

Hadoop Ecosystem By Songera Group Setup Guide

Required Software:

First, install the following software:

- a. JDK 1.7 or higher, JDK 1.8 recommended
<http://www.oracle.com/technetwork/java/javase/downloads>
- b. Eclipse IDE for Java EE Developers 4.5.0 (Mar 2) or higher:
<https://www.eclipse.org/downloads>
- c. Python 2.7 (optional)
<https://www.python.org/downloads/>
- d. R (optional)
<https://www.r-project.org/>

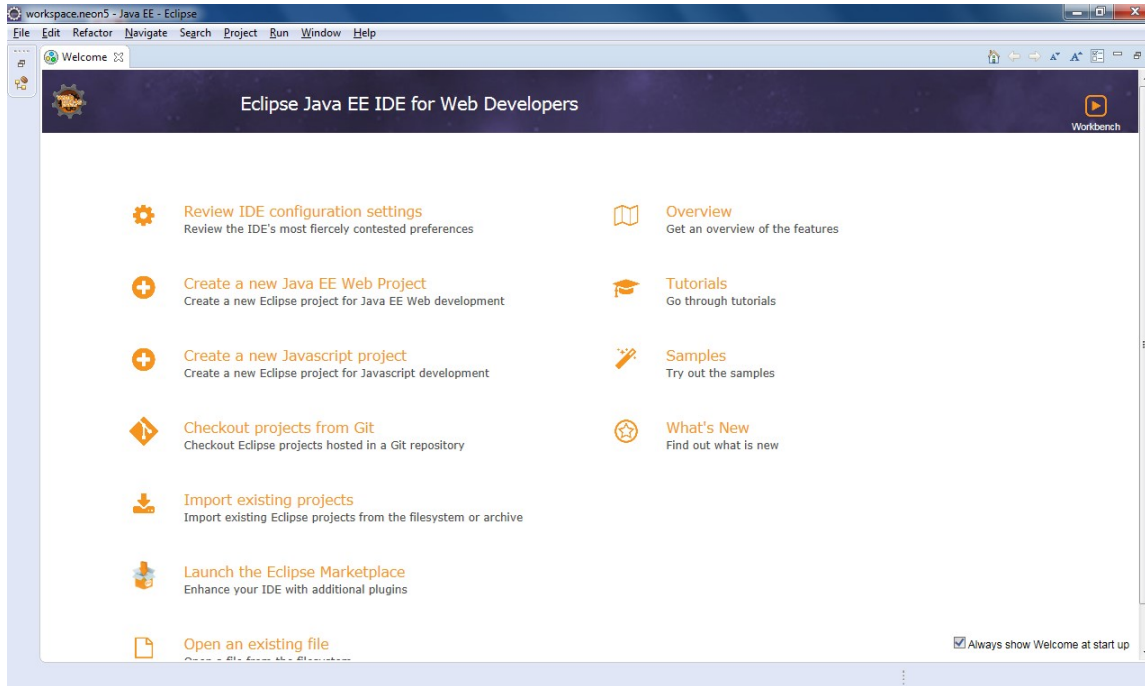
The following instruction will show you how to install and configure the Hadoop Ecosystem in Windows Platform. The installation process for Mac OS X and Linux are similar to that of Windows.

The Hadoop Ecosystem has two major steps to configure all Hadoop modules:

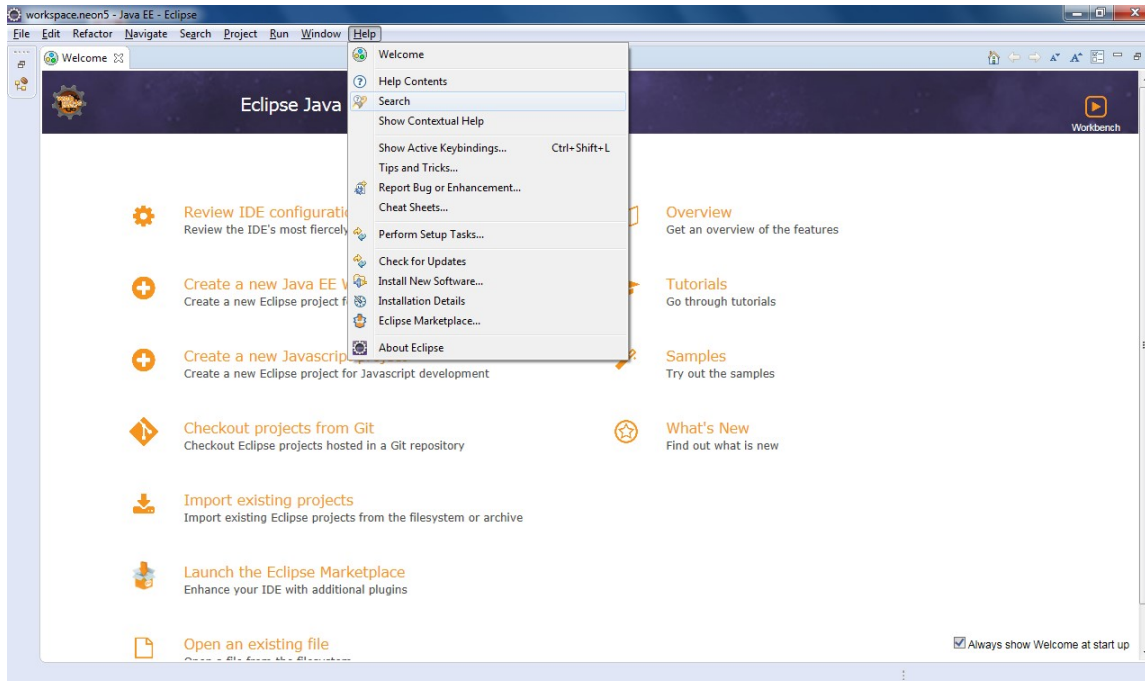
1. Runtime Environment – Install and deploy all Hadoop modules to hosts.
2. Server – Setup each module's configuration files. One Runtime instance allows the setup of multiple sets of server configurations for different developing or testing purposes.

1. Install “Hadoop Ecosystem By Songera Group” plugin.
Download plugin software from: site_assembly.zip at
<http://www.songeragroup.com/products>

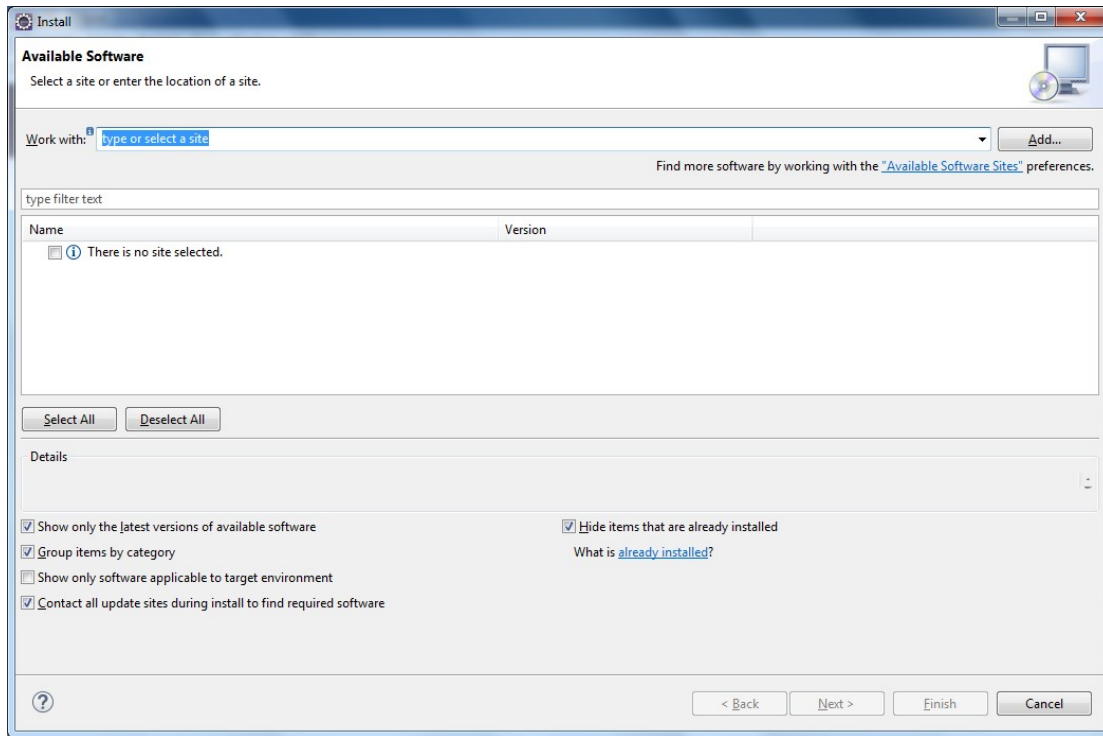
Start Eclipse:



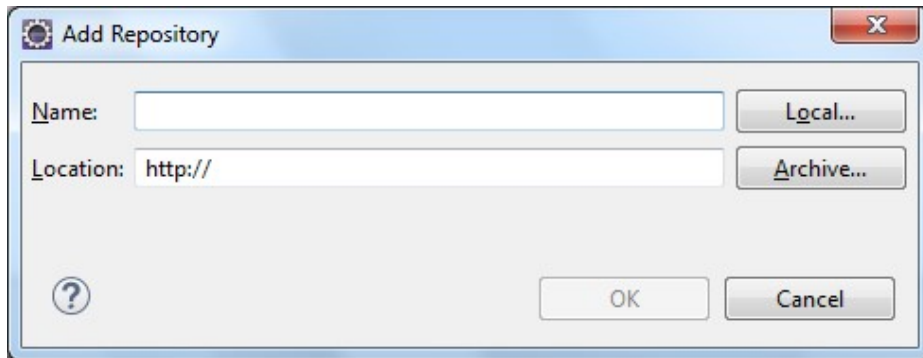
Launch plugin install window by going to Menu: Help -> Install New Software



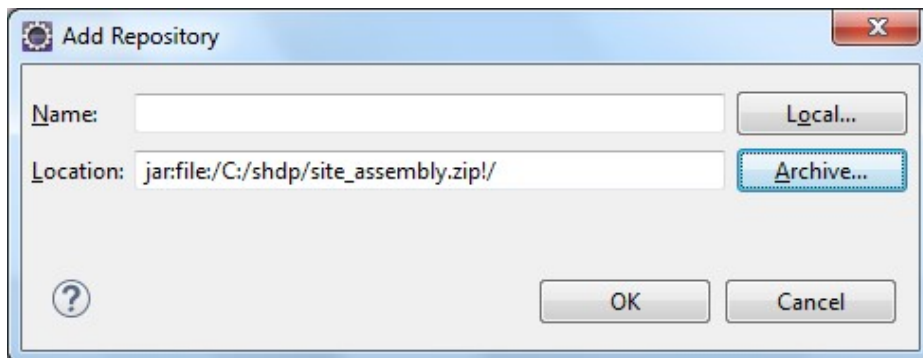
Click the Add Button in the Install window:



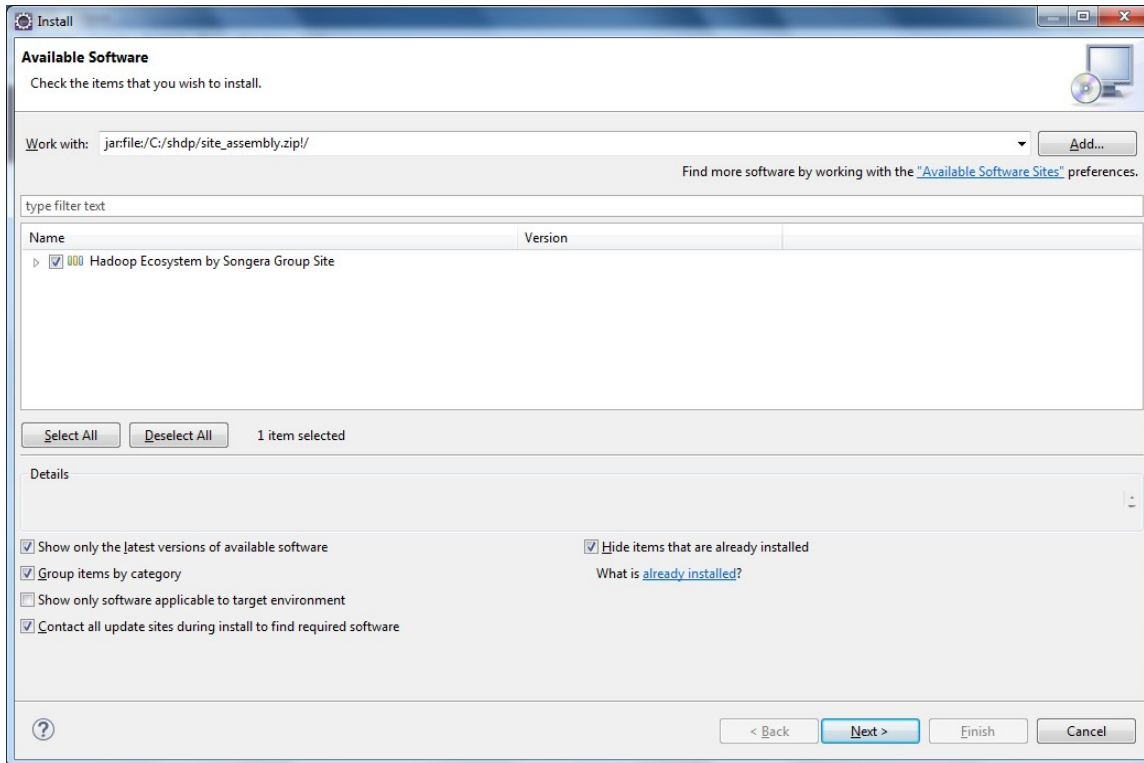
Click Archive button to select site_assembly.zip



