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/

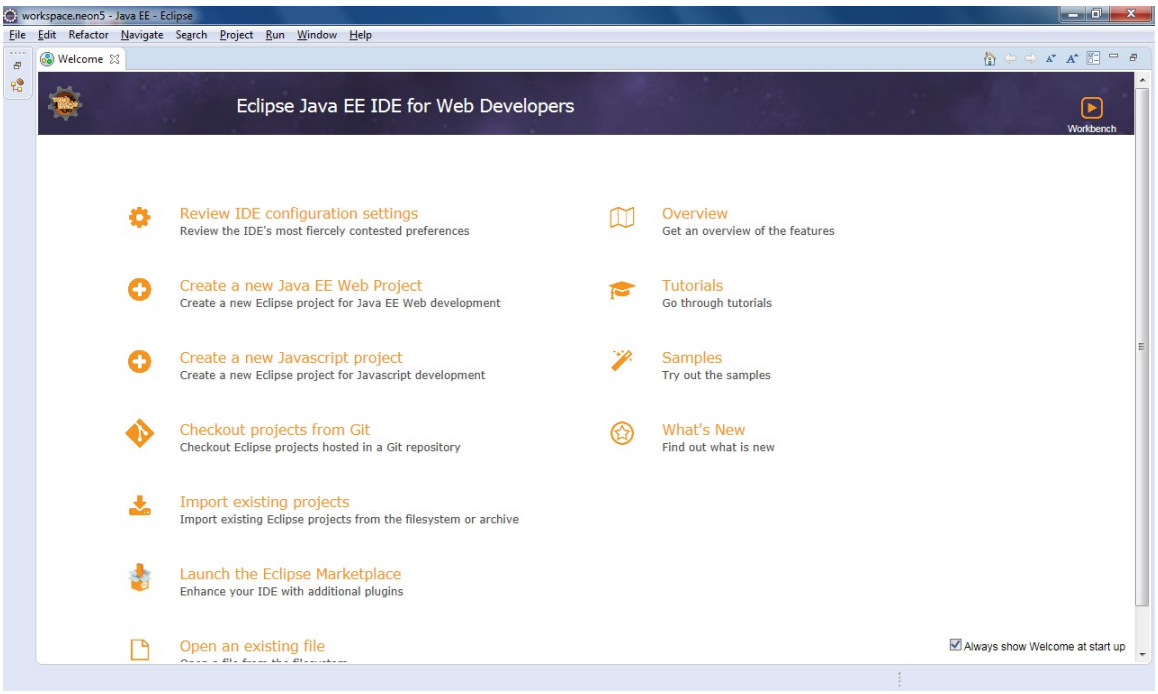
The following instruction will show you how to install and configure the Hadoop Ecosystem in Windows Platform. The installation process for Mac OS X and Linux are similar to that of Windows.

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

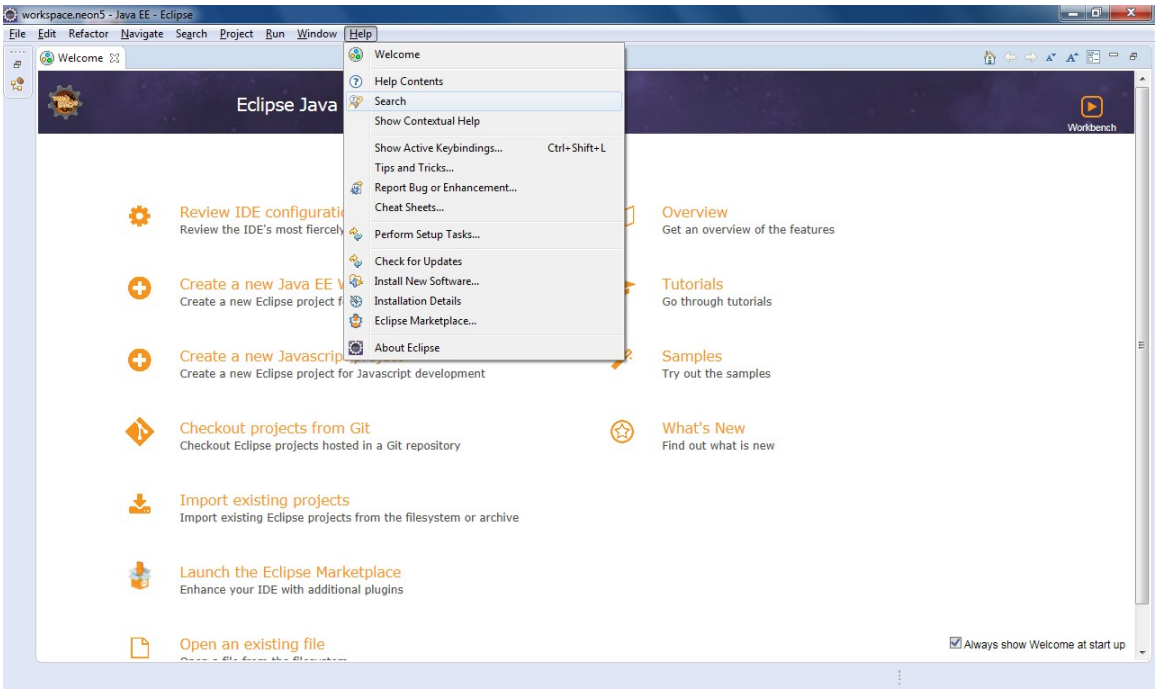
1. Runtime Environment – Install and deploy all Hadoop modules to hosts.
2. Server – Setup each module's configuration files. One Runtime instance allows the setup of multiple sets of server configurations for different developing or testing purposes.

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

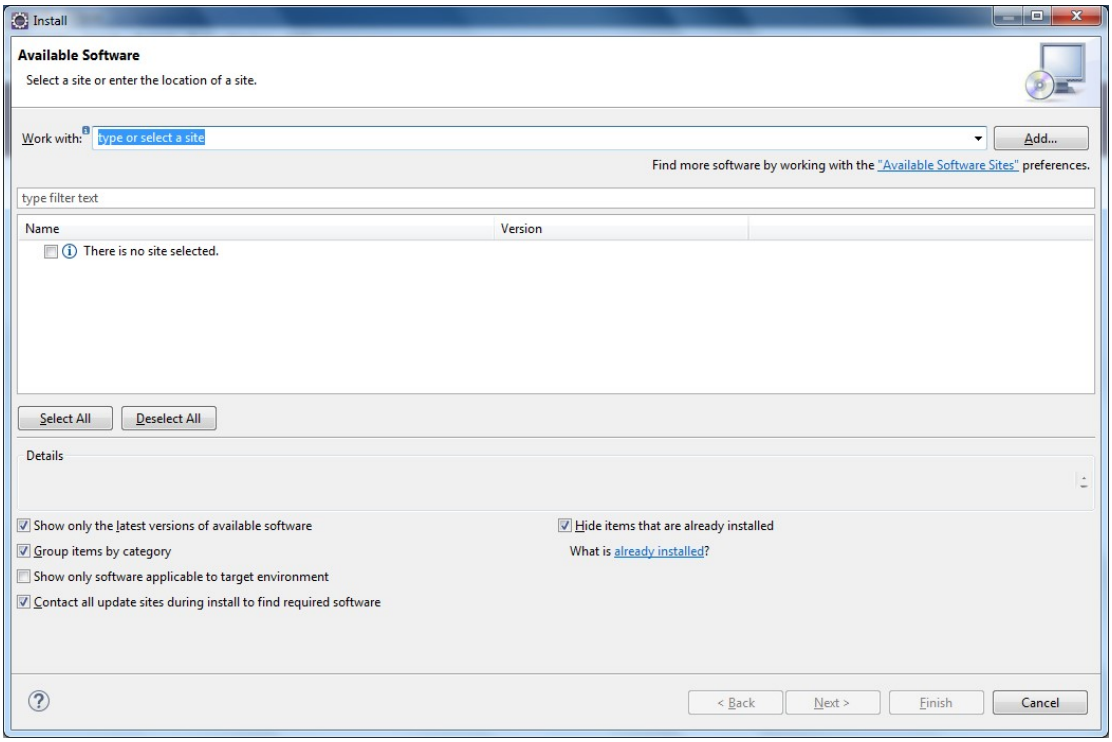
   Start Eclipse:



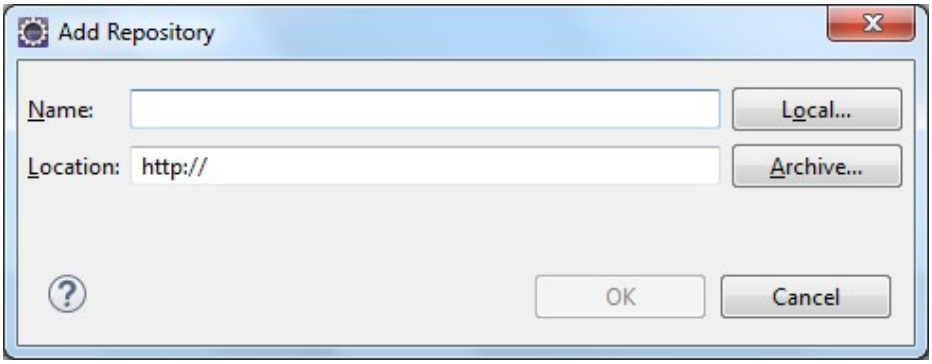
   Launch plugin install window by going to Menu: Help -> Install New Software



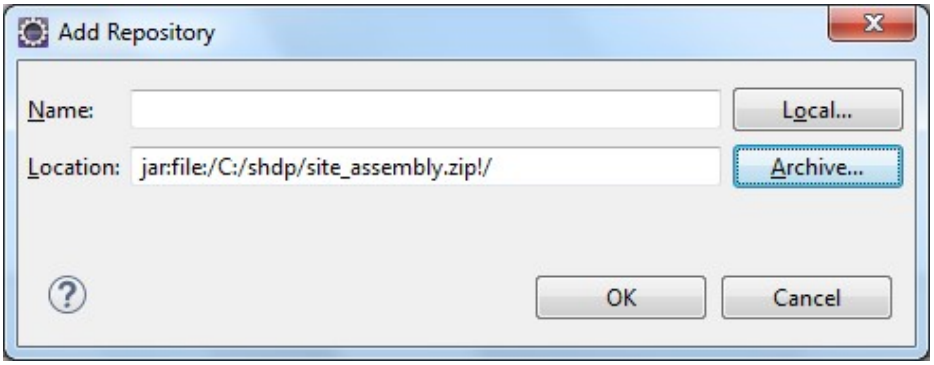
Click the Add Button in the Install window:



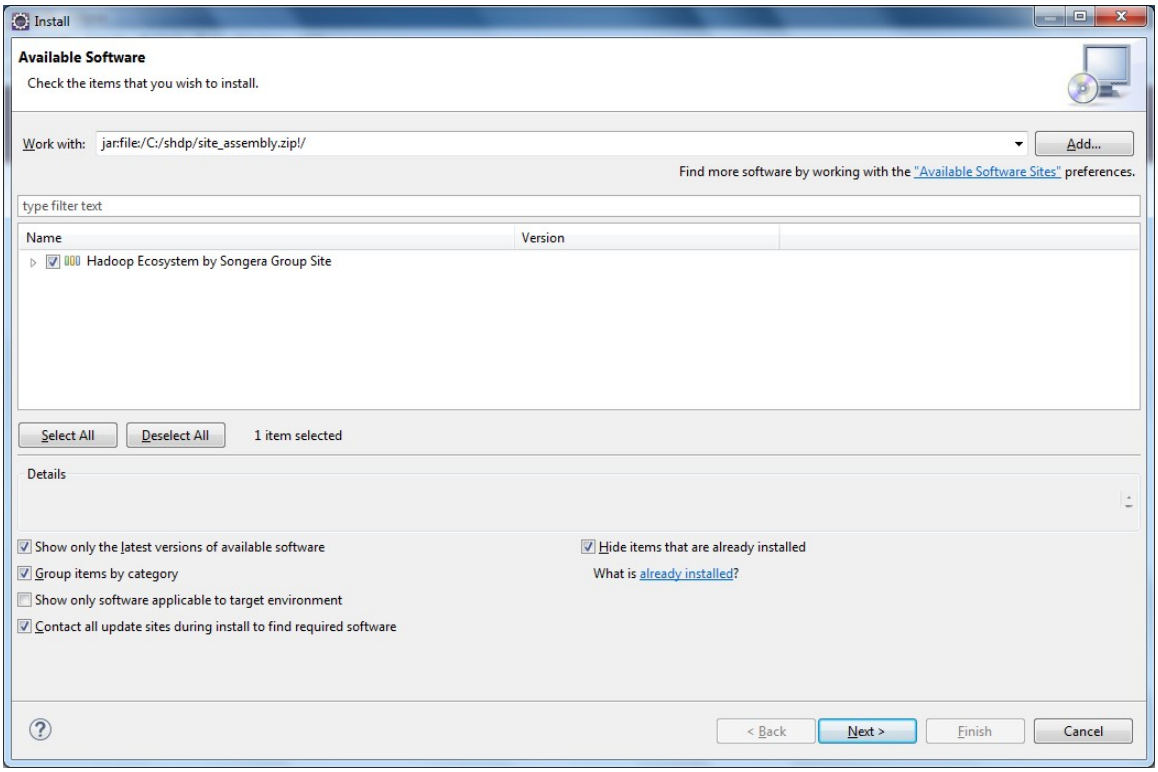
Click Archive button to select site_assembly.zip



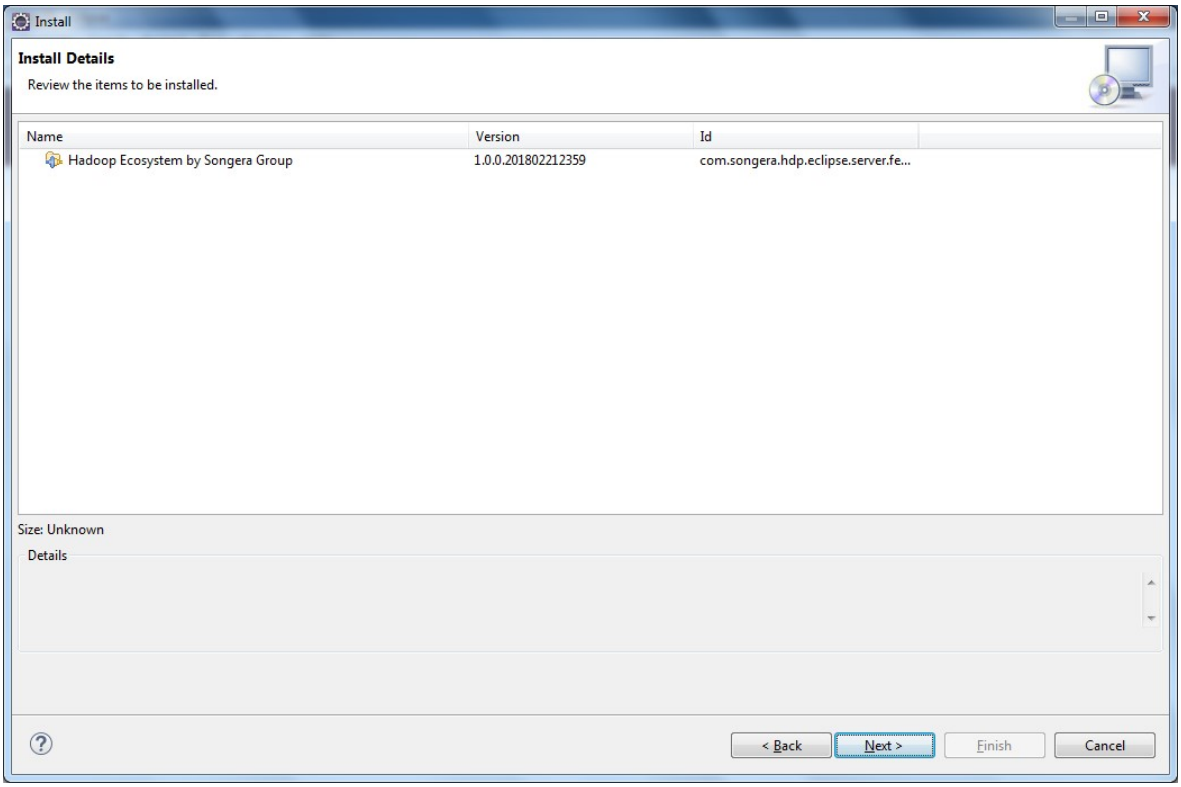
Click OK



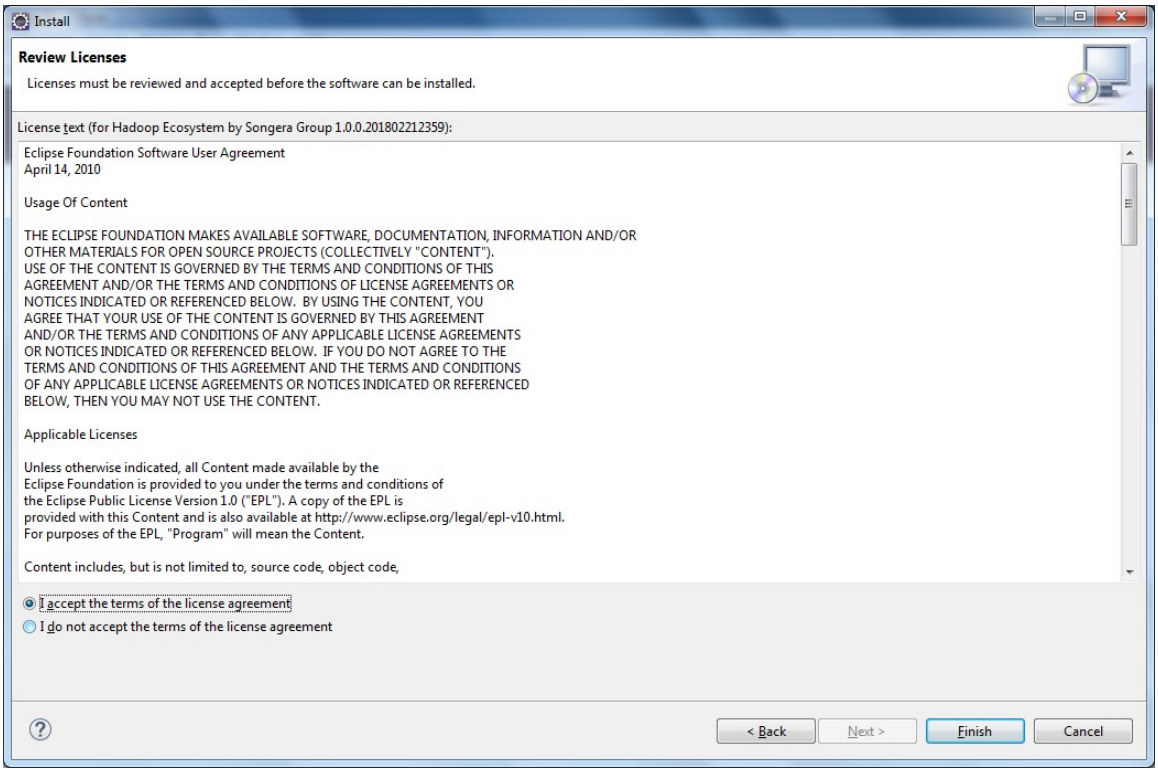
Select “Hadoop Ecosystem by Songera Group Site” and click Next button.



