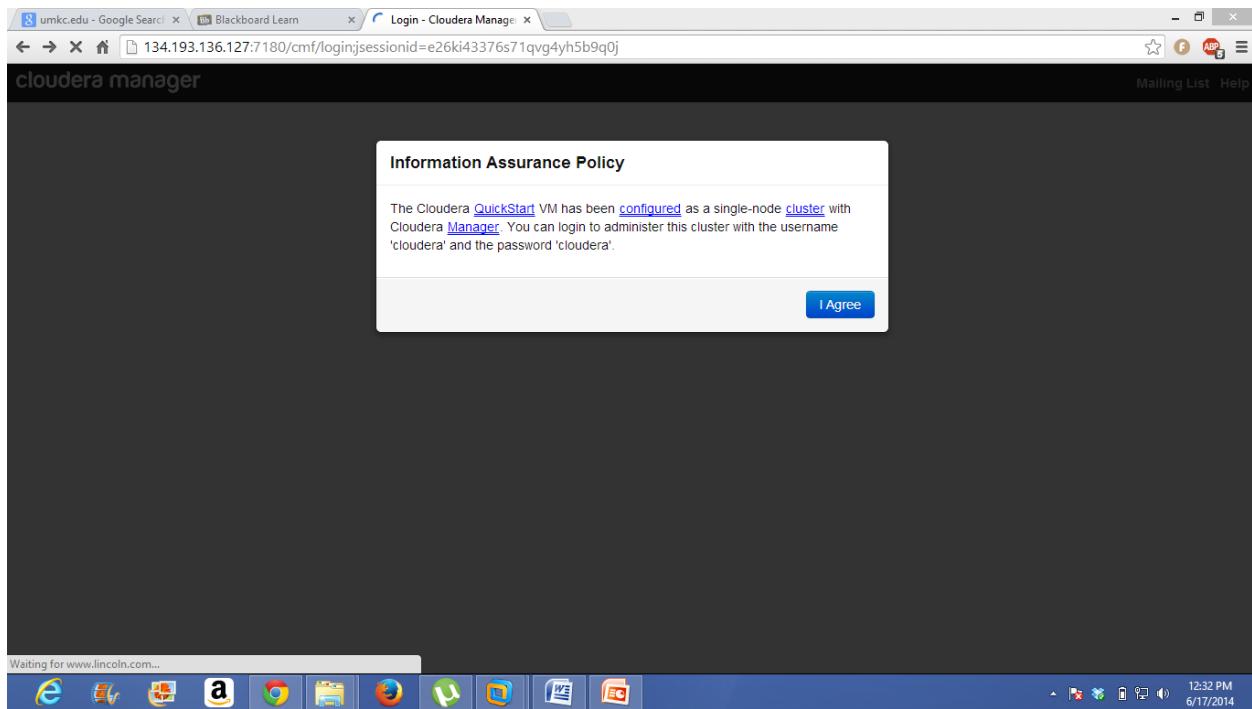


# **Big Data Analytics and Applications**

## **Lab-1**

**N.AYYAPPA KUMAR(16157522)**

In order to login into Cloudera Manager use one of the web link provided by professor here I used the link <http://134.193.136.127:7180>



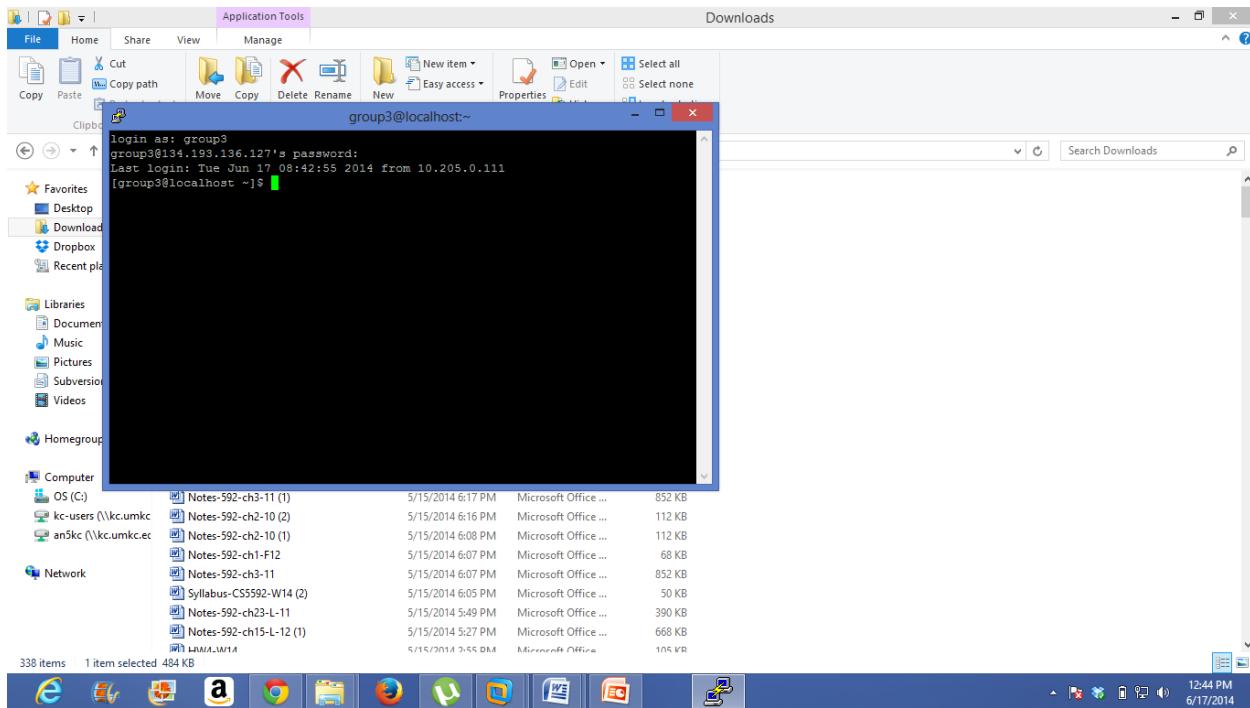
After that click on I agree button then login with your SSO ID as both user name and password then below screen will appear.

The screenshot shows the Cloudera Manager Home page. In the Status section, under Cluster 1 - CDH4, there are 13 hosts listed: Hosts, flume1, hbase1, hdfs1, hive1, hue1, impala1, ks\_indexer1, mapreduce1, oozie1, solr1, and sqoop1. In the Charts section, there are three graphs: Cluster CPU, Cluster Disk IO, and Cluster Network IO, all showing a 'QUERY ERROR' event between 10:15 and 10:30. The top right corner shows the date and time: June 17 2014, 10:34:05 PDT.

Next down load putty and provide the web link you used at specific location.

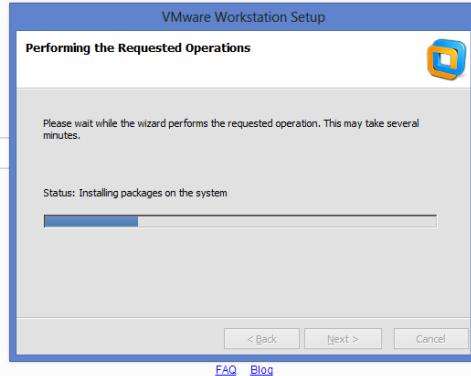
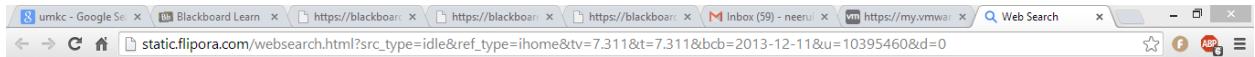
The screenshot shows the Cloudera Manager Home page with a Microsoft PowerPoint presentation titled 'CS590BD-Tutorial 2-Cloudera.pptx' displayed in the background. A Putty Configuration dialog box is overlaid on the screen. The dialog is set to connect to host 134.193.136.127:7180 on port 22 using SSH. The connection type is selected as SSH. The background shows the same status and charts as the previous screenshot.

Then login as group3 and use the password also as group3. here you can change your password.

