**Big data Analytics**

# **CS5590BD**

**Lab-1 Assignment**

BY

L V SANDEEP, BATHINA (16174348)

Task1: Apps

**Submitted as a group**

Task 2: Cloudera

* **Subtask 1: How to access UMKC Cloudera servers ?**

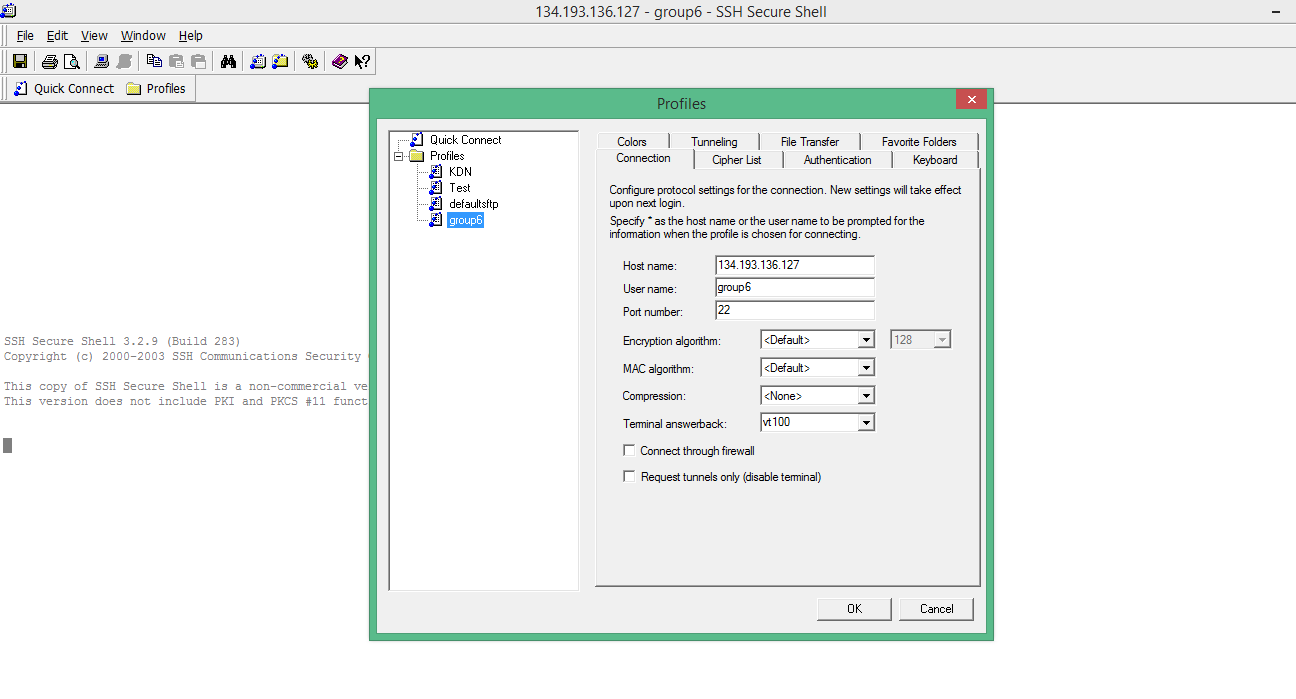
**Steps:**

1. Download and Install Putty/SSH from the web and enter the

