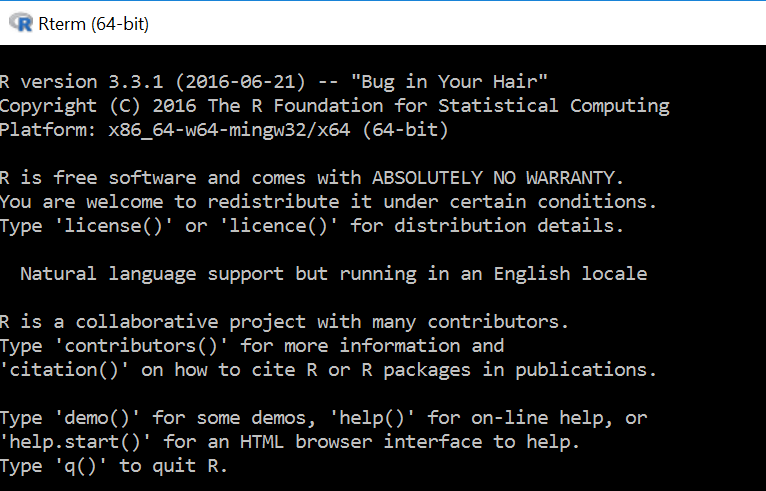
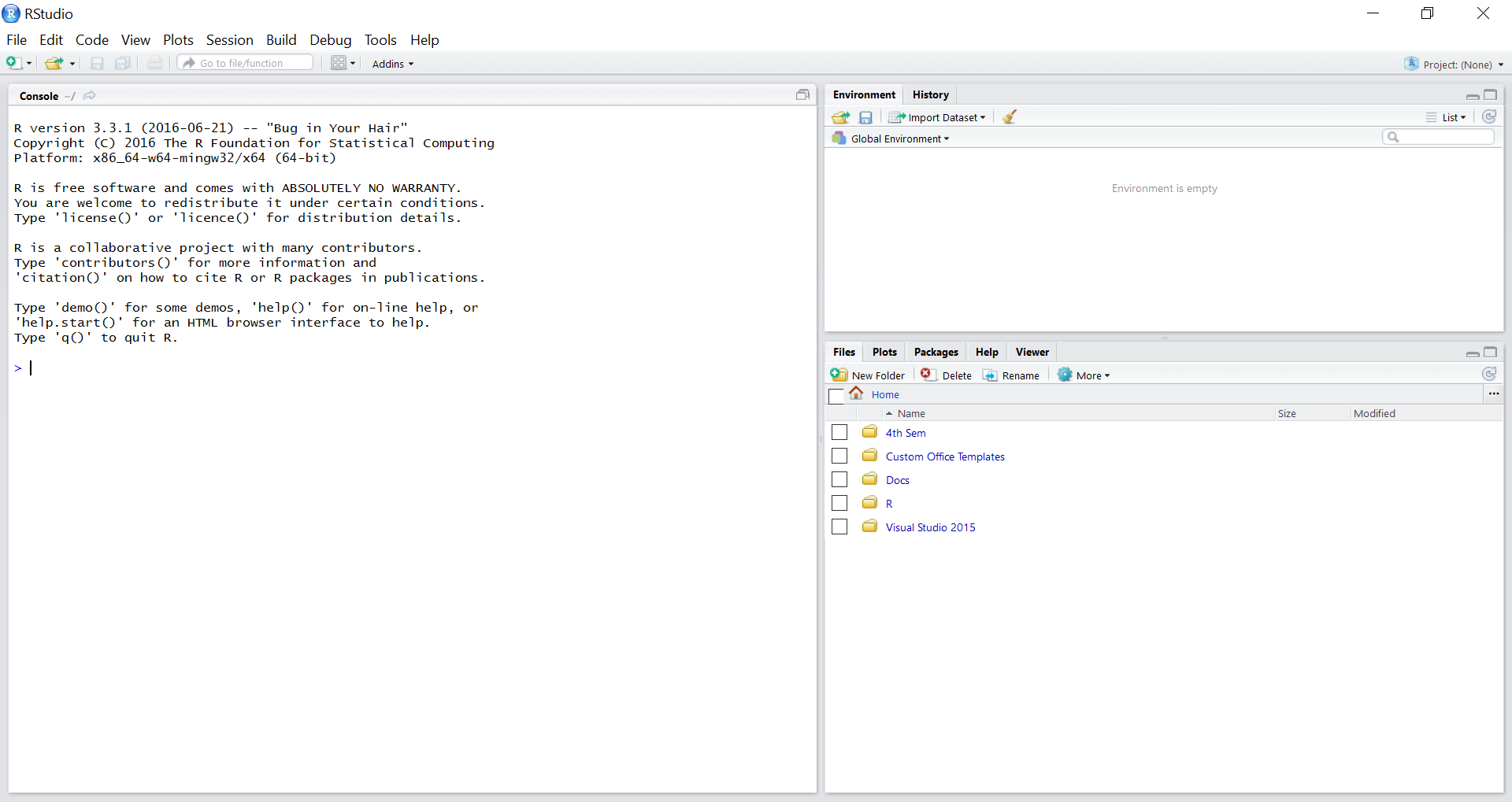
**R Installation**

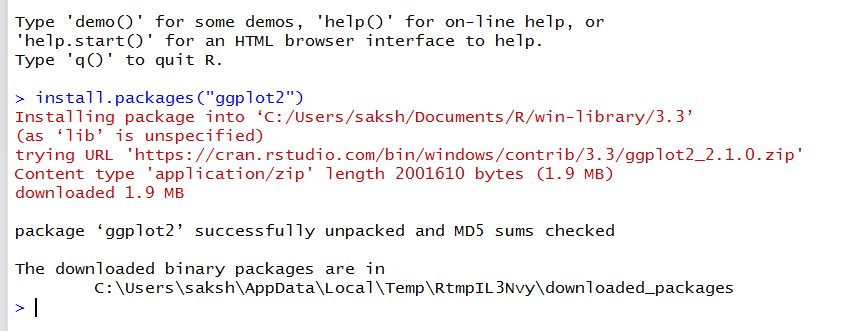
1. Downloaded **R-3.3.1-win.exe** and installed
2. After that downloaded R-Studio (**RStudio-0.99.903.exe**) and installed



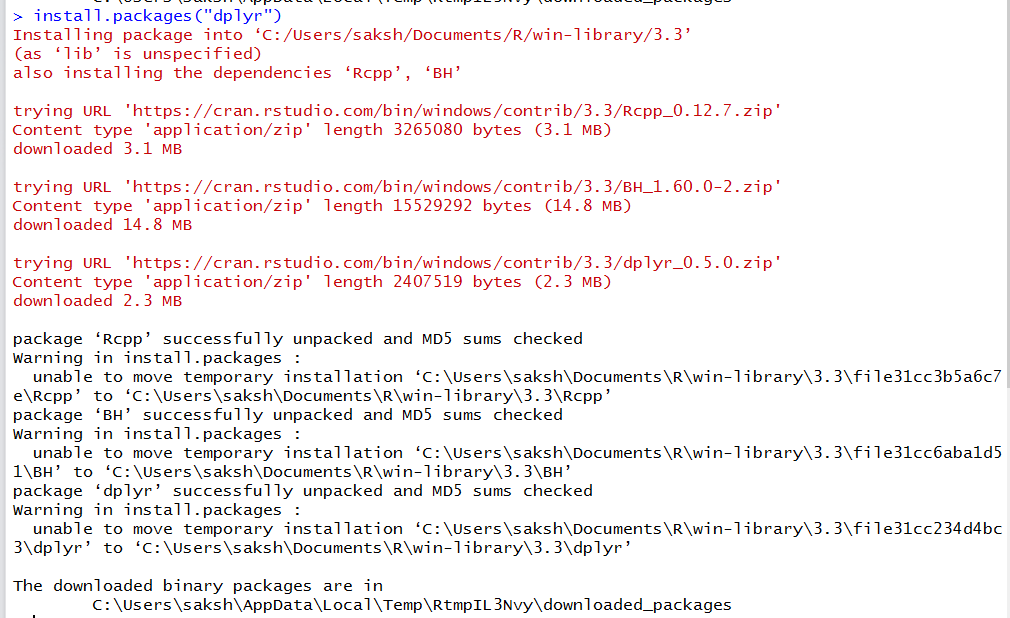
RStudio



1. After this use “**install.packages(“ggplot2”)** to install package ggplot



1. After this use “**install.packages(“dplyr”)** to install package dplyr

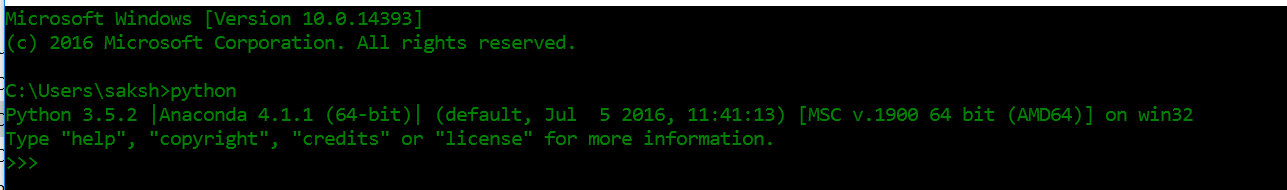


Reference Link:

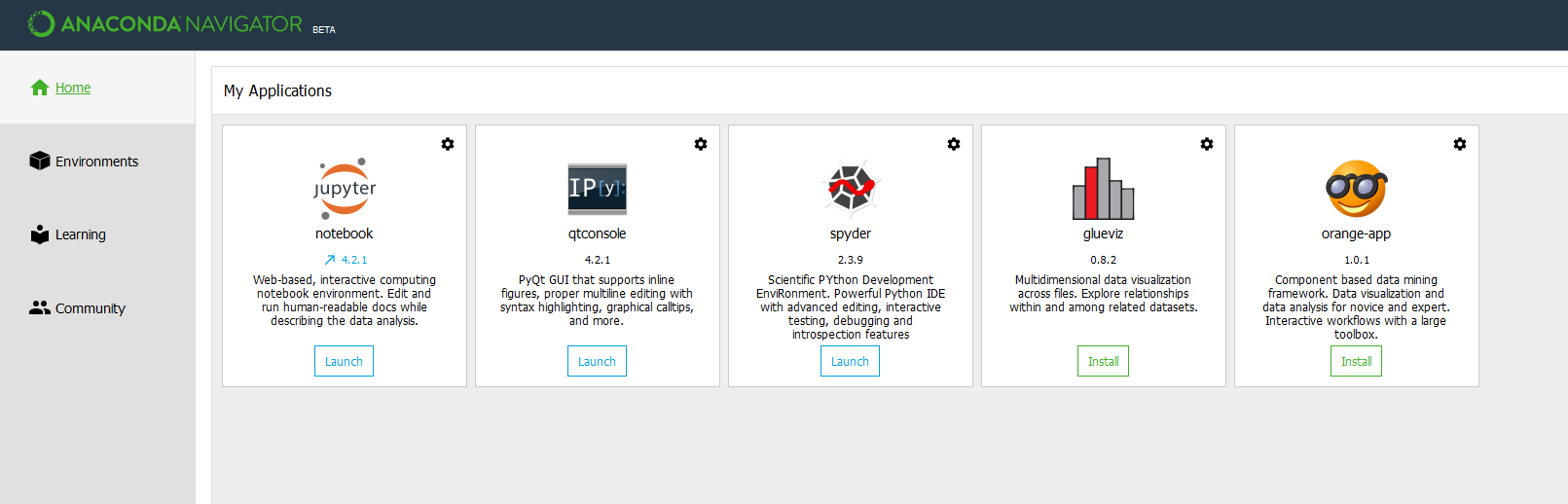
<https://cran.r-project.org/bin/windows/base/>

**Python Installation**

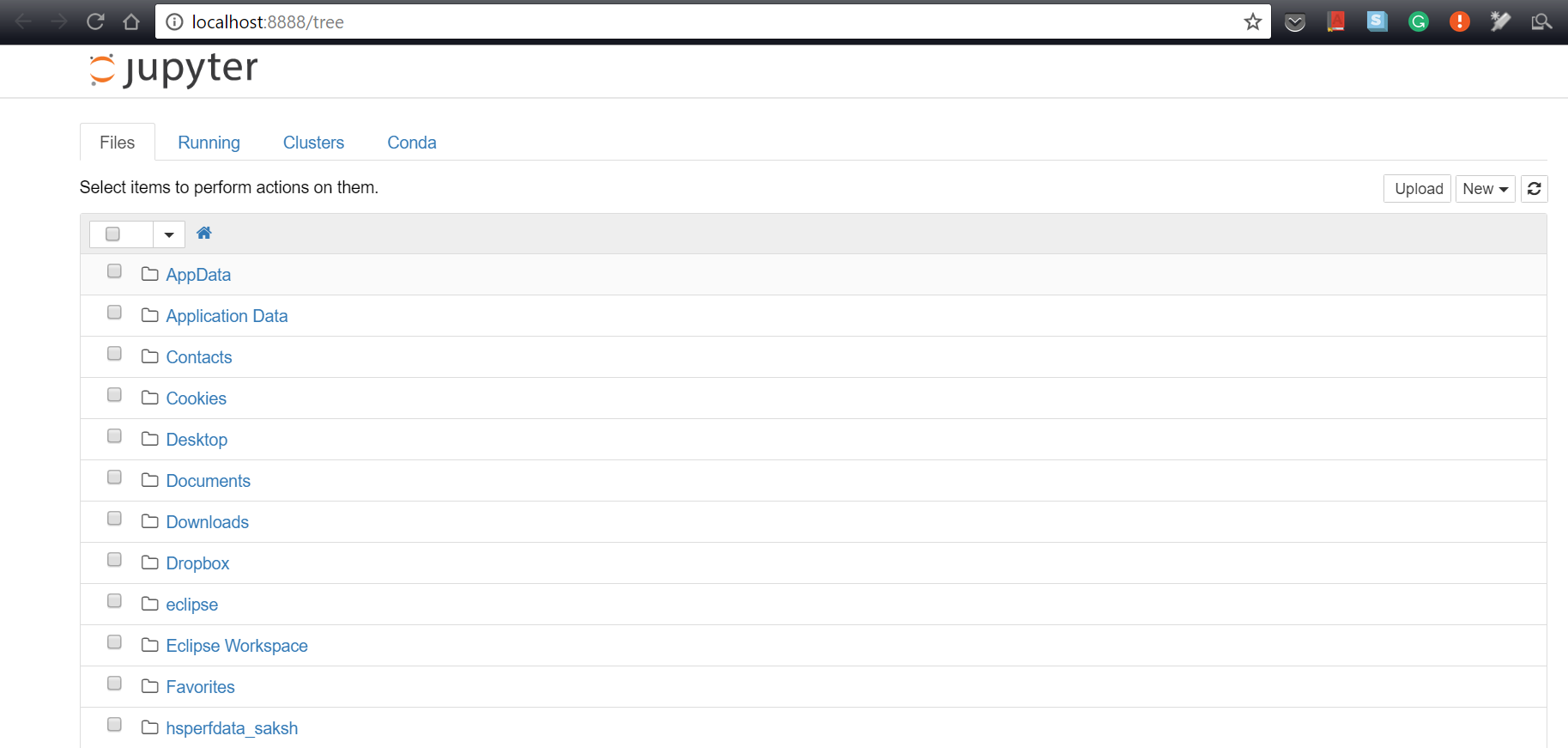
1. Downloaded Anaconda from the Anaconda website (Anaconda3-4.1.1-Windows-x86\_64.exe)
2. It will install Python 3.5



1. Anaconda Navigator (an IDE for python)



1. Jupyter notebook is also installed for Python. By clicking launch, we can start **localhost:8888/tree**

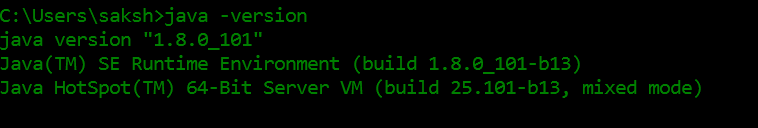


Reference:

<https://www.continuum.io/downloads>

**Java 8 Installation**

1. Use Oracle website for installing the java latest version



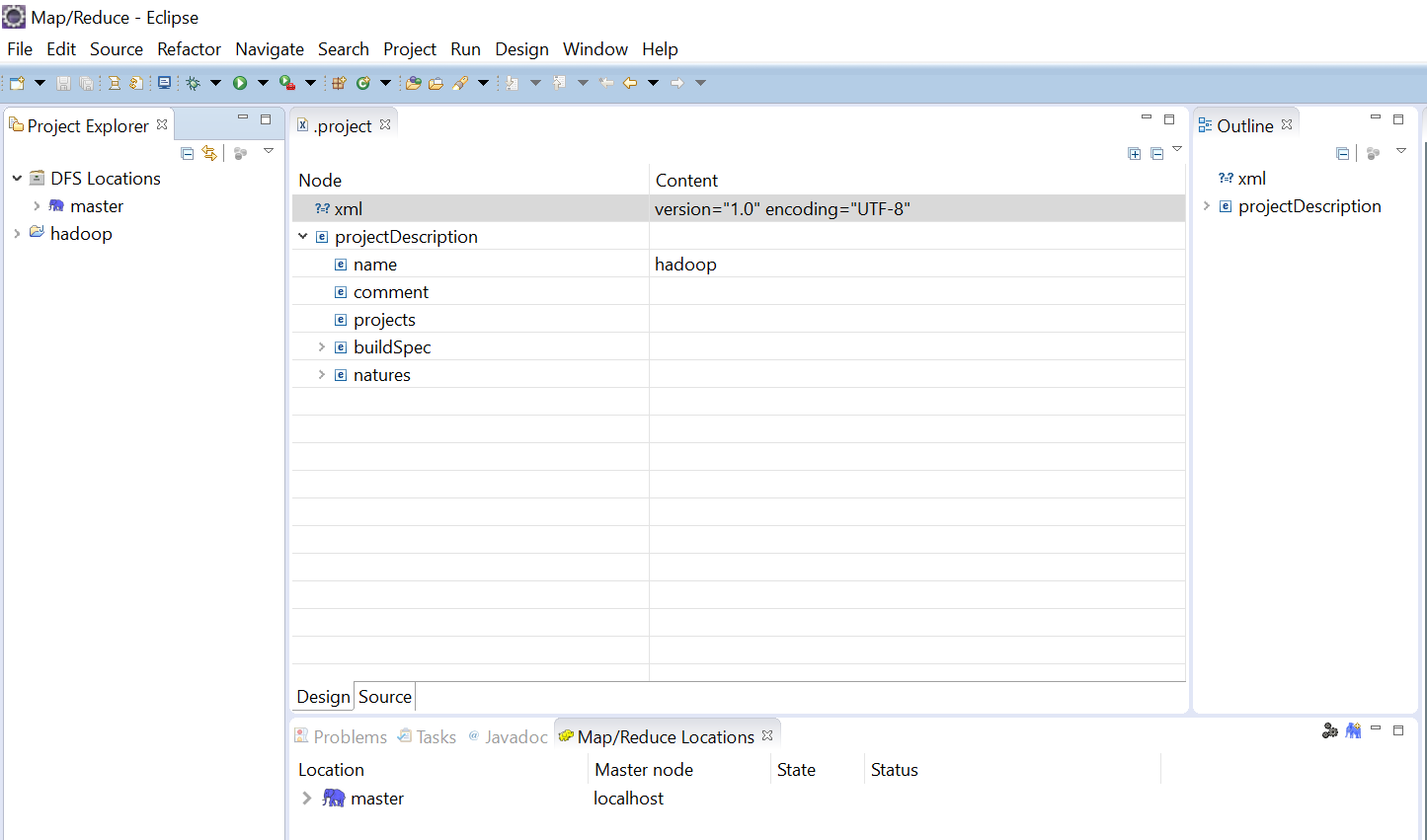
Reference:

<http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>

**Eclipse Mars Installation**

1. Download and install the Eclipse Mars version from the eclipse website



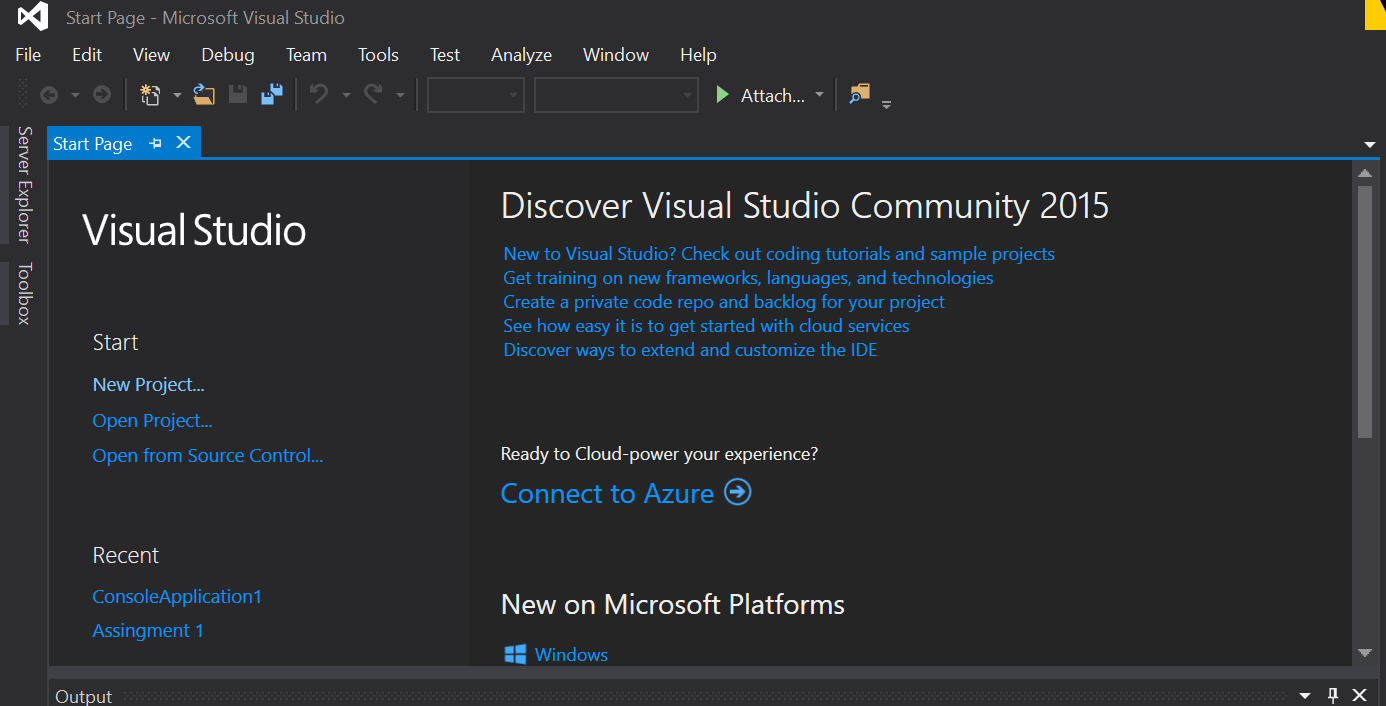


Reference:

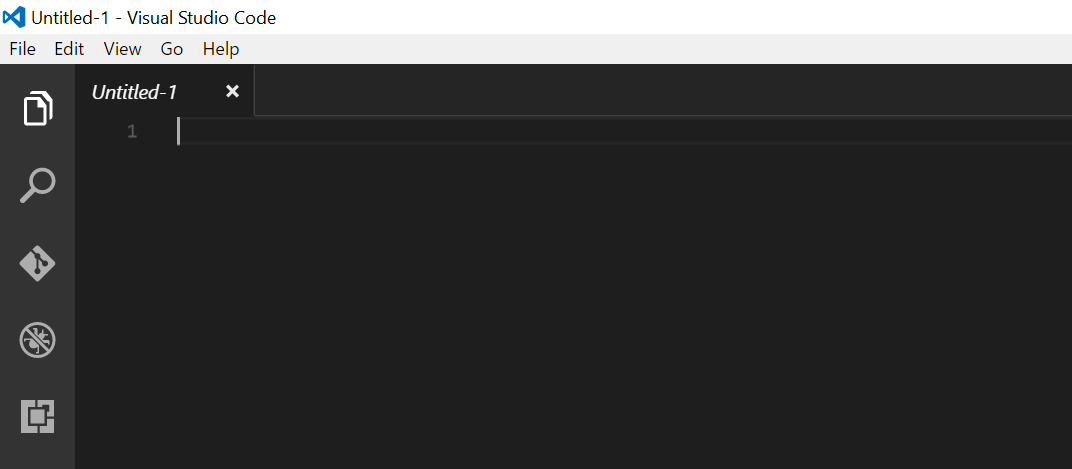
<https://eclipse.org/mars/>

**Visual Studio 15 CE Installation**

1. Download and install Visual Studio 15 community edition from dreamspark account. It is free for students.



1. Download and Install Visual Studio Code from the dreamspark website.



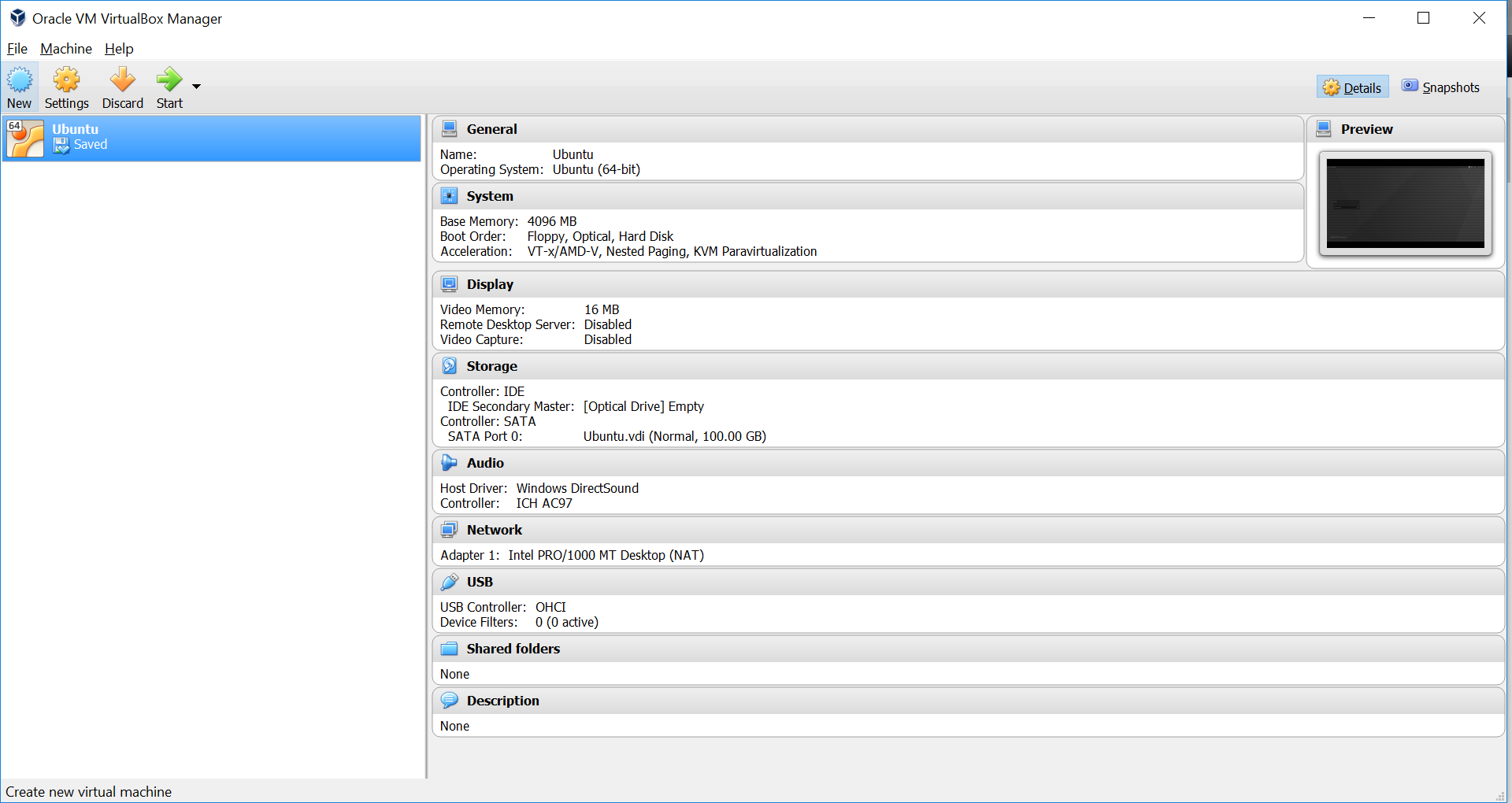
Reference:

<https://catalog.imagine.microsoft.com/en-us/catalog>

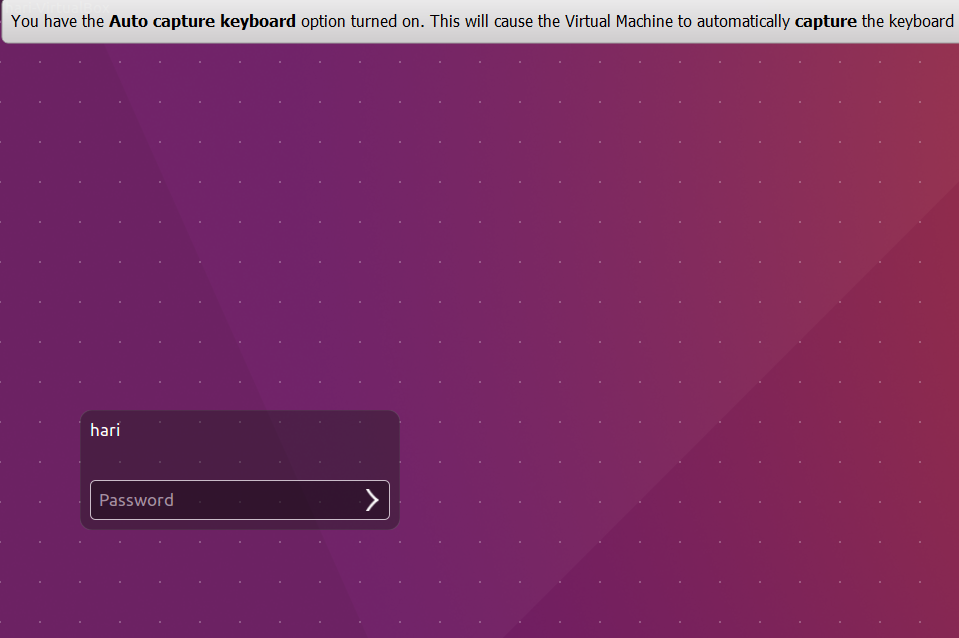
<https://www.visualstudio.com/en-us/downloads/download-visual-studio-vs.aspx>

**Virtual Box and Ubuntu Installation**

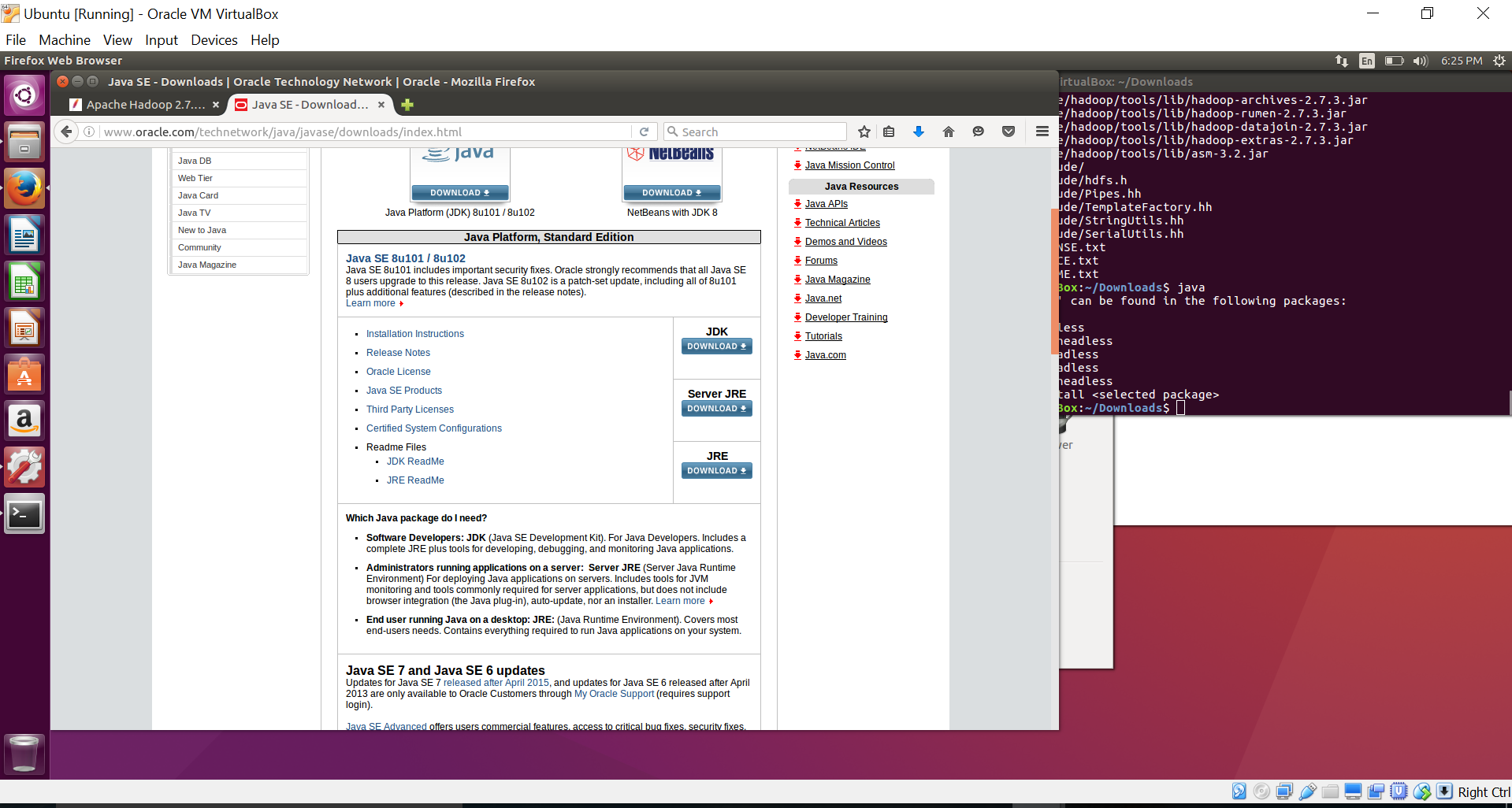
1. Downloaded and installed **VirtualBox 5.1.6 for Windows hosts** from the virtualbox website.
2. Downloaded the latest version of Ubuntu from the ubutu website
3. Allocated 100 GB Hard Disk and 6 GB RAM
4. After the setup is complete the VirutalBox looks like the below image.



1. After starting the virtual box the login screen look like this.



1. This is the desktop of the VM with firefox opened in it.



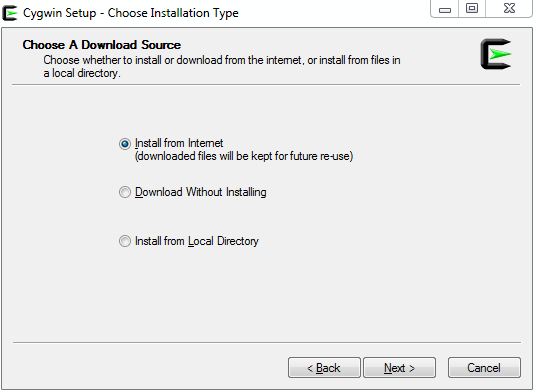
Reference:

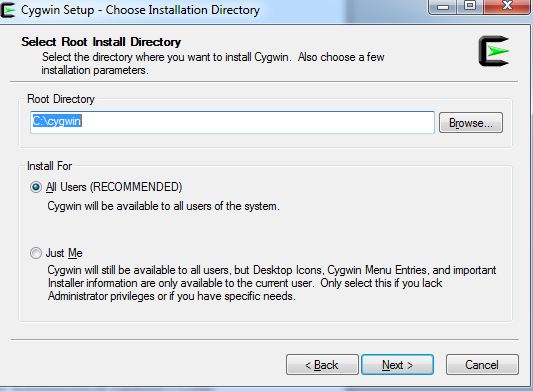
<https://www.virtualbox.org/wiki/Downloads>

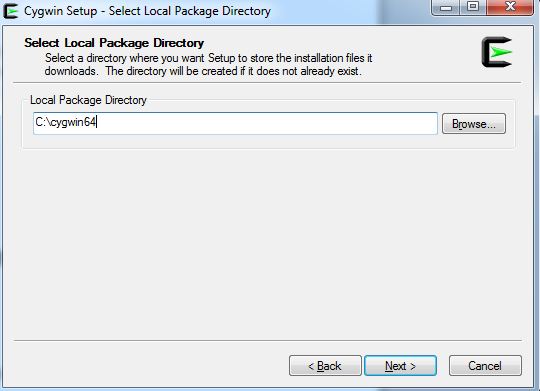
<http://www.ubuntu.com/download/desktop>

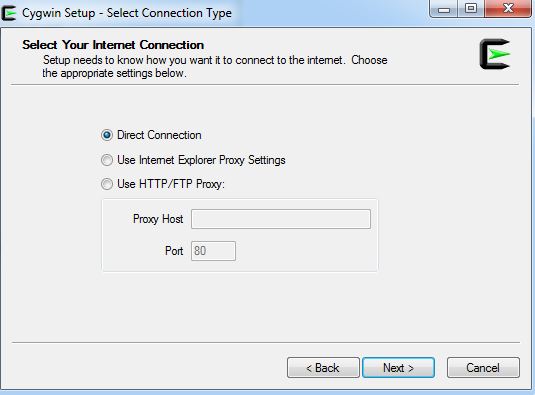
**Cgywin Installation**

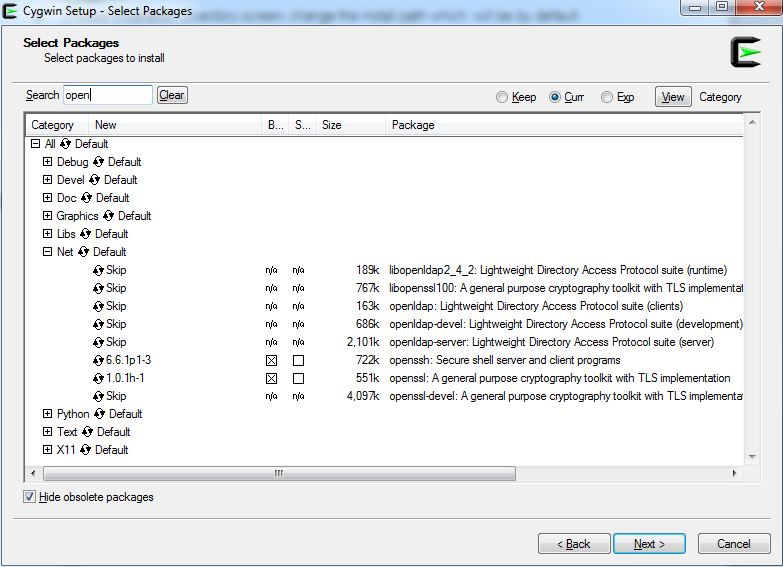
1. Download cgywin exe and install, while installing cgywin download packages: **dos2unix, openssh, openssl**