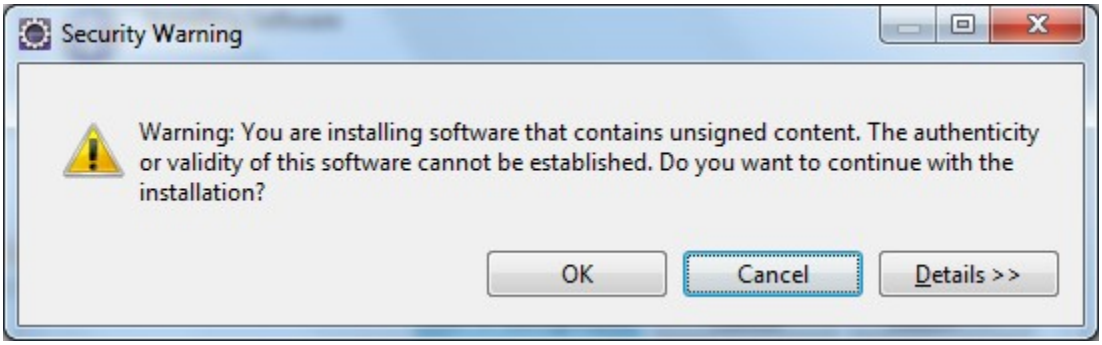
Click the Next Button.



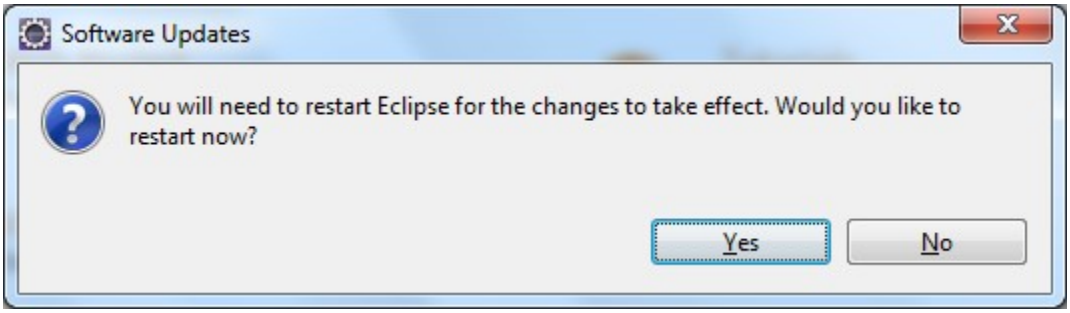
Read Licenses content carefully. To accept the conditions, click the Finish Button



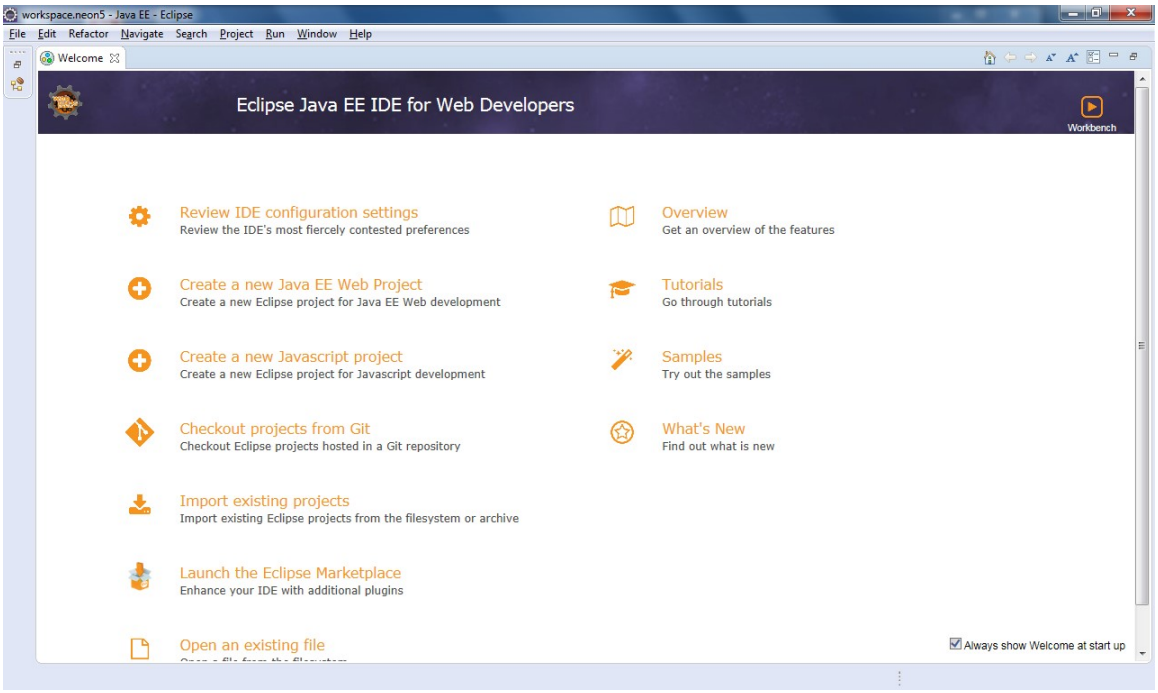
A Warning message will be displayed, as shown below. Click the OK button:



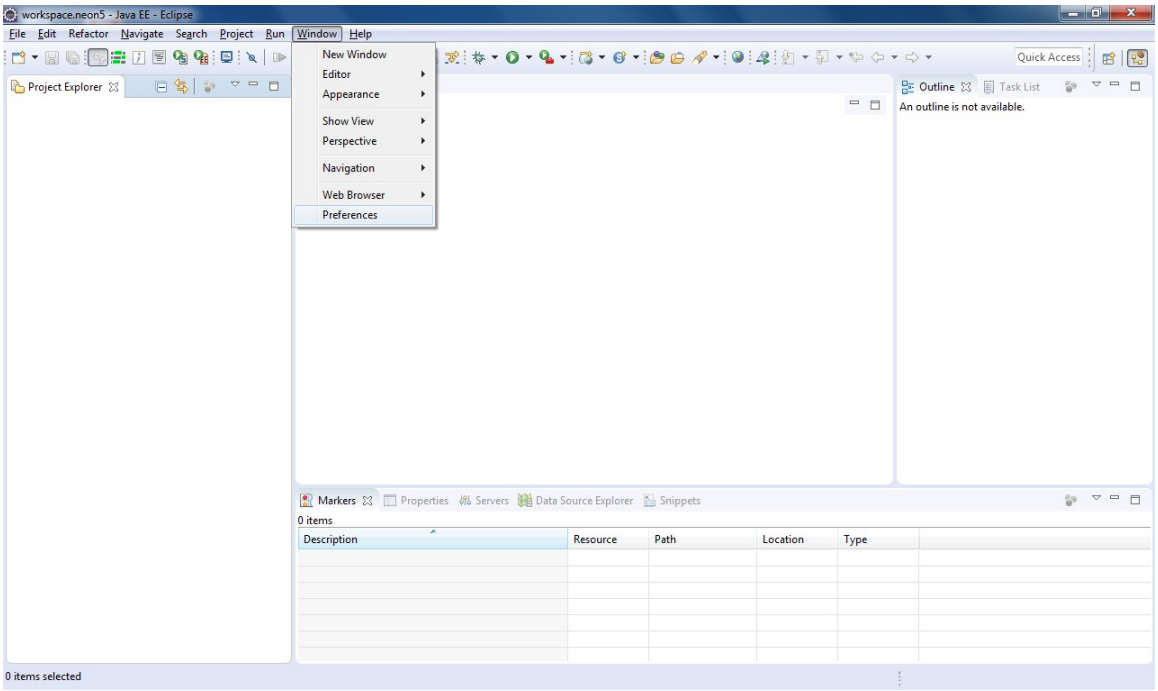
After the plugin is installed, restart Eclipse:



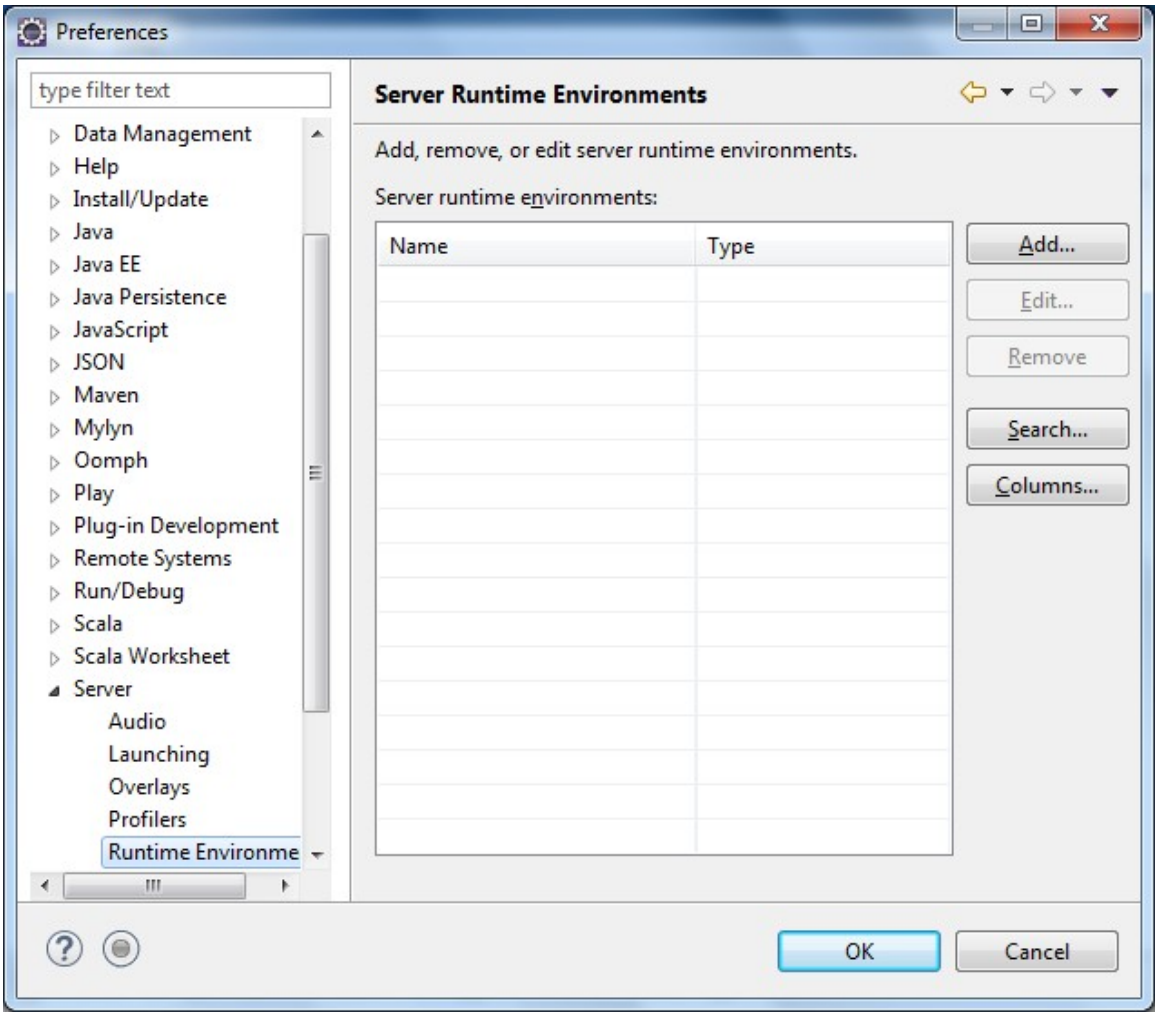
2. Install Hadoop module Runtime environment.
   Start Eclipse Java EE IDE:



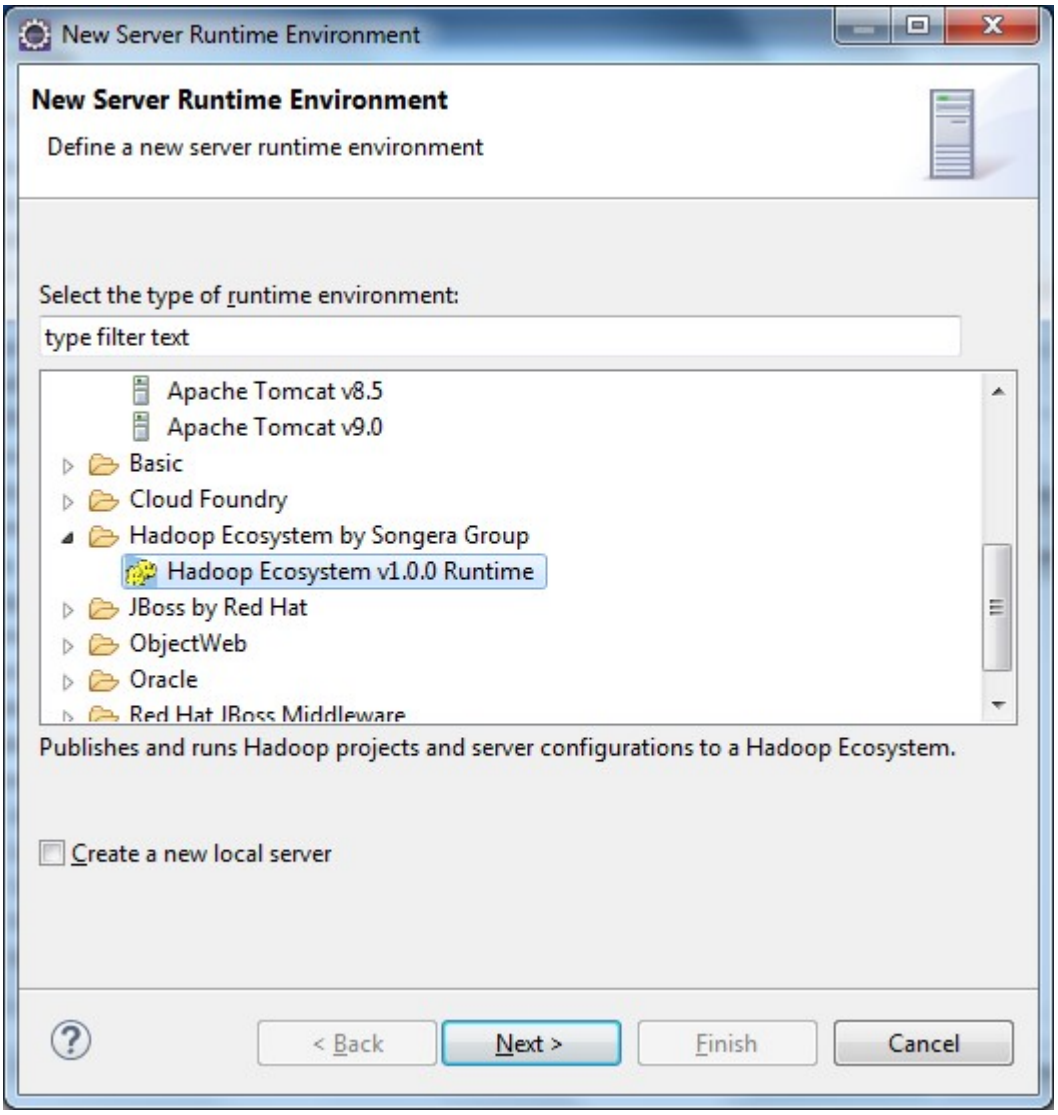
Click Menu: Window -> Preferences



In Preferences Screen, find Server and select “Runtime Environment”, then click Add button.

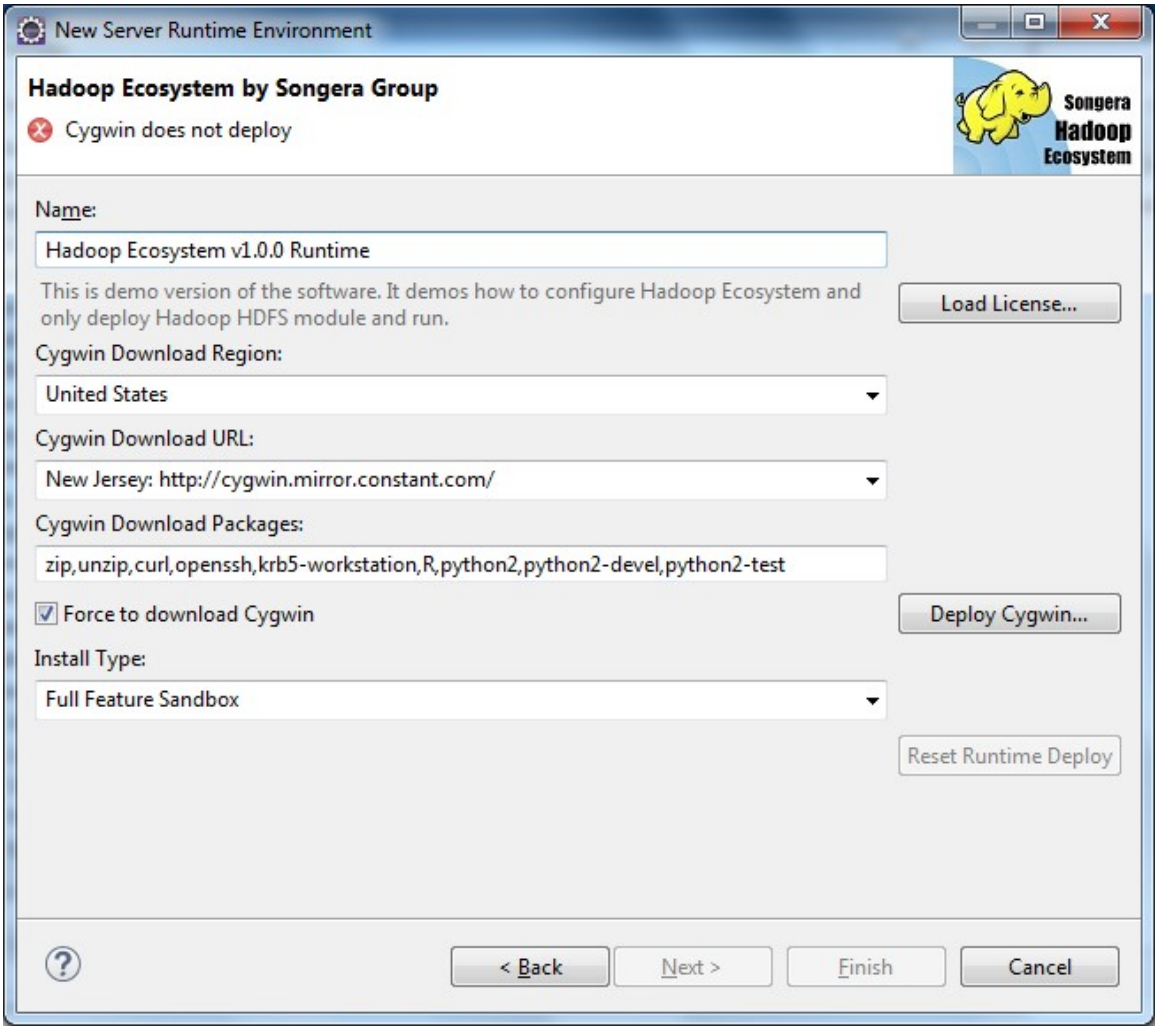


In New Server runtime Enviroment screen, find “Hadoop Ecosystem by Songera Group” and select “Hadoop Ecosystem v1.0.0 Runtime.” Click the Next button.

