Lab Sheet 1

Sabaragamuwa University of Sri Lanka

Faculty of Computing

Software Engineering

SE6103 - Parallel and Distributed Systems

Name	: R.D.M. Perera
Reg No	: 19APSE4287
Academic Period	: 3rd Year 2nd Semester
Due Date	: 18/11/2024

Check Docker Version

docker –version

```
PS C:\Users\User> docker --version
Docker version 24.0.5, build ced0996
```

Checking the Images which are already downloaded

docker images

```
PS C:\Users\User> docker images
REPOSITORY                TAG       IMAGE ID       CREATED         SIZE
dockerapp                 1.1       e251ed860845   3 weeks ago     159MB
nginx                     latest    3b25b682ea82   6 weeks ago     192MB
alpine                    latest    2df64a0493bc   8 months ago    7.38MB
mevin                     0.0.1     2df64a0493bc   8 months ago    7.38MB
mevin                     0.0.2     2df64a0493bc   8 months ago    7.38MB
<none>                    <none>    c3a1088472dd   8 months ago    7.38MB
postgres                  latest    b9390dd1ea18   9 months ago    431MB
dpage/pgadmin4            7.7       ce8ae42da4a0   14 months ago   545MB
postgres                  15.4      68a92c148701   14 months ago   411MB
hello-world               latest    d2c94e258dcb   18 months ago   13.3kB
bde2020/hadoop-namenode   latest    b638307a2119   4 years ago     1.37GB
sequenceiq/hadoop-docker  2.7.1     42efa33d1fa3   8 years ago     1.76GB
```

Run the Docker Container

docker run -it --name hadoop-cluster -p 9870:9870 -p 8088:8088 -p 50070:50070 bde2020/hadoop-namenode:latest /bin/bash

```
PS C:\Users\User> docker run -it --name hadoop-cluster -p 9870:9870 -p 8088:8088 -p 50070:50070 bde2020/hadoop-namenode:latest /bin/bash
Configuring core
 - Setting fs.defaultFS=hdfs://47102cf40a88:8020
Configuring hdfs
 - Setting dfs.namenode.name.dir=file:///hadoop/dfs/name
Configuring yarn
Configuring httpfs
Configuring kms
Configuring mapred
Configuring for multihomed network
root@47102cf40a88:/# hdfs namenode -format
2024-11-18 17:05:28,420 INFO namenode.NameNode: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting NameNode
```

Configure the Hadoop file system

hdfs namenode -format
hdfs namenode &
hdfs datanode &

```
root@47102cf40a88:/# hdfs namenode -format
2024-11-18 17:05:28,420 INFO namenode.NameNode: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = 47102cf40a88/172.17.0.2
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 3.2.1
STARTUP_MSG:   classpath = /etc/hadoop:/opt/hadoop-3.2.1/share/hadoop/common/lib/jersey-core-1.19.jar:/opt/hadoo
re.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/re2j-1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/log4j-1.
on/lib/jackson-databind-2.9.8.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jetty-xml-9.3.24.v20180605.jar:/opt/
ent-4.5.6.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jetty-servlet-9.3.24.v20180605.jar:/opt/hadoop-3.2.1/sha
jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/animal-sniffer-annotations-1.17.jar:/opt/hadoop-3.2.1/share/hadoop
```

```
root@47102cf40a88:/# hdfs namenode &
[1] 169
root@47102cf40a88:/# 2024-11-18 17:06:47,990 INFO namenode.NameNode: STARTUP_MSG:
/*****************************************************
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = 47102cf40a88/172.17.0.2
STARTUP_MSG:   args = []
STARTUP_MSG:   version = 3.2.1
STARTUP_MSG:   classpath = /etc/hadoop:/opt/hadoop-3.2.1/share/hadoop/common/lib/jersey-core-1.19.jar:/opt/hadoop-3.2.1/share
re.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/re2j-1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/log4j-1.2.17.jar:/opt
on/lib/jackson-databind-2.9.8.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jetty-xml-9.3.24.v20180605.jar:/opt/hadoop-3.2.1/
```

```
root@47102cf40a88:/# hdfs datanode &
[2] 293
root@47102cf40a88:/# 2024-11-18 17:07:48,112 INFO datanode.DataNode: STARTUP_MSG:
/*****************************************************
STARTUP_MSG: Starting DataNode
STARTUP_MSG:   host = 47102cf40a88/172.17.0.2
STARTUP_MSG:   args = []
STARTUP_MSG:   version = 3.2.1
STARTUP_MSG:   classpath = /etc/hadoop:/opt/hadoop-3.2.1/share/hadoop/common/lib/jersey-core-1.19.jar:/opt/hadoop-3.2
re.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/re2j-1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/log4j-1.2.17
on/lib/jackson-databind-2.9.8.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jetty-xml-9.3.24.v20180605.jar:/opt/hadoo
ent-4.5.6.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jetty-servlet-9.3.24.v20180605.jar:/opt/hadoop-3.2.1/share/ha
jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/animal-sniffer-annotations-1.17.jar:/opt/hadoop-3.2.1/share/hadoop/comm
```