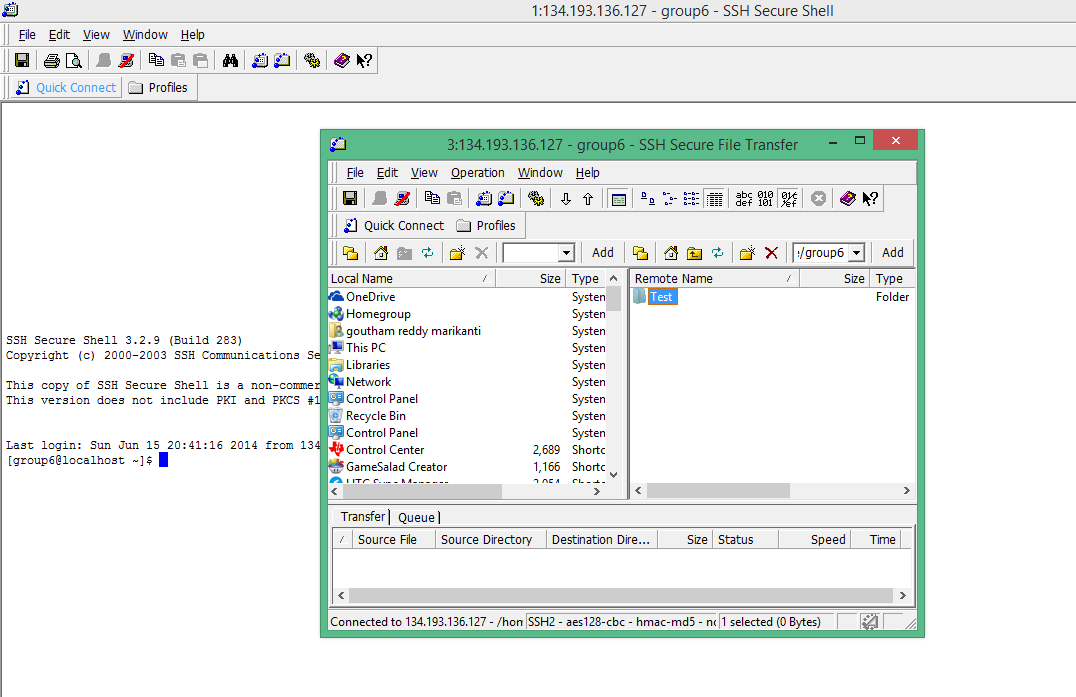
IP address: 134.193.136.127

Username: group6

Port: 22



1. Now after getting connected enter your group password. In this way we can connect to the UMKC Cloudera server.



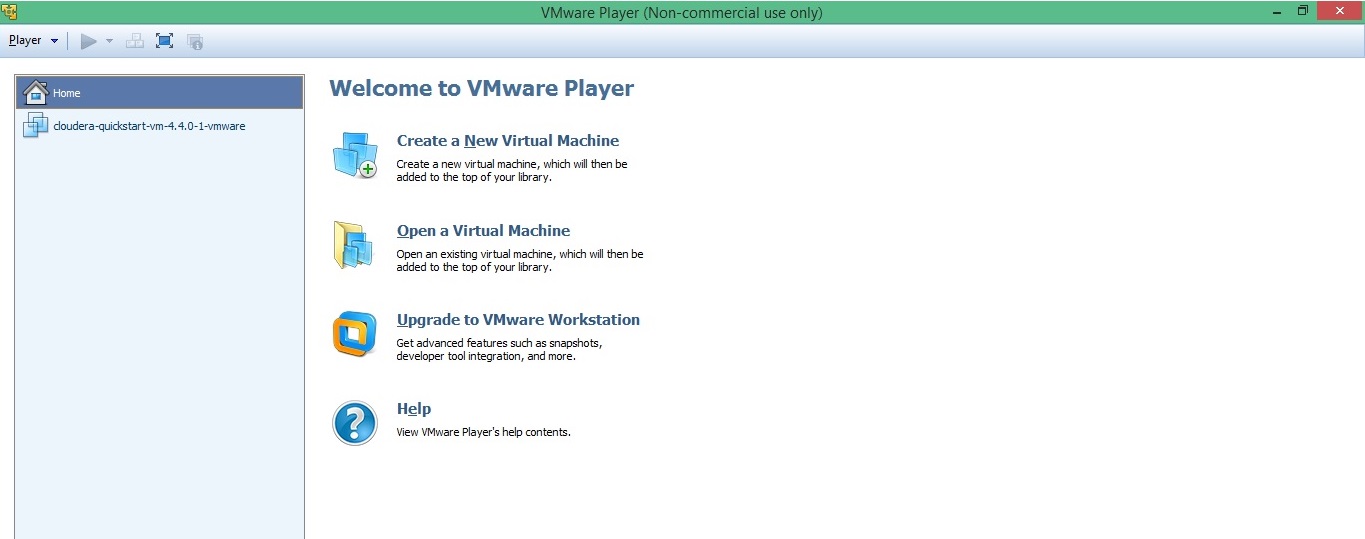
1. In order to enter the home page of Cloudera enter the below URL

