

[More](#) [Next Blog»](#)[Create Blog](#) [Sign In](#)

# Everything related to Hadoop

MONDAY, SEPTEMBER 16, 2013

## Write and run map reduce jobs using Eclipse IDE!!

I started working in Hadoop field since last 1.5 years and i can see a picture where every one will be moving to Hadoop and related big data processing tools and suddenly it will become even much more popular than it is now and at that time it will be a field where everyone will be somehow related. As a developer, i really feel that everyone, at least every developer, should know how to write a MapReduce job and run it over a Hadoop cluster. I tried to find some help to write MapReduce jobs over the internet, but i found very few and even there there was no complete information.

So, i felt an urge to write this blog, so that every body should be able to know, in very easy and faster manner.

In this example i am using latest stable release of Hadoop which is 1.2.1. If you are using a different version then it may not work.

### Requirements to run this exercise:

1. Hadoop cluster (even a single node cluster will work) . I found [Michael G. Noll's website](#) very useful for setting up a single node cluster. You can follow the steps on this site to setup your cluster for testing.
2. It is very good practice to keep source files of an opensource software, so that you will be able to consult sources or even change, if required. You can download sources of Hadoop 1.2.1 from [here](#).
3. Eclipse IDE 3.7. (I have tested it on Eclipse 3.7. It may not work on higher versions.)

### Steps:

1. Download MapReduce eclipse plugin from [here](#)
2. Copy it into "Eclipse installation directory -> plugins" directory
3. Restart eclipse
4. Open MapReduce perspective in Eclipse

### BLOG ARCHIVE

▼ 2013 (2)

▼ September (2)

[How to write MapReduce program in JAVA](#)

[Write and run map reduce jobs using Eclipse IDE!!](#)