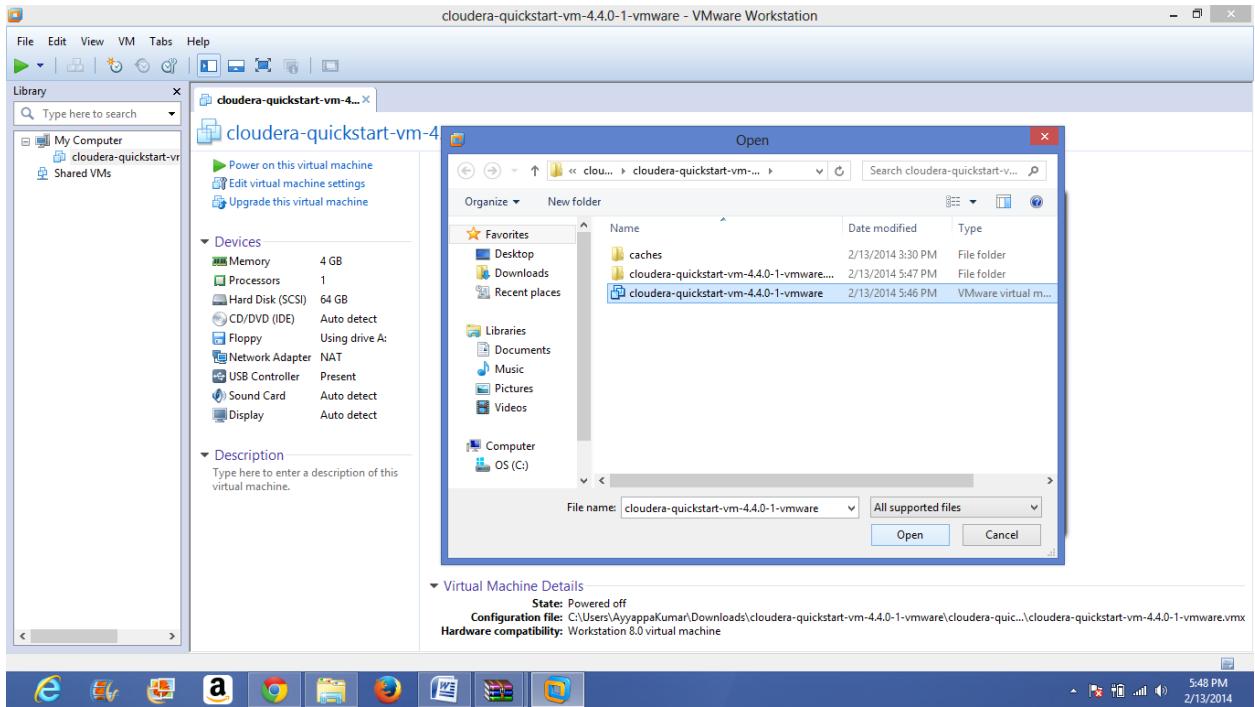


### Creating own local Cloudera and implementing word count application:

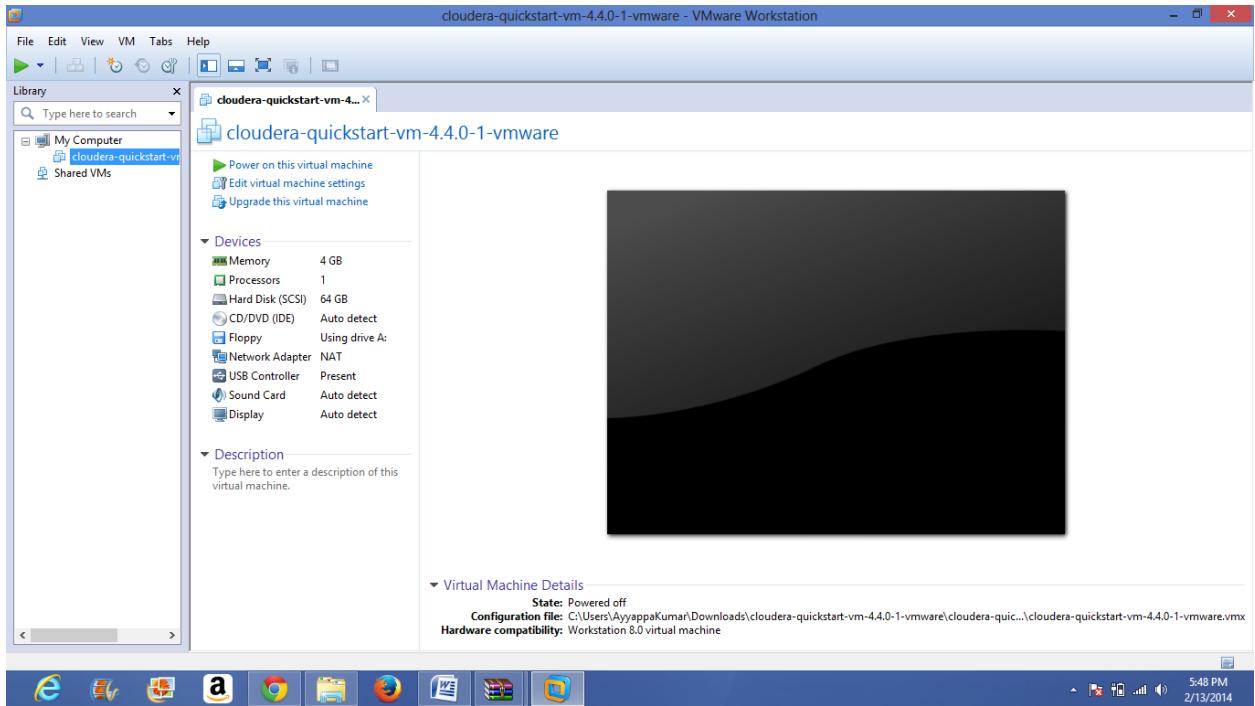
In order to complete the tasks assigned in lab-1 assignment first of all download the Cloudera image from the provided URL In addition to this Install VMware in your systems. below figure represents the installation process of VMware.



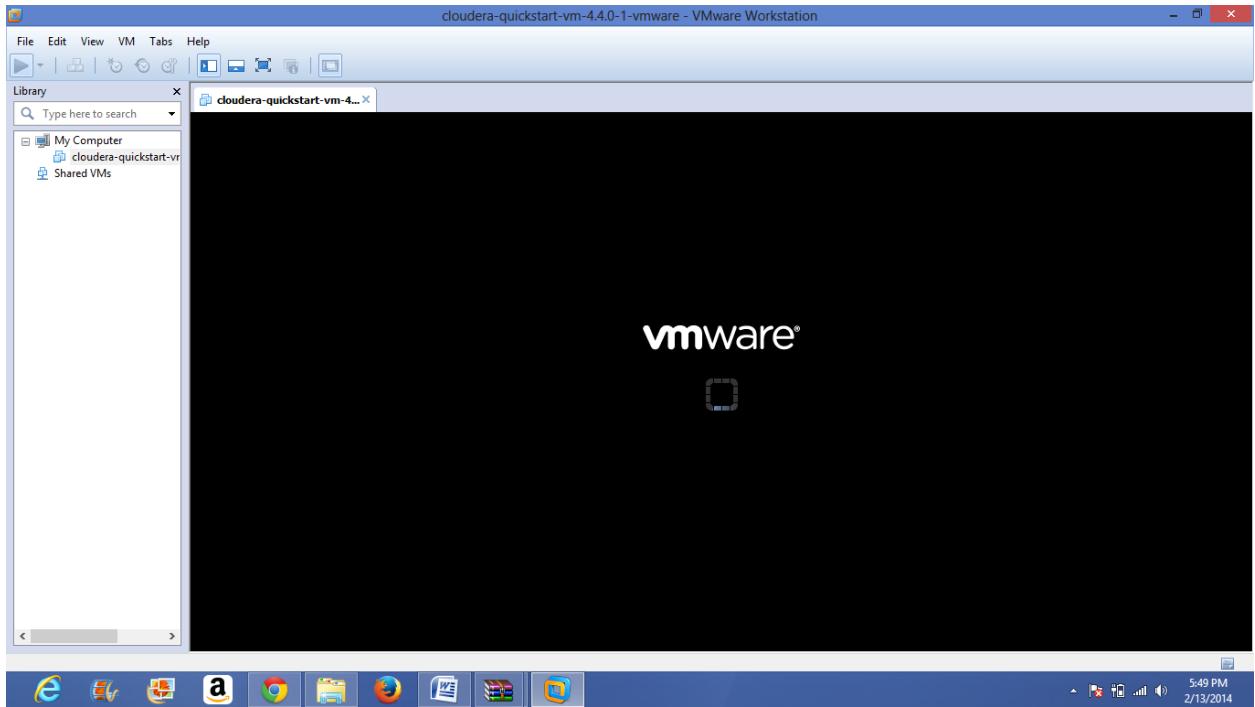
Then open the Cloudera image in VMware.