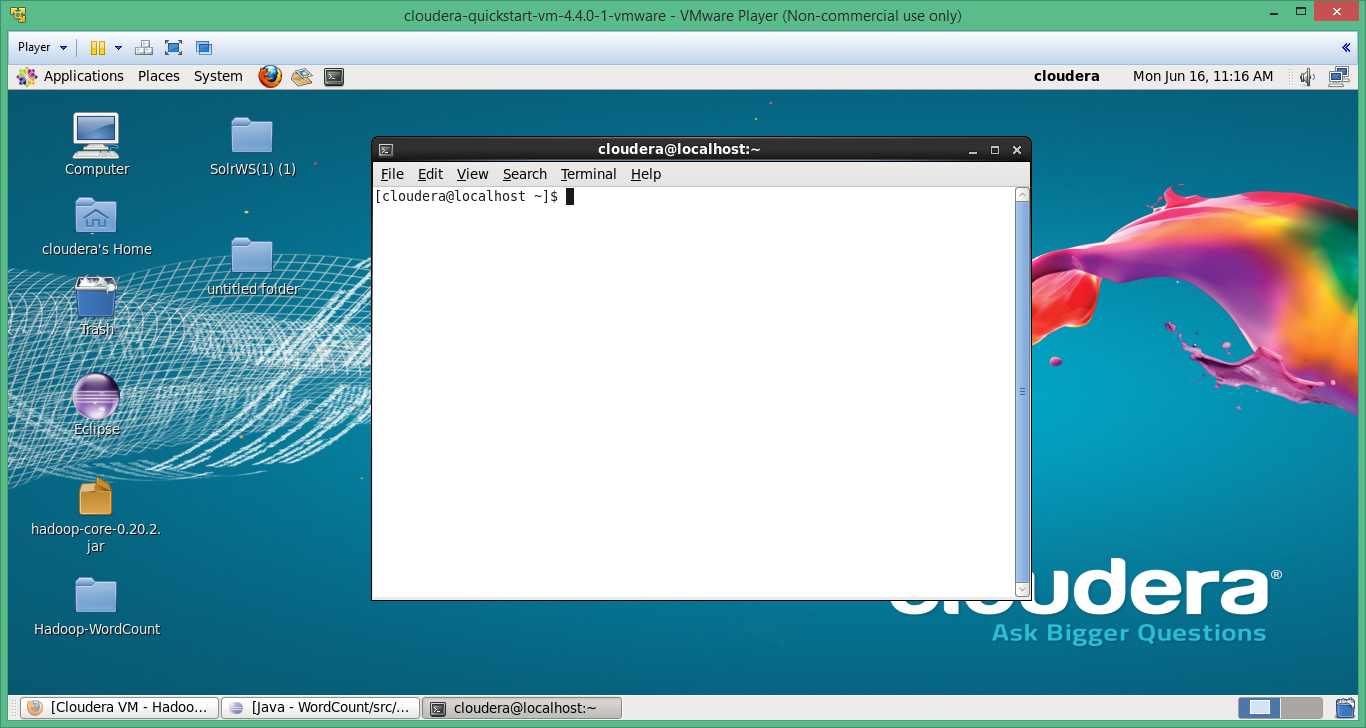
<http://134.193.136.127:7180/home> and enter the credentials

* **Subtask 2: How to install your own Cloudera?**

**Steps:**

1. Download and install VMPlayer from the internet
2. Download Cloudera 4.5 image from the net and upload this image in VMPlayer
3. Now Open the Cloudera image

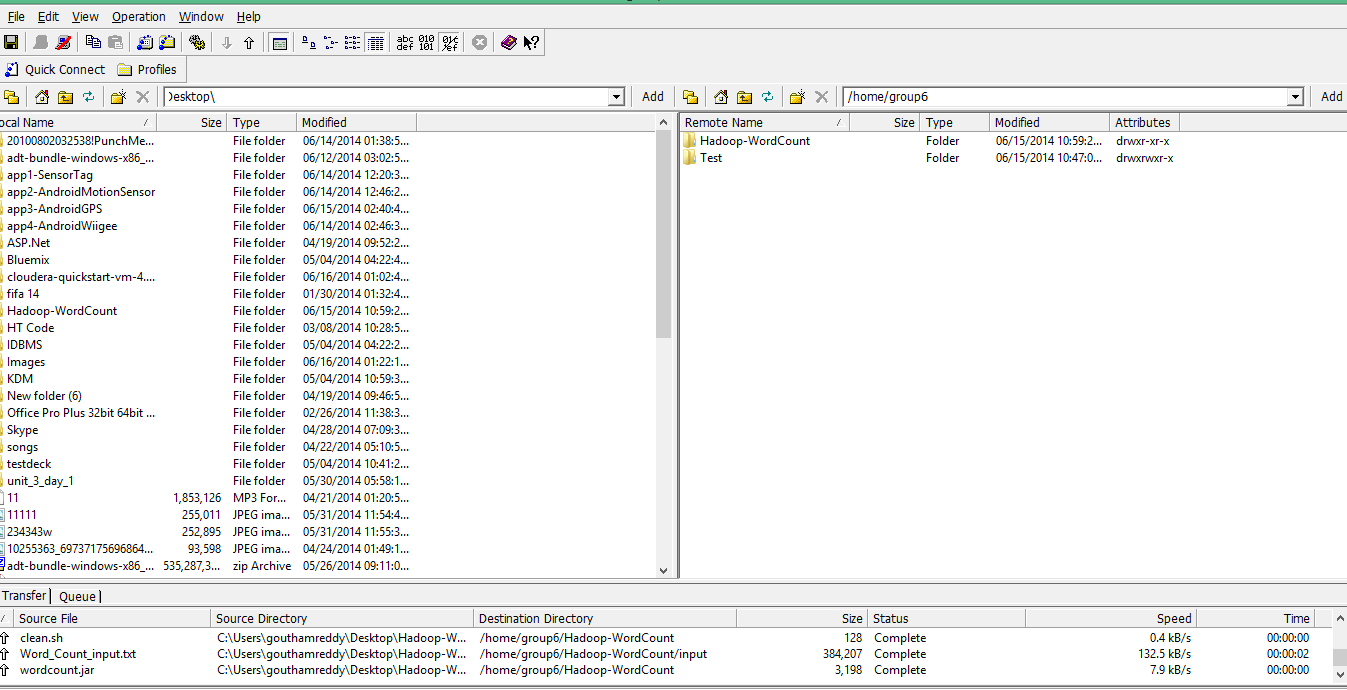




* **Subtask 3: How to transfer files to Cloudera?**

**Steps:**

1. In order to transfer files from our local machine to the Cloudera server, we need to install winscp for putty.
2. After that just drag and drop your local files to the Cloudera window.

