

27/08/24

Procedure for Hadoop installation along with commands using virtual box and ubuntu

PREREQUISITES:-

- * Download and install virtual Box.
- * Download the latest LTS (Long-Term Support) version of Ubuntu.

STEP-BY-STEP PROCEDURE:-

1. Create a new virtual Machine.

- * Open VirtualBox and click "New"
- * Choose a name for your VM and select "Ubuntu" as the OS.
- * Select memory allocation atleast 2GB for single node cluster.
- * Create a new virtual hard disk and specify the size atleast 20GB.

2. Install Ubuntu

- * Attach the downloaded ubuntu ISO image to VM's optical drive.
- * Start VM & install Ubuntu
- * During installation, choose suitable platform scheme and set a root password.

3. Update & Upgrade Packages

* After installation, open terminal in Ubuntu & run following commands.

→ Bash

→ sudo apt update

→ sudo apt upgrade

4. Install Java

* Download JDK

* Set JAVA_HOME environment variable by adding
Bash export JAVA_HOME = /path/to/jdk to .bashrc.
file

5. INSTALL HADOOP:

* Download hadoop

* Extract the downloaded tar file to directory
Bash

tar -xvf hadoop-3.x.x.tar.gz -C /opt

* Set HADOOP_HOME environment variable

Bash

export HADOOP_HOME = /opt/hadoop-3.x.x to
.bashrc file.

* Add hadoop bin directory to your PATH

Bash

export PATH = \$HADOOP_HOME/bin : \$PATH

CONFIGURE HADOOP

* Copy the core-site.xml and hdfs-site.xml files from the etc/hadoop directory to your home directory.

Bash

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

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

* Edit files to configure hadoop

```
<configuration>
```

```
<property>
```

```
<name> fs.defaultFS </name>
```

```
<value> hdfs://localhost:9000 </value>
```

```
</property>
```

```
<property>
```

```
<name> dfs.replication </name>
```

```
<value> 1 </value>
```

```
</property>
```

```
</configuration>
```

7. Format the NameNode

* Run the command to format NameNode
Bash

`hdfs namenode -format`.

8. Start Hadoop Services:-

* Run the commands to start NameNode & DataNode services.

Bash

`start -d fs.sh`

9. Verify hadoop Installation.

* Run to check status of hadoop services.

Bash

`jps`

10. Access Hadoop web UI

* Open web browser & navigate to `http://localhost:50070`. you should see the Hadoop web UI.
