CS590BD Big Data Analytics and Apps

**LAB - 1**

**Report on Installation of Cloudera virtually and in own machine, Execution of Word Count, Restful services and Github and Scrum do Account Deployment**

**By**

**Ponnam Balakrishna**

**16177831(BP8G6)**

**Task-2:**

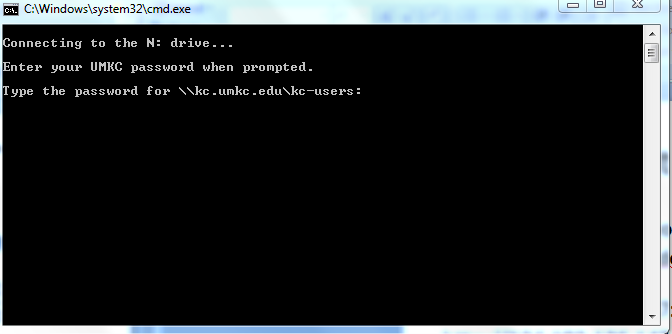
**Subtask – 1:**

**UMKC VM – Installation:**

To install UMKC VM first we need to connect UMKC VPN. To connect to VPN we need to download VPN and need to login with the SSO.

The login screen looks as below:





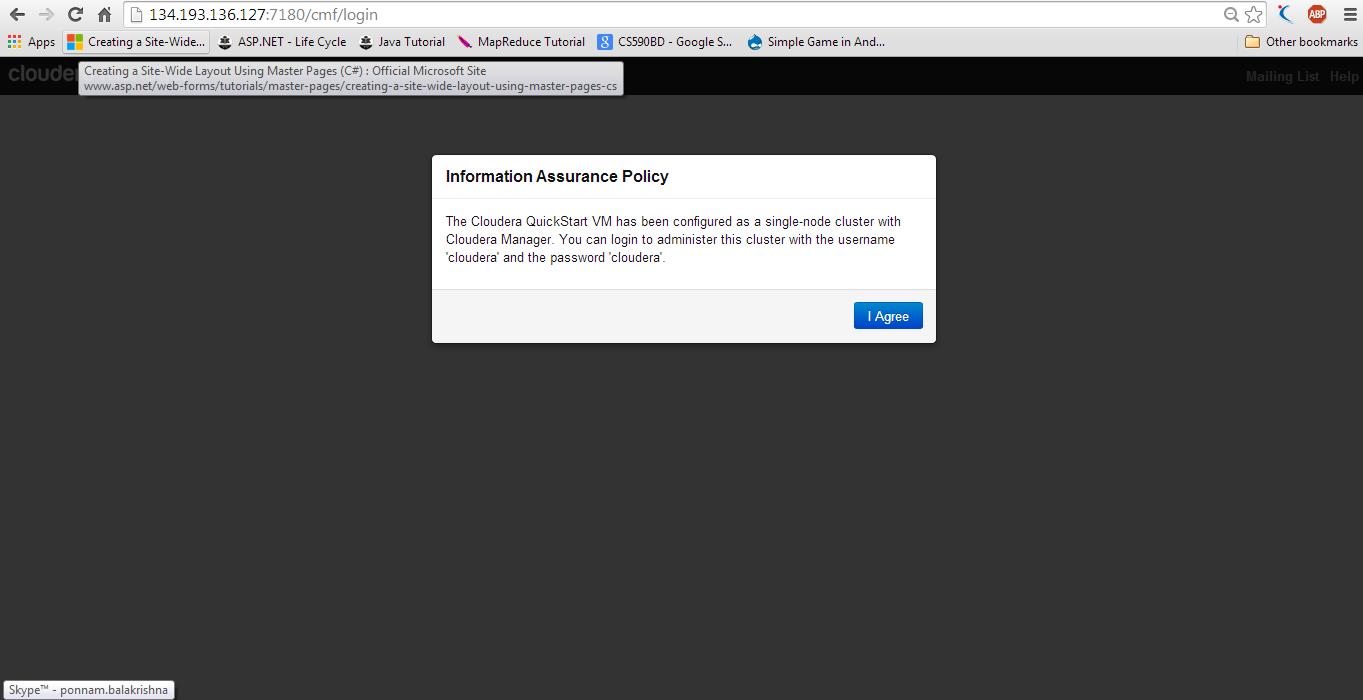
To login to the cloudera manger, we need access the following URL

<http://134.193.136.127:7180>

7180 is the port number to reach the cloudera manager page.

On accessing the above URL we get the cloudera manager page.

The cloudera manager page looks as below:

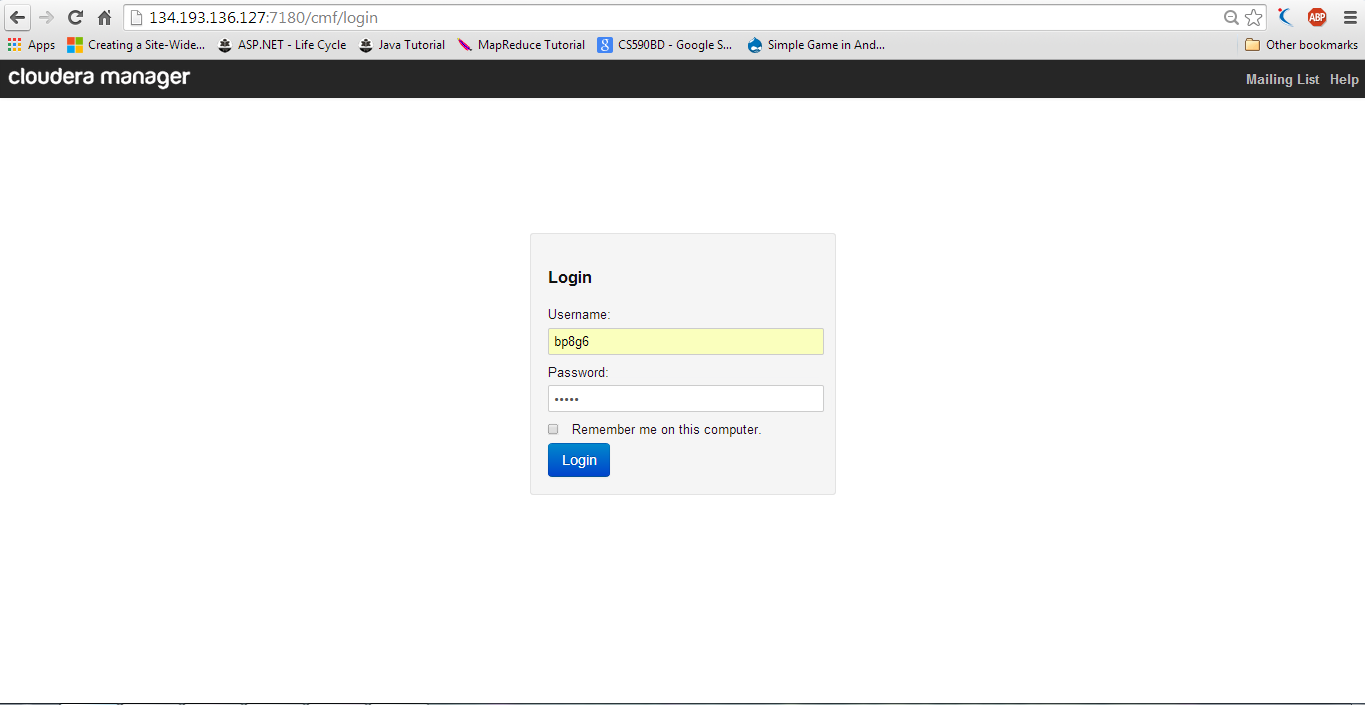


On clicking I Agree we will be redirecting to the login page of cloudera where we need to login with our SSO as our username and password.

Those SSO is the college credentials where the access is available to connect the virtual machine remotely.

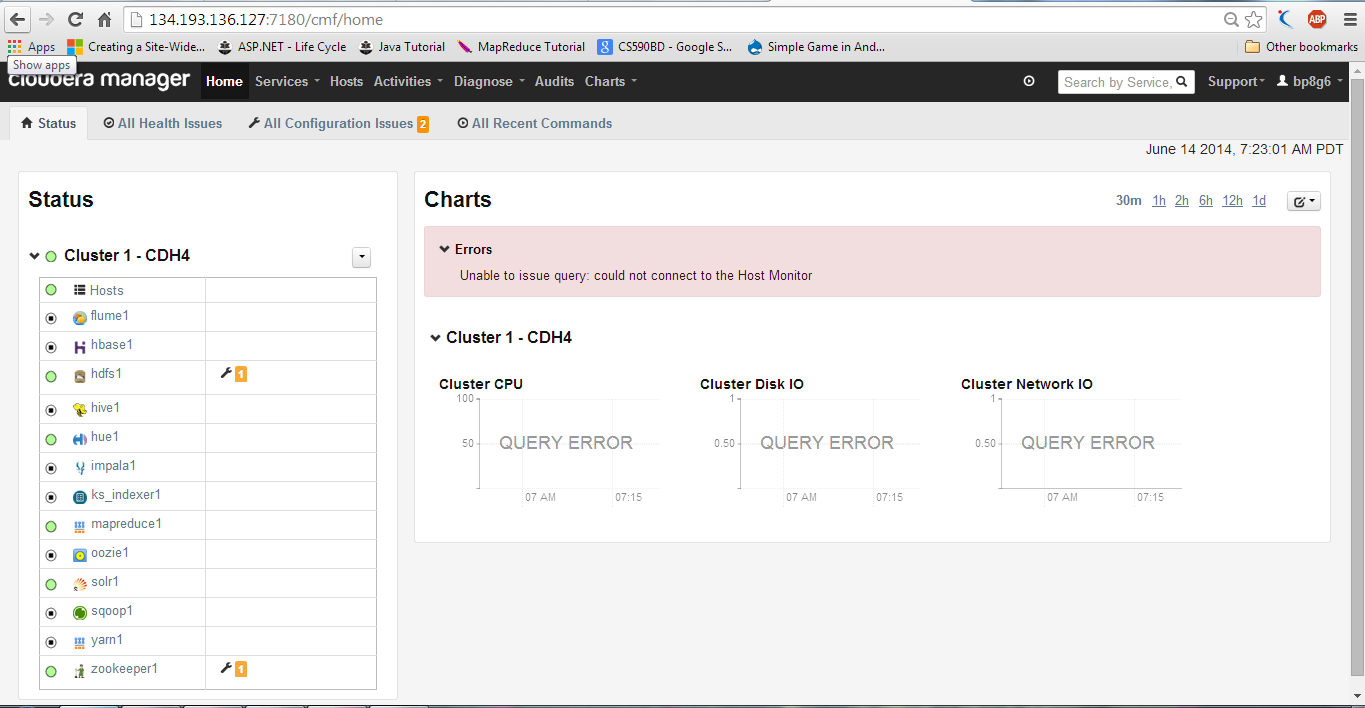
After the authentication it opens the list of services provided where we can use all the services.

The login page looks as below:



On logging to cloudera manager we get the list of all hadoop ecosystems which are the services used in the cloudera listed in the page indicating the health. If it says it is having good health we can work on it and if it says bad health we need to restart and then after making it as good health we can work on it.

After logging to the cloudera manager the screen shows as below:



**Subtask -2:**

**Installation of cloudera in your local machine:**

To install Cloudera we have to download VM Player.

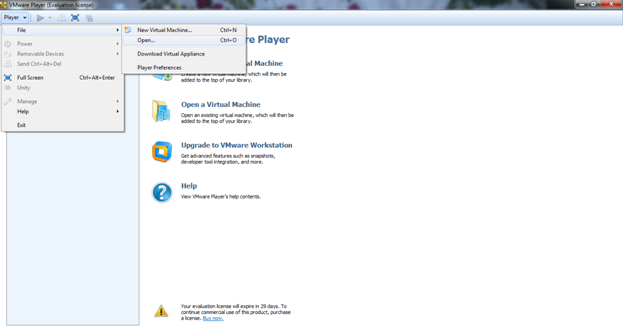
VM player can be downloaded from the <http://www.vmware.com/products/player>

Then after installing VM player we need to download cloudera image 4.5 which is in zip format, unzip the image and then need to open the image from VM player.