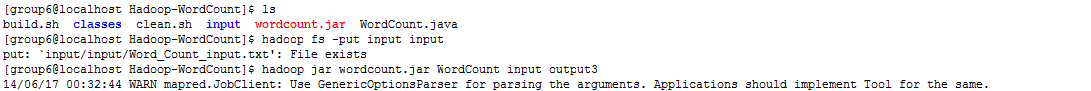


* **Subtask 4: How to run a program Word Count on Cloudera?**

**Steps:**

1. open the putty/ssh shell, run the below commands

Initially download word count source code from the provided link and copy the word count folder to the Cloudera server.



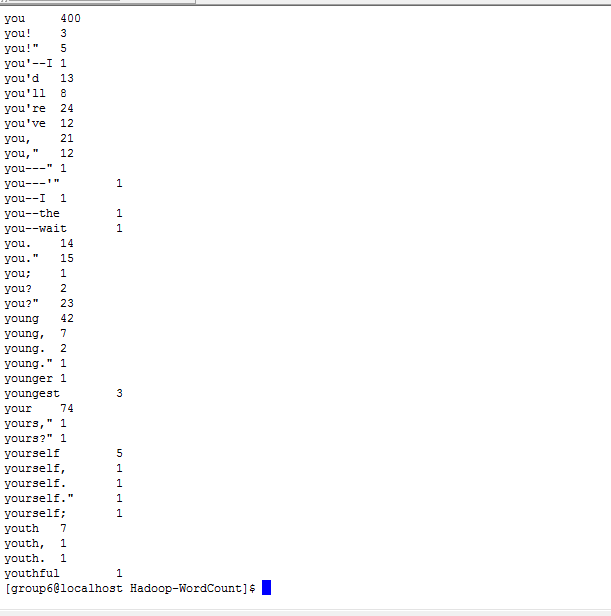
1. $ ls

$hadoop fs –put input input

$ hadoop jar wordcount.jar WordCount input output

$hadoop fs –cat output/\*

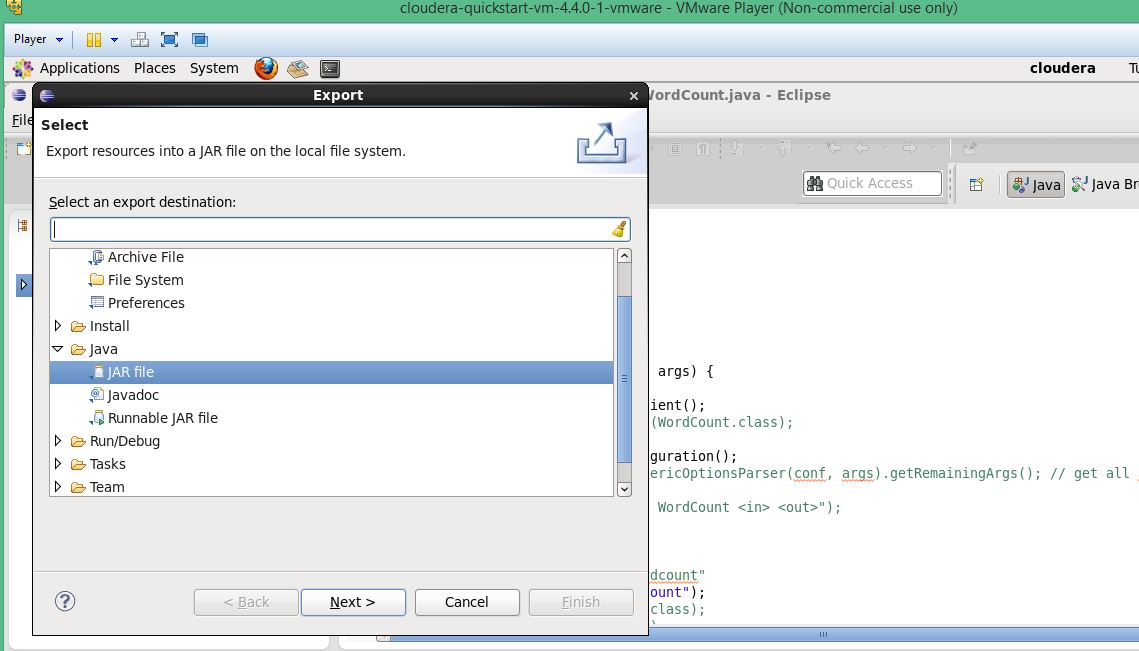
Input folder contains the input file and output folder contains the output of the word count code. The first command copies input folder from local to hadoop file system. The second command runs the code in hadoop environment and the third command displays the output present in output folder.



