**(1) \*\*\*\*\*\*\*\*\*\*\*\*\* Spark commands installation\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***

**Spark commands installation**

Spark is Hadoop’s sub-project. Therefore, it is better to install Spark into a Linux based system. The following steps show how to install Apache Spark.

**Download the latest version of java**

http://openjdk.java.net/install/index.html

sudo apt-get install openjdk-8-jre

<http://www.scala-lang.org/download/>

**Verifying Java Installation**

Java installation is one of the mandatory things in installing Spark. Try the following command to verify the JAVA version.

$java -version

If Java is already, installed on your system, you get to see the following response –

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)

In case you do not have Java installed on your system, then Install Java before proceeding to next step.

**Edit Bashrc**

gedit ~/.bashrc

export JAVA\_HOME="/usr"

export PATH="$PATH:$JAVA\_HOME/bin"

source ~/.bashrc

**Downloading Scala and Installing Scala**

Follow the below given steps for installing Scala.

<http://downloads.lightbend.com/scala/2.12.0/scala-2.12.0.tgz>

**Extract the Scala tar file**

Type the following command for extracting the Scala tar file.

$ tar -xvf scala-2.11.6.tgz

**Move Scala software files**

Use the following commands for moving the Scala software files, to respective directory **(/usr/local/scala**).

$ su –

Password:

# cd /home/Hadoop/Downloads/

# mv scala-2.11.6 /usr/local/scala

# exit

**Set PATH for Scala**

Use the following command for setting PATH for Scala.

gedit ~/.bashrc

export SCALA\_HOME="/usr/local/scala"

export PATH="$PATH:$SCALA\_HOME/bin"

source ~/.bashrc

**Verifying Scala Installation**

After installation, it is better to verify it. Use the following command for verifying Scala installation.

$scala -version

If Scala is already installed on your system, you get to see the following response –

Scala code runner version 2.11.6 -- Copyright 2002-2013, LAMP/EPFL

**Step 5: Downloading Apache Spark**

Download the latest version of Spark by visiting the following link Download Spark. For this tutorial, we are using **spark-1.3.1-bin-hadoop2.6** version. After downloading it, you will find the Spark tar file in the download folder.

**Step 6: Installing Spark**

Follow the steps given below for installing Spark.

**Extracting Spark tar**

The following command for extracting the spark tar file.

$ tar xvf spark-1.3.1-bin-hadoop2.6.tgz

**Moving Spark software files**

The following commands for moving the Spark software files to respective directory

**(/usr/local/spark**).

$ su –

Password:

# cd /home/Hadoop/Downloads/

# mv spark-1.3.1-bin-hadoop2.6 /usr/local/spark

# exit

**Setting up the environment for Spark**

Add the following line to **~/.bashrc** file. It means adding the location, where the spark software file are located to the PATH variable.

Gedit ~/.bashrc

export SPARK\_HOME="/usr/local/spark"

export PATH="$PATH:$SPARK\_HOME/bin"

source ~/.bashrc

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***for standalone installation it’s not required**\*\*\*\*\*\*\*\*\*\*\*\*\*

cd conf  
cp spark-env.sh.template spark-env.sh  
/conf/spark-env.sh  
export SPARK\_MASTER\_IP=  
export SPARK\_WORKER\_CORES=1  
export SCALA\_HOME=$SCALA\_HOME  
export SPARK\_WORKER\_MEMORY=1g

export SPARK\_WORKER\_INSTANCES=2  
export SPARK\_WORKER\_DIR=/home/home/spark\_work\_Dir