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)

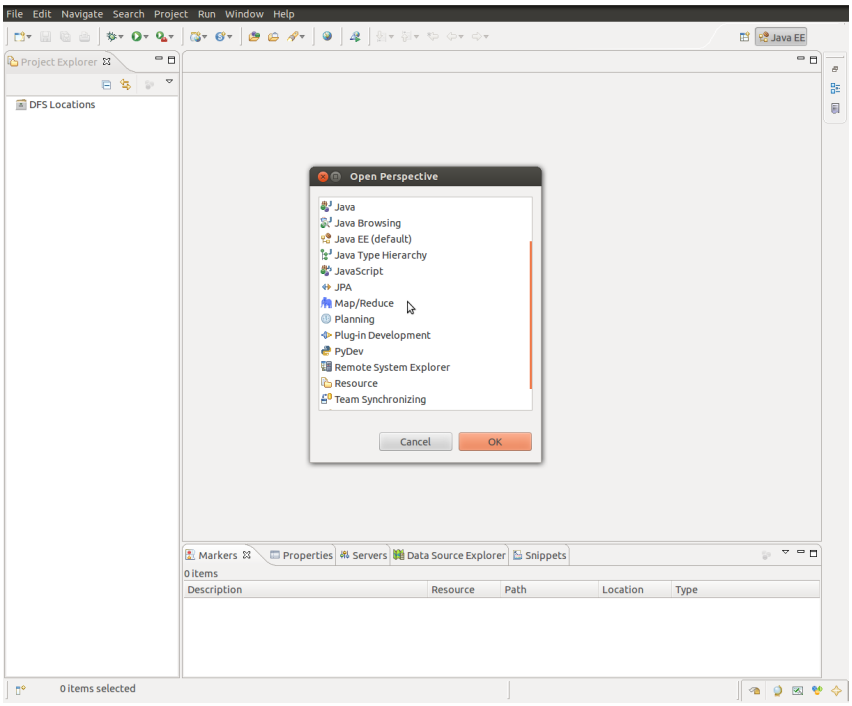
▼ September (2)

How to write MapReduce program in JAVA

Write and run map reduce jobs using Eclipse IDE!!