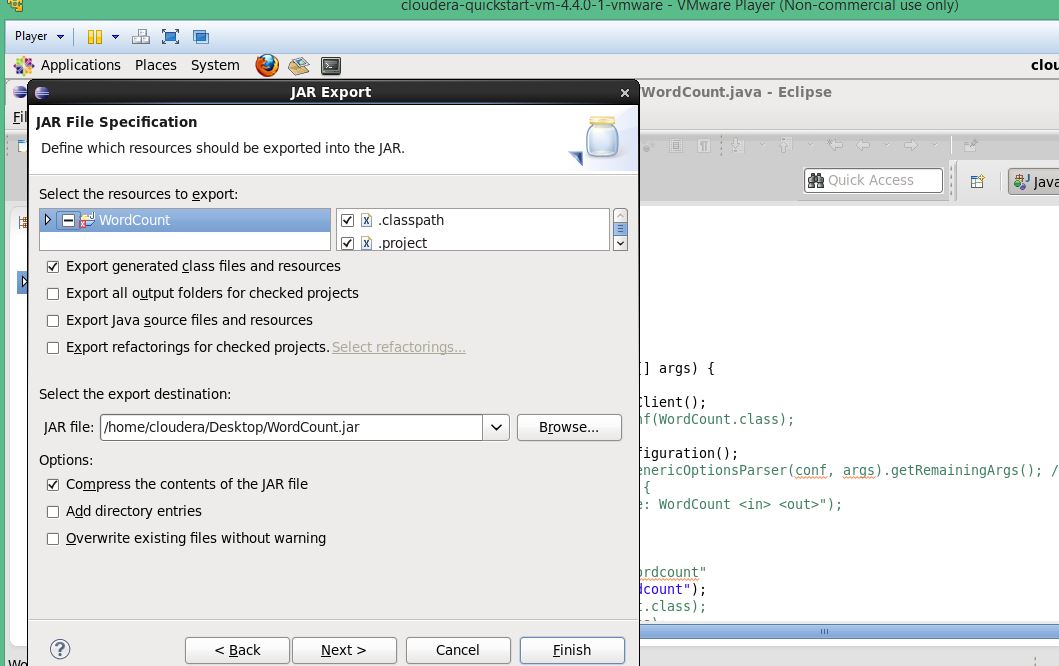
* **Subtask 5: How to make a hadoop jar and run it in Cloudera**

**Steps:**

1. Import the word count project file in to eclipse.
2. Export the project to jar file by right click on project-> export->Jar



1. Continue with the steps as prompted



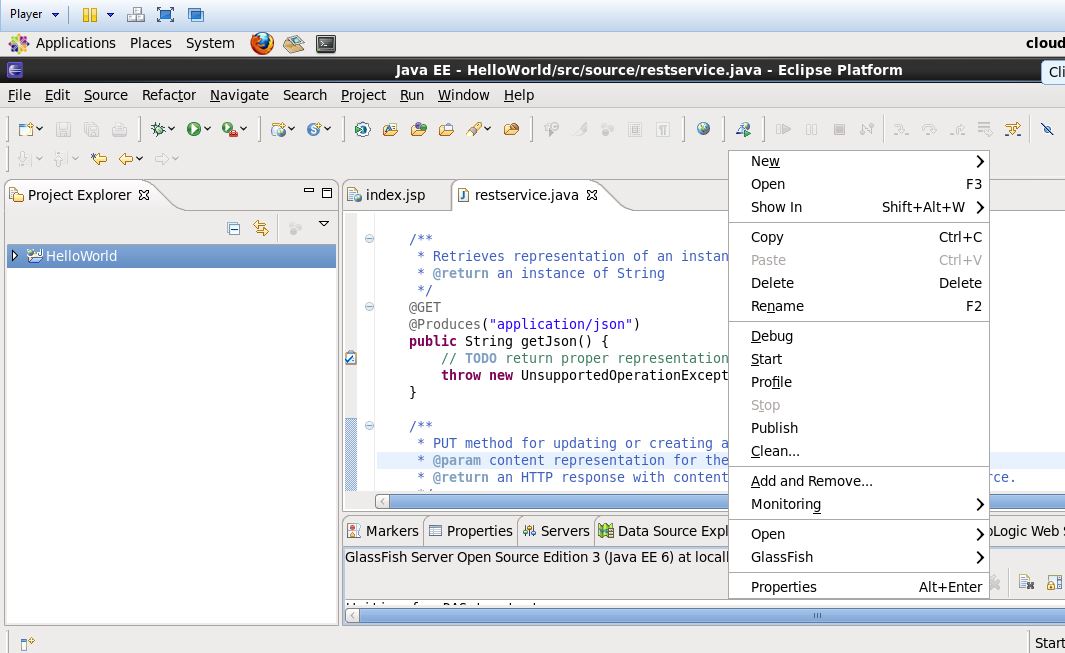
1. Now run the below command in Cloudera to display the output