### ABOUT ME



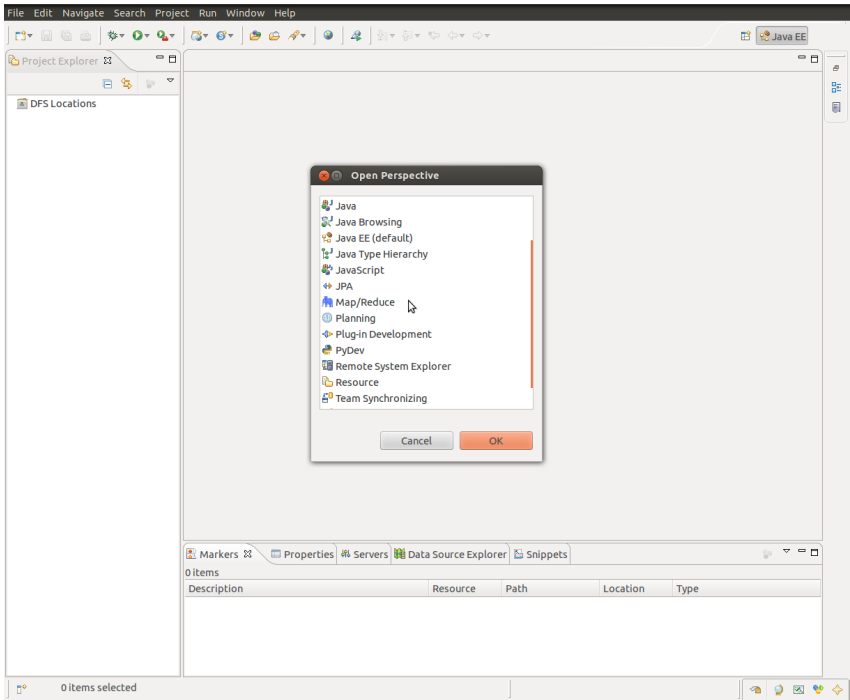
**Deepak Shrivastava**



Follow

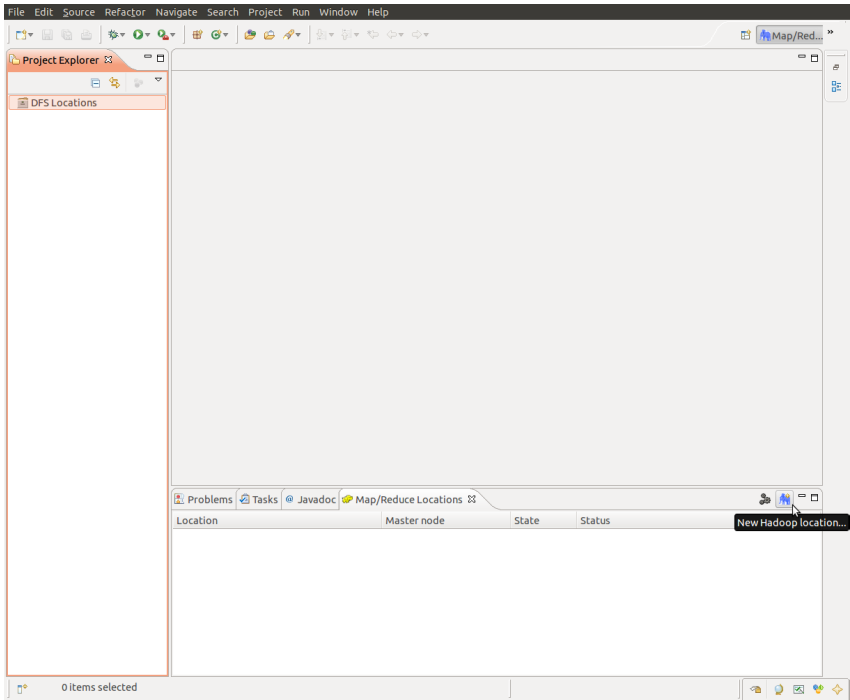
173

[View my complete profile](#)



5.

Add MapReduce locations



6. Click on add new locations

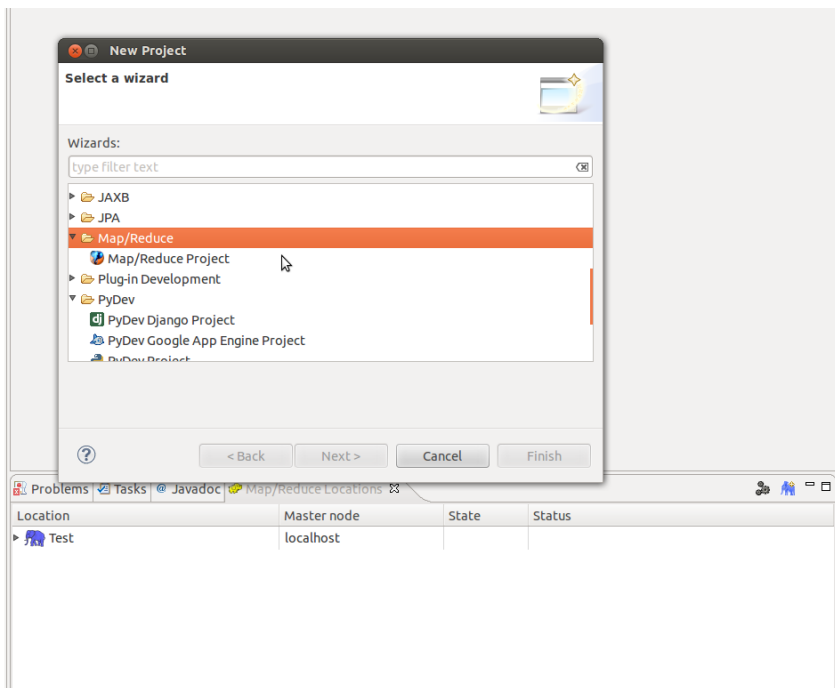
The screenshot shows the 'Define Hadoop location' dialog box with the 'General' tab selected. The 'Location name' field contains 'Test'. The 'Map/Reduce Master' section has 'Host' set to 'localhost' and 'Port' set to '50030'. The 'DFS Master' section has 'Host' set to 'localhost' and 'Port' set to '54310'. The 'User name' field contains 'hdusr'. The 'SOCKS proxy' section has 'Enable SOCKS proxy' unchecked, 'Host' set to 'host', and 'Port' set to '1080'. At the bottom, there are 'Load from file' and 'Validate location' buttons, and 'Cancel' and 'Finish' buttons.