Check the Hadoop Health using local port

http://localhost:9870/dfshealth.html#tab-overview

Hadoop	Overview	Datanodes	Datanode Volume Failures	Snapshot	Startup Progress	Utilities ▾

Overview '47102cf40a88:8020' (active)

Started:	Mon Nov 18 22:36:50 +0530 2024
Version:	3.2.1, rb3cbbb467e22ea829b3808f4b7b01d07e0bf3842
Compiled:	Tue Sep 10 21:26:00 +0530 2019 by rohithsharmaks from branch-3.2.1
Cluster ID:	CID-be0e1c97-5f9e-463e-a234-b1c9786acb10
Block Pool ID:	BP-1755187175-172.17.0.2-1731949530365

Summary

Security is off.

Safemode is off.

1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).

Heap Memory used 52.92 MB of 285.5 MB Heap Memory. Max Heap Memory is 1.71 GB.

Non Heap Memory used 44.7 MB of 46.15 MB Commited Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	1006.85 GB

Start the node manager and resource manager

yarn nodemanager &
yarn resourcemanager &

Upload sample data to hdfs

hdfs dfs -mkdir -p /user/hadoop/input

```
root@47102cf40a88:/# hdfs dfs -mkdir -p /user/hadoop/input
2024-11-18 17:17:52,322 INFO namenode.FSEditLog: Number of transactions: 4 Total time for transactions(ms): 33 Number of
 transactions batched in Syncs: 0 Number of syncs: 2 SyncTimes(ms): 11
```

hdfs dfs -put $HADOOP_HOME/etc/hadoop/*.xml /user/hadoop/input

```
root@47102cf40a88:/# hdfs dfs -put $HADOOP_HOME/etc/hadoop/*.xml /user/hadoop/input
2024-11-18 17:19:07,616 INFO namenode.FSEditLog: Number of transactions: 5 Total time for transactions(ms): 34 Number of
mber of syncs: 3 SyncTimes(ms): 13
2024-11-18 17:19:07,694 INFO hdfs.StateChange: BLOCK* allocate blk_1073741825_1001, replicas=172.17.0.2:9866 for /user/ha
COPYING_
2024-11-18 17:19:07,745 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHo
2024-11-18 17:19:07,948 INFO datanode.DataNode: Receiving BP-1755187175-172.17.0.2-1731949530365:blk_1073741825_1001 src:
2:9866
2024-11-18 17:19:08,072 INFO DataNode.clienttrace: src: /172.17.0.2:37194, dest: /172.17.0.2:9866, bytes: 8260, op: HDFS_
E_-833154509_1, offset: 0, srvID: ac52af37-d20d-41bd-b05c-c11cfc2cddbf, blockid: BP-1755187175-172.17.0.2-1731949530365:b
3831810
```

Run the word counter job

hadoop   jar $HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-*.jar
wordcount /user/hadoop/input /user/hadoop/output

```
root@47102cf40a88:/# hadoop jar $HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-*.jar wordcount /user/hadoop/input /user/hadoop/outpu
2024-11-18 17:20:32,228 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2024-11-18 17:20:32,339 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2024-11-18 17:20:32,339 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2024-11-18 17:20:32,935 INFO input.FileInputFormat: Total input files to process : 9
2024-11-18 17:20:32,990 INFO mapreduce.JobSubmitter: number of splits:9
2024-11-18 17:20:33,241 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1204723782_0001
2024-11-18 17:20:33,241 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-11-18 17:20:33,494 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2024-11-18 17:20:33,495 INFO mapreduce.Job: Running job: job_local1204723782_0001
2024-11-18 17:20:33,499 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2024-11-18 17:20:33,523 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-11-18 17:20:33,523 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore c
ilures: false
```

Check the output

hdfs dfs -cat /user/hadoop/output/part-r-00000

```
root@47102cf40a88:/# hdfs dfs -cat /user/hadoop/output/part-r-00000
2024-11-18 17:21:38,347 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
"*"      21
"AS      9
"License");     9
"alice,bob      21
"clumping"      1
(ASF)    1
(root    1
(the     9
-->      18
-1       1
-1,      1
0.0      1
1-MAX_INT.      1
```

Stop the Cluster

exit

docker stop hadoop-cluster

```
root@47102cf40a88:/# exit
exit
PS C:\Users\User> docker stop hadoop-cluster
hadoop-cluster
PS C:\Users\User>
```