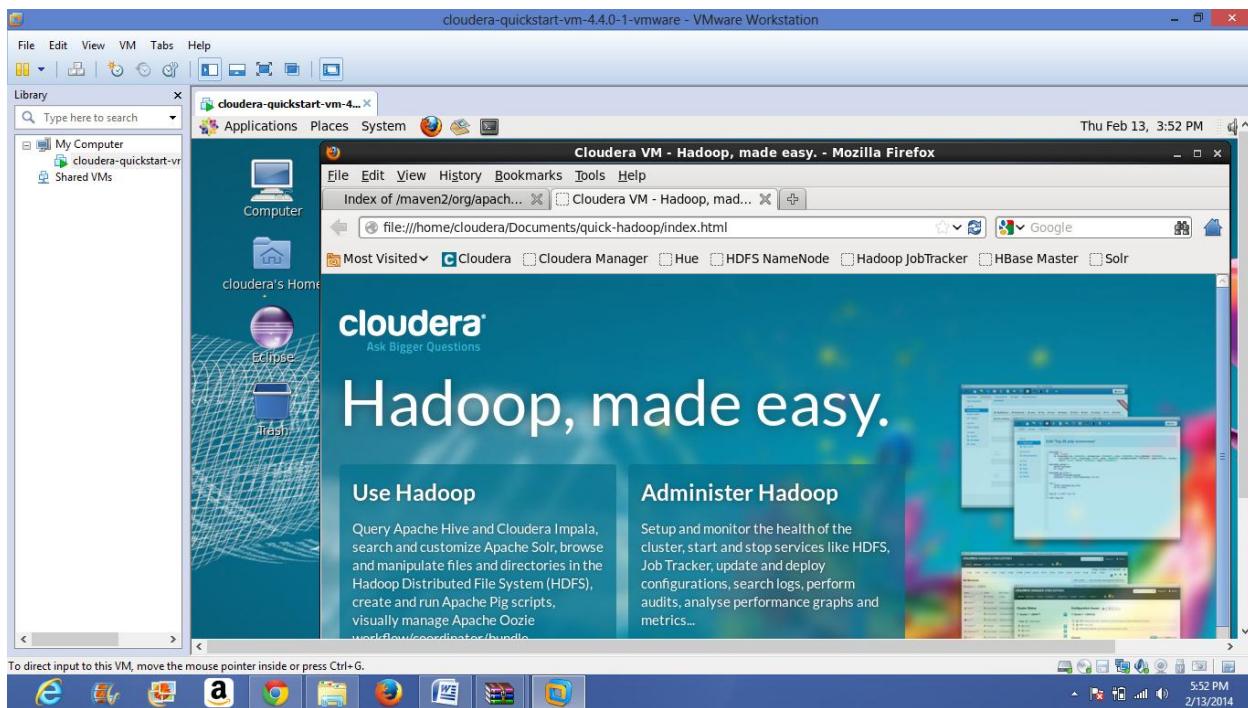
After opening the machine click on power on this virtual machine. Mean While you have to use the login details like username and password as Cloudera.



Then the image gets initialized as shown in below following two figures.



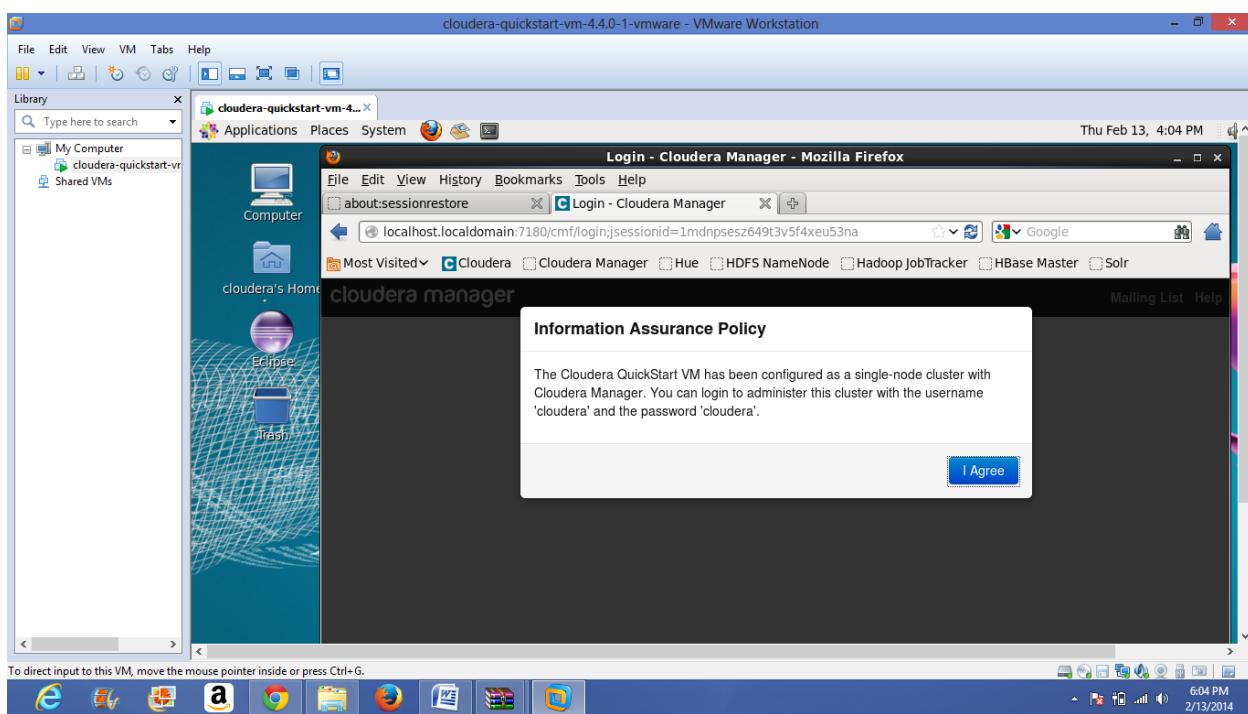
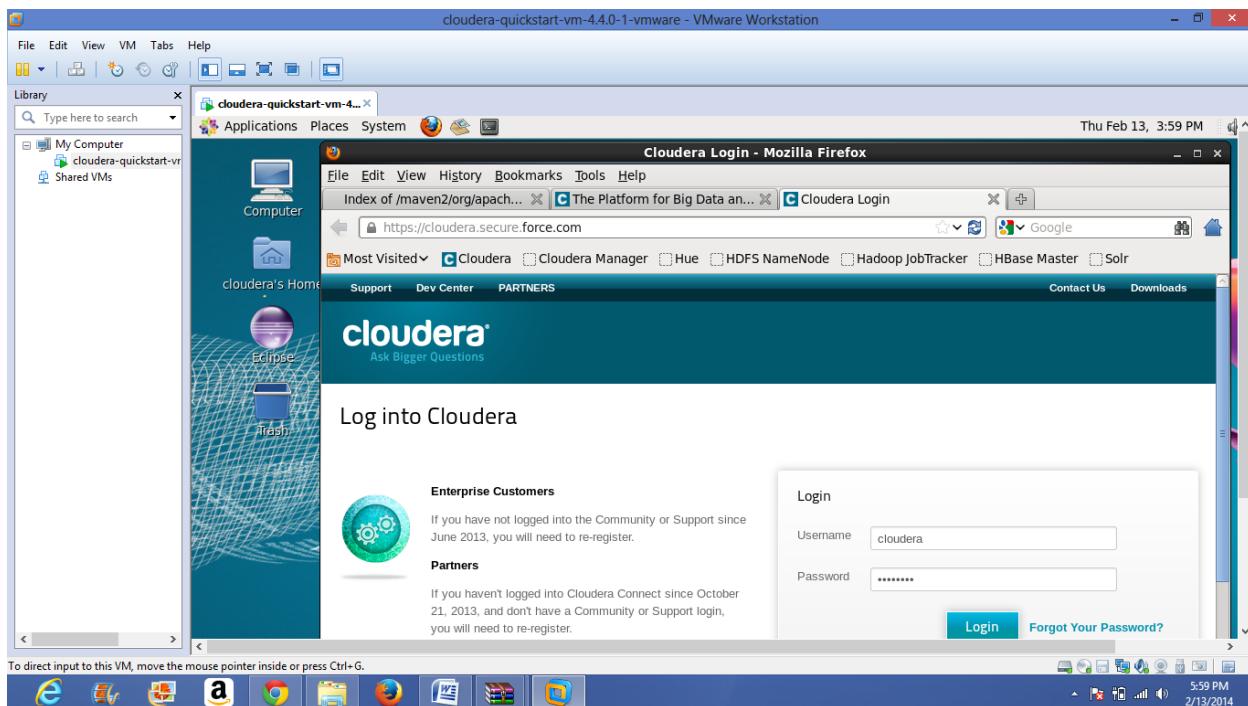
There are some features already initialized in Firefox , you can see them in below figure.