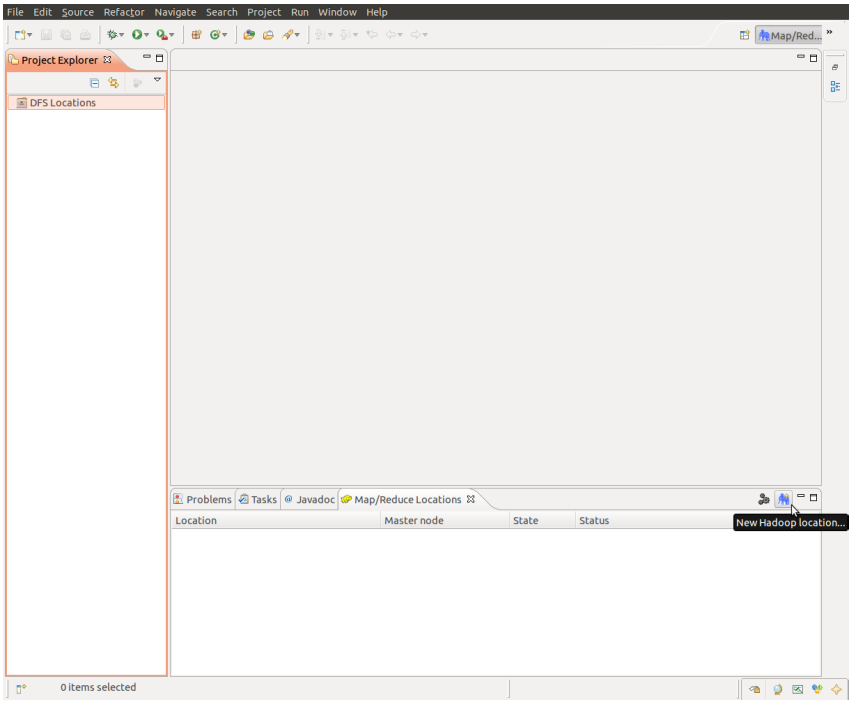
ABOUT ME

Deepak Shrivastava

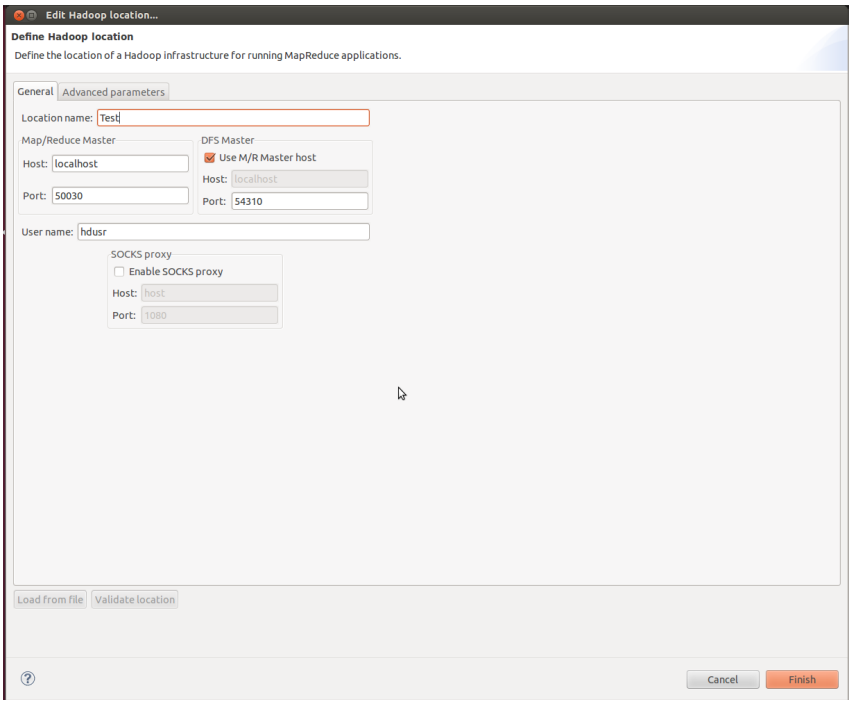


5.

