

Exp-1**Downloading and installing Hadoop on Ubuntu, Understanding different Hadoop modes, Startup scripts, Configuration files****Aim:**

To successfully install, configure, and run Hadoop on a local system using a single-node setup.

Procedure:**1. Install Java and SSH:**

- Update your package lists and install OpenJDK 8 and SSH.

```
sudo apt update
```

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

```
java -version # Verify Java installation
```

```
sudo apt install ssh
```

2. Create Hadoop User:

- Add a dedicated user for Hadoop and generate SSH keys for passwordless SSH.

```
sudo adduser hadoop
```

```
su - hadoop # Switch to Hadoop user
```

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

```
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
```

```
chmod 640 ~/.ssh/authorized_keys
```

```
ssh localhost # Test SSH connection to localhost
```

3. Download and Install Hadoop:

- Download the latest Hadoop version (3.3.6), extract the tarball, and move it to the desired location.

```
wget https://downloads.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz
```

```
tar -xvzf hadoop-3.3.6.tar.gz
```

```
mv hadoop-3.3.6 hadoop
```

4. Configure Environment Variables:

- **Update .bashrc to include Hadoop and Java paths.**

```
nano ~/.bashrc
```

```
# Add the following lines at the end
```

```
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
```

```
export HADOOP_HOME=$HOME/hadoop
export PATH=$PATH:$HADOOP_HOME/bin:$HADOOP_HOME/sbin
source ~/.bashrc # Apply changes
```

5. Edit Hadoop Configuration Files:

- Modify configuration files to set up the necessary Hadoop directories and services.
- **core-site.xml:**

```
nano $HADOOP_HOME/etc/hadoop/core-site.xml
```

Add between <configuration></configuration>:

```
<property>
  <name>fs.defaultFS</name>
  <value>hdfs://localhost:9000</value>
</property>
```

- **hdfs-site.xml:**

```
nano $HADOOP_HOME/etc/hadoop/hdfs-site.xml
```

Add:

```
<property>
  <name>dfs.replication</name>
  <value>1</value>
</property>
<property>
  <name>dfs.namenode.name.dir</name>
  <value>file:///home/hadoop/hadoopdata/hdfs/namenode</value>
</property>
<property>
  <name>dfs.datanode.data.dir</name>
  <value>file:///home/hadoop/hadoopdata/hdfs/datanode</value>
</property>
```

- **mapred-site.xml:**

```
cp $HADOOP_HOME/etc/hadoop/mapred-site.xml.template
  $HADOOP_HOME/etc/hadoop/mapred-site.xml
```

```
nano $HADOOP_HOME/etc/hadoop/mapred-site.xml
```

Add:

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

- **yarn-site.xml:**

```
nano $HADOOP_HOME/etc/hadoop/yarn-site.xml
```

Add:

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

6. Format the NameNode:

- Format the HDFS NameNode.

```
hdfs namenode -format
```

7. Start Hadoop:

- Start Hadoop services (NameNode, DataNode, ResourceManager, and NodeManager).

```
start-all.sh
```

```
jps # Verify running services
```

8. Access Web Interfaces:

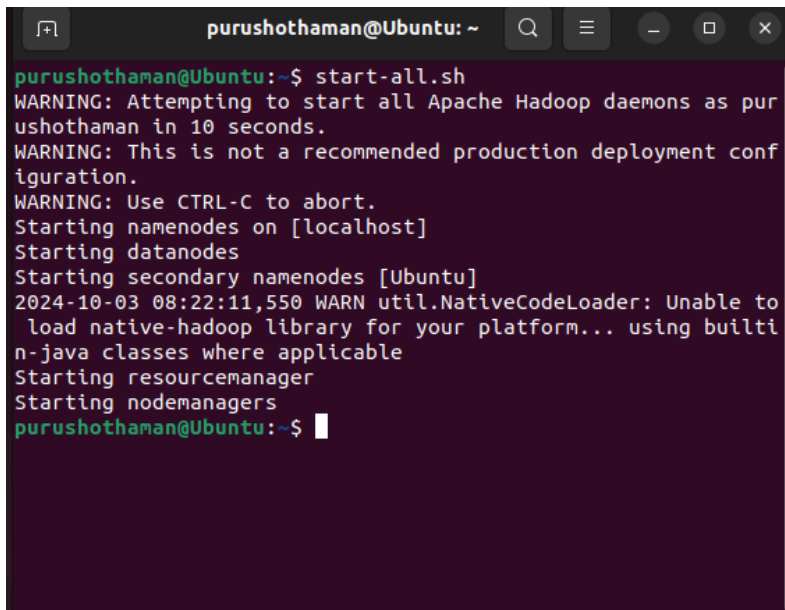
- Verify that Hadoop is running by accessing the following URLs:

- **NameNode:** <http://localhost:9870>
- **Resource Manager:** <http://localhost:8088>

9. Stop Hadoop Cluster:

- Stop all Hadoop services.

```
stop-all.sh
```

A terminal window titled 'purushothaman@Ubuntu: ~' with standard window controls. The terminal shows the execution of 'start-all.sh' and the subsequent startup of Hadoop daemons. The output includes several warnings and status messages for starting namenodes, datanodes, secondary namenodes, the resource manager, and node managers. A timestamp '2024-10-03 08:22:11,550' is visible. The prompt returns to the user after the process completes.

```
purushothaman@Ubuntu:~$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as pur
ushothaman in 10 seconds.
WARNING: This is not a recommended production deployment conf
iguration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [Ubuntu]
2024-10-03 08:22:11,550 WARN util.NativeCodeLoader: Unable to
load native-hadoop library for your platform... using builti
n-java classes where applicable
Starting resourcemanager
Starting nodemanagers
purushothaman@Ubuntu:~$
```

RESULT:

The step-by-step installation and configuration of Hadoop on Ubuntu system have been successfully completed.