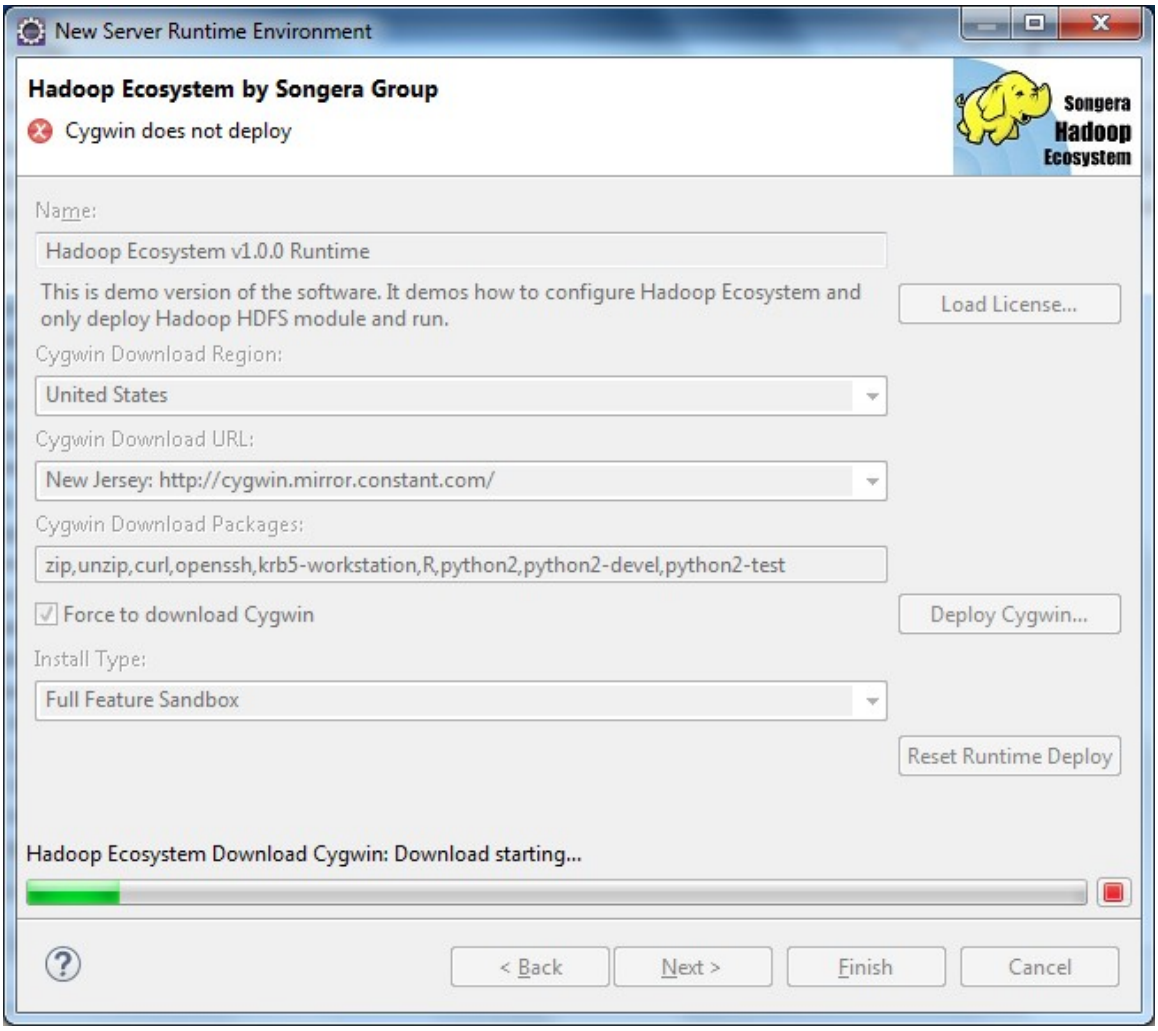


This is the first page of the Hadoop Ecosystem setup. When you receive the license file, you can load the license file by clicking the "Load License…" button. Without the license file, it will run DEMO version with limited functionality.

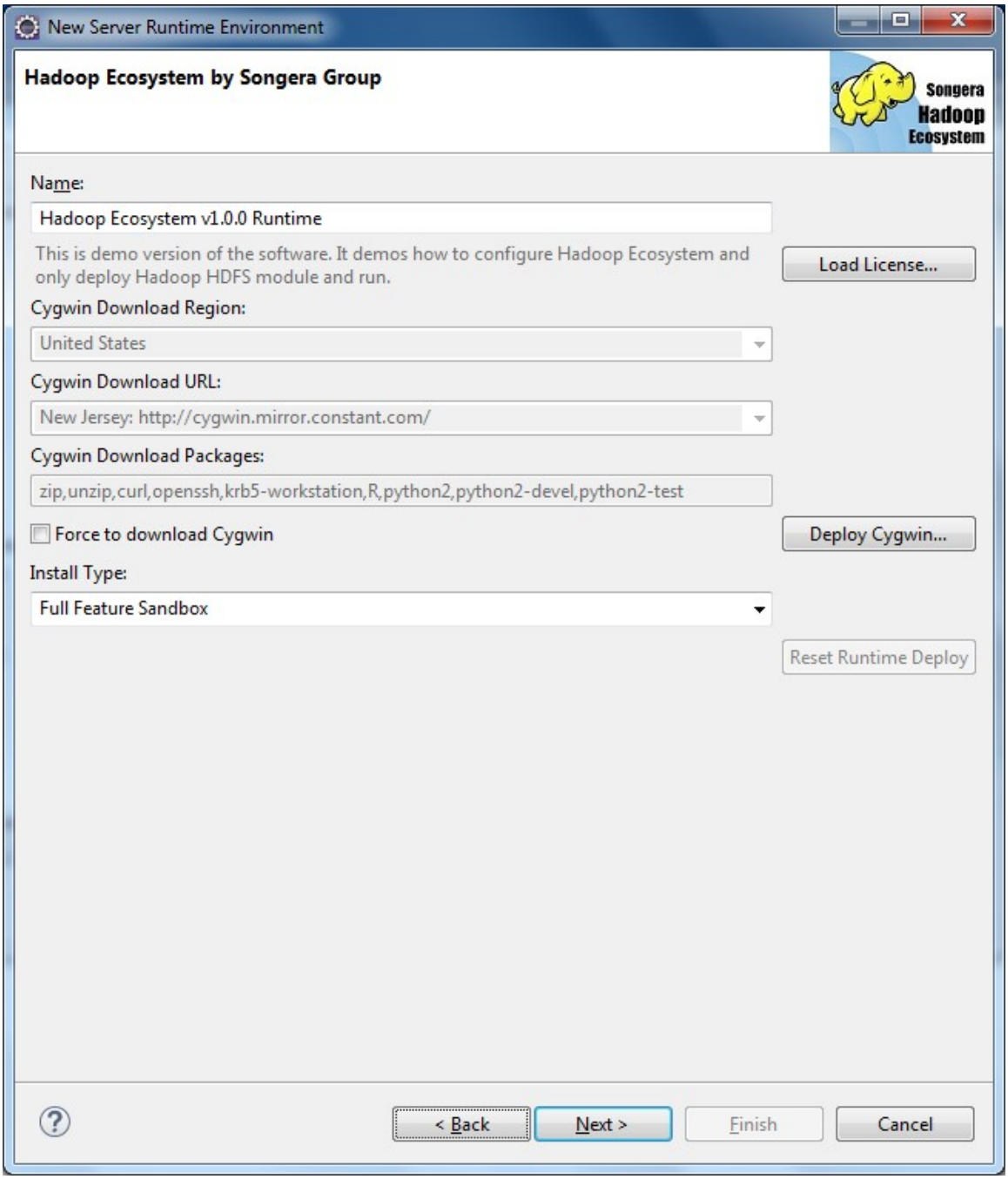
For Windows, the Hadoop Ecosystem requires cygwin software. Select your region and download URL, and then click the "Deploy Cygwin" button to start downloading and deploying cygwin.



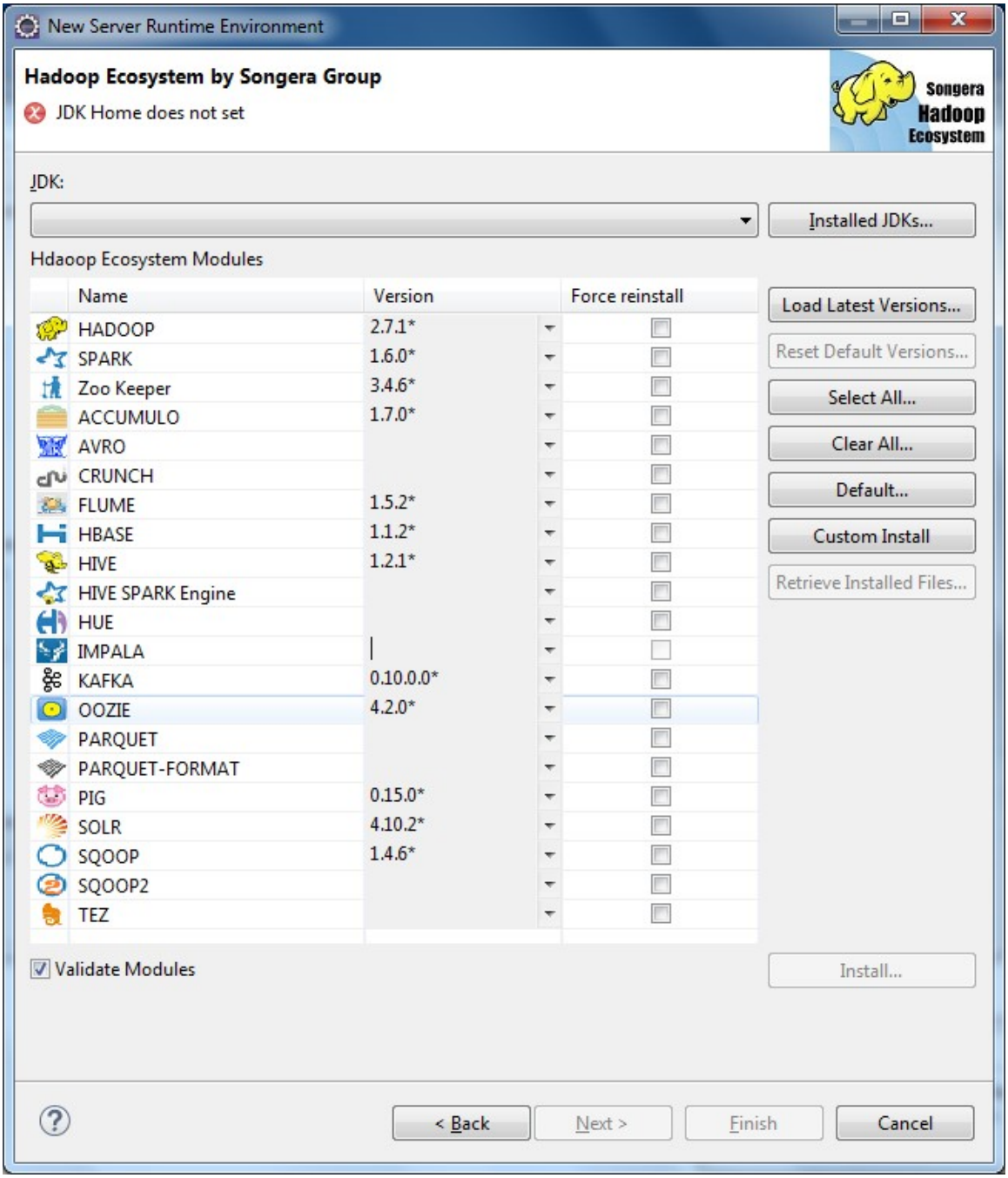
Cygwin install and deploy:



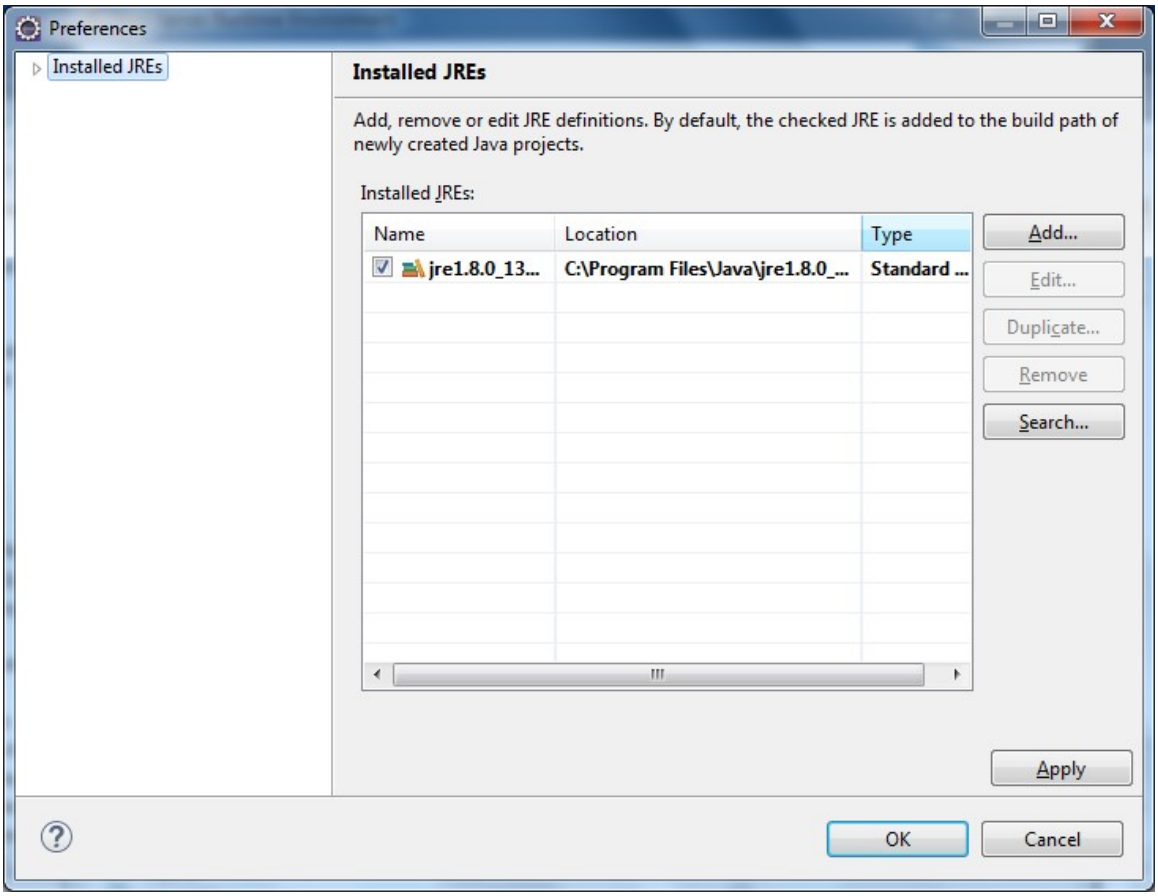
When cygwin finish deploying, the "Next" button will enable the software. Click the Next button to go next page.