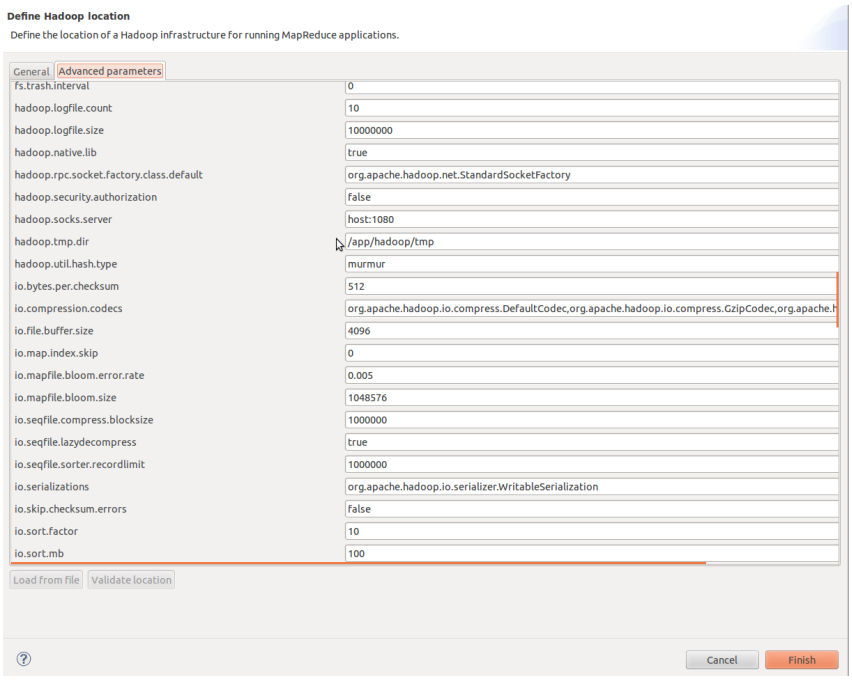
Add MapReduce locations



6. Click on add new locations

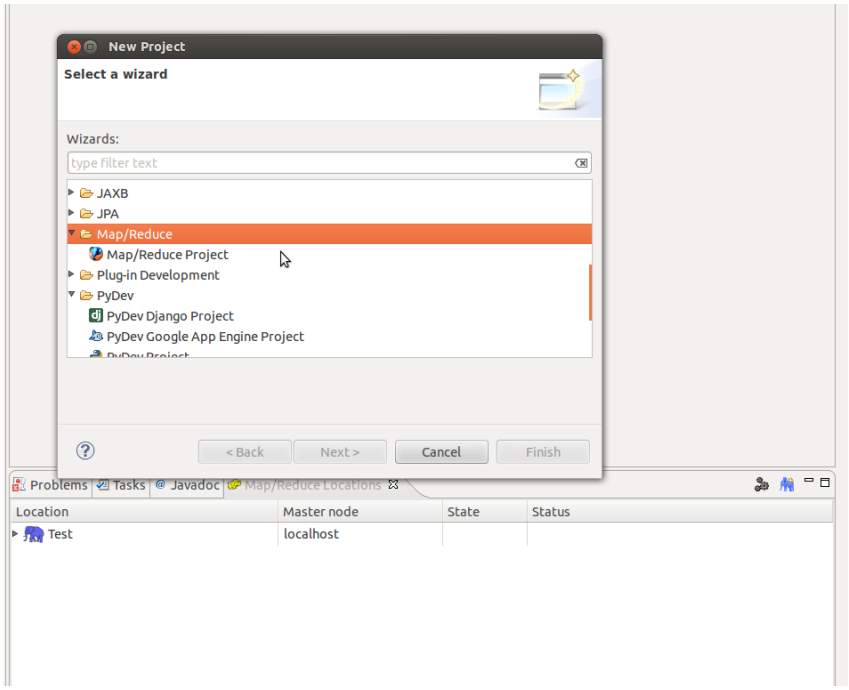


7. There are few more advanced parameters also which may be changed depending upon your environment

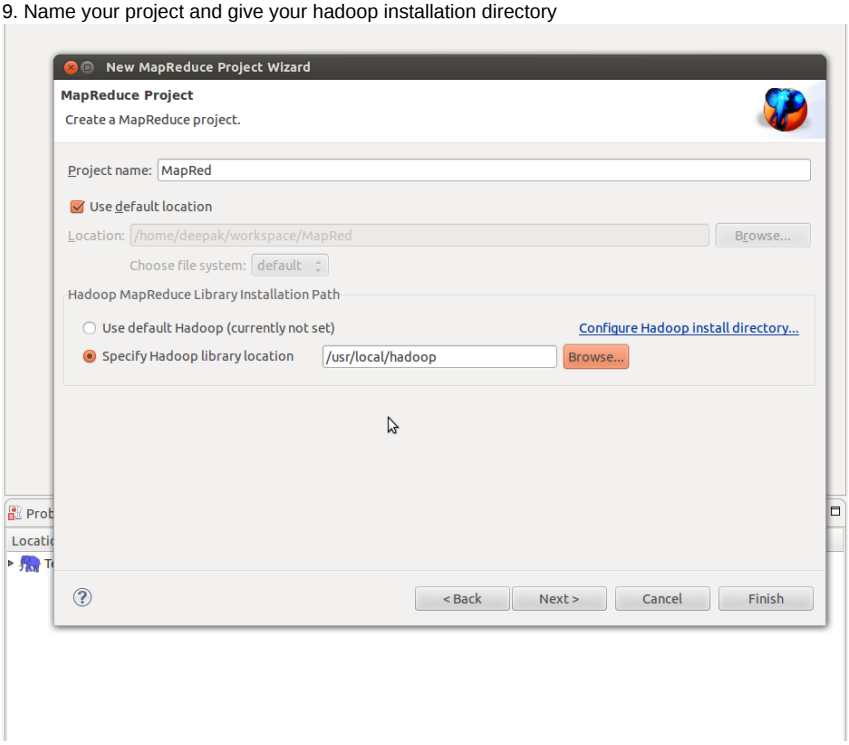


Now, development environment setup is done. You are ready to write and run your MapReduce job. To create a new MapReduce project and run over hadoop cluster follow below steps

