 0%
20/09/19 15:46:32 INFO mapreduce.Job: map 31% reduce 0%
20/09/19 15:46:38 INFO mapreduce.Job: map 49% reduce 0%
20/09/19 15:46:44 INFO mapreduce.Job: map 64% reduce 0%
20/09/19 15:46:46 INFO mapreduce.Job: map 82% reduce 0%
20/09/19 15:46:47 INFO mapreduce.Job: map 100% reduce 0%
20/09/19 15:46:51 INFO mapreduce.Job: map-100%-reduce-100%
20/09/19 15:46:51 INFO mapreduce.Job: Job job_1600520943120_0009 completed successfully
20/09/19 15:46:51 INFO mapreduce.Job: Counters: 49

File System Counters
  FILE: Number of bytes read=5203497
  FILE: Number of bytes written=7536568
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=223923324
  HDFS: Number of bytes written=1092600
  HDFS: Number of read operations=9
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2

Job Counters
  Launched map tasks=2
  Launched reduce tasks=1
  Data-local map tasks=2
  Total time spent by all maps in occupied slots (ms)=60333
  Total time spent by all reduces in occupied slots (ms)=2537
  Total time spent by all map tasks (ms)=60333
  Total time spent by all reduce tasks (ms)=2537
  Total vcore-milliseconds taken by all map tasks=60333
  Total vcore-milliseconds taken by all reduce tasks=2537
  Total megabyte-milliseconds taken by all map tasks=92671488
```

5. Duplicate the output into a .txt file

```
1 $ hadoop fs -cat /user/ubuntu/wordcount/output/part-r-00000 >  
resultofshakespear.txt
```



```
93316 zulfikar 11
93317 zulu 32
93318 zululand 22
93319 zulus 2
93320 zum 13
93321 zumbo 13
93322 zumpt 3
93323 zumstein 3
93324 zumsteinspitze 3
93325 zungeru 3
93326 zuni 4
93327 zuntz 4
93328 zunyan 8
93329 zupitza 14
93330 zupo 2
93331 zur 111
93332 zurbriggen 2
93333 zurich 77
93334 zusammenhang 4
93335 zutphen 8
93336 zwaarteberg 9
93337 zwagger'd 43
93338 zwei 2
93339 zwanzip 8
93340 zweckmassigkeit 1
93341 zwei 3
93342 zwickau 16
93343 zwingli 1
93344 zwinglians 3
93345 zwinglius 4
93346 zwischenbergen 1
93347 zwischen 1
93348 zwolf 7
93349 zwolferkofel 1
93350 zwolle 4
93351 zygnema 4
93352 zygnemaceae 7
93353 zygogonium 5
93354 zygo 3
93355 zygospore 31
93356 zygospores 1
93357 zygote 10
93358 zyryanovsk 1
93359
```

6. Compare the running time with task c(almost the same)



All Applications

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved
7	0	0	7	0	0 B	80 GB	0 B

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes
4	0	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation
Capacity Scheduler	[MEMORY]	<memory:1536, vCores:1>	<memory:20480, vCores:4>

Show 20 entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCores	Allocated Memory MB	Reserved CPU VCore
application_1600520943120_0009	ubuntu	word count	MAPREDUCE	default	0	Sat Sep 19 23:46:09 +0800 2020	Sat Sep 19 23:46:50 +0800 2020	FINISHED	SUCCEEDED	N/A	N/A	N/A	N/A
application_1600520943120_0008	ubuntu	streamjob4069683088770483485.jar	MAPREDUCE	default	0	Sat Sep 19 22:28:39 +0800 2020	Sat Sep 19 22:29:01 +0800 2020	FINISHED	SUCCEEDED	N/A	N/A	N/A	N/A
application_1600520943120_0007	ubuntu	streamjob331727319078999419.jar	MAPREDUCE	default	0	Sat Sep 19 22:26:43 +0800 2020	Sat Sep 19 22:27:08 +0800 2020	FINISHED	FAILED	N/A	N/A	N/A	N/A

Overview 'hadoop-master:9000' (active)

Started:	Thu Sep 17 22:16:30 +0800 2020
Version:	2.9.2, r826afbeae31ca687bc2f8471dc841b66ed2c6704
Compiled:	Tue Nov 13 20:42:00 +0800 2018 by ajisaka from branch-2.9.2
Cluster ID:	CID-d9947b4f-c252-4b78-8245-385ad99223e6
Block Pool ID:	BP-1931485071-127.0.1.1-1600351832416

Summary

Security is off.

Safemode is off.

1 files and directories, 0 blocks = 1 total filesystem object(s).

Heap Memory used 76.7 MB of 185.5 MB Heap Memory. Max Heap Memory is 889 MB.

Non Heap Memory used 38.43 MB of 39.5 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	23.08 GB
DFS Used:	72 KB (0%)
Non DFS Used:	10.04 GB
DFS Remaining:	12.99 GB (56.27%)
Block Pool Used:	72 KB (0%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
Live Nodes	3 (Decommissioned: 0, In Maintenance: 0)
Dead Nodes	0 (Decommissioned: 0, In Maintenance: 0)
Decommissioning Nodes	0
Entering Maintenance 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	Thu Sep 17 22:16:30 +0800 2020
Last Checkpoint Time	Thu Sep 17 22:12:53 +0800 2020

NameNode Journal Status

Current transaction ID: 4	
Journal Manager	State
FileJournalManager(root=/home/ubuntu/hadoop_data/hdfs/namenode)	EditLogOutputStream(/home/ubuntu/hadoop_data/hdfs/namenode/current/edits_inproc)

NameNode Storage

Storage Directory	Type	State
/home/ubuntu/hadoop_data/hdfs/namenode	IMAGE_AND_EDITS	Active

DFS Storage Types

Storage Type	Configured Capacity	Capacity Used	Capacity Remaining	Block Pool Used	Nodes In Service
DISK	23.08 GB	72 KB (0%)	12.99 GB (56.27%)	72 KB	3

Hadoop, 2018.