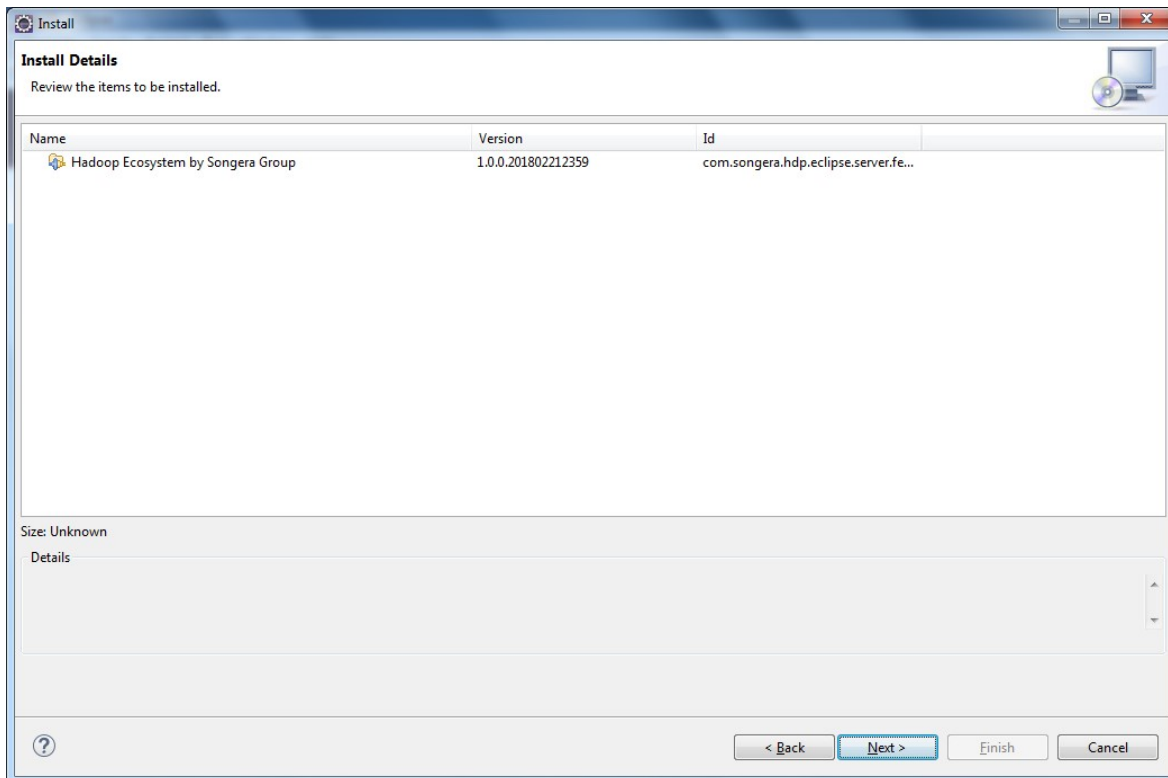
Click OK



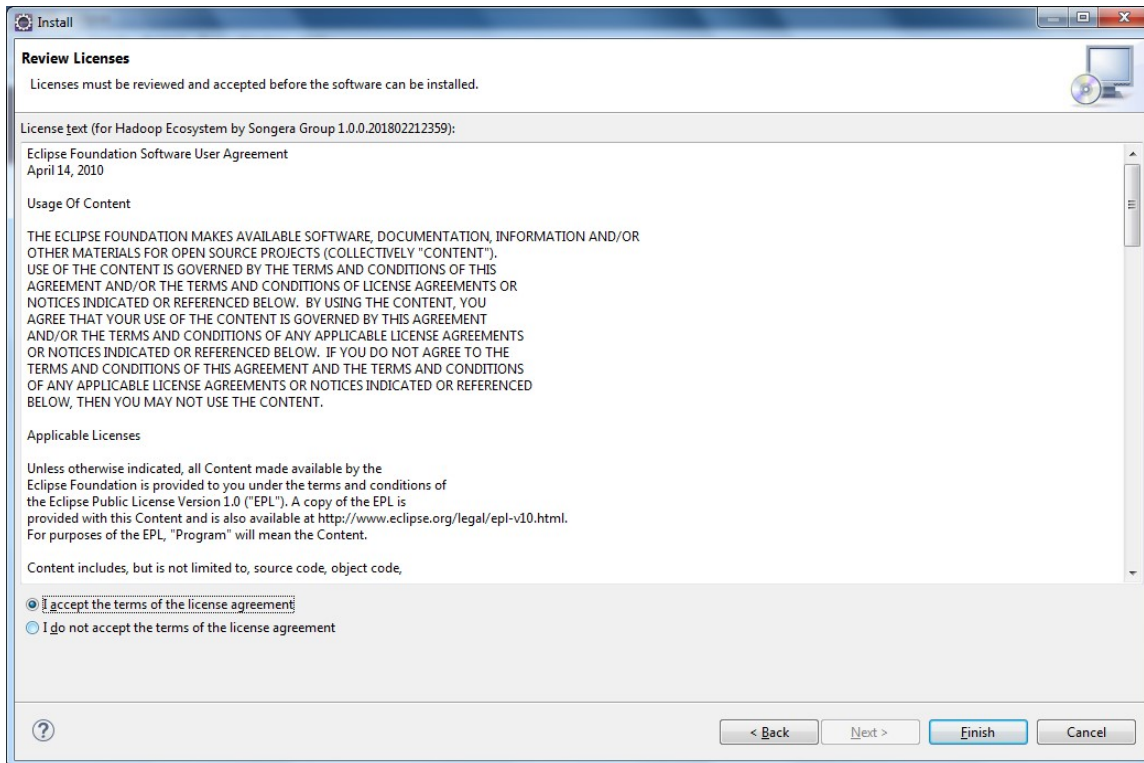
Select “Hadoop Ecosystem by Songera Group Site” and click Next button.



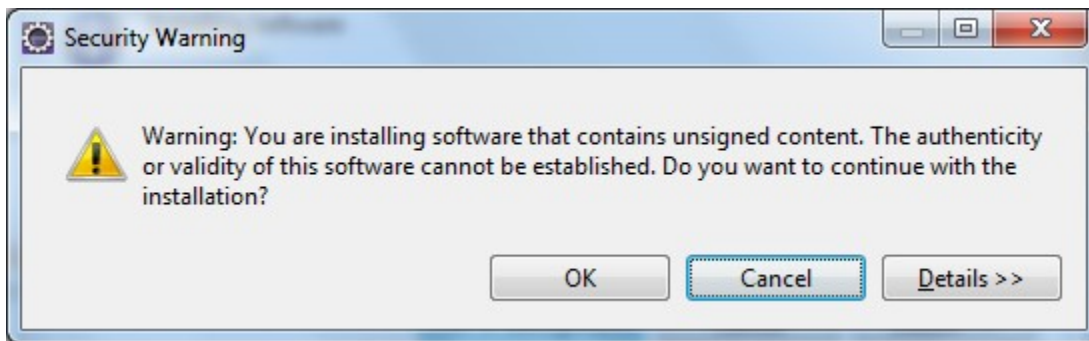
Click the Next Button.



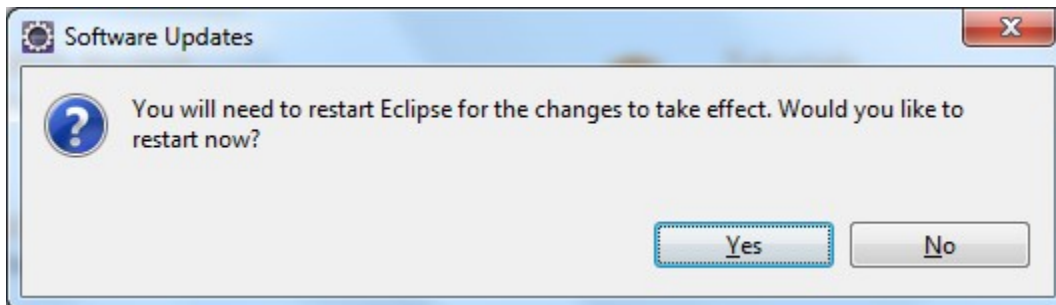
Read Licenses content carefully. To accept the conditions, click the Finish Button



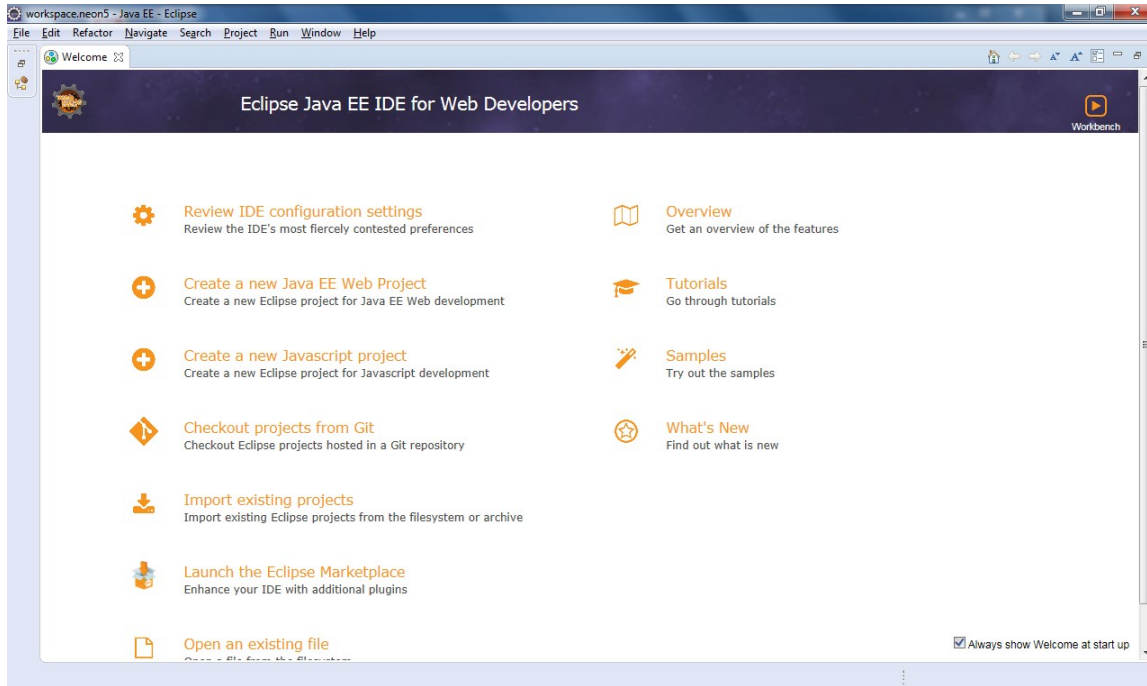
A Warning message will be displayed, as shown below. Click the OK button:



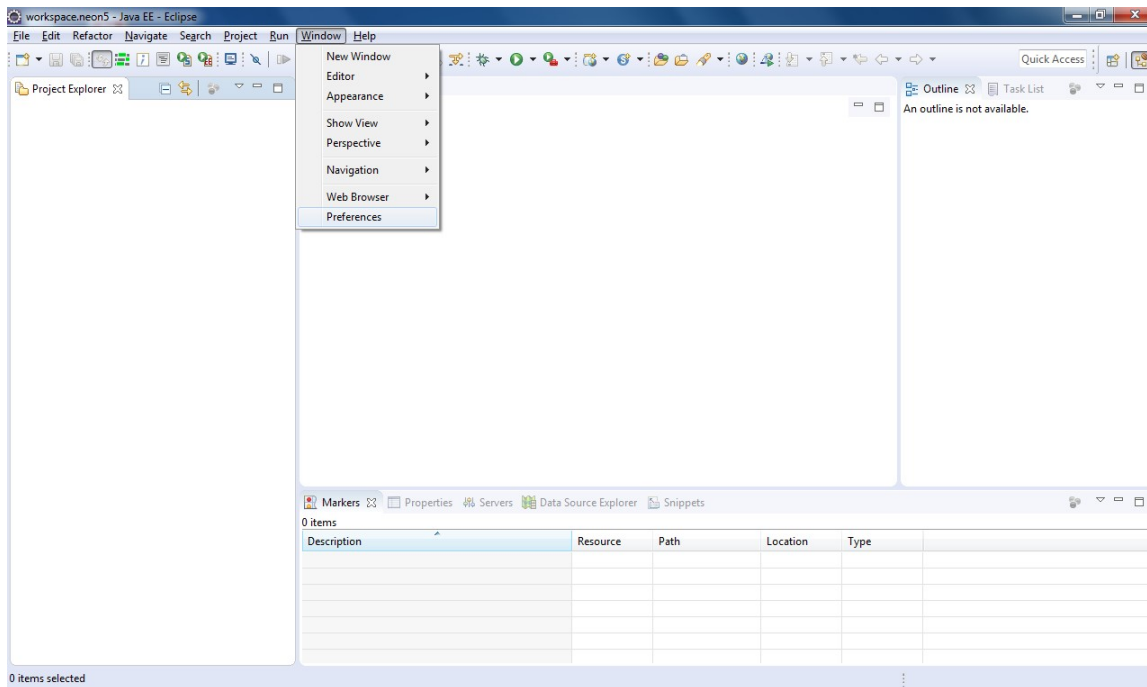
After the plugin is installed, restart Eclipse:



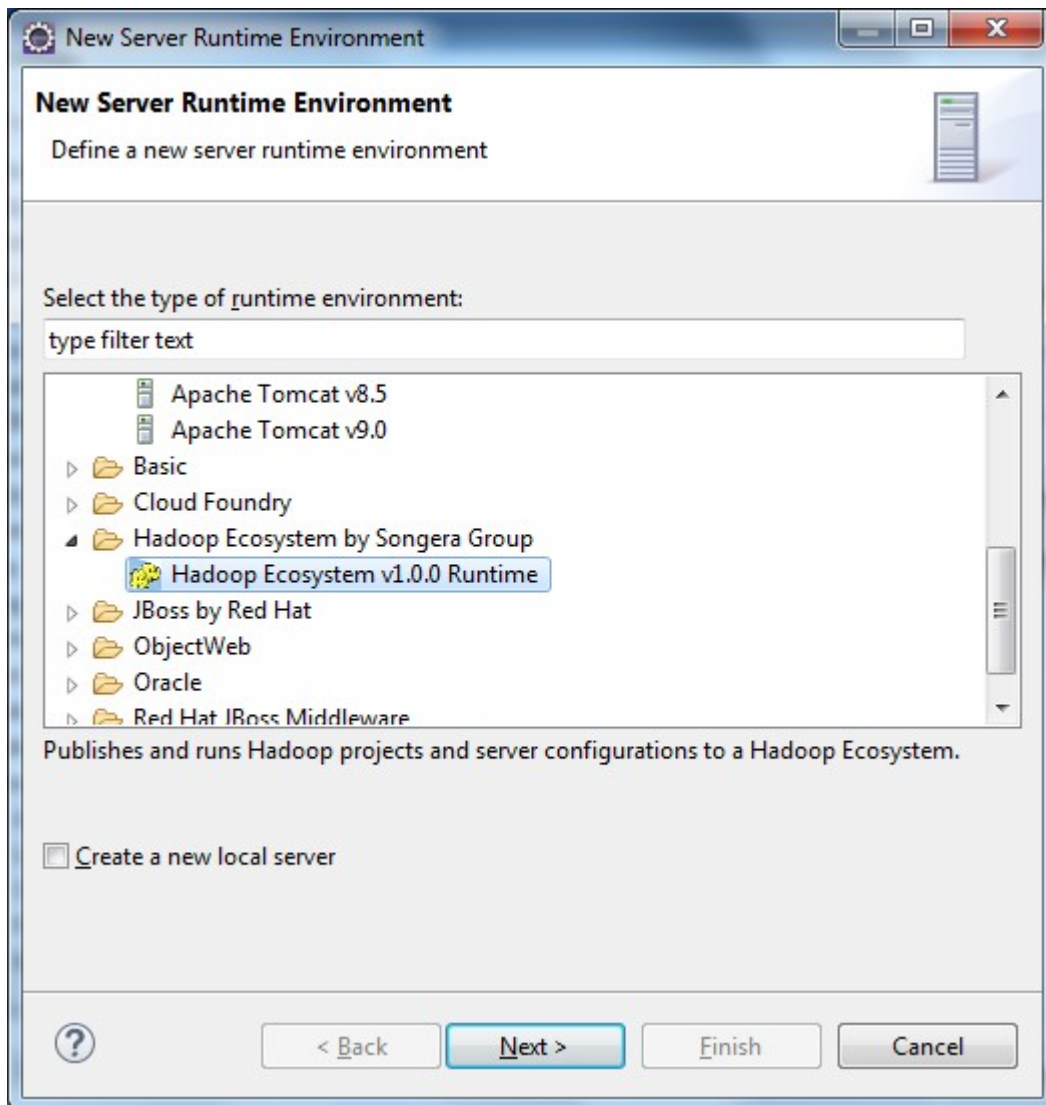
2. Install Hadoop module Runtime environment.
Start Eclipse Java EE IDE:



Click Menu: Window -> Preferences



In New Server runtime Environment screen, find “Hadoop Ecosystem by Songera Group” and select “Hadoop Ecosystem v1.0.0 Runtime.” Click the Next button.