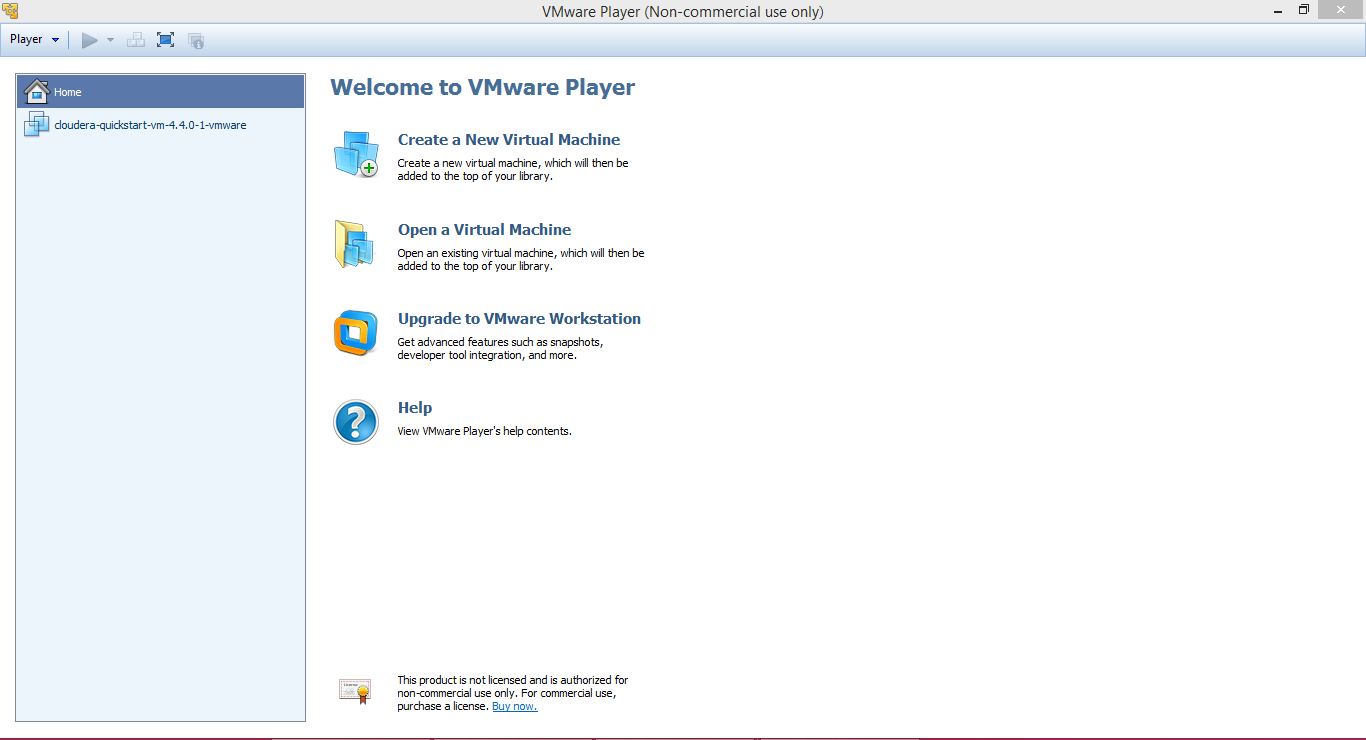
Cloudera image can be downloaded from <http://www.cloudera.com/content/support/en/downloads/download-components/download-products.html?productID=F6mO278Rvo>

After installing VM Player goto File ---> Open open the image cloudera-quickstart-vm-4.4.0-1-vmware.

The below screen shot shows how to open the image file.

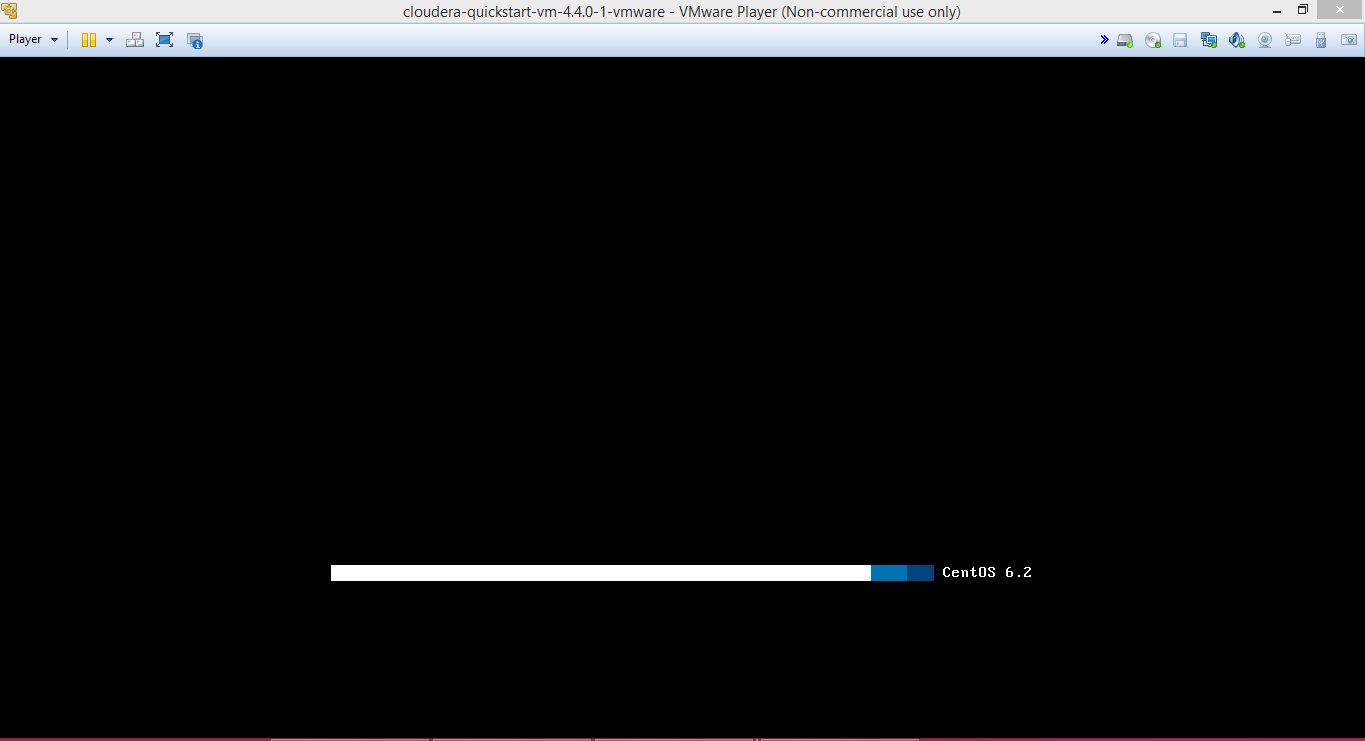


After the Image is opened once it is stored and from second time onwards it shows as in the below screen



After opening the image cloudera will be installed and will boot to open the virstual machine.

The below screen shot shows the booting of virtual machine

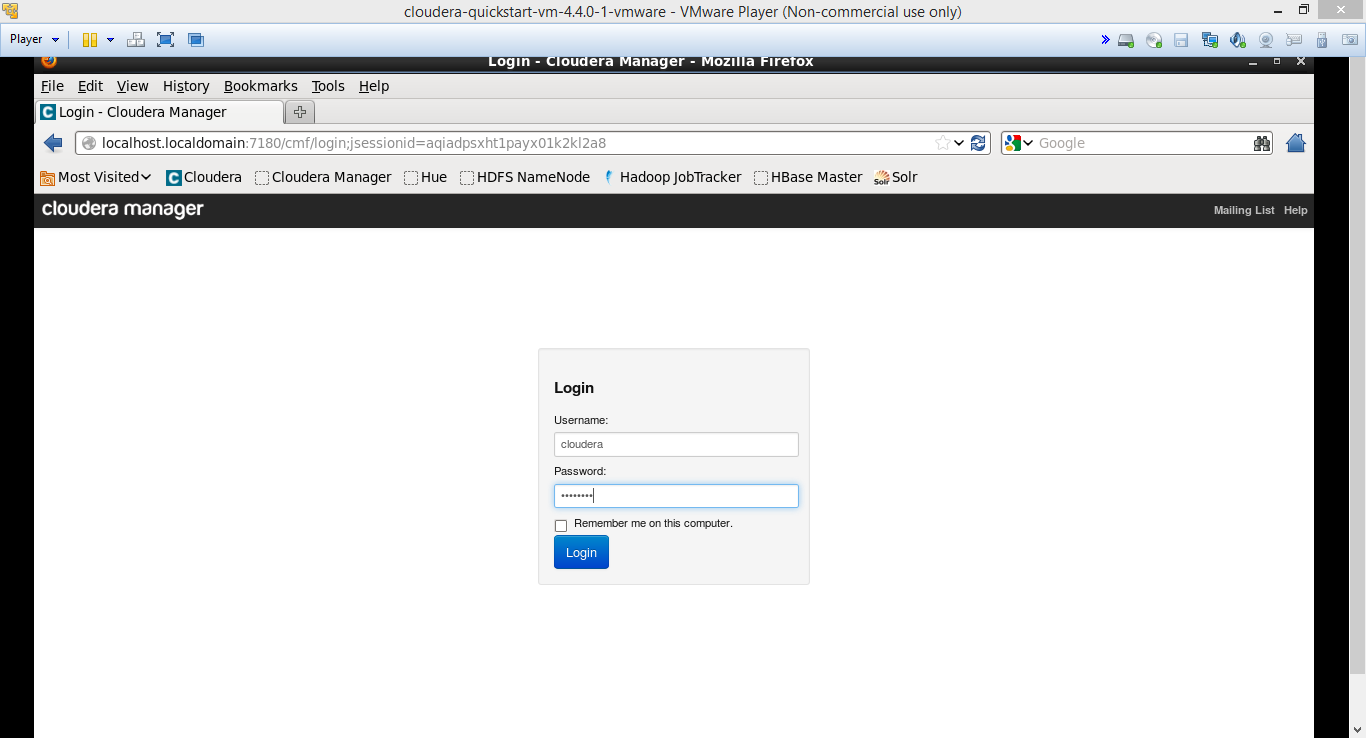


After booting is done it opens the firefox browser with different tabs named as cloudera,cloudera manager,Hue,HDFS NameNode,etc….

It asks for the login.

Login with Username and Password as “Cloudera”.

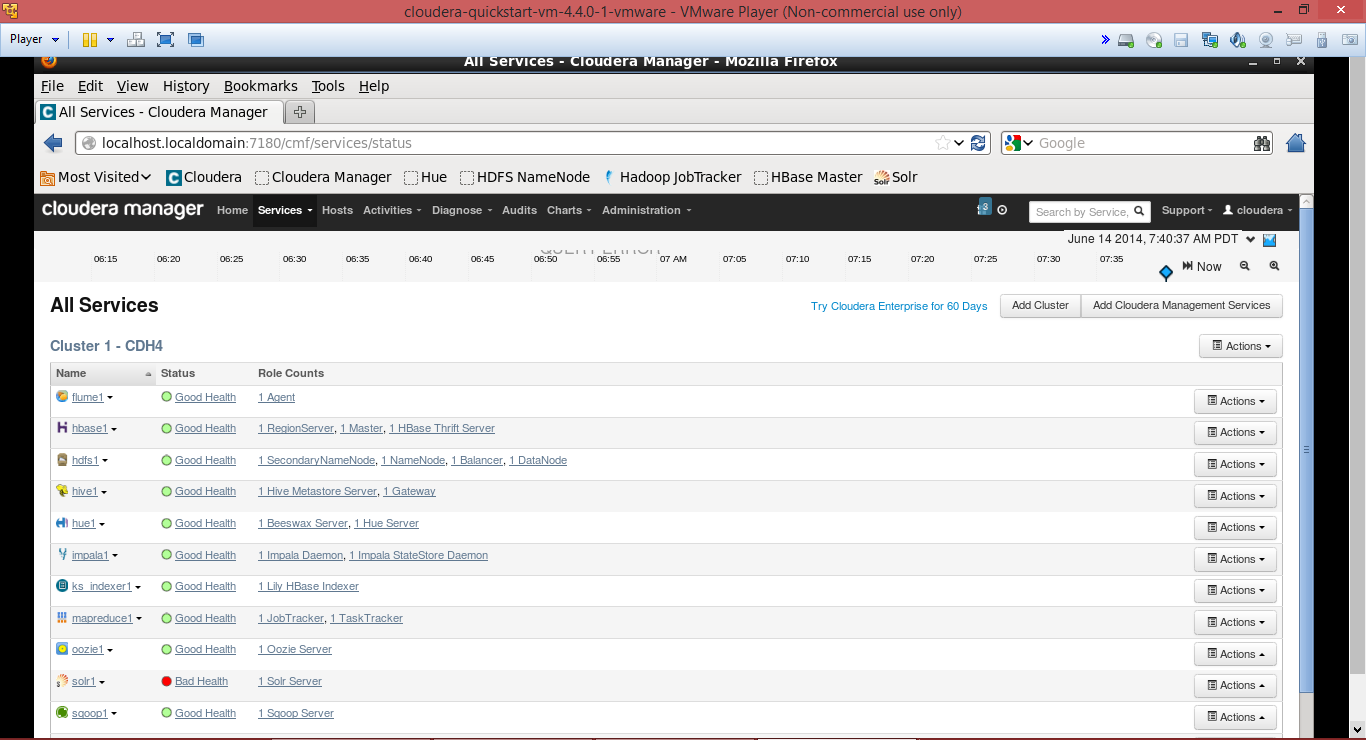
The below screen shows the firefox browser with different tabs,logged screen and all the services.



