LAB SHEET 02

Sabaragamuwa University of Sri Lanka

Faculty of Computing

Software Engineering

SE6103 – Parallel and Distributed Systems

Name: Disara Mapalagama
Reg. No: 19APSE4302
Academic Period: 3rd Year 2nd Semester
Degree Program: BSc (Hons) in Software Engineering
Due Date: 25.11.2024

Lab Sheet: Distributed Hadoop Application

1. Confirming Pre-requisites

Command:

Docker images

Output:

```
PS C:\Users\HP> docker images
REPOSITORY                     TAG       IMAGE ID       CREATED         SIZE
nginx                          latest    3b25b682ea82   6 weeks ago     192MB
alpine                         latest    91ef0af61f39   2 months ago    7.8MB
hello-world                    latest    d2c94e258dcb   18 months ago   13.3kB
bde2020/hadoop-namenode        latest    b638307a2119   4 years ago     1.37GB
bde2020/hadoop-historyserver   latest    9d3d45ba75ae   4 years ago     1.37GB
bde2020/hadoop-datanode        latest    9e7b07fcba31   4 years ago     1.37GB
```

Start containers:

```
PS C:\Users\HP> docker run -d --name hadoop-namenode bde2020/hadoop-namenode
d6c54ebc9a48f6fd421b8cf1d1e0f1c9fb8ae6c547da9394dba69a54ffb93160
PS C:\Users\HP> docker run -d --name hadoop-datanode bde2020/hadoop-datanode
16cb27b41b06aac4f47227bcc3aa677852b54136420d4146c04c5d34877b91b2
PS C:\Users\HP> docker run -d --name hadoop-historyserver bde2020/hadoop-historyserver
15d247e5af8c297c8ae3e909d5209915548130abbf763f2cb7224a7c050e7d2c
```
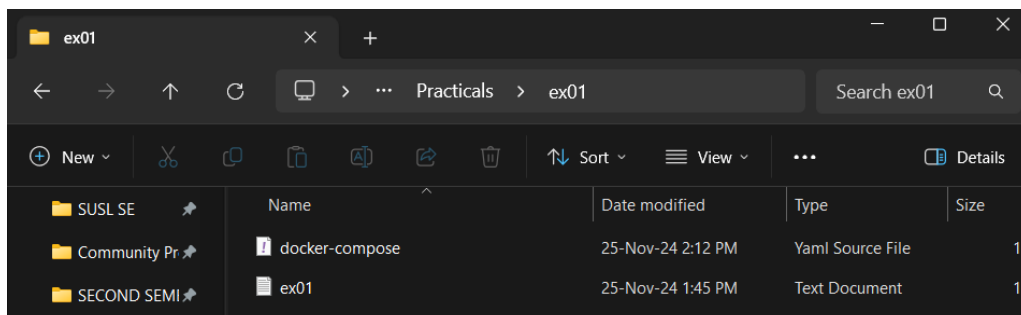
Confirmed running:

```
PS C:\Users\HP> docker ps
CONTAINER ID   IMAGE                          COMMAND               CREATED         STATUS
          PORTS         NAMES
15d247e5af8c   bde2020/hadoop-historyserver   "/entrypoint.sh /run…"  47 seconds ago  Up 46 seconds (h
ealthy)   8188/tcp   hadoop-historyserver
16cb27b41b06   bde2020/hadoop-datanode        "/entrypoint.sh /run…"  55 seconds ago  Up 54 seconds (h
ealthy)   9864/tcp   hadoop-datanode
```

2. Tasks

2.1. Created docker-compose.yaml and a text file (ex01.txt) in exo1 directory:

docker-compose.yaml:

```yaml
services:
  namenode:
    image: bde2020/hadoop-namenode:latest
    container_name: namenode
    environment:
      - CLUSTER_NAME=ShopSmartCluster
      - CORE_CONF_fs_defaultFS=hdfs://namenode:8020
    ports:
      - "9870:9870"
      - "9000:9000"
    volumes:
      - namenode-data:/hadoop/dfs/namenode

  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"
    environment:
      - CORE_CONF_fs_defaultFS=hdfs://namenode:8020
    volumes:
      - namenode-data:/hadoop/dfs/namenode
      - datanode-data:/hadoop/dfs/data

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

ex01.txt:

```
EX01
docker-compose.yaml
ex01.txt

ex01.txt
1    I am Disara. Disara is my name. This is a random text. This text file will be
```

"I am Disara. Disara is my name. This is a random text. This text file will be used to count the number of words in this, to demonstrate how Hadoop can be used to show wordcount in a text file."

3. Task 2: Deployed the cluster and uploaded files:

```
PS E:\SUSL SE\THIRD YEAR\SECOND SEMESTER\Parallel and Distributed Systems\Practicals\ex01> docker exec
it namenode hdfs dfs -ls /output1
>>
Found 2 items
-rw-r--r--   3 root supergroup          0 2024-11-25 09:53 /output1/_SUCCESS
-rw-r--r--   3 root supergroup        223 2024-11-25 09:53 /output1/part-r-00000
```

4. Task 3: Run the MapReduce job and got the wordcount:

```
PS E:\SUSL SE\THIRD YEAR\SECOND SEMESTER\Parallel and Distributed Systems\Practicals\ex01> docker exec -
it namenode hdfs dfs -cat /output1/part-r-00000
>>
2024-11-25 09:53:29,846 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted
= false, remoteHostTrusted = false
Disara  1
Disara. 1
Hadoop  1
I       1
This    2
a       2
am      1
be      2
can     1
count   1
demonstrate     1
file    1
file.   1
how     1
in      2
```

```
is      2
my      1
name.   1
number  1
amenode
    Learn more at https://docs.docker.com/go/debug-cli/
```

5. Task 4: Analyzed and cleaned up:

```
PS E:\SUSL SE\THIRD YEAR\SECOND SEMESTER\Parallel and Distributed Systems\Practicals\ex01> docker-compo
se down
[+] Running 4/4
 ✓ Container historyserver  Removed                                                    10.6s
 ✓ Container datanode       Removed                                                    10.5s
 ✓ Container namenode       Removed                                                    10.6s
 ✓ Network ex01_default     Removed                                                     0.4s
PS E:\SUSL SE\THIRD YEAR\SECOND SEMESTER\Parallel and Distributed Systems\Practicals\ex01>
```