**Instruction Guide**

**JAVA**

1. Verify whether you have JAVA installed on your system using the following command.

**$ javac --version**

1. If not, to install JDK in Linux, use the following command.

**$ sudo apt install default-jdk**

1. Then, the JRE File of Java will be installed using the following command.

**$ sudo apt install default-jre**

1. To verify the installation, the following command you can use. It will prompt the Java version used there.

**$ javac --version**

1. Create a new folder named <1MS23SCS/SCN\*\*> and navigate inside the folder using the following command.

**$ cd <1MS23SCS/SCN\*\*>**

**Hadoop**

1. Now download the hadoop tar file to install using the given url.
2. Once you have downloaded hadoop-3.2.2.tar.gz, extract this file with the below command (make sure to check your tar filename).

**$ tar xvzf hadoop-3.2.2.tar.gz**

1. Now navigate inside the folder using the below command.

**$ cd hadoop-3.2.2/**

1. Create and open a new ***bash.sh*** file inside the directory.

**$ gedit bash.sh**

1. We configure the file, copy the below command inside this file and save it.

**export JAVA\_HOME=$(readlink -f $(which javac) | awk 'BEGIN {FS="/bin"} {print $1}')**

**export PATH=$(echo $PATH):$(pwd)/bin**

**export CLASSPATH=$(hadoop classpath)**

1. Execute the bash.sh File using following command

**$ source bash.sh**

1. Verify ***JAVA\_HOME*** variable to be set to Java Path and ***PATH*** variable has your Hadoop Folder.
2. Verify Hadoop is Installed or not by executing hadoop command. If command gives Information about Hadoop command, then Hadoop is Successfully Installed.

**$ hadoop**

**Spark**

1. Verify that hadoop is installed and running in your system.

**$ hadoop**

1. Download Apache Spark from the given url inside your folder.
2. Now we extract this tar file with the help of below command (make sure to check your tar filename).

**$ tar xvzf spark-3.5.2-bin-hadoop3.tgz**

1. Now navigate inside the folder using the below command.

**$ cd spark-3.5.2-bin-hadoop3/**

1. Create a new file named bash.sh inside your folder by using any text editor.

**$ gedit bash.sh**

1. Copy below code and paste inside bash.sh file

**export PATH=$(echo $PATH):$(pwd)/bin**

1. In terminal, execute bash.sh file using the following command.

**$ source bash.sh**

1. Verify spark version with the below command.

**$ spark-shell --version**

**PIG**

1. Verify that hadoop is installed and running in your system.

**$ hadoop**

1. Download Apache Pig from the given url inside your folder.
2. Now we extract this tar file with the help of below command (make sure to check your tar filename).

**$ tar -xvf pig-0.17.0.tar.gz**

1. Now navigate inside the folder using the below command.

**$ cd pig-0.17.0/**

1. Create and open a new bash.sh file inside the directory.

**$ gedit bash.sh**

1. We configure file, copy the below command inside this file and save it.

**export PIG\_INSTALL=$(pwd)**

**export PATH=$PATH:$(pwd)/bin**

1. Execute the bash.sh File using following command

**$ source bash.sh**

1. You can check your pig version with the below command.

**$ pig -version**

1. Once you get it correct that’s it we have successfully install pig to our Hadoop single node setup, now we start pig with below pig command.

**$ pig**