Once the Login is done it lists all the services of cloudera and also indicates the health of the particular service.

If it shows as bad health we can restart that service after it becomes good health we can work on them.

The below screen shot shows the list of all the services.



Those in the above list are the services. And those services also show the health of particular service.

The health is represented in terms of good and bad. If it shows good health it is ready to use.

If it shows health as bad we can restart the service and after coming to good health we can use that service.

**Subtask -3:**

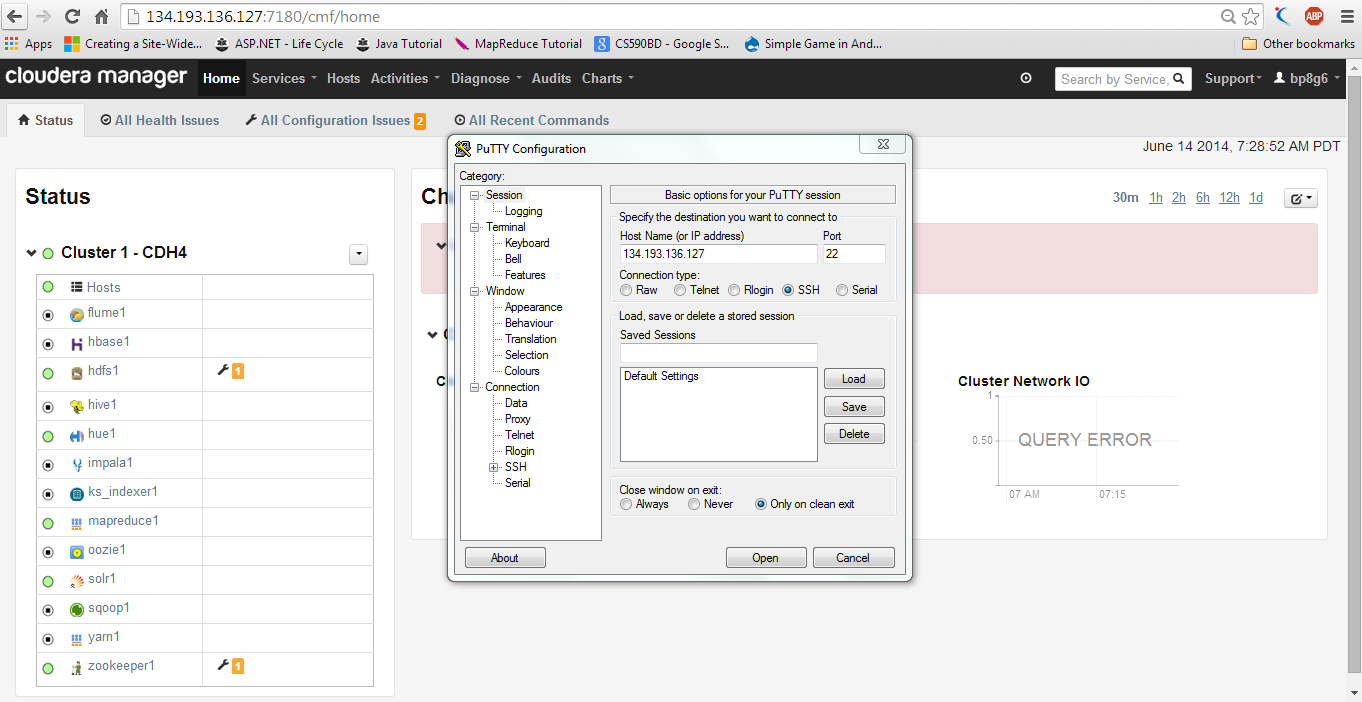
**File transfer to Cloudera:**

We need to download the putty and install it.

Putty can be downloaded from the following URL:

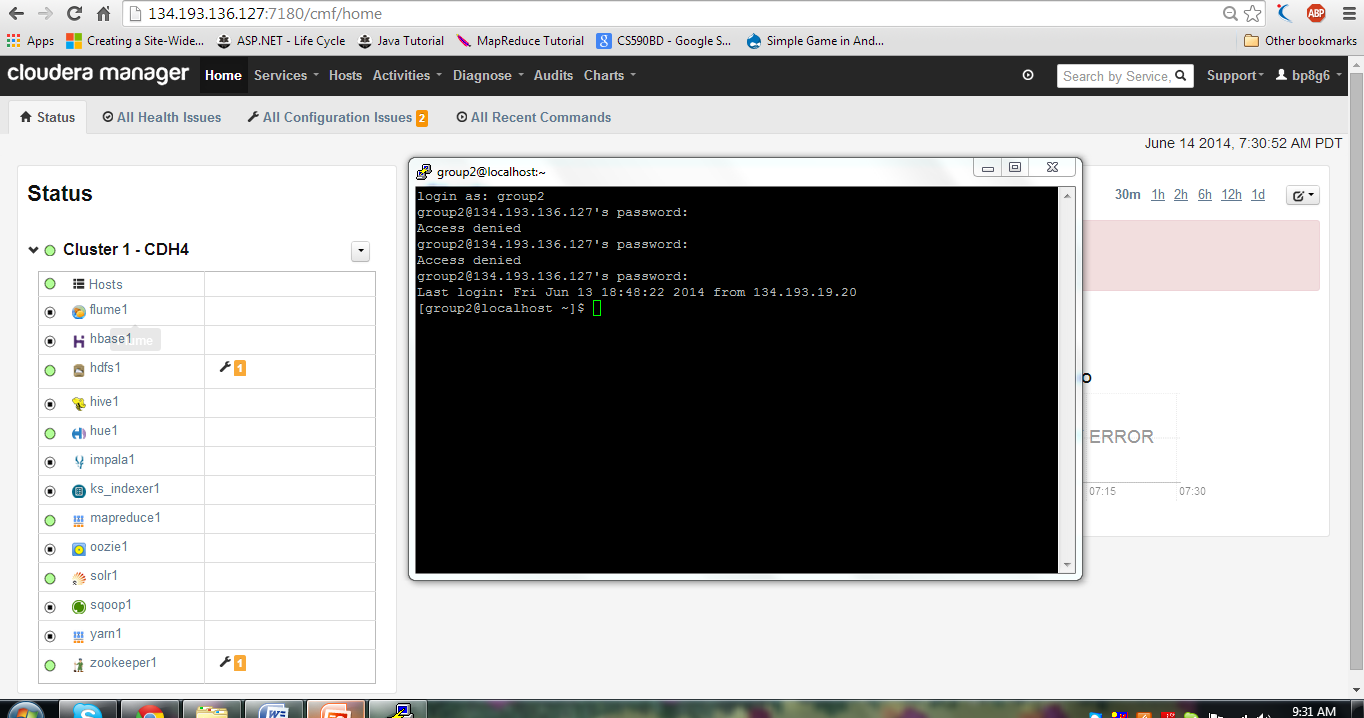
<http://www.chiark.greenend.org.uk/~sgtatham/putty/download.html>

For accessing remotely you need to give the host name to which you are connecting



After connecting you can acess the group where there is some space allocation where we can store the data.

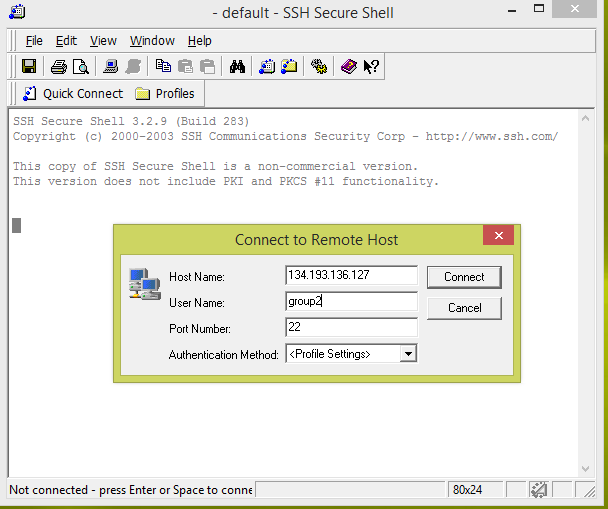
The below screen shows the connecting to the group by giving the user name and pwd.



WinSCP is a very easy way to transfer files remotely from win to linux

It can be downloaded from <http://winscp.net/eng/download.php>

If we press Quick Connect present on the window it will show a pop-up for connection to the remote host.Enter the host name and username as shown below and connect it.



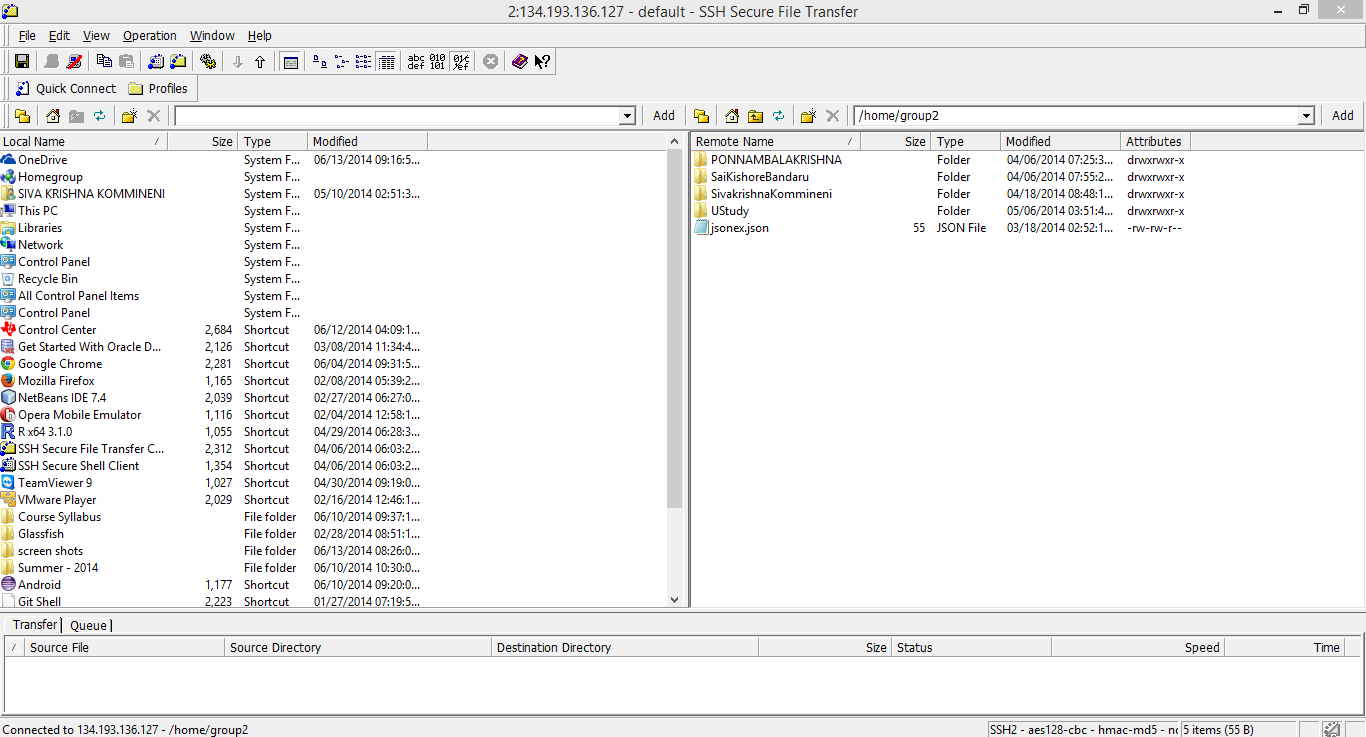
After connecting it then asks for the password for authentication

The authentication screen looks as follows:



Later you can transfer files directly using the file transfer .If we click on that icon shows the both areas side by side, just we need to drag and drop the files required from left window to right window.

The below screen shows the file transfer:



**Subtask – 4:**

**Executing word count example:**

Now to execute word count example Download the source code of word count which is downloaded in Zip format.

The word count source code can be downloaded from the following link.

