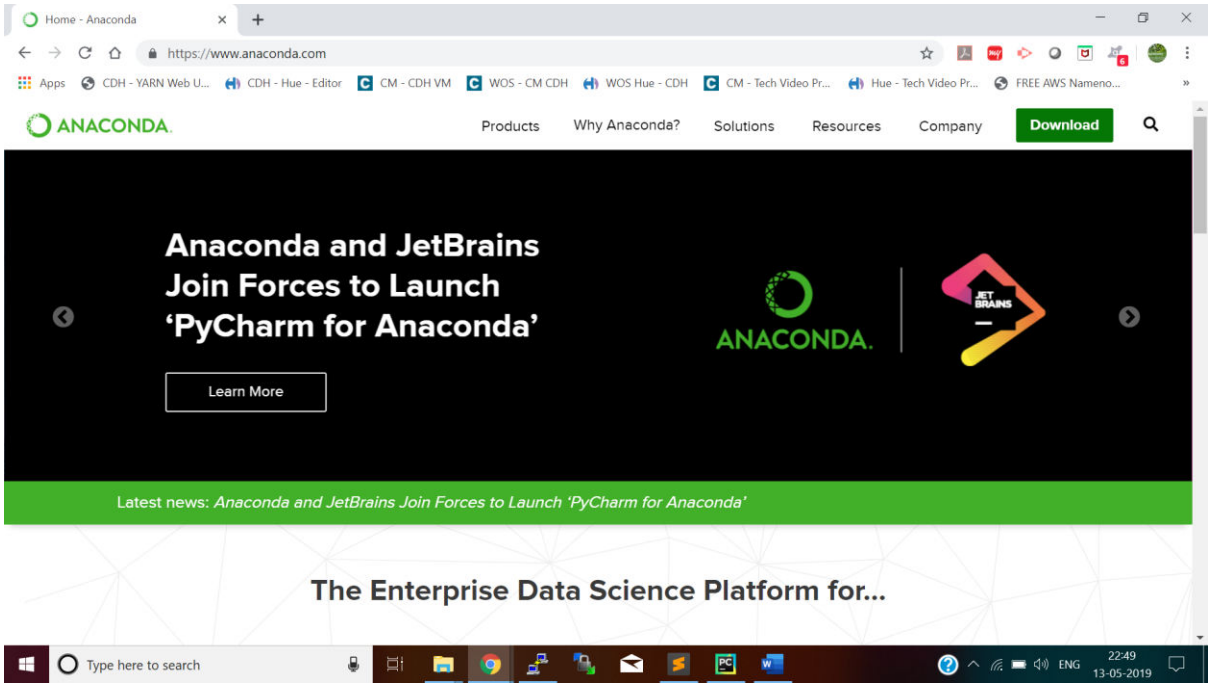
Anaconda Python Installation Guide

Table of Contents

Anaconda Python 2
Download Anaconda Python 2
Install Anaconda Python 4
Issues face while importing Python packages and resolution for the same 10
How to run Jupyter Notebook 14
Configure Spark in Jupyter 18

Anaconda Python

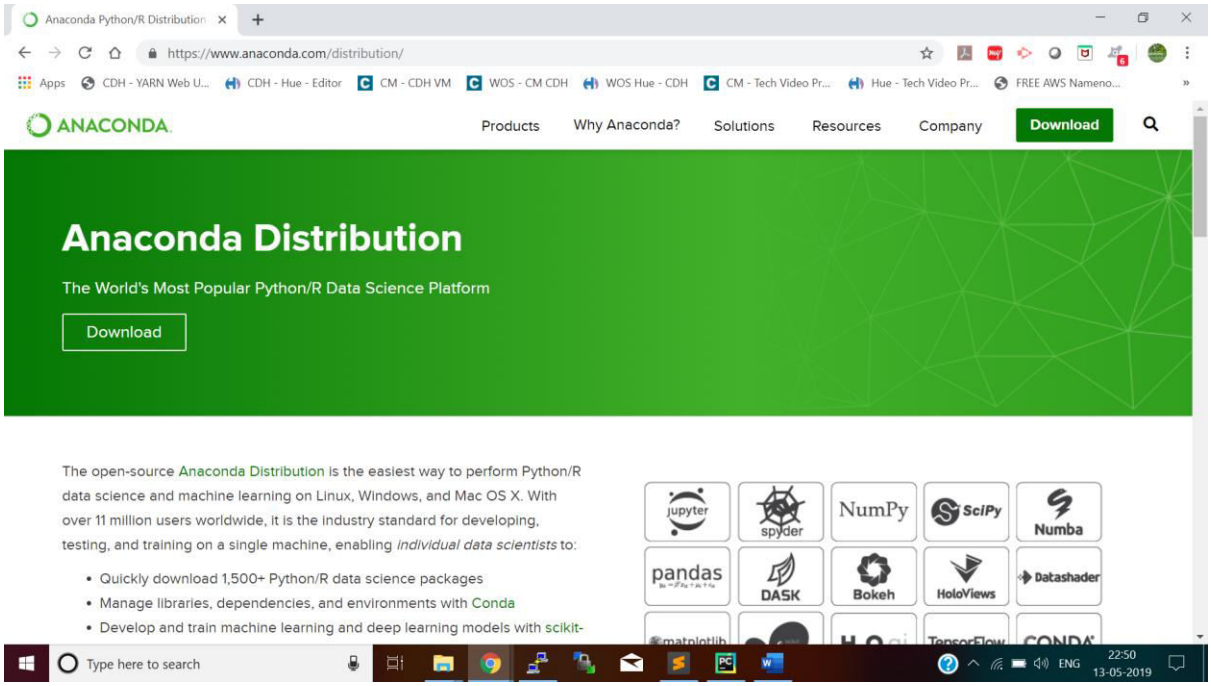
The open-source Anaconda Distribution is the easiest way to perform Python/R data science and machine learning on Linux, Windows, and Mac OS X.



