# SSN COLLEGE OF ENGINEERING
# Department of Computer Science and Engineering
# CS8711 Cloud Computing Laboratory

## Assignment - 5 : Installation of Single Node Hadoop and Executing Word Count Program

**Assigned Date: 29.09.2020.** **Due Date: 07.10.2020 & 08.10.2020**

---

**I. Pre-requisites**

1. Operating System: Ubuntu 16.04 LTS Desktop (64-Bit only) OS
2. Installation Mode : Install in Guest OS (**Note**: Virtual Machine can be slower to work when Hadoop Cluster starts)
3. Java: jdk 1.8
4. Download latest version of Apache Hadoop package
5. Eclipse Luna 64-bit for Linux (Can be downloaded from SSN Intranet – Tech Support). Install Eclipse. (If Necessary)
6. Create a Virtual Machine and install Ubuntu 16.04 Desktop amd64.iso in VM. Name the VM as HadoopVM and set network configuration.

**Please refer Manual for detailed installation and execution steps.**

**Step 1**: Install Java 8 and verify that it is working.
]$ **java -version**
]$ **apt-get install default-jdk**
]$ **apt-get install default-jre**

7. Download the JDK1.8 tar.gz file from the following URL
**http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html**
Download: Linux x64 Compressed Archive (jdk-8u261-linux-x64.tar.gz)

8. Download latest version of Apache Hadoop package
https://hadoop.apache.org/releases.html
https://www.apache.org/dyn/closer.cgi/hadoop/common/hadoop-3.1.4/hadoop-3.1.4.tar.gz

II. Create user named as hduser and add hduser in a group named Hadoop.
Login as hduser using ]$ su – hduser command

III. Generate ssh key-pairs and move id_rsa.pub key to authorized_keys

IV. Install openssh-server and openssh-client packages using root privilege.

V. Download Hadoop and Java packages from internet and move them to the path /usr/local/hadoop

VI. Install Java and set path in ~/.bashrc file.
(Take atmost care while making changes in ~/.bashrc file)
To check java version: ]$ java –version

VII. Install Hadoop and set path in ~/.bashrc file
To check Hadoop version: ]$ hadoop version

VIII. Make changes in following Configuration files.
1. core-site.xml
2. yarn-site.xml
3. mapred-site.xml
4. hdfs-site.xml

IX. Login to hadoop / hduser user. Format the namenode

X. Start all Hadoop service.

XI. Check the running services in Web Interface.

XII. Managing files in HDFS. Follow below link.

Create an input file with few sentences in it. Upload input file into HDFS using put command / copyFromLocal command
***put*** *command to store file in HDFS,* ***get*** *command to read / retrieve file from HDFS.*

hduser ]$ /usr/local/hadoop/bin/hadoop dfs –copyFromLocal /tmp/MapReduceInput /user/hduser/MapReduceInput

hduser ]$ /usr/local/hadoop/bin/hadoop dfs –ls /tmp/MapReduceInput /user/hduser/MapReduceInput

https://hadoop.apache.org/docs/r2.4.1/hadoop-project-dist/hadoop-common/FileSystemShell.html

XIII. Write word count program in Java using Map and Reduce functions. Create word count program into a jar file using Eclipse.
Follow link in reference section. Either download wordcount.jar file and execute it or use Eclipse IDE to export java program into a .jar file.

https://hadoop.apache.org/docs/stable/hadoop-mapreduce-client/hadoop-mapreduce-client-core/MapReduceTutorial.html#Example:_WordCount_v2.0

http://hortonworks.com/hadoop-tutorial/using-commandline-manage-files-hdfs/

XIV. Execute the word count program's jar file using below command in hduser.

hduser]$ ls /usr/local/hadoop/

hduser ]$ /usr/local/hadoop/bin/hadoop jar /usr/local/Hadoop/Hadoop-examples-2.7.1.jar wordcount /user/hduser/MapReduceInput /user/hduser/MapReduce.output

OR

hadoop jar wordcount.jar /usr/local/hadoop/input /usr/local/hadoop/output

hduser ]$ /usr/local/hadoop/bin/hadoop dfs –ls /user/hduser
hduser ]$ /usr/local/hadoop/bin/hadoop dfs –ls /user/hduser/ MapReduce.output
hduser ]$ /usr/local/hadoop/bin/hadoop dfs –cat /user/hduser/ MapReduce.output/part-r-00000

Finally stop all Hadoop services.