$ hadoop jar wordcount.jar

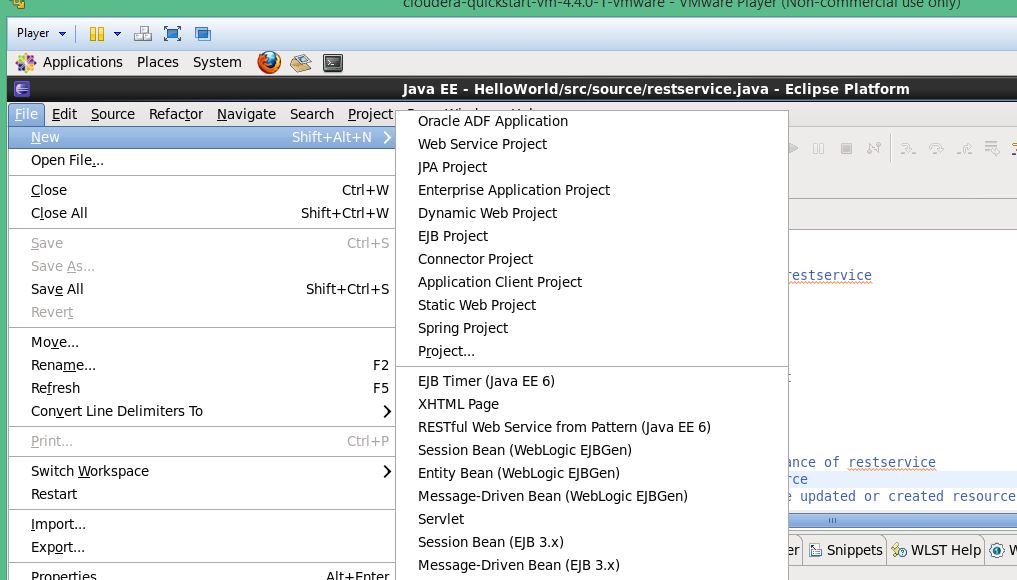
* **Subtask 6: How to Build a java based restful service in Cloudera**

**Steps:**

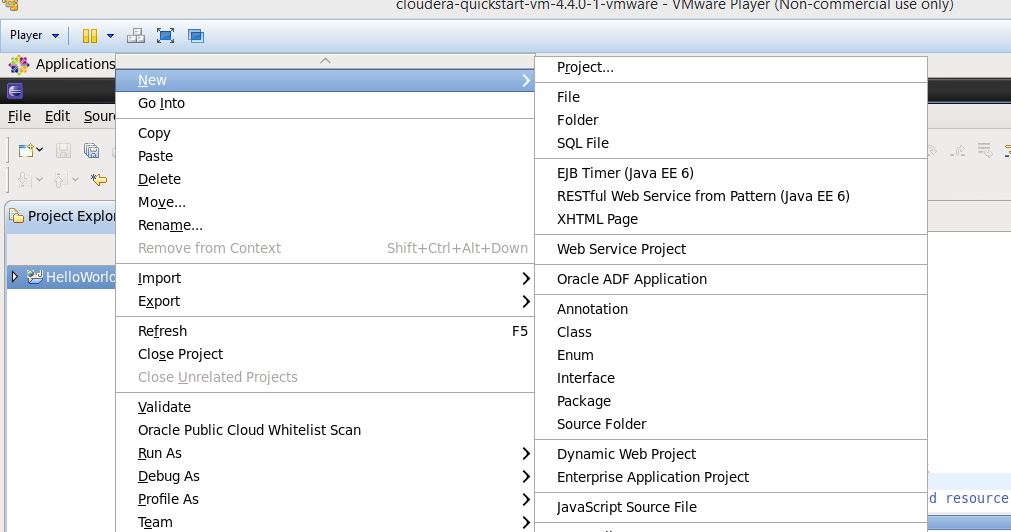
1. Initially Install glassfish server in the eclipse



