

# Lab Sheet 2

**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	: 3 <sup>rd</sup> Year 2 <sup>nd</sup> Semester
Due Date	: 25/11/2024

## Setting Up the Distributed Hadoop Cluster

```
version: '3'
services:
  namenode:
    image: bde2020/hadoop-namenode:latest
    container_name: namenode
    environment:
      - CLUSTER_NAME=testhadoop
      - CORE_CONF_fs_defaultFS=hdfs://namenode:8020
    ports:
      - "9870:9870" # Web UI
      - "9000:9000" # NameNode port
    volumes:
      - namenode-data:/hadoop/dfs/name

  datanode:
    image: bde2020/hadoop-datanode:latest
    container_name: datanode
    environment:
      - CORE_CONF_fs_defaultFS=hdfs://namenode:8020
    volumes:
      - datanode-data:/hadoop/dfs/data

  depends_on:
    - namenode

  historyserver:
    image: bde2020/hadoop-historyserver:latest
    container_name: historyserver
    depends_on:
      - namenode
      - datanode
    ports:
      - "8188:8188" # Job History server UI
    environment:
      - CORE_CONF_fs_defaultFS=hdfs://namenode:8020

volumes:
  namenode-data:
  datanode-data:
```

## Deploy the Cluster

```
PS C:\Docker> docker-compose up -d
[+] Running 18/5
✓ historyserver 9 layers [██████████] 0B/0B Pulled
✓ datanode 7 layers [██████████] 0B/0B Pulled

[+] Running 5/5
[+] Running 5/6er_default Created
[+] Running 5/6er_default Created
[+] Running 5/6er_default Created
[+] Running 5/6er_default Created
[+] Running 5/6er_default Created
[+] Running 5/6er_default Created
[+] Running 5/6er_default Created

• PS C:\Docker> docker exec -it namenode hdfs dfs -mkdir -p /input
```

## Verify Cluster Status

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities ▾

### Overview 'namenode:8020' (active)

Started:	Mon Nov 25 13:44:15 +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-d15712e9-520a-41aa-8fed-d77fcdc62f23
Block Pool ID:	BP-880098654-172.19.0.2-1732522451910

## Download Sample Data

### Upload Data to the HDFS

```
• PS C:\Docker> docker exec -it namenode hdfs dfs -put /sample.txt /input
2024-11-25 08:19:46,116 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false

• PS C:\Docker> docker exec -it namenode ls /sample.txt
/sample.txt
PS C:\Docker> docker exec -it namenode hadoop jar
```

## Running a MapReduce Job

```
PS C:\Docker> docker exec -it namenode bash
root@61e13ba3012a:/# hadoop jar /opt/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-*.jar wordcount /input /output
```

## View Job Output

```
root@61e13ba3012a:/# hadoop jar /opt/hadoop-3.2.1/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.2.1.jar wordcount /input /output
2024-11-25 08:31:05,217 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2024-11-25 08:31:05,424 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2024-11-25 08:31:05,424 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2024-11-25 08:31:05,790 INFO input.FileInputFormat: Total input files to process : 1
2024-11-25 08:31:05,817 INFO mapreduce.JobSubmitter: number of splits:1
2024-11-25 08:31:06,120 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1280096721_0001
```

```
root@61e13ba3012a:/# exit
exit
PS C:\Docker> docker exec -it namenode hdfs dfs -ls /output
Found 2 items
-rw-r--r-- 3 root supergroup 0 2024-11-25 08:31 /output/_SUCCESS
-rw-r--r-- 3 root supergroup 124 2024-11-25 08:31 /output/part-r-00000
```

## Analyze and Cleanup

```
PS C:\Docker> docker exec -it namenode hdfs dfs -cat /output/part-r-00000
2024-11-25 08:33:48,085 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
"Hadoop 1
> 1
> 1
a 1
and 1
big 1
data," 1
distributed 1
echo 1
for 1
framework 1
is 1
of 1
processing 1
framework 1
is 1
framework 1
is 1
of 1
framework 1
is 1
framework 1
framework 1
is 1
of 1
processing 1
sample.txt 1
storage 1
```

```
PS C:\Docker> docker-compose down
[+] Running 4/4
✓ Container historyserver Removed
✓ Container datanode Removed
✓ Container namenode Removed
✓ Network docker_default Removed
```