7. There are few more advanced parameters also which may be changed depending upon your environment

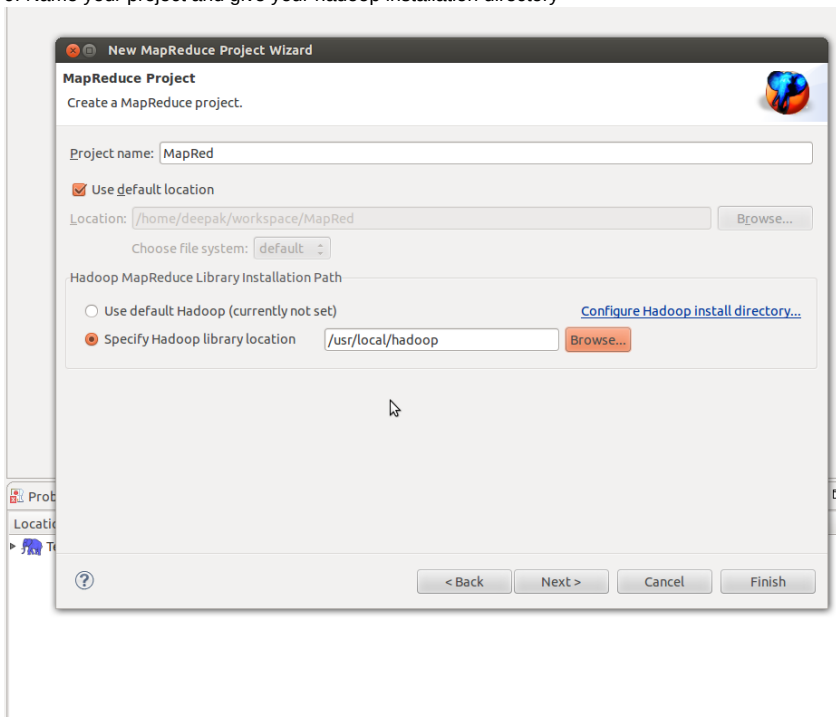
The screenshot shows the 'Define Hadoop location' dialog box with the 'Advanced parameters' tab selected. The 'fs.trash.interval' is set to 0. The 'hadoop.logfile.count' is set to 10. The 'hadoop.logfile.size' is set to 10000000. The 'hadoop.native.lib' is set to true. The 'hadoop.rpc.socket.factory.class.default' is set to org.apache.hadoop.net.StandardSocketFactory. The 'hadoop.security.authorization' is set to false. The 'hadoop.socks.server' is set to host:1080. The 'hadoop.tmp.dir' is set to /app/hadoop/tmp. The 'hadoop.util.hash.type' is set to murmur. The 'io.bytes.per.checksum' is set to 512. The 'io.compression.codecs' is set to org.apache.hadoop.io.compress.DefaultCodec,org.apache.hadoop.io.compress.GzipCodec,org.apache.h. The 'io.file.buffer.size' is set to 4096. The 'io.map.index.skip' is set to 0. The 'io.mapfile.bloom.error.rate' is set to 0.005. The 'io.mapfile.bloom.size' is set to 1048576. The 'io.seqfile.compress.blocksize' is set to 1000000. The 'io.seqfile.lazydecompress' is set to true. The 'io.seqfile.sorter.recordlimit' is set to 1000000. The 'io.serialization' is set to org.apache.hadoop.io.serializer.WritableSerialization. The 'io.skip.checksum.errors' is set to false. The 'io.sort.factor' is set to 10. The 'io.sort.mb' is set to 100. At the bottom, there are 'Load from file' and 'Validate location' buttons, and 'Cancel' and 'Finish' buttons.

Now, development environment setup is done. You are ready to write and run your MapReduce job. To create a new MapReduce project and run over hadoop cluster follow below steps