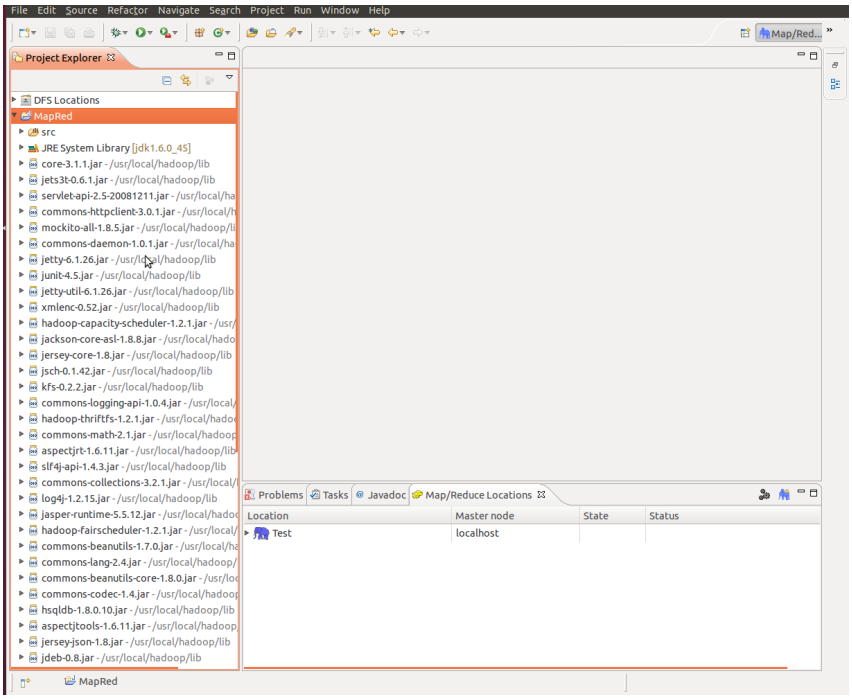
8. Create new Map reduce project



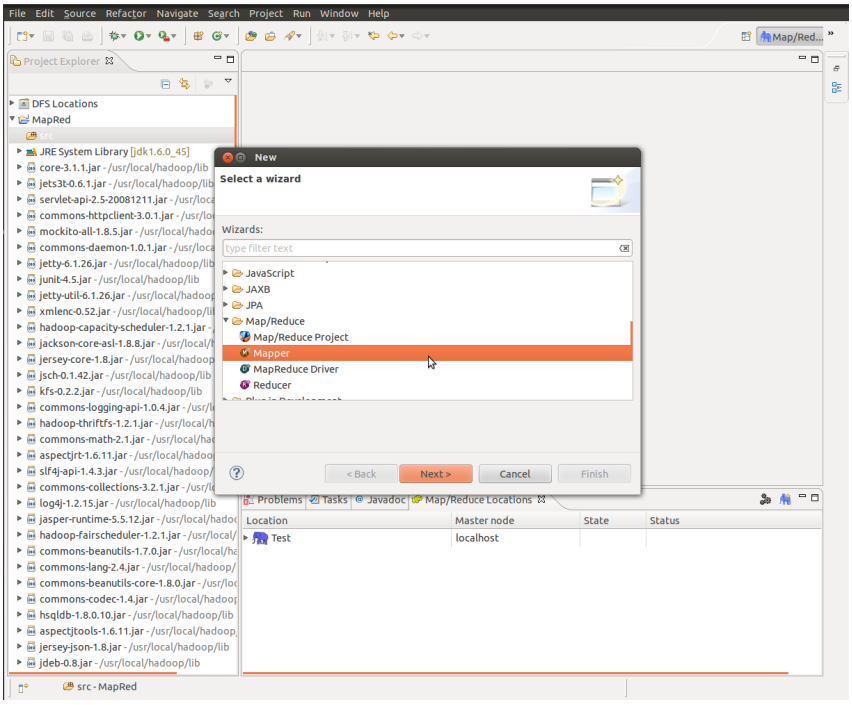
9. Name your project and give your hadoop installation directory