1. Configure ssh for localhost, so that we can make the connection and run the Hadoop as a Single Node Cluster
2. After you install Cygwin, navigate to the C:\cygwin directory, open the Cygwin.bat file in edit mode using any editor, and add the following line before invoking the bash shell

**set CYGWIN=binmode ntsec**

1. To verify if Cygwin (cygrunsrv) is installed properly, run C:\cygwin\Cygwin.bat, and execute the following command:

cygrunsrv -h

1. If Cygwin is installed properly, then all the Cygwin help options are displayed on the screen. However, if this command returns an error message, then you may have to reinstall Cygwin.
2. To configure the SSHD service, run C:\cygwin\Cygwin.bat, and execute the following command:

ssh-host-config

1. After running the command, you are prompted the following questions:

\*\*\* Query: Should privilege separation be used? <yes/no>: **yes**

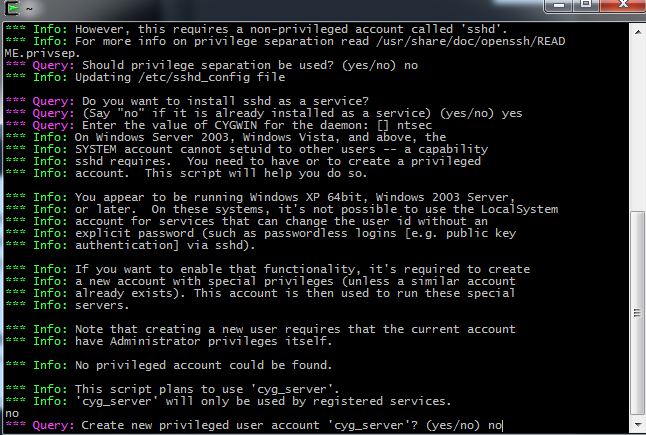
\*\*\* Query: New local account 'sshd'? <yes/no>: **yes**

\*\*\* Query: Do you want to install sshd as a service?

\*\*\* Query: <Say "no" if it is already installed as a service> <yes/no>: **yes**

\*\*\* Query: Enter the value of CYGWIN for the deamon: [] **binmode ntsec**

\*\*\* Query: Do you want to use a different name? (yes/no) **yes/no**



1. At this point, if you want to use the same name, that is cyg\_server, enter no. You are then prompted the following questions:

\*\*\* Query: Create new privileged user account 'cyg\_server'? (yes/no) **yes**

\*\*\* Query: Please enter the password:

\*\*\* Query: Renter:

1. However, if you want to use a different name, enter yes. You are then prompted the following questions:

\*\*\* Query: Enter the new user name: cyg\_server1

\*\*\* Query: Reenter: cyg\_server1

\*\*\* Query: Create new privileged user account 'cyg\_server1'? (yes/no) **yes**

\*\*\* Query: Please enter the password:

\*\*\* Query: Reenter:

1. If the configuration is successful, you will see the following message:

Host configuration finished. Have fun!

1. Execute command **ssh-keygen**, **ssh-keygen** is used to generate that key pair for you. Here is a session where your own personal private/public key pair is created:

ssh-keygen -t rsa

Generating public/private rsa key pair.

Enter file in which to save the key (/home/saksh/.ssh/id\_rsa):

Enter passphrase (empty for no passphrase):

Enter same passphrase again:

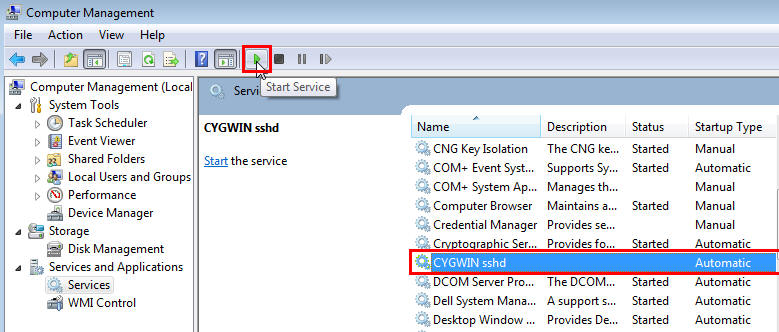
Your identification has been saved in /home/saksh/.ssh/id\_rsa.

Your public key has been saved in /home/saksh/.ssh/id\_rsa.pub.

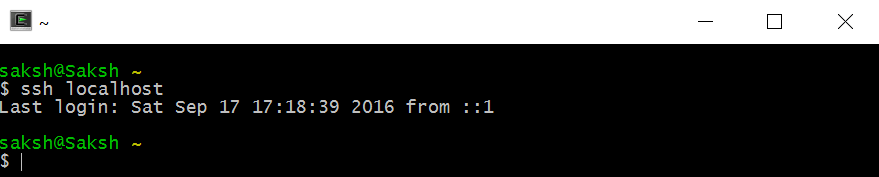
The key fingerprint is:

f6:61:a8:27:35:cf:4c:6d:13:22:70:cf:4c:c8:a0:23 saksh@Saksh

1. Type run in the search bar, execute “**services.misc**” and go to **CGYWIN sshd,** click start



1. Open the Cgywin terminal, and execute **ssh localhost**



Reference:

<https://cygwin.com/install.html>

<https://docs.oracle.com/cd/E24628_01/install.121/e22624/preinstall_req_cygwin_ssh.htm#EMBSC340>

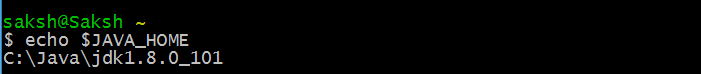
**Hadoop-0.20.2 Installation**

1. Prerequsite:
   1. Update version of Java is installed
   2. Set the Path of Java and Cgywin in the Enviornment variable.
   3. Append the bin folder path of Installed Cygwin C:\cygwin64\bin; and click OK.
   4. Be default the path will be C:\cygwin64\bin; In case the installation path is different, you may have to select a different path.
2. Download Hadoop-0.20.2 from the Apache website.
3. Open Cgywin Terminal (Run as Administrator)
4. Execute the tar command as below to start unpacking the Hadoop package.

tar -xzvf Hadoop-0.20.2.tar.gz

This process might take some time and after that you see Cygwin command prompt again. There is no success message.

1. Execute echo $JAVA\_HOME



1. Open Cygwin and type the command **"explorer ."** to open Home folder
2. Create a folder with name "hadoop-dir". And inside "hadoop-dir" folder create 2 folder with names "datadir" and "namedir"

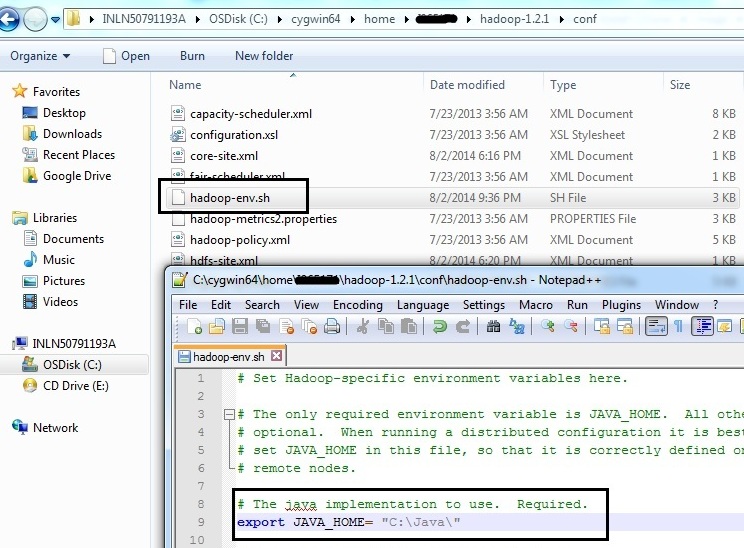
$chmod 755 hadoop-dir

cd hadoop-dir

$chmod 755 datadir

$chmod 755 namedir

1. Go within the “**Hadoop-0.20.2**” folder and go to the “**conf**” folder
2. Open **hadoop-env.sh** file to set Java home as you did it before for environmental variable setup