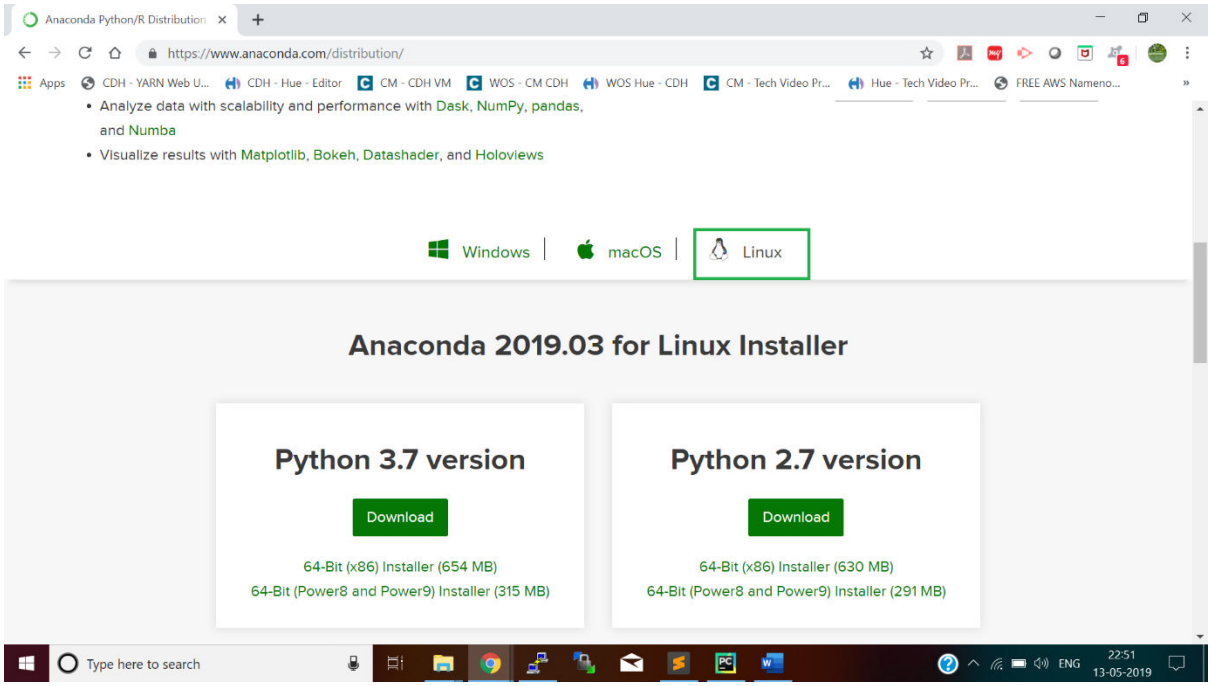
Url: https://www.anaconda.com

Download Anaconda Python

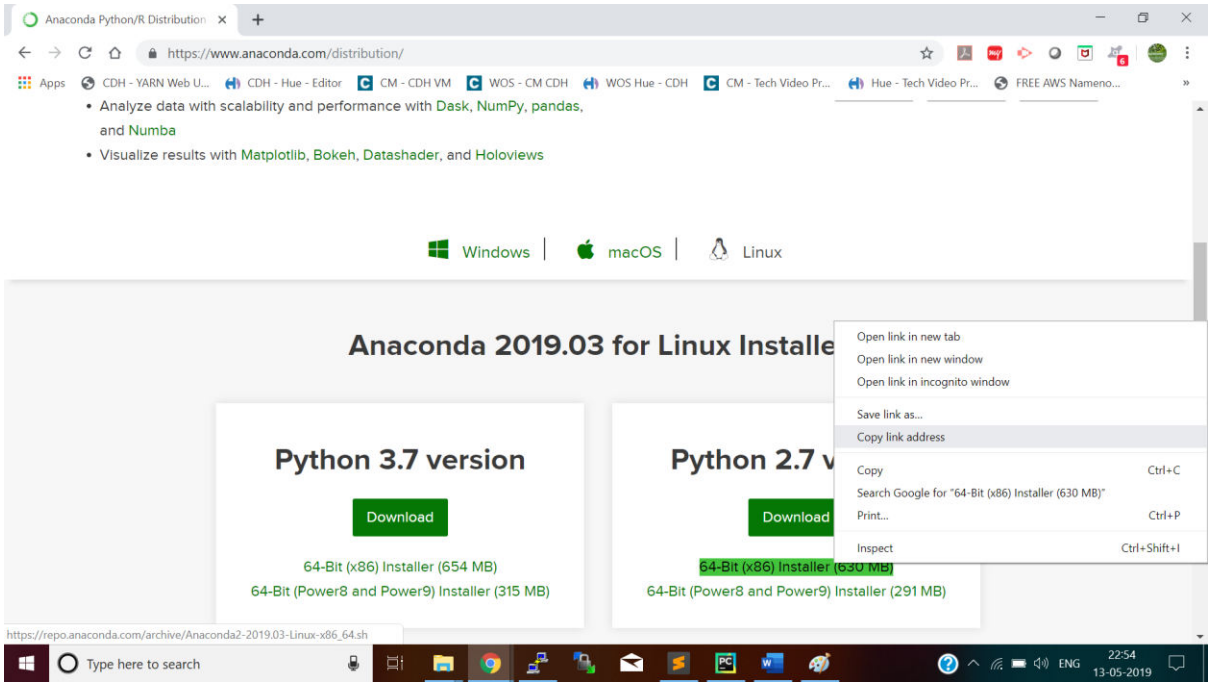
Click on “Download” button to navigate to Download page



Scroll down page



Click respective operating system tab to download Python installer, again choose Python version which you would like to install