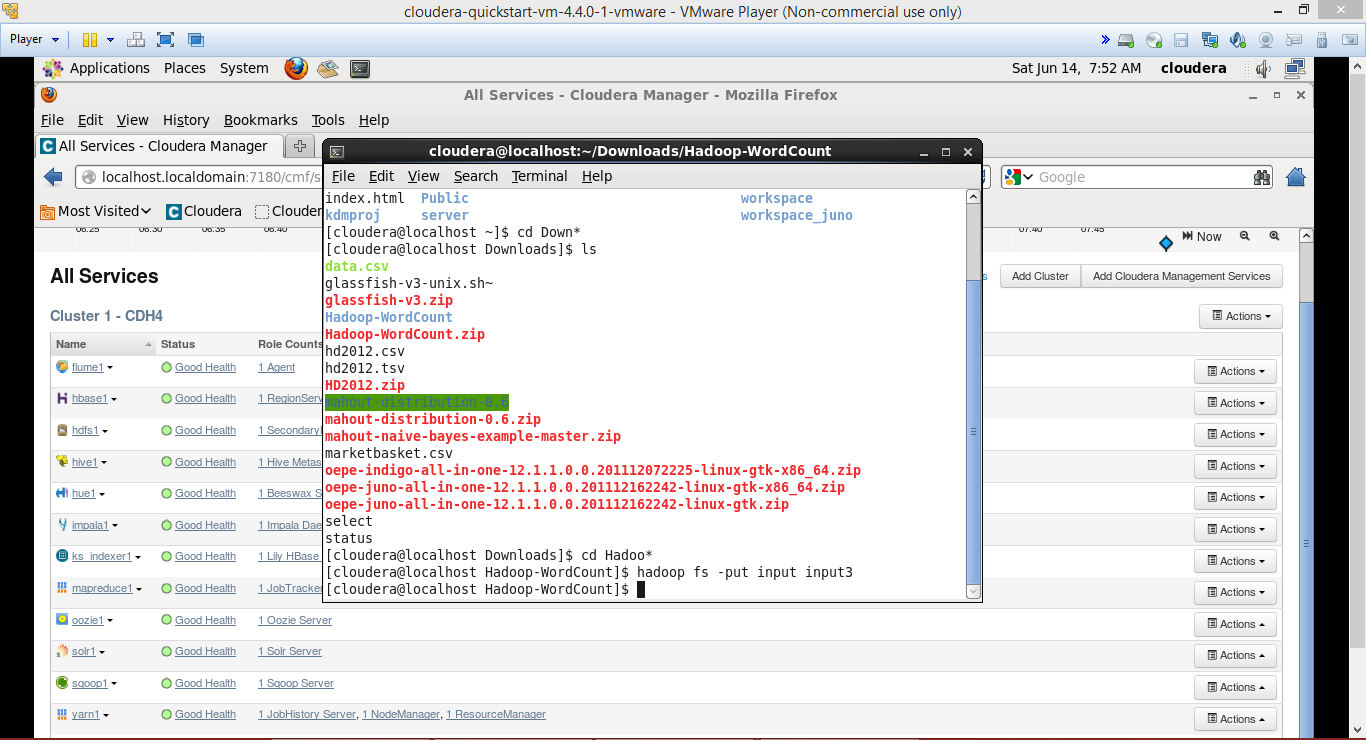
<https://portal.futuregrid.org/manual/hadoop-wordcount>

Unzip the word Count File by running the command **Unzip WordCount.Zip in** the terminal.

Then put the local input file to the hadoop input directory using the command

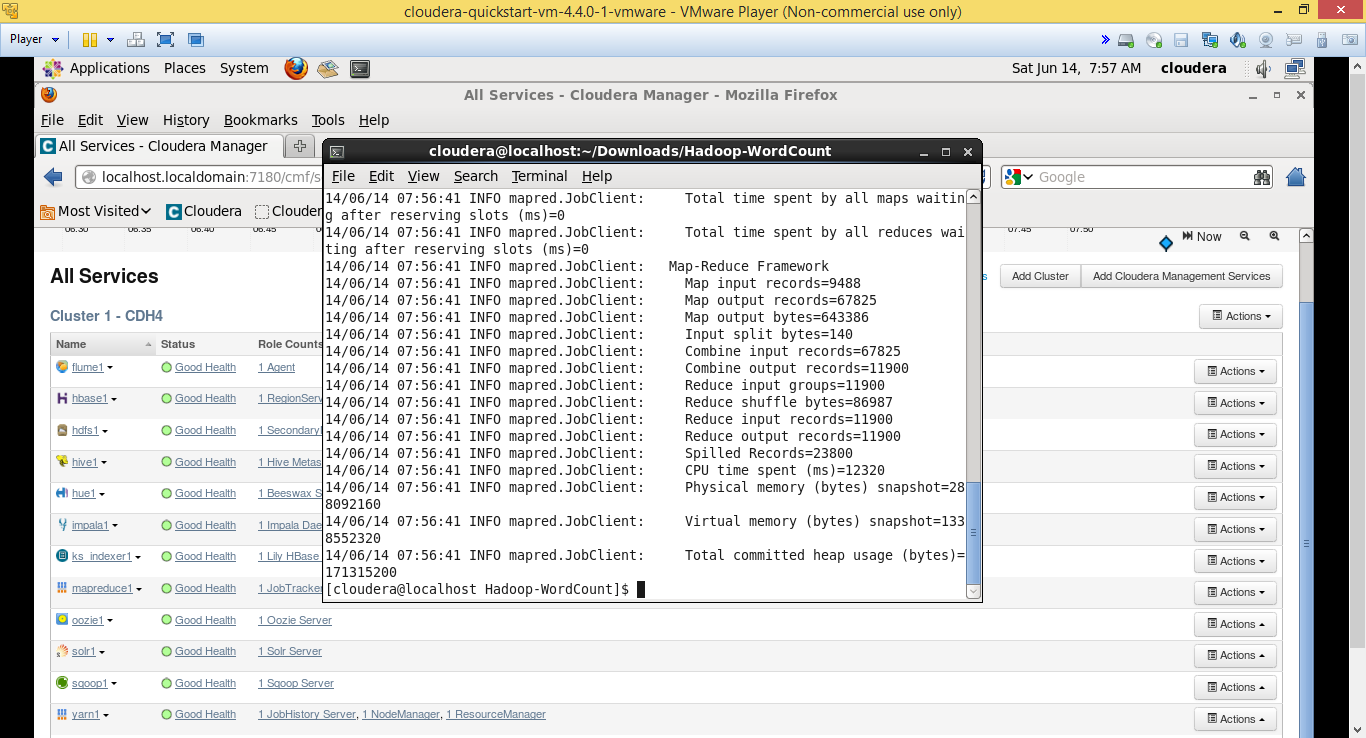
**hadoop fs –put input input.**

On executing the above command the screen shows as below



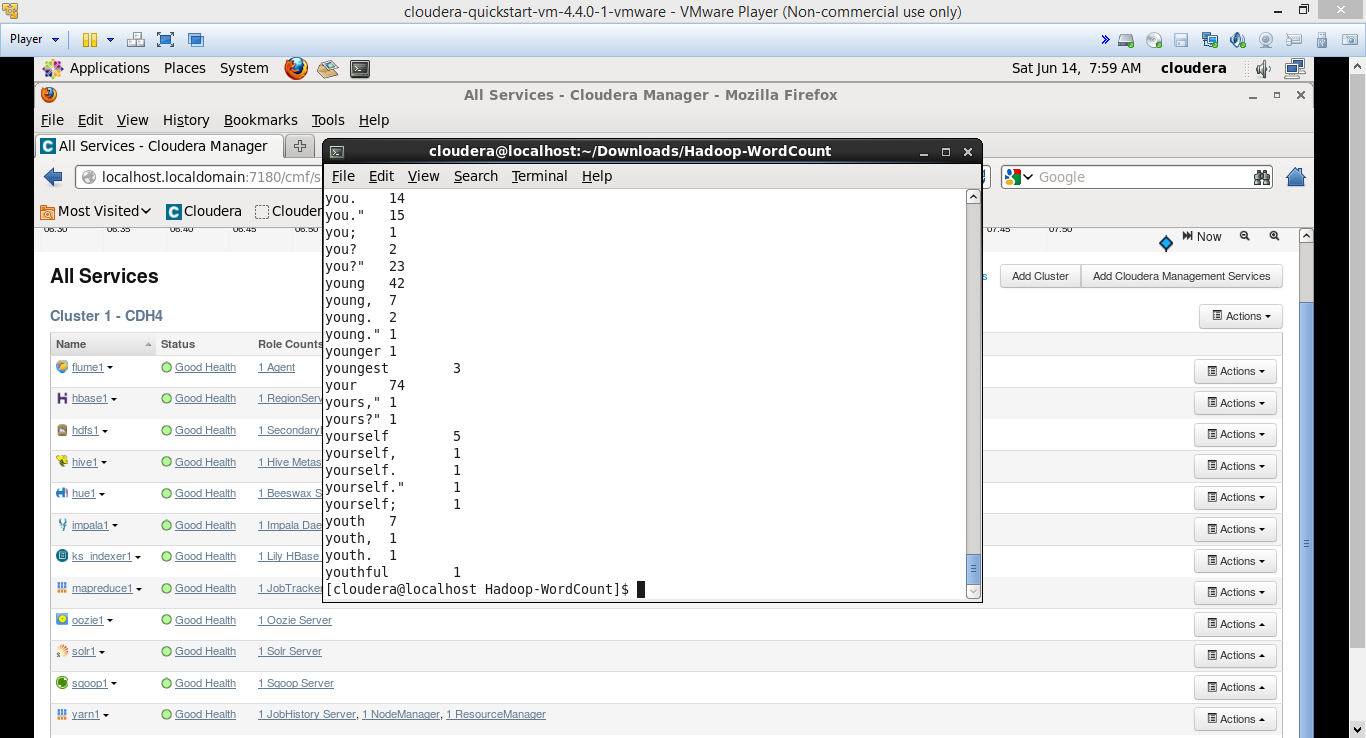
Next run the hadoop using the command **hadoop** **jar wordcount.jar WordCount input output**.

On Executing the above command the screen is shown as below:



To see the result from output run the command **hadoop fs –cat output/\***

The below screen shows the result of word count example:

****

The above word Count Example reads all the sentences in the text file and gives the result as number of words used in the total document.

It shows the each word used number of times in the coulomb.

The above screen shot shows the Word and count of the words used in the text document.

**Subtask – 5:**

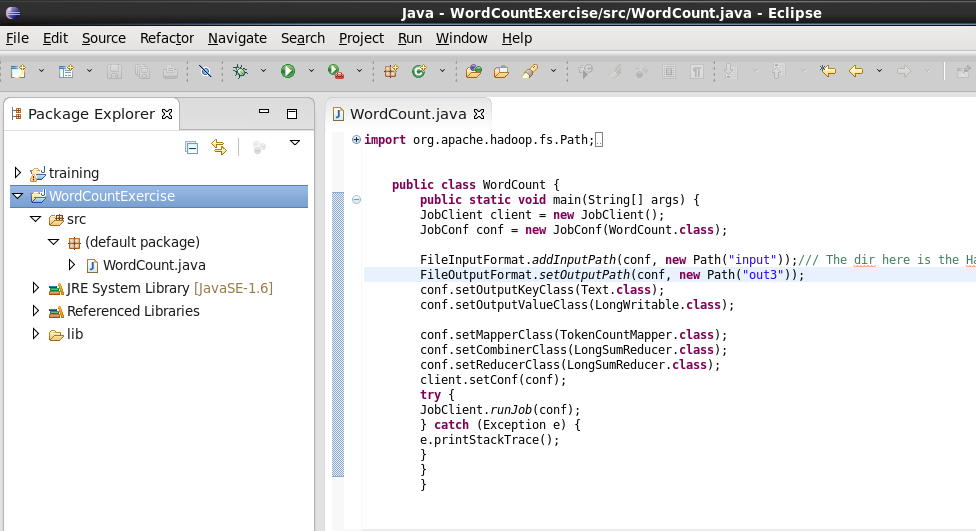
**Hadoop wordcount and run in cloudera:**

To run Hadoop word count jar in cloudera we need to import the word count exercise project in eclipse.