1. Open **core-site.xml** file and add below code.

<configuration>

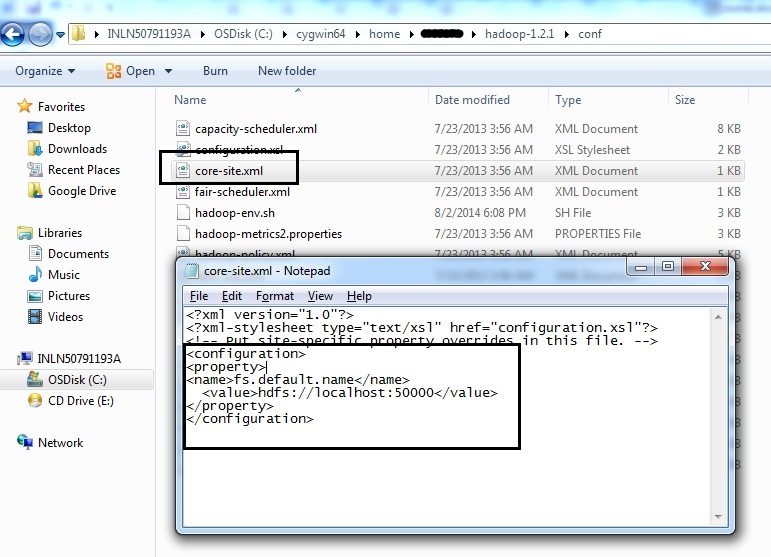
<property>

<name>fs.default.name</name>

<value>hdfs://localhost:50000</value>

</property>

</configuration>



1. Open **mapred -site.xml** file and add below code

<configuration>

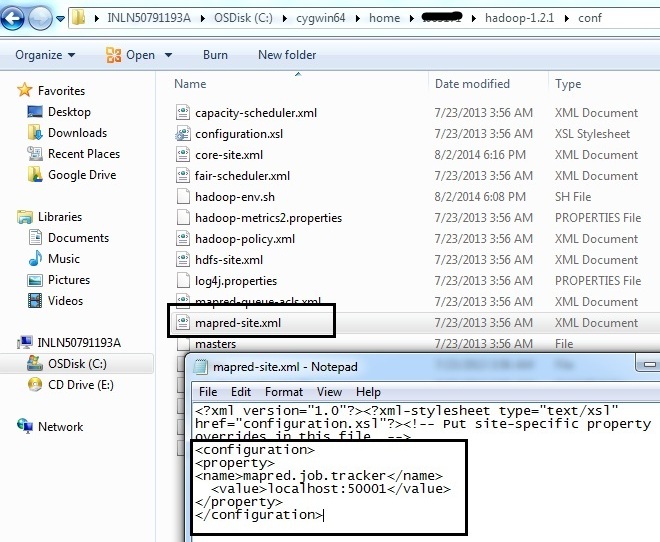
<property>

<name>mapred.job.tracker</name>

<value>localhost:50001</value>

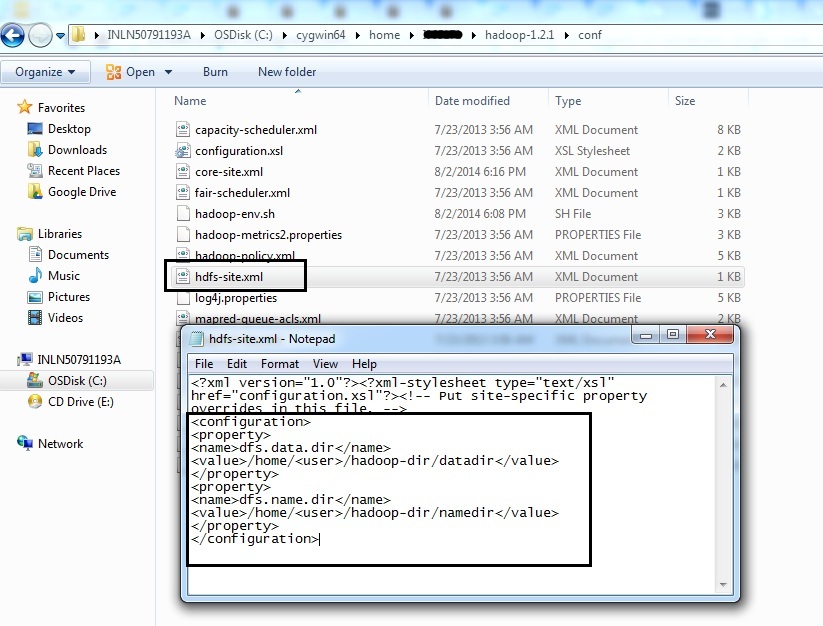
</property>

</configuration>



1. Open **hdfs -site.xml** file and add below code

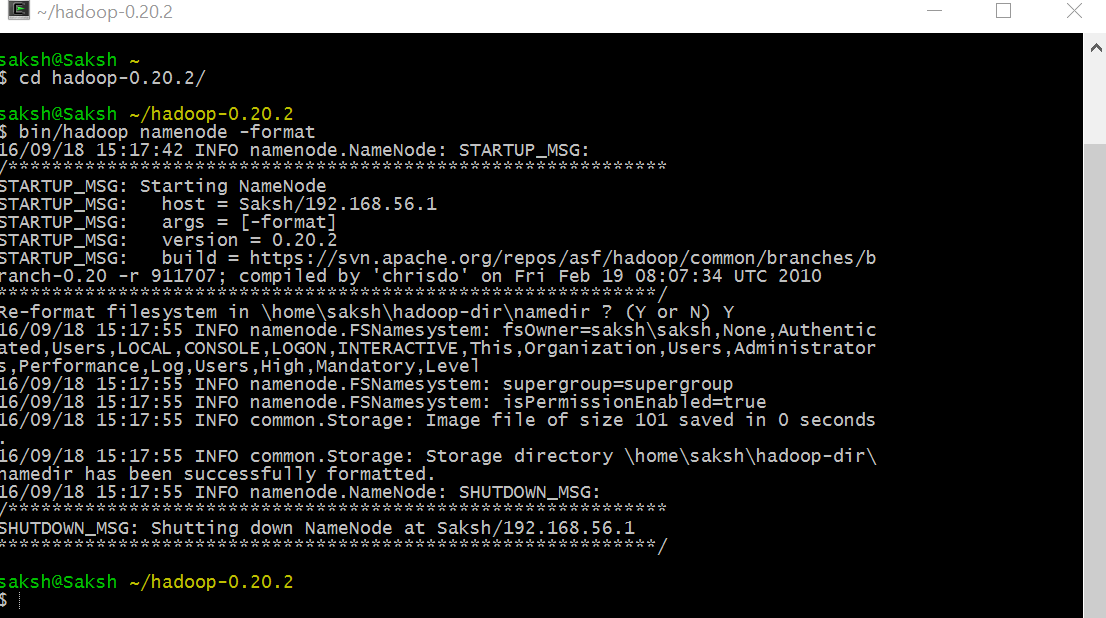
<configuration>   
<property>   
<name>dfs.data.dir</name>   
<value>/home/USER\_FOLDER\_NAME/hadoop-dir/datadir</value>   
</property>   
<property>   
<name>dfs.name.dir</name>   
<value>/home/USER\_FOLDER\_NAME/hadoop-dir/namedir</value>   
</property>   
</configuration>



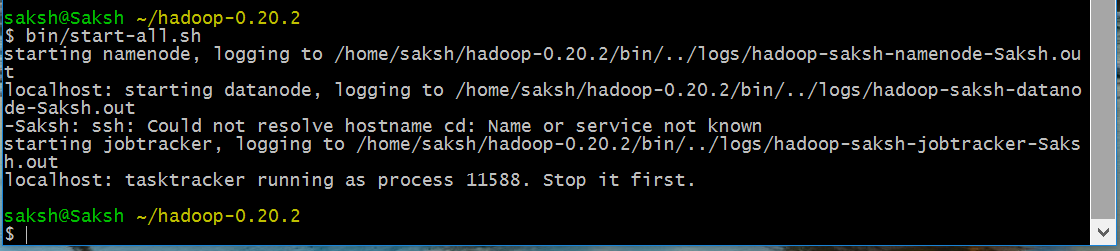
1. For every file we changed, run **dos2unix** command.

**Format the namenode and Run Hadoop Daemons**

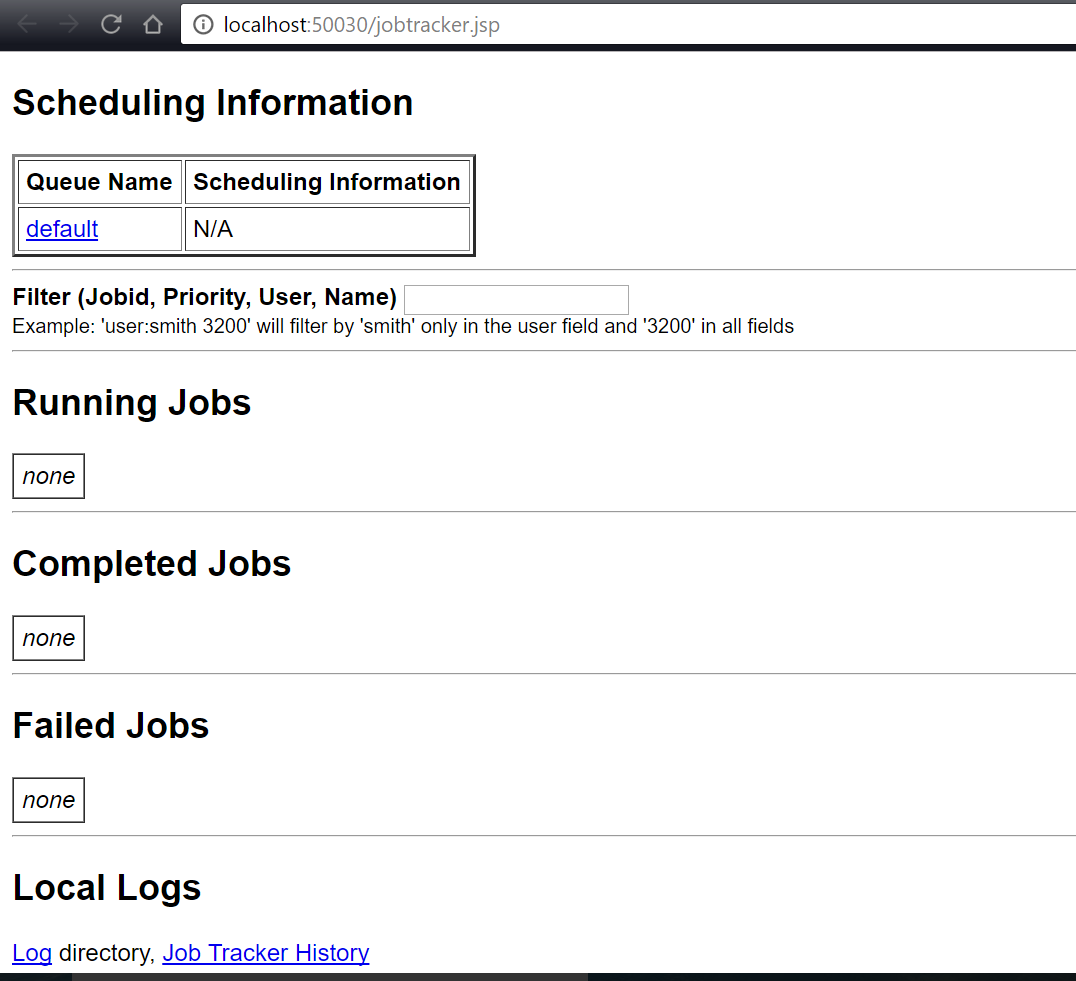
1. Open the Cgywin Terminal (Run as Administrator)
2. Execute **cd Hadoop-0.20.2** and after that “**bin/hadoop namenode -format”**