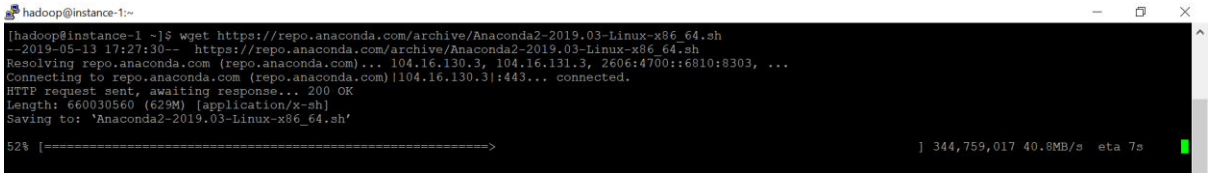


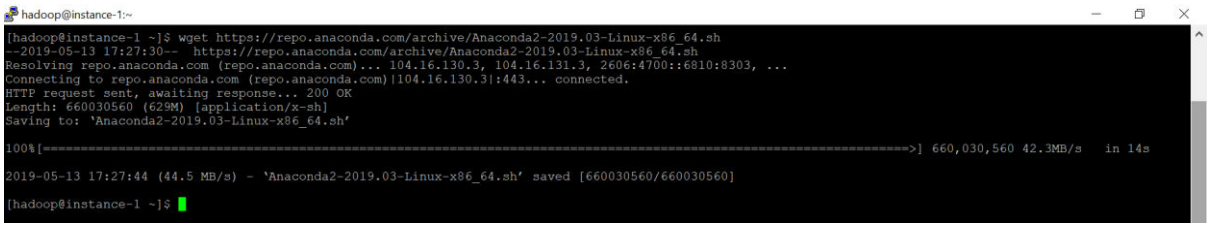
Copy the Python installer address

https://repo.anaconda.com/archive/Anaconda2-2019.03-Linux-x86_64.sh

Use wget command to download the Python installer in your CentOS command line

wget https://repo.anaconda.com/archive/Anaconda2-2019.03-Linux-x86_64.sh

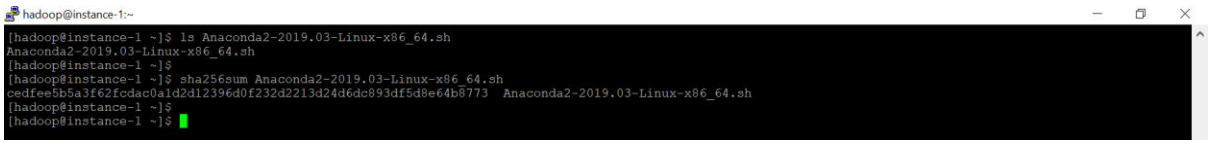




Install Anaconda Python

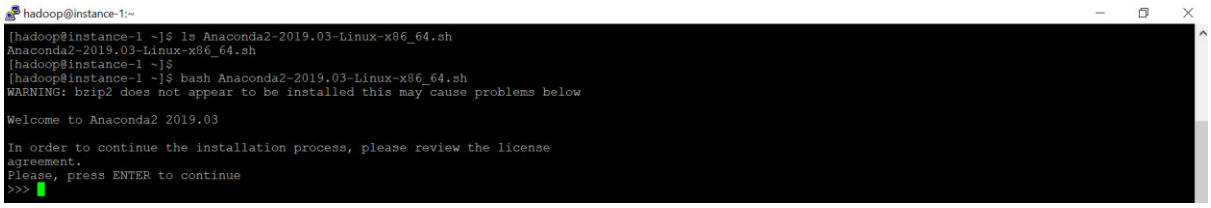
Run the below command

sha256sum Anaconda2-2019.03-Linux-x86_64.sh

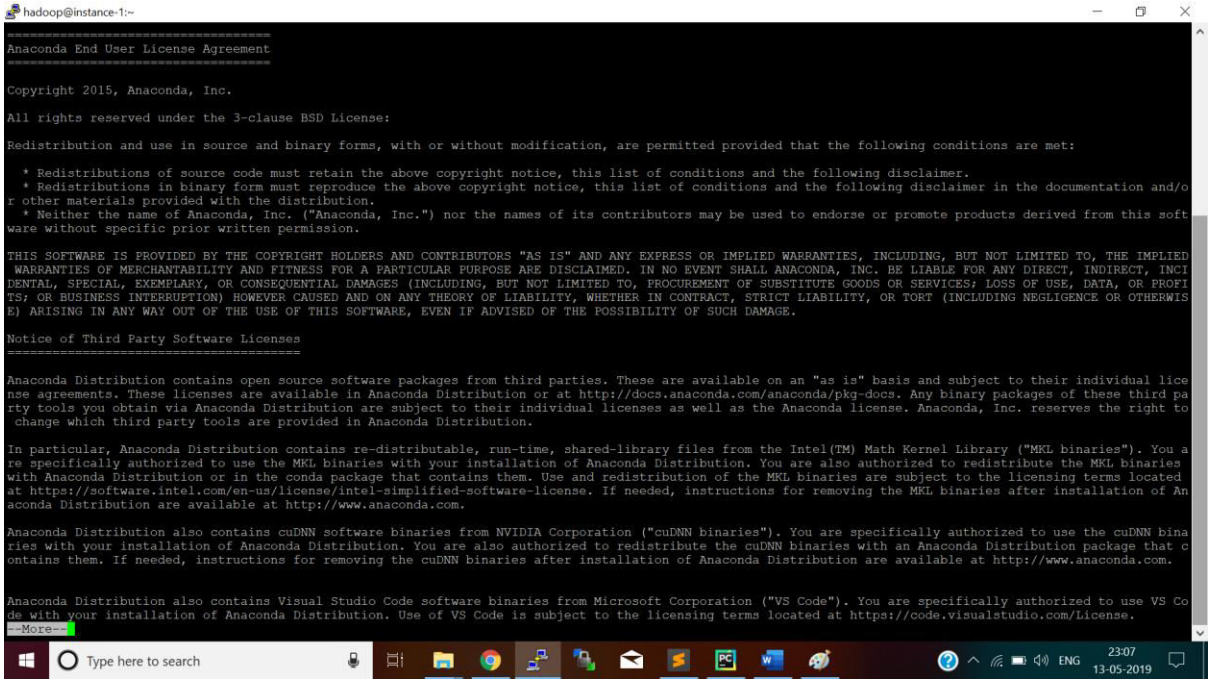


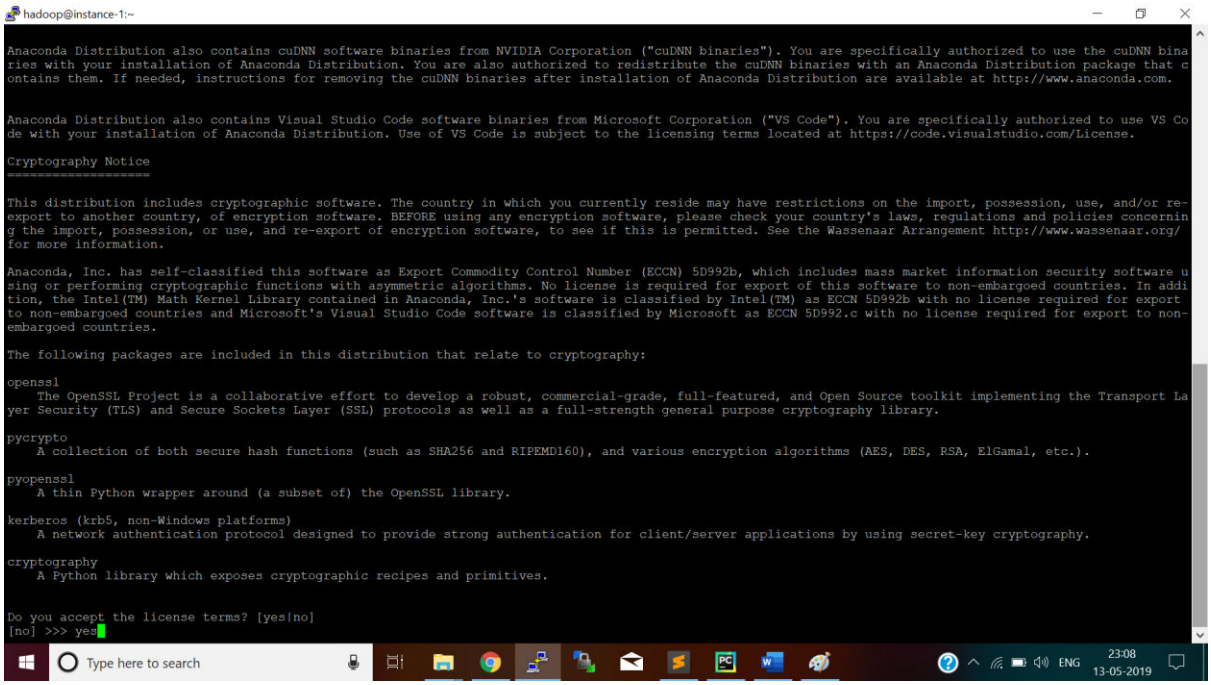
Run the below command to install Anaconda Python

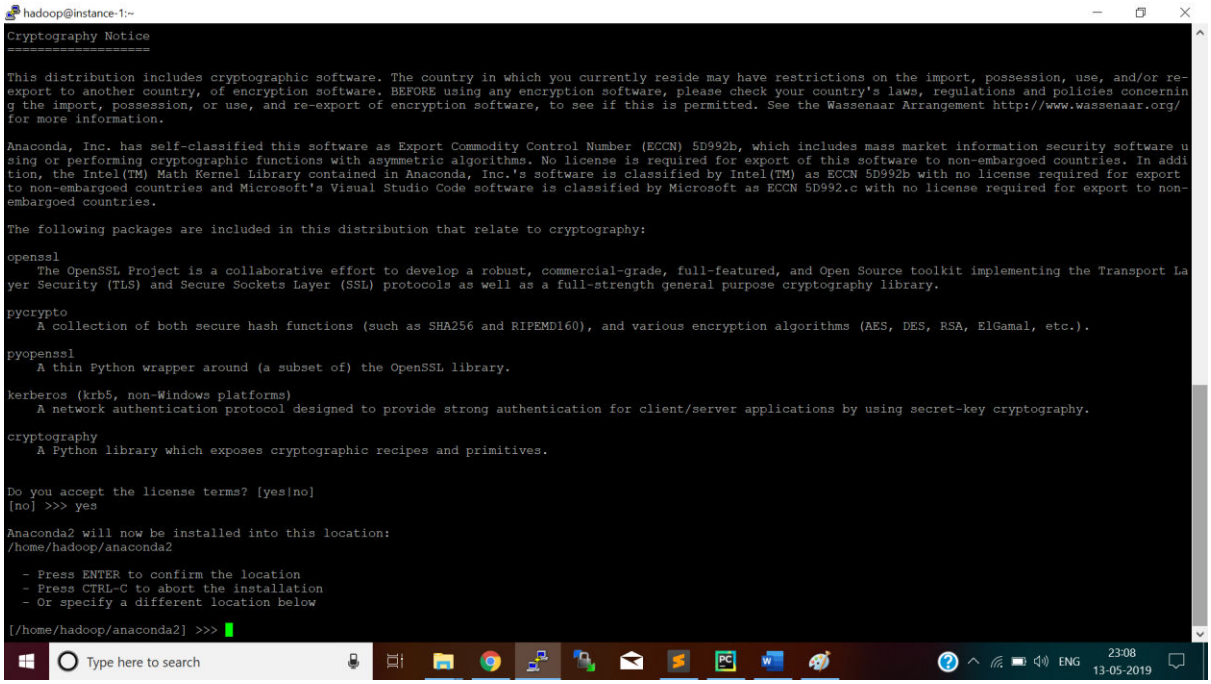
bash Anaconda2-2019.03-Linux-x86_64.sh



Preses “Enter” key







Choose your own path to install the Anaconda Python or press “Enter” to continue the installation at “/home/Hadoop/anaconda2” location (This is user’s home directory)

If you get the below error,

Anaconda2-2019.03-Linux-x86_64.sh: line 353: bunzip2: command not found

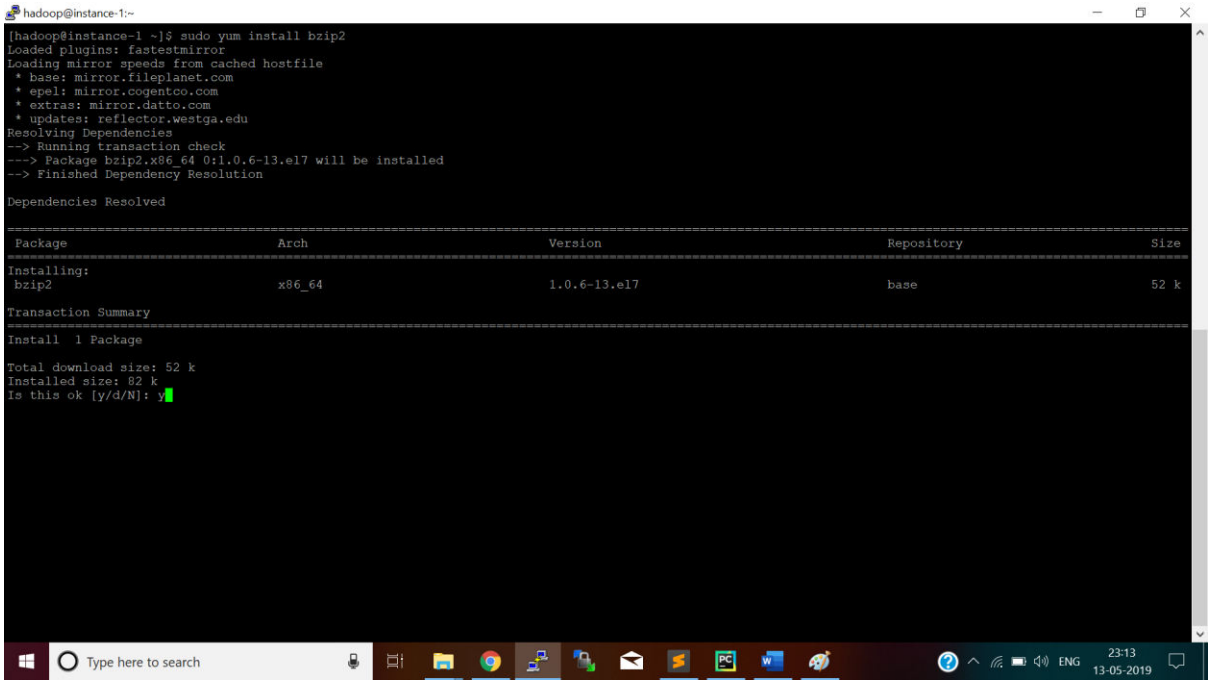
tar: This does not look like a tar archive