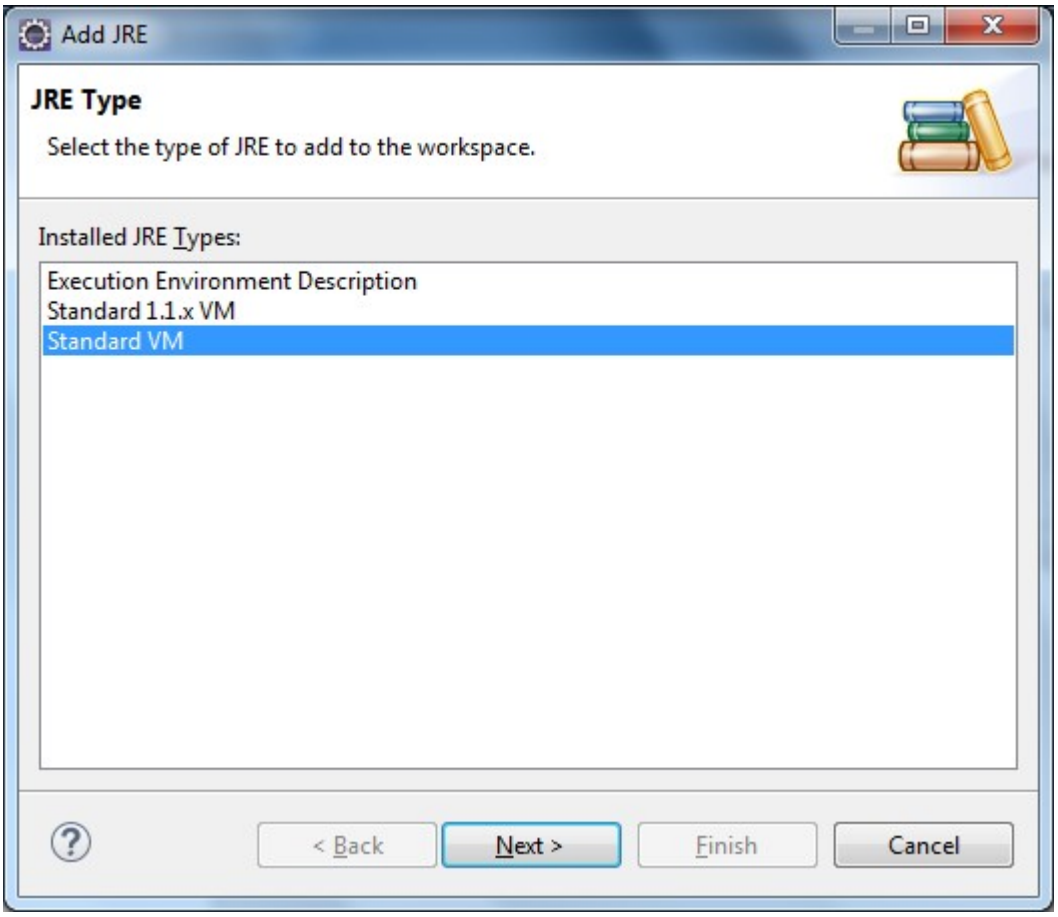
The Hadoop Ecosystem requires you to setup JDK first in eclipse. Click "Installed JDKs …" button



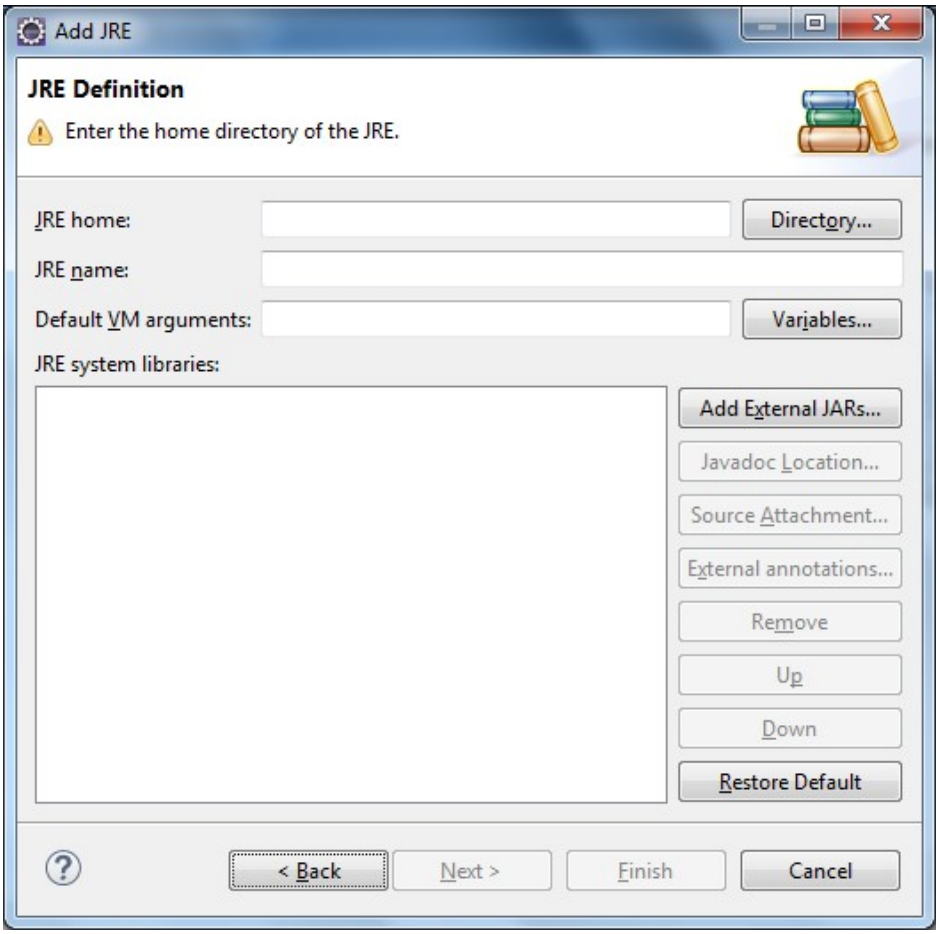
The Eclipse Installed JREs Screen will show up. Click the Add button.

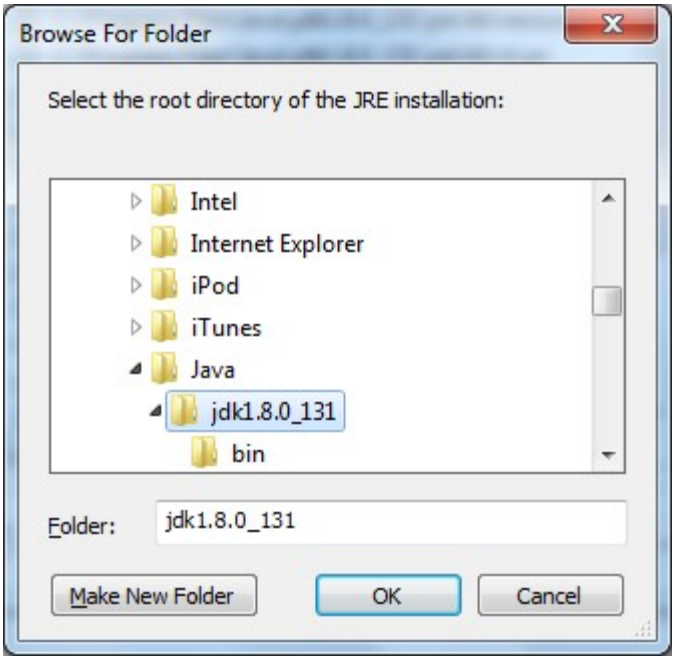


Select Standard VM and Click the Next button

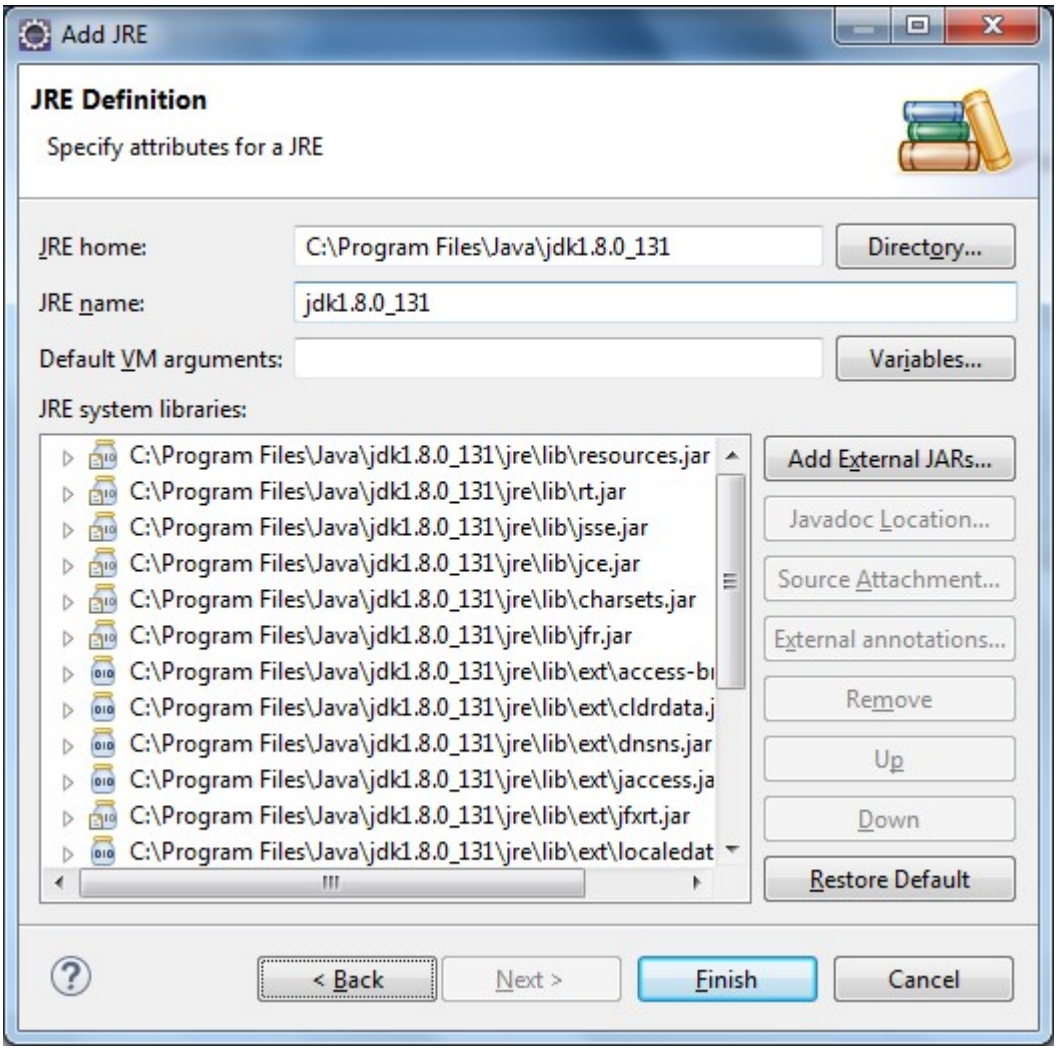


Fill JDK location by “Directory…” or “Variables….”

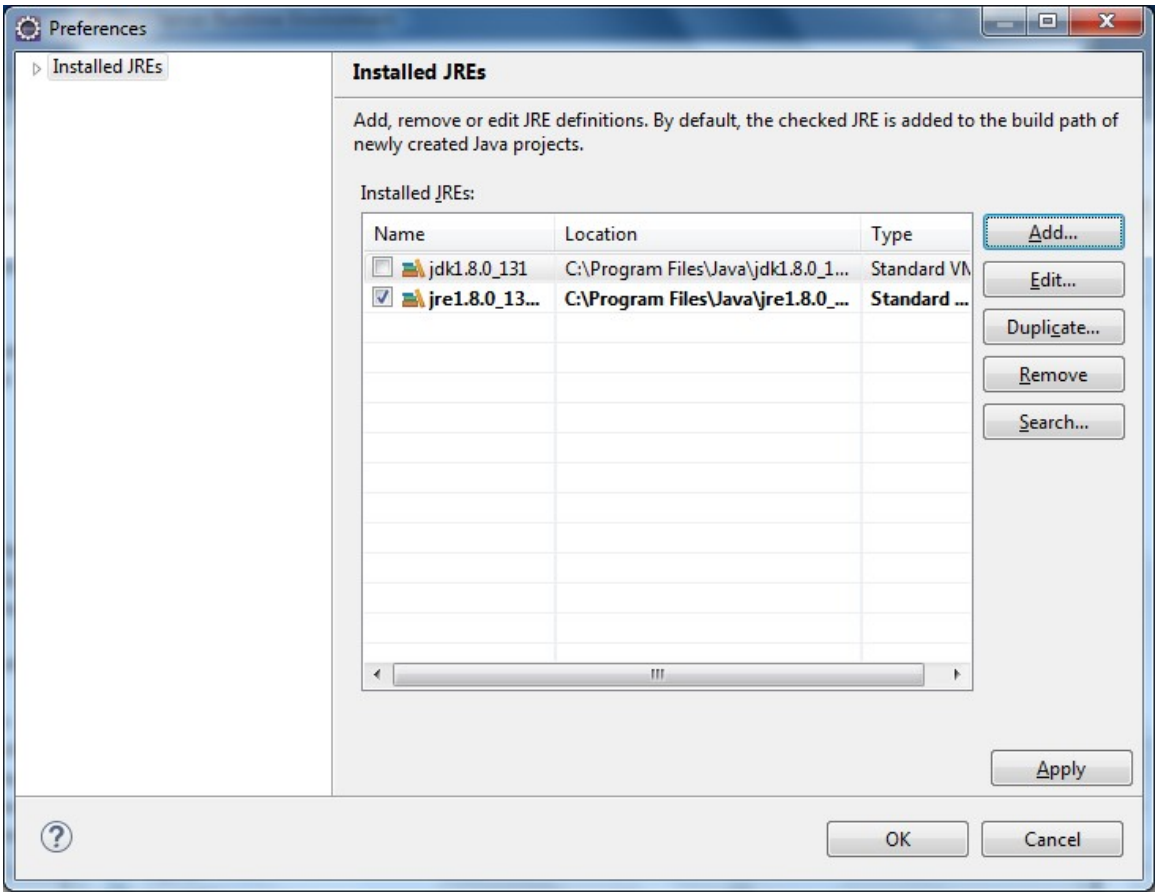




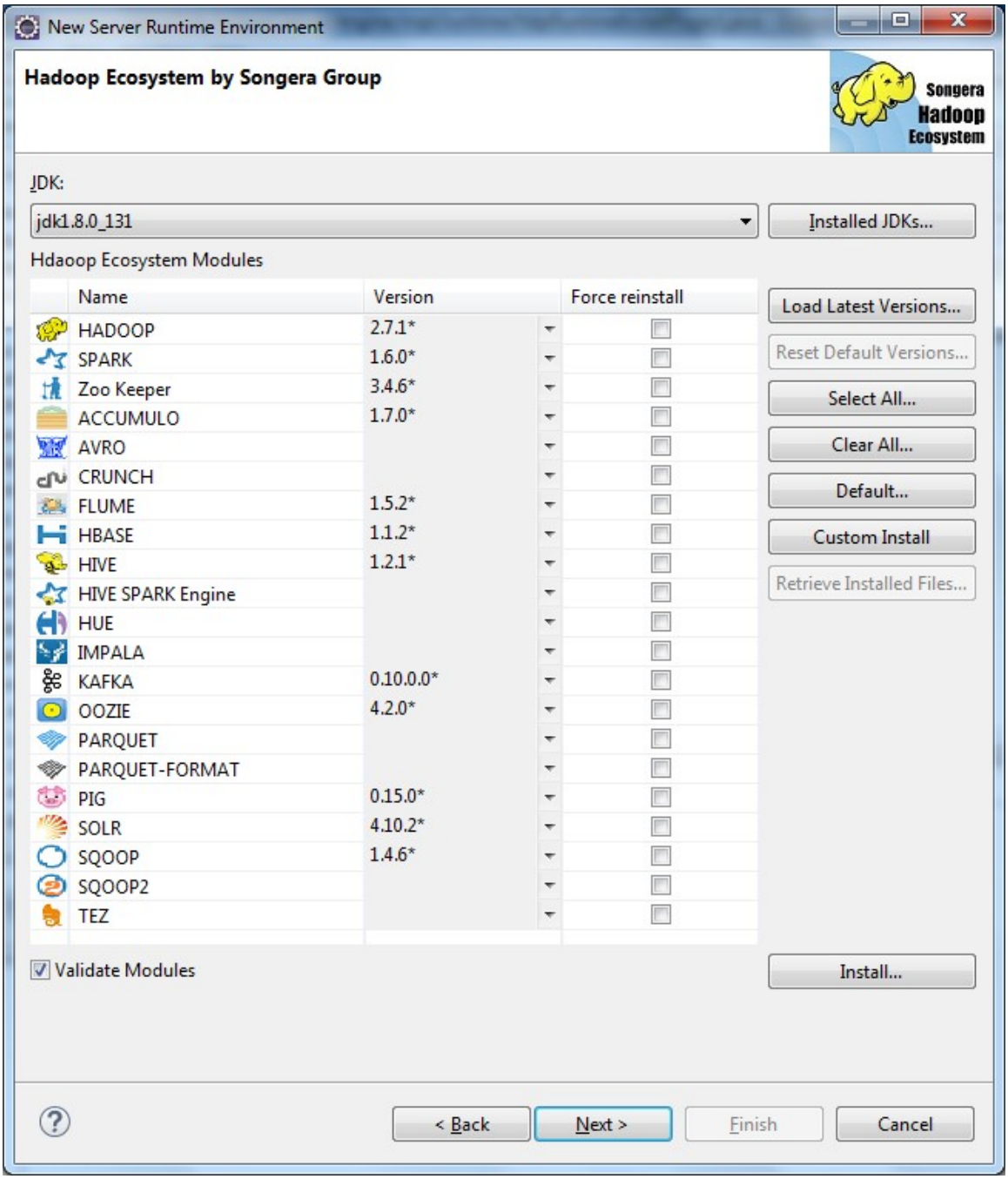
After JDK information is filled, check the Finish button



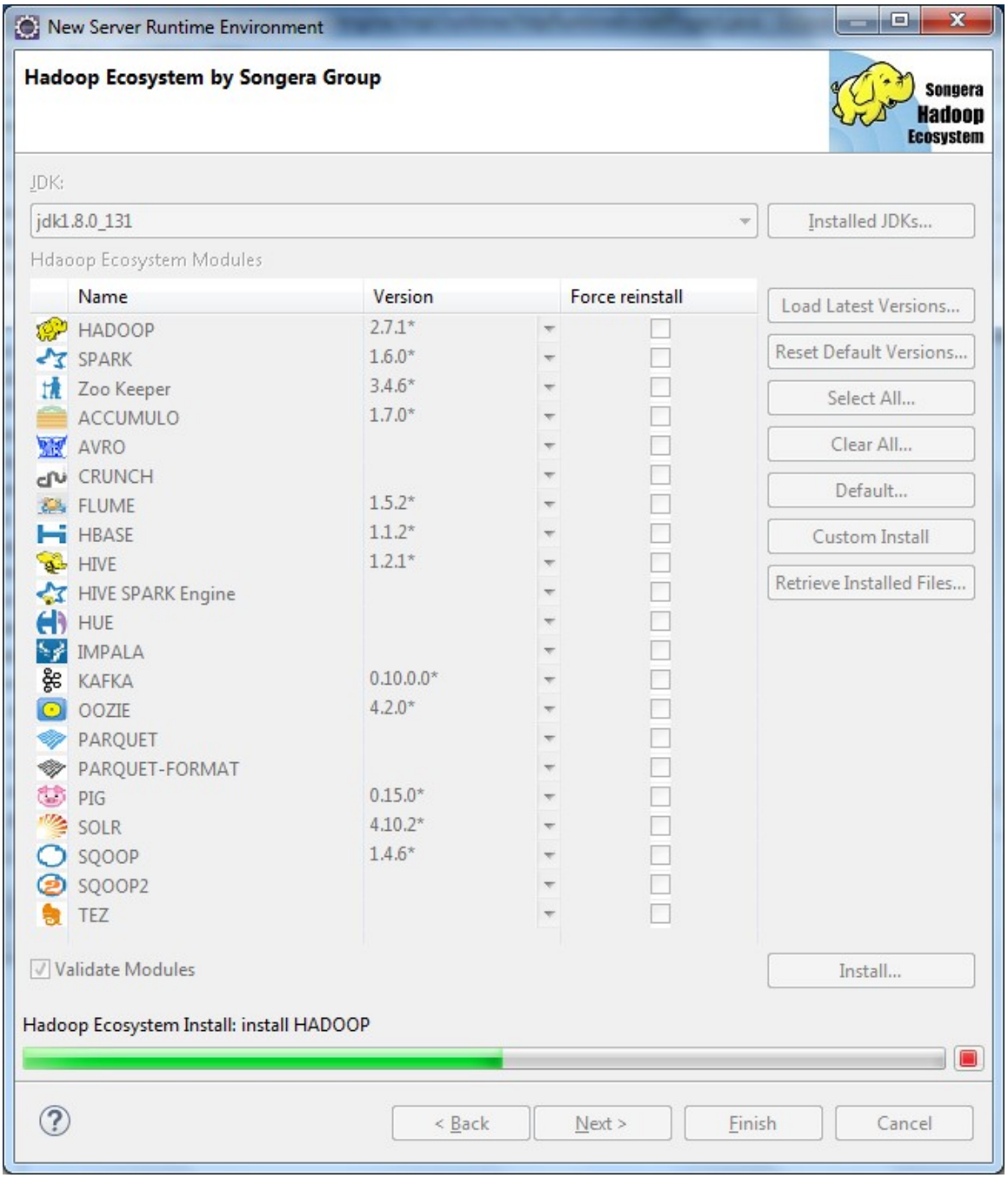
Now, the JDK entry is shown in the list. Click the OK button