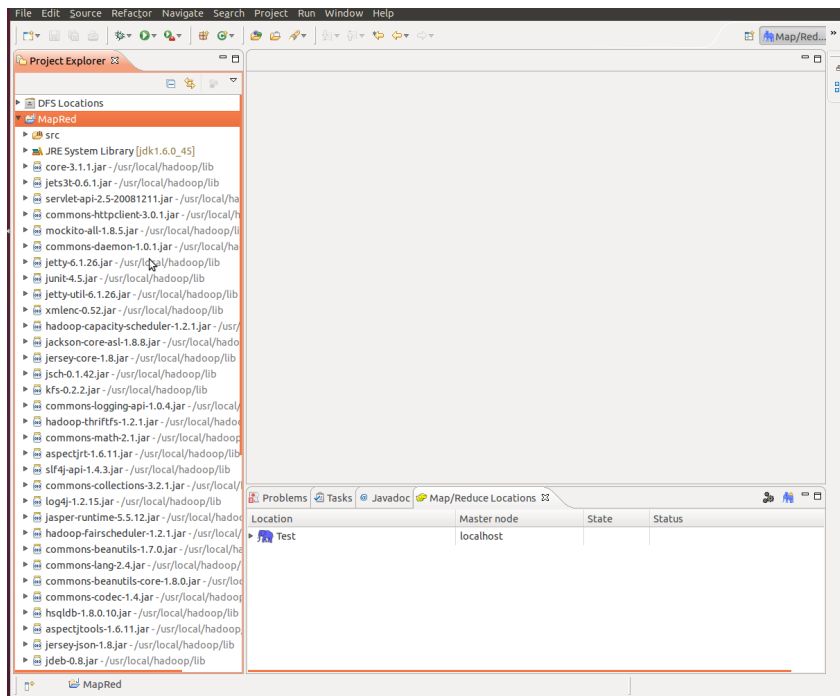
8. Create new Map reduce project



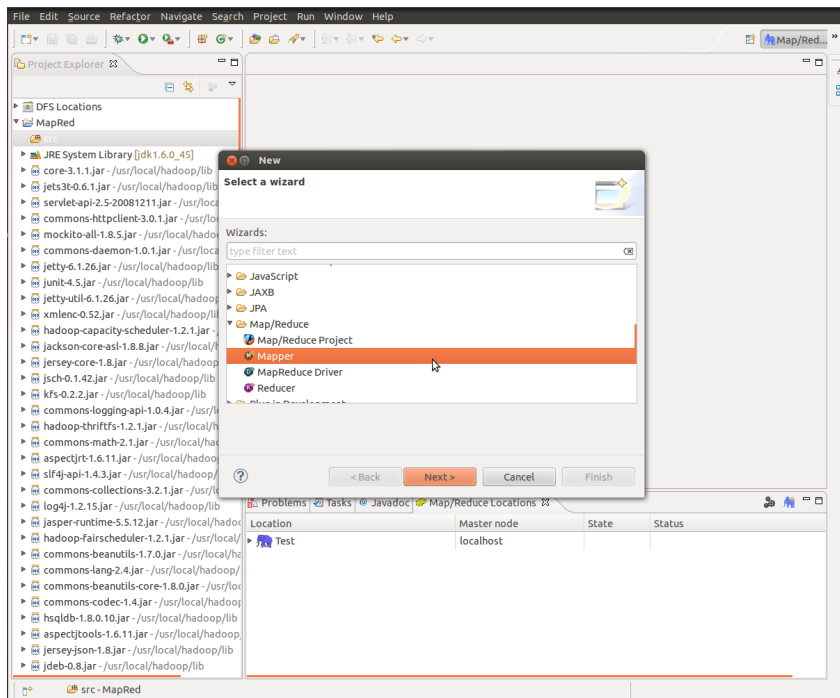
9. Name your project and give your hadoop installation directory

