

Lab Sheet - 01

Sabaragamuwa University of Sri Lanka

Faculty of Computing

Software Engineering

SE5110 – Business Intelligence

Name : I.K.A.D.Thilakshana

Reg. No : 19APSE4293

Academic Period : 3rd Year 1st Semester

Degree Program : Software Engineering

1. Check docker version

```
PS C:\Users\ASUS> docker --version
Docker version 27.2.0, build 3ab4256
PS C:\Users\ASUS>
```

2. Pull and verify the Hadoop image

```
PS C:\Users\ASUS> docker pull bde2020/hadoop-namenode:latest
latest: Pulling from bde2020/hadoop-namenode
Digest: sha256:fd74110805132d646cf6f12635efc0919e1fb2ac5bd376c5366272fc261301e
Status: Image is up to date for bde2020/hadoop-namenode:latest
docker.io/bde2020/hadoop-namenode:latest
```

3. Run the docker container

```
PS C:\Users\ASUS> 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://a45eb3a929cd: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@a45eb3a929cd:/#
```

4. Configure the Hadoop file system

```
root@010e629ba610:/#
root@010e629ba610:/# hdfs namenode -format
2024-11-18 15:24:23,200 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG: host = 010e629ba610/172.17.0.2
STARTUP_MSG: args = [-format]
STARTUP_MSG: version = 3.2.1
STARTUP_MSG: classpath = /etc/hadoop:/opt/hadoop-3.2.1/share/hadoop/common/lib/jackson-core-asl-1.
9.13.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/javax.servlet-api-3.1.0.jar:/opt/hadoop-3.2.1/sha
re/hadoop/common/lib/nimbus-jose-jwt-4.41.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/netty-3.10
.5.Final.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jackson-databind-2.9.8.jar:/opt/hadoop-3.2.1/
share/hadoop/common/lib/audience-annotations-0.5.0.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/ker
```

5. Check the Hadoop health using local port

```
root@010e629ba610:/#
root@010e629ba610:/# hdfs dfs -mkdir -p /user/hadoop/input
2024-11-18 15:51:03,682 INFO namenode.FSEditLog: Number of transactions: 4 Total time for transactio
ns(ms): 12 Number of transactions batched in Syncs: 0 Number of syncs: 2 SyncTimes(ms): 8
root@010e629ba610:/# |
```

6. Start the node manager and resource manager

Upload sample data to HDFS

```

root@010e629ba610:/#
root@010e629ba610:/# hdfs dfs -mkdir -p /user/hadoop/input
2024-11-18 15:51:03,682 INFO namenode.FSEditLog: Number of transactions: 4 Total time for transactions(ms): 12 Number of transactions batched in Syncs: 0 Number of syncs: 2 SyncTimes(ms): 8
root@010e629ba610:/#

```

```

root@010e629ba610:/#
root@010e629ba610:/# hdfs dfs -put $HADOOP_HOME/etc/hadoop/*.xml /user/hadoop/input
2024-11-18 15:53:03,143 INFO namenode.FSEditLog: Number of transactions: 5 Total time for transactions(ms): 13 Number of transactions batched in Syncs: 2 Number of syncs: 3 SyncTimes(ms): 10
2024-11-18 15:53:03,184 INFO hdfs.StateChange: BLOCK* allocate blk_1073741825_1001, replicas=172.17.0.2:9866 for /user/hadoop/input/capacity-scheduler.xml._COPYING_
2024-11-18 15:53:03,203 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2024-11-18 15:53:03,288 INFO datanode.DataNode: Receiving BP-1887826647-172.17.0.2-1731943463769:blk

```

7. Run the word counter job

0

```

root@010e629ba610:/#
root@010e629ba610:/# hadoop jar $HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-*.jar wordcount /user/hadoop/input /user/hadoop/output
2024-11-18 15:55:06,555 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2024-11-18 15:55:06,615 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2024-11-18 15:55:06,615 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2024-11-18 15:55:06,789 INFO input.FileInputFormat: Total input files to process : 9
2024-11-18 15:55:06,807 INFO mapreduce.JobSubmitter: number of splits:9
2024-11-18 15:55:06,914 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local313506135_0001

```

8. Check the output

```

root@010e629ba610:/#
root@010e629ba610:/# hdfs dfs -cat /user/hadoop/output/part-r-00000
2024-11-18 15:58:36,955 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

```