SE256 : Scalable Systems for Data Science

Lab Session: 2

Maven setup:

Run the following commands to download and extract maven.

wget http://www.eu.apache.org/dist/maven/maven-3/3.3.9/binaries/apache-maven-3.3.9-bin.tar.gz

tar -xvf apache-maven-3.3.9-bin.tar.gz

Update proxy settings in the config file present in /apache-maven-3.3.9/conf/setting.xml Make sure that you uncomment proxy section only if you are using college wifi. Below is snippet of the xml file.

<!-- proxies

| This is a list of proxies which can be used on this machine to connect to the network.| Unless otherwise specified (by system property or command-line switch), the first proxy| specification in this list marked as active will be used.

|-->

<proxies>

<!-- proxy

| Specification for one proxy, to be used in connecting to the network.

<proxy>

<id>optional</id>

- <active>true</active>
- <protocol>http</protocol>

<username>proxyuser</username> /* required only for wifi connection*/
 <password>proxypass</password> /* required only for wifi connection*/
 <host>proxy.iisc.ernet.in</host>
 <port>3128</port>
 <nonProxyHosts>local.net|some.host.com</nonProxyHosts>
 </proxy>

</proxies>

Add the following lines to bashrc file

export M2_HOME=<Path to extracted folder>/apache-maven-3.3.9/ export PATH=\$M2_HOME/bin:\$PATH

Run the following command source ~/.bashrc

Install **eclipse** from their website.

First set proxy in eclipse (only for HTTP and HTTPS) and then edit /etc/eclipse.ini as per following page

find / -name eclipse.ini /*command to search for eclipse.ini file*/

Append the following lines in the file eclipse.ini -Dhttp.proxyPort=3128 -Dhttp.proxyHost=proxy.iisc.ernet.in -Dhttp.nonProxyHosts=localhost|127.0.0.1

Additionally you can refer the following link for understanding how to configure proxy settings on Eclipse.

http://stackoverflow.com/questions/5857499/how-do-i-have-to-configure-the-proxy-settings-so-e clipse-can-download-new-plugin

Under windows->Preferences->General-> Network add the proxy settings as follows:

| | | | | | Preferences |
|--|------------------|------------------|----------|----------|-------------|
| type filter text | Network Con | nections | | | ⇔ < |
| General Appearance Compare/Patch | Active Provider: | Native Y | | | |
| Content Types | Schen Ho | ost Port | Provic A | uth User | Password |
| + Editors | 🔲 НТТР | | Manua N | 0 | |
| Error Reporting | HTTPS | | Manua N | 0 | |
| Globalization | SOCKS | | Manua N | 0 | |
| Keys | HTTP pr | roxy.serc.i 3128 | Native N | 0 | |
| Network Connections | MTTPS pr | roxy.serc.i 3128 | Native N | 0 | |
| Notifications | | | | | |
| Perspectives | | | | | |
| Search | | | | | |
| + Security | | | | | |
| \pm Startup and Shutdow | Proxy bypass | | | | |

then in eclipse install new software use link

In eclipse, go to **Help->Install New Software**, use the link given below to integrate maven in eclipse.

http://download.eclipse.org/technology/m2e/releases/1.0

| Check the i | tems that you wish to install. | | |
|------------------------|---|---|----------------------------|
| Work with: | maven - http://download.eclipse.org/technology/m2e/re | leases/1.0 ¥ | Add |
| | | Find more software by working with the <u>"Available Software S</u> | <u>Sites"</u> preferences. |
| type filter te | ext | | C |
| Name | | Version | |
| = 🖾 Ш м | laven Integration for Eclipse | | |
| P | m2e - Maven Integration for Eclipse | 1.0.200.20111228-1245 | |
| A | m2e - slf4j over logback logging (Optional) | 1.0.200.20111228-1245 | |
| Details Maven Integ | gration for Eclipse 1.0.0.4-EAUcLS4nh7E7AAAEEc3AEMQ | | More |
| Show or | nly the latest versions of available software | Hide items that are already installed | |
| Group it | tems by category | What is <u>already installed</u> ? | |
| Show or | nly software applicable to target environment | | |
| Contact | all update sites during install to find required software | | |
| | | | |
| | | | |
| ? | | < Back Next > Cancel | Finish |

After performing the above step, go to **Window->Preferences**, expand maven from left menu, select Installations and add the path to your maven folder which you had extracted in the initial steps.