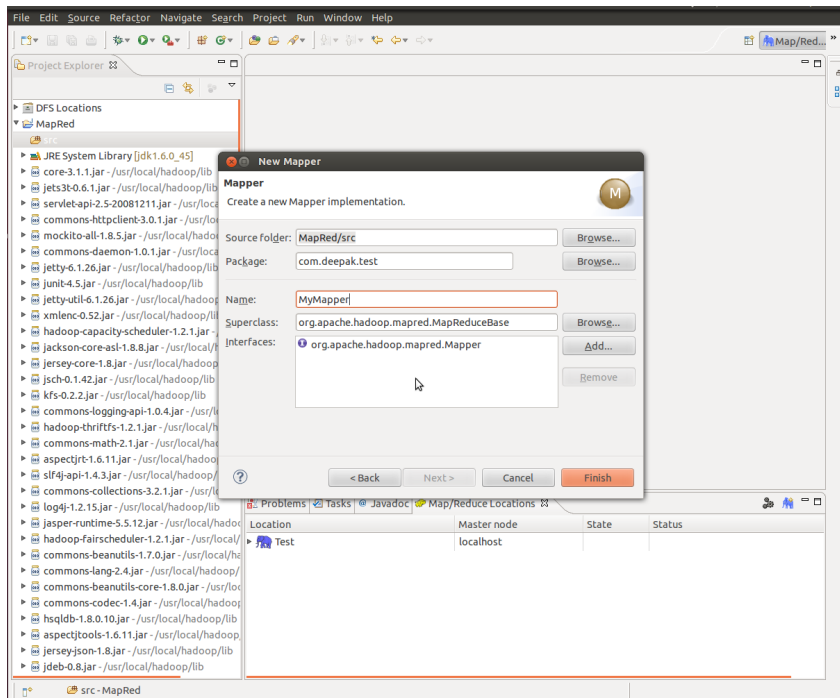


10. Your project structure may look like this



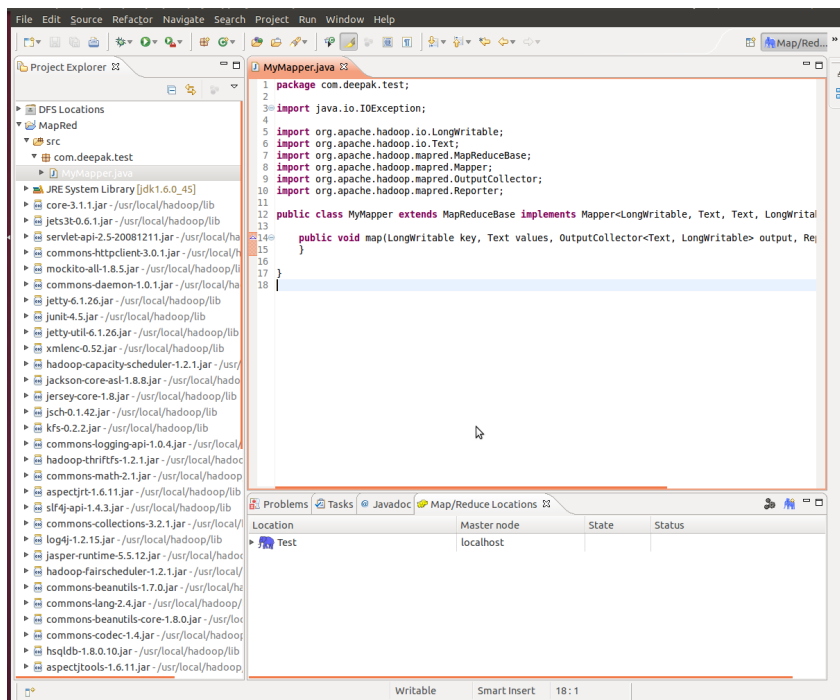
11. Now you need to add new Mapper class





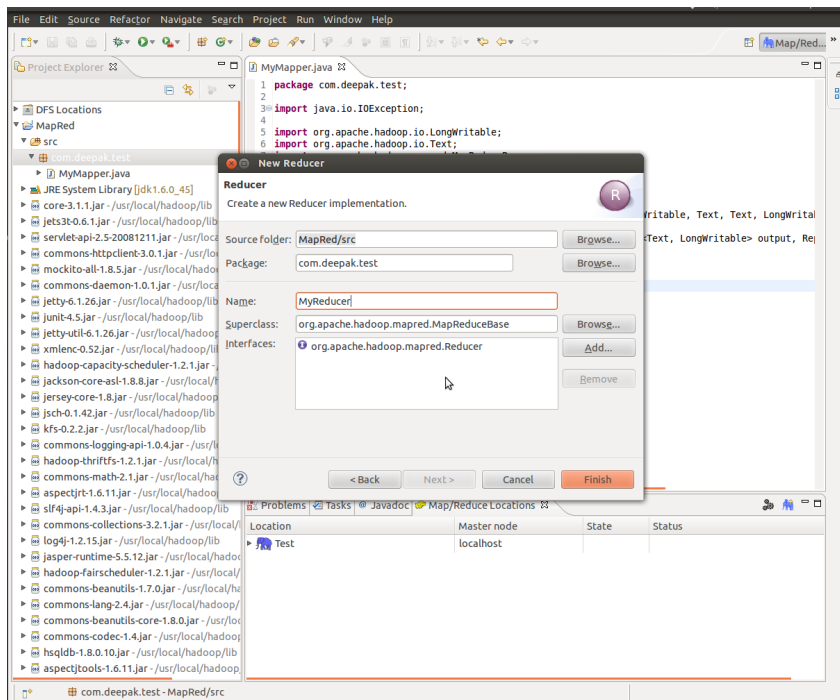
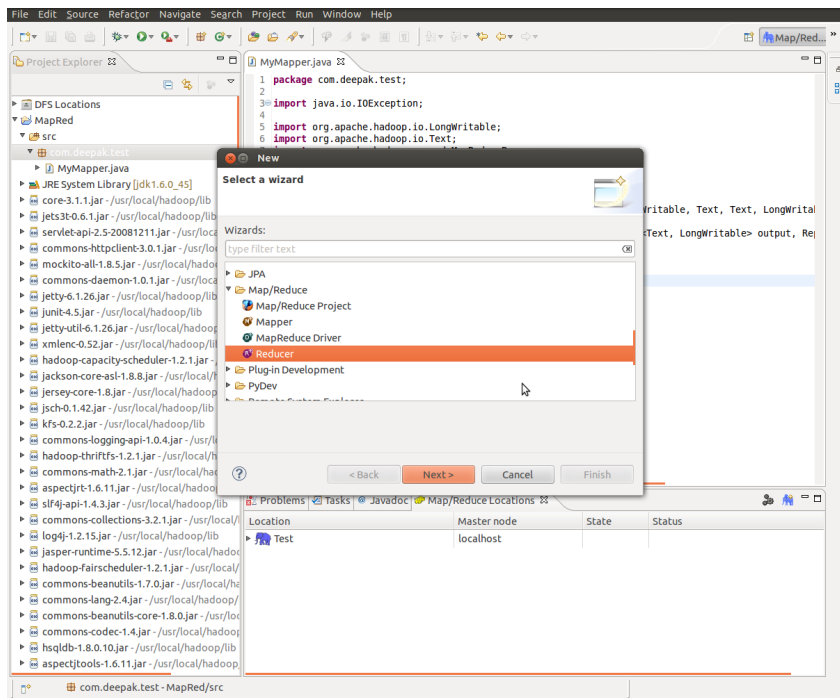
12.  
It