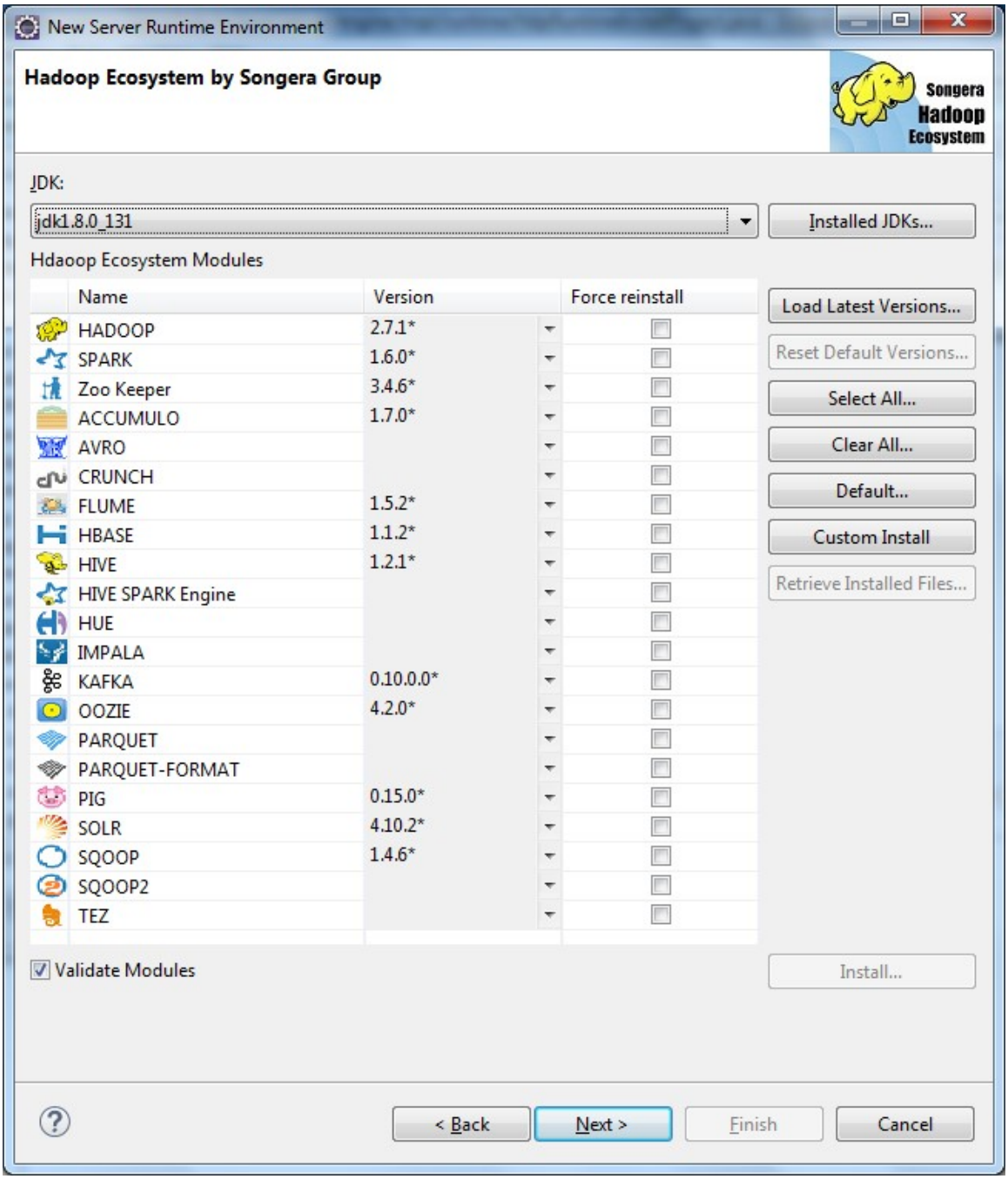
You can select different Hadoop modules or versions to deploy. Then, click the Install button to install the Hadoop module distribution software.



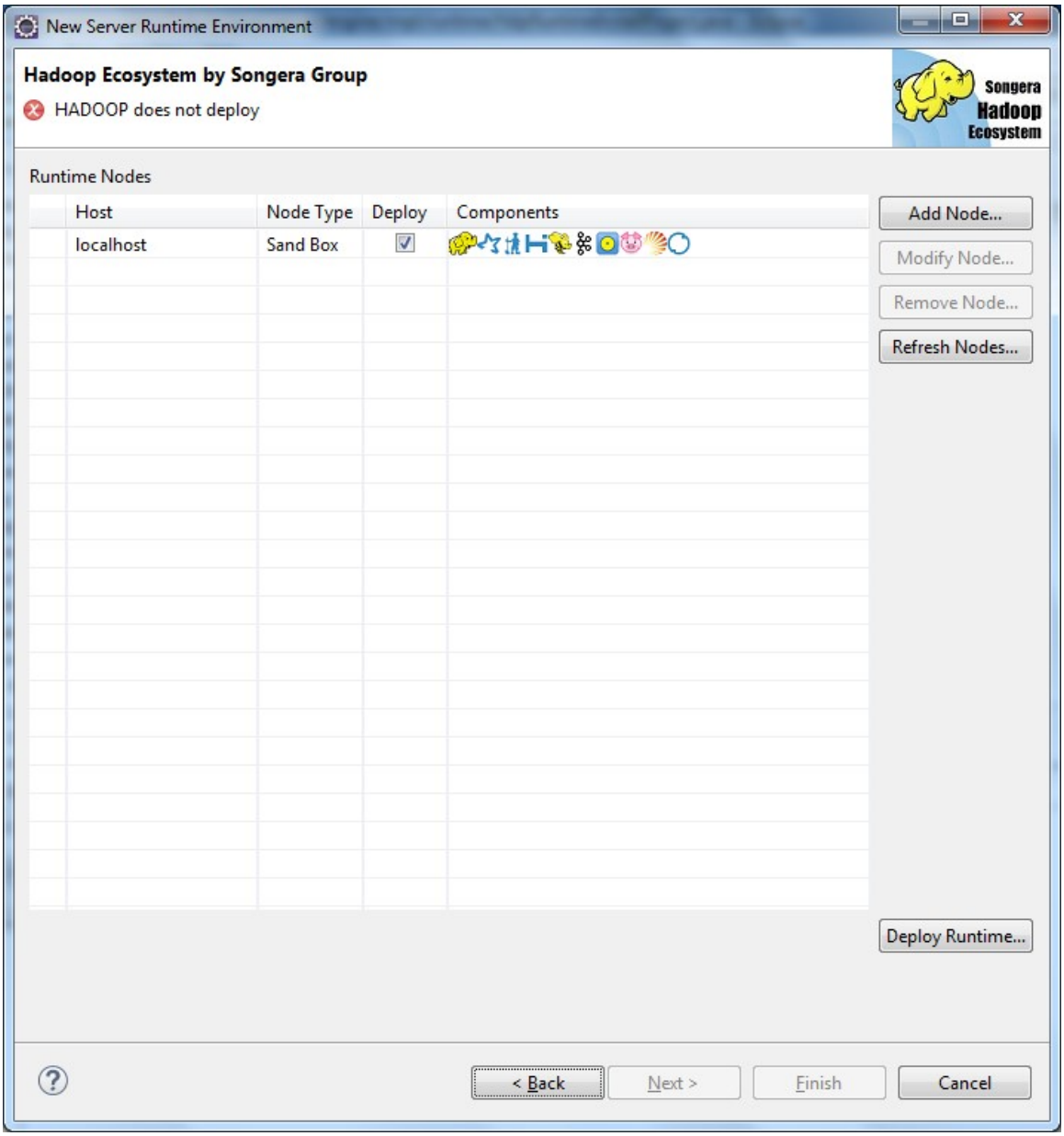
Installation processing might take some time depending on how fast the internet access speed is.



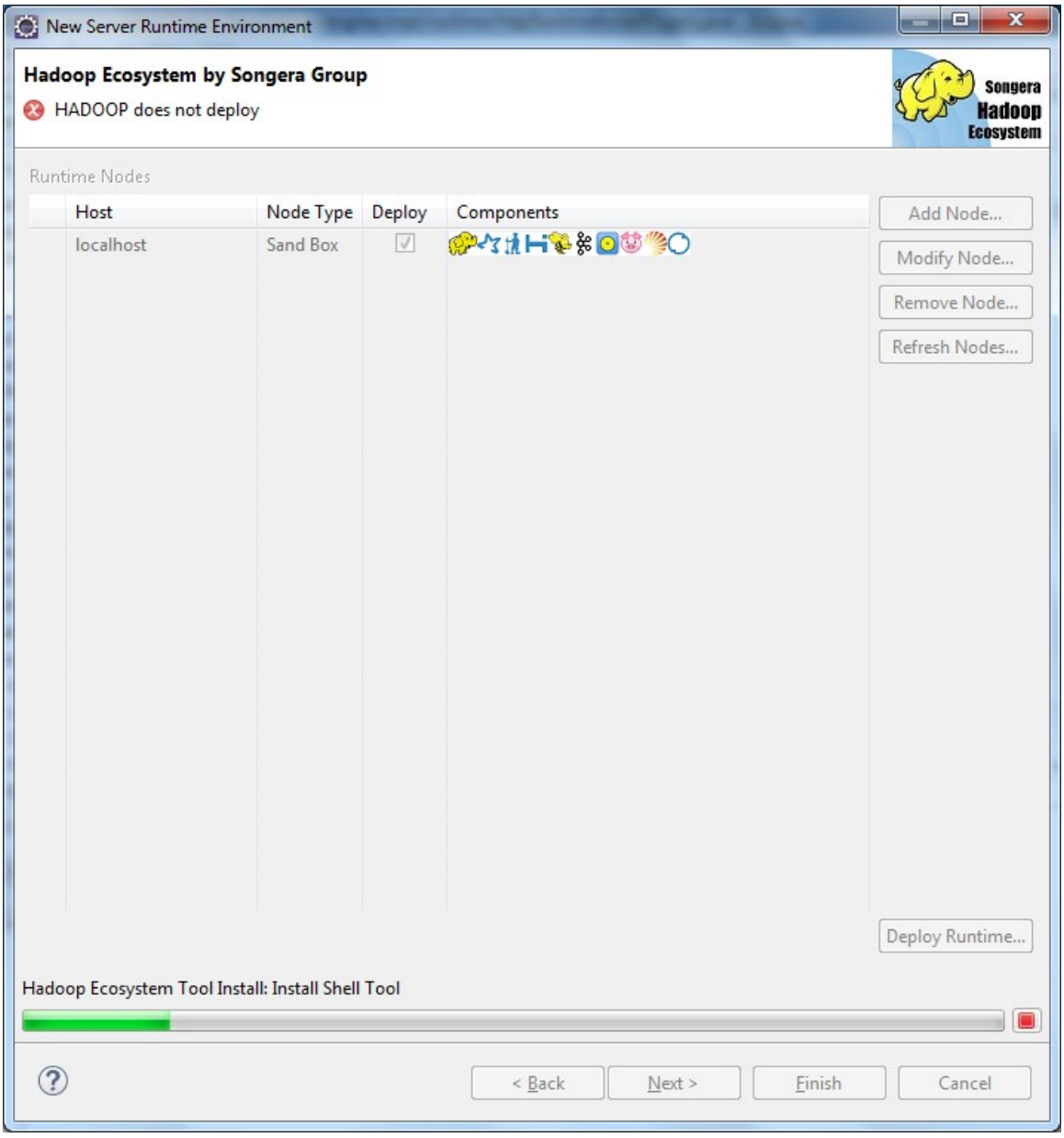
When finished installing, click the Next button.



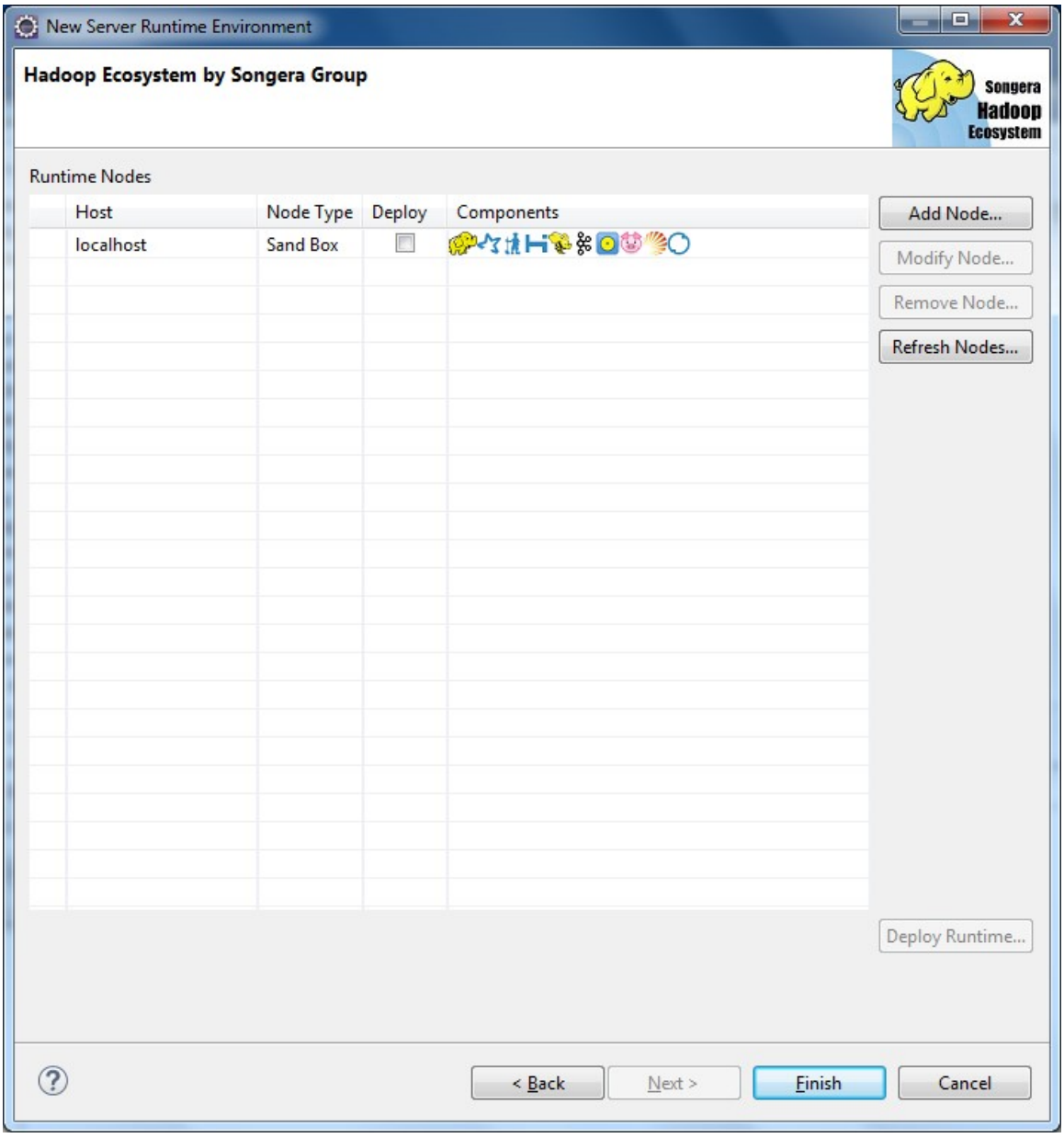
This screen allows you to configure where to deploy Hadoop modules. By default, it deploys to local host. Click "Add Node" button if you need to deploy to a different host. If you click the "Deploy Runtime" button, it deploys the Hadoop module distribution to the configured host.



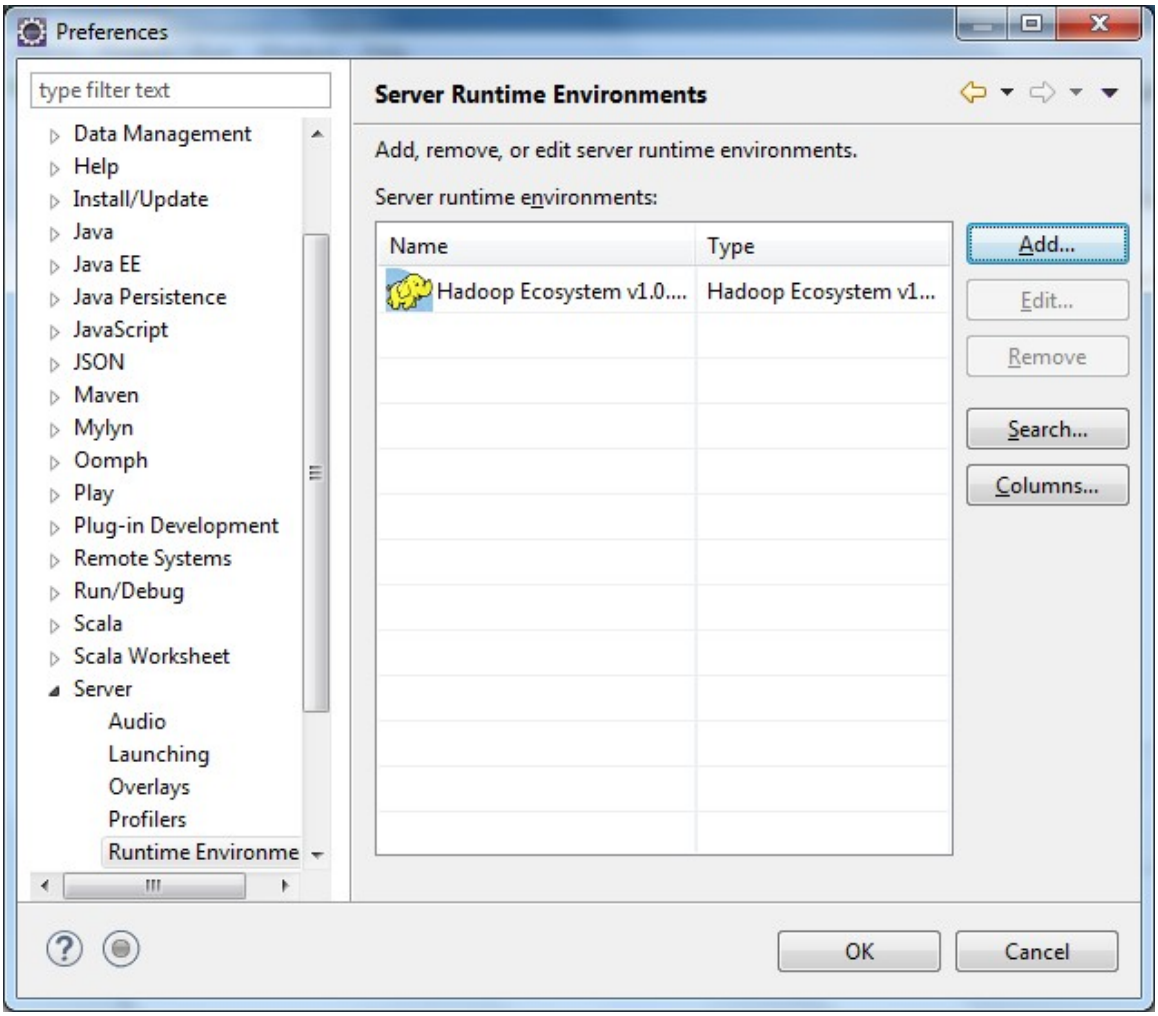
Deploy processing takes time. The progress bar shows the current status.



