

Installing a Single Node Hadoop Cluster

This document assumes that you have a ubuntu system already installed. The user has a sudo permission assigned. Set hostname and add it to the /etc/hosts file.

Create a new user.

```
sudo adduser hduser
sudo passwd hduser
```

Provide sudo permissions to the above user.

```
sudo visudo
```

Add following line

```
hduser    ALL:(ALL)    NOPASSWD: ALL
```

Save the file.

Login as the hduser .

1. Install Java

The Hadoop release supports Java 11 or Java 8 only. Here we install Java 8.

```
sudo apt install openjdk-8-jdk
```

Confirm the Java installation using

```
java -version
```

2. Set variables in ~/.bashrc file.

```
#JAVA
export JAVA_HOME=/usr/java/jdk1.8.0_212-amd64/
export JRE_HOME=/usr/java/jdk1.8.0_212-amd64/jre

#Hadoop Environment Variables
export HADOOP_HOME=/usr/local/hadoop
export HADOOP_CONF_DIR=/usr/local/hadoop/etc/hadoop
export HADOOP_LOG_DIR=$HADOOP_HOME/logs

# Add Hadoop bin/ directory to PATH
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
export PDSH_RCMD_TYPE=ssh
```

3. Reload the .bashrc file using

```
source ~/.bashrc
```

4. Setup SSH

```
sudo apt install ssh -y
sudo apt install pdsh -y

sudo systemctl start ssh
sudo systemctl enable ssh
```

5. Enable passwordless SSH

```
ssh-keygen -t rsa
```

Press Enter on all prompts. This will create a .ssh folder and id_rsa and id_rsa.pub file in the user home directory.

Copy the public key file

```
ssh-copy-id -i ~/.ssh/id_rsa.pub hduser@localhost
```

Check using command , it should not ask for the password.

```
ssh hduser@localhost
```

6. Download Hadoop

```
cd
```

```
Wget -c -O hadoop.tar.gz
https://dlcdn.apache.org/hadoop/common/hadoop-3.2.4/hadoop-3.2.4.tar.gz
```

```
sudo mkdir /usr/local/hadoop
```

```
tar xvzf hadoop.tar.gz
```

```
sudo mv hadoop-3.2.4/* /usr/local/hadoop
```

```
sudo chown -R hduser:hduser /usr/local/hadoop
```

```
sudo chmod 755 -R /usr/local/hadoop
```

7. Configure Hadoop

Create directories for Namenode and datanode.

```
mkdir -p /usr/local/hadoop/hadoop_store/tmp
mkdir -p /usr/local/hadoop/ho_store/namenode
mkdir -p /usr/local/hadoop/hd_store/datanode
chmod -R 755 /usr/local/hadoop/hadoop_store/tmp
chmod -R 755 /usr/local/hadoop/hd_store/namenode
```

```
chmod -R 755 /usr/local/hadoop/hd_store/datanode
```

```
cd $HADOOP_HOME/etc/hadoop
```

A. Edit the `hadoop-env.sh` file and define following variables.

```
#JAVA
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64/
export JRE_HOME=/usr/lib/jvm/java-8-openjdk-amd64/jre
#Hadoop Environment Variables
export HADOOP_HOME=/usr/local/hadoop
export HADOOP_CONF_DIR=/usr/local/hadoop/etc/hadoop
export HADOOP_LOG_DIR=$HADOOP_HOME/logs
export HDFS_NAMENODE_USER=hduser
export HDFS_DATANODE_USER=hduser
export HDFS_SECONDARYNAMENODE_USER=hduser
export YARN_RESOURCEMANAGER_USER=hduser
export YARN_NODEMANAGER_USER=hduser
export YARN_NODEMANAGER_USER=hduser
# Add Hadoop bin/ directory to PATH
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
```

B. Edit `core-site.xml` and add following.

```
<configuration>
<property>
<name>hadoop.tmp.dir</name>
<value>/usr/local/hadoop/hadoop_store/tmp</value>
</property>
<property>
<name>fs.defaultFS</name>
<value>hdfs://vlsi1:9000</value>
</property>
</configuration>
```

C. Edit `yarn-site.xml` and add following.

```
<configuration>
<property>
<name>yarn.nodemanager.aux-services</name>
<value>mapreduce_shuffle</value>
</property>
</configuration>
```

D. Edit hdfs-site.xml and add following.

```
<configuration>
<property>
<name>dfs.replication</name>
<value>1</value>
</property>
<property>
<name>dfs.name.dir</name>
<value>/usr/local/hadoop/hadoop_store/namenode</value>
</property>
<property>
<name>dfs.data.dir</name>
<value>/usr/local/hadoop/hadoop_store/datanode</value>
</property>
</configuration>
```

E. Edit mapred-site.xml and add following

```
<configuration>
<property>
<name>mapreduce.framework.name</name>
<value>yarn</value>
</property>
</configuration>
```

F. Edit workers file and add localhost entry.

8. Start Hadoop daemons

```
cd $HADOOP_HOME/sbin
```

```
hadoop namenode -format
```

```
start-dfs.sh
```

```
Check using jps command
```

```
Start-yarn.sh
```

```
Check using jps command.
```

If you get all daemons running, it means your single node Hadoop cluster is ready.