should look like this

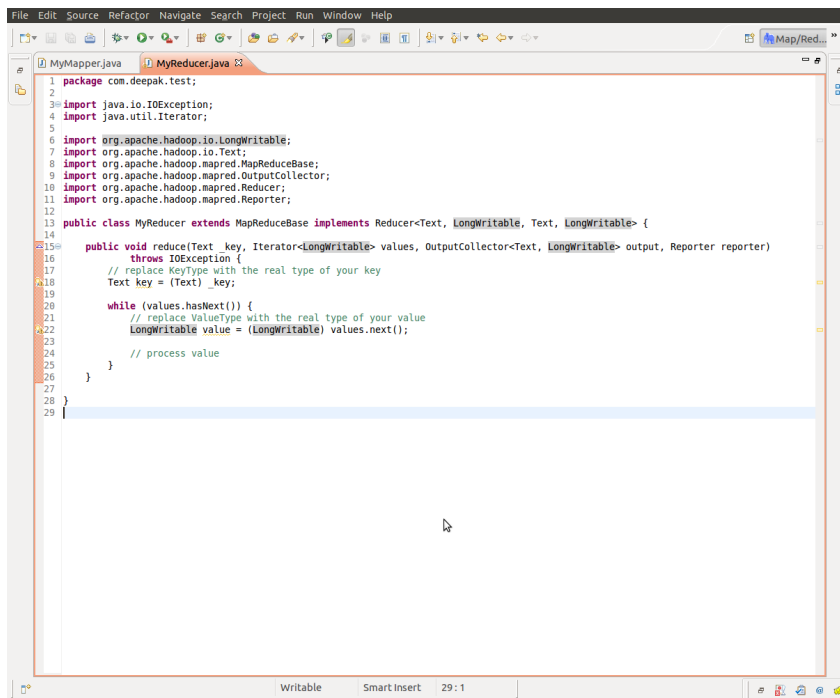


You can write your code for mapper class here.

13. Add Reducer class



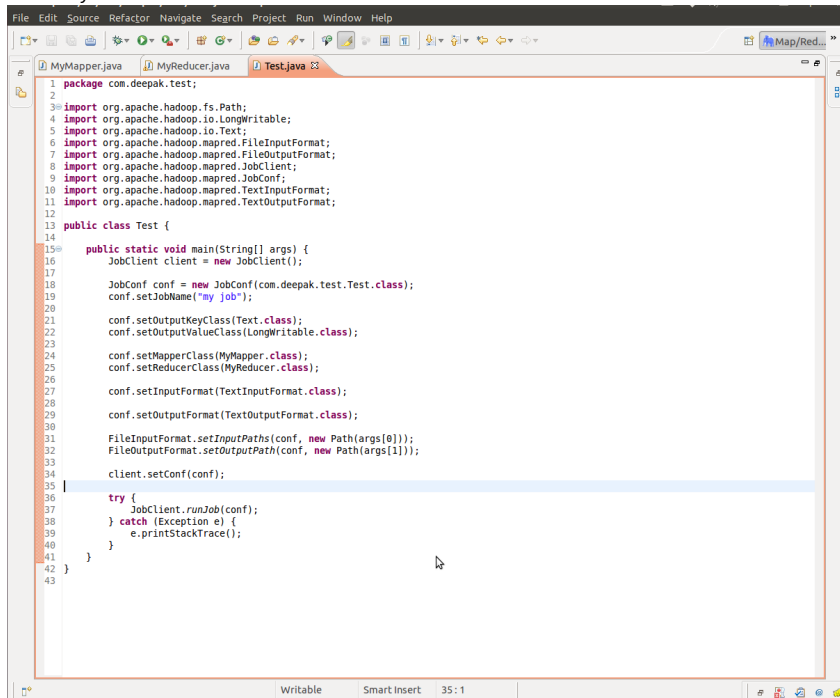
14. Your Reducer class should look like this



```
1 package com.deepak.test;
2
3 import java.io.IOException;
4 import java.util.Iterator;
5
6 import org.apache.hadoop.io.LongWritable;
7 import org.apache.hadoop.io.Text;
8 import org.apache.hadoop.mapred.MapReduceBase;
9 import org.apache.hadoop.mapred.OutputCollector;
10 import org.apache.hadoop.mapred.Reducer;
11 import org.apache.hadoop.mapred.Reporter;
12
13 public class MyReducer extends MapReduceBase implements Reducer<Text, LongWritable, Text, LongWritable> {
14
15     public void reduce(Text key, Iterator<LongWritable> values, OutputCollector<Text, LongWritable> output, Reporter reporter)
16         throws IOException {
17         // replace KeyType with the real type of your key
18         Text key = (Text) _key;
19
20         while (values.hasNext()) {
21             // replace ValueType with the real type of your value
22             LongWritable value = (LongWritable) values.next();
23
24             // process value
25         }
26     }
27 }
28
29
```

