



JAYOTI VIDYAPEETH WOMEN'S UNIVERSITY, JAIPUR
Government of Rajasthan established
Through ACT No. 17 of 2008 as per UGC ACT 1956
NAAC Accredited University

Faculty of Education and methodology

Department of Computer Science and Engineering

Faculty Name- Jv'n Narendra Kumar Chahar (Assistant Professor)

Program- B.Tech 6thSemester

Course Name – Web Intelligence, HADOOP and Big Data Analysis

Session no.: 7

Session Name- Installing Hadoop in Pseudo Distributed Mode

Academic Day starts with –

- Greeting with saying '**Namaste**' by joining Hands together following by 2-3 Minutes Happy session, Celebrating birthday of any student of respective class and **National Anthem**.

Lecture starts with- quotations' answer writing

- Review of previous Session- **Hadoop installation in pseudo distributed Modes**

Topic to be discussed today- Today We will discuss about – **Installing Hadoop in pseudo distributed Mode**

- Lesson deliverance (ICT, Diagrams & Live Example)-
➤ Diagrams

Introduction & Brief Discussion about the Topic – **An overview about installation of Hadoop**

Installing Hadoop in Pseudo Distributed Mode

Follow the steps given below to install Hadoop 2.4.1 in pseudo distributed mode.

Step 1 – Setting Up Hadoop

You can set Hadoop environment variables by appending the following commands to `~/.bashrc` file.

```
export HADOOP_HOME=/usr/local/hadoop
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME

export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
export HADOOP_INSTALL=$HADOOP_HOME
```

Now apply all the changes into the current running system.

```
$ source ~/.bashrc
```

Step 2 – Hadoop Configuration

You can find all the Hadoop configuration files in the location “`$HADOOP_HOME/etc/hadoop`”. It is required to make changes in those configuration files according to your Hadoop infrastructure.

```
$ cd $HADOOP_HOME/etc/hadoop
```

In order to develop Hadoop programs in java, you have to reset the java environment variables in **hadoop-env.sh** file by replacing **JAVA_HOME** value with the location of java in your system.

```
export JAVA_HOME=/usr/local/jdk1.7.0_71
```

The following are the list of files that you have to edit to configure Hadoop.

core-site.xml

The **core-site.xml** file contains information such as the port number used for Hadoop instance, memory allocated for the file system, memory limit for storing the data, and size of Read/Write buffers.

Open the core-site.xml and add the following properties in between <configuration>, </configuration> tags.

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

hdfs-site.xml

The **hdfs-site.xml** file contains information such as the value of replication data, namenode path, and datanode paths of your local file systems. It means the place where you want to store the Hadoop infrastructure.

Let us assume the following data.

```
dfs.replication (data replication value) = 1

(In the below given path /hadoop/ is the user name.
hadoopinfra/hdfs/namenode is the directory created by hdfs file
system.)
namenode path = //home/hadoop/hadoopinfra/hdfs/namenode

(hadoopinfra/hdfs/datanode is the directory created by hdfs
file system.)
datanode path = //home/hadoop/hadoopinfra/hdfs/datanode
```

Open this file and add the following properties in between the <configuration> </configuration> tags in this file.

```
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>

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

```
<value>file:///home/hadoop/hadoopinfra/hdfs/namenode
</value>
</property>

<property>
  <name>dfs.data.dir</name>
  <value>file:///home/hadoop/hadoopinfra/hdfs/datanode
</value>
</property>
</configuration>
```

Note – In the above file, all the property values are user-defined and you can make changes according to your Hadoop infrastructure.

yarn-site.xml

This file is used to configure yarn into Hadoop. Open the yarn-site.xml file and add the following properties in between the <configuration>, </configuration> tags in this file.

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

mapred-site.xml

This file is used to specify which MapReduce framework we are using. By default, Hadoop contains a template of yarn-site.xml. First of all, it is required to copy the file from **mapred-site.xml.template** to **mapred-site.xml** file using the following command.

```
$ cp mapred-site.xml.template mapred-site.xml
```

Open **mapred-site.xml** file and add the following properties in between the <configuration>, </configuration> tags in this file.

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

References-

1. **Book:** Hadoop: The Definitive Guide by Tom White, 3rd Edition, O'reilly Hadoop in Action by Chuck Lam, MANNING Publications
2. **Online:** <https://www.tutorialspoint.com/>
3. **Online:** <http://www.oracle.com>

QUESTIONS: -

Q1. What is pre-requisite of installing HADOOP in pseudo distributed mode?

Q2. Write steps for setup of HADOOP in Pseudo Distributed Mode.

Next, we will discuss about Verifying Hadoop installation.

- Academic Day ends with-
National song 'Vande Mataram'