This is the first page of the Hadoop Ecosystem setup. When you receive the license file, you can load the license file by clicking the “Load License...” button. Without the license file, it will run DEMO version with limited functionality.

For Windows, the Hadoop Ecosystem requires cygwin software. Select your region and download URL, and then click the “Deploy Cygwin” button to start downloading and deploying cygwin.

New Server Runtime Environment

Hadoop Ecosystem by Songera Group

 Songera Hadoop Ecosystem

 Cygwin does not deploy

Name:
Hadoop Ecosystem v1.0.0 Runtime

This is demo version of the software. It demos how to configure Hadoop Ecosystem and only deploy Hadoop HDFS module and run. Load License...

Cygwin Download Region:
United States

Cygwin Download URL:
New Jersey: <http://cygwin.mirror.constant.com/>

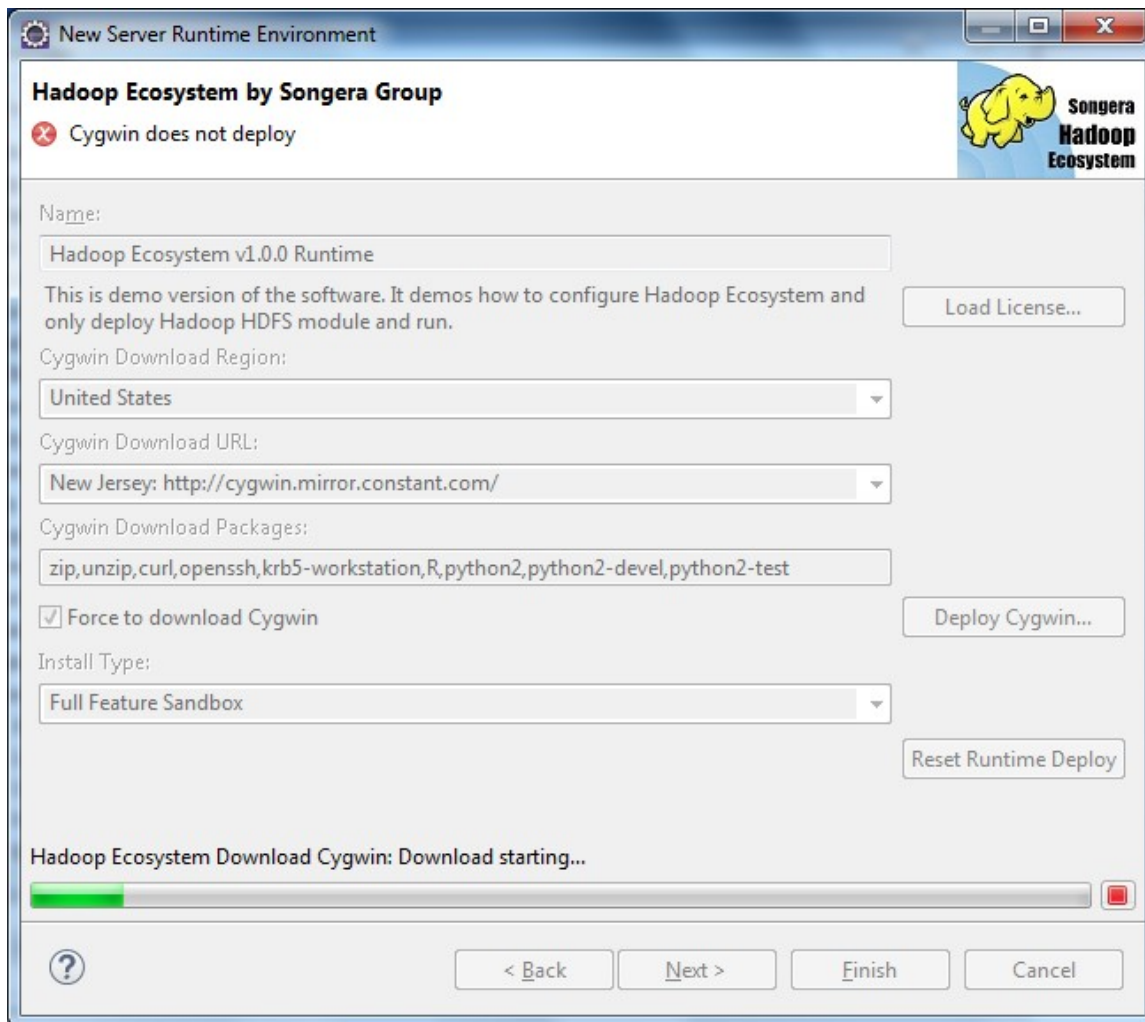
Cygwin Download Packages:
zip,unzip,curl,openssh,krb5-workstation,R,python2,python2-devel,python2-test

Force to download Cygwin Deploy Cygwin...

Install Type:
Full Feature Sandbox Reset Runtime Deploy

? < Back Next > Finish Cancel


Cygwin install and deploy:



When cygwin finish deploying, the “Next” button will enable the software. Click the Next button to go next page.

New Server Runtime Environment

Hadoop Ecosystem by Songera Group



Name:
Hadoop Ecosystem v1.0.0 Runtime

This is demo version of the software. It demos how to configure Hadoop Ecosystem and only deploy Hadoop HDFS module and run. Load License...

Cygwin Download Region:
United States

Cygwin Download URL:
New Jersey: <http://cygwin.mirror.constant.com/>

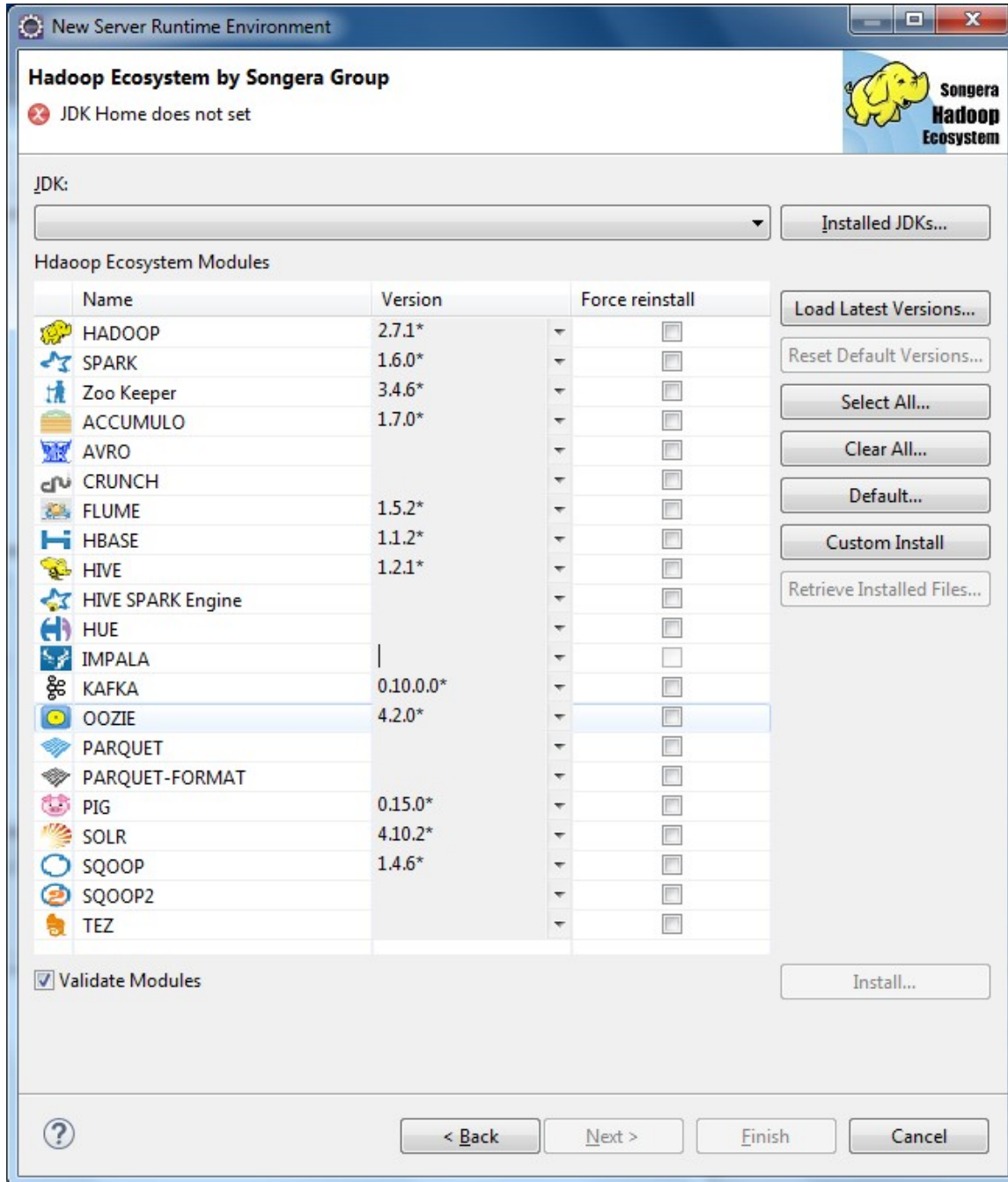
Cygwin Download Packages:
zip,unzip,curl,openssh,krb5-workstation,R,python2,python2-devel,python2-test

Force to download Cygwin Deploy Cygwin...

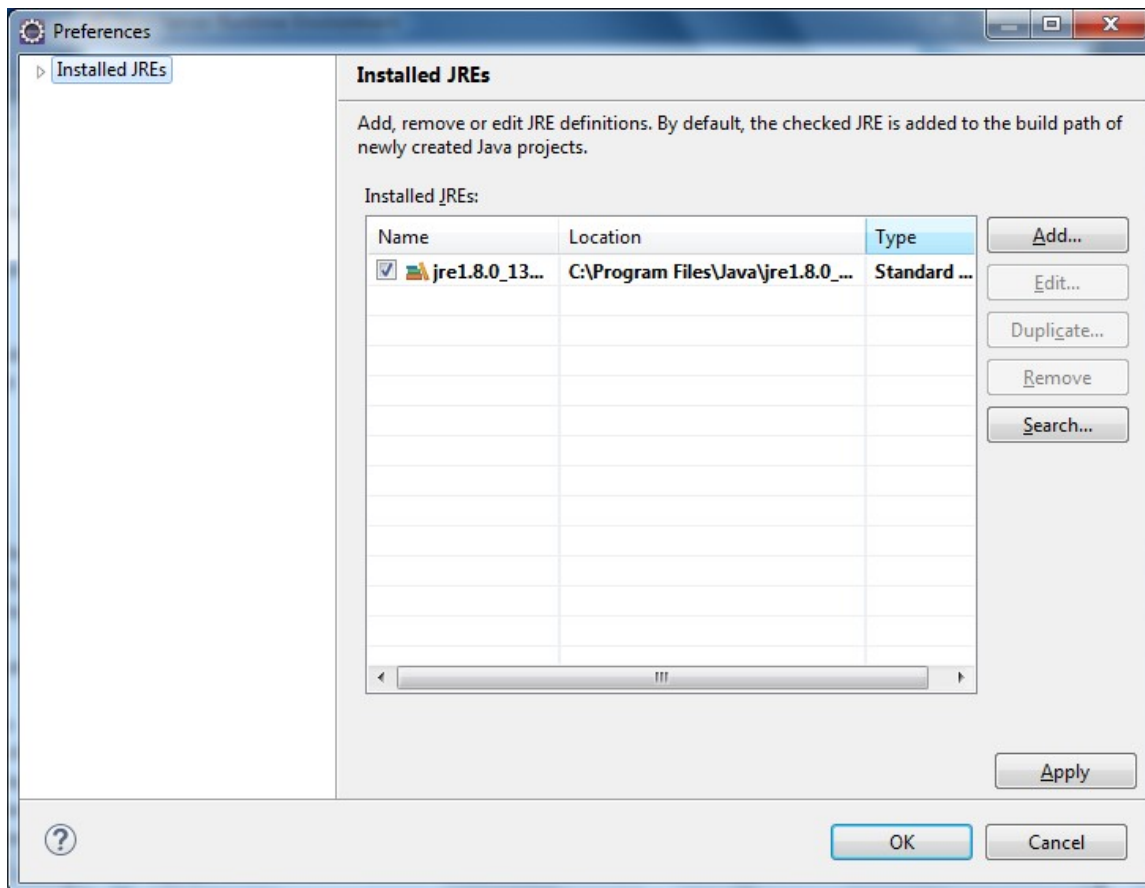
Install Type:
Full Feature Sandbox Reset Runtime Deploy

? < Back Next > Finish Cancel

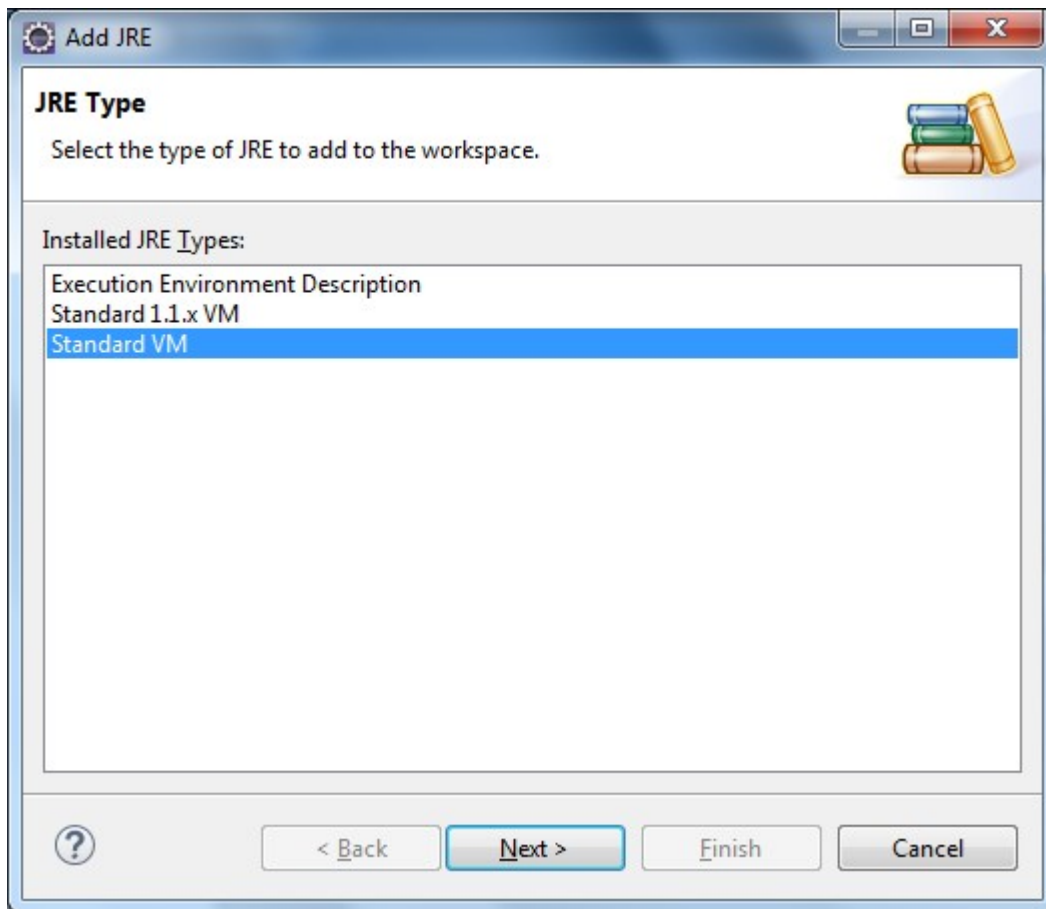
The Hadoop Ecosystem requires you to setup JDK first in eclipse. Click “Installed JDKs ...” button



