Assignment I **Hadoop Exercise** 19APSE4314 Tharushi Wijethunga Department of Software Engineering

First Attempt

Step 1

Prerequisites

Docker: Ensure Docker is installed and running on your system. You can check if Docker is installed by running:

bash

Copy code

docker --version

Output

```
Microsoft Windows (Warsion 18.8.22631.4468)
(c) Microsoft Corporation All rights reserved.

C) Wars DELL> docker - version
Docker version 25.8.3, build 4debf41
```

Step 2

Step 1: Pull the Hadoop Docker Image

- 1. Choose a Hadoop Docker Image:
 - Many Docker images are available for Hadoop single-node setups. For this lab, we'll use the bde2020/hadoop-namenode image from the Big Data Europe repository on Docker Hub.

Pull the Docker image:

docker pull bde2020/hadoop-namenode:latest

Then it generates error

```
C:\Users\DELL*docker pull bde2829/hadoop-namenode:latest error during connect: this error may indicate that the docker dammen is not running. Post "http://k2FN3F %2Fpipe%2Fdocker_engine/vi_24/images/create?frumImagerbde2828%2Fhadoop-namenode&tagrlatest": open //./pipe/docker_engine: The system cannot find the file specified.

C:\Users\DELL*docker pull bde2828/hadoop-namenode
Using default tag: latest
error during connect: this error may indicate that the docker dammen is not running. Post "http://k2FN3F %2Fpipe%2Fdocker_engine/vi_24/images/create?frumImage=bde2828%2Fhadoop-namenode&tag=latest": open //./pipe/docker_engine: The system cannot find the file specified.
```

Start docker

```
C:\Users\DELL'docker pull bde2026/hadoop-namenode
Using default tig: latest
Latest: Pulling fram bde2028/hadoop-namenode
3192219af604: Pulling fa layer
3192219af604: Pulling fa layer
3192219af604: Pull complete
81392219af604: Pull complete
```

Step 3

C

2. Verify the Download:

List Docker images to confirm the Hadoop image is downloaded:

docker images

Step 4

Step 2: Start the Hadoop Container

1. Run the Container:

Start a container with the necessary configurations to act as a single-node Hadoop cluster:

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

```
C:\Users\DELL>docker run -it --name hadoop-cluster -p 9879:9878 -p 8888:8888 -p 50070:50078 bde2028/hadoop-namenode:latest/bin/bash
Configuring core
- Setting fs. defaultFS=hdfs://42d338f88c67:8928
Configuring hdfs
- Setting dfs. namenode.name.dir=file:///hadoop/dfs/name
Configuring parn
Configuring yarn
Configuring hdfs
Configuring yarn
Configuring far autithosed network
rectpu2d338f88c67:/s
```

Step 5

0

2. Start Hadoop Services:

Once inside the container's shell, start the Hadoop services:

```
start-all.sh
```

0

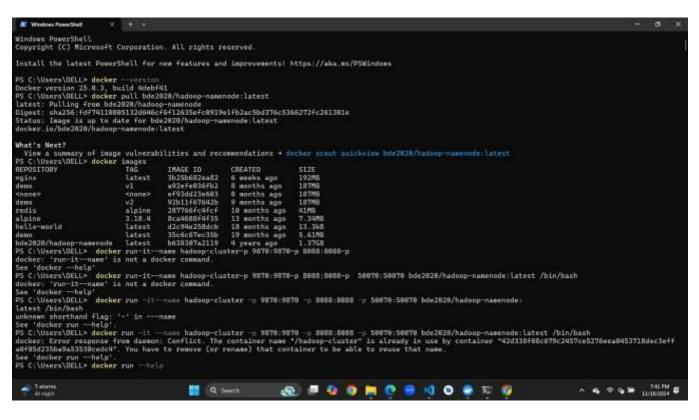
This will initialize HDFS and YARN.

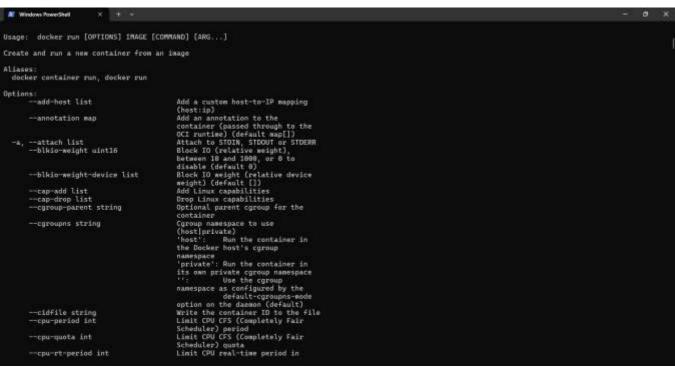
```
root@42d338f08c67:/# find / -name start-all.sh
/opt/hadoop-3.2.1/sbin/start-all.sh
root@42d338f08c67:/# /opt/hadoop-3.2.1/sbin/start-all.sh
Starting namenodes on [42d338f08c67]
ERROR: Attempting to operate on hdfs namenode as root
ERROR: but there is no HDFS_NAMENODE_USER defined. Aborting operation.
Starting datanodes
ERROR: Attempting to operate on hdfs datanode as root
ERROR: but there is no HDFS_DATANODE_USER defined. Aborting operation.
Starting secondary namenodes [42d338f08c67]
ERROR: Attempting to operate on hdfs secondarynamenode as root
```

From this moments onwards the above error remains same

Second Attempt

Then I have start from the beginning in the powershell





| Modes | New | Head | Head