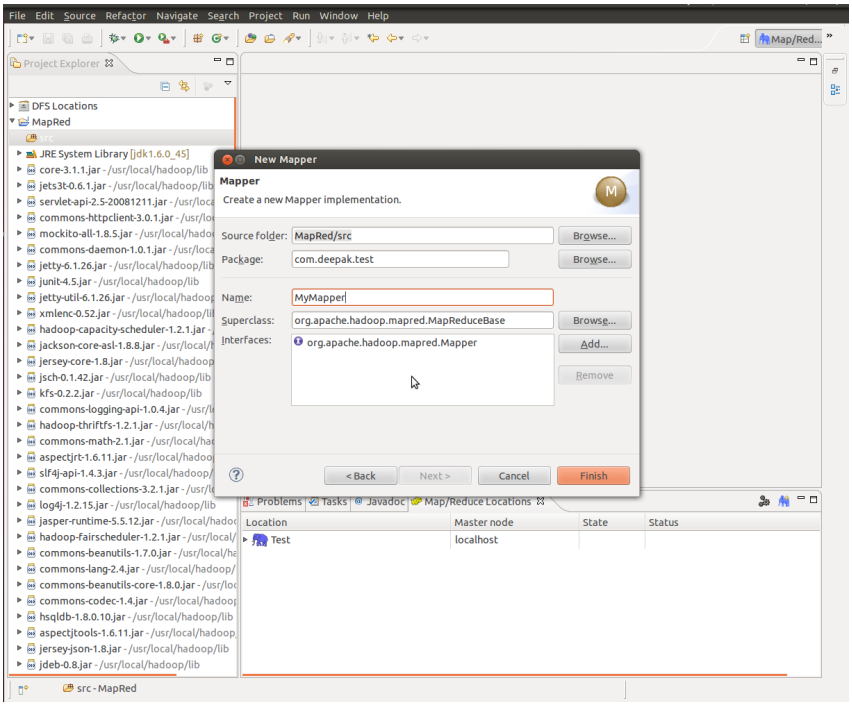


10. Your project structure may look like this

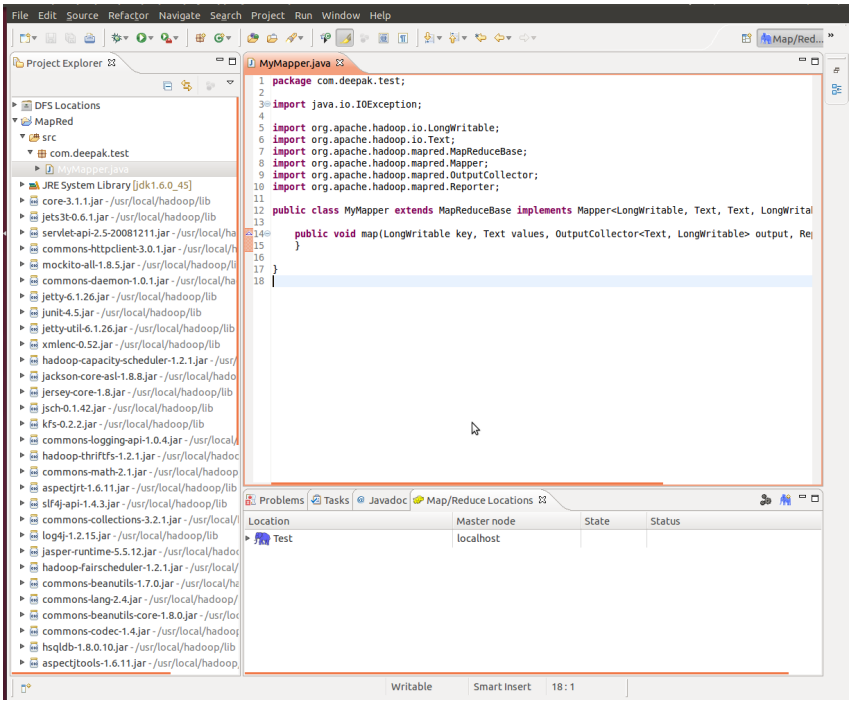


11. Now you need to add new Mapper class