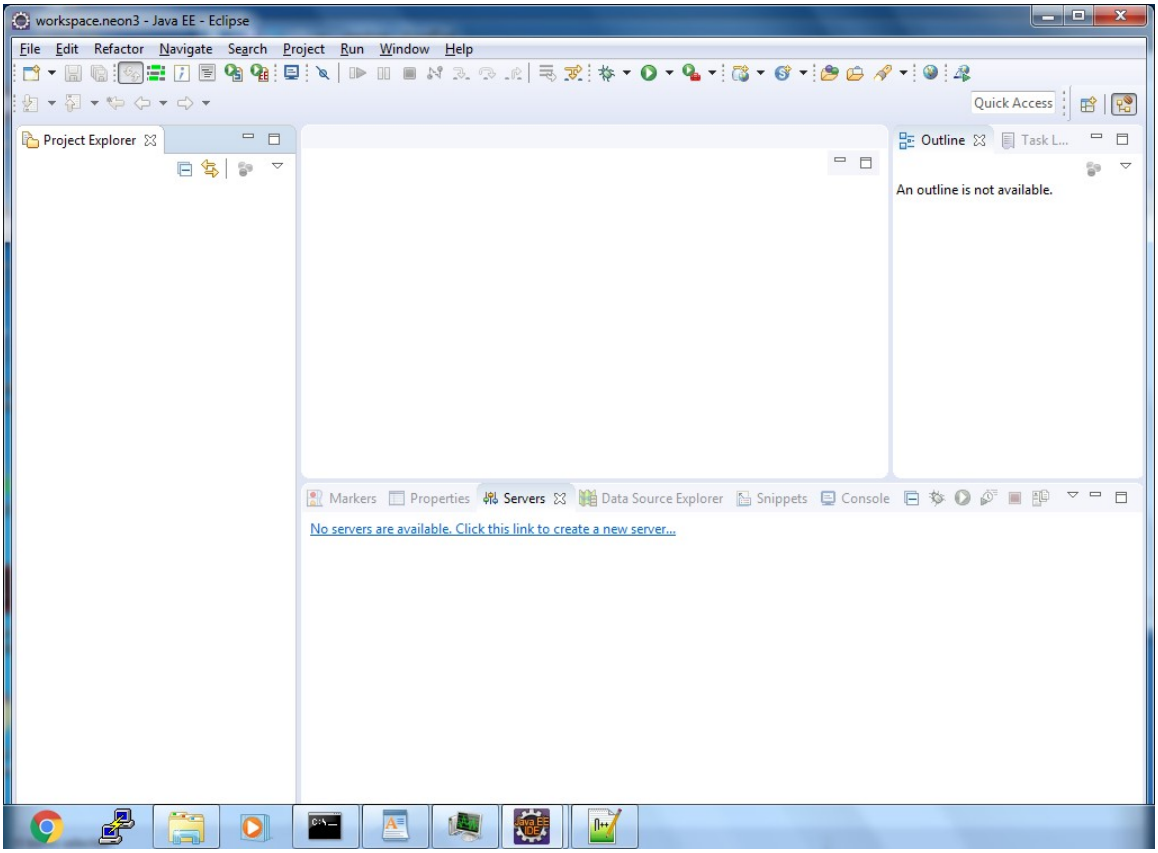
Now click the Finish button to complete the deployment of the Runtime environment



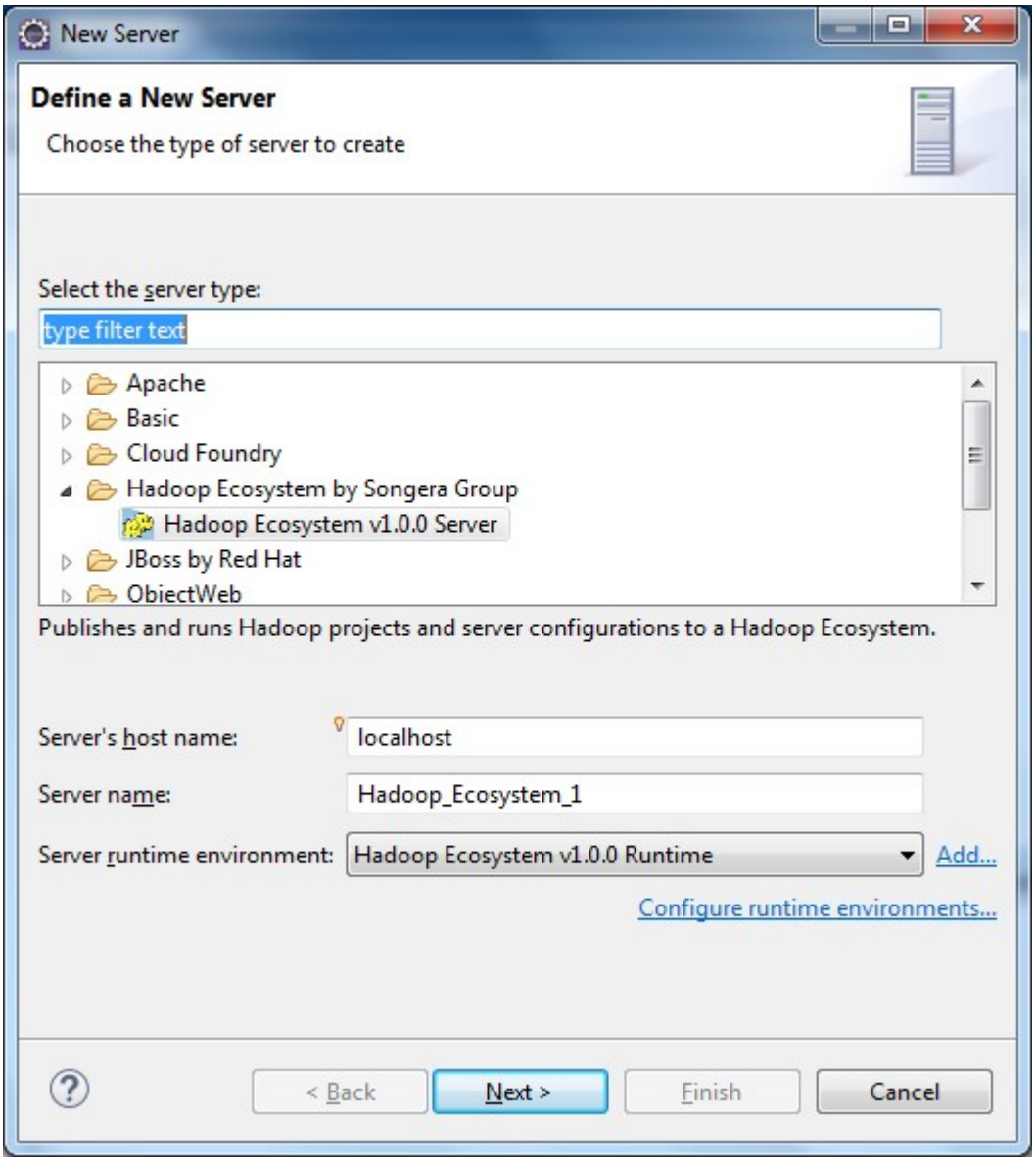
Click the OK button to close the Preferences screen.



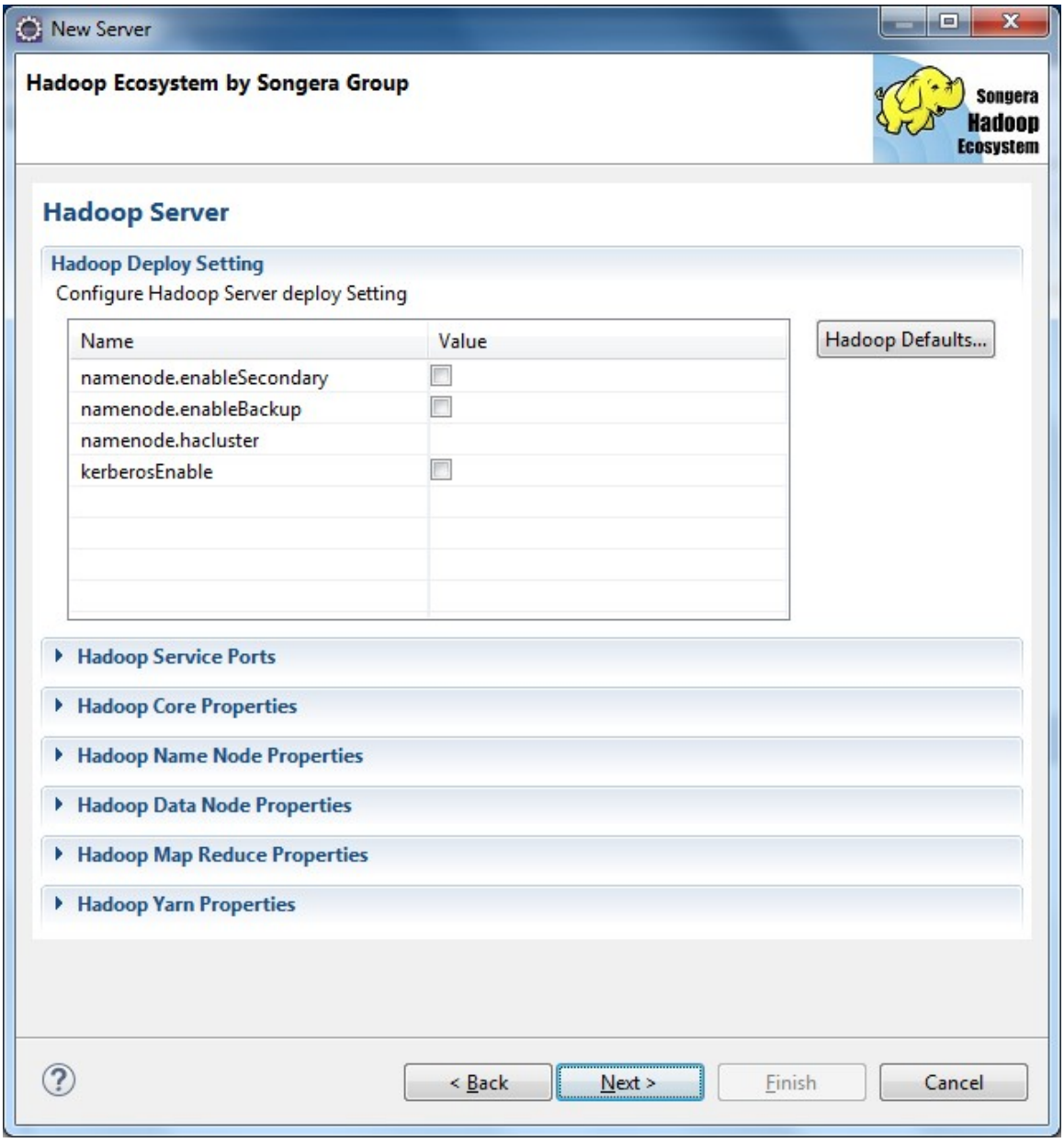
3. Server setup
   After the Runtime environment is deployed, Select eclipse Servers tab and click “No servers are available. Click this link to create a new server…”



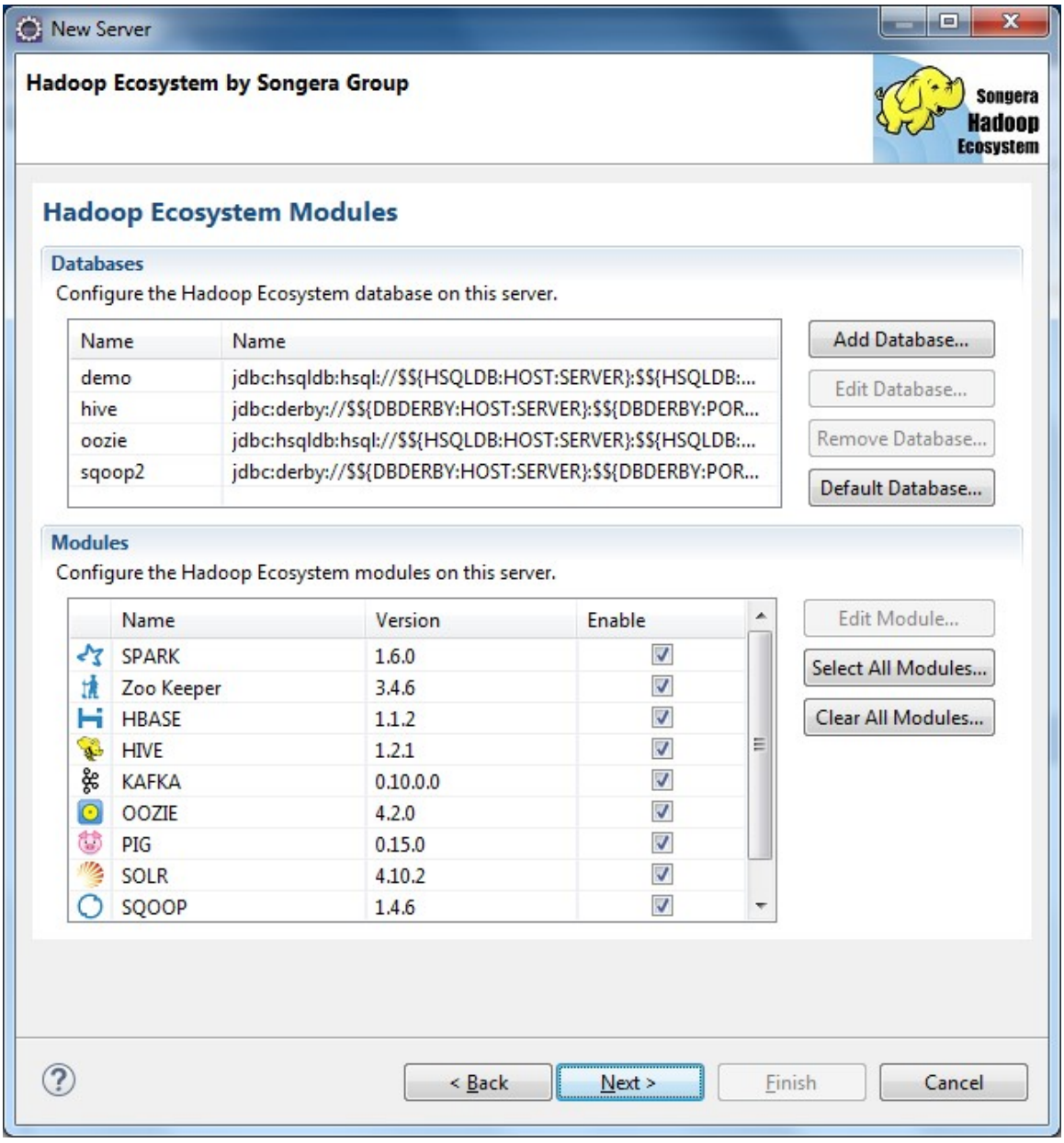
On the New Server dialog, find the "Hadoop Ecosystem by Songera Group" tree item and select "Hadoop Ecosystem v1.0.0 Server," and then click the Next button.



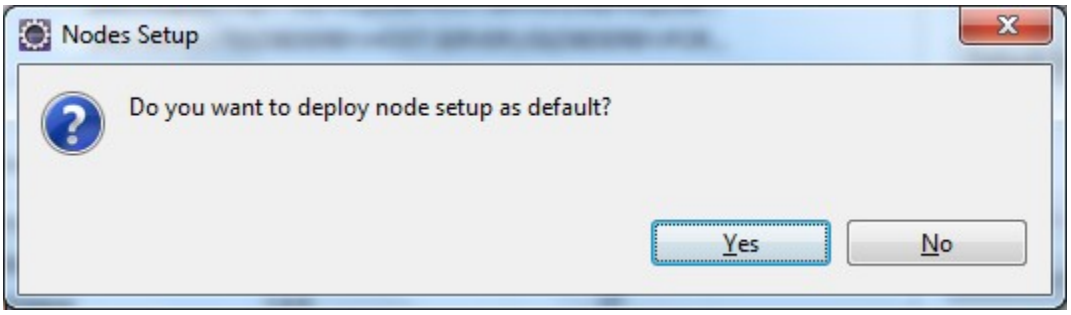
This “Hadoop Server” screen allows you to configure Hadoop Core, HDFS, Mapred and Yarn. Click the Next button when finished changing the configurations.