You can add your code for reducer class here

#### 15. Add your Test runner class like this

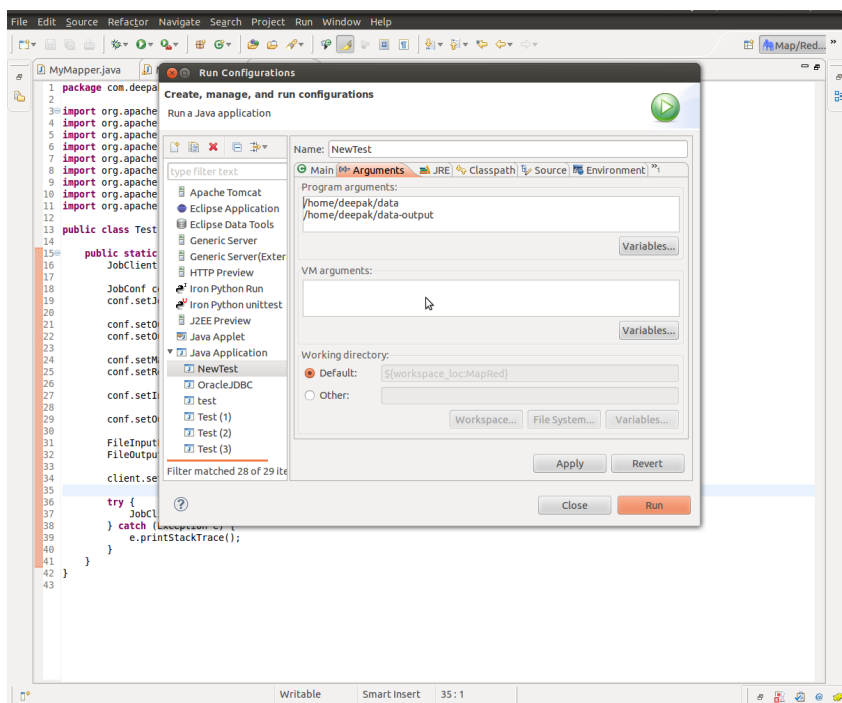
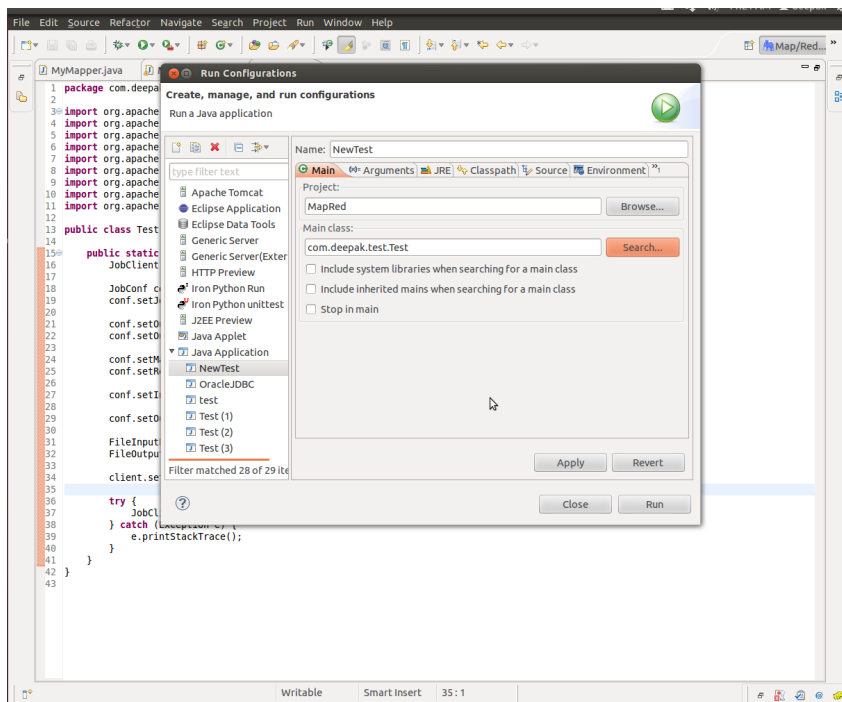


```
1 package com.deepak.test;
2
3 import org.apache.hadoop.fs.Path;
4 import org.apache.hadoop.io.LongWritable;
5 import org.apache.hadoop.io.Text;
6 import org.apache.hadoop.mapred.FileInputFormat;
7 import org.apache.hadoop.mapred.FileOutputFormat;
8 import org.apache.hadoop.mapred.JobClient;
9 import org.apache.hadoop.mapred.JobConf;
10 import org.apache.hadoop.mapred.TextInputFormat;
11 import org.apache.hadoop.mapred.TextOutputFormat;
12
13 public class Test {
14
15     public static void main(String[] args) {
16         JobClient client = new JobClient();
17
18         JobConf conf = new JobConf(com.deepak.test.Test.class);
19         conf.setJobName("my job");
20
21         conf.setOutputKeyClass(Text.class);
22         conf.setOutputValueClass(LongWritable.class);
23
24         conf.setMapperClass(MyMapper.class);
25         conf.setReducerClass(MyReducer.class);
26
27         conf.setInputFormat(TextInputFormat.class);
28         conf.setOutputFormat(TextOutputFormat.class);
29
30         FileInputFormat.setInputPaths(conf, new Path(args[0]));
31         FileOutputFormat.setOutputPath(conf, new Path(args[1]));
32
33         client.setConf(conf);
34
35         try {
36             JobClient.runJob(conf);
37         } catch (Exception e) {
38             e.printStackTrace();
39         }
40     }
41 }
42
43
```

