

## Assignment-2

HADOOP INSTALLATION USING COMMANDS IN  
VIRTUAL BOX AND UBUNTU

## Prerequisites:

- Download and Install Virtual Box.
- Download the latest LTS version 22.4 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 atleast 24B for single node cluster.
- Create a new virtual hard disk and specify the size atleast 20GB.

2. Install ~~Ja~~ Ubuntu:

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

## 3. Update and Upgrade Packages:

- After installation, open Terminal in Ubuntu and run the following commands:
  - ↳ bash
  - ↳ sudo apt update

↳ sudo apt upgrade.

#### 4. Install Java:

→ Download Java.

→ Set JAVA\_HOME environment variable by adding the line below:

bash export JAVA\_HOME = /path to /jdk to bashrc file.

#### 5. Install Hadoop:

→ Download Hadoop from its official site.

→ Extract the downloaded .tar file to directory bash.

`tar -xvf hadoop-3.x.x.tar.gz -ctapt`

→ Set HADOOP\_HOME environment variable in the bash.

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

→ Add hadoop bin directory to your PATH.

`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 the core-site.xml file.

```
<configuration>
```

```
<property>
```

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

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

```
</property>
```

```
<property>
```

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

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

```
</property>
```

```
</configuration>
```

→ Edit hdfs-site.xml file.

```
<configuration>
```

```
<property>
```

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

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

```
</property>
```

```
<property>
```

```
<name>dfs.namenode.name.dir</name>
```

```
<value>C://hadoop/hdfs/namenode</value>
```

```
</property>
```

```
<property>
```

```
<name>dfs.datanode.data.dir</name>
```

```
<value>C://hadoop/hdfs/datanode</value>
```



</property>

</configuration>

→ Edit yarn-site.xml file.

<configuration>

<property>

<name> yarn.nodemanager.aux-services </name>

<value> mapreduce\_shuffle </value>

</property>

</configuration>

7. Format the namenode:

hdfs namenode -format

8. Start Hadoop services:

start-all.sh

9. Verify hadoop installation:

jps

10. Access the Hadoop Namenode:

http://localhost:9870

Open the web browser and enter the above URL.