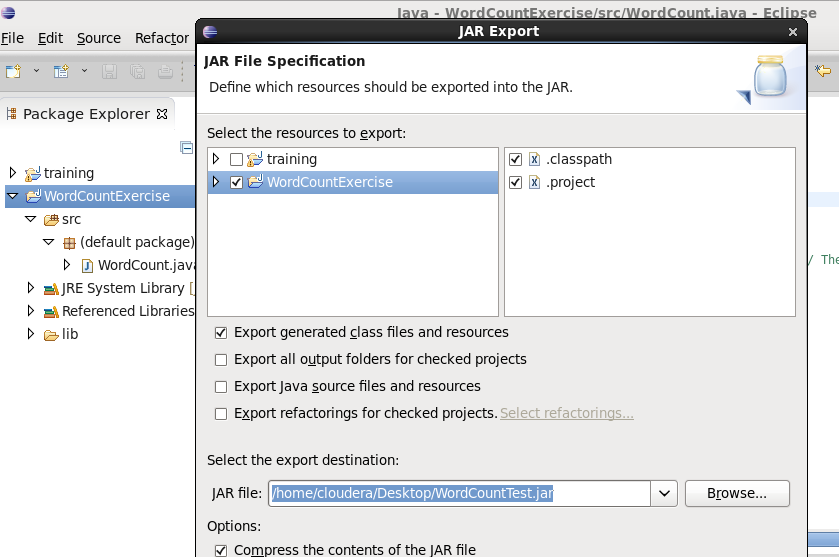
We can import to the eclipse using File 🡪 import 🡪Existing projects into workspace 🡪 “ Choose Word Count Exercise Project”.

Then Export the project as jar file. For exporting Right click on the proect 🡪 Export 🡪 Jar file

The below screen shows the code which is imported to the eclipse:

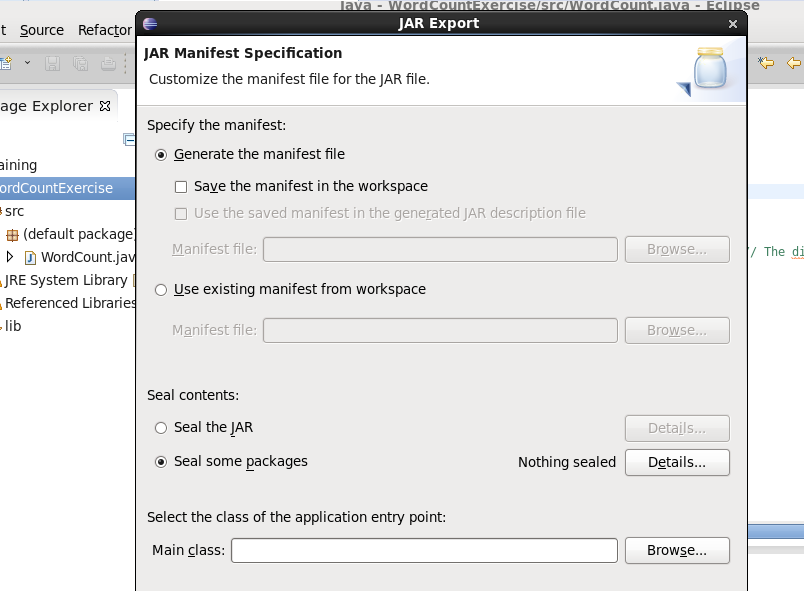


That above code is the code of word count example and which is imported to the java eclipse.

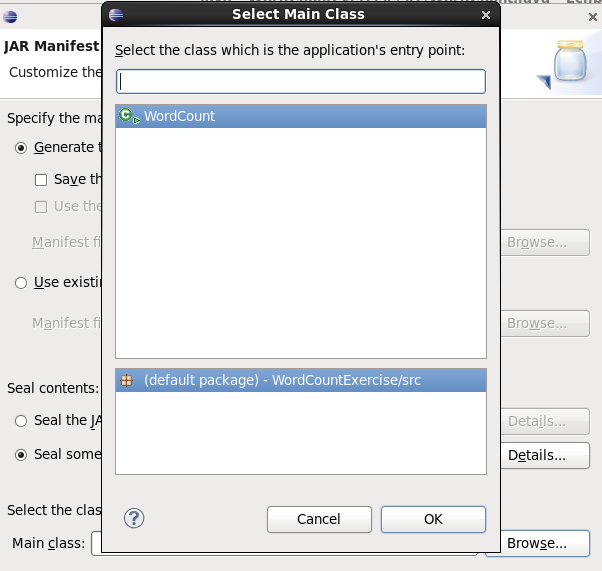
****

The above screen shows the exporting the project as a jar file.

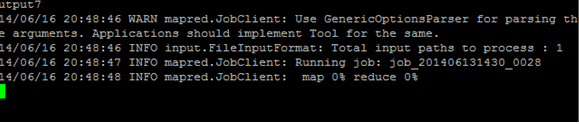
On clicking next you will be getting the next screen shown below



For selecting the main class you need to browse where you can select the main class which is shown in below serene.

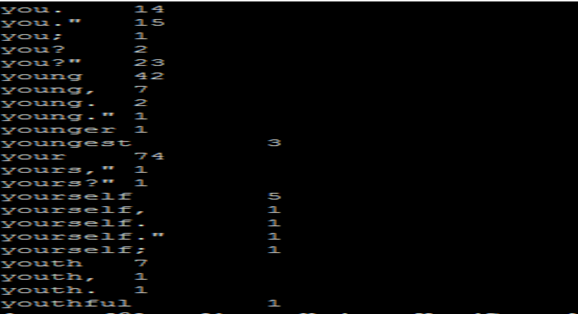
****

Now run the command **hadoop jar WordCountExercise.jar** after executing that command the result can be checked using the command **hadoop fs –cat output/\***

****

The result of the word count is given as counting the words which is shown below.

The below screen shot shows the output of the word count jar program



**Subtask – 6:**

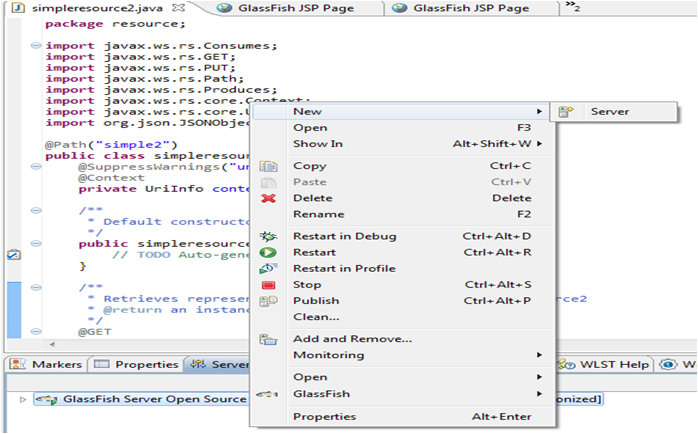
**Java Based Restful Service:**

For building java based restful service we need to install the Glass fish server

For develop web application we need to open the Java EE perspective.

Then create a new glass fish server. To create new glass fish server need to right click on server tab 🡪 new 🡪select the server

The below screen shows the creation of new glassfish server

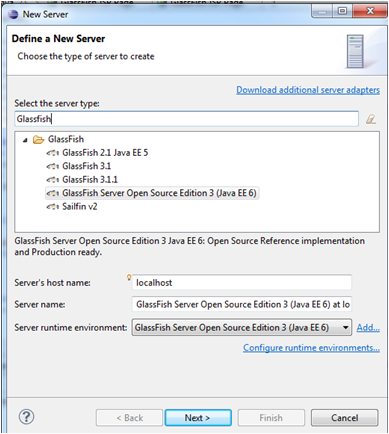
****

Then choose the glass fish server. Type the glass fish server and choose the one from all the servers available.

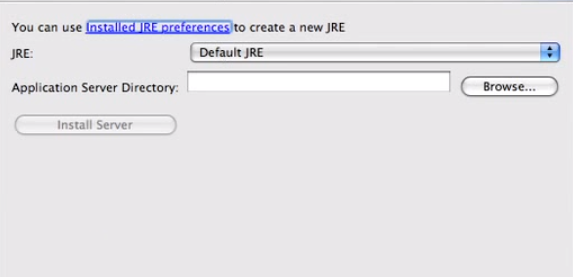
Select the Glass fish server open source Editiion 3 and then install.

After that you need to select the GRE and then unstall.

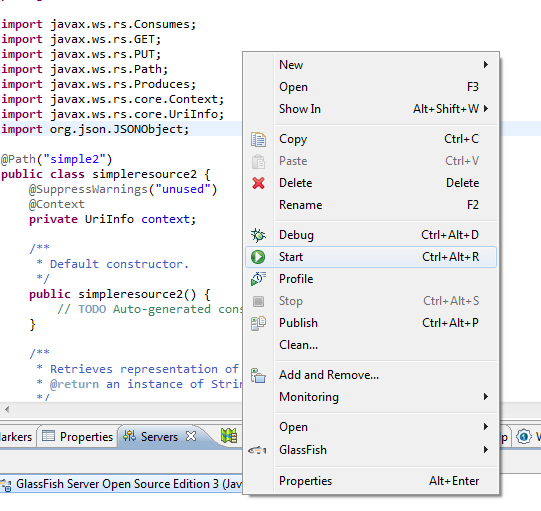
The below screen shows all the glass fish server available



The below screen shows the selection of GRE and its installation

****

After installing the the server then start the server to run.

****

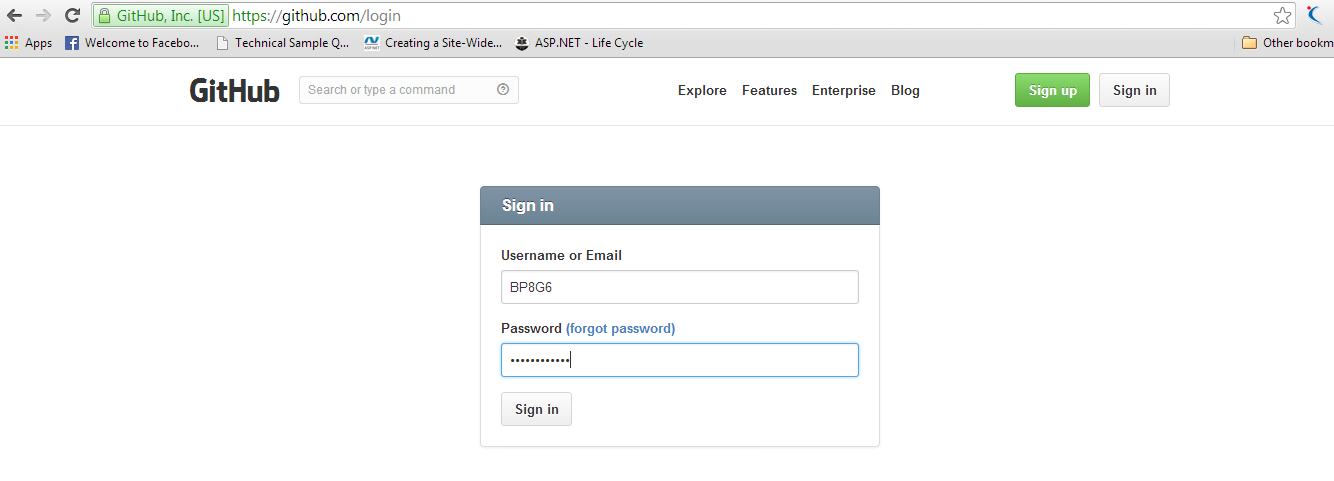
**Task – 3:**

**Subtask – 1:**

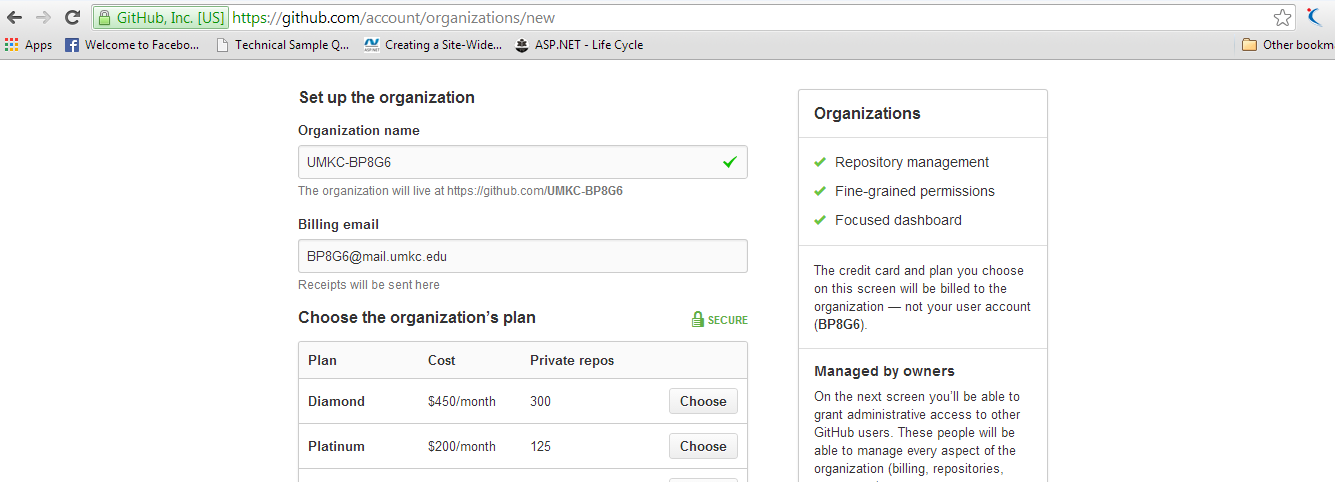
**Create a Github Account and deploy files to Github:**

Go to www.github.com and Sign up.