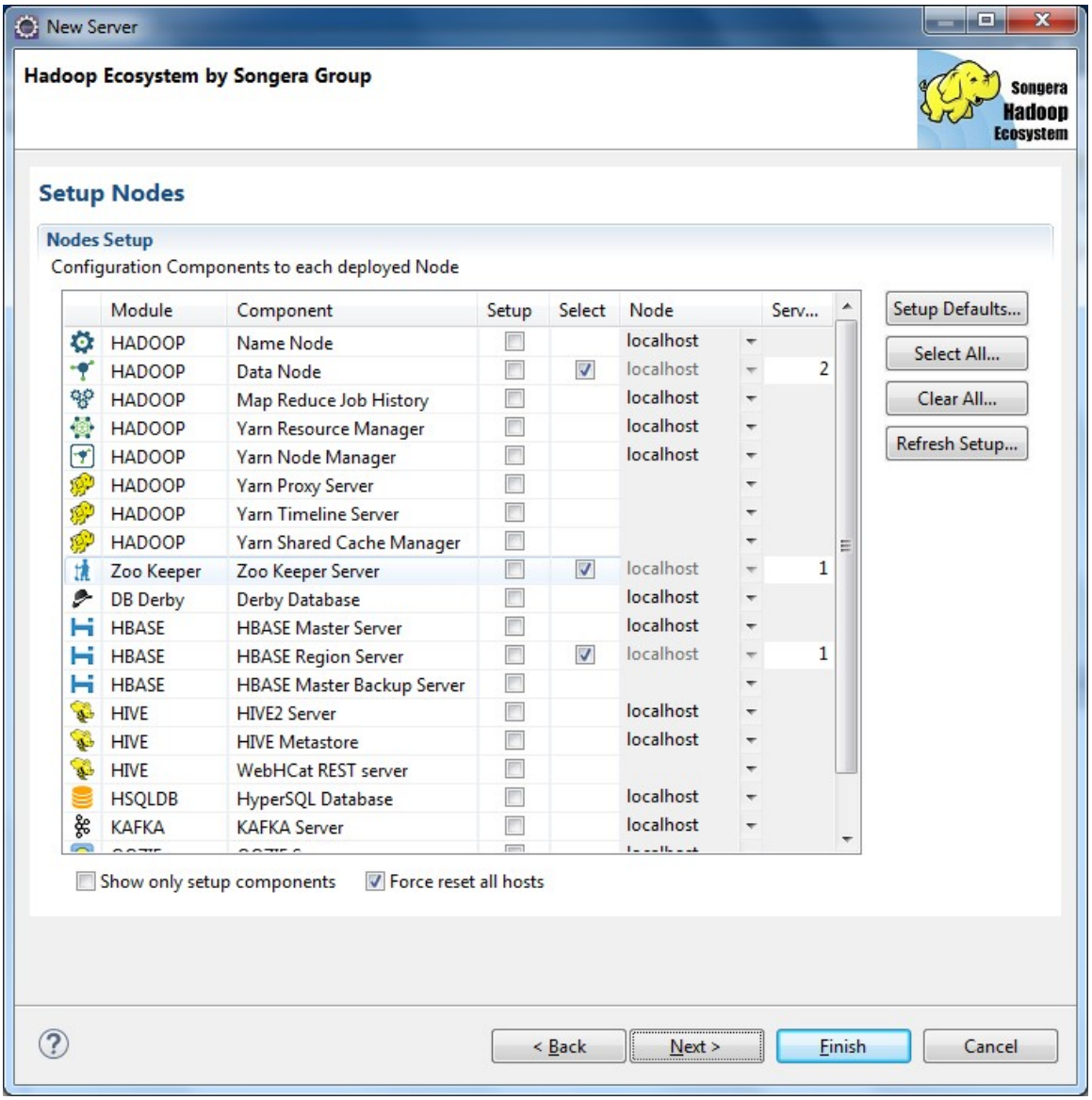
This screen allows you to change the database and Hadoop Module setup. Click the Next button when finish with the setup.



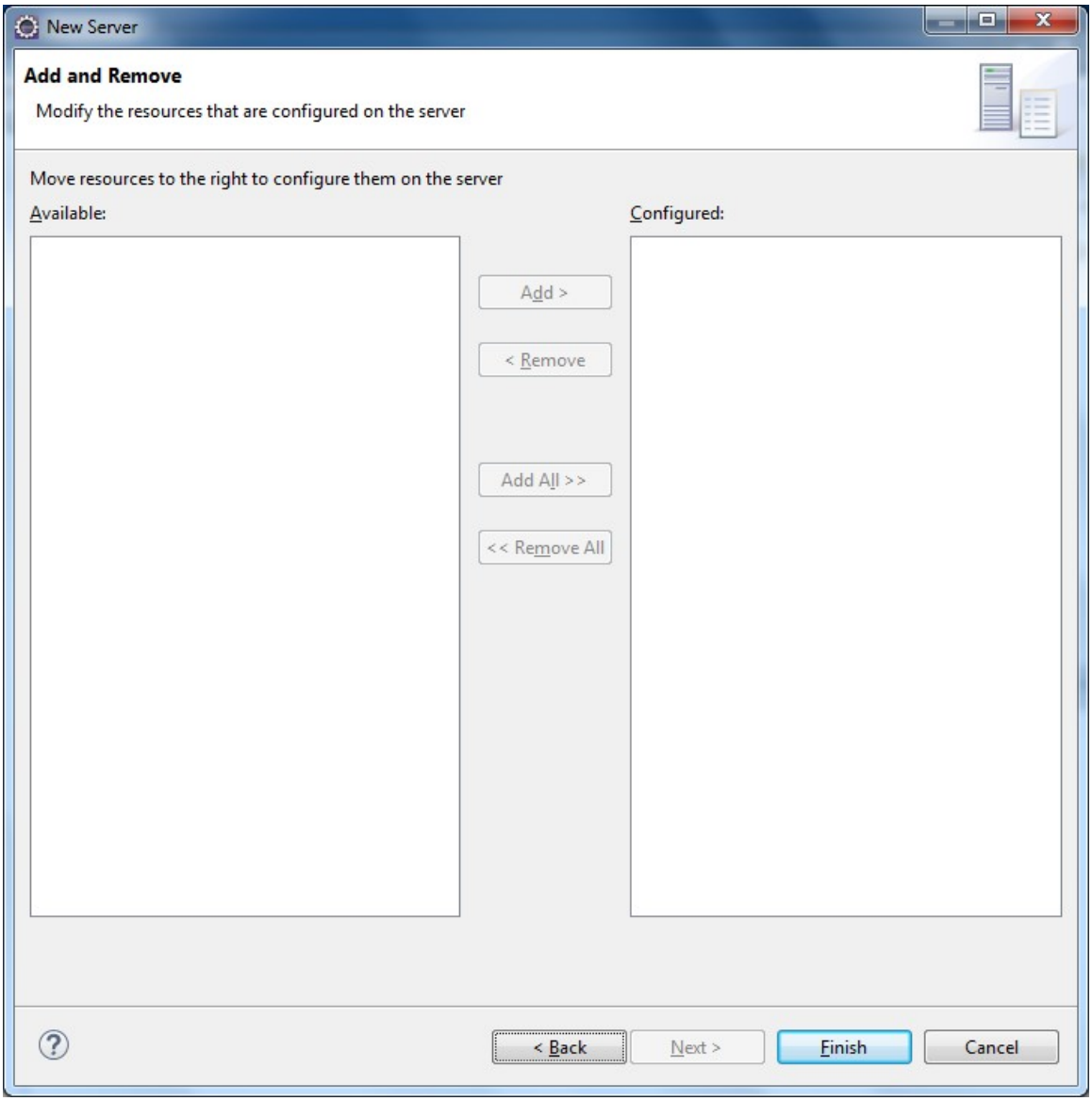
An alert screen will prompt ask you if you want to deploy the settings. Click Yes to allow the Hadoop Ecosystem to use them as the default.



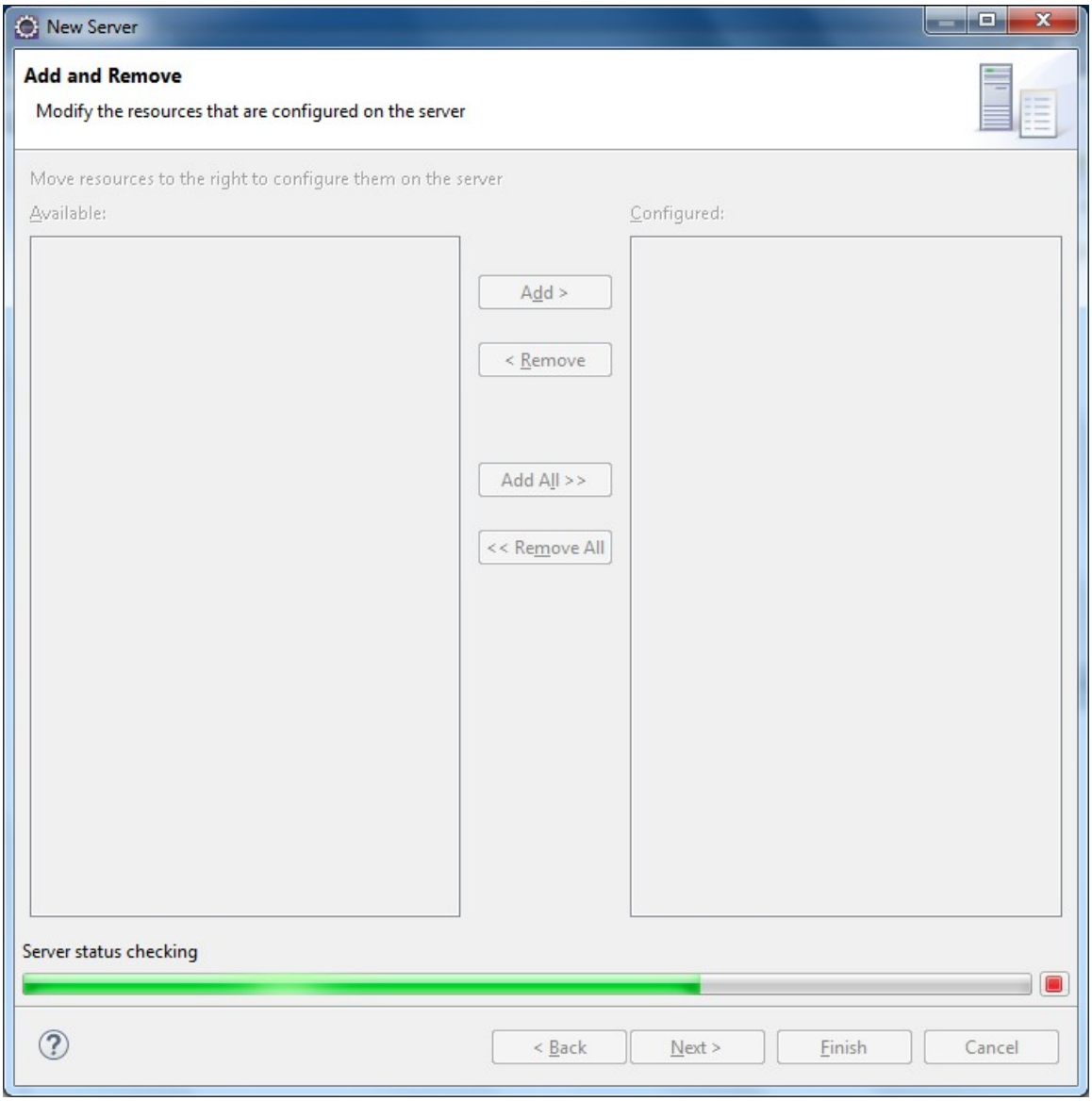
The Setup Nodes Screen displays information on where each Hadoop module service is hosted. Click Next when finish with changes.



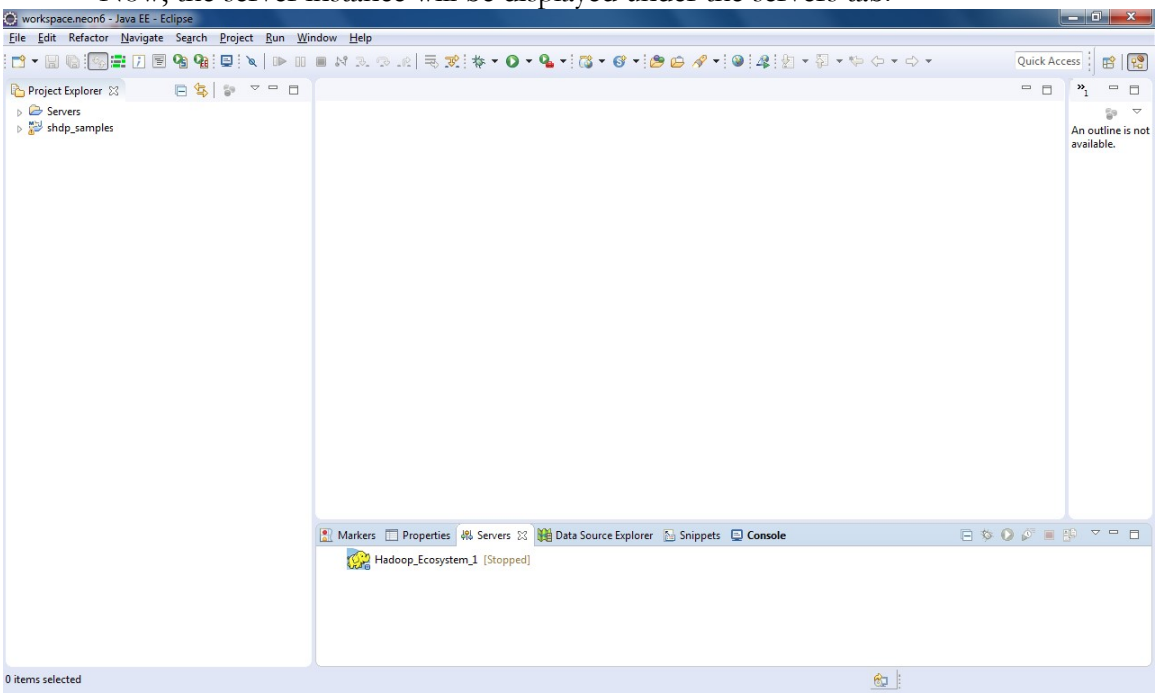
Hadoop Ecosystem allows you to automatically sync current workspace directories and files to Hadoop HDFS file system. This screen allows you to configure those setups. Click the Finish button.



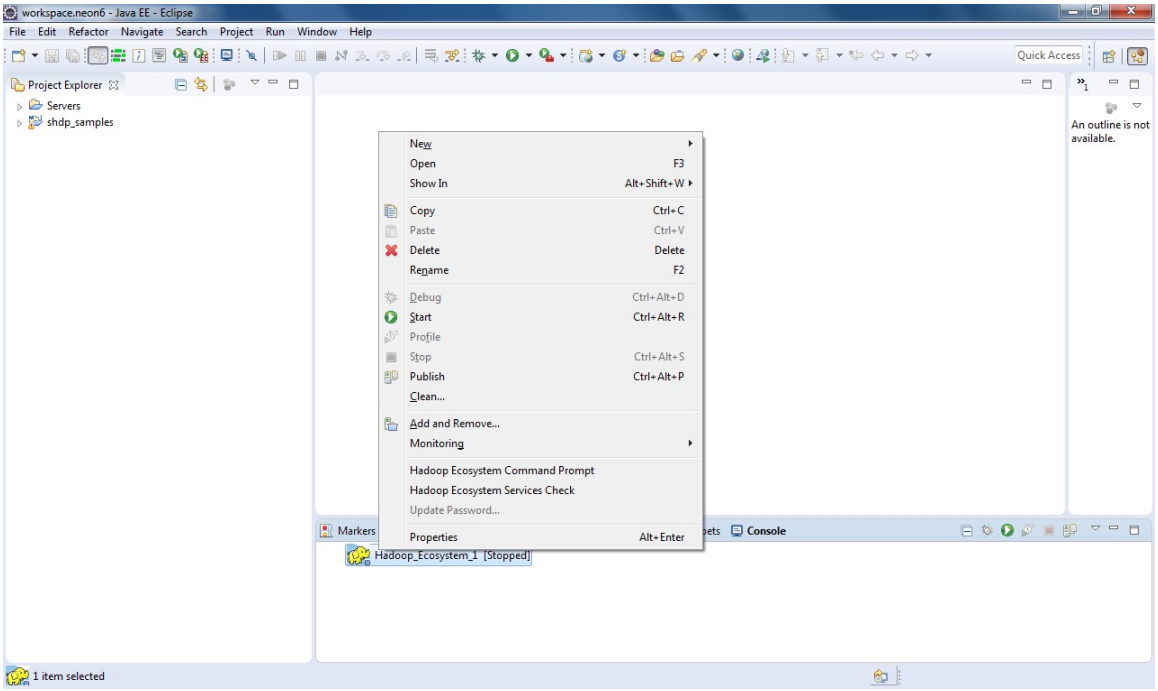
It takes a while to setup each Hadoop module configuration to the hosts.



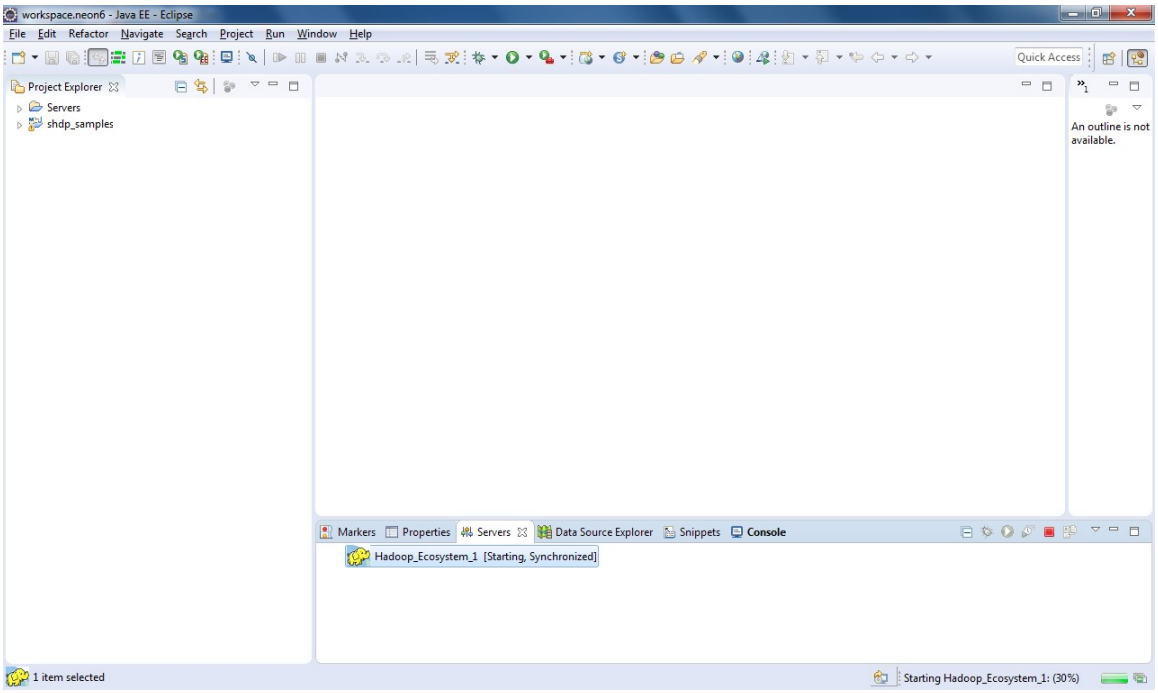
Now, the server instance will be displayed under the servers tab.