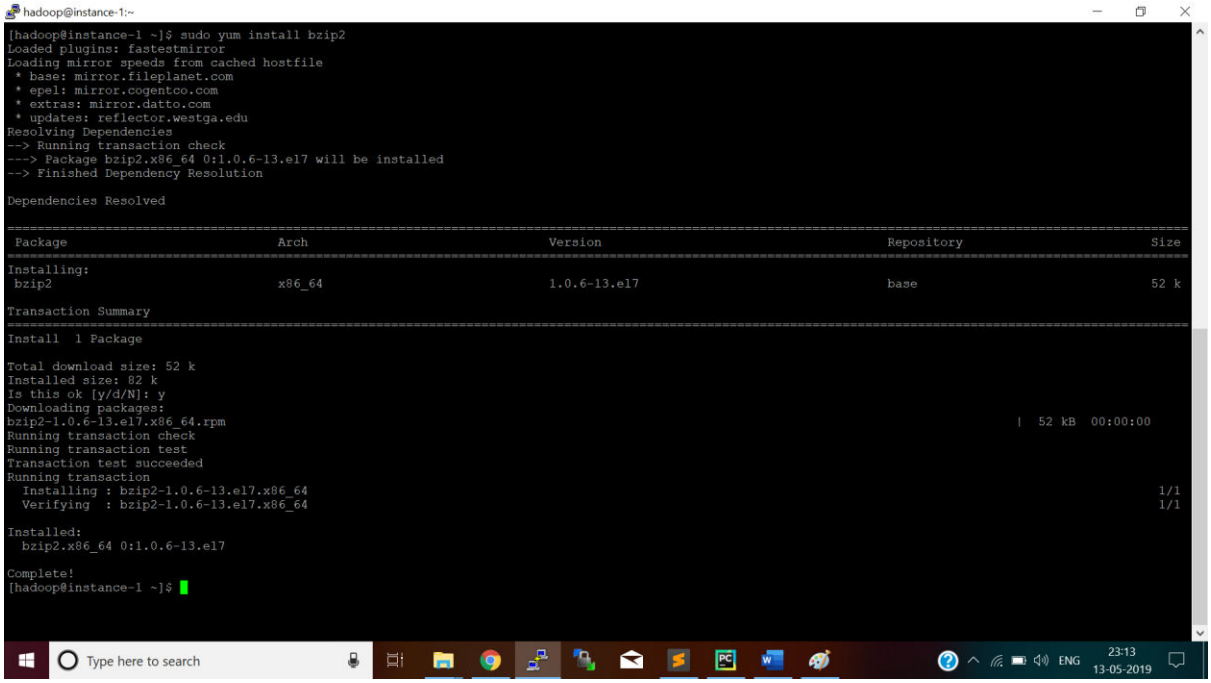
tar: Exiting with failure status due to previous errors

