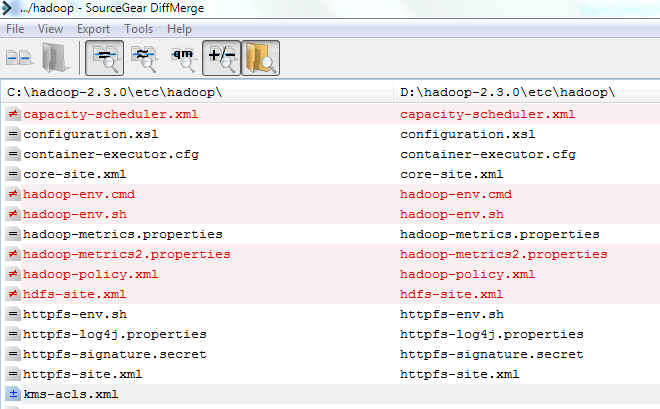
**Prerequisites**

1. Download, Install **Java 8** and Set Java Path.
   1. <http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>
   2. <http://download.oracle.com/otn-pub/java/jdk/8u92-b14/jdk-8u92-windows-x64.exe>
2. Download **Maven version 3.3.9** unzip and set the Path.
   1. <https://maven.apache.org/download.cgi>
   2. <http://mirror.nus.edu.sg/apache/maven/maven-3/3.3.9/binaries/apache-maven-3.3.9-bin.zip>
3. Download **eclipse JEE mars 2 (latest)** version with **m2e** plugin and unzip.
   1. <https://eclipse.org/downloads/>
   2. <https://eclipse.org/downloads/download.php?file=/technology/epp/downloads/release/mars/2/eclipse-jee-mars-2-win32-x86_64.zip>
4. **Download SourceGear Diffmerge and compare** Hadoop 2.3 version **with your code and make the necessary changes as mentioned below in Step 5.** [**https://sourcegear.com/diffmerge/downloaded.php**](https://sourcegear.com/diffmerge/downloaded.php)
5. Download and unzip **Hadoop 2.3** and set the path.
   1. <https://archive.apache.org/dist/hadoop/core/>
   2. <https://archive.apache.org/dist/hadoop/core/hadoop-2.3.0/hadoop-2.3.0.tar.gz>

**Note:** Unzip hadoop-2.3.0.tar.gz into c:\ and d:\ for compare with diffmerge and make necessary changes as below

**For ex:**



* 1. **D:\hadoop-2.3.0\etc\hadoop\hadoop-env.cmd**

set JAVA\_HOME=%JAVA\_HOME%

or

set JAVA\_HOME=C:/Progra~1/Java/jdk1.8.0\_92

5.2**) D:\hadoop-2.3.0\etc\hadoop\core-site.xml**

<configuration>

<property>

<name>fs.defaultFS</name>

<!-- <value>hdfs://127.0.0.1:9000</value> -->

<value>hdfs://localhost:9000</value>

</property>

</configuration>

**5.3) C:\hadoop-2.3.0\etc\hadoop\hdfs-site.xml**

<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

<name>dfs.namenode.name.dir</name>

<value>file:/hadoop/data/dfs/namenode</value>

</property>

<property>

<name>dfs.datanode.data.dir</name>

<value>file:/hadoop/data/dfs/datanode</value>

</property>

<property>

<name>dfs.permissions</name>

<value>false</value>

</property>

<property>

<name>dfs.http.address</name>

<!-- <value>127.0.0.1:50070</value> -->

<value>localhost:50070</value>

</property>

<property>

<name>dfs.webhdfs.enabled</name>

<value>true</value>

<description>to enable webhdfs</description>

<final>true</final>

</property>

</configuration>

5.4) **copy D:\hadoop-2.3.0\etc\hadoop\mapred-site.xml.template and rename mapred-site.xml.template -> mapred-site.xml D:\hadoop-2.3.0\etc\hadoop\mapred-site.xml**