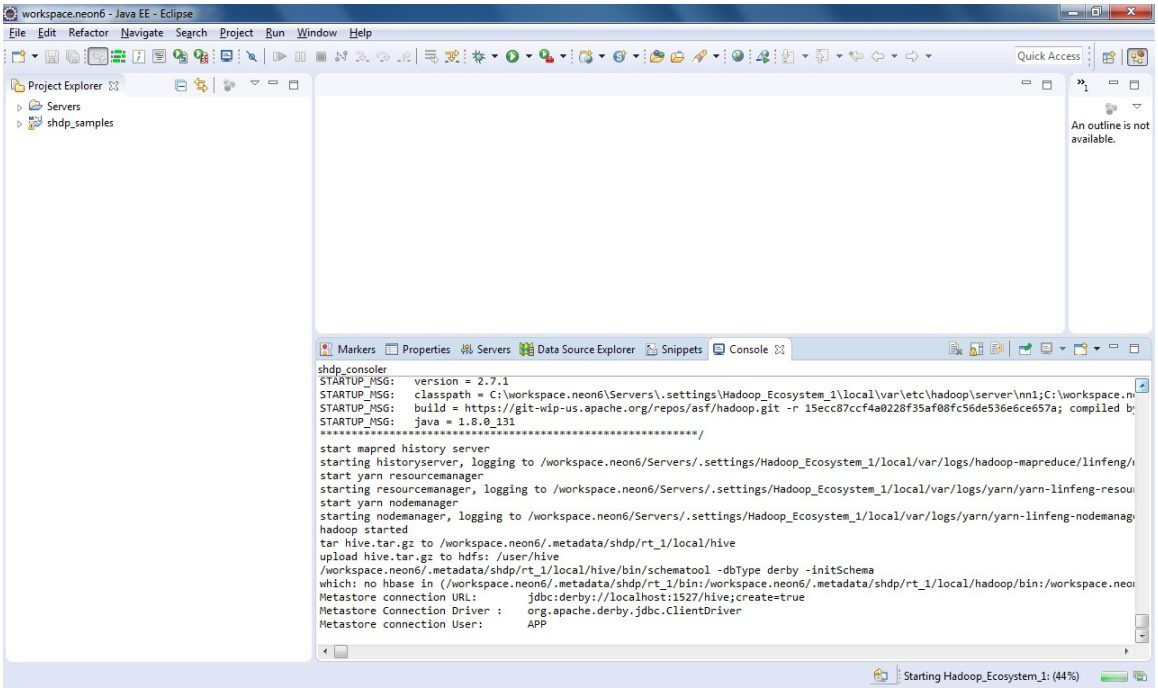
Right click the item to prompt the menu and click "start" menu to start services.



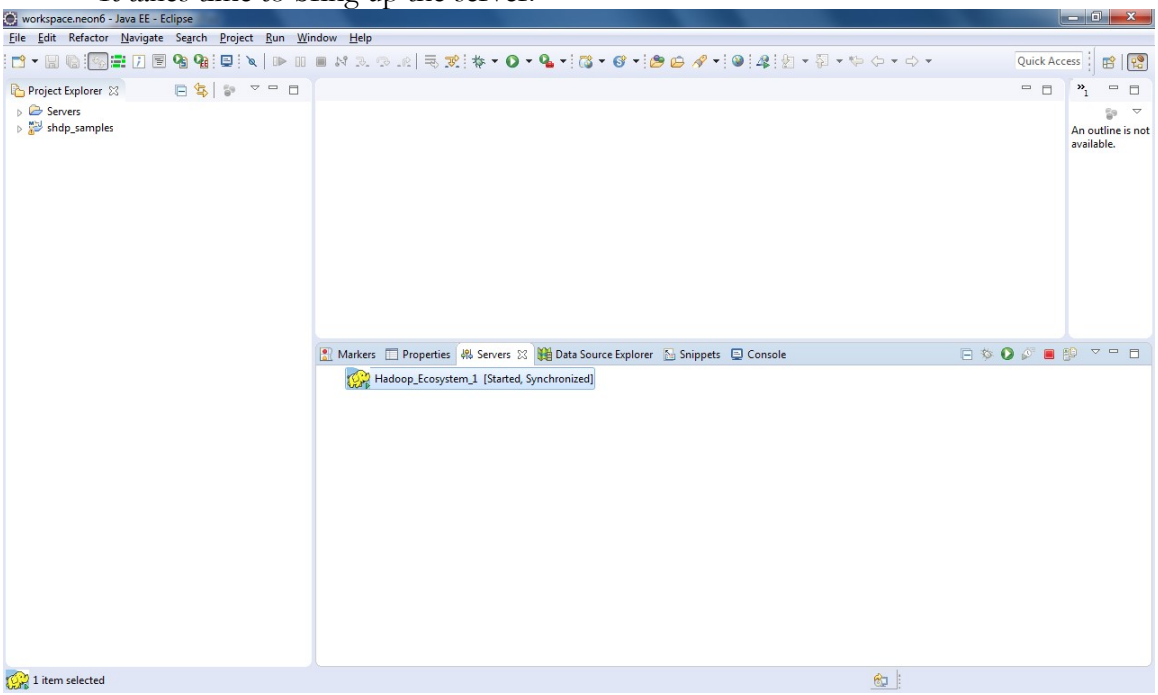
Server status shows the service starting.



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



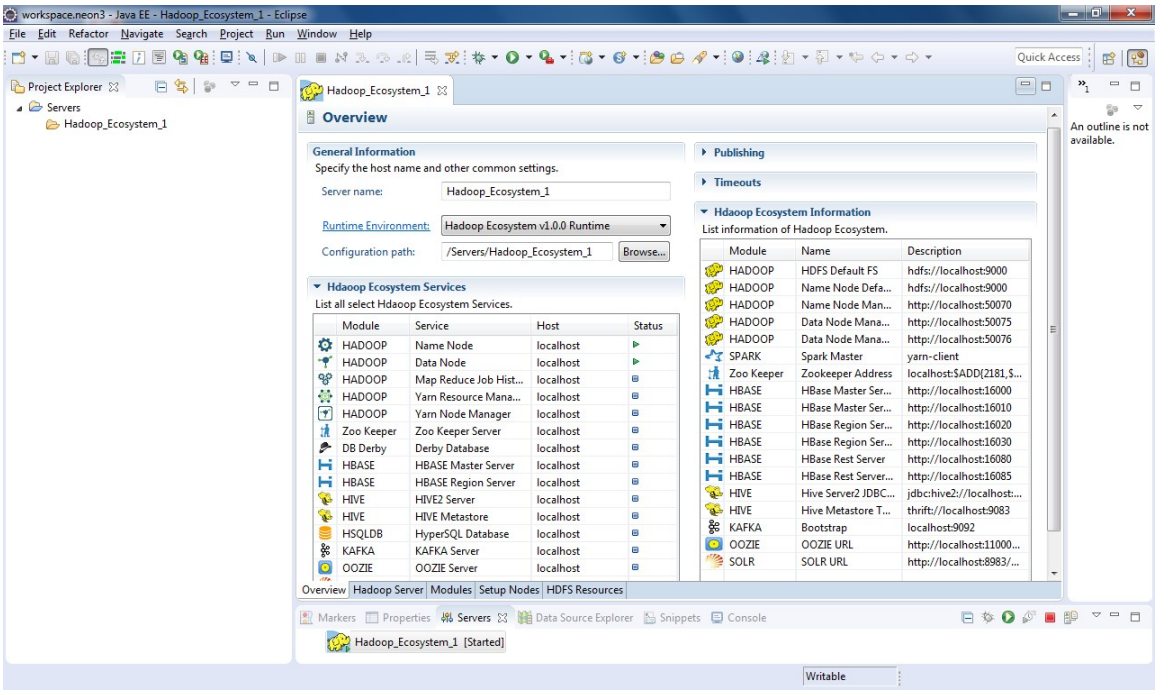
It takes time to bring up the server.



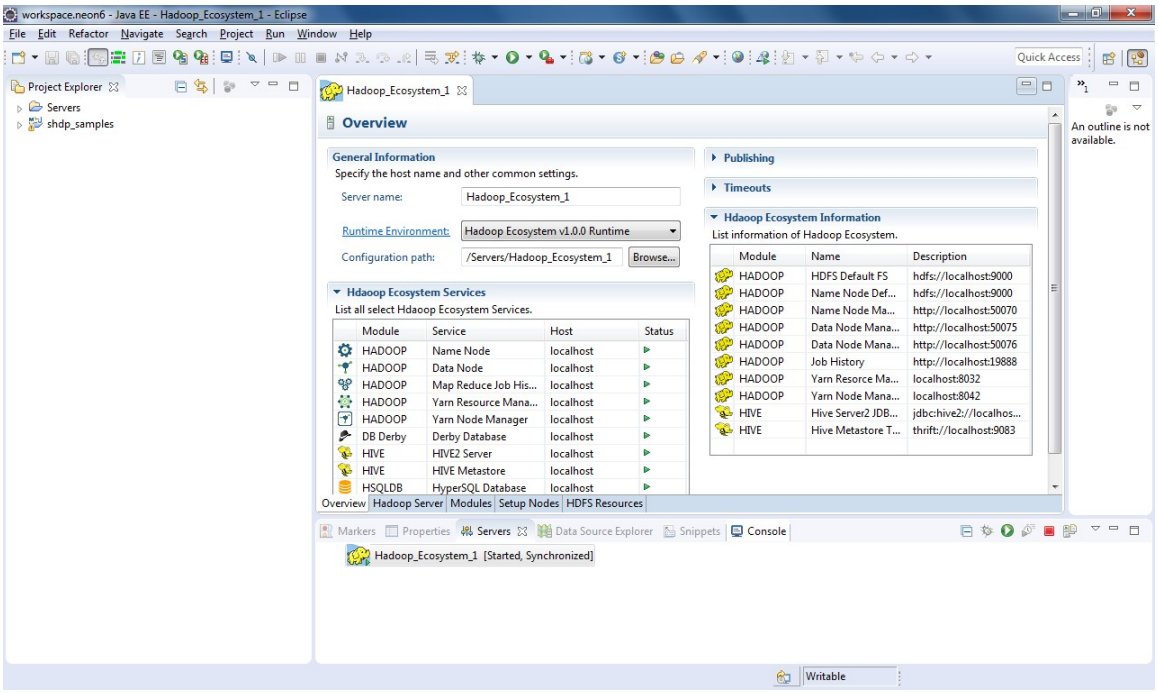
Double Click "Hadoop_Ecosystem_1" item to display the server information. The list on the left shows service start/stop information. The list on the right shows service access information for applications or users.

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

Demo License:

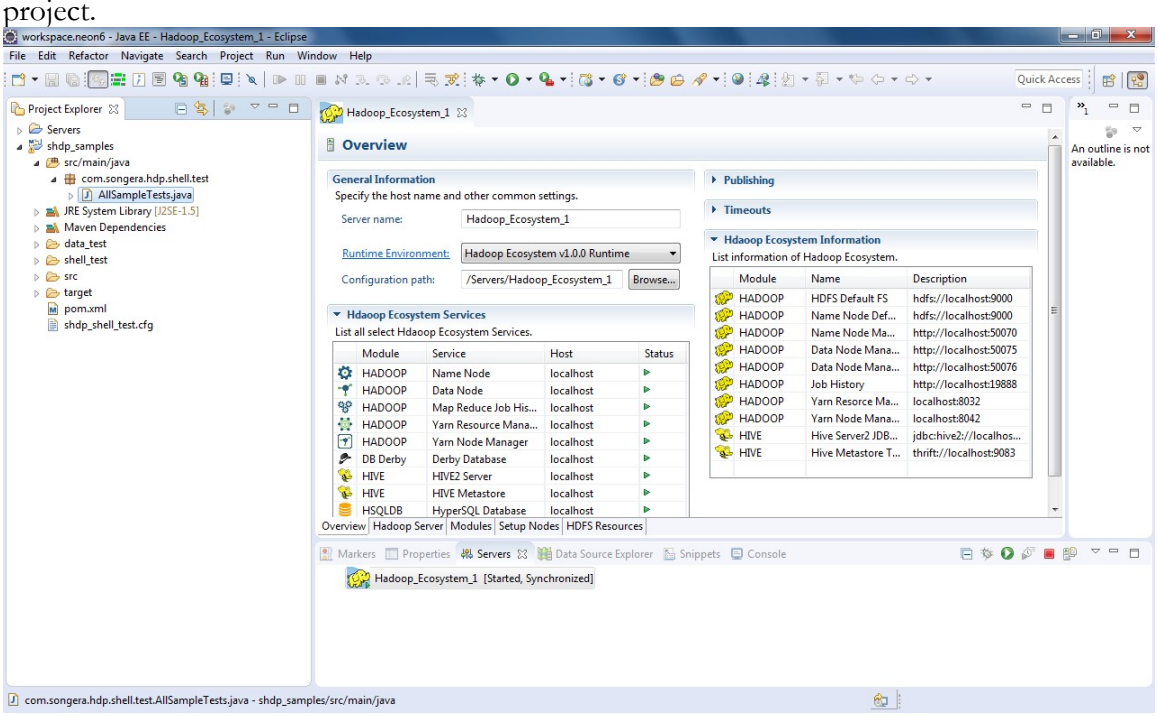


Trial, Test or Commercial license:

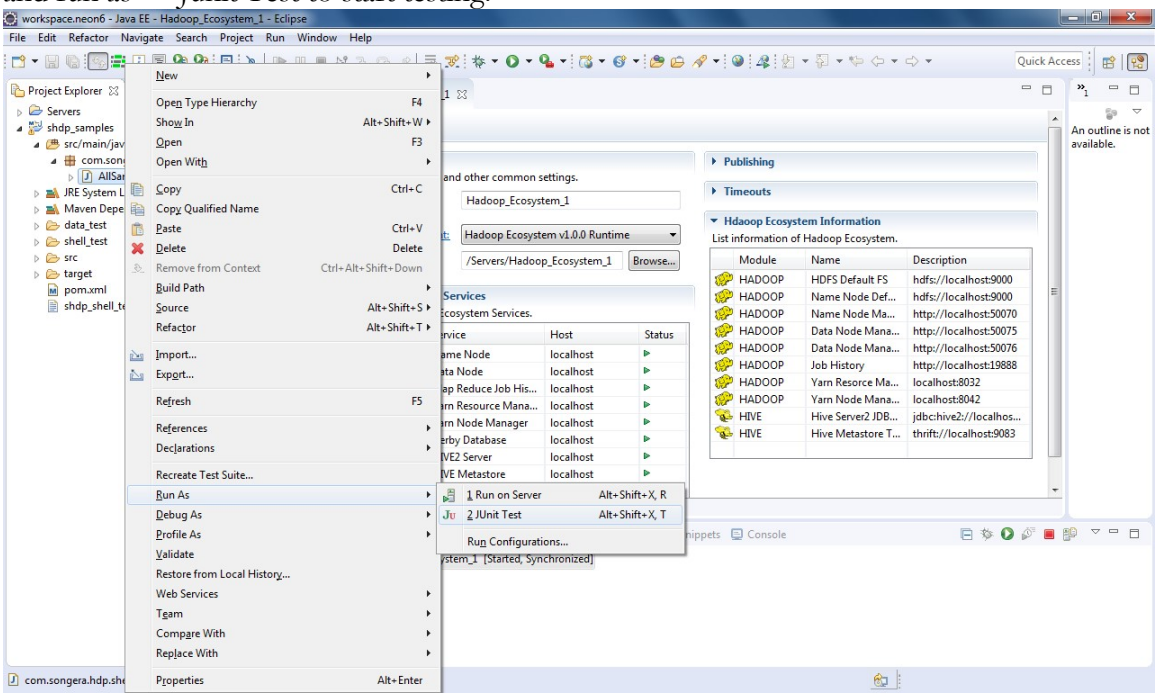


4. Test Hadoop Modules

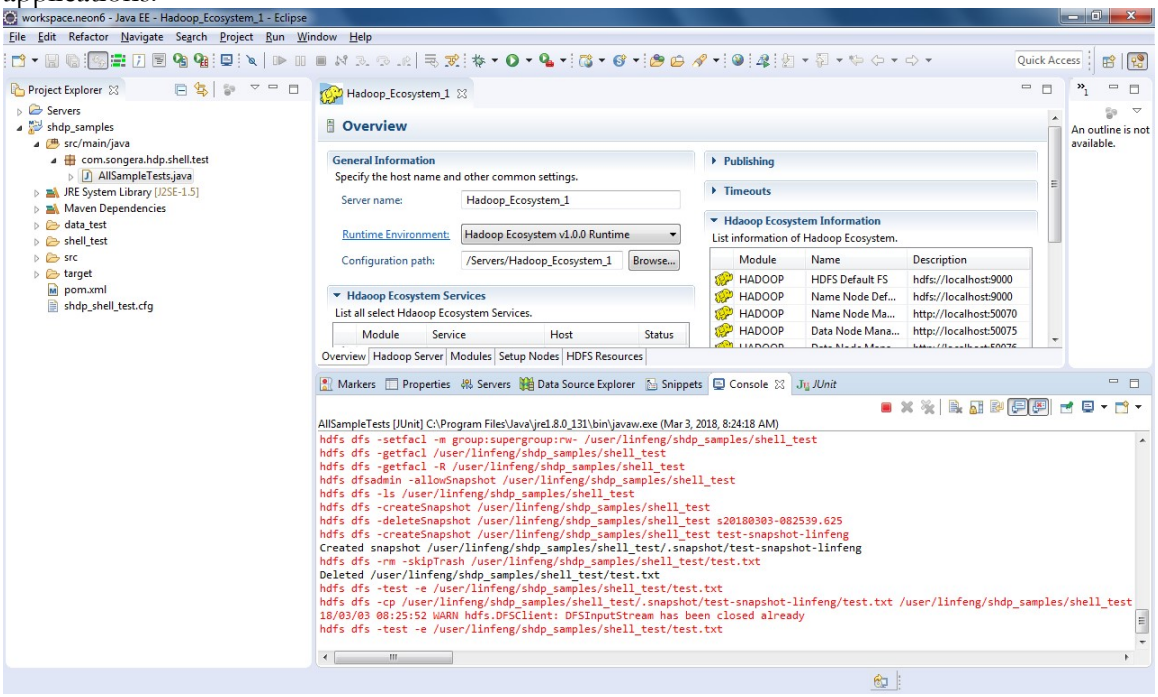
When the Hadoop Ecosystem server instance is created, it automatically import sample project: shdp_samples to current workspace. One Junit test suite java class is under that project.



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 a pass or failure.