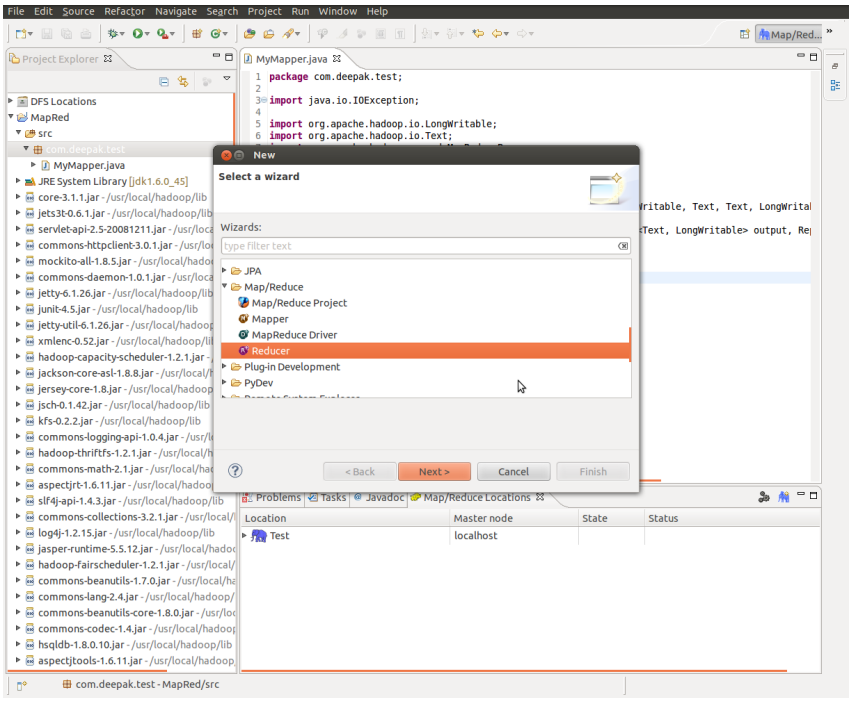


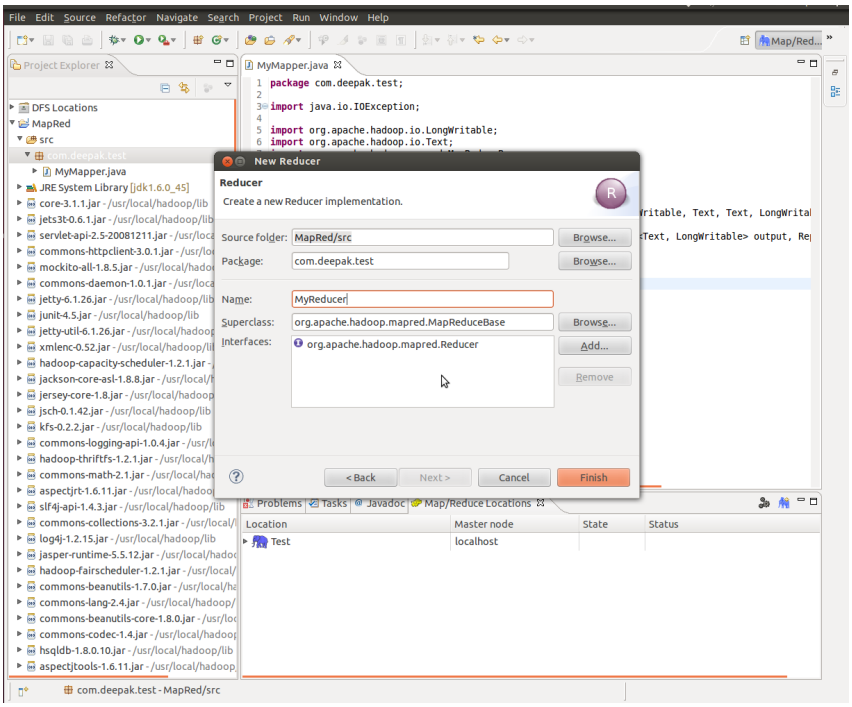
12. It should look like this



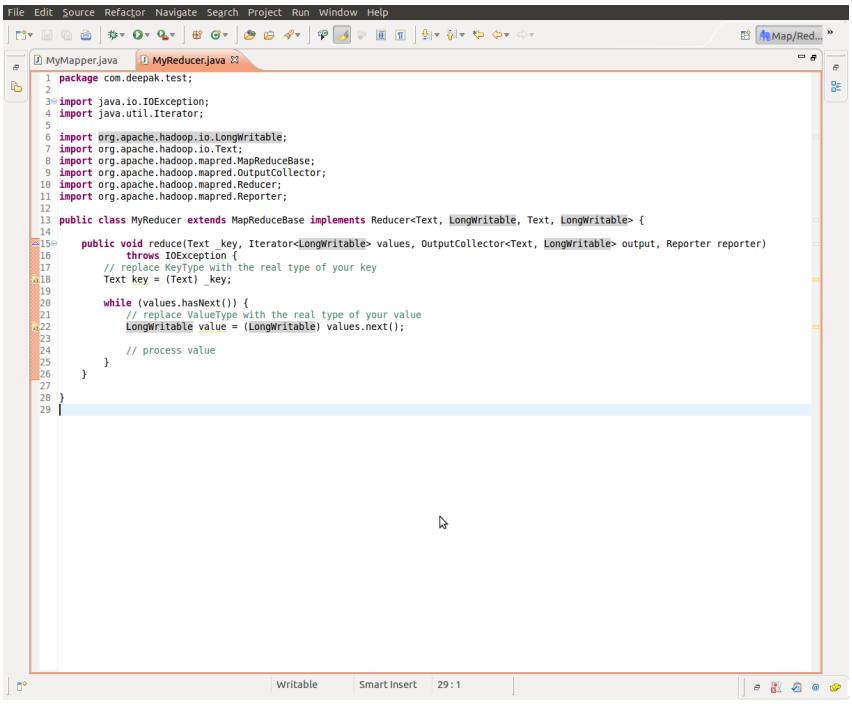
You can write your code for mapper class here.