Within the maven Menu, select User Settings, now in the form opened, change user settings to point to the settings.xml file that we changed. Now apply the settings, it may take some time.

Refer to the following screenshot for this.

| /home/user/apache-maven-3.3.9/conf/settings.xml | | Browse. |
|---|--|---------|
| Update Settings | | |

After this, we can now import a maven project, Go to File->Import,

| | | Import | | + × |
|-----------------------------|---------------------------------------|------------------|--------|--------|
| Select Import Existing I | Maven Projects | | | ry. |
| Select an impor | t source: | | | |
| type filter text | | | | |
| 🗄 🗁 GIT | | | | |
| 🕀 🗁 Gradle | | | | |
| 🕀 🗁 Install | | | | |
| 🗆 🗁 Maven | | | | |
| 🗐 Check | out Maven Projects f | rom SCM | | |
| 👼 Existi | ng Maven Projects | | | |
| 🗟 Instal | l or <mark>deplo</mark> y an artifact | to a Maven repos | itory | |
| 🗐 Mater | ialize Maven Projects | from SCM | | |
| 🕀 🗁 Oomph | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| ? | < Back | Next > | Cancel | Finish |
| _ | | | | |

Now point to path of project that you want to import, it is to be noted that the path should contain a pom.xml file which is used by maven. Refer screenshot below for example.

| | Select Root Folde | er | × |
|---------------------|-------------------|--------|---------------|
| | se256 | | Create Folder |
| Places | Name | Size | Modified ^ |
| Q Search | 🚞 src | | 18:25 |
| € Recently Used | 🚞 target | | 18:25 |
| 📄 eclipse-installer | pom.xml | 1.6 kB | 18:25 |
| | | | |
| 📰 Desktop | | | |
| File System | | | |
| New Volume | | | |
| 🔄 33 GB Volume | | | |
| 21 GB Volume | • | | |
| 295 GB Volume | | | |
| New Volume | | | |
| 82 GB Volume | | | |
| 524 MB Volume | | | |
| + * | | | |
| | | Cancel | ОК |

Now before building, go to **Run->Run configurations**, select maven build from left menu and create a configuration by entering parameters as given below.

| | | Run C | onfigurations | i | | |
|---|--|---|--|-----------|--------------------|---------------------------|
| reate, manage, and run co | onfigurations | | | | | |
| Image: Second system Image: Second system Image: Second | Name: se256 Main Base directory /home/ Goals: Profiles: User settings: Parameter M | JRE Refresh Image: Constraint of the second se | Source En ache-maven-3.3 Update Sn Skip Tests bace artifacts | vironment | Browse File System | Variables File Edit |
| ter matched 7 of 7 items | | | | | Revert | Apply |
| Filter matched 7 of 7 items | | | | | Close | Run |

Click Run which builds the project by creating a jar file in the target folder.

Now we can use this jar file to run hadoop application, make sure all the processes are running.

\$jps

11899 DataNode27396 Jps11800 NameNode12033 ResourceManager12284 NodeManager

Now run the following commands:

Now we are going to run the wordCount example , we studied in last lab session

- 1. Create a text file which will be input to our program
- 2. Create a folder in hdfs
- 3. Put the input file inside this folder in hdfs

\$hadoop hdfs -mkdir /WordCountInput

\$hadoop hdfs -put input_file.txt /WordCountInput

From the source code provided, generate the jar file and run the following command:

\$HADOOP_HOME/bin/hadoop jar /home/user/se256/target/se256-1.0-jar-with-dependencies.jar in.dreamlab.iisc.se256.Driver /WordCountInput **/output**

where /WordCountInput is path to input folder in hdfs which has input text files, output is directory in hdfs where result is generated<should not be present before running the job>.