



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.: 4

Session Name- HADOOP – ENVIRONMENT SETUP

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- **Introduction**

Topic to be discussed today- Today We will discuss about - **HADOOP Environment Setup**

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

Introduction & Brief Discussion about the Topic - **An overview of Big Data and Hadoop**

HADOOP – ENVIRONMENT SETUP

Hadoop is supported by GNU/Linux platform and its flavors. Therefore, we have to install a Linux operating system for setting up Hadoop environment. In case you have an OS other than Linux, you can install a Virtualbox software in it and have Linux inside the Virtualbox.

Pre-installation Setup

Before installing Hadoop into the Linux environment, we need to set up Linux using ssh (Secure Shell). Follow the steps given below for setting up the Linux environment.

Creating a User

At the beginning, it is recommended to create a separate user for Hadoop to isolate Hadoop file system from Unix file system. Follow the steps given below to create a user:

- Open the root using the command “su”.
- Create a user from the root account using the command “useradd username”.
- Now you can open an existing user account using the command “su username”. Open the Linux terminal and type the following commands to create a user.

```
$ su
password
:
# useradd
hadoop# passwd
hadoop New
passwd: Retype
new passwd
```

SSH Setup and Key Generation

SSH setup is required to do different operations on a cluster such as starting, stopping, distributed daemon shell operations. To authenticate different users of Hadoop, it is required to provide public/private key pair for a Hadoop user and share it with different users.

The following commands are used for generating a key value pair using SSH. Copy the public keys from id_rsa.pub to authorized_keys, and provide the owner with read and write permissions to authorized_keys file respectively.

```
$ ssh-keygen -t rsa
$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
$ chmod 0600 ~/.ssh/authorized_keys
```

Installing Java

Java is the main prerequisite for Hadoop. First of all, you should verify the existence of java in your system using the command “java -version”. The syntax of java version command is given below.

```
$ java -version
```

If everything is in order, it will give you the following output.

```
java version "1.7.0_71"  
Java (TM) SE Runtime Environment (build 1.7.0_71-  
b13) Java Hotspot(TM) Client VM (build 25.0-b02,  
mixed mode)
```

If java is not installed in your system, then follow the steps given below for installing java.

Step 1

Download java (JDK <latest version> - X64.tar.gz) by visiting the following link

<http://www.oracle.com/technetwork/java/javase/downloads/jdk7-downloads-1880260.html>.

Then **jdk-7u71-linux-x64.tar.gz** will be downloaded into your system.

Step 2

Generally, you will find the downloaded java file in Downloads folder. Verify it and extract the **jdk-7u71-linux-x64.gz** file using the following commands.

```
$ cd Downloads/  
$ ls  
jdk-7u71-linux-x64.gz  
  
$ tar zxf jdk-7u71-linux-x64.gz  
$ ls  
jdk1.7.0_71  jdk-7u71-linux-x64.gz
```

Step 3

To make java available to all the users, you have to move it to the location “/usr/local/”. Open root, and type the following commands.

```
$ su
password
:
# mv jdk1.7.0_71
/usr/local/# exit
```

Step 4

For setting up PATH and JAVA_HOME variables, add the following commands to ~/.bashrc file.

```
export
JAVA_HOME=/usr/local/jdk1.7.0_71
export PATH=PATH:$JAVA_HOME/bin
```

Now apply all the changes into the current running system.

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

Step 5

Use the following commands to configure java alternatives:

```
# alternatives --install /usr/bin/java java usr/local/java/bin/java 2
# alternatives --install /usr/bin/javac javac
usr/local/java/bin/javac 2# alternatives --install /usr/bin/jar
jar usr/local/java/bin/jar 2

# alternatives --set java usr/local/java/bin/java
# alternatives --set javac
usr/local/java/bin/javac# alternatives --set
jar usr/local/java/bin/jar
```

Now verify the installation using the command `java -version` from the terminal as explained above.

Downloading Hadoop

Download and extract Hadoop 2.4.1 from Apache software foundation using the following commands.

```
$ su
```

```
password:  
# cd /usr/local  
# wget http://apache.claz.org/hadoop/common/hadoop-  
2.4.1/hadoop-2.4.1.tar.gz  
# tar xzf hadoop-  
2.4.1.tar.gz # mv hadoop-  
2.4.1/* to hadoop/# exit
```

Reference-

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?

Q2. How to create a user for HADOOP?

Q3. Write steps for setup of HADOOP.

Next, we will discuss about HADOOP Operation Modes.

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