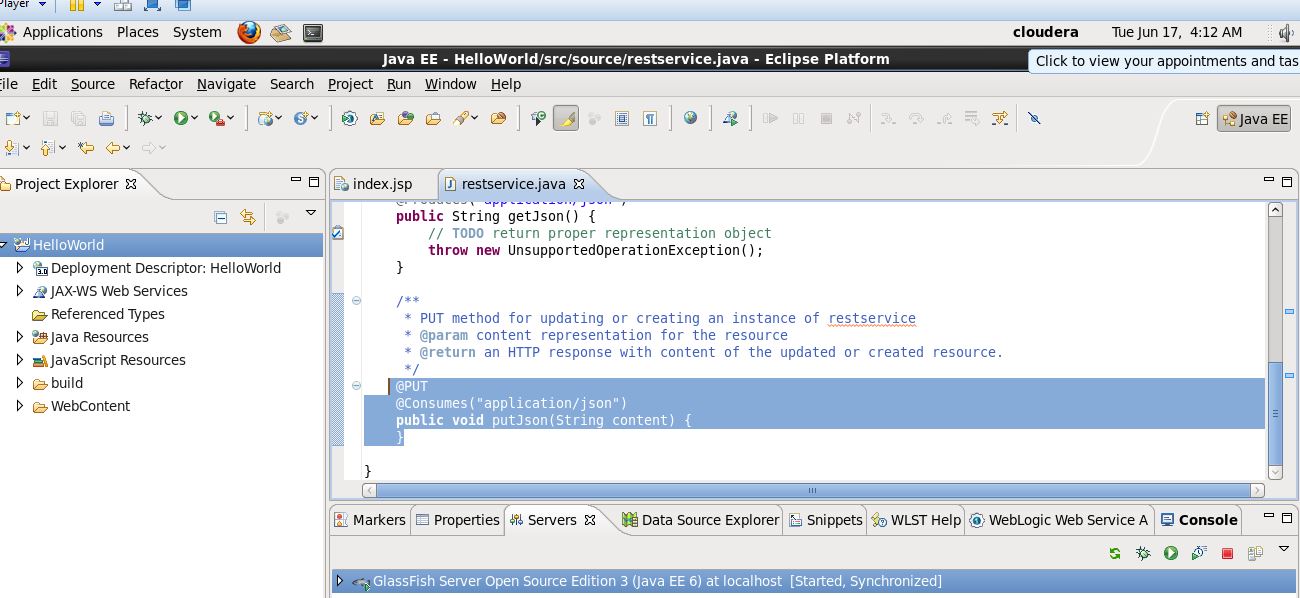
1. Create a Dynamic web project



1. Start the glassfish server and now create a new restful web service



1. Import the word count project file and export this to WAR file



1. Now deploy this war file on to the glassfish server and run the below

<http://192.168.133.129:8080/WordCount/jaxrs/generic/ViewResult/output>

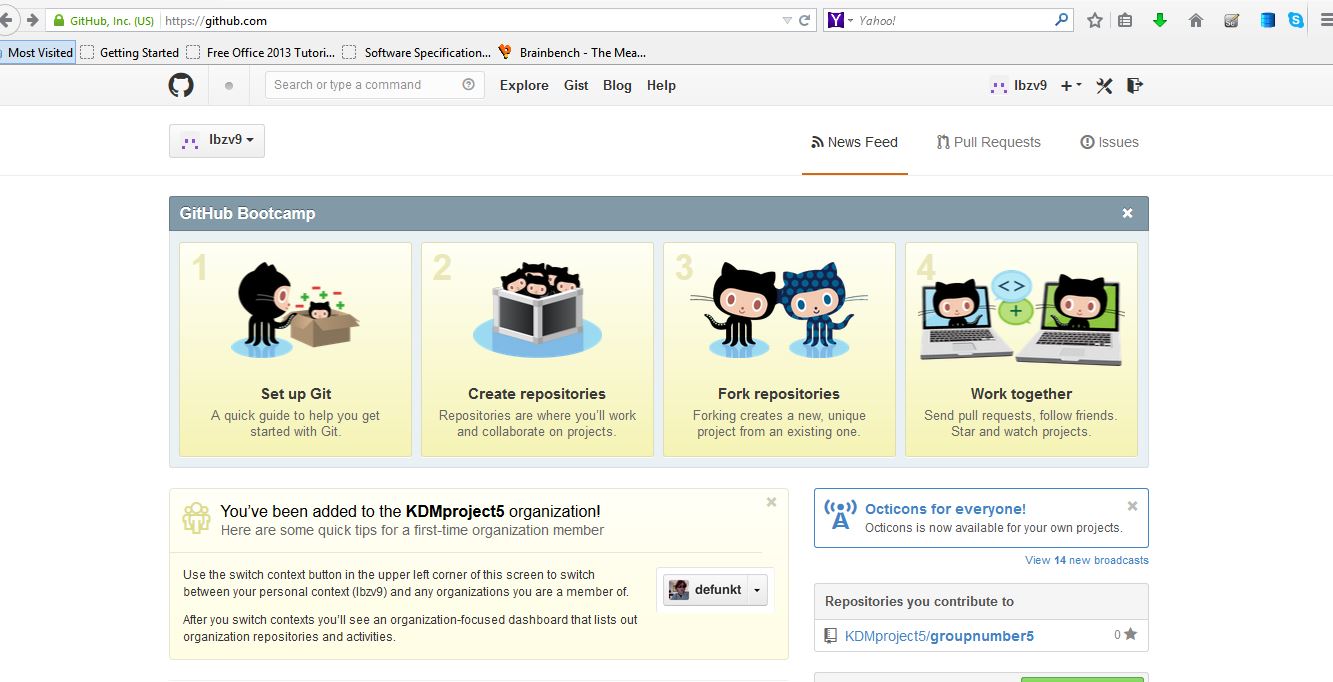
Task 3: Github and Scrumdo

* **Subtask 1: How to create an account and deploy files to Github**

**Steps:**

1. Create an account in Github site.

Username: lbzv9



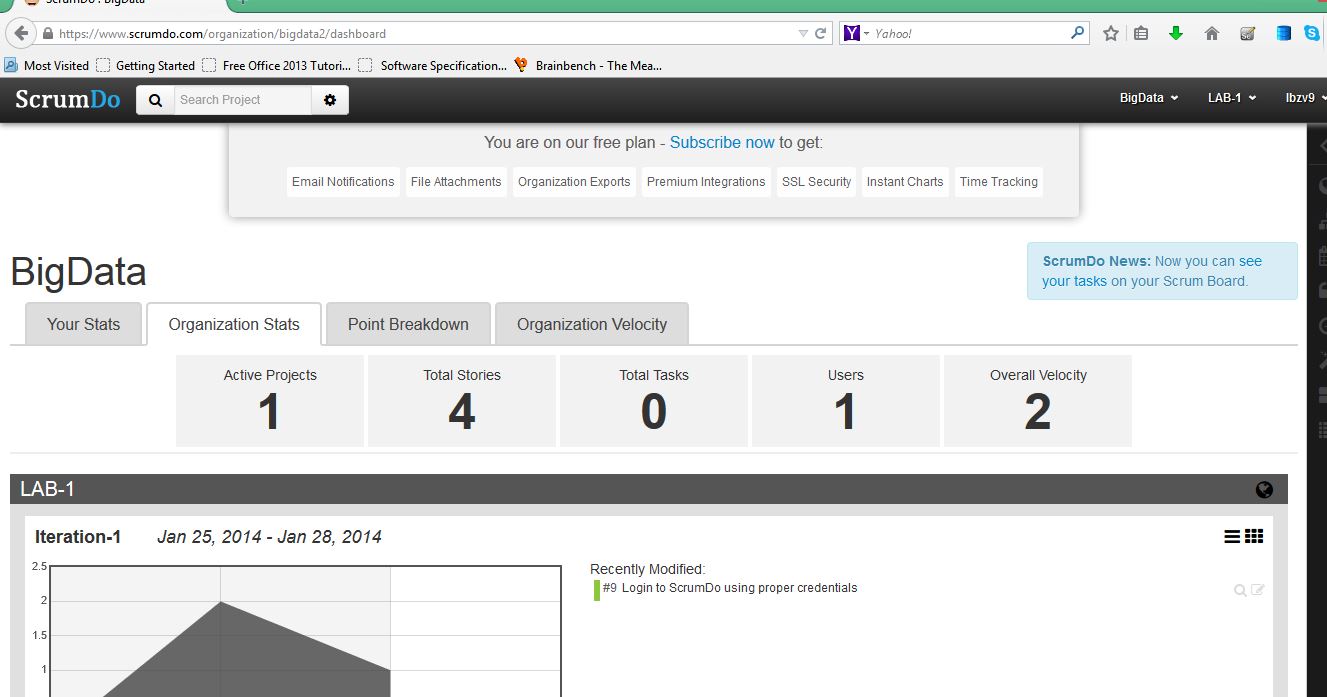
1. Install Github on windows