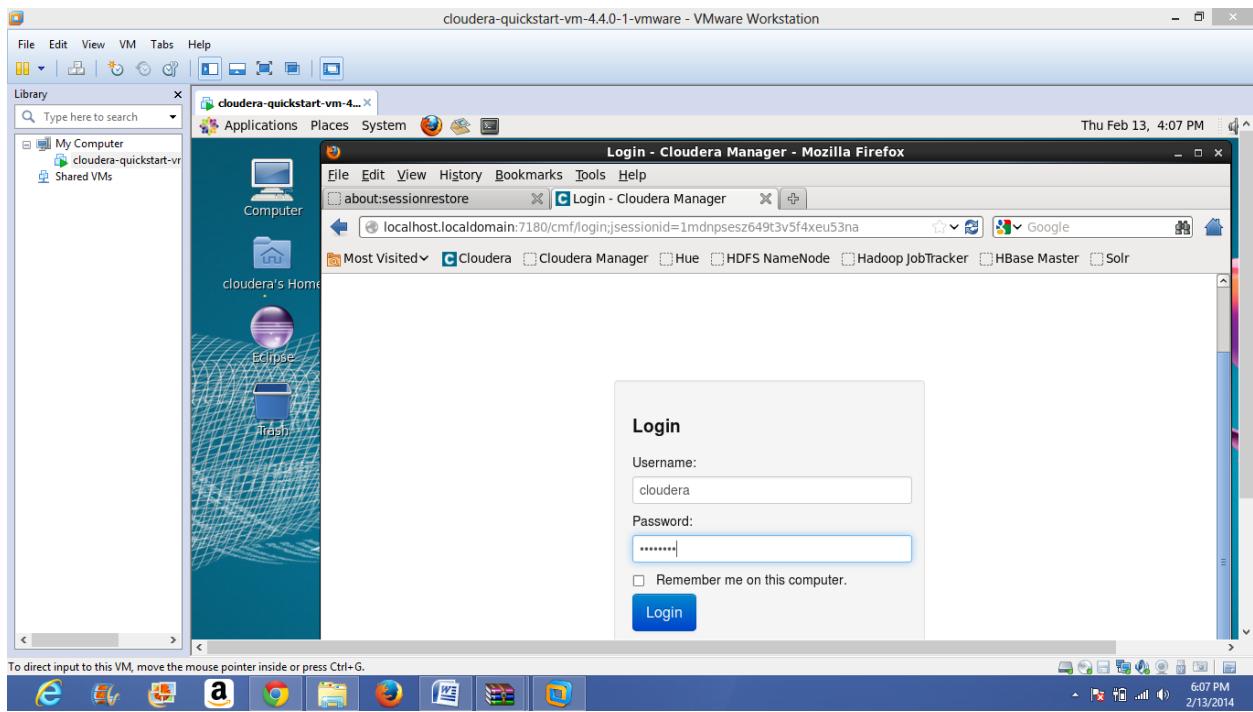
Then click on any option like Cloudera, Cloudera manager etc.



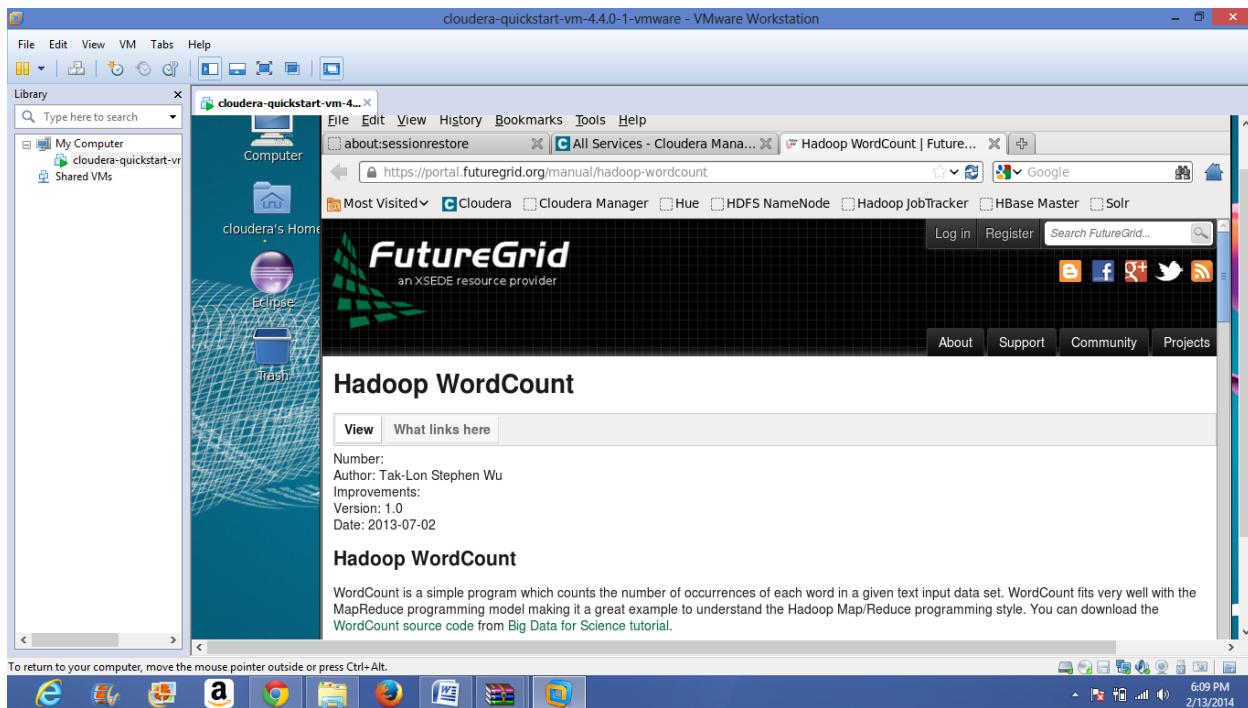
you have to login into Cloudera by using both username and password as Cloudera.



