**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.**: 6

 **Session Name-** Installing Hadoop in Standalone 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 Operation Modes**

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

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

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

# Installing Hadoop in Standalone Mode

Here we will discuss the installation of Hadoop 2.4.1 in standalone mode.

There are no daemons running and everything runs in a single JVM. Standalone mode is suitable for running MapReduce programs during development, since it is easy to test and debug them.

**Setting Up Hadoop**

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

```
export HADOOP_HOME=/usr/local/hadoop
```

Before proceeding further, you need to make sure that Hadoop is working fine. Just issue the following command:

```
$ hadoop version
```

If everything is fine with your setup, then you should see the following result:

```
Hadoop 2.4.1
Subversion https://svn.apache.org/repos/asf/hadoop/common -r
1529768Compiled by hortonmu on 2013-10-07T06:28Z
Compiled with protoc 2.5.0
From source with checksum 79e53ce7994d1628b240f09af91e1af4
```

It means your Hadoop's standalone mode setup is working fine. By default, Hadoop is configured to run in a non-distributed mode on a single machine.

**Example**

Let's check a simple example of Hadoop. Hadoop installation delivers the following example MapReduce jar file, which provides basic functionality of MapReduce and can be used for calculating, like Pi value, word counts in a given list of files, etc.

```
$HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.2.0.jar
```

Let's have an input directory where we will push a few files and our requirement is to count the total number of words in those files. To calculate the total number of words, we do not need to write our MapReduce, provided the .jar file contains the implementation for word count. You can try other examples using the same .jar file; just issue the following commands to check supported MapReduce functional programs by *hadoop-mapreduce-examples- 2.2.0.jar* file.

```
$ hadoop jar $HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-
examples-2.2.0.jar
```

**Step 1**

Create temporary content files in the input directory. You can create this input directory anywhere you would like to work.

```
$ mkdir input
$ cp $HADOOP_HOME/*.txt input
$ ls -l input
```

It will give the following files in your input directory:

```
total 24
-rw-r--r-- 1 root root 15164 Feb 21 10:14 LICENSE.txt
-rw-r--r-- 1 root root 101 Feb 21 10:14 NOTICE.txt
-rw-r--r-- 1 root root  1366 Feb 21 10:14 README.txt
```

These files have been copied from the Hadoop installation home directory. For your experiment, you can have different and large sets of files.

**Step 2**

Let's start the Hadoop process to count the total number of words in all the files available in the input directory, as follows:

```
$ hadoop jar $HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-
examples-2.2.0.jar wordcount input ouput
```

**Step 3**

Step-2 will do the required processing and save the output in output/part-r00000 file, which you can check by using –

```
$cat output/*
```

It will list down all the words along with their total counts available in all the files available in the input directory.

```
"AS           4
"Contribution" 1
"Contributor" 1
"Derivative 1
"Legal 1
"License"       1
"License");     1
"Licensor"      1
"NOTICE"        1
"Not        1
"Object"        1
"Source"        1
"Work"     1
"You"      1
"Your")    1
"[]"       1
"control"       1
"printed        1
"submitted"     1
(50%)      1
(BIS),     1
(C)        1
(Don't)    1
(ECCN)     1
(INCLUDING      2
(INCLUDING,     2
.............
```

# 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?**

**Q2. Write steps for setup of HADOOP in Standalone Mode.**

Next, we will discuss about Installing Hadoop in Pseudo Distributed Mode

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