[hadoop@instance-1 ~]$

Then install “bzip2” using below command

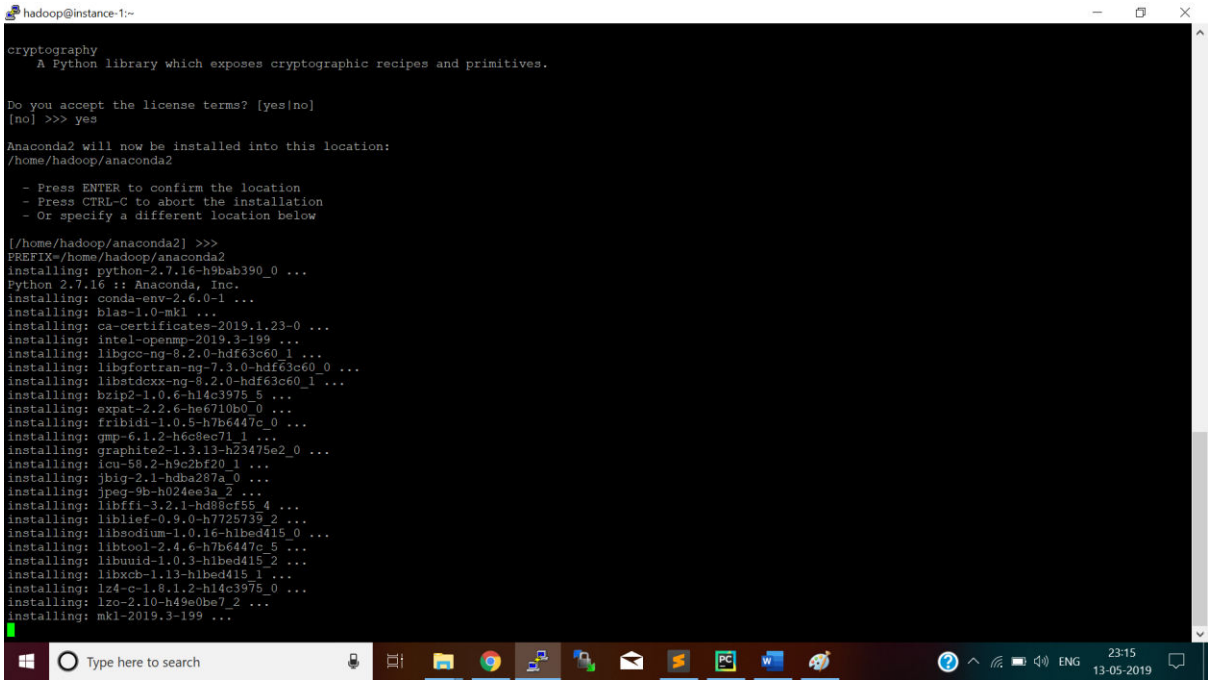
sudo yum install bzip2

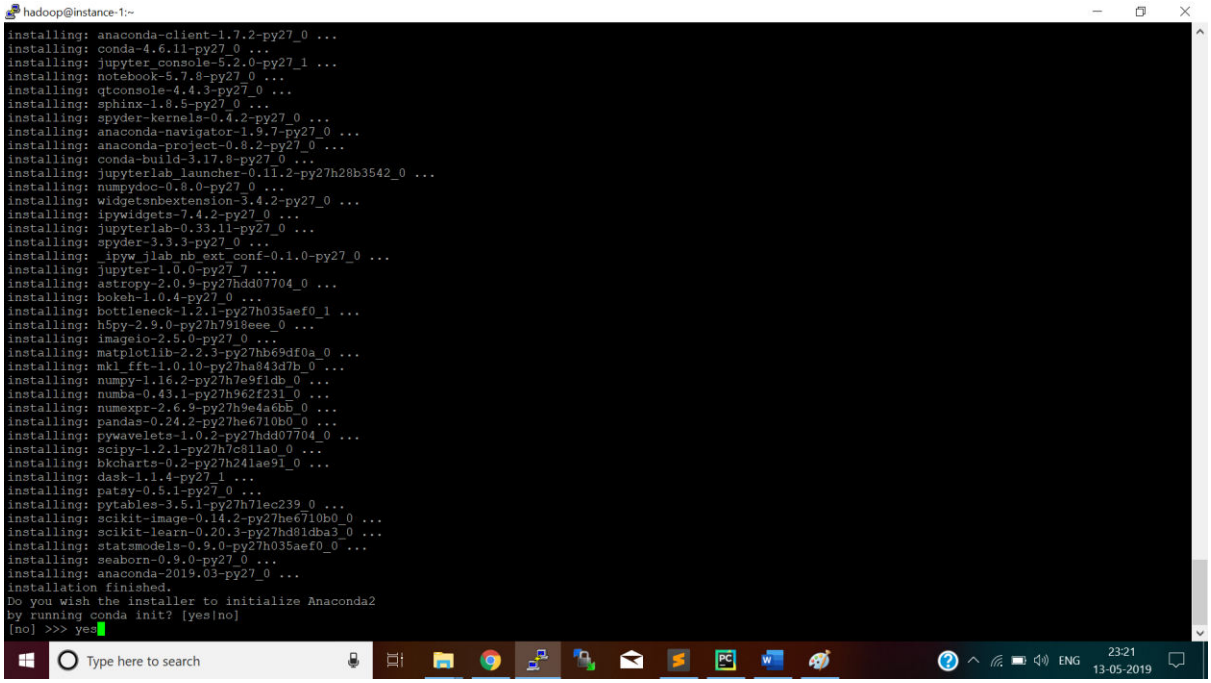


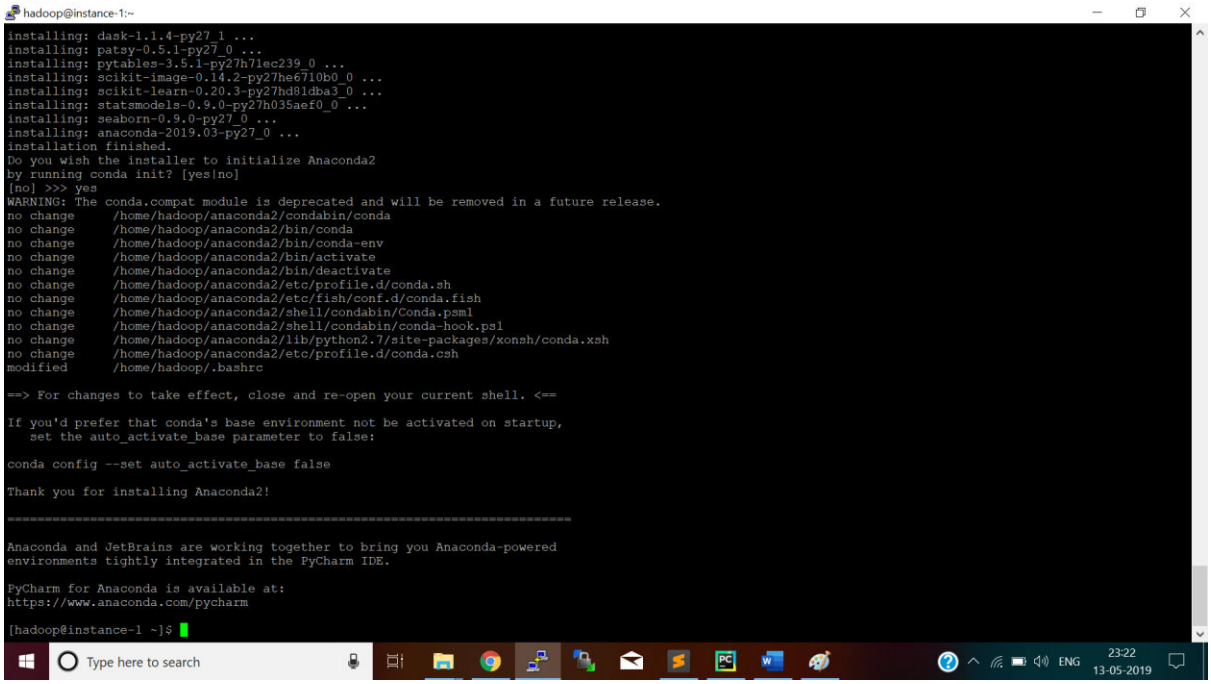


Again, run the below command

bash Anaconda2-2019.03-Linux-x86_64.sh


hadoop@instance-1:~
cryptography
A Python library which exposes cryptographic recipes and primitives.
Do you accept the license terms? [yes|no]
[no] >>> yes
Anaconda2 will now be installed into this location:
/home/hadoop/anaconda2
- Press ENTER to confirm the location
- Press CTRL-C to abort the installation
- Or specify a different location below
[/home/hadoop/anaconda2] >>>
PREFIX=/home/hadoop/anaconda2
installing: python-2.7.16-h9bab390_0 ...
Python 2.7.16 :: Anaconda, Inc.
installing: conda-env-2.6.0-1 ...
installing: blas-1.0-mkl ...
installing: ca-certificates-2019.1.23-0 ...
installing: intel-openmp-2019.3-199 ...
installing: libgcc-ng-8.2.0-hdf63c60_1 ...
installing: libgfortran-ng-7.3.0-hdf63c60_0 ...
installing: libstdcxx-ng-8.2.0-hdf63c60_1 ...
installing: bzip2-1.0.6-h14c3975_5 ...
installing: expat-2.2.6-he6710b0_0 ...
installing: fribidi-1.0.5-h7b6447c_0 ...
installing: gmp-6.1.2-h6c8ec71_1 ...
installing: graphite2-1.3.13-h23475e2_0 ...
installing: icu-58.2-h9c2bf20_1 ...
installing: jbig-2.1-hdba287a_0 ...
installing: jpeg-9b-h024ee3a_2 ...
installing: libffi-3.2.1-hd88cf55_4 ...
installing: liblief-0.9.0-h7725739_2 ...
installing: libsodium-1.0.16-h1bed415_0 ...
installing: libtool-2.4.6-h7b6447c_5 ...
installing: libuuid-1.0.3-h1bed415_2 ...
installing: libxcb-1.13-h1bed415_1 ...
installing: lz4-c-1.8.1.2-h14c3975_0 ...
installing: lzo-2.10-h49e0be7_2 ...
installing: mkl-2019.3-199 ...
Type here to search
ENG
23:15
13-05-2019


hadoop@instance-1:~
installing: anaconda-client-1.7.2-py27_0 ...
installing: conda-4.6.11-py27_0 ...
installing: jupyter_console-5.2.0-py27_1 ...
installing: notebook-5.7.8-py27_0 ...
installing: qtconsole-4.4.3-py27_0 ...
installing: sphinx-1.8.5-py27_0 ...
installing: spyder-kernels-0.4.2-py27_0 ...
installing: anaconda-navigator-1.9.7-py27_0 ...
installing: anaconda-project-0.8.2-py27_0 ...
installing: conda-build-3.17.8-py27_0 ...
installing: jupyterlab_launcher-0.11.2-py27h28b3542_0 ...
installing: numpydoc-0.8.0-py27_0 ...
installing: widgetsnbextension-3.4.2-py27_0 ...
installing: ipywidgets-7.4.2-py27_0 ...
installing: jupyterlab-0.33.11-py27_0 ...
installing: spyder-3.3.3-py27_0 ...
installing: _ipyw_jlab_nb_ext_conf-0.1.0-py27_0 ...
installing: jupyter-1.0.0-py27_7 ...
installing: astropy-2.0.9-py27hdd07704_0 ...
installing: bokeh-1.0.4-py27_0 ...
installing: bottleneck-1.2.1-py27h035aef0_1 ...
installing: h5py-2.9.0-py27h7918eee_0 ...
installing: imageio-2.5.0-py27_0 ...
installing: matplotlib-2.2.3-py27hb69df0a_0 ...
installing: mkl_fft-1.0.10-py27ha843d7b_0 ...
installing: numpy-1.16.2-py27h7e9f1db_0 ...
installing: numba-0.43.1-py27h962f231_0 ...
installing: numexpr-2.6.9-py27h9e4a6bb_0 ...
installing: pandas-0.24.2-py27he6710b0_0 ...
installing: pywavelets-1.0.2-py27hdd07704_0 ...
installing: scipy-1.2.1-py27h7c811a0_0 ...
installing: bkcharts-0.2-py27h241ae91_0 ...
installing: dask-1.1.4-py27_1 ...
installing: patsy-0.5.1-py27_0 ...
installing: pytables-3.5.1-py27h71ec239_0 ...
installing: scikit-image-0.14.2-py27he6710b0_0 ...
installing: scikit-learn-0.20.3-py27hd81dba3_0 ...
installing: statsmodels-0.9.0-py27h035aef0_0 ...
installing: seaborn-0.9.0-py27_0 ...
installing: anaconda-2019.03-py27_0 ...
installation finished.
Do you wish the installer to initialize Anaconda2
by running conda init? [yes|no]
[no] >>> yes
Type here to search
ENG
23:21
13-05-2019



Finally, Anaconda Python is installed successfully.

[hadoop@instance-1 ~]$ ls anaconda2/

bin condabin doc etc lib LICENSE.txt mkspecs pkgs qml sbin shell translations x86_64-conda_cos6-linux-gnu

compiler_compat conda-meta envs include libexec man phrasebooks plugins resources share ssl var

[hadoop@instance-1 ~]$

[hadoop@instance-1 ~]$ cd anaconda2/

[hadoop@instance-1 anaconda2]$

[hadoop@instance-1 anaconda2]$ bin/python

Python 2.7.16 |Anaconda, Inc.| (default, Mar 14 2019, 21:00:58)

[GCC 7.3.0] on linux2

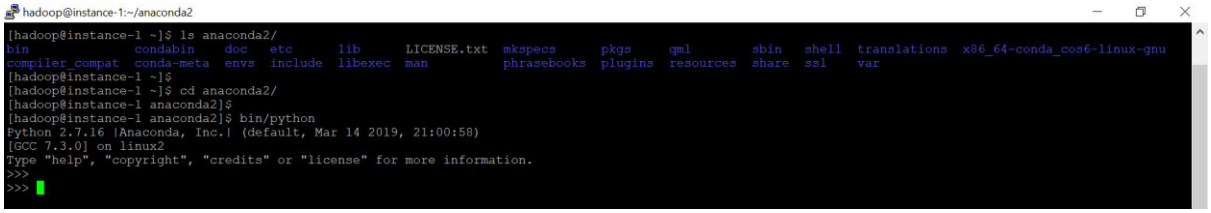
Type "help", "copyright", "credits" or "license" for more information.

>>>

>>> print("Welcome to Python World!!!")

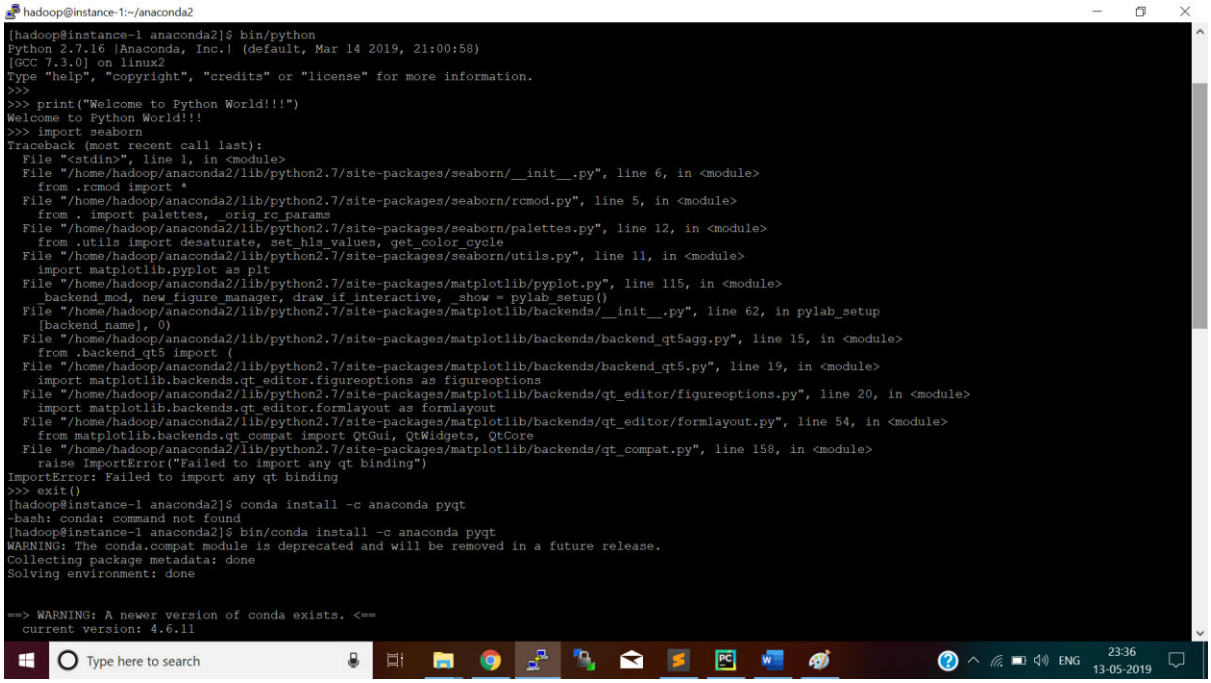
Welcome to Python World!!!