The below screen shot shows the Login page of the gitbub account.

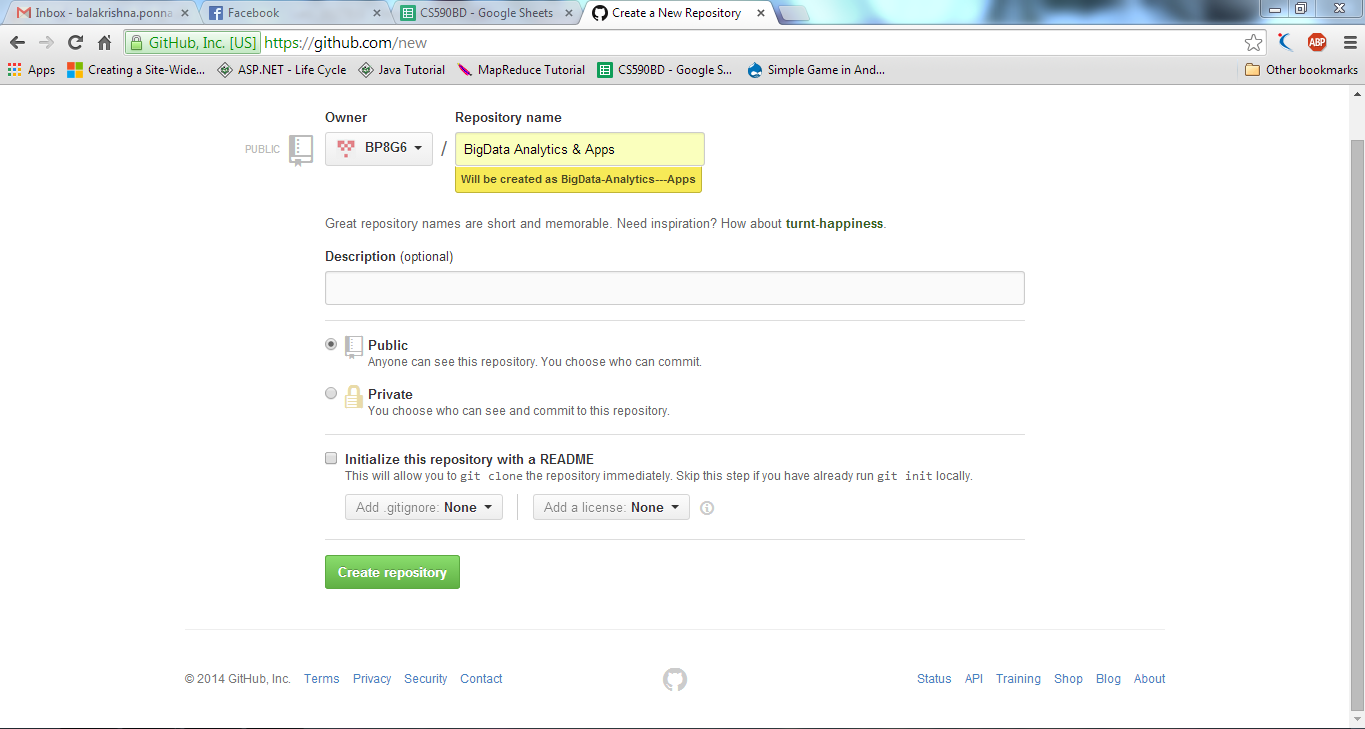


Set the Organisation name and the Organisation Plan for the Github account.

****

Then Create the repositories and upload the files which you need to submit to the github.

The below screen shot shows the creation of new repository:

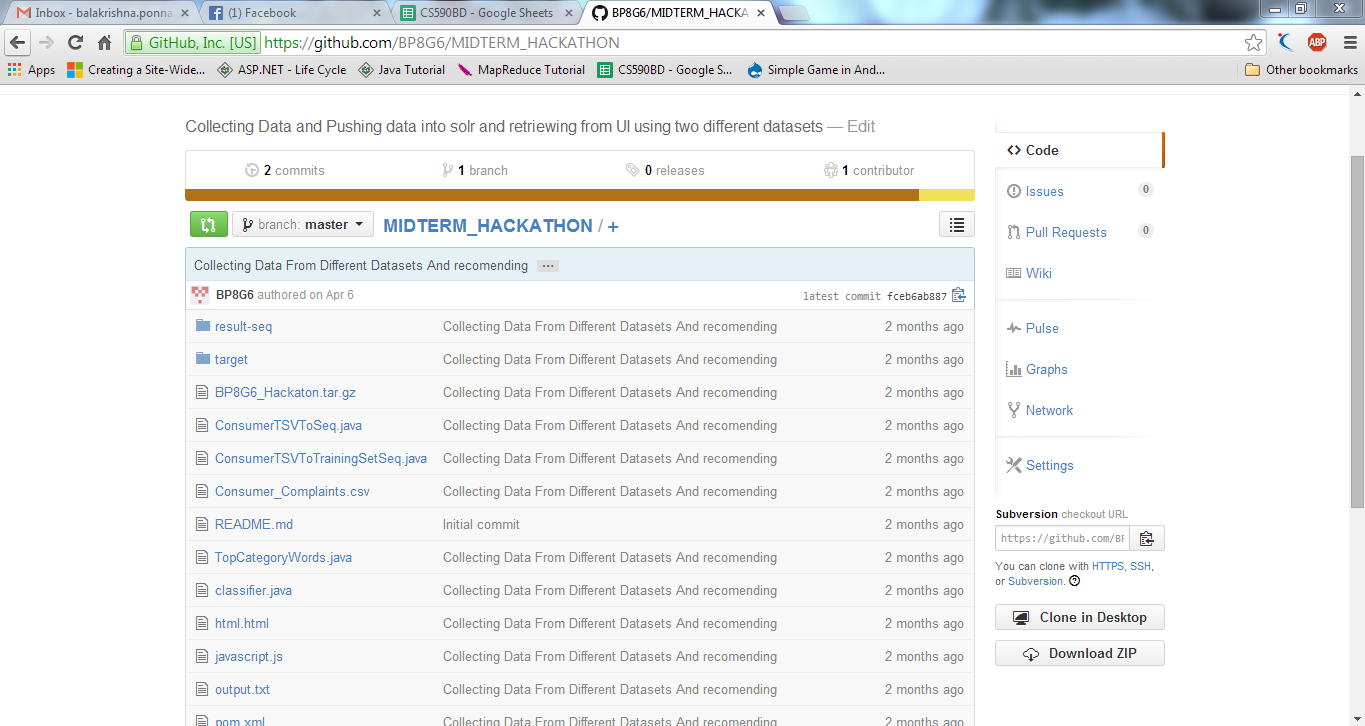


After creating a repository open the repository and there will be creating a github folder in your system.

Upload all the files which you need to submit to the github and you find the option clone to desktop after opening the new repository you have created. Then click the Sync button and submit.

On submiting, all the documents uploaded in the github folder will be in the repository where you can acessing using the url.

The below screen shows the files after uploading in the git hub



Then you need to copy the URL of that page by pasting that in URL in address you can see the uploaded documents.

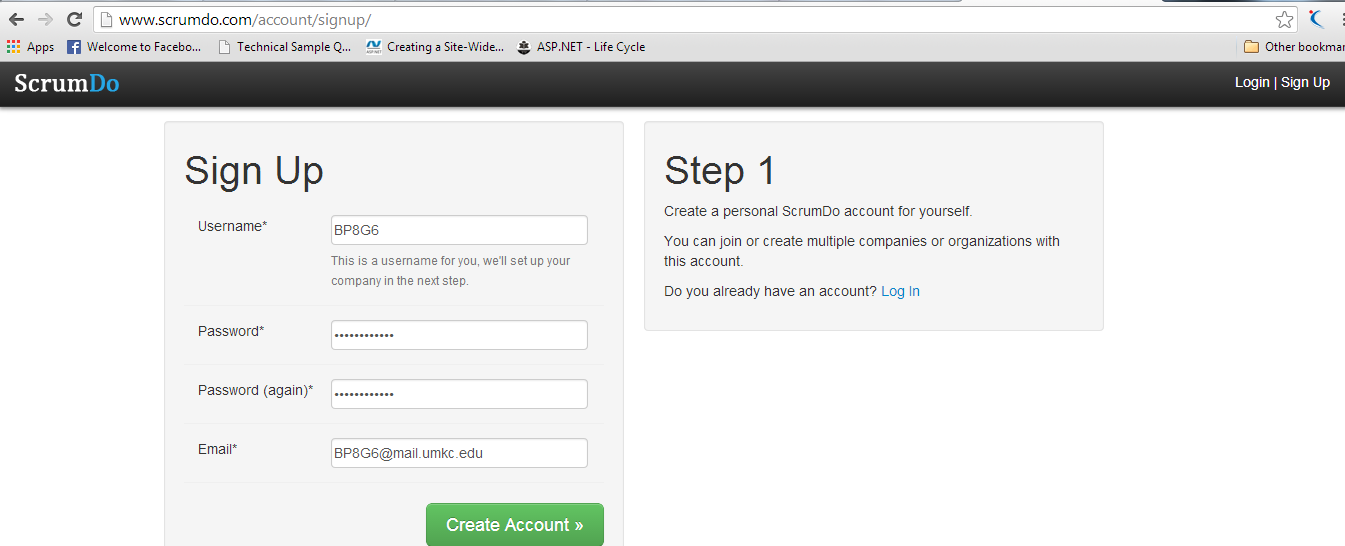
The repository can be set to public or private where b y setting it to the public everyone can get access.

**Subtask – 2:**

**Create an account and design projects with scrumdo:**

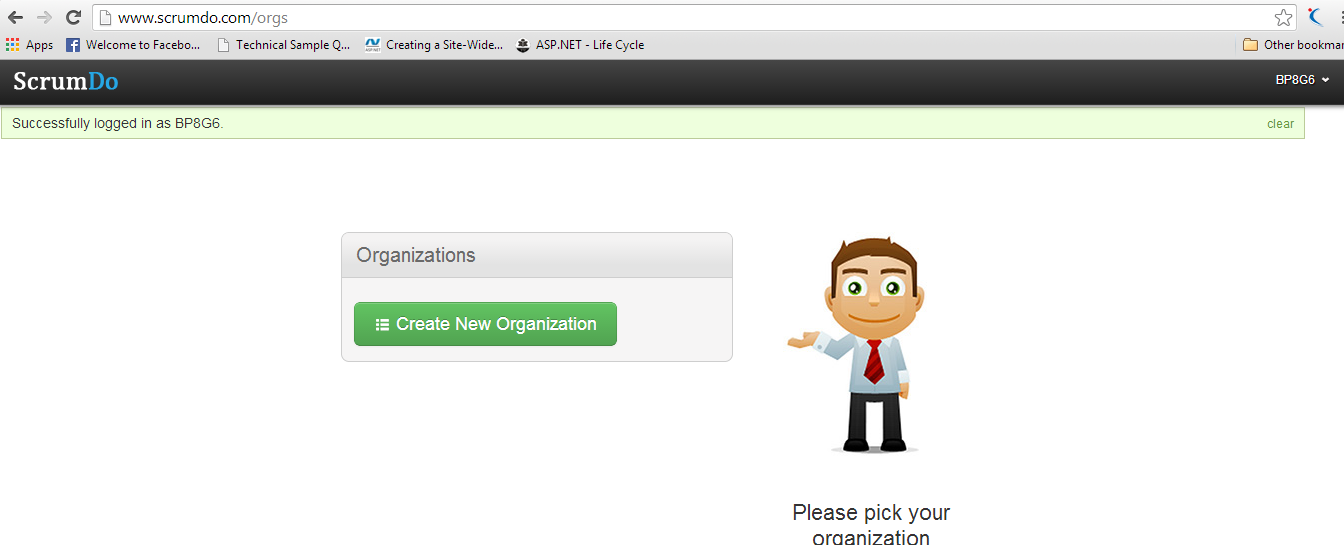
Go to www.Scrumdo.com and Signup for the account and get login.