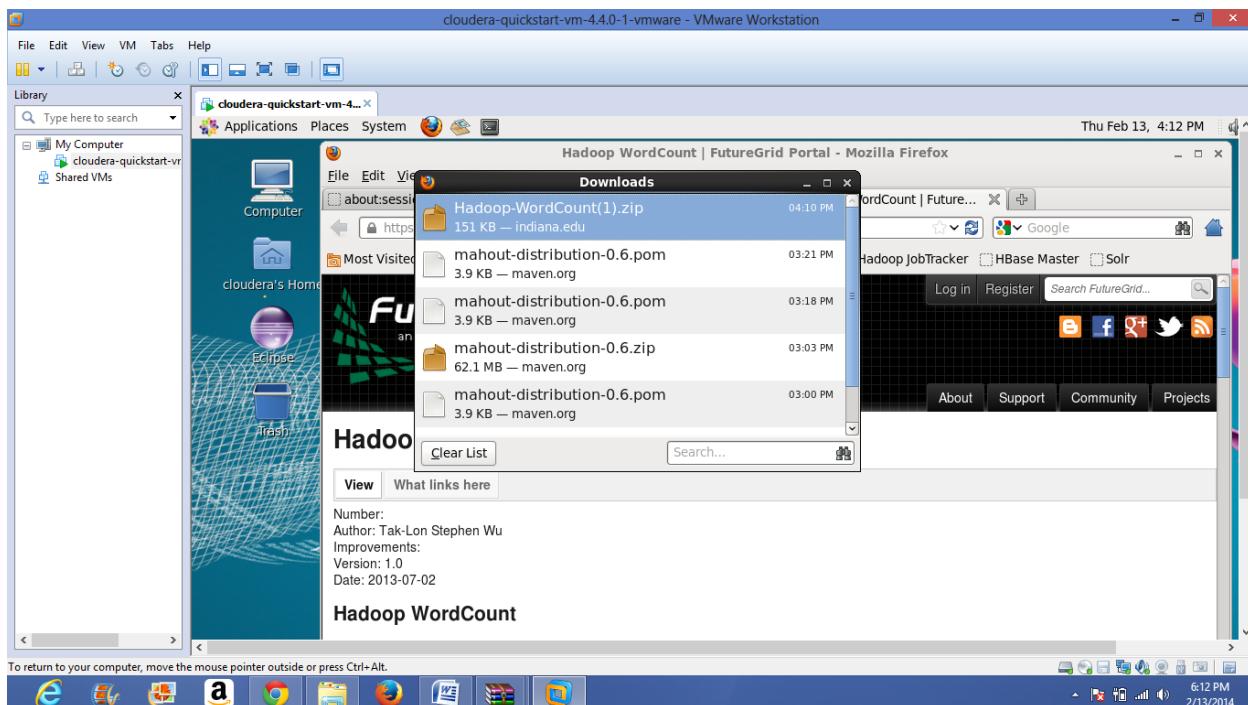
For Cloudera manager also use both username and password as Cloudera.

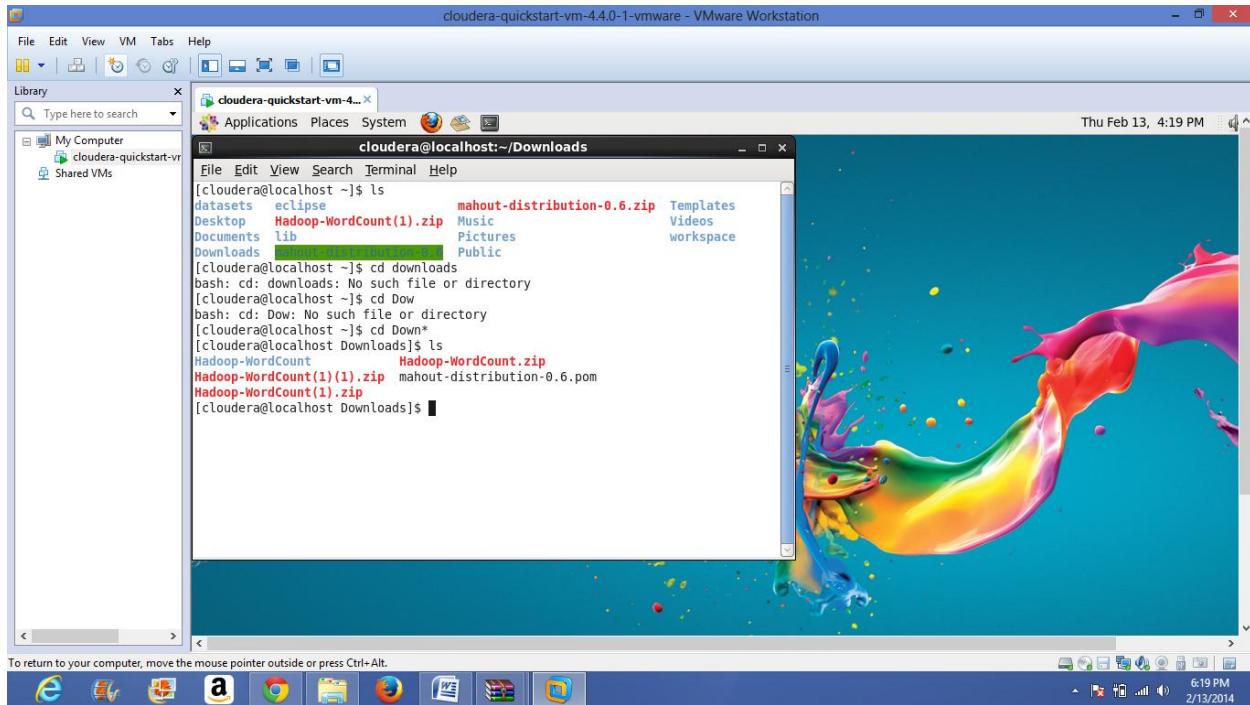
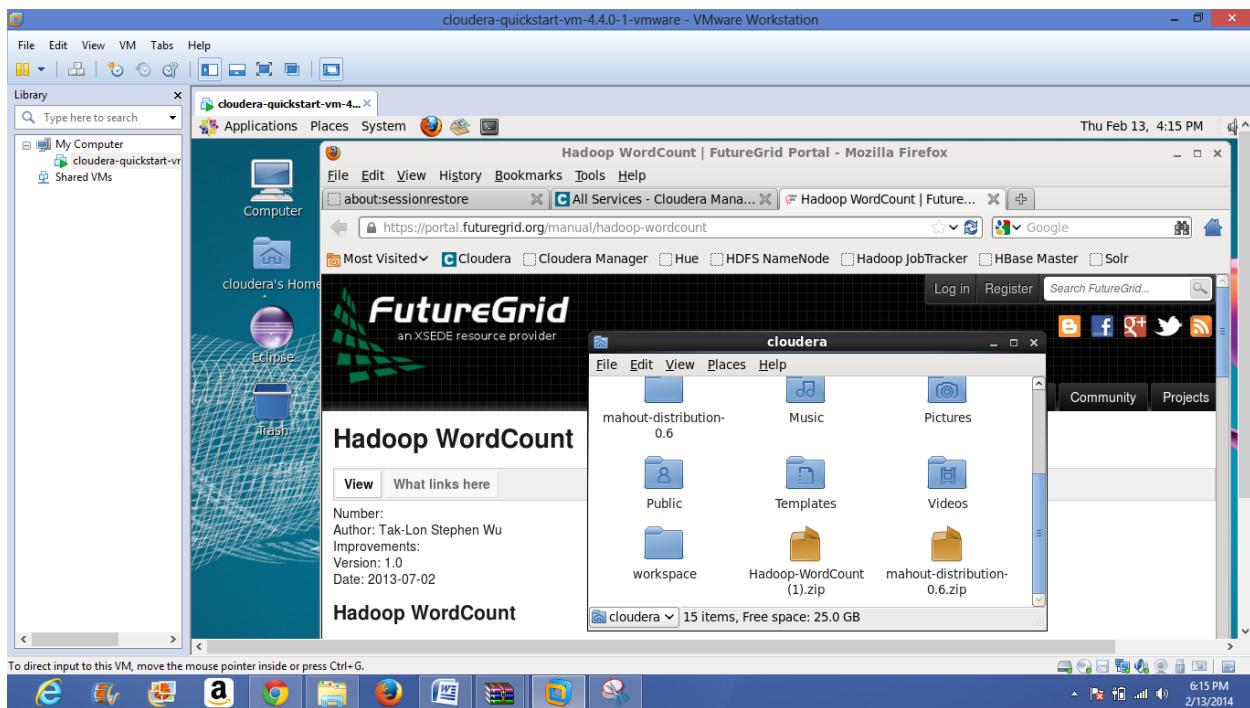


Then download word count example from the link <https://portal.futuregrid.org/manual/hadoop-wordcount> in Firefox.