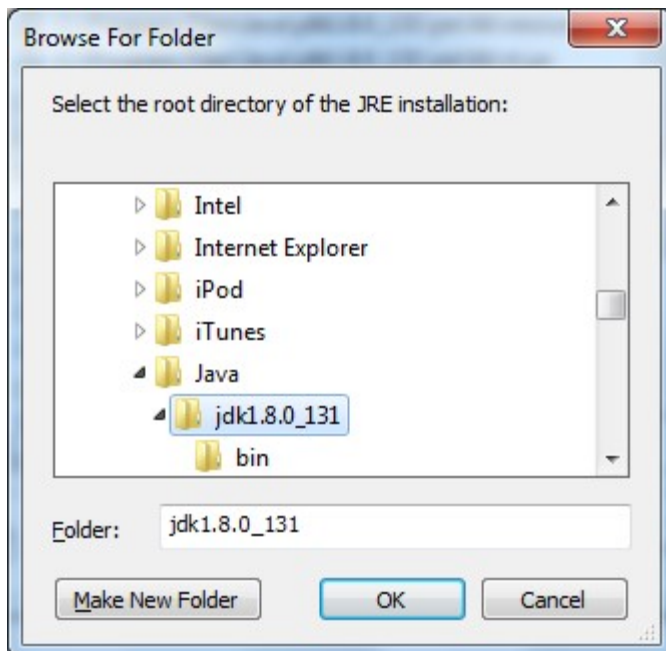
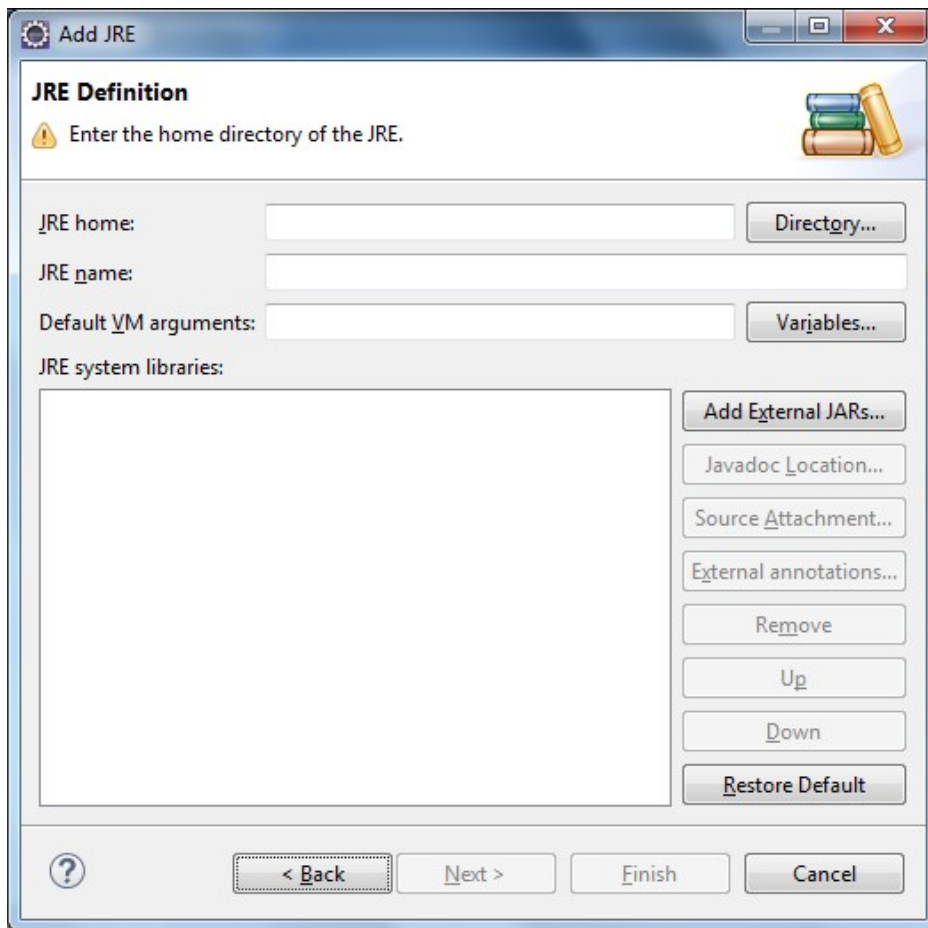
The Eclipse Installed JREs Screen will show up. Click the Add button.



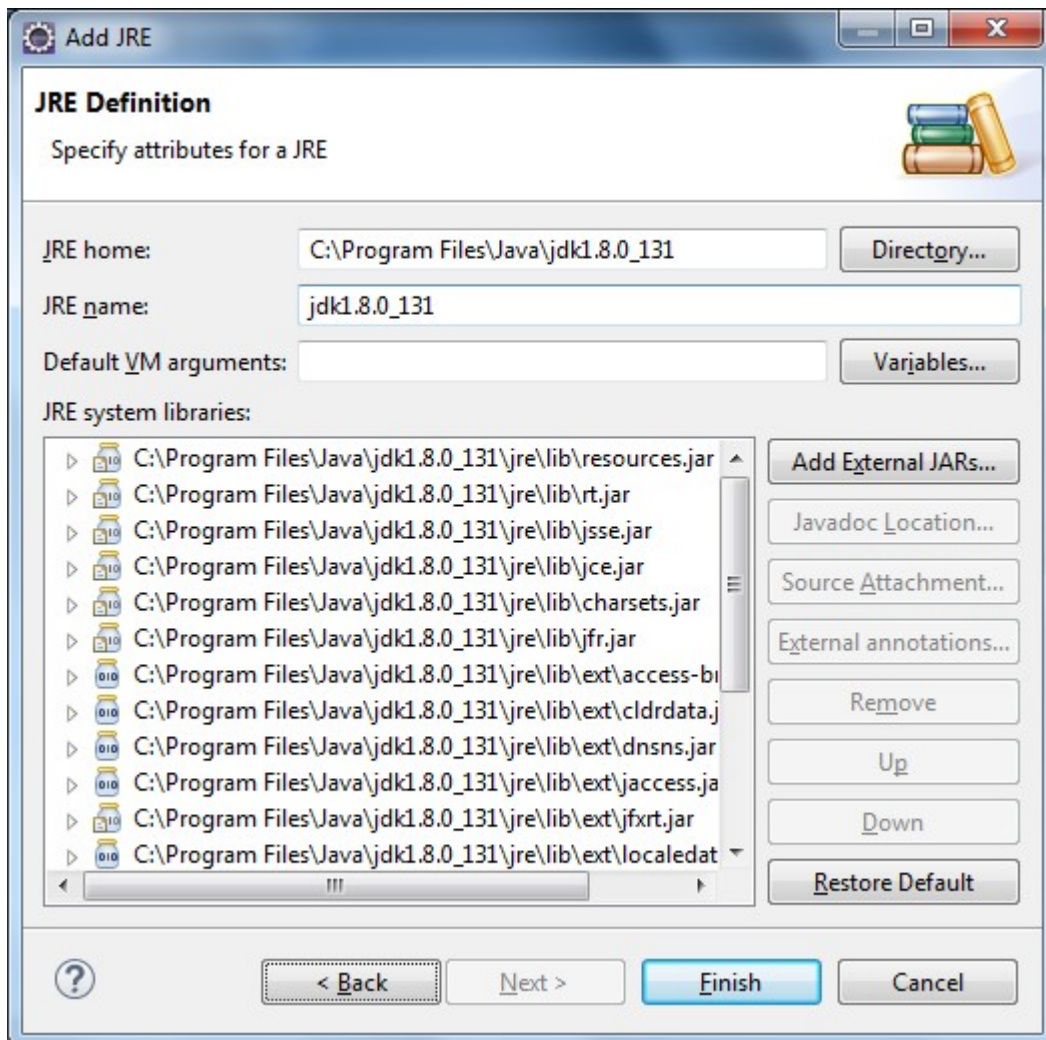
Select Standard VM and Click the Next button



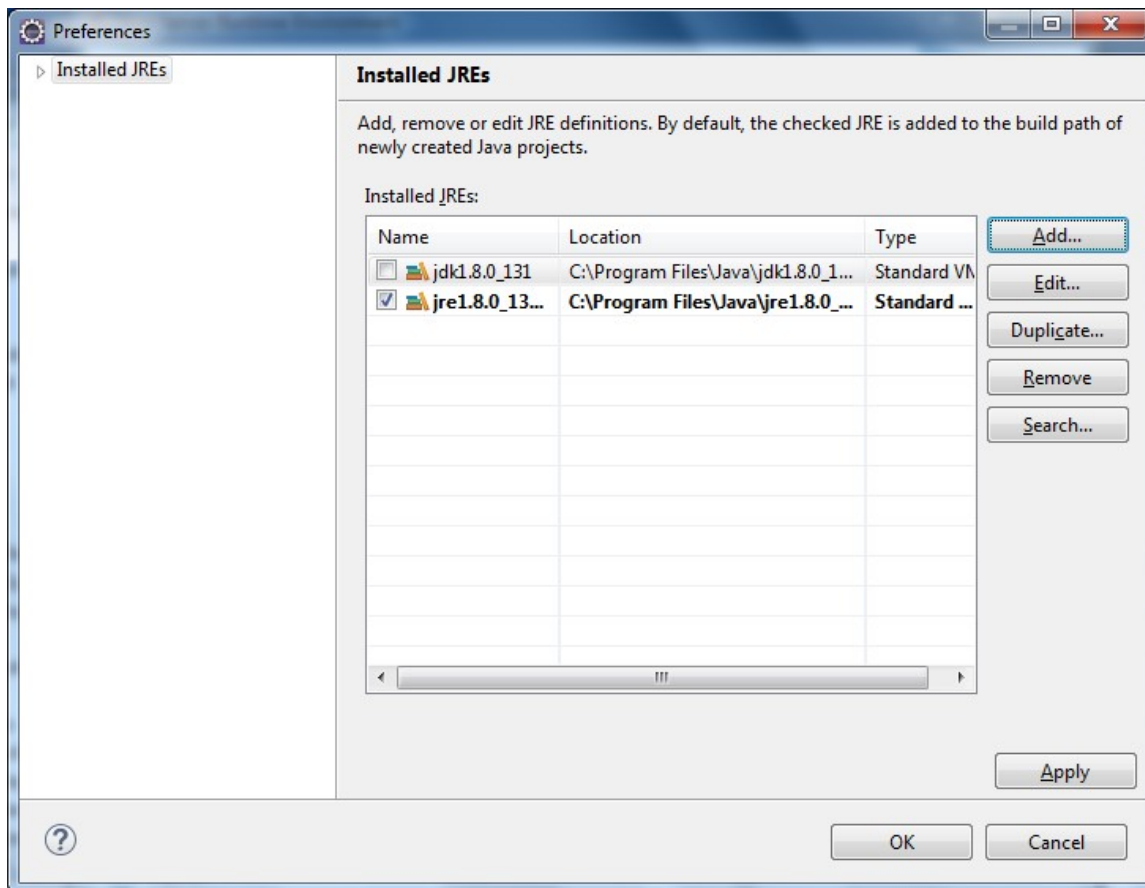
Fill JDK location by “Directory...” or “Variables...”



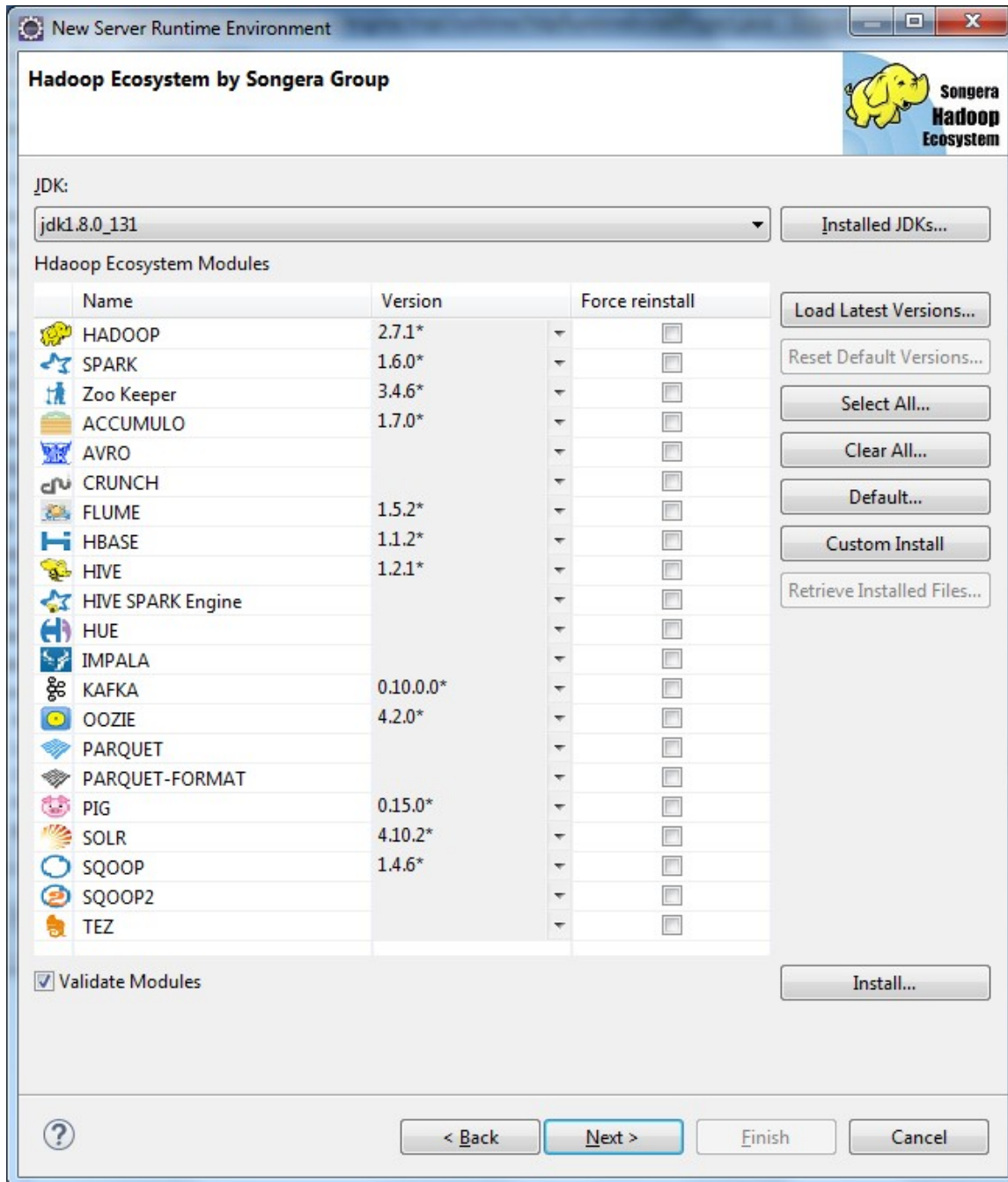
After JDK information is filled, check the Finish button