The below is the Screenshot for signup:

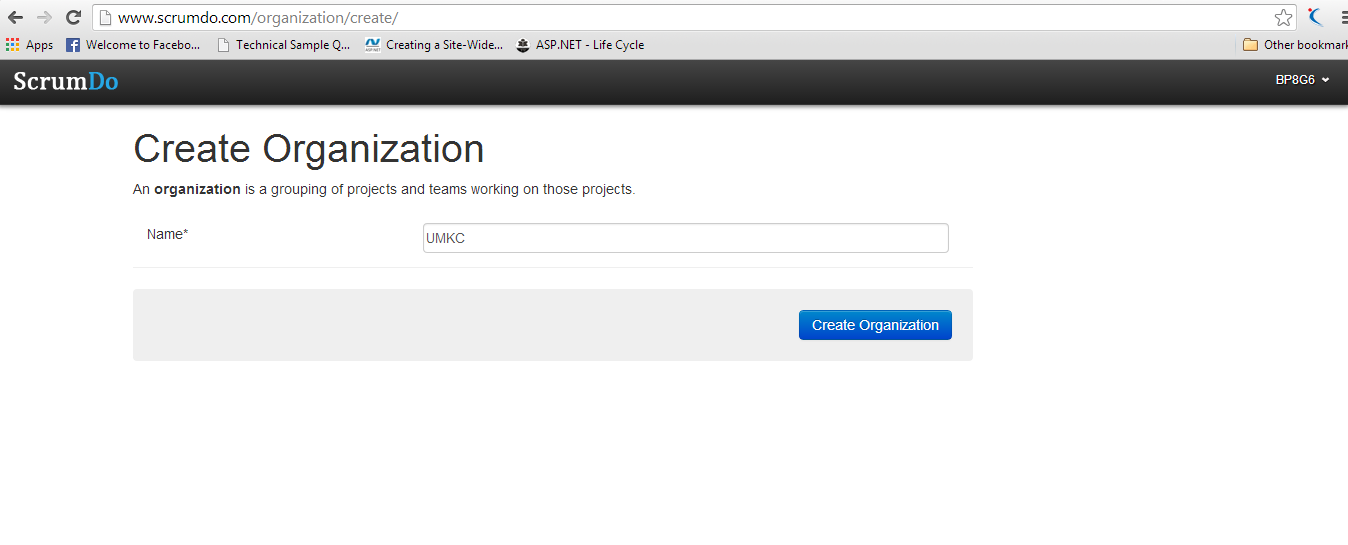


Then you need to create an organization for scrumdo

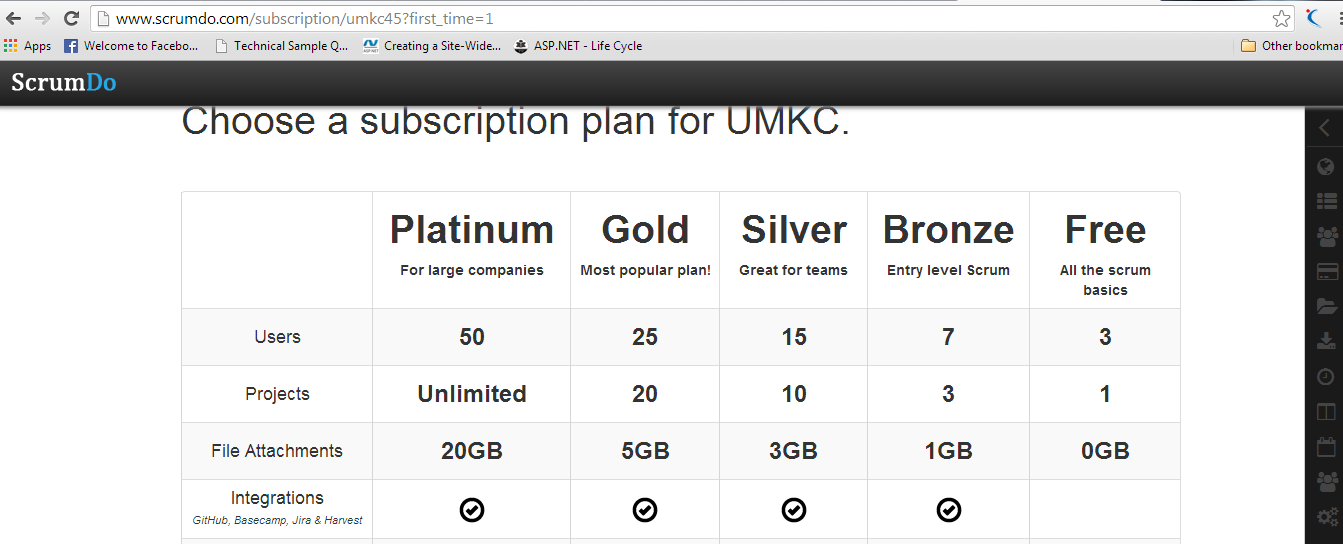
The below is the screen shot fro creating organisation in scrumdo



We need to name the Organization.



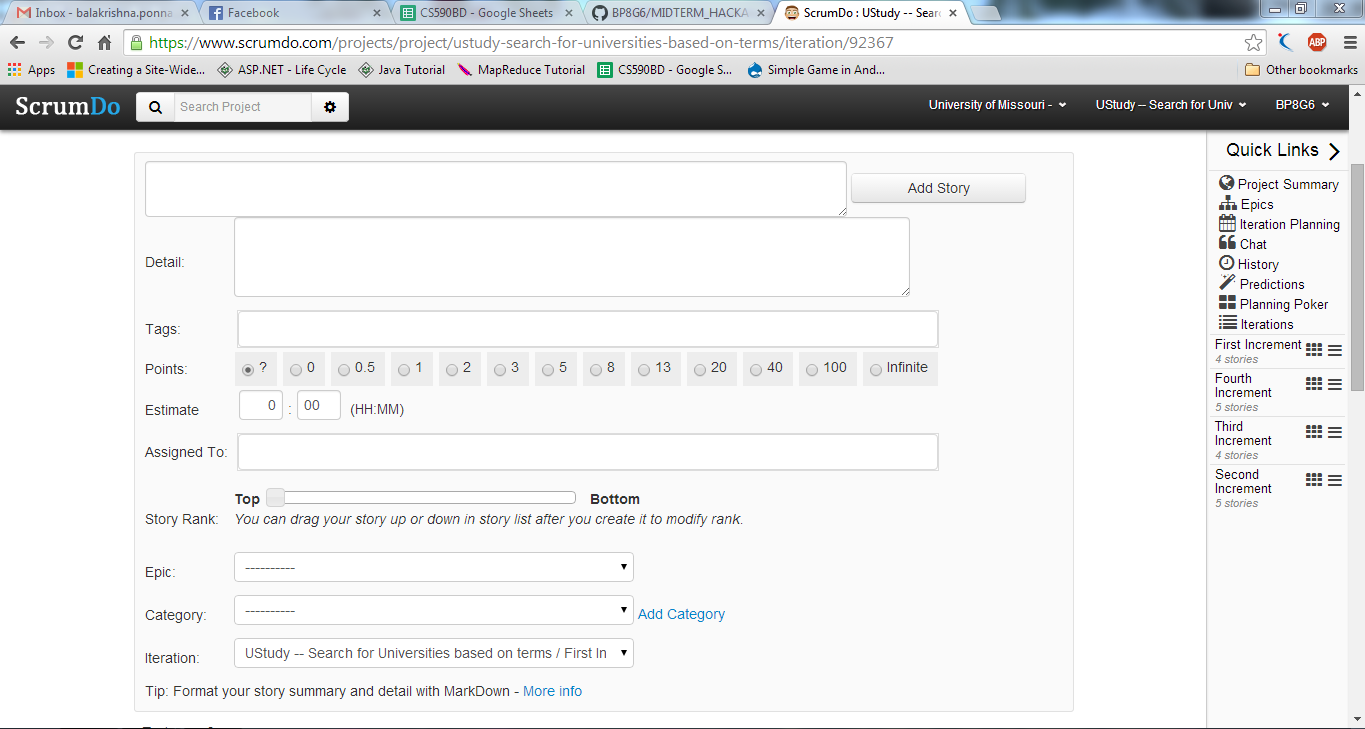
Choose the Subscription Plan for the Organization



We can add some stories what need to perform and in which iteration it is going to be done and tag the person who will do it.

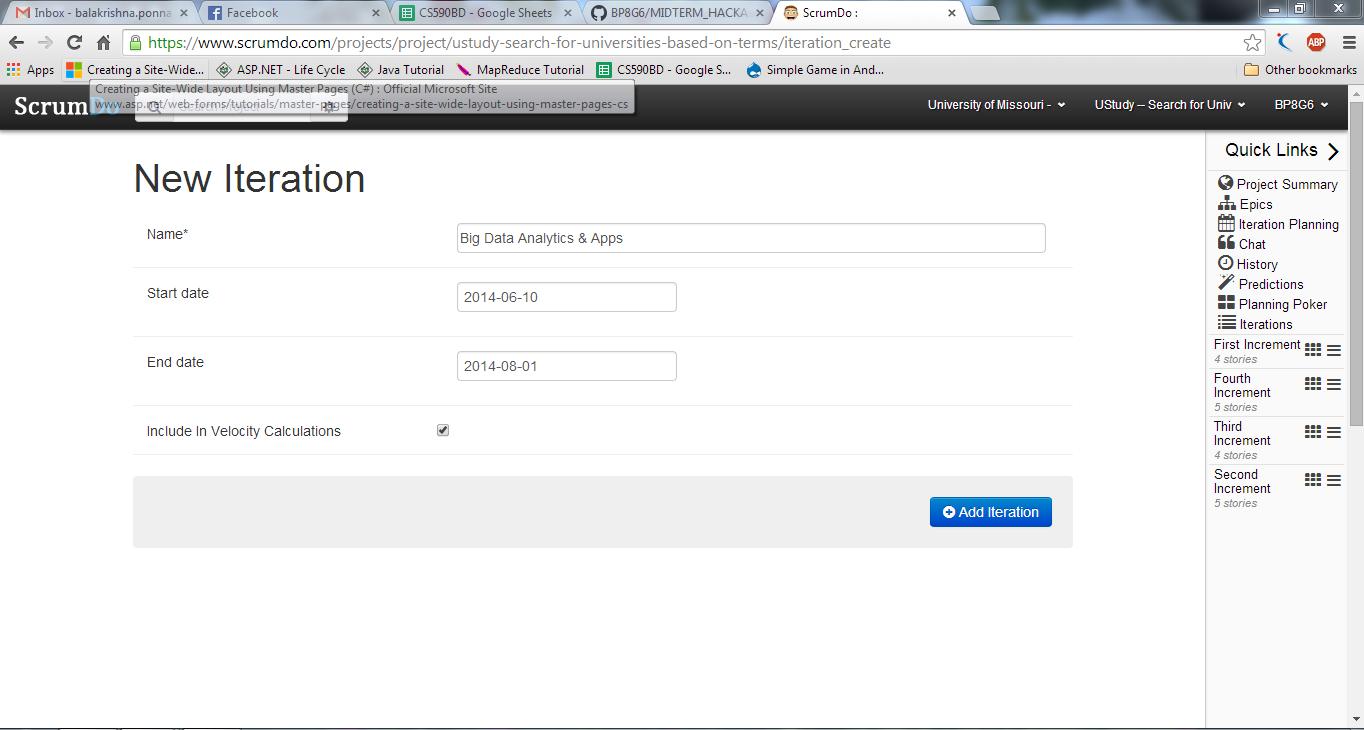
Add a Story, Details of the story and tag the users who are working and assign number of hours required, Difficulty rate ,Category to which it belongs if category is not added add a new category and select the Iteration.