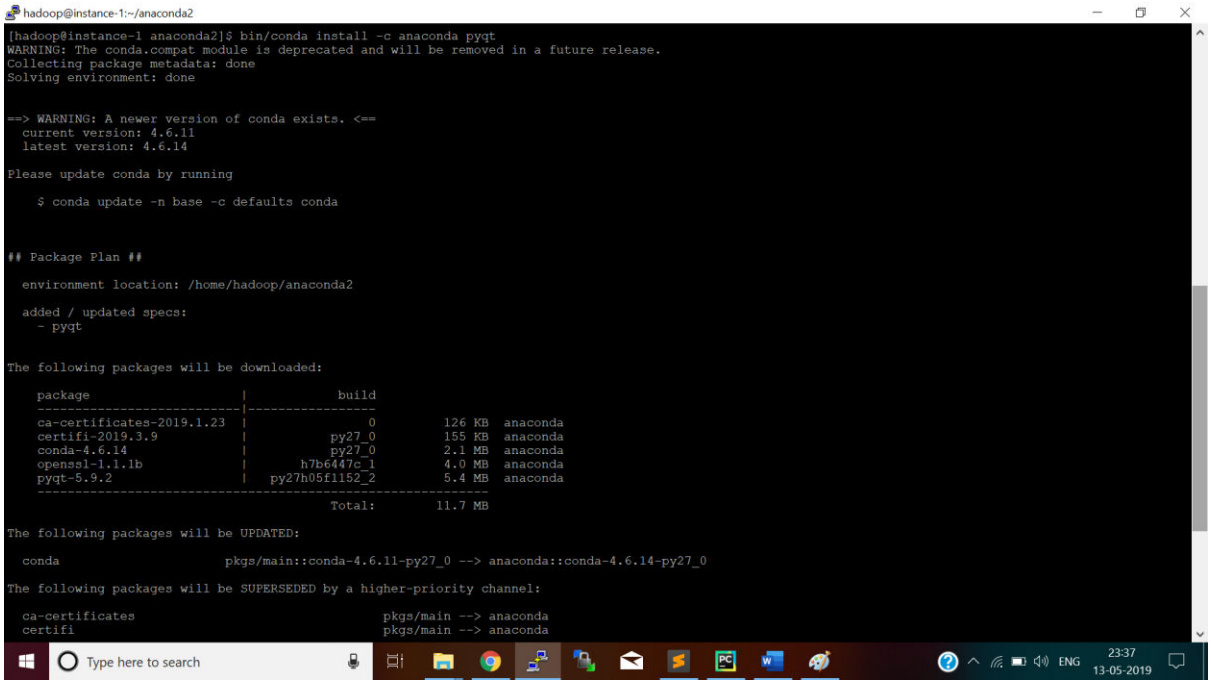
>>>

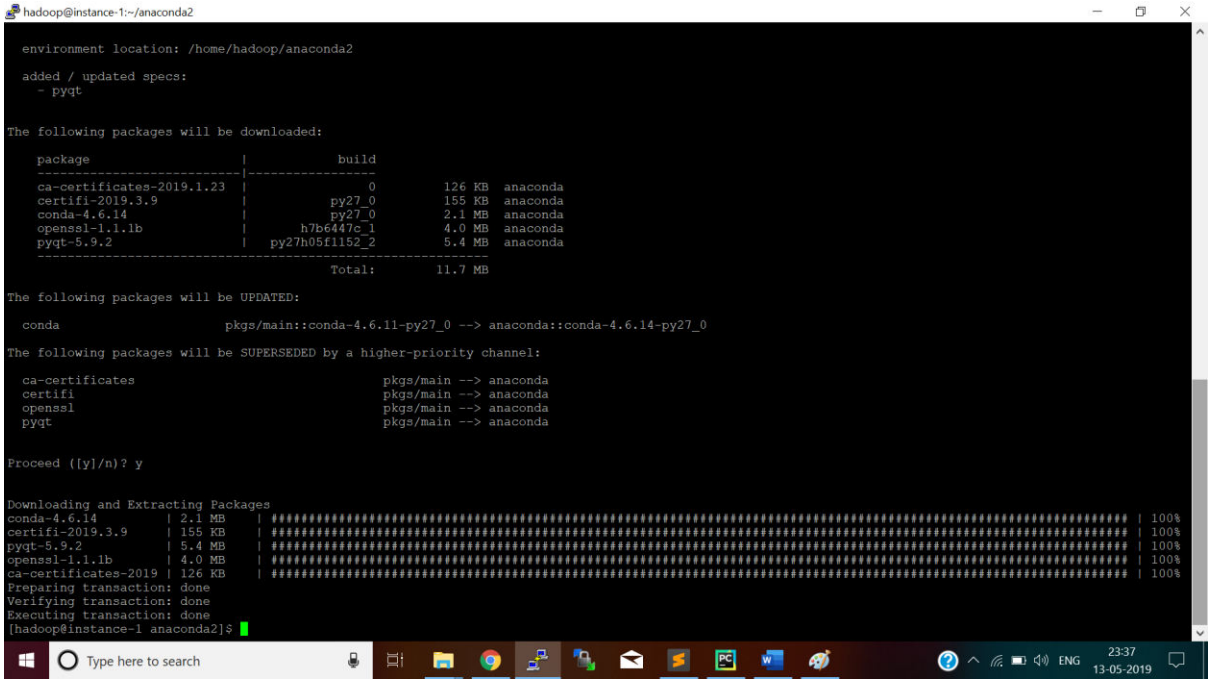


Issues face while importing Python packages and resolution for the same

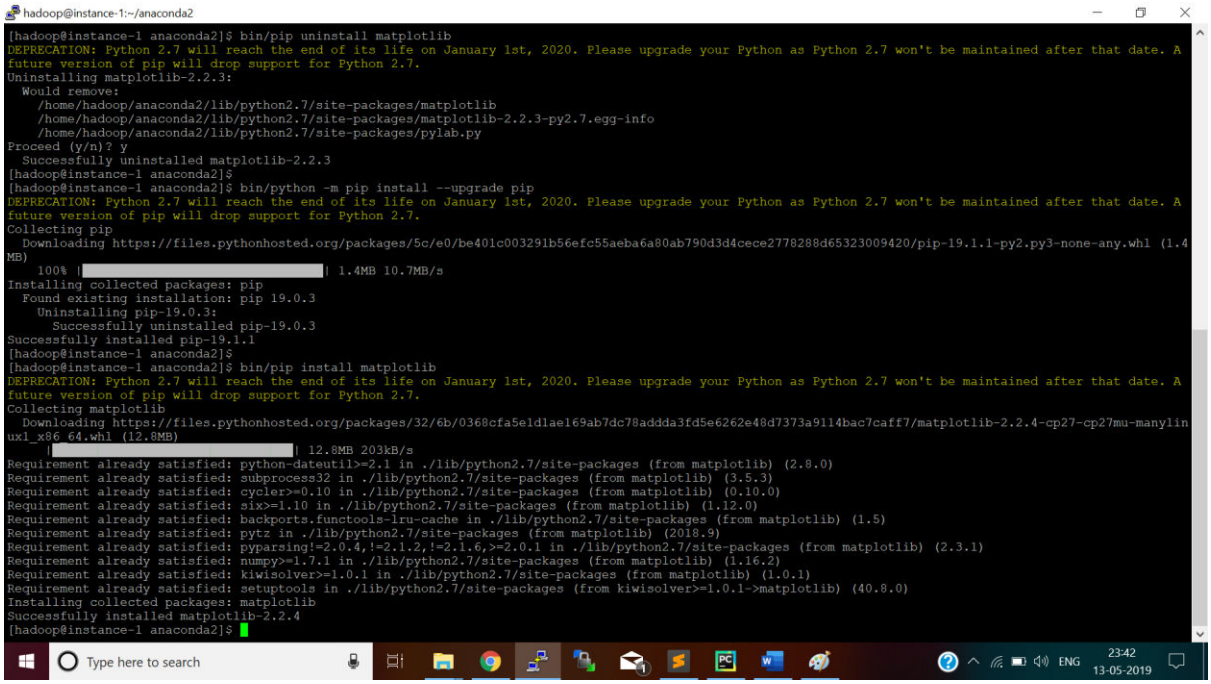
FYI.



