**Sipna College of Engineering & Technology, Amravati.**

**Department of Computer Science and Engineering**

**Branch: - Computer Science and Engineering Class: - III Year**

**Subject: - Big Data Analytics Lab Sem: - VI**

**Student Manual**

**PRACTICAL NO. 8**

**Aim: Installation of Hadoop in Linux operating system**

**Software Requirement: Jupyter**

**Theory:**

1.Installation of Hadoop:

Hadoop software can be installed in three modes of

operation:

**• Stand Alone Mode:** Hadoop is a distributed software and is designed to run on a commodity of machines. However, we can install it on a single node in stand-alone mode. In this mode, Hadoop software runs as a single monolithic java process. This mode is extremely useful for debugging purpose. You can first test run your Map-Reduce application in this mode on small data, before actually executing it on cluster with big data.

• **Pseudo Distributed Mode**: In this mode also, Hadoop software is installed on a Single Node.

Various daemons of Hadoop will run on the same machine as separate java processes. Hence all the daemons namely NameNode, DataNode, SecondaryNameNode, JobTracker, TaskTracker run on single machine.

• **Fully Distributed Mode**: In Fully Distributed Mode, the daemons

NameNode, JobTracker,

SecondaryNameNode (Optional and can be run on a separate node) run on the Master Node.

The daemons DataNode and TaskTracker run on the Slave Node.

**Hadoop Installation: Ubuntu Operating System in stand-alone mode**

Steps for Installation

1. **sudo apt-get update**

2.In this step, we will install latest version of JDK

(1.8) on the machine.

The Oracle JDK is the official JDK; however, it is no longer provided by Oracle as a default installation for Ubuntu. You can still install it using apt-get.To install any version, first execute the following commands:

**a. sudo apt-get install python-softwareproperties**

**b. sudo add-apt-repository ppa:webupd8team/**

**java**

**c. sudo apt-get update**

Then, depending on the version you want to install, execute one of the following commands:

**Oracle JDK 7: sudo apt-get install oraclejava7-installer**

**Oracle JDK 8: sudo apt-get install oraclejava8-installer**

**2**.Now, let us setup a new user account for Hadoop installation. This step is optional, but recommended because it gives you flexibility to have a separate account for Hadoop installation by separating this installation from other software installation

**a. sudo adduser hadoop\_dev** ( Upon executing this command, you will prompted to enter the new password for this user. Please enter the password and enter other details. Don’t forget to save the details at the end)

**b. su - hadoop\_dev** ( Switches the user from current user to the new user created i.e Hadoop\_dev)

**3**. Download the latest Hadoop distribution.

a. Visit this URL and choose one of the mirror sites. You can copy the download link and also use

“wget” to download it from command prompt:

**Wget http://apache.mirrors.lucidnetworks.net/hadoop/common/hadoop-2.7.0/hadoop-2.7.0.tar.gz**

**4**.Untar the file **:**

**tar xvzf hadoop-2.7.0.tar.gz**

**5.**Rename the folder to hadoop2

**mv hadoop-2.7.0 hadoop2**

**6.** Edit configuration file /home/hadoop\_dev/hadoop2/etc/hadoop/hadoop-env.sh and set JAVA\_HOME in that file.

**a. vim /home/hadoop\_dev/hadoop2/etc/hadoop/hadoop-env.sh**

b. uncomment JAVA\_HOME and update it following line:

**export JAVA\_HOME=/usr/lib/jvm/java-8-oracle** ( Please check for your relevant java

installation and set this value accordingly. Latest versions of Hadoop require > JDK1.7)

**7**. Let us verify if the installation is successful or not( change to home directory

**cd /home/hadoop\_dev/hadoop2**/):

1. **bin/hadoop** ( running this command should prompt you with various options)

**8**.. This finishes the Hadoop setup in stand-alone mode.

**9.** Let us run a sample hadoop programs that is provided to you in the download package:

**$ mkdir input** (create the input directory)

**$ cp etc/hadoop /\*.xml**

**input** ( copy over all the xml files to input folder)

**$ bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.0.jar grep**

**input output 'dfs[a-z.]+' (grep/find all the files matching the pattern ‘dfs[a-z.]+’** and copy

those files to output directory)

**$ cat output/\*** (look for the output in the output directory that Hadoop creates for you).

This completes the installation part of Hadoop.

**Conclusion**: Thus we have studied hadoop Installation