The below is the screen shot for adding a story:



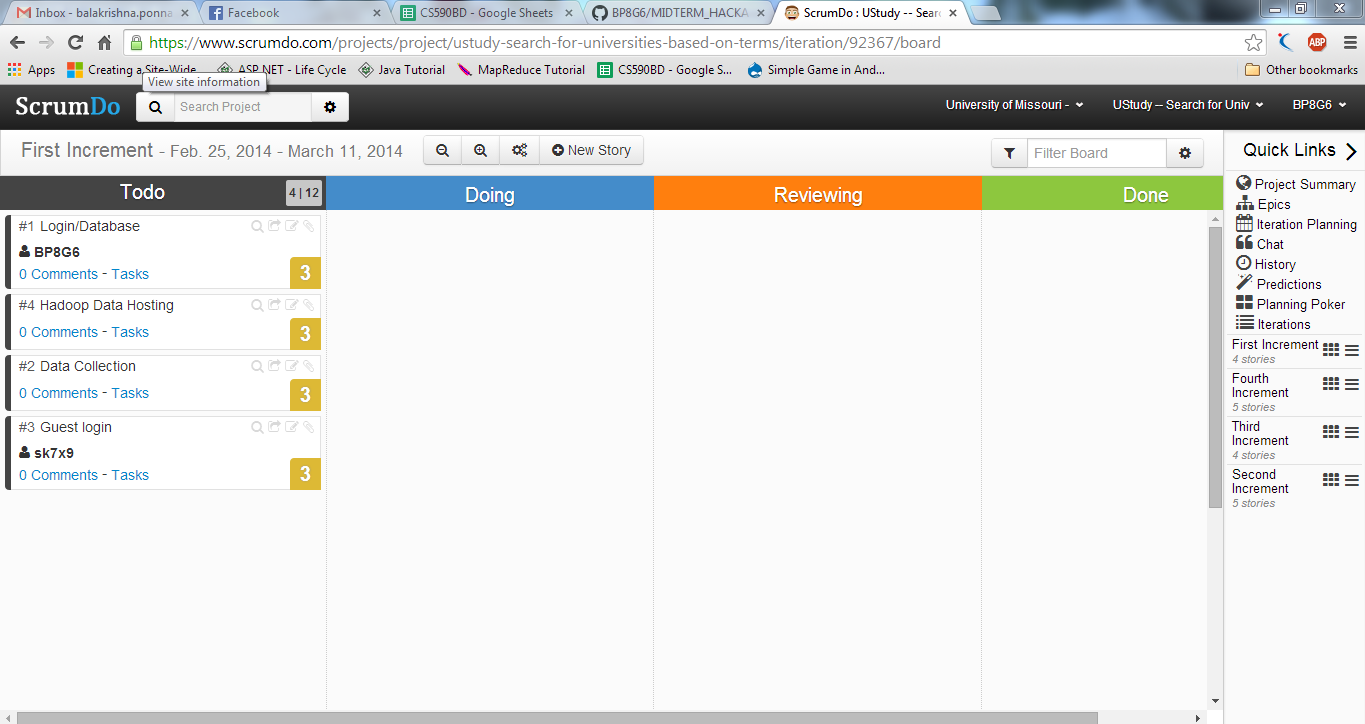
Creating the New Iteration

Create the new Iteration and assign the Start Date and the End Date of the Particular Iteration

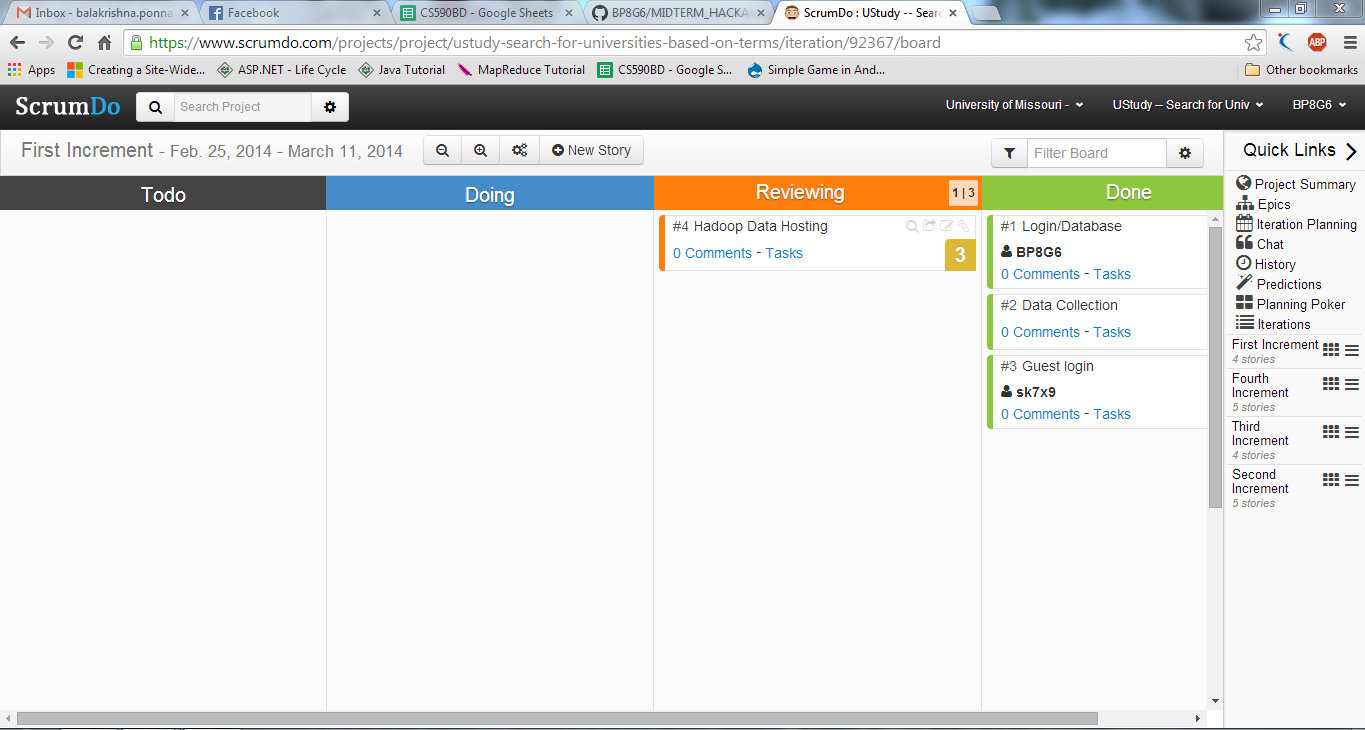


To Do in Backlog

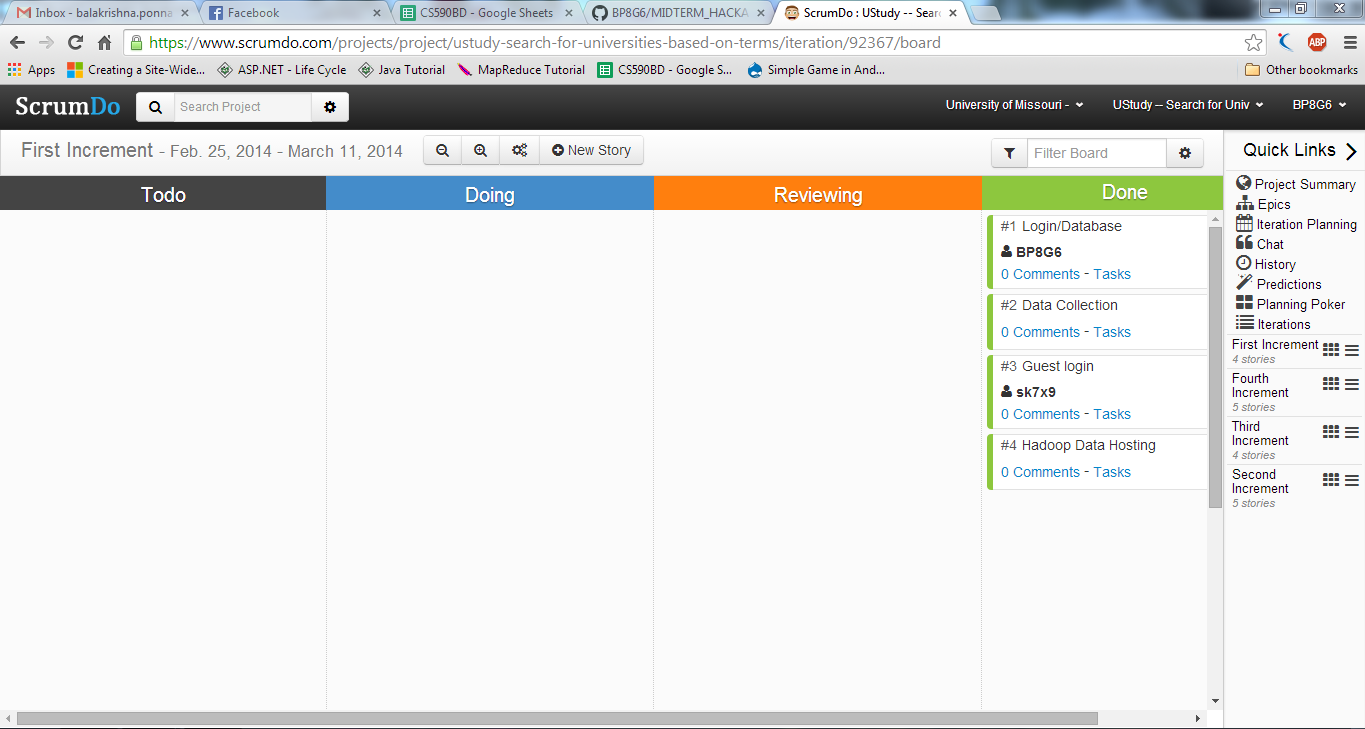
Job To do in the to do list. This is the Screenshot of the Todo .Todo Indicates the Pending jobs to be done



The Below Screen describes one story has be completed(done) and one story is Reviewing



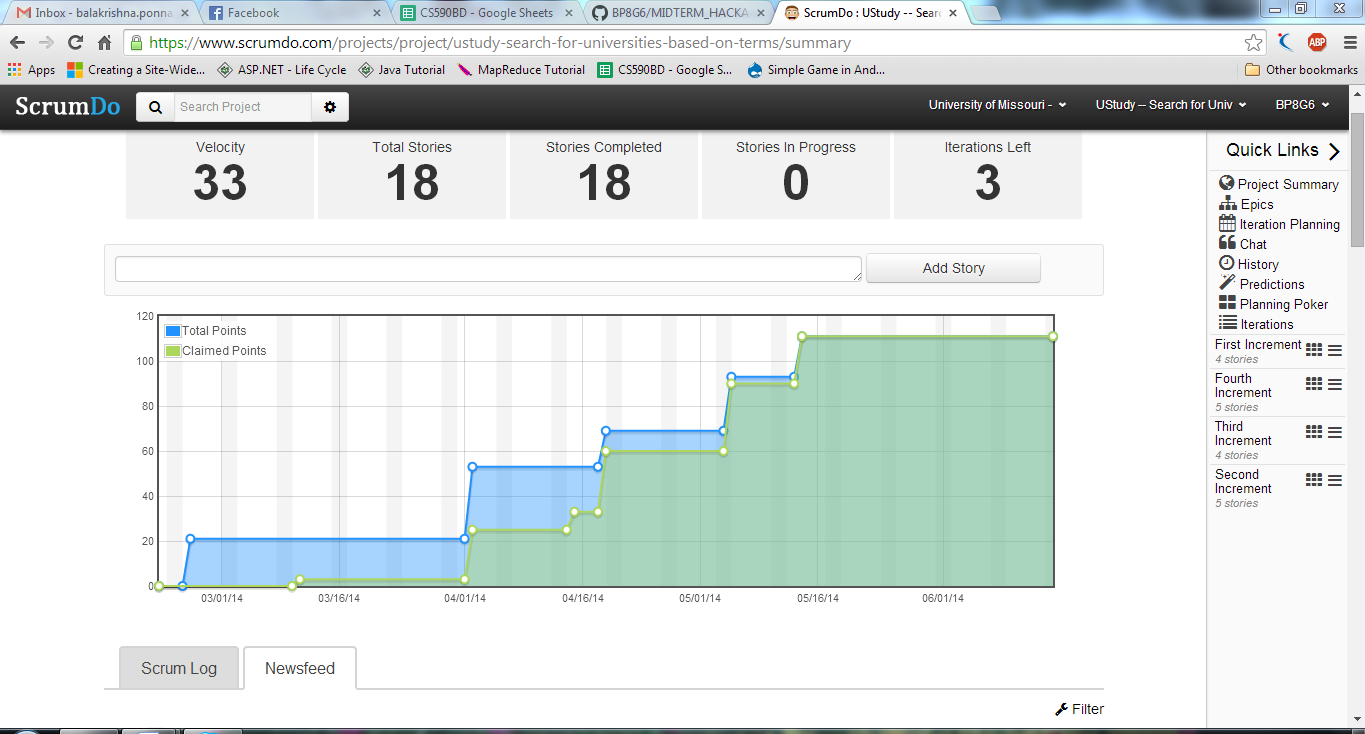
The below Screen Shows that all the Stories are done



The below Graph describes all the jobs are done and the screen indicates the total points, Points in progress ,points completed and number of days left.

It shows two graphs indicating total points and the clamed points and based on that it draws the graph.

The below screen shows the graph of the total work:



There will be also some log and feeds which update whatever did to the account.

The below screen shows the Scrum log and feeds:

