

1. Introduction

- 2. Prerequistes
- 3. <u>Set Environment Variables</u>
- 4. Setup SSH daemon
- 5. <u>Download hadoop and place it</u> in the home directory.
- 6. <u>Unpack hadoop</u>
- 7. Configure Hadoop
- 8. Format the namenode
- 9. <u>Setup hadoop plugin</u>
- 10. Start the cluster
- 11. <u>Setup hadoop location</u>
- 12. Upload data
- 13. Create and run a test project.

🟮 add this 📲 😭 🏘 ...)

Create and run Hadoop project

Now we are ready to create and run out first Hadoop project.

Creating and configuring Hadoop eclipse project.

- 1. Launch Eclipse.
- Right-click on the blank space in the **Project Explorer** window and select **New Project..** to create a new project.
- 3. Select **Map/Reduce Project** from the list of project types as shown in the image below.

	🖷 New Project 📃 🗖 🚺	
'indow H	Select a wizard 🛛 🛁 🔶	
) 🗁 🚿		
	Wizards:	
	M	
	Map/Reduce Project Eclipse Modeling Framework Map/Reduce Map/Reduce Project	
		=
ay at this		
	(?) < Back Next > Finish C~	

- 4. Press the **Next** button.
- 5. You will see the project properties window similar to the one shown below

🖨 New MapReduce Project Wizard
MapReduce Project Create a MapReduce project.
Project name: HadoopTest

Encerion: C:(Dava) wispace.europa/macoopresc	
Hadoop MapReduce Library Installation Path	
💿 Use default Hadoop	Configur
O Specify Hadoop library location	Browse
0	< <u>B</u> ack

6. Fill in the project name and click on **Configure Hadoop Installation** link on the right hand side of the project configuration window. This will bring up the project **Preferences** window shown in the image below.

Preferences	
Hadoop Map/Reduce Tools	Select Hadoop Installation Directory
	Hadoop installation directory: C:\cygwin\home\User\hadoop-0.19.1 Browse
7. In the project Prefere in the Hadoop instal	ences window enter the location of the Hadoop directory lation directory field as shown above.

If you are not sure of the location of the Hadoop home directory, refer to Step 1 of <u>this section</u>. Hadoop home directory is one level up from the **conf** directory.

- 8. After entering the location close the **Preferences** window by pressing the **OK** button. Then close the **Project** window with the **Finish** button.
- 9. You have now created your first Hadoop Eclipse project. You should see its name in the **Project Explorer** tab.

Creating Map/Reduce driver class

- 1. Right-click on the newly created Hadoop project in the **Project Explorer** tab and select **New -> Other** from the context menu.
- 2. Go to the Map/Reduce folder, select MapReduceDriver, then press the Next

button as shown in the image below.

	🖨 New
F	Select a wizard
<i>^</i>	
	<u>W</u> izards:
	type filter text
	CVS Eclipse Modeling Framework Eclipse Modeling Framework EJB Eclipse Example EMF Model Creation Wizards Eclipse Example EMF model Creation Wizards Eclipse Dava Eclipse Dava
@ Jav	
5 6 7	O
	💽 ~/hadoop-0.19.1

3. When the **MapReduce Driver** wizard appears, enter TestDriver in the **Name** field and press the **Finish** button. This will create the skeleton code for the **MapReduce Driver**.

🖨 New MapRed	uce Driver		
MapReduce Dri	ver efault package is		D
Source fol <u>d</u> er:	HadoopTest/src		Browse
Package:	(default)	Browse
Na <u>m</u> e:	TestDriver		
Superclass:	java.lang.Object		Brows <u>e</u>
Interfaces.		(<u>A</u> dd
		(<u>R</u> emove
Mapper:			Browse
<u>R</u> educer:			Browse
?	< <u>B</u> ack <u>N</u> ext >	Einish	_

4. Unfortunately the Hadoop plug-in for Eclipse is slightly out of step with the recent Hadoop API, so we need to edit the driver code a bit.

Find the following two lines in the source code and comment them out:

```
conf.setInputPath(new Path("src"));
conf.setOutputPath(new Path("out"));
```

Enter the following code immediatly after the two lines you just commented out (see image below):

```
conf.setInputFormat(TextInputFormat.class);
conf.setOutputFormat(TextOutputFormat.class);
FileInputFormat.setInputPaths(conf, new Path("In"));
FileOutputFormat.setOutputPath(conf, new Path("Out"));
```

```
🕖 TestDriver.java 🛛 📉
           // TODO: specify output types
           conf.setOutputKeyClass(Text.class);
           conf.setOutputValueClass(IntWritable.class);
           // TODO: specify input and output DIRECTORIES (not files)
       // conf.setInputPath(new Path("src"));
       // conf.setOutputPath(new Path("out"));
            conf.setInputFormat(TextInputFormat.class);
            conf.setOutputFormat(TextOutputFormat.class);
             FileInputFormat.setInputPaths(conf, new Path("In"));
            FileOutputFormat.setOutputPath(conf, new Path("Out"));
            // TODO: enacifu a mannar
   1
🖹 Problems 🖉 Tasks 🙋 Javadoc 🎲 Map/Reduce Locations 🛛 🔪 📃 Console
Location
                                        Master node
🗉 👧 localhost
                                       localhost
```

5. After you have changed the code, you will see the new lines marked as incorrect by Eclipse. Click on the error icon for each line and select Eclipse's suggestion to import the missing class.

You need to import the following classes: *TextInputFormat, TextOutputFormat, FileInputFormat, FileOutputFormat.*

6. After the missing classes are imported you are ready to run the project.

Running Hadoop Project

 Right-click on the **TestDriver** class in the **Project Explorer** tab and select **Run** As --> Run on Hadoop. This will bring up a window like the one shown below.

🖨 Run on Hadoop	
Select Hadoop location Select a Hadoop location to run on.	
Select a Hadoop Server to run on.	

localhost	
Einish	
	Einish

- 2. In the window shown above select "*Choose existing Hadoop location*", then select **localhost** from the list below. After that click **Finish** button to start your project.
- 3. If you see console output similar to the one shown below, Congratulations! You have started the project successfully!



Continue

🚺 ADD THIS 📲 🎡 🧤 ...)

If you have questions comments suggestions regarding this tutorial you $\frac{\text{can post}}{\text{them here}}$ or you can write me an email to tutorials **AT** v-lad.org.

2008 - Vlad Korolev