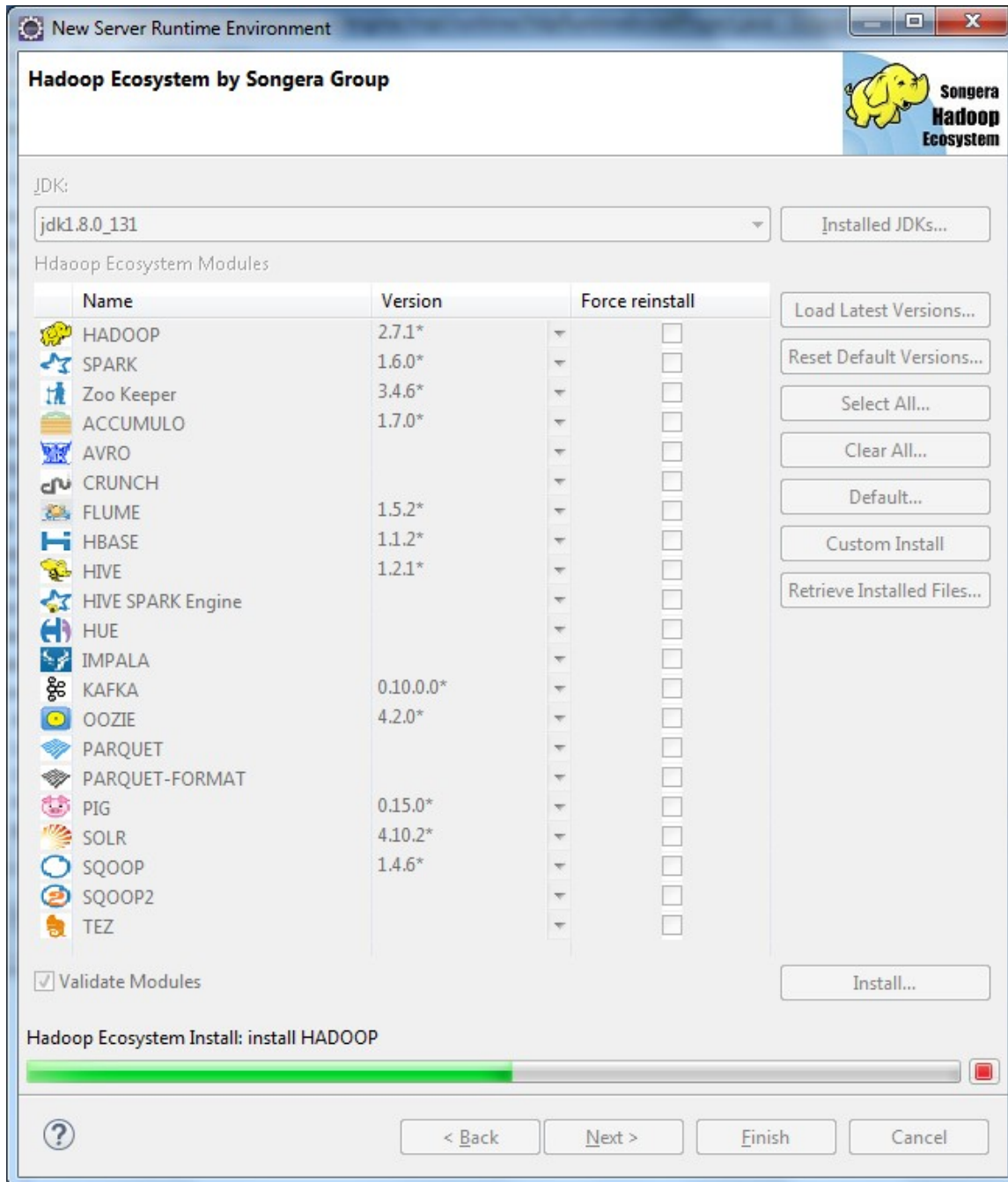
Now, the JDK entry is shown in the list. Click the OK button



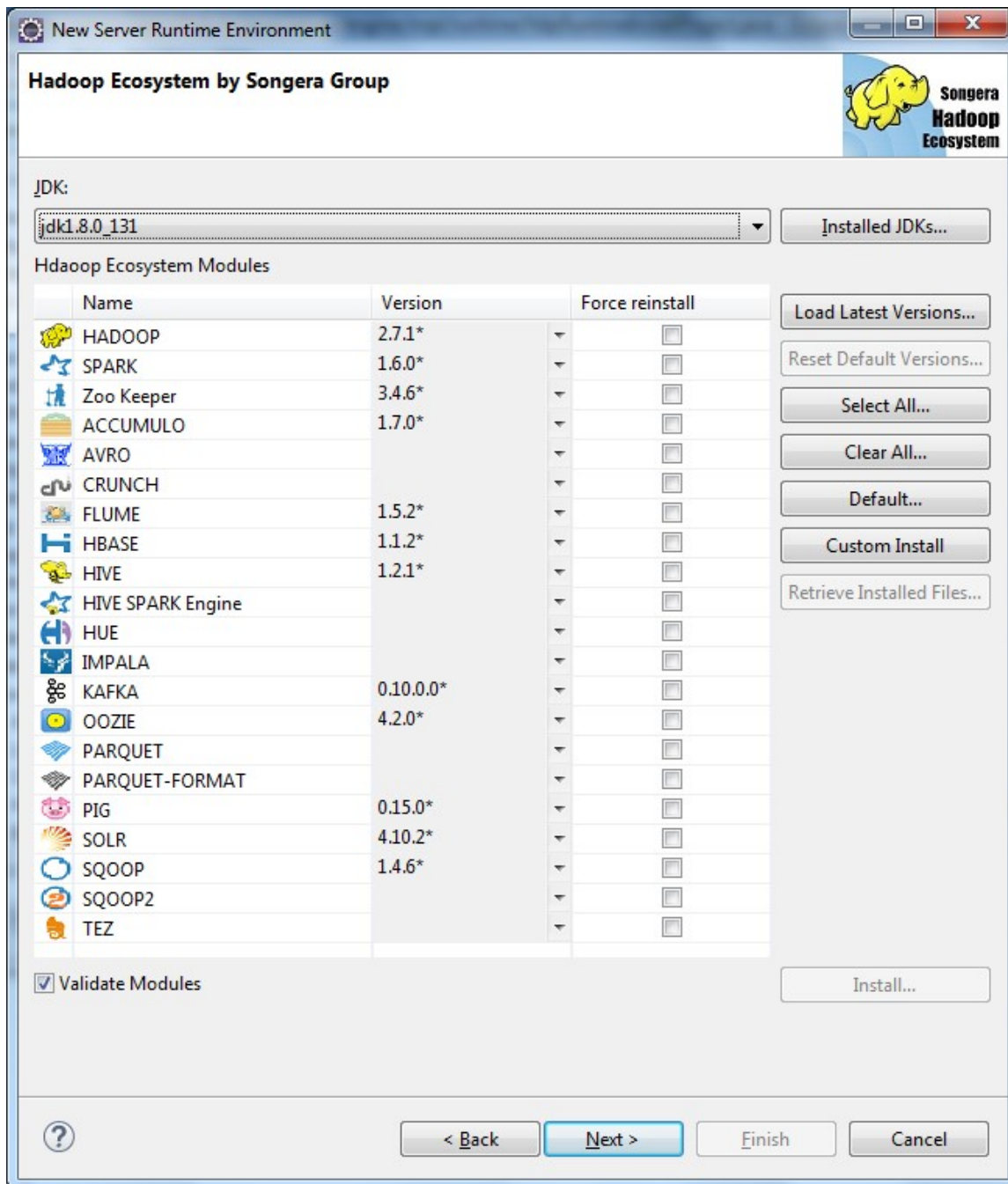
You can select different Hadoop modules or versions to deploy. Then, click the Install button to install the Hadoop module distribution software.



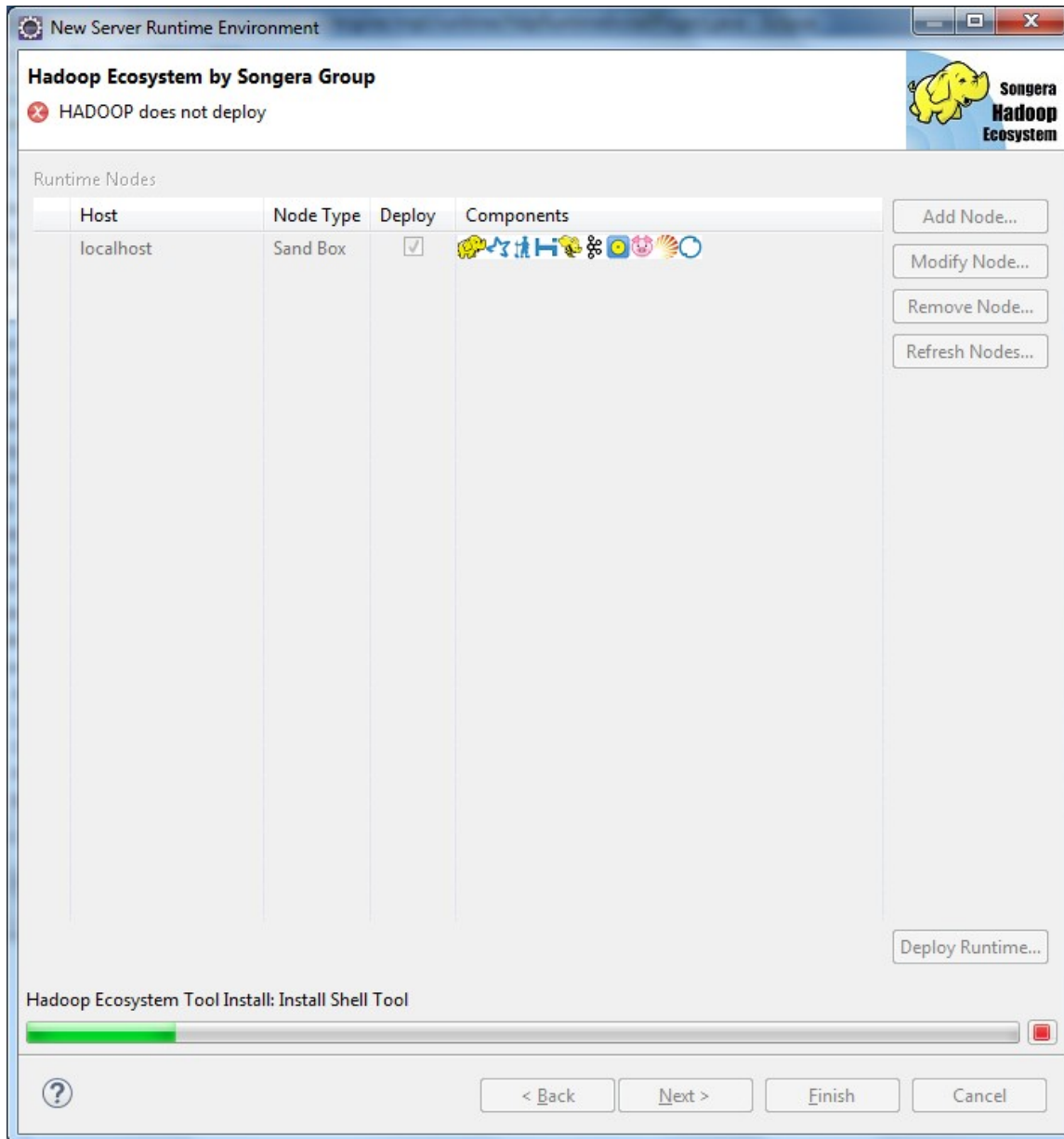
Installation processing might take some time depending on how fast the internet access speed is.



When finished installing, click the Next button.

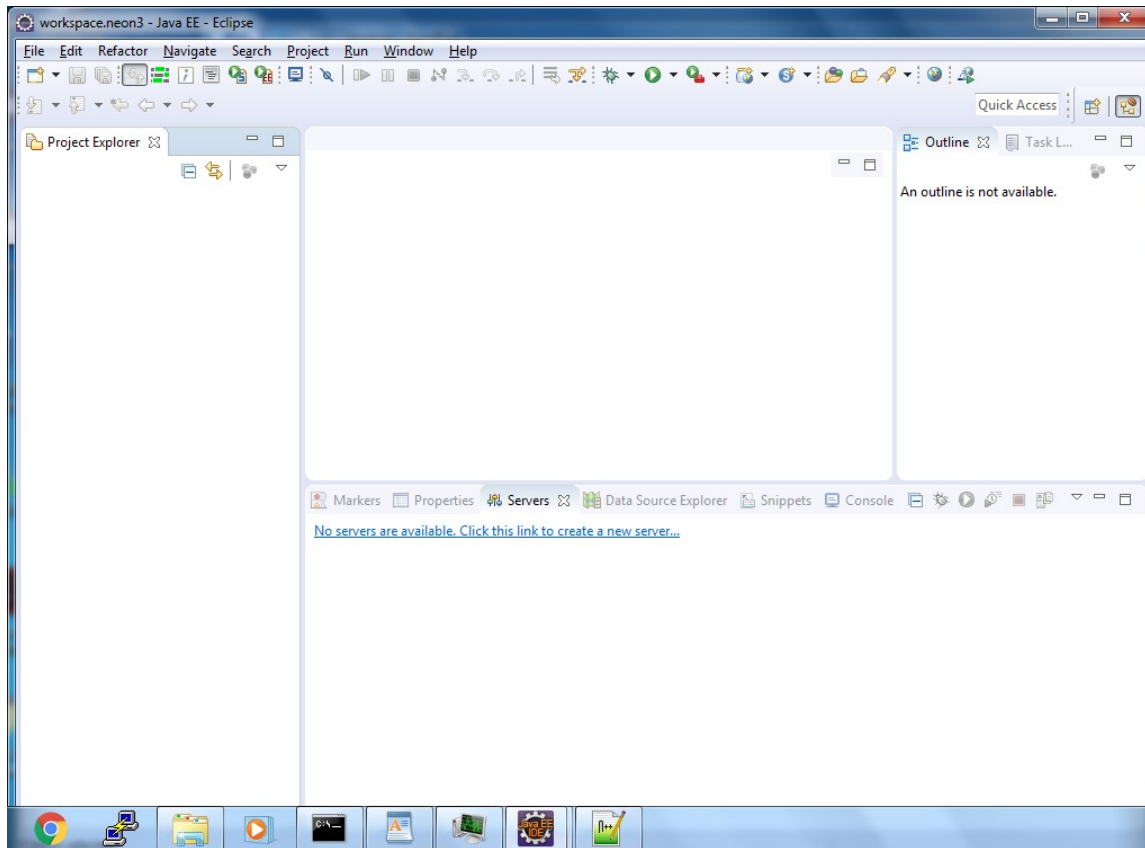


Deploy processing takes time. The progress bar shows the current status.