#### 16. To run the program Click on Run As->Run Configurations...

Your screen should look like this





Click on Run and your program should start running.


You can find the output of the program at your configured location.

Next blog post: [How to write MapReduce program in JAVA](#)

Please let me know if you find any difficulty in running the above said steps.

Thanks

Posted by **Deepak Shrivastava** at 9/16/2013 12:42:00 PM

 +2 Recommend this on Google

Labels: [Hadoop MapReduce Development Environment setup](#), [Hadoop MapReduce Example](#), [Java](#), [MapReduce Development](#), [MapReduce Java](#), [MapReduce on Eclipse](#), [Programing Map-Reduce \( Hadoop \) with Eclipse](#), [Running MapReduce from eclipse](#)

7 comments:



**alok** September 22, 2013 at 9:38 PM

great work....a very simple step by step representation of the facts.

[Reply](#)



**typedef uchar\_t pluto**; March 1, 2014 at 11:25 PM

Brilliant man. Very useful and saved a lot of time for me.

Thanks very much.

[Reply](#)



**besantvignesh M** June 27, 2014 at 11:40 AM

Its really helpful for me to understand where we i lost in my previous interview. Thanks.  
If anyone wants to Learn Hadoop in Chennai go to the Besant Technologies which is No.1 Training Institute in Chennai.

<http://www.hadooptrainingchennai.co.in>

[Reply](#)



**Lobna Tonn** April 8, 2015 at 4:55 PM

Thank you for this step by step tutorial.

But what if i have Eclipse in windows system and hadoop cluster on CentOS server machines and no GUI available. How can i run eclipse with hadoop eclipse plugin on windows machine ?

[Reply](#)



**Andrew Son** August 11, 2015 at 11:42 AM

There are lots of information about latest technology and how to get trained in them, like [Hadoop Training Chennai](#) have spread around the web, but this is a unique one according to me. The strategy you have updated here will make me to get trained in future technologies([Hadoop Training in Chennai](#)). By the way you are running a great blog. Thanks for sharing this ([Salesforce Training in Chennai](#)).

[Reply](#)



**Roshini RS** September 16, 2015 at 12:22 PM

I have read your blog, it was good to read & I am getting some useful info's through your blog keep sharing... Informatica is an ETL tools helps to transform your old business leads into new vision. Learn [Informatica training in chennai](#) from corporate professionals with very good experience in informatica tool.

Regards,

[Best Informatica Training In Chennai](#)|[Informatica training center in Chennai](#)|[Informatica training chennai](#)

[Reply](#)



**Amirtha rao** December 22, 2015 at 3:42 PM

Pretty article! I found some useful information in your blog, it was awesome to read, thanks for sharing this great content to my vision, keep sharing.

Regards,

[sas training in Chennai](#)[sas course in Chennai](#)[sas training institute in Chennai](#)  
[Reply](#)

Enter your comment...

Comment as: Unknown (Google)

[Sign out](#)

[Publish](#)

[Preview](#)

☐ [Notify me](#)

[Newer Post](#)

[Home](#)

Subscribe to: [Post Comments \(Atom\)](#)

Picture Window template. Powered by Blogger.