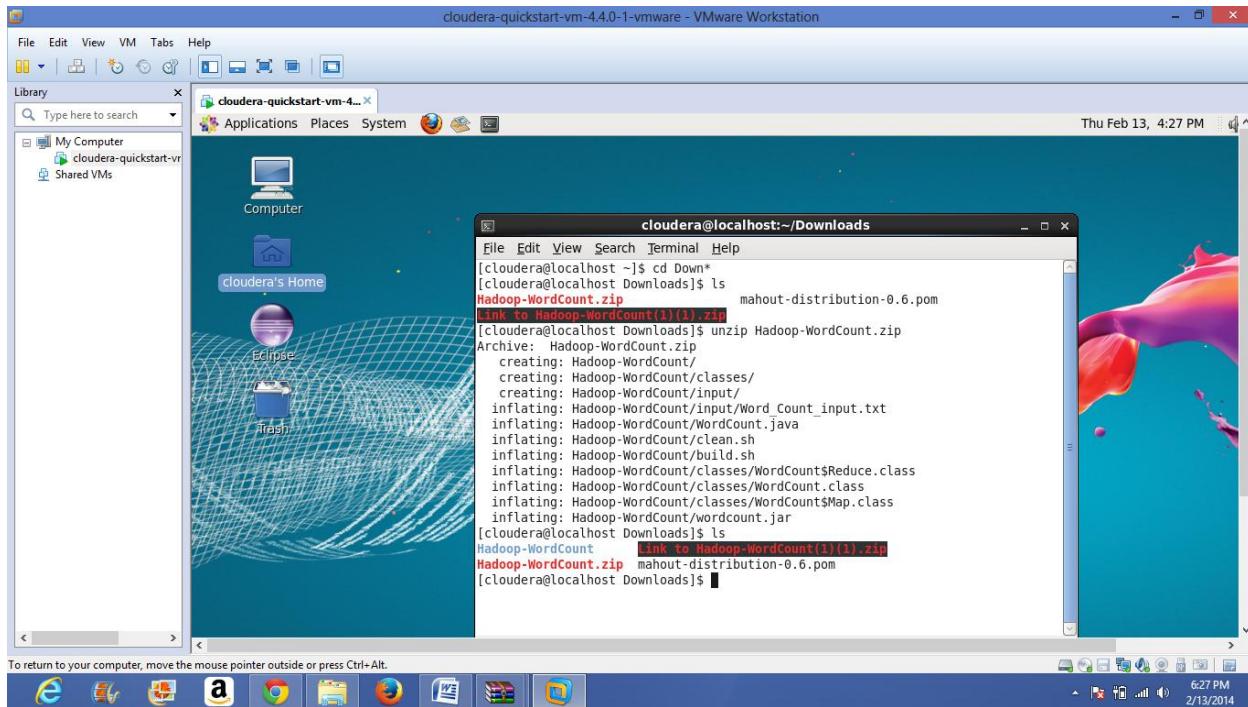


then the download completion of word count example is shown in below figure.



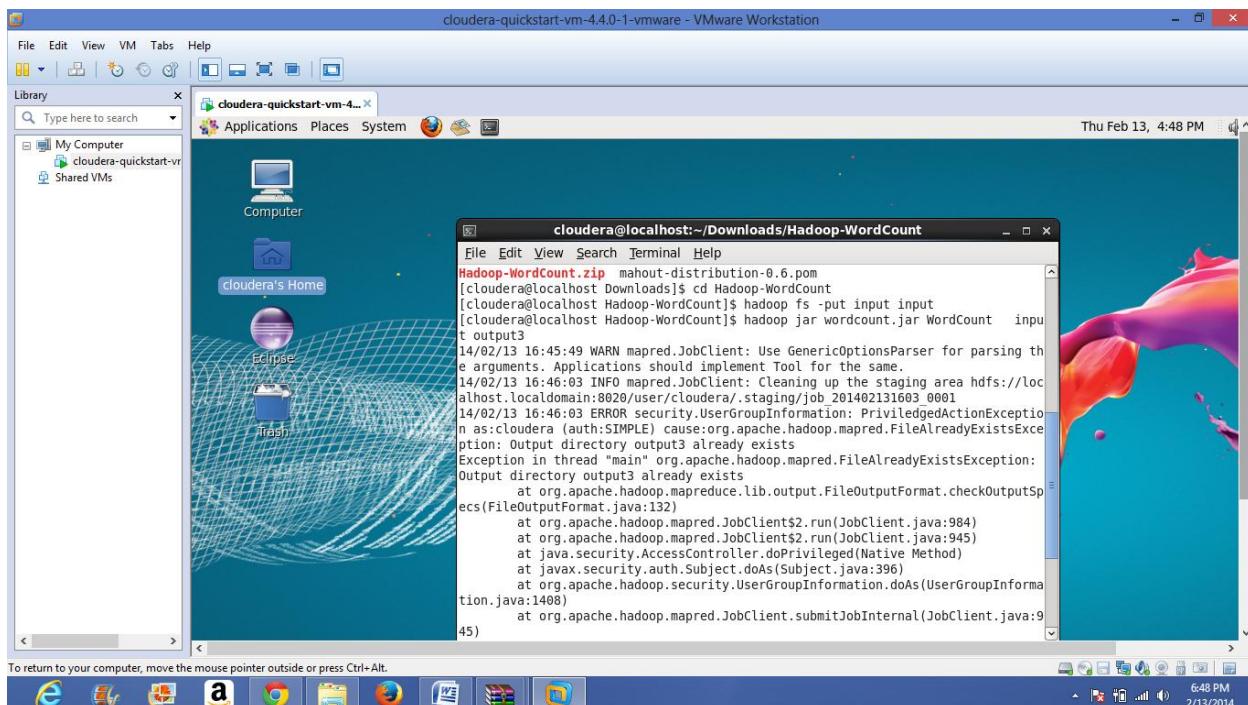


Then unzip the wordcount.zip and go to word count folder for this use command like cd wordcount.