13. Add Reducer class



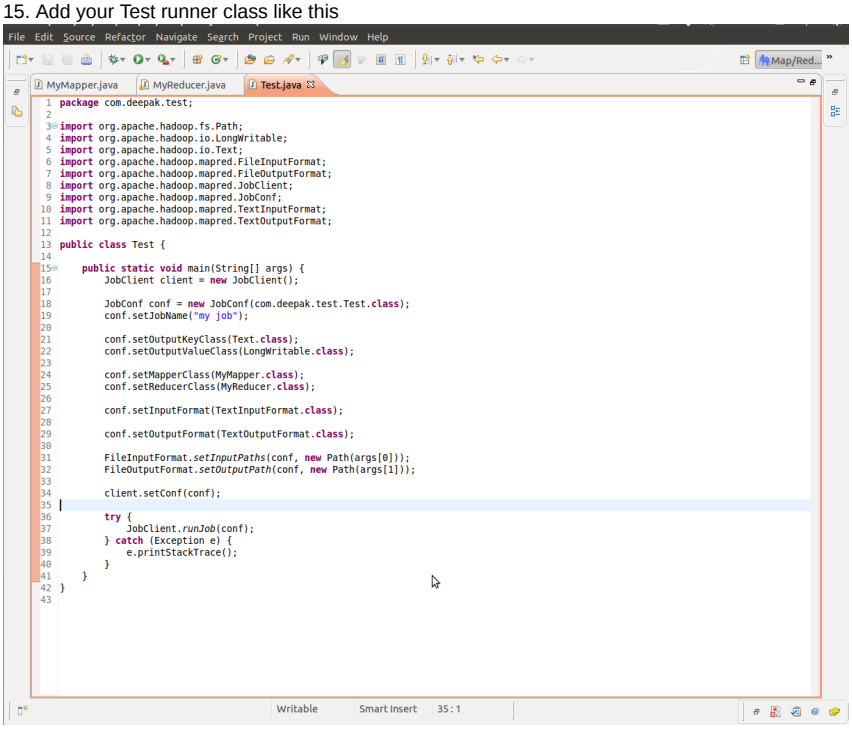


14. Your Reducer class should look like this

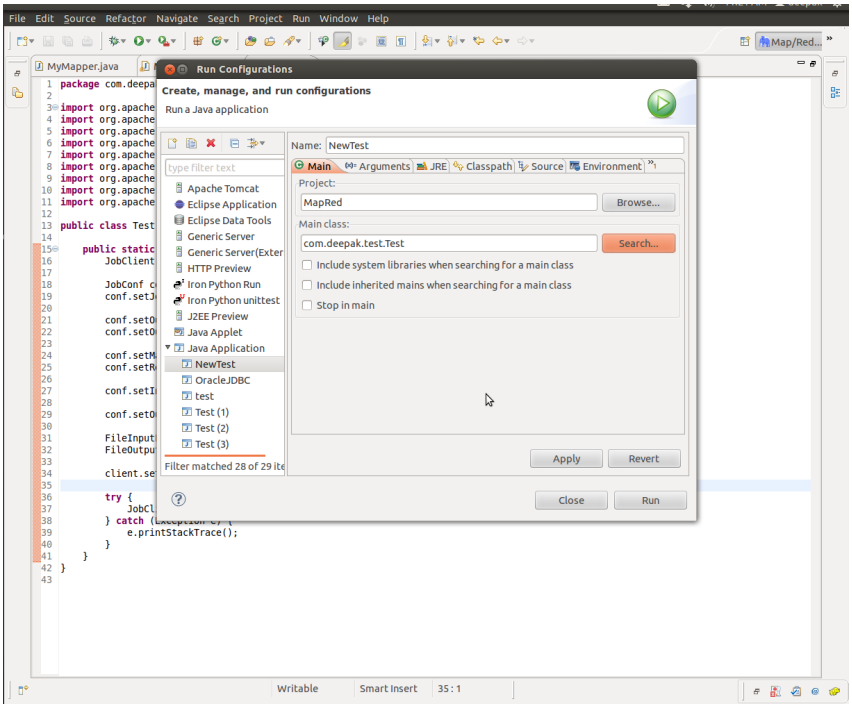


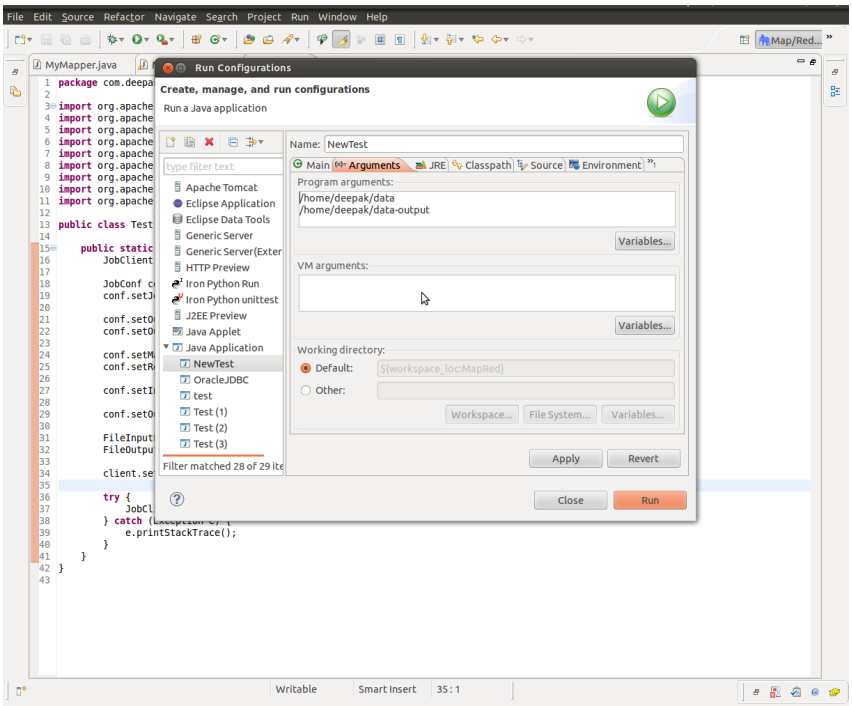
You can add your code for reducer class here

15. Add your Test runner class like this



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

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

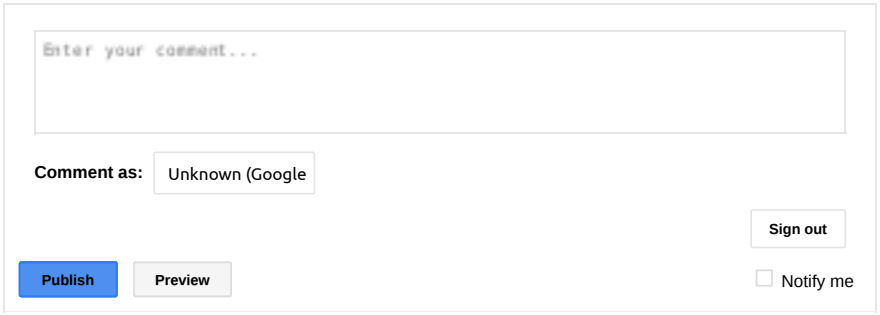
Amirtha rao December 22, 2015 at 3:42 PM

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



Newer Post Home

Subscribe to: Post Comments (Atom)

Picture Window template. Powered by Blogger.