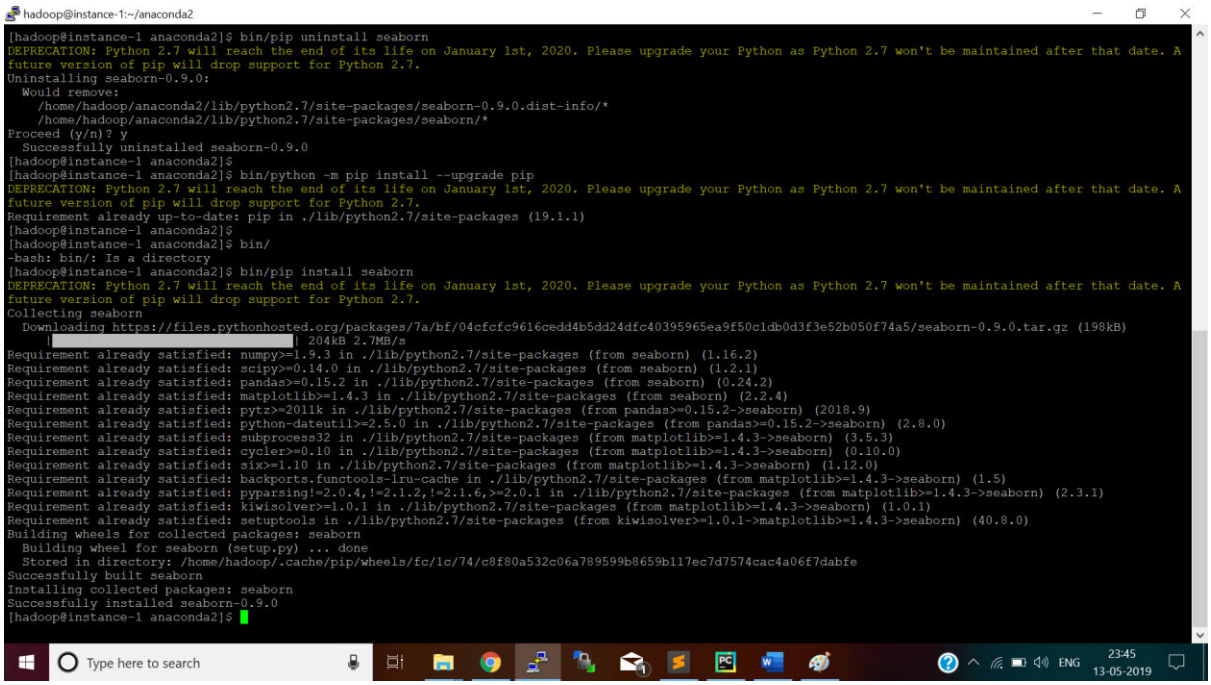

hadoop@instance-1:~/anaconda2
[hadoop@instance-1 anaconda2]$ bin/conda install -c anaconda pyqt
WARNING: The conda.compat module is deprecated and will be removed in a future release.
Collecting package metadata: done
Solving environment: done
==> WARNING: A newer version of conda exists. <==
current version: 4.6.11
latest version: 4.6.14
Please update conda by running
$ conda update -n base -c defaults conda
Package Plan
environment location: /home/hadoop/anaconda2
added / updated specs:
- pyqt
The following packages will be downloaded:
package | build
ca-certificates-2019.1.23 | 0 126 KB anaconda
certifi-2019.3.9 | py27_0 155 KB anaconda
conda-4.6.14 | py27_0 2.1 MB anaconda
openssl-1.1.1b | h7b6447c_1 4.0 MB anaconda
pyqt-5.9.2 | py27h05f1152_2 5.4 MB anaconda
Total: 11.7 MB
The following packages will be UPDATED:
conda pkgs/main::conda-4.6.11-py27_0 --> anaconda::conda-4.6.14-py27_0
The following packages will be SUPERSEDED by a higher-priority channel:
ca-certificates pkgs/main --> anaconda
certifi pkgs/main --> anaconda
Type here to search
ENG
23:37
13-05-2019


hadoop@instance-1:~/anaconda2
environment location: /home/hadoop/anaconda2
added / updated specs:
- pyqt
The following packages will be downloaded:
package | build
ca-certificates-2019.1.23 | 0 126 KB anaconda
certifi-2019.3.9 | py27_0 155 KB anaconda
conda-4.6.14 | py27_0 2.1 MB anaconda
openssl-1.1.1b | h7b6447c_1 4.0 MB anaconda
pyqt-5.9.2 | py27h05f1152_2 5.4 MB anaconda
Total: 11.7 MB
The following packages will be UPDATED:
conda pkgs/main::conda-4.6.11-py27_0 --> anaconda::conda-4.6.14-py27_0
The following packages will be SUPERSEDED by a higher-priority channel:
ca-certificates pkgs/main --> anaconda
certifi pkgs/main --> anaconda
openssl pkgs/main --> anaconda
pyqt pkgs/main --> anaconda
Proceed ([y]/n)? y
Downloading and Extracting Packages
conda-4.6.14 | 2.1 MB | 100%
certifi-2019.3.9 | 155 KB | 100%
pyqt-5.9.2 | 5.4 MB | 100%
openssl-1.1.1b | 4.0 MB | 100%
ca-certificates-2019 | 126 KB | 100%
Preparing transaction: done
Verifying transaction: done
Executing transaction: done
[hadoop@instance-1 anaconda2]$
Type here to search
ENG
23:37
13-05-2019

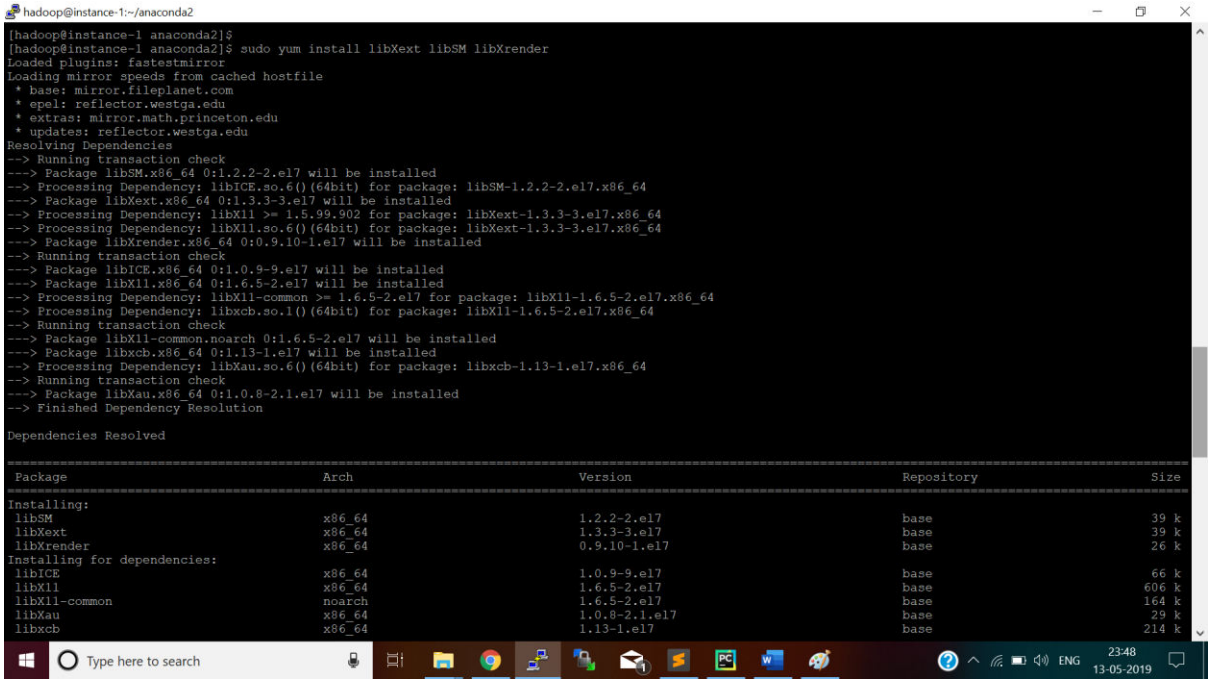
FYI.