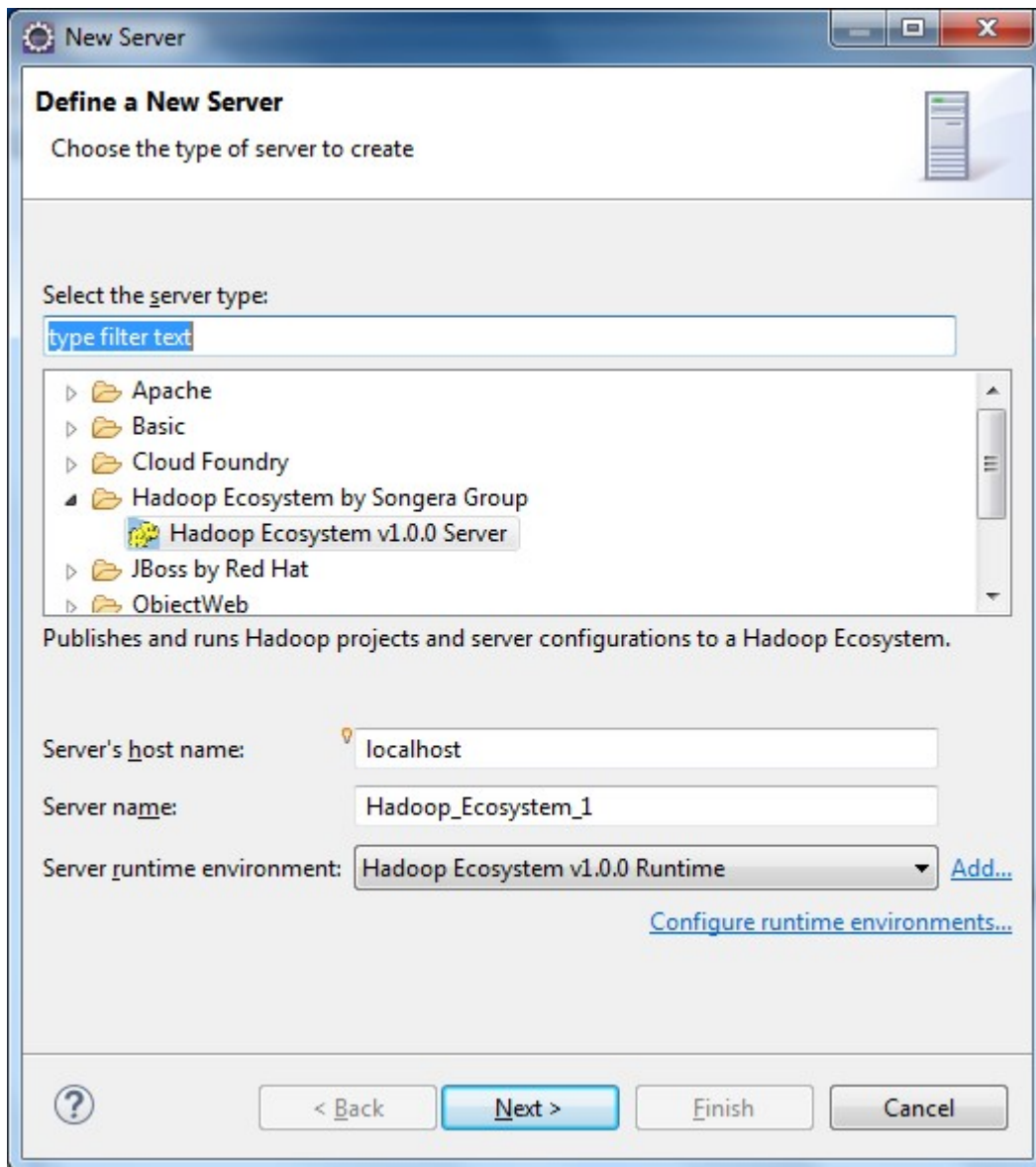


3. Server setup

After the Runtime environment is deployed, Select eclipse Servers tab and click “No servers are available. Click this link to create a new server...”



On the New Server dialog, find the “Hadoop Ecosystem by Songera Group” tree item and select “Hadoop Ecosystem v1.0.0 Server,” and then click the Next button.



This “Hadoop Server” screen allows you to configure Hadoop Core, HDFS, Mapred and Yarn. Click the Next button when finished changing the configurations.

Hadoop Ecosystem by Songera Group

Hadoop Server

Hadoop Deploy Setting
Configure Hadoop Server deploy Setting

Name	Value
namenode.enableSecondary	<input type="checkbox"/>
namenode.enableBackup	<input type="checkbox"/>
namenode.hacluster	
kerberosEnable	<input type="checkbox"/>

Hadoop Defaults...

- ▶ Hadoop Service Ports
- ▶ Hadoop Core Properties
- ▶ Hadoop Name Node Properties
- ▶ Hadoop Data Node Properties
- ▶ Hadoop Map Reduce Properties
- ▶ Hadoop Yarn Properties

? < Back Next > Finish Cancel

This screen allows you to change the database and Hadoop Module setup. Click the Next button when finish with the setup.

Hadoop Ecosystem by Songera Group

Hadoop Ecosystem Modules

Databases
Configure the Hadoop Ecosystem database on this server.

Name	Name
demo	jdbc:hsqldb:hsqldb://\${HSQLDB:HOST:SERVER};\${HSQLDB:...
hive	jdbc:derby://\${DBDERBY:HOST:SERVER};\${DBDERBY:POR...
oozie	jdbc:hsqldb:hsqldb://\${HSQLDB:HOST:SERVER};\${HSQLDB:...
sqoop2	jdbc:derby://\${DBDERBY:HOST:SERVER};\${DBDERBY:POR...

Buttons: Add Database..., Edit Database..., Remove Database..., Default Database...

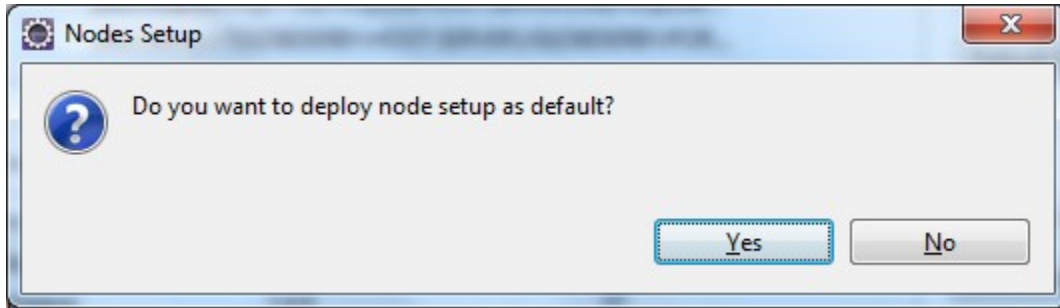
Modules
Configure the Hadoop Ecosystem modules on this server.

Name	Version	Enable
SPARK	1.6.0	<input checked="" type="checkbox"/>
Zoo Keeper	3.4.6	<input checked="" type="checkbox"/>
HBASE	1.1.2	<input checked="" type="checkbox"/>
HIVE	1.2.1	<input checked="" type="checkbox"/>
KAFKA	0.10.0.0	<input checked="" type="checkbox"/>
OOZIE	4.2.0	<input checked="" type="checkbox"/>
PIG	0.15.0	<input checked="" type="checkbox"/>
SOLR	4.10.2	<input checked="" type="checkbox"/>
SQOOP	1.4.6	<input checked="" type="checkbox"/>

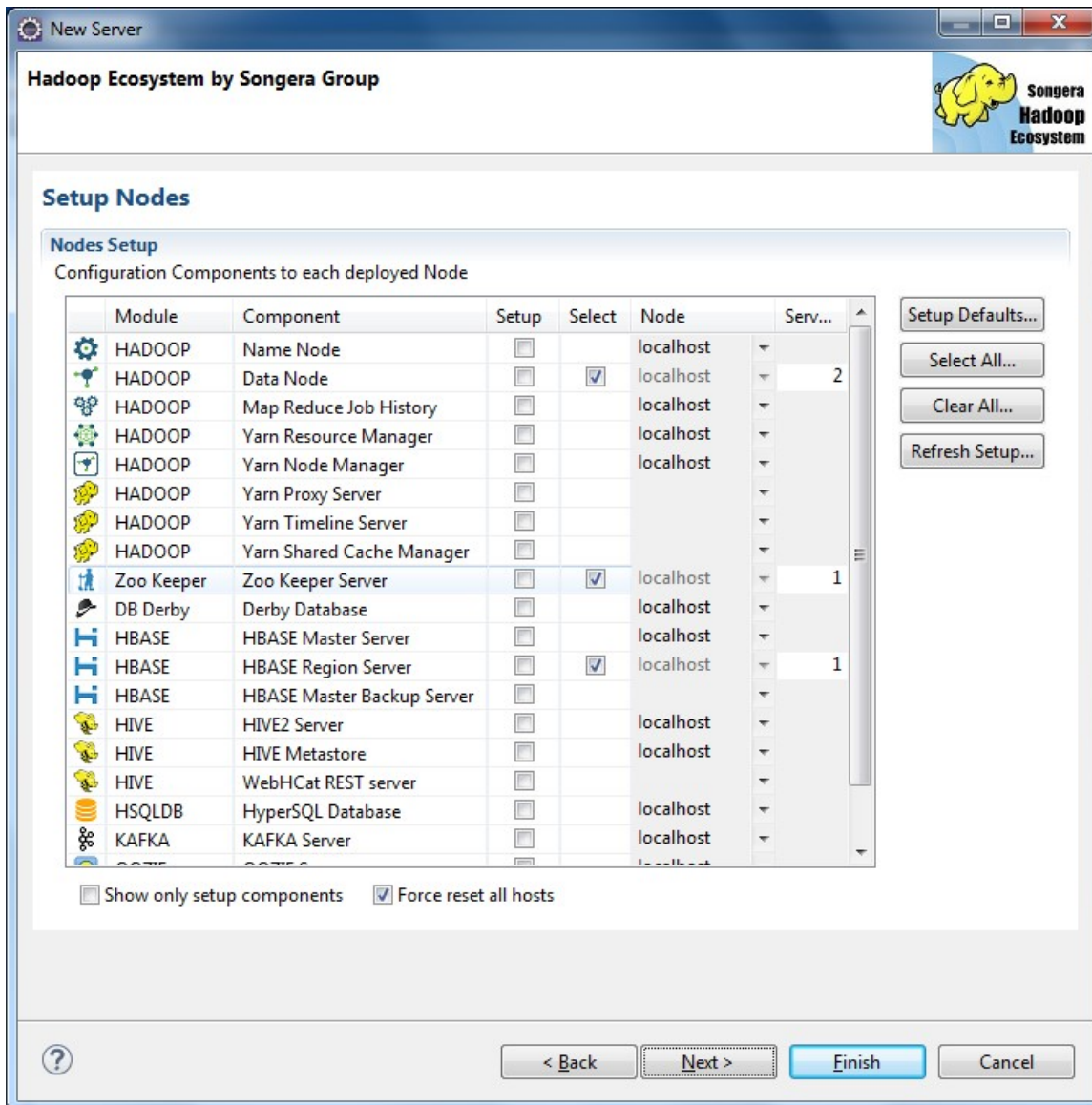
Buttons: Edit Module..., Select All Modules..., Clear All Modules...

Navigation: ? < Back **Next >** Finish Cancel

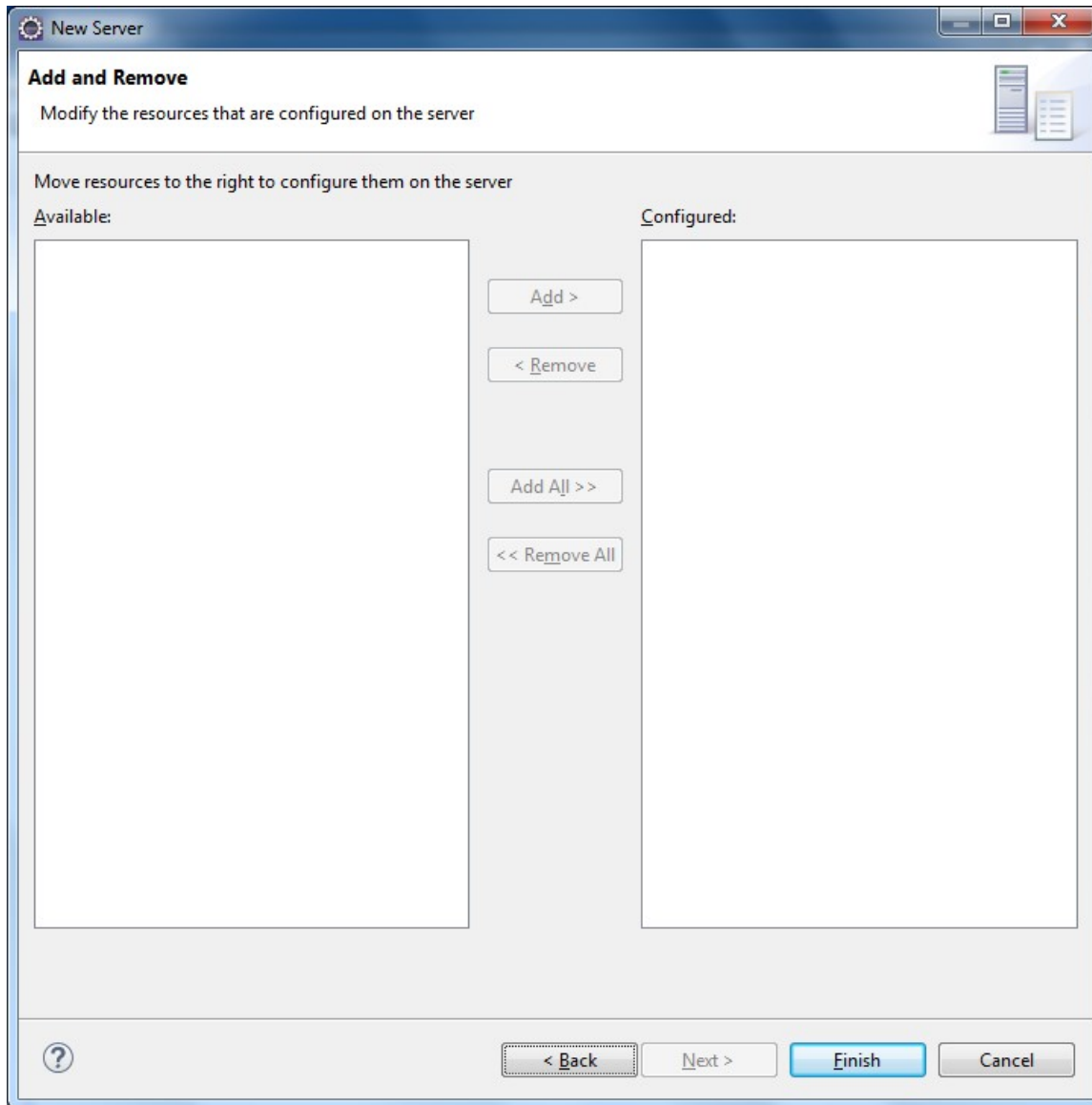
An alert screen will prompt ask you if you want to deploy the settings. Click Yes to allow the Hadoop Ecosystem to use them as the default.



