

CS19716 - Data Analytics
Assignment - 2

Sreena R
210701256
CSE - D

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

2. Install Ubuntu

- * Attach the downloaded ubuntu ISO image to VM's optical drive.
- * Start VM & install ubuntu

* During installation, choose suitable partition 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
```

6. 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-dfs.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.