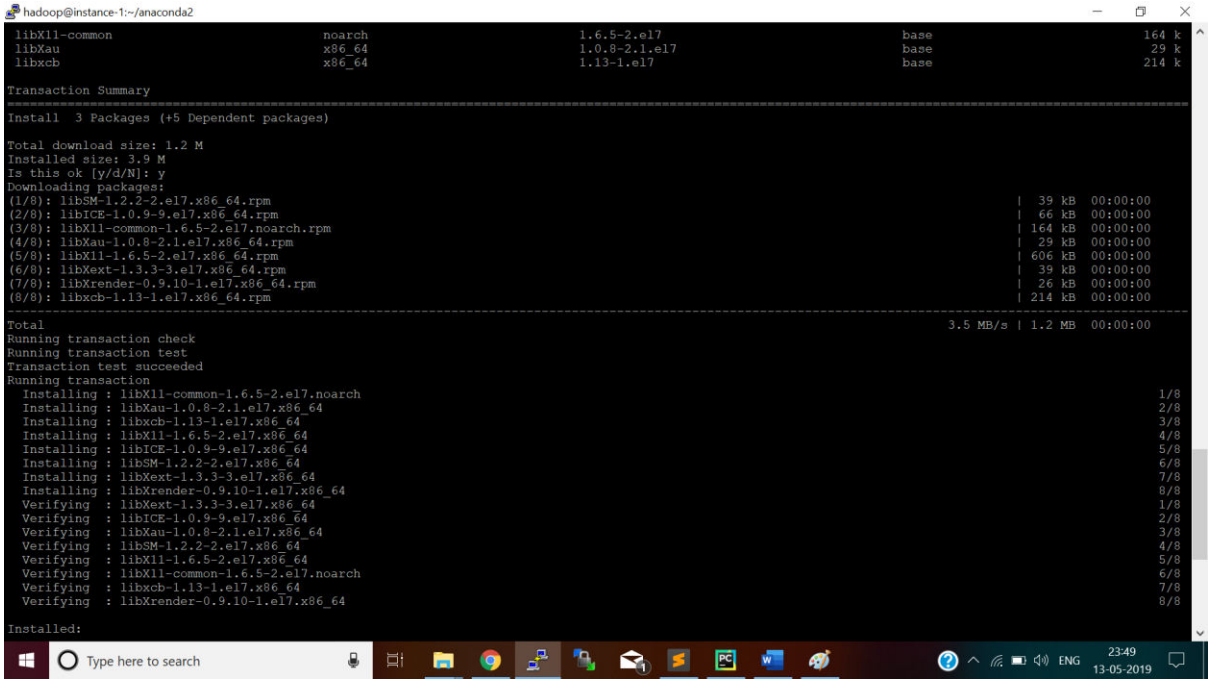

hadoop@instance-1:~/anaconda2
[hadoop@instance-1 anaconda2]$ bin/pip uninstall matplotlib
DEPRECATION: Python 2.7 will reach the end of its life on January 1st, 2020. Please upgrade your Python as Python 2.7 won't be maintained after that date. A future version of pip will drop support for Python 2.7.
Uninstalling matplotlib-2.2.3:
Would remove:
/home/hadoop/anaconda2/lib/python2.7/site-packages/matplotlib
/home/hadoop/anaconda2/lib/python2.7/site-packages/matplotlib-2.2.3-py2.7.egg-info
/home/hadoop/anaconda2/lib/python2.7/site-packages/pylab.py
Proceed (y/n)? y
Successfully uninstalled matplotlib-2.2.3
[hadoop@instance-1 anaconda2]$
[hadoop@instance-1 anaconda2]$ bin/python -m pip install --upgrade pip
DEPRECATION: Python 2.7 will reach the end of its life on January 1st, 2020. Please upgrade your Python as Python 2.7 won't be maintained after that date. A future version of pip will drop support for Python 2.7.
Collecting pip
Downloading https://files.pythonhosted.org/packages/5c/e0/be401c003291b56efc55aeba6a80ab790d3d4cece2778288d65323009420/pip-19.1.1-py2.py3-none-any.whl (1.4MB)
100% | 1.4MB 10.7MB/s
Installing collected packages: pip
Found existing installation: pip 19.0.3
Uninstalling pip-19.0.3:
Successfully uninstalled pip-19.0.3
Successfully installed pip-19.1.1
[hadoop@instance-1 anaconda2]$
[hadoop@instance-1 anaconda2]$ bin/pip install matplotlib
DEPRECATION: Python 2.7 will reach the end of its life on January 1st, 2020. Please upgrade your Python as Python 2.7 won't be maintained after that date. A future version of pip will drop support for Python 2.7.
Collecting matplotlib
Downloading https://files.pythonhosted.org/packages/32/6b/0368cfa5e1d1ae169ab7dc78addda3fd5e6262e48d7373a9114bac7caff7/matplotlib-2.2.4-cp27-cp27mu-manylinux1_x86_64.whl (12.8MB)
| 12.8MB 203kB/s
Requirement already satisfied: python-dateutil>=2.1 in ./lib/python2.7/site-packages (from matplotlib) (2.8.0)
Requirement already satisfied: subprocess32 in ./lib/python2.7/site-packages (from matplotlib) (3.5.3)
Requirement already satisfied: cycler>=0.10 in ./lib/python2.7/site-packages (from matplotlib) (0.10.0)
Requirement already satisfied: six>=1.10 in ./lib/python2.7/site-packages (from matplotlib) (1.12.0)
Requirement already satisfied: backports.functools-lru-cache in ./lib/python2.7/site-packages (from matplotlib) (1.5)
Requirement already satisfied: pytz in ./lib/python2.7/site-packages (from matplotlib) (2018.9)
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in ./lib/python2.7/site-packages (from matplotlib) (2.3.1)
Requirement already satisfied: numpy>=1.7.1 in ./lib/python2.7/site-packages (from matplotlib) (1.16.2)
Requirement already satisfied: kiwisolver>=1.0.1 in ./lib/python2.7/site-packages (from matplotlib) (1.0.1)
Requirement already satisfied: setuptools in ./lib/python2.7/site-packages (from kiwisolver>=1.0.1->matplotlib) (40.8.0)
Installing collected packages: matplotlib
Successfully installed matplotlib-2.2.4
[hadoop@instance-1 anaconda2]$
Type here to search
ENG
23:42
13-05-2019

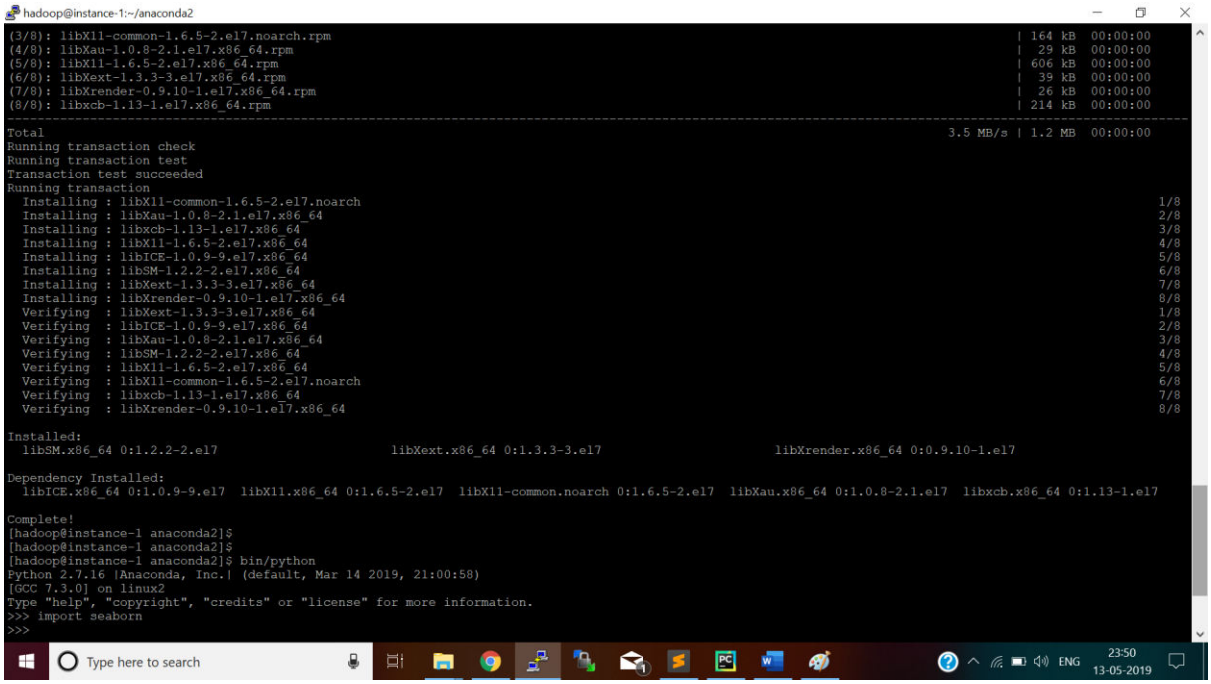
FYI.


hadoop@instance-1:~/anaconda2
[hadoop@instance-1 anaconda2]$ bin/pip uninstall seaborn
DEPRECATION: Python 2.7 will reach the end of its life on January 1st, 2020. Please upgrade your Python as Python 2.7 won't be maintained after that date. A future version of pip will drop support for Python 2.7.
Uninstalling seaborn-0.9.0:
Would remove:
/home/hadoop/anaconda2/lib/python2.7/site-packages/seaborn-0.9.0.dist-info/*
/home/hadoop/anaconda2/lib/python2.7/site-packages/seaborn/*
Proceed (y/n)? y
Successfully uninstalled seaborn-0.9.0
[hadoop@instance-1 anaconda2]$
[hadoop@instance-1 anaconda2]$ bin/python -m pip install --upgrade pip
DEPRECATION: Python 2.7 will reach the end of its life on January 1st, 2020. Please upgrade your Python as Python 2.7 won't be maintained after that date. A future version of pip will drop support for Python 2.7.
Requirement already up-to-date: pip in ./lib/python2.7/site-packages (19.1.1)
[hadoop@instance-1 anaconda2]$
[hadoop@instance-1 anaconda2]$ bin/
-bash: bin/: Is a directory
[hadoop@instance-1 anaconda2]$ bin/pip install seaborn
DEPRECATION: Python 2.7 will reach the end of its life on January 1st, 2020. Please upgrade your Python as Python 2.7 won't be maintained after that date. A future version of pip will drop support for Python 2.7.
Collecting seaborn
Downloading https://files.pythonhosted.org/packages/7a/bf/04cfcfc9616cedd4b5dd24dfc40395965ea9f50c1db0d3f3e52b050f74a5/seaborn-0.9.0.tar.gz (198kB)
| 204kB 2.7MB/s
Requirement already satisfied: numpy>=1.9.3 in