2. Now create new repository and open explorer window in the settings
3. Copy the files in to the window and commit the changes made.
4. Now we need to synchronize the files on github cloud, so click on sync in github window.

* **Subtask 2: How to create an account and design projects with scrumdo**

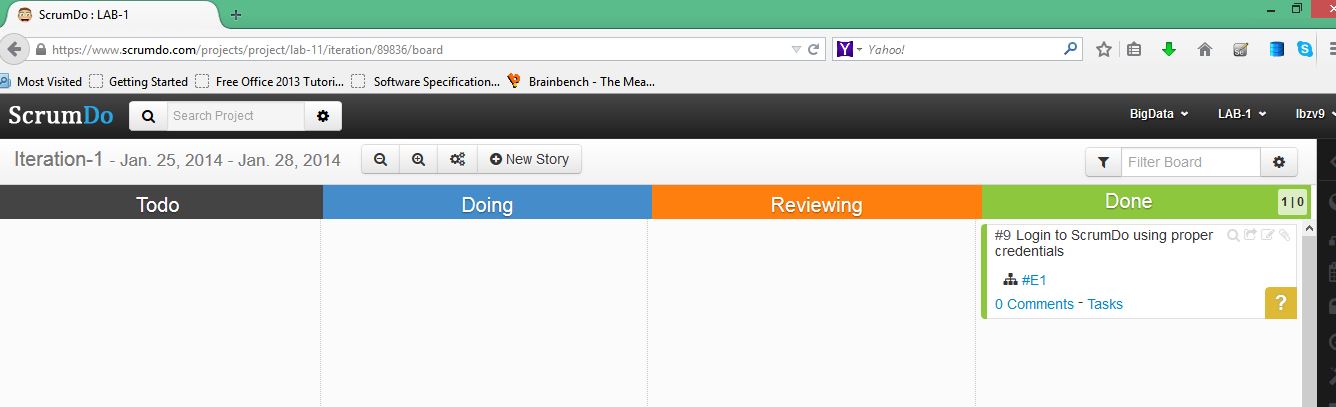
**Steps:**

1. Create Scrumdo account

Username: lbzv9



1. Create an organization in the account
2. Now we can create iterations and stories over there which will be helpful for the project.



1. On a regular basis we need to check our scrumboard and make changes accordingly.