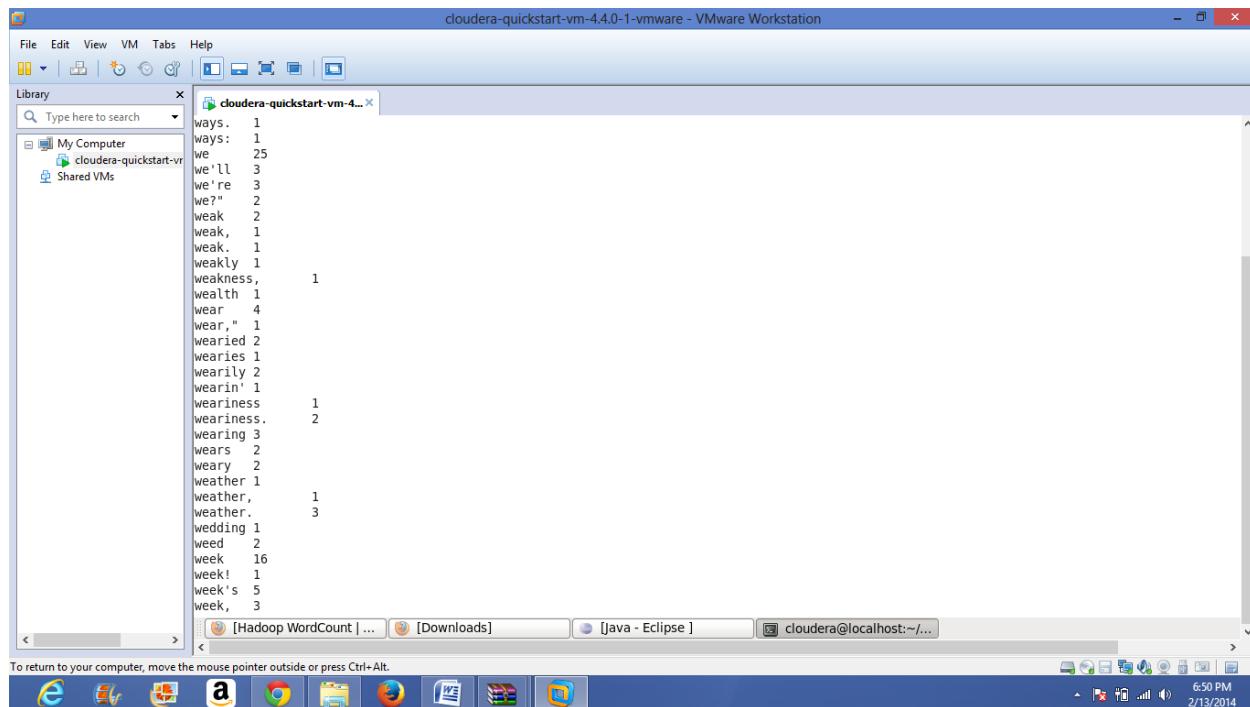


Then apply the commands

hadoop fs -put input input      hadoop jar wordcount.jar WordCount input output3



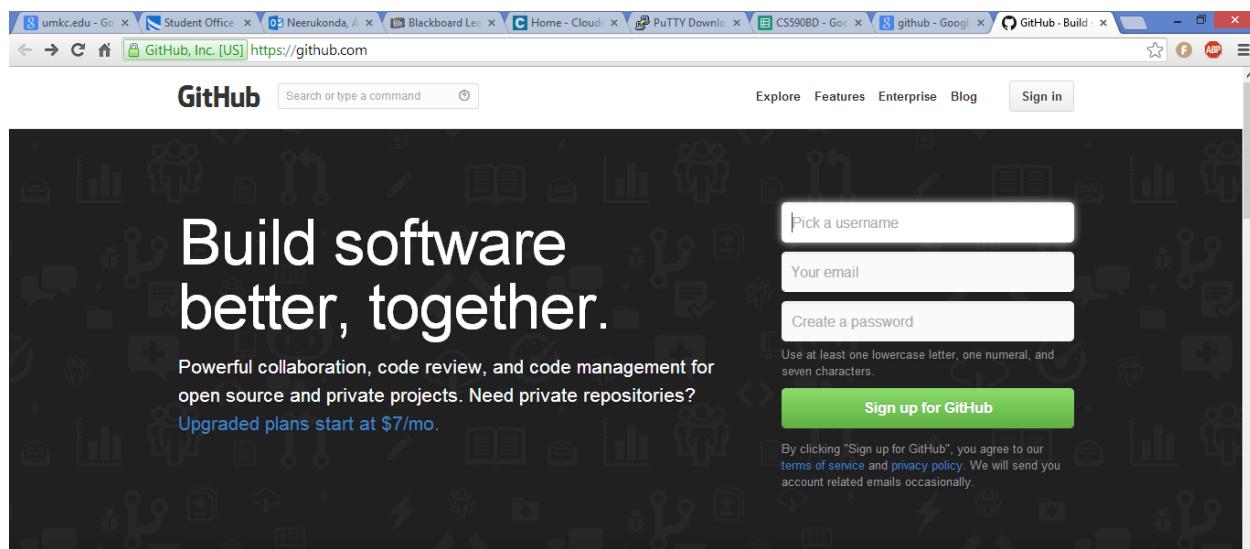
view the result from output3 by using command hadoop fs -cat output3/\*



```
cloudera-quickstart-vm-4.4.0-1-vmware - VMware Workstation
File Edit View VM Tabs Help
Library Type here to search
My Computer cloudera-quickstart-v...
Shared VMS
cloudera-quickstart-vm-4... ways: 1
ways: 1
we: 25
we'll: 3
we're: 3
we?" 2
weak: 2
weak, 1
weak. 1
weakly: 1
weakness, 1
wealth: 1
wear: 4
wear," 1
wearied: 2
wearies: 1
wearily: 2
wearin': 1
weariness: 1
weariness. 2
wearing: 3
wears: 2
weary: 2
weather: 1
weather, 1
weather. 3
wedding: 1
weed: 2
week: 16
week!: 1
week's: 5
week,: 3
To return to your computer, move the mouse pointer outside or press Ctrl+Alt.
[Hadoop WordCount | ...] [Downloads] [Java - Eclipse] cloudera@localhost:~/...
6:50 PM 2/13/2014
```

## Creating GITHUB Account:

go to web page [github.com](https://github.com) and click on sign up for github.



Why you'll love GitHub.



Then provide all the required information as of necessary.

The screenshot shows the first step of creating a personal account on GitHub. It includes fields for Username (bdan5kc), Email Address (an5kc@mail.umkc.edu), Password, and Confirm your password. A red error box at the top states "There were problems creating your account." Below the form, a note says "You will occasionally receive account related emails. We promise not to share your email with anyone." To the right, a sidebar lists benefits: "Unlimited collaborators", "Unlimited public repositories", "Great communication", "Friction-less development", and "Open source community". At the bottom, there's a link to the Terms of Service and Privacy Policy, and a green "Create an account" button.

After that choose for free plan.

The screenshot shows the second step of choosing a personal plan. It lists five options: Large (\$50/month, 50 repos), Medium (\$22/month, 20 repos), Small (\$12/month, 10 repos), Micro (\$7/month, 5 repos), and Free (\$0/month, 0 repos). The "Free" plan is selected and highlighted. To the right, a sidebar titled "Each plan includes:" lists benefits: "Unlimited collaborators", "Unlimited public repositories", "Free setup", "SSL Protection", "Email support", and "Wikis, Issues, Pages, & more". At the bottom, there's a note about canceling or upgrading, a checkbox for "Help me set up an organization next", and a green "Finish sign up" button.

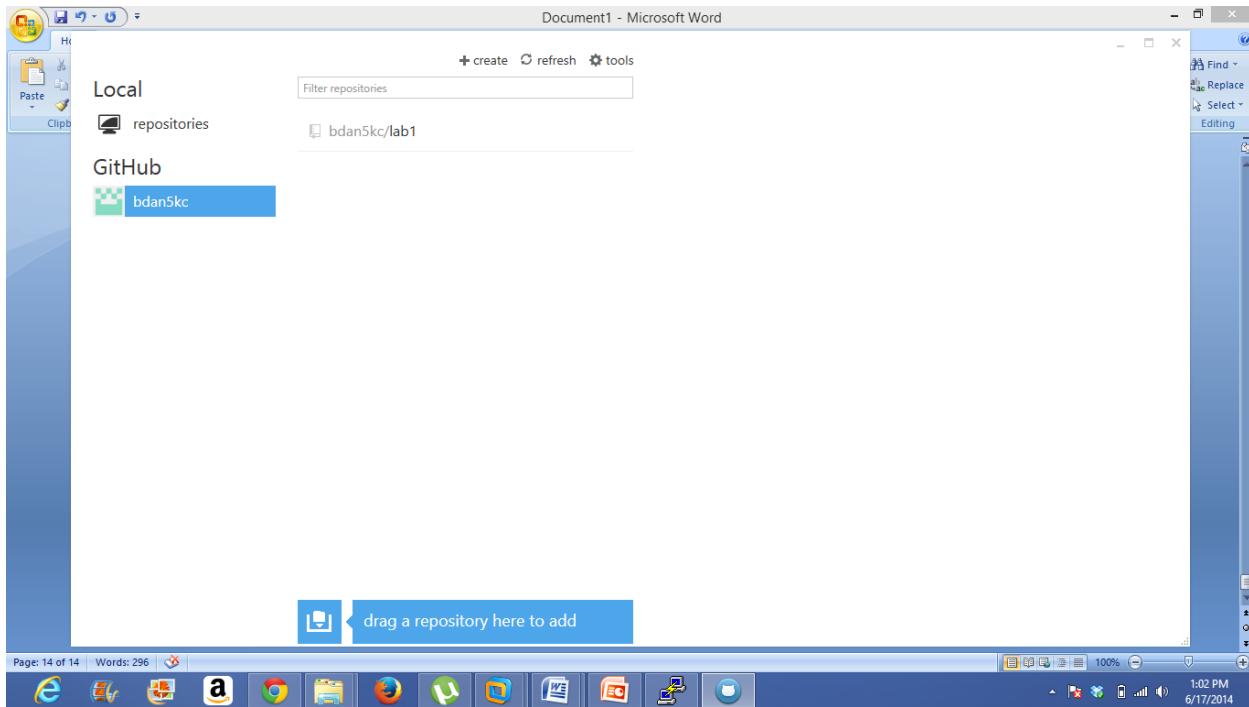
Then create a new repository.

The screenshot shows a browser window with multiple tabs open. The active tab is 'GitHub, Inc. [US] https://github.com/new'. The page displays a form for creating a new repository. The 'Owner' dropdown is set to 'bdan5kc'. The 'Repository name' field contains 'lab1'. Below the name, a note says 'Great repository names are short and memorable. Need inspiration? How about [scaling-meme](#)'. The 'Description (optional)' field is empty. Under 'Visibility', the 'Public' radio button is selected, with the note 'Anyone can see this repository. You choose who can commit.' The 'Private' radio button is also available. There is an unchecked checkbox for 'Initialize this repository with a README', which includes a note about skipping if already run locally. At the bottom are buttons for 'Add .gitignore: None' and 'Add a license: None'. A large green 'Create repository' button is at the very bottom.

Then login into github in your local machine.