**and add**

<configuration>

<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

</configuration>

5.5) **Edit and copy D:\hadoop-2.3.0\etc\hadoop\yarn-site.xml**

<configuration>

<property>

<name>yarn.nodemanager.aux-services</name>

<value>mapreduce\_shuffle</value>

</property>

<property>

<name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>

<value>org.apache.hadoop.mapred.ShuffleHandler</value>

</property>

<property>

<name>yarn.application.classpath</name>

<value>

%HADOOP\_HOME%\etc\hadoop,

%HADOOP\_HOME%\share\hadoop\common\\*,

%HADOOP\_HOME%\share\hadoop\common\lib\\*,

%HADOOP\_HOME%\share\hadoop\hdfs\\*,

%HADOOP\_HOME%\share\hadoop\hdfs\lib\\*,

%HADOOP\_HOME%\share\hadoop\mapreduce\\*,

%HADOOP\_HOME%\share\hadoop\mapreduce\lib\\*,

%HADOOP\_HOME%\share\hadoop\yarn\\*,

%HADOOP\_HOME%\share\hadoop\yarn\lib\\*

</value>

</property>

</configuration>

1. **Running the Hadoop Sample Programs with Maven and Eclipse.**

**Setting Java, Maven, Hadoop Path(s) in windows 7 64 bit systems**

Click on **Advanced system settings -> Environment Variables -> System Variables**

set JAVA\_HOME=C:\Program Files\Java\jdk1.8.0\_92 (Actual path modified to below)

set JAVA\_HOME=C:\Progra~1\Java\jdk1.8.0\_92 or

set JAVA\_HOME=C:/Progra~1/Java/jdk1.8.0\_92 (I prefer / (forward slash) )

set JAVA\_OPTS=-Xms512m -Xmx2048m

set MAVEN\_HOME=D:/sw/maven-3.3.9

set MAVEN\_OPTS=-Xms512m -Xmx2048m

set HADOOP\_HOME=D:/hadoop-2.3.0

set HADOOP\_OPTS=-Djava.library.path=%HADOOP\_HOME%/lib

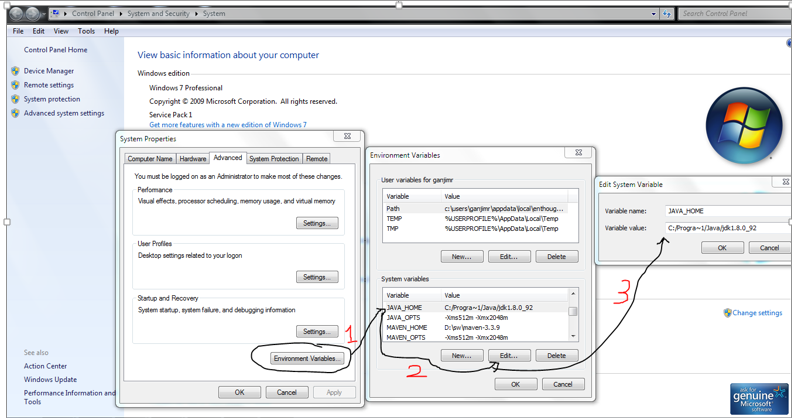
Finally edit PATH variable

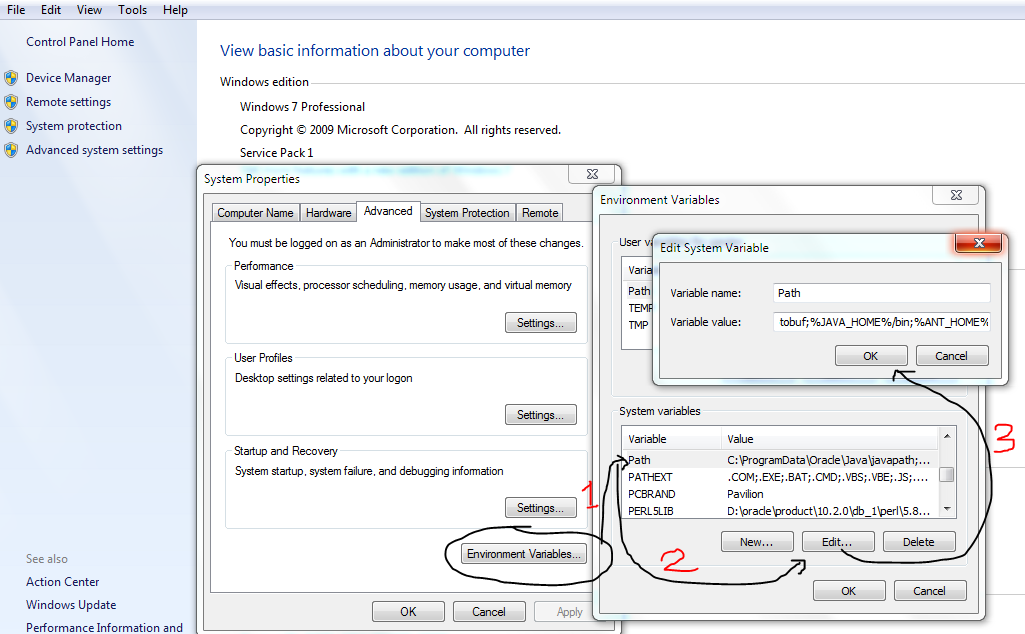
PATH=%PATH%;%JAVA\_HOME%/bin;%MAVEN\_HOME%/bin;%HADOOP\_HOME%/bin;.;

Note: =%PATH% and ;.; (at the end of the path) appended to PATH variable.

**This is my system/laptop windows 7 64 bit path variable entries:**

C:\ProgramData\Oracle\Java\javapath;C:\Progra~1\Common Files\Microsoft Shared\Windows Live;C:\Program Files (x86)\Common Files\Microsoft Shared\Windows Live;C:\Program Files (x86)\Windows Live\Shared;c:\Program Files (x86)\Intel\iCLS Client\;c:\Progra~1\Intel\iCLS Client\;C:\Progra~1\Broadcom\Broadcom 802.11;%SystemRoot%\system32;%SystemRoot%;%SystemRoot%\System32\Wbem;%SYSTEMROOT%\System32\WindowsPowerShell\v1.0\;c:\Progra~1\WIDCOMM\Bluetooth Software\;c:\Program Files\WIDCOMM\Bluetooth Software\syswow64;C:\Progra~1\Intel\Intel(R) Management Engine Components\DAL;C:\Program Files (x86)\Intel\Intel(R) Management Engine Components\DAL;C:\Progra~1\Intel\Intel(R) Management Engine Components\IPT;C:\Program Files (x86)\Intel\Intel(R) Management Engine Components\IPT;D:\oracle\product\10.2.0\db\_1\bin;C:\Progra~1\TortoiseSVN\bin;C:\cygwin64\bin**; ;**%ANT\_HOME%/bin;%GRADLE\_HOME%/bin;C:\Progra~1\Git\cmd**; %JAVA\_HOME%/bin;%MAVEN\_HOME%/bin;%HADOOP\_HOME%/bin**;.;

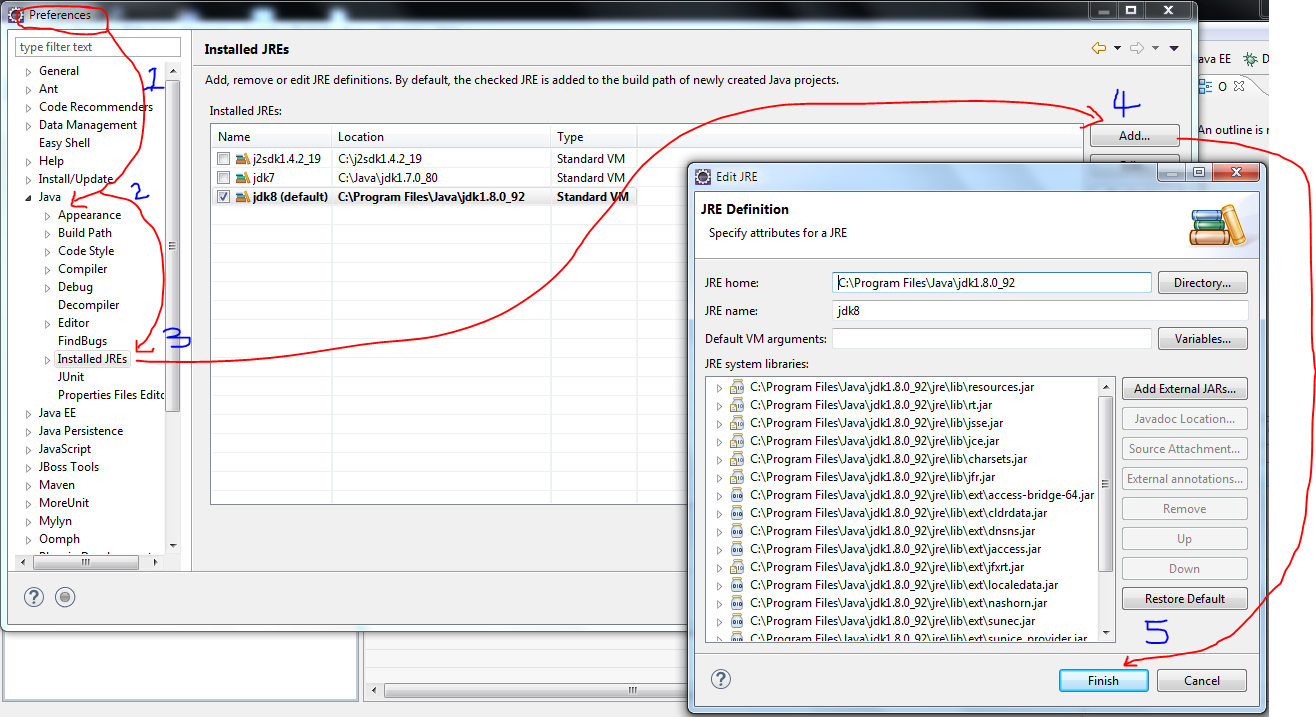




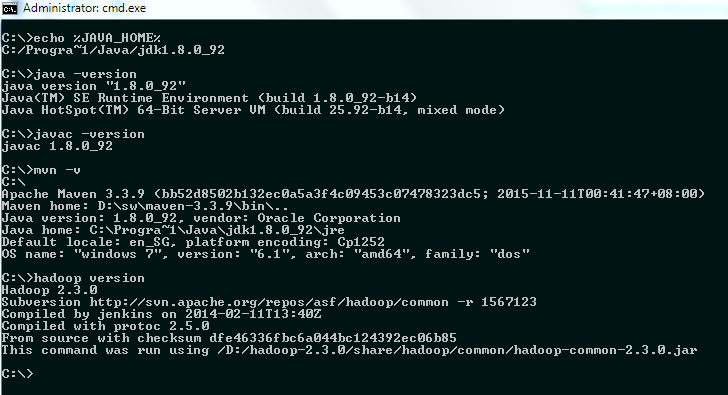
Open Command mode and check the java version

Open Eclipse (Mars 2)-> Window -> Preferences -> Java -> Installed JREs -> JRE home -> Give your Java Installation Path (**C:\Program Files\Java\jdk1.8.0\_92**)

JRE name **jdk1.8.0\_92** -> edited to **jdk8** for easy configuration in future if you want upgrade to another version of Java just change the above JDK path only useful when you choose run time environment etc for multiple projects.







C:\>echo %JAVA\_HOME%

C:/Progra~1/Java/jdk1.8.0\_92

C:\>java -version

java version "1.8.0\_92"

Java(TM) SE Runtime Environment (build 1.8.0\_92-b14)

Java HotSpot(TM) 64-Bit Server VM (build 25.92-b14, mixed mode)

C:\>javac -version

javac 1.8.0\_92

C:\>mvn -v

C:\

Apache Maven 3.3.9 (bb52d8502b132ec0a5a3f4c09453c07478323dc5; 2015-11-11T00:41:47+08:00)

Maven home: D:\sw\maven-3.3.9\bin\..

Java version: 1.8.0\_92, vendor: Oracle Corporation

Java home: C:\Progra~1\Java\jdk1.8.0\_92\jre

Default locale: en\_SG, platform encoding: Cp1252

OS name: "windows 7", version: "6.1", arch: "amd64", family: "dos"

C:\>hadoop version

Hadoop 2.3.0

Subversion http://svn.apache.org/repos/asf/hadoop/common -r 1567123

Compiled by jenkins on 2014-02-11T13:40Z

Compiled with protoc 2.5.0

From source with checksum dfe46336fbc6a044bc124392ec06b85

This command was run using /D:/hadoop-2.3.0/share/hadoop/common/hadoop-common-2.3.0.jar

C:\>

-- Too install the tools.jar to Maven local repository issue below command

mvn install:install-file -DgroupId=jdk.tools -DartifactId=jdk.tools -Dpackaging=jar -Dversion=1.8 -Dfile=${java.home}/../lib/tools.jar -DgeneratePom=true

Download the project from GitHub to **D:\github\**MapReduce directory

And import into eclipse

Final step before running MapReduce Program

**Step 7: Copy winutils.exe D:\hadoop-2.3.0\bin** directory

**Note: Step 1 to 7 are one time activities only.**

It will download all the required jar files into your respective C:\Users\<UserID-XXX**>**\.m2\repository directory.

