**Setting up Hadoop for Lab**

1. Open a terminal and check the java version in your system

Command: **java --version**

1. Check if the java version is >=8 if not then install java version 8 or more
2. Download the tar gz file for Hadoop using the command

**wget http://apache.mirrors.lucidnetworks.net/hadoop/common/hadoop-3.2.2/hadoop-3.2.2.tar.gz**If the download speed is taking more time you can manually download the file from <http://apache.mirrors.lucidnetworks.net/hadoop/common/>

1. Once the file gets downloaded open the directory where the file is downloaded, if the file is downloaded using a terminal then it will be present in the same directory.
2. Unzip the file **tar xvzf filename**[eg: hadoop-2.7.0.tar.gz]
3. Rename the file to hadoop2 using the command: **mv filename hadoop2**
4. Setting Environment variables:
   1. Open the directory **/hadoop2** in terminal
   2. Go to directory etc/hadoop/ command: **cd etc/hadoop/**
   3. Open hadoop-env.sh using vim

vi hadoop-env.sh

* 1. Press insert to modify the file and then navigate to line

**export JAVA\_HOME=${....}**

If the line is commented (has ‘#’ before it) then uncomment it by removing ‘#’.

* 1. Modify the line **export JAVA\_HOME=/usr/lib/jvm/default-java**
  2. Press **:** (colan), type wq and press enter. This command closes the file

1. Let us verify if the installation is successful or not
   1. Again navigate to **/hadoop2**
   2. Run the command: **bin/hadoop** ( running this command should prompt you with various options). This indicates hadoop is properly installed

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

Running sample hadoop program provided in the package downloaded.

1. Make directory input command: **mkdir input**
2. Run the command to copy all the xml files, command: **cp etc/hadoop/\*.xml input**
3. Now lets execute simple wordCount program from examples,

**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)

1. Finally run the command: **cat output/\*** (look for the output in the output directory that Hadoop creates for you)