The screenshot shows a browser window with multiple tabs open. The active tab is 'github - Google Chrome'. The page displays a 'log in' form. On the left, there is a sidebar with 'Local' and 'GitHub' sections. The main area has 'Log in' and 'Cancel' buttons. Below them is a 'Sign up' section with the text 'Powerful collaboration, review, and code management for open source and private development projects. Build software better, together.' and a 'Sign up' link. The status bar at the bottom shows the date and time as '6/17/2014 1:01 PM'.

Then clone the repository that is created in website into your local machine.



After cloning the folder for the repository will appear in machine then paste all the files you want to publish in to that folder after that commit the changes and sync the repository then the content will also be available in github webpage. Then copy that link and paste it to the Google documents page.

### Creating ScrumDo account:

A screenshot of the ScrumDo website. On the left, there is a sign-up form with fields for 'Username', 'Password', and 'Email', and a 'Create Account' button. Below the form is a link 'Free Trial Available'. On the right, there is a 'ScrumDo' interface showing a 'Stories' backlog board. The board has a header for 'Iteration 4' with metrics: 8 stories, 60 total points, 11 points in progress, and 49 points completed. Below the header is a burndown chart. The backlog board lists several stories with their details. At the bottom of the page, there is a large graphic displaying '4,293,133 Stories Completed'.

After going to scrumdo page click on create account button and provide all the necessary information.  
After that choose free plan.

The screenshot shows a Microsoft Edge browser window with the URL [https://www.scrumdo.com/subscription/bigdata-analytics-and-applications?first\\_time=1](https://www.scrumdo.com/subscription/bigdata-analytics-and-applications?first_time=1). The main content is a table comparing six subscription plans: Diamond, Platinum, Gold, Silver, and Bronze. The table includes a header row with descriptions for each plan. The 'Diamond' plan is described as 'All you could need'. The 'Platinum' plan is 'For large companies'. The 'Gold' plan is 'Most popular plan!'. The 'Silver' plan is 'Great for teams'. The 'Bronze' plan is 'Entry level Scrum'. The table rows list various features with corresponding icons or status indicators (e.g., checkmarks, question marks). The bottom of the browser window shows the Windows taskbar with various pinned icons.

	Diamond All you could need	Platinum For large companies	Gold Most popular plan!	Silver Great for teams	Bronze Entry level Scrum
Users	150	60	30	15	7
Projects	Unlimited	Unlimited	20	10	3
File Attachments	50GB	20GB	5GB	3GB	1GB
Integrations <small>GithHub, Basecamp, Jira &amp; Harvest</small>	?	?	?	?	?
Planning Poker	?	?	?	?	?

Create a project by providing the name.

The screenshot shows a Microsoft Edge browser window with the URL <https://www.scrumdo.com/projects/project/bigdata-analytics-and-applications/create#>. The main content is a 'Create Project' form. It has fields for 'Name\*' (containing 'big data'), 'Project type\*', and a 'Tip' box. The 'Project type\*' field has two options: 'Time Boxed' (selected) and 'Continuous Flow (beta)'. A tip box states: 'Tip: Time boxed projects are perfect for traditional Scrum projects with fixed iterations. Continuous flow projects are best for Scrumban projects, which layer Kanban principles onto Scrum. You can read more here.' At the bottom right of the form is a 'Create Project' button. The bottom of the browser window shows the Windows taskbar with various pinned icons.

Create iterations to the project.

The screenshot shows the ScrumDo interface for a project named 'big data'. The main dashboard displays 'Total Stories' and 'Stories Complete' both at 0. Below this, there are links for 'Scrum Log' and 'Newsfeed'. A message at the bottom indicates that 'bdan5kc created project big data' 0 minutes ago. On the right side, a sidebar titled 'big data' contains links for Project Summary, History, Chat, Project Admin, Epics, Predictions, Export Project, Files, Iteration Planning, Planning Poker, Extras, and Track Time. It also includes sections for Backlog, Story List, Export Iteration, Iteration Admin, Scrum Board, Import Iteration, Files, and Print. A prominent green button labeled 'New iteration' is located in the 'Iterations' section.

While creating the iteration provide the name and starting and ending dates of iteration.

The screenshot shows the 'New Iteration' creation form. The 'Name\*' field is filled with 'iteration-1'. The 'Start date' is set to '2014-06-17' and the 'End date' is set to '2014-06-24'. The 'Include in Velocity Calculations' checkbox is checked. At the bottom right of the form is a blue button labeled 'Add Iteration'.

Then add the members to the project.

The screenshot shows the ScrumDo interface for managing team members. At the top, there's a banner for a free plan with features like Email Notifications, File Attachments, Organization Exports, Premium Integrations, SSL Security, Instant Charts, and Time Tracking. Below this, the main title is "bigdata analytics and applications Teams". A sidebar on the left lists "Members (0 members)" and "Staff (1 members)". A green button labeled "New Team" is visible. The main area is titled "Members" with a pencil icon. It shows a pending invite for "gvishnu146@gmail.com". Below this is a section for "Add User" with an "Invitation Sent" message and a text input field for "Username or Email Address". A green "Add User" button is next to it. Another section titled "Grants write access to:" has a dropdown menu for selecting a project and a green "Add Project To Team" button. The Windows taskbar at the bottom shows various pinned icons and the date/time as 1:13 PM on 6/17/2014.

Create stories to the iteration. Each story represents what work has to be done and assign this story to each member of the team.

The screenshot shows the ScrumDo interface for creating a new story. The URL in the address bar is https://www.scrumdo.com/projects/project/big-data5/iteration/103371. The main form has a title "creating web page" and a "Detail:" text area containing "creating login details". There are fields for "Tags", "Points" (radio buttons for 0, 0.5, 1, 2, 3, 5, 8, 13, 20, 40, 100, and Infinite), "Estimate" (a time input field showing 0 : 00), and "Assigned To" (a text input field with "vishnu"). Below the form are "Story Rank" controls (Top, Bottom) with a note about dragging stories. There are also "Epic:" and "Iteration:" dropdown menus. A tip at the bottom suggests using Markdown for story summaries. The Windows taskbar at the bottom shows various pinned icons and the date/time as 1:15 PM on 6/17/2014.

The status of the stories change from time to time in order to see the status of all stories click on scrum board. At starting the status of story is to do then it will change from to do to doing then to reviewing then to done.

A screenshot of a web browser window showing the ScrumDo interface. The URL is <https://www.scrumdo.com/projects/project/big-data5/iteration/103371/board#>. The main area shows a 'Todo' board with one item: '#1 creating web page'. A 'Board Options' dialog box is open, titled 'Configure your board here.' It contains sections for 'Project Summary', 'History', 'Chat', 'Project Admin', 'Iterations', and 'Iterations Admin'. On the right side of the screen, there are navigation links for 'big data analytics and api', 'big data', and 'bdan5kc'. The bottom of the screen shows a taskbar with various icons and the system tray indicating the date and time as 6/17/2014 at 1:16 PM.

we can also change the status of the story manually.

A screenshot of the ScrumDo interface showing a 'Stories' board. A story card for '#1 creating web page' is selected, showing its current status as 'Todo'. To the right of the card is a status dropdown menu with four options: 'Todo' (dark grey), 'Doing' (blue), 'Reviewing' (orange), and 'Done' (green). Above the stories, there is a header with 'Add Story' and 'Filter Board' buttons, and a navigation bar with 'Burnup | Burndown | Stacked | Time' links. The bottom of the screen shows a taskbar and the system tray indicating the date and time as 6/17/2014 at 1:17 PM.

In this manner we can plan the progress of the project.

The screenshot shows a web browser window with multiple tabs open at the top, including 'umkc.edu', 'Student', 'Neerukon', 'Blackboard', 'Home', 'PuTTY', 'CS590BD', 'github', 'ScrumDo', and 'Inbox'. The main content is a ScrumDo project management interface for 'iteration-1' from June 17, 2014, to June 24, 2014. The board has three columns: 'Todo' (dark grey), 'Reviewing' (orange), and 'Done' (light green). A modal window titled 'Board Options' is open over the 'Reviewing' column, with the sub-instruction 'Configure your board here.' Below the columns, there's a search bar and a section for '#1 creating web page' with '0 Comments - Tasks'. The bottom of the screen shows the Windows taskbar with various pinned icons and the system tray indicating the date and time as 1:17 PM on 6/17/2014.