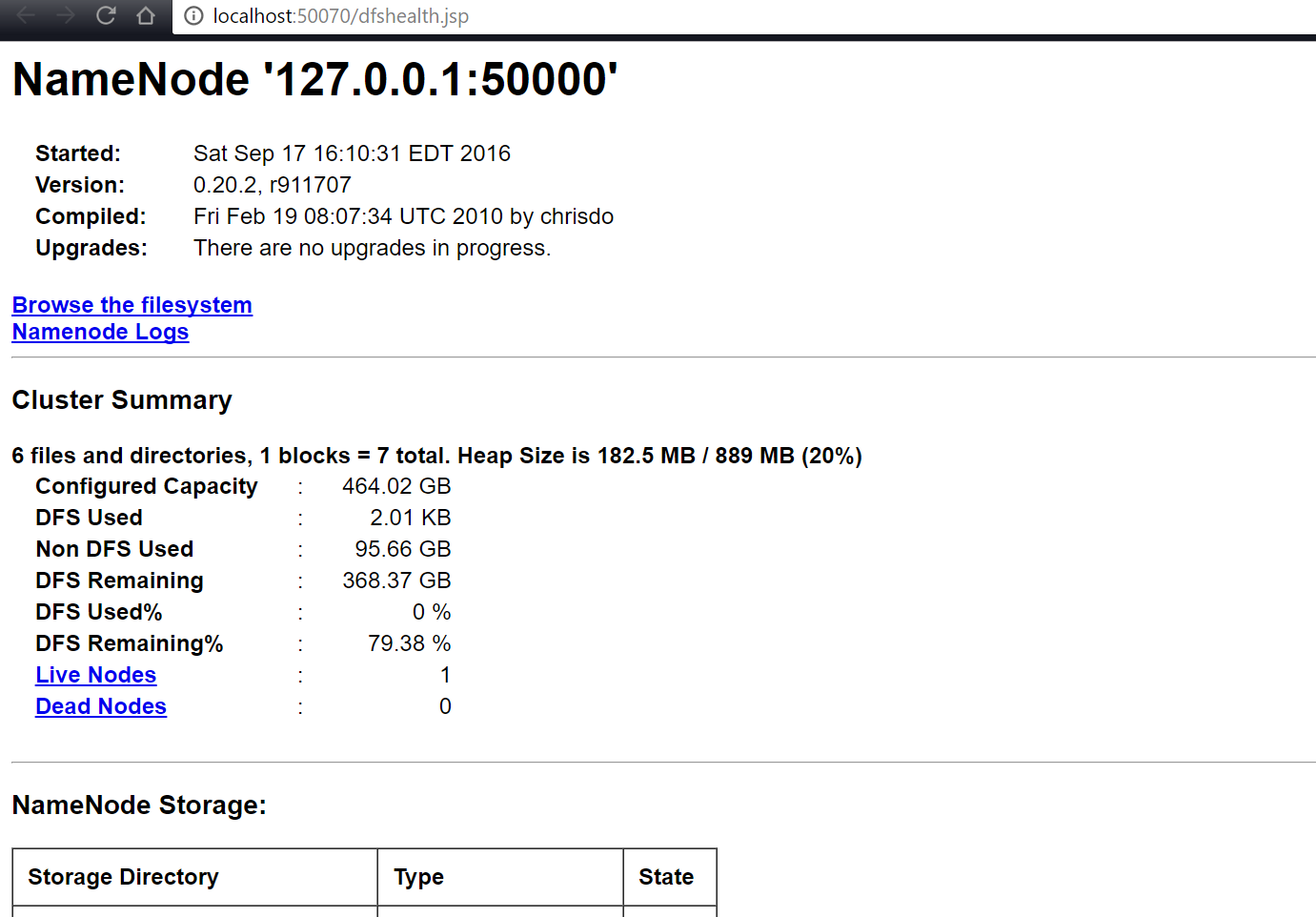
1. Start Hadoop Daemons: Once the filesystem has been created . Next step would be to check and start Hadoop Cluster Daemons NameNode, DataNode, SecondaryNameNode, JobTracker, TaskTracker.
2. Execute **bin/start-all.sh**



1. By default they are available at below address.
   1. **JobTracker :**[**http://localhost:50030/**](http://localhost:50030/)

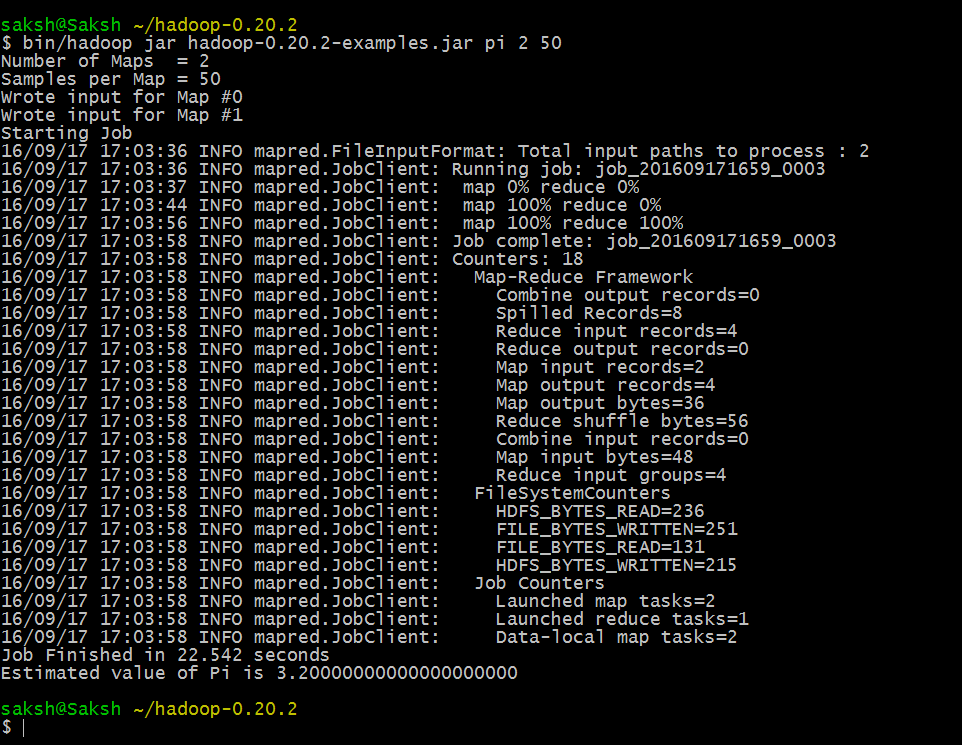


* 1. **NameNode:**[**http://localhost:50070/**](http://localhost:50070/)



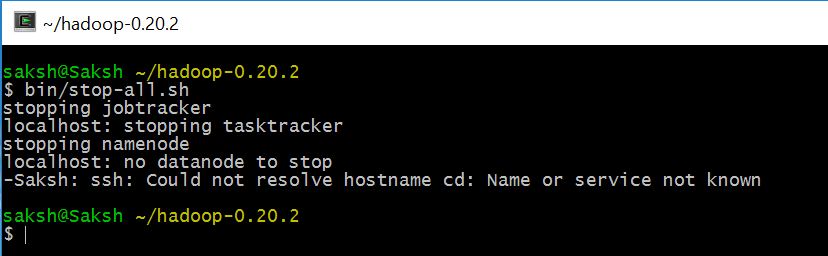
1. Executing the given given example after running the Hadoop

bin/hadoop jar Hadoop-0.20.2-examples.jar pi 2 50



1. **Stop Hadoop Daemons:**To stop all the daemons, we can execute the command

bin/stop-all.sh



**Hadoop-2.7.3 Installation**

1. Hadoop-2.7.3 is installed in the MAC machine, as winutils.exe was not available and cgywin require winutils.exe for the Hadoop-2.7.3 installation.
2. Prerequisite
   1. Java™ must be installed.
   2. ssh must be installed and sshd must be running to use the Hadoop scripts that manage remote Hadoop daemons.
3. **Set JAVA\_HOME** in the terminal for MAC and confirm it is set by printing **echo $JAVA\_HOME**
4. Unzip the Hadoop, using the command **tar -xzvf Hadoop-2.7.3.ta.gz**
5. Execute **bin/Hadoop**, to get the Hadoop documentation
6. Hadoop Configuration for Single Node Cluster
7. In the distribution, edit the file **etc/Hadoop/core-site.xml** to define some

<configuration>

<property>

<name>fs.defaultFS</name>

<value>hdfs://localhost:9000</value>

</property>

</configuration>

1. In the distribution, edit the file **etc/Hadoop/hdfs-site.xml** to define some

<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

</configuration>

1. Set passphraseless ssh, execute **ssh localhost**. If you cannot ssh to localhost without a passphrase, execute the following commands:

$ ssh-keygen -t rsa -P ‘’ -f ~/.ssh/id\_rsa

$ cat ~/.ssh/id\_rsa.pub >> ~/.ssh/authorized\_keys

$ chmod 600 ~/.ssh/authorized\_keys

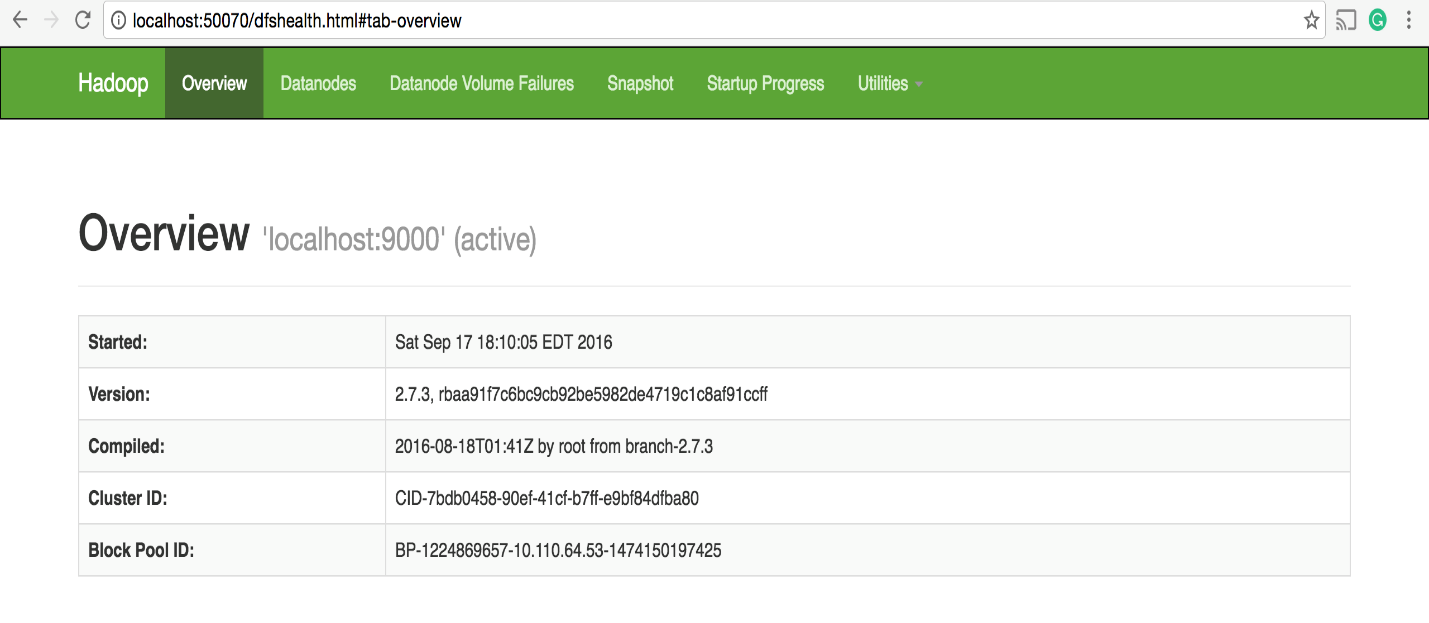
1. Execution
   1. Format the filesystem

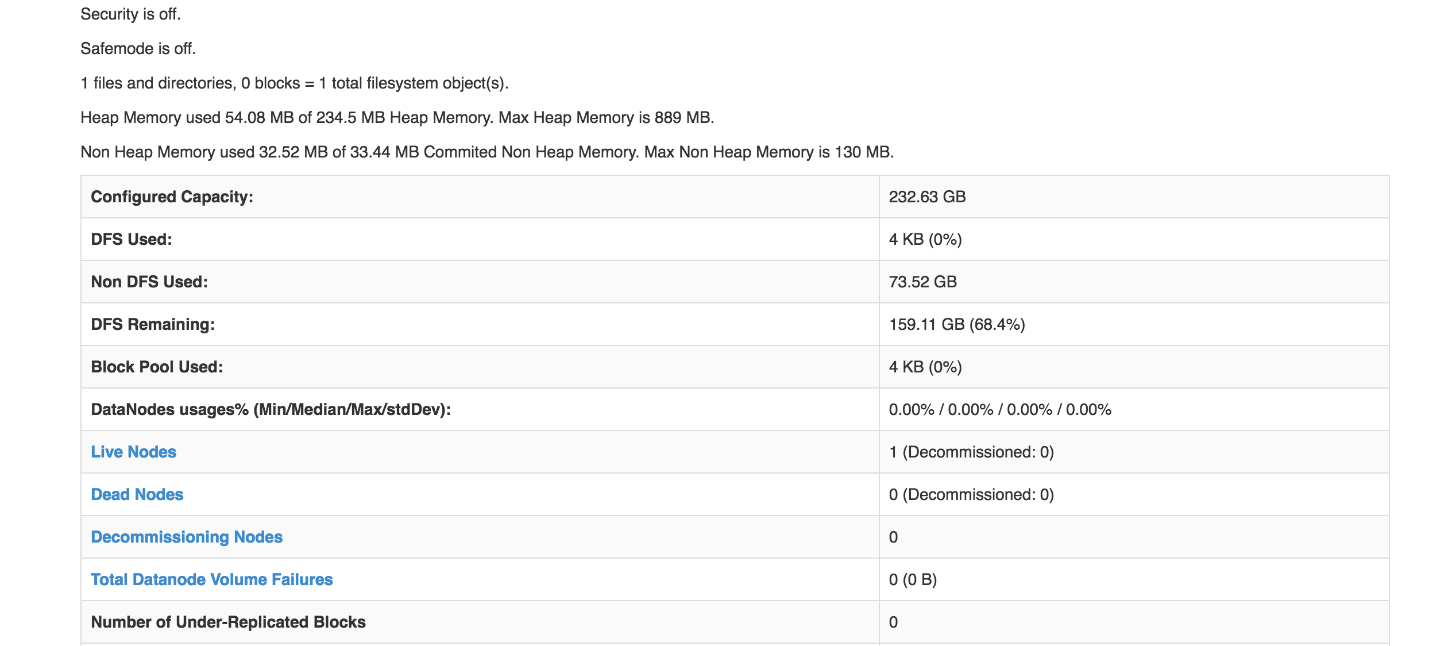
bin/hdfs namenode -format

* 1. Start Namenode and Datanode daemon

sbin/start-dfs.sh

1. Browse the web interface for the NameNode; by default it is available at:
   1. **NameNode -**[**http://localhost:50070/**](http://localhost:50070/)





1. Run a MapReduce job on YARN in a pseudo-distributed mode by setting a few parameters and running ResourceManager daemon and NodeManager daemon in addition.
2. Configure parameters as follows: **etc/hadoop/mapred-site.xml:**

<configuration>

<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

</configuration>

1. Edit **etc/hadoop/yarn-site.xml:**

<configuration>

<property>

<name>yarn.nodemanager.aux-services</name>

<value>mapreduce\_shuffle</value>

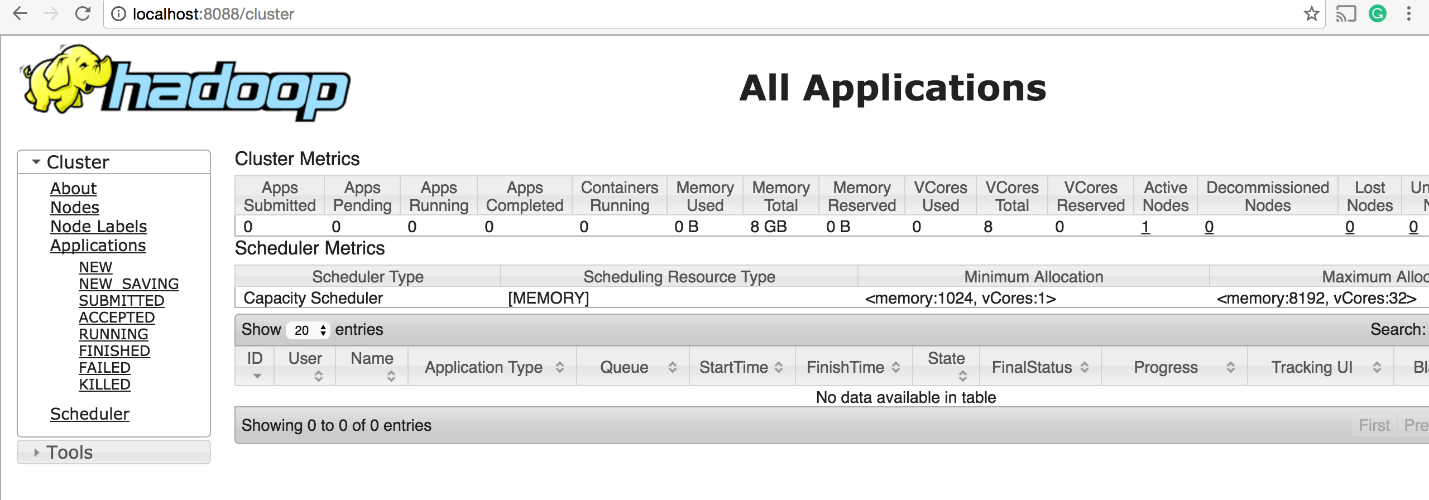
</property>

</configuration>

1. Start ResourceManager daemon and NodeManager daemon:

sbin/start-yarn.sh

1. Browse the web interface for the ResourceManager; by default it is available at:
   1. **ResourceManager -**[**http://localhost:8088/**](http://localhost:8088/)



1. Stop the daemons

sbin/stop-yarn.sh

Reference

<http://saphanatutorial.com/hadoop-installation-on-windows-7-using-cygwin/>

<https://hadoop.apache.org/releases.html>

**Visual Studio Code – HomeWork 4**

**Note:** For the Homework 4, the documentation of the prime number generation and primonacci number is in the Visual Studio code file itself. Explained every part of the code in the comment and the process I followed.