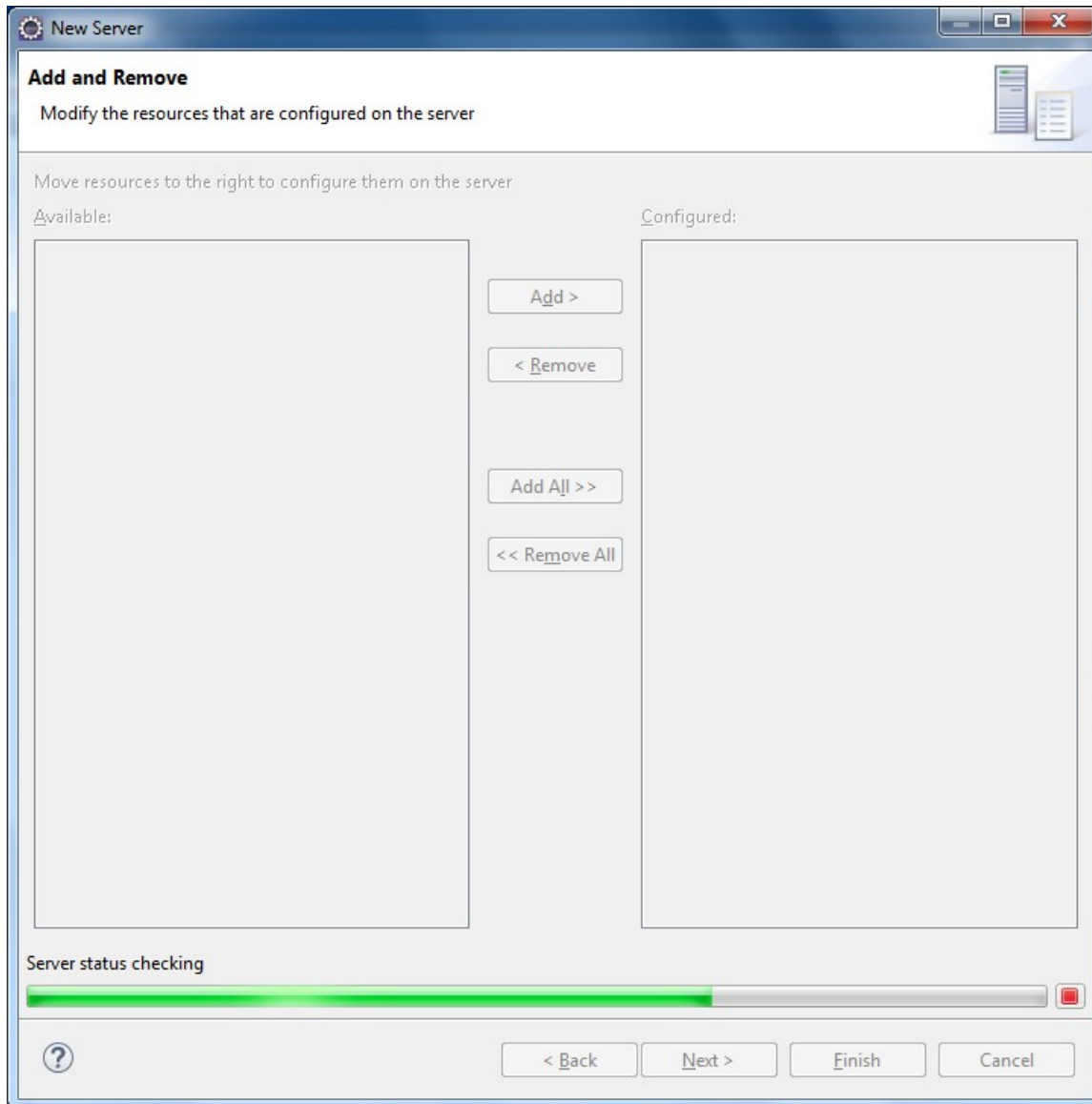
The Setup Nodes Screen displays information on where each Hadoop module service is hosted. Click Next when finish with changes.



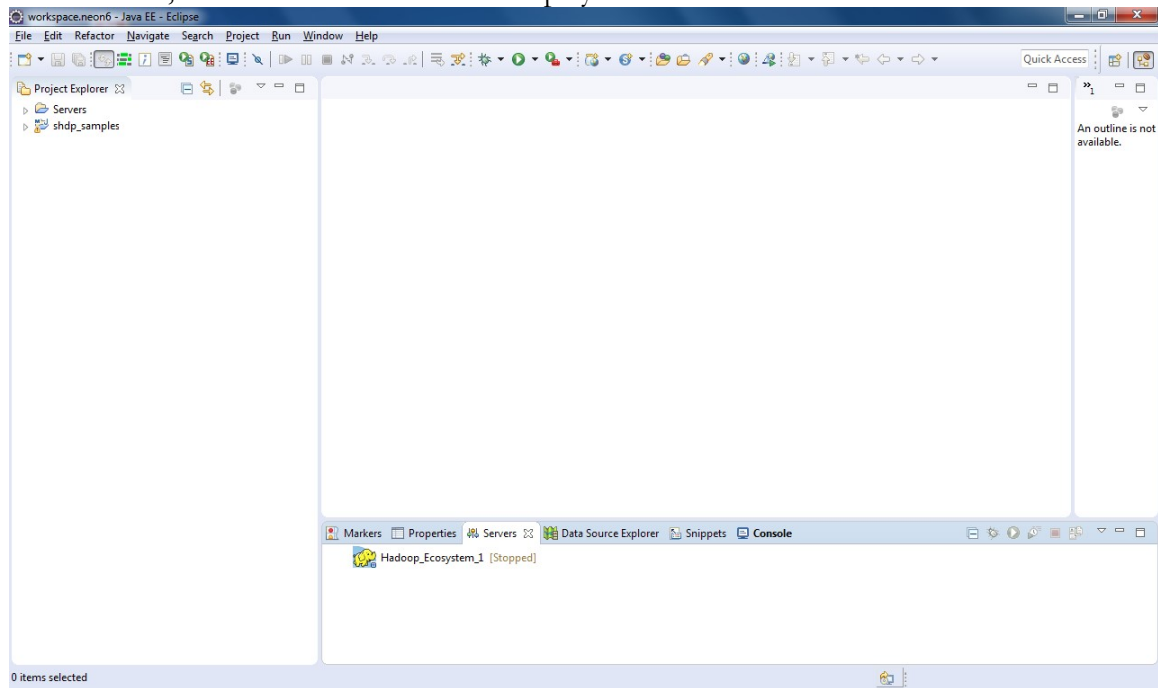
Hadoop Ecosystem allows you to automatically sync current workspace directories and files to Hadoop HDFS file system. This screen allows you to configure those setups. Click the Finish button.



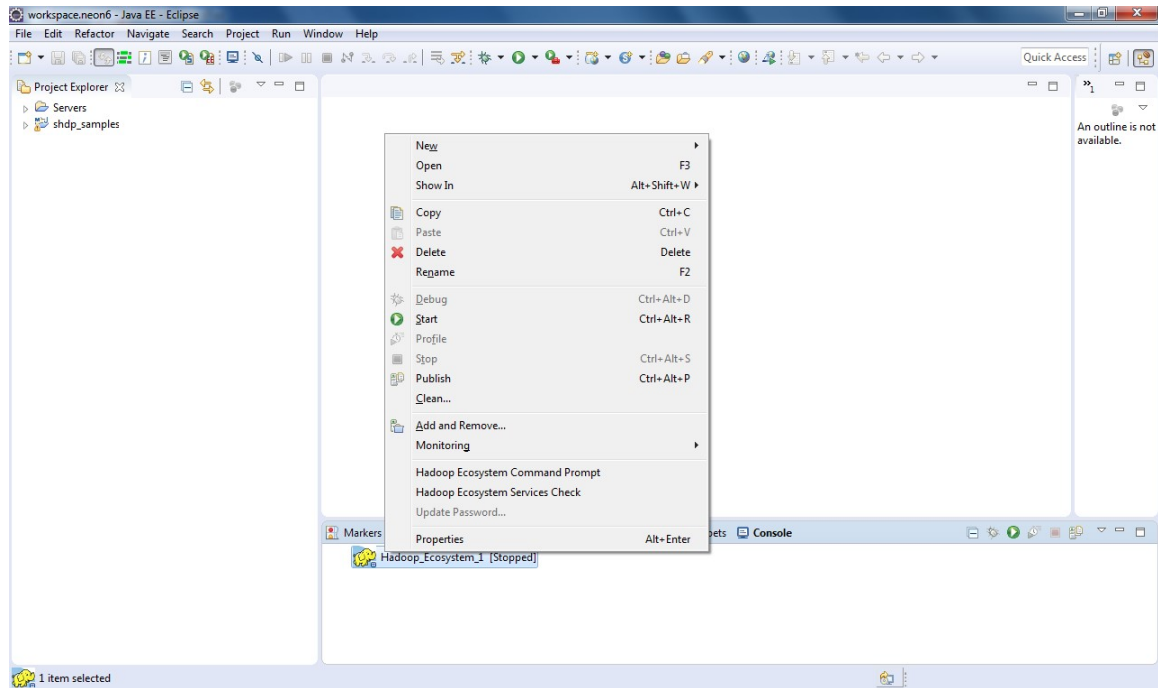
It takes a while to setup each Hadoop module configuration to the hosts.



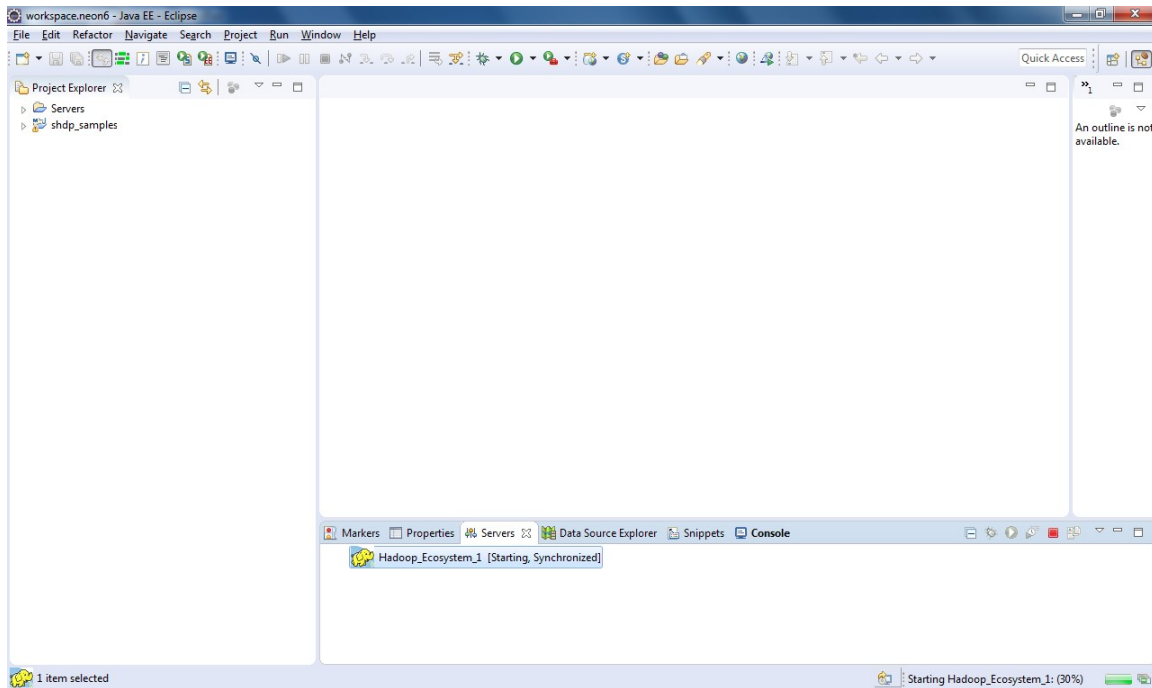
Now, the server instance will be displayed under the servers tab.



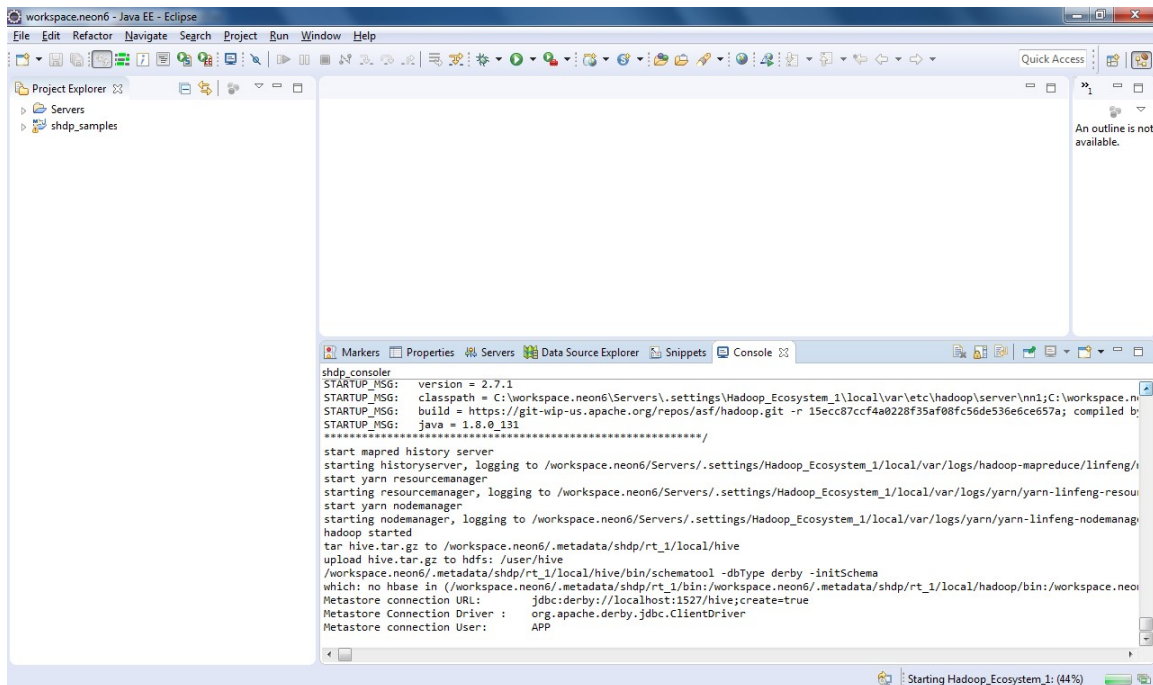
Right click the item to prompt the menu and click “start” menu to start services.



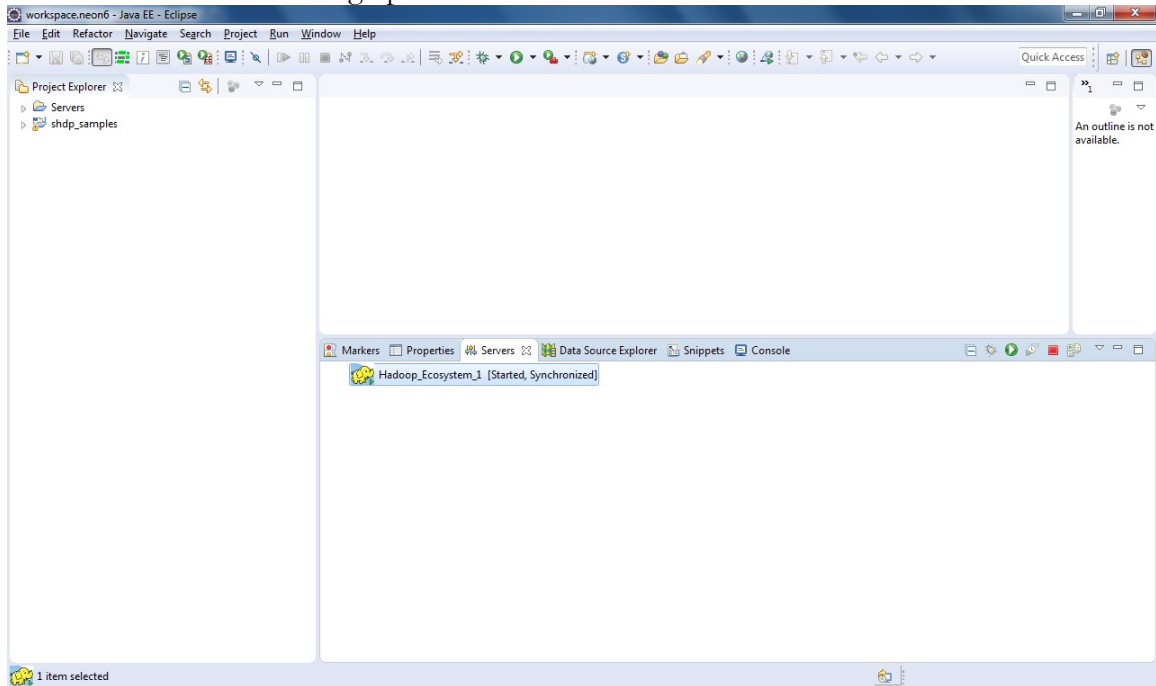
Server status shows the service starting.



Switch to the Console tab to show server startup in a detailed log.



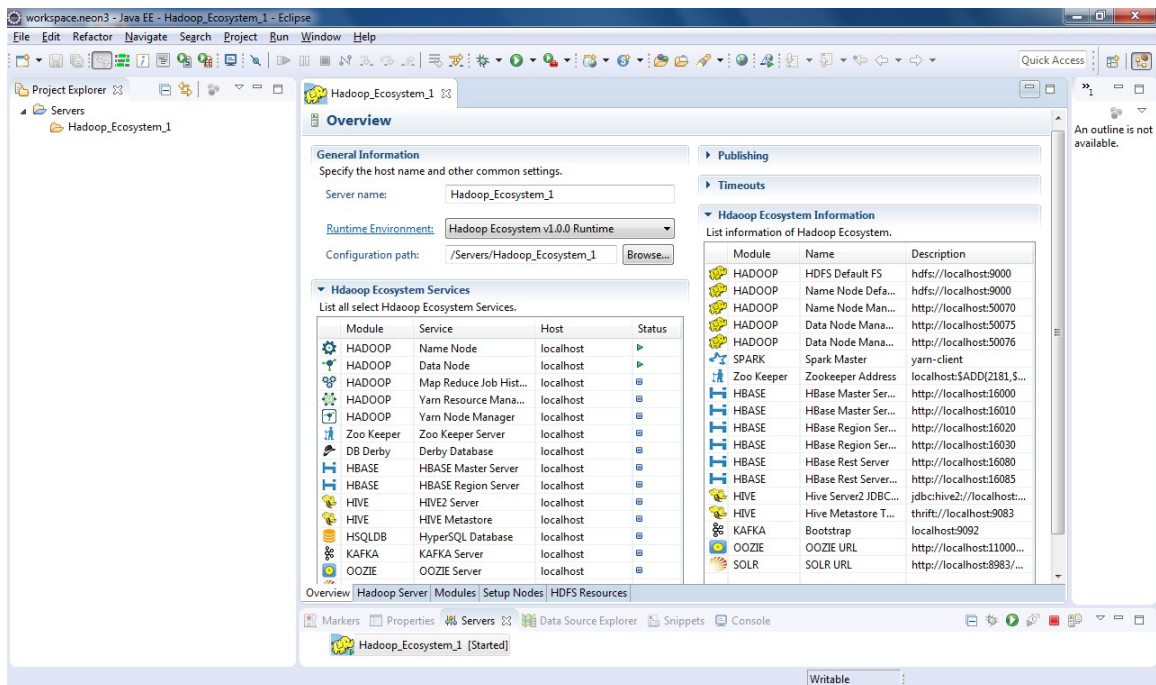
It takes time to bring up the server.



Double Click “Hadoop_Ecosystem_1” item to display the server information. The list on the left shows service start/stop information. The list on the right shows service access information for applications or users.

The Demo license only starts Hadoop Name Node and Data Node service. However, for other licenses, all services will be started.

Demo License:



Trial, Test or Commercial license:

The screenshot shows the Eclipse IDE interface with the Hadoop Ecosystem Overview page. The page is divided into several sections:

- General Information:** Server name: Hadoop_Ecosystem_1. Runtime Environment: Hadoop Ecosystem v1.0.0 Runtime. Configuration path: /Servers/Hadoop_Ecosystem_1.
- Publishing:** Timeouts section.
- Hadoop Ecosystem Information:** List information of Hadoop Ecosystem. A table lists modules and their descriptions.
- Hadoop Ecosystem Services:** List all select Hadoop Ecosystem Services. A table lists modules, services, hosts, and statuses.

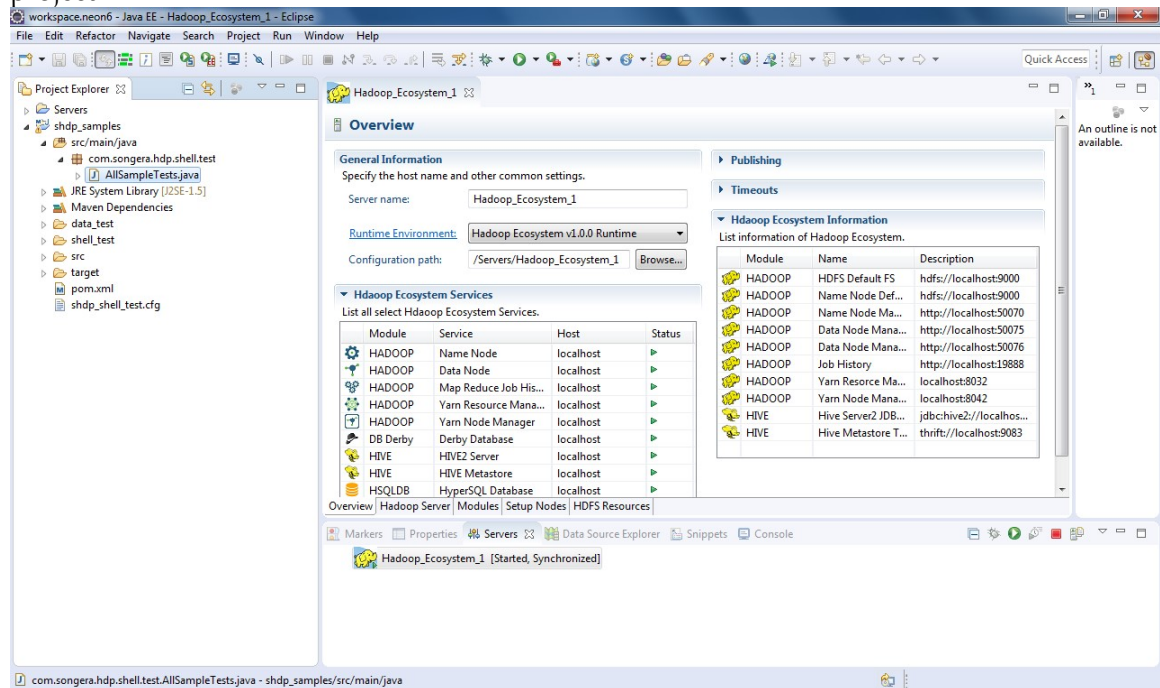
The console at the bottom shows: Hadoop_Ecosystem_1 [Started, Synchronized]

Module	Name	Description
HADOOP	HDFS Default FS	hdfs://localhost:9000
HADOOP	Name Node Def...	hdfs://localhost:9000
HADOOP	Name Node Ma...	http://localhost:50070
HADOOP	Data Node Mana...	http://localhost:50075
HADOOP	Data Node Mana...	http://localhost:50076
HADOOP	Job History	http://localhost:19888
HADOOP	Yarn Resorce Ma...	localhost:8032
HADOOP	Yarn Node Mana...	localhost:8042
HIVE	Hive Server2 JDB...	jdbc:hive2://localhos...
HIVE	Hive Metastore T...	thrift://localhost:9083

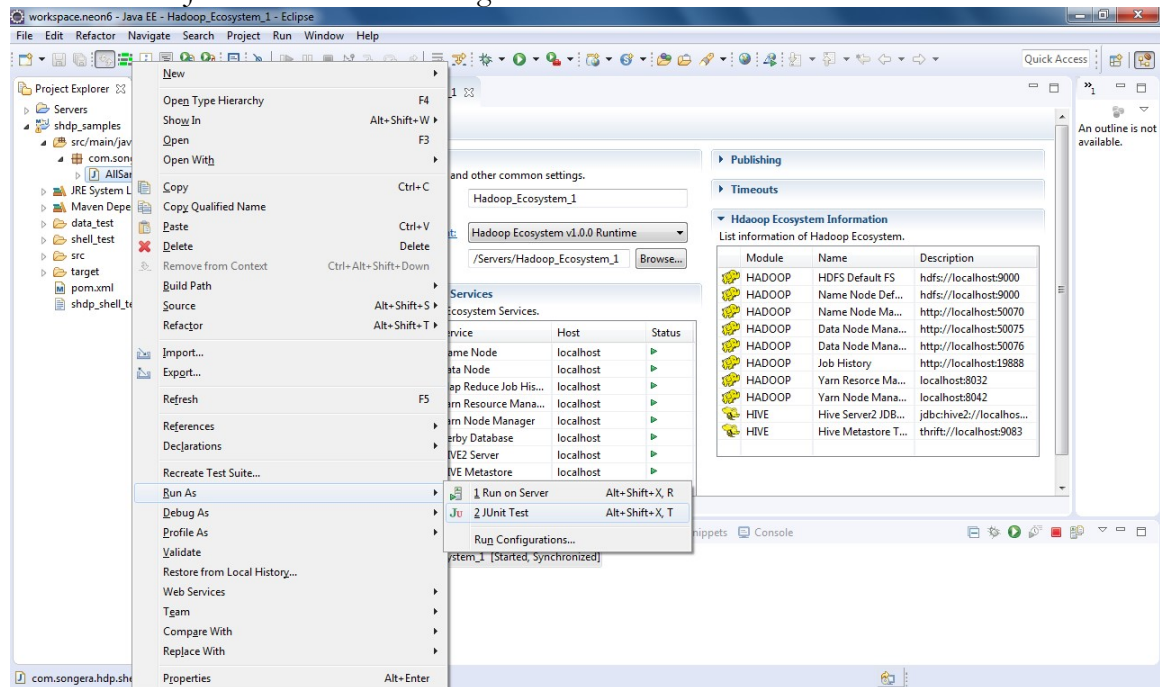
Module	Service	Host	Status
HADOOP	Name Node	localhost	▶
HADOOP	Data Node	localhost	▶
HADOOP	Map Reduce Job His...	localhost	▶
HADOOP	Yarn Resource Mana...	localhost	▶
HADOOP	Yarn Node Manager	localhost	▶
DB Derby	Derby Database	localhost	▶
HIVE	HIVE2 Server	localhost	▶
HIVE	HIVE Metastore	localhost	▶
HSQldb	HyperSQL Database	localhost	▶

4. Test Hadoop Modules

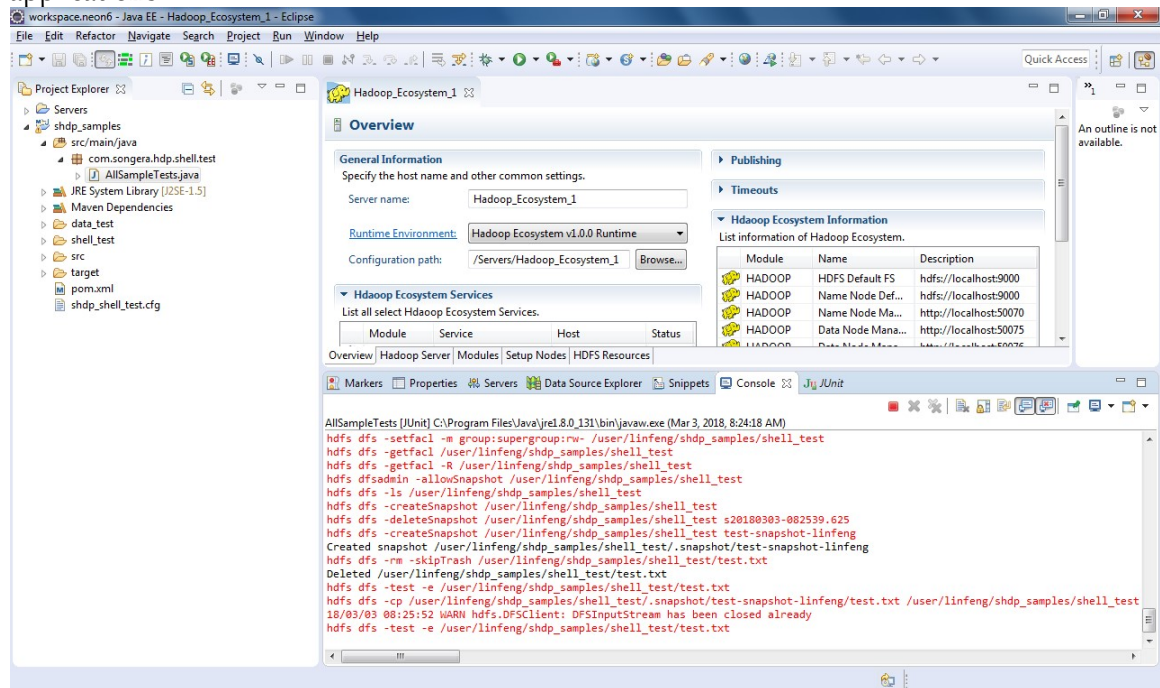
When the Hadoop Ecosystem server instance is created, it automatically import sample project: shdp_samples to current workspace. One Junit test suite java class is under that project.



Run the Junit test suit to test all deployed services and modules. Right click java class and run as -> Junit Test to start testing.



Test cases are written in Java and launch in the Unix Shell under the server. Switch to the console tab to display a detailed test command. For commercial version, source code of those test cases is shipped in packages. User can modify or write their own test cases or applications.



Once the test is done, it will pass or failure.

The screenshot displays the Eclipse IDE interface for a project named 'Hadoop_Ecosystem_1'. The 'Overview' tab is active, showing configuration details for the Hadoop ecosystem. The 'General Information' section includes the server name 'Hadoop_Ecosystem_1' and the runtime environment 'Hadoop Ecosystem v1.0.0 Runtime'. The 'Hadoop Ecosystem Services' section lists various services and their hosts.

Module	Service	Host	Status
HADOOP	HDFS Default FS	hdfs://localhost:9000	
HADOOP	Name Node Def...	hdfs://localhost:9000	
HADOOP	Name Node Ma...	http://localhost:50070	
HADOOP	Data Node Mana...	http://localhost:50075	
HADOOP	Data Node Mana...	http://localhost:50076	

The bottom of the IDE shows the test results. The console indicates that the tests finished after 567.179 seconds, with 46 runs, 0 errors, and 0 failures. The test suite 'com.songera.hdp.shell.command.test.ShellTestSuite' is expanded, showing a list of individual tests and their execution times.

Test Class	Execution Time
com.songera.hdp.shell.core.test.ShellTest	2.217 s
com.songera.hdp.shell.command.test.PrepareTest	24.935 s
com.songera.hdp.shell.command.test.output.FileInfoTest	0.016 s
com.songera.hdp.shell.command.test.output.MapredJobInfoTest	0.016 s
com.songera.hdp.shell.command.test.UnixTest	1.690 s
com.songera.hdp.shell.command.test.RemoteTest	0.514 s
com.songera.hdp.shell.command.test.HadoopTest	1.346 s
com.songera.hdp.shell.command.test.HdfsTest	13.370 s
com.songera.hdp.shell.command.test.HdfsDfsTest	154.551 s
com.songera.hdp.shell.command.test.MapredTest	207.649 s
com.songera.hdp.shell.command.test.YarnTest	32.770 s