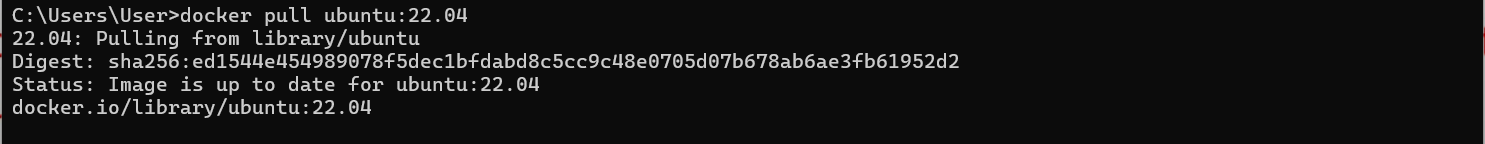
**Solution to installing hadoop:**

1)first pull the image to use :

----docker pull ubuntu:22.04

22.04: Pulling from library/ubuntu

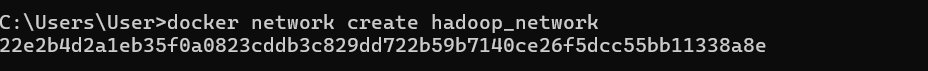
The Digest: sha256:ed1544e454989078f5dec1bfdabd8c5cc9c48e0705d07b678ab6ae3fb61952d2



2)create network to use ?

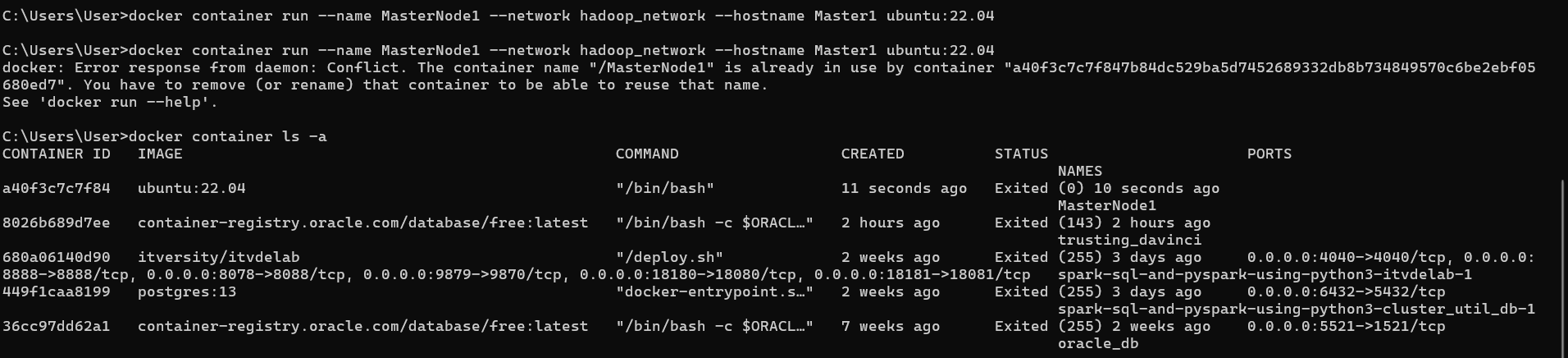
-----docker network create hadoop\_network

The Digest: 22e2b4d2a1eb35f0a0823cddb3c829dd722b59b7140ce26f5dcc55bb11338a8e



3) build the container based on the image and attach a shell to it

(take on mind to name it /name host /give it the network/…)



4) steps to setup

# 1. Update and install dependencies

1. package update and upgrade all
2. sudo apt update -y
3. sudo apt upgrade -y
4. sudo apt install sudo
5. apt install ssh -y
6. apt install openjdk-8-jdk -y
7. apt install nano -y
8. apt install curl -y

# 2. Set root password (optional)

1. echo "root:123" | sudo chpasswd

# 3. Create hadoop user and group

1. addgroup hadoop
2. adduser --ingroup hadoop --disabled-password --gecos "" Hadoop
3. usermod -aG sudo hadoop

# 4. Set environment variables

1. sudo -u hadoop bash -c 'echo " or nano ~/.bashrc
2. export HADOOP\_HOME=/usr/local/hadoop
3. export PATH=\$HADOOP\_HOME/bin:\$HADOOP\_HOME/sbin:\$PATH:/usr/local/zookeeper/bin
4. export JAVA\_HOME=/usr/lib/jvm/java-8-openjdk-amd64
5. export HDFS\_NAMENODE\_USER=hadoop
6. export HDFS\_DATANODE\_USER=hadoop
7. export HDFS\_SECONDARYNAMENODE\_USER=hadoop
8. export YARN\_RESOURCEMANAGER\_USER=hadoop
9. export YARN\_NODEMANAGER\_USER=hadoop
10. " >> /home/hadoop/.bashrc'
11. source ~/.bashrc

# 5. Install Hadoop

1. sudo wget <https://dlcdn.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz> /tmp
2. sudo tar -xzf /tmp/hadoop-3.3.6.tar.gz -C /usr/local/
3. sudo mv /usr/local/hadoop-3.3.6 /usr/local/hadoop
4. sudo rm -f /tmp/hadoop-3.3.6.tar.gz
5. sudo chown -R hadoop:hadoop /usr/local/hadoop
6. sudo chmod -R 755 /usr/local/Hadoop
7. nano /usr/local/hadoop/etc/hadoop/core-site.xml
8. nano /usr/local/hadoop/etc/hadoop/hdfs-site.xml
9. nano /usr/local/hadoop/etc/hadoop/mapred-site.xml
10. nano /usr/local/hadoop/etc/hadoop/yarn-site.xml
11. nano /usr/local/hadoop/etc/hadoop/hadoop-env.sh

# 6. Install ZooKeeper

1. sudo wget https://archive.apache.org/dist/zookeeper/zookeeper-3.6.3/apache-zookeeper-3.6.3-bin.tar.gz -P /tmp
2. sudo tar -xvf /tmp/apache-zookeeper-3.6.3-bin.tar.gz -C /usr/local/
3. sudo mv /usr/local/apache-zookeeper-3.6.3-bin /usr/local/zookeeper
4. sudo rm -f /tmp/apache-zookeeper-3.6.3-bin.tar.gz
5. sudo chown -R hadoop:hadoop /usr/local/zookeeper
6. sudo chmod -R 755 /usr/local/zookeeper
7. sudo mkdir -p /usr/local/zookeeper/data
8. sudo chown -R hadoop:hadoop /usr/local/zookeeper/data
9. sudo touch /usr/local/zookeeper/data/myid
10. echo "1" | sudo tee /usr/local/zookeeper/data/myid
11. cp /usr/local/zookeeper/conf/zoo\_sample.cfg /usr/local/zookeeper/conf/zoo.cfg
12. nano /usr/local/zookeeper/zoo.cfg

# 7. Grant sudo access to hadoop user

1. echo "hadoop ALL=(ALL) NOPASSWD:ALL" | sudo tee -a /etc/sudoers

# 8. Setup SSH

1. sudo mkdir -p /var/run/sshd
2. mkdir -p /home/hadoop/.ssh
3. chown hadoop:hadoop /home/hadoop/.ss
4. sudo -u hadoop ssh-keygen -t rsa -f /home/hadoop/.ssh/id\_rsa -N ""
5. sudo -u hadoop cat /home/hadoop/.ssh/id\_rsa.pub >>/home/hadoop/.ssh/authorized\_keys
6. sudo chmod 600 /home/hadoop/.ssh/authorized\_keys

# 9. Create Hadoop directories

1. sudo mkdir -p /usr/local/hadoop/yarn\_data/hdfs/{namenode,datanode,journalnode}
2. sudo chown -R hadoop:hadoop /usr/local/hadoop/yarn\_data
3. sudo chmod -R 755 /usr/local/hadoop/yarn\_data

# 10. Format HDFS

1. sudo -u hadoop /usr/local/hadoop/bin/hdfs namenode -format

# 11. Start services

1. ZooKeeper → JournalNodes → NameNode format → Start other services on one node
2. ZooKeeper → JournalNodes → Standby format → Start other services on one node

echo "YAY 😊Hadoop and ZooKeeper setup complete!"

Scale ?

--docker run -it --network Hadoop\_network -v ${PWD}:/mnt --hostname worker2 --name worker2 -e NODE\_TYPE=worker Hadoop-image

Test the active Namenode and active rm

* hdfs haadmin -getAllServiceState
* yarn rmadmin -getAllServiceState

/usr/local/zookeeper/bin/zkServer.sh status

hdfs haadmin -getAllServiceState

yarn rmadmin -getAllServiceState

echo hadoooop >> test

hdfs dfs -mkdir /mydir

hdfs dfs -put test /

hdfs dfs -cat /test