APACHE SPARK SETUP GUIDELINE

September 19th, 2015

Professor: **Prem Nair**
Student Name: **Bao Pham**
Student ID: **984588**

Install Apache Spark

1. Download Apache Spark “pre-built for hadoop 2.6 or later” at https://spark.apache.org/downloads.html

2. Extract the “spark-1.5.0-bin-hadoop2.6.tar” file into the “spark-1.5.0-bin-hadoop2.6” folder
3. Then Apache Spark is ready for usage

Install Maven

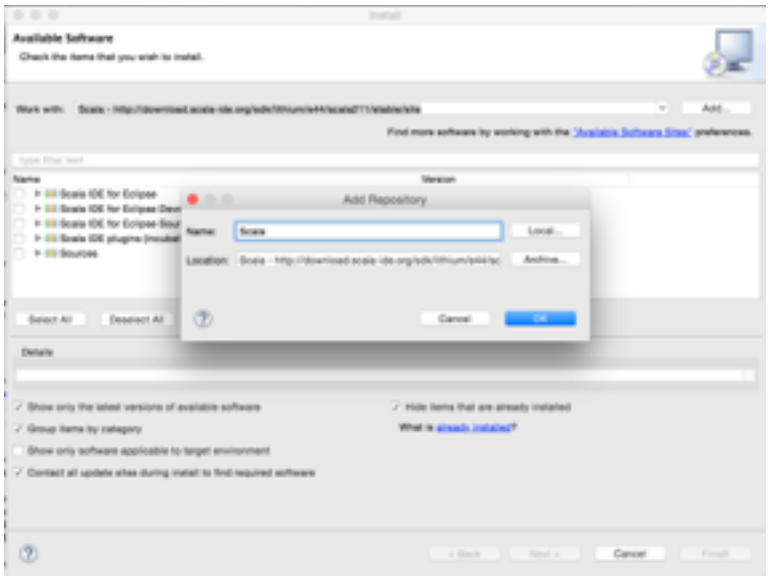
Start the terminal in Mac and then type *“brew install maven”* then maven is ready for usage

In the pom.xml of the project, configure the dependencies like below

```
<project xmlns="http://maven.apache.org/POM/4.0.0" xmlns:xsi="http://www.w3.org/2001/
XMLSchema-instance" xsi:schemaLocation="http://maven.apache.org/POM/4.0.0 http://
maven.apache.org/xsd/maven-4.0.0.xsd">
  <modelVersion>4.0.0</modelVersion>
  <groupId>bigdata.project.spark.maximumvote</groupId>
  <artifactId>MaxVote</artifactId>
  <version>0.0.1-SNAPSHOT</version>
  <dependencies>
        <dependency>
                <groupId>org.apache.spark</groupId>
                <artifactId>spark-core_2.10</artifactId>
                <version>1.5.0</version>
        </dependency>
  </dependencies>
</project>
```

Install Scala in eclipse

Select Help -> Install New Software.., then update Scala at this site http://download.scala-ide.org/sdk/lithium/e44/scala211/stable/site



Then click “Next” to install it

Project description

Assume that there are the list of candidates with the number of votes. Calculate the maximum vote that a candidate has

Run project

1. Create project MaxVote in eclipse
2. Create the jar file by using maven command “mvn package”, the jar file will be located at <MaxVote>/target/wordcount-0.0.1-SNAPSHOT.jar
3. Execute the jar file on Apache Spark locally by using the below command

```
./bin/spark-submit --class bigdata.project.spark.MaxVote --master local[2] /Users/baopham/Documents/workspace/MaxVote/target/MaxVote-0.0.1-SNAPSHOT.jar /Users/baopham/Documents/workspace/MaxVote/data/input.txt /Users/baopham/Documents/workspace/MaxVote/output
```

4. After executing successfully, the result is put at /Users/baopham/Documents/workspace/MaxVote/output/part-00000

Test Data

```
Bao 20
Kaylee 20
Bao 15
Tori 14
Xuan 2
```

```
Xuan 5
Loan 10
Hue 200
Loan 100
Hue 100
```

Data output

```
(Loan,100)
(Kaylee,20)
(Hue,200)
(Tori,14)
(Xuan,5)
(Bao,20)
```