This Project covers:

**Setting up of Spark on Desktop.**

**Running Spark-Shell in RPEL.**

**Running Spark through IDE using both SBT and Maven (Specially using Eclipse IDE but also similar apply to Intellij IDE)**

Help is available in the following links:

Help with setting spark and configuring it on Desktop:

Setup Spark on Windows – for development (<http://www.itversity.com/2016/10/17/setup-spark-on-windows-for-development/>)

Configuring of log4j.properties in conf of spark folder (<https://www.youtube.com/watch?v=4sO-VgqHLp4>)

Setup Spark in eclipse (Scala IDE) by building using Maven (<http://www.devinline.com/2016/01/apache-spark-setup-in-eclipse-scala-ide.html>)

Setup Scala IDE for Eclipse and Integrate SBT with Scala IDE for Eclipse (<http://www.itversity.com/topic/setup-scala-ide-for-eclipse/>)

**Setting up of Spark on Desktop:**

They are two ways to install/run Spark on Windows

One is running spark-shell in RPEL on windows through spark-shell (this is quick to set-up but tedious while developing)

Other is running spark through IDE’s (It is tedious, but recommended to have IDE such Eclipse, IntelliJ etc...)

**Running Spark-Shell in RPEL (Installing Spark on Desktop):**

Go to <http://spark.apache.org/downloads.html> and download spark release along with Hadoop pre-build and correct Scala version (2.10)

Choose download type as Direct Download and then click the Download Spark link marked by arrow to download Spark pre build binary in download folder

Extracted the gzip and then the tar folder (twice) in C Drive.

Create SPARK\_HOME Environment Variable pointing to the extracted folder.

Validate Spark by running spark-shell (go to CMD C:\local\_hadoop\spark-1.6.0-bin-hadoop2.6\bin and run spark-shell)

HDFS:

Download winutils.exe (<https://github.com/steveloughran/winutils/raw/master/hadoop-2.6.0/bin/winutils.exe>)

Create the following folders C:\local\_hadoop\HDFS\_Spark\bin and save winutils.exe to the previous path.

Set HADOOP\_HOME to reflect the directory with an upper level to winutils i.e HADOOP\_HOME = C:\local\_hadoop\HDFS\_Spark

Append PATH environment variable to include %HADOOP\_HOME%\bin

HADOOP\_HOME is new Environment Variable and PATH needs to be appended and both are System Variables

For going to Spark UI open <http://localhost:4040/> in browser

To test run run-example SparkPi 10 using CMD in the path C:\local\_hadoop\spark-1.6.0-bin-hadoop2.6\bin

To reduce logs:-

Goto C:\local\_hadoop\spark-1.6.0-bin-hadoop2.6\conf

Copy log4j.properties.template and paste in the same location and rename it to log4j.properties.

Change INFO to WARN i.e log4j.root category = INFO, console to log4j.root category = WARN, console

Setting this to WARN will reduce the log info we get on command line. In command line you can only see warnings and errors messages.

**Running Spark through IDE:**

Preparing Eclipse IDE for Spark Scala project:

Help->install new software->check for Available software->copy m2e releases repository link->browse for it available software-> Next

Go to help->check for updates (install updates and restart)

Go to help->eclipse market place-> find-> scala-> install scala->next-> next-> accept-> finish (restart).

**Create Maven Project:**

Create Maven Project by selecting a simple type.

Rename src/main/java and src/test/java to scala

Project root-> Configure-> Add Scala nature

Change Scala library container to 2.10 from 2.11 in properties (or) Project root-> properties-> scala complier-> use project settings-> select the wanted versions.

Configure pom.xml for scala-library, hadoop-common, spark-core, spark-streaming dependencies via mvnrepository.com.

To format xml in pom.xml go to Source-> Format. Use Builds providing by devinline.com.

To remove additional scala library container from build path (has jars required is already added via spark core (via pom.xml)). Right click on the project root-> Build path-> configure build path-> remove scala library container.

Add code in src/main/scala/……

Right click on main project-> Run as-> Scala application and check output.

To create jar file:-

Project-> Export-> Java (jar) -> Select project and give destination-> Finish

**Integrate sbt with Scala IDE for Eclipse:**

Installing SBT:

Install SBT.

Go to C:\Users\Ranganadakishore\.sbt\0.13 folder in home directory and create directory named “plugins”

Create a file plugin.sbt and add addSbtPlugin("com.typesafe.sbteclipse" % "sbteclipse-plugin" % "4.0.0") to it.

Building Project:

Create a root directory for the project

Mkdir –P /src/main/scala in root directory

Create and Configure build.sbt in root directory

CMD in root directory and run sbt eclipse. For project layout to be ready.

Use Eclipse IDE to import the project with source has root directory

Create the code in src/main/scala using the IDE

Run as scala application for running on Desktop

To complie for changes use sbt ~compile (or) delete all files except src, build.sbt and run sbt eclipse.

To run the program in sbt use sbt “run-main Main\_object\_name”

Use sbt package command for creation of jar file.

To run the application using jar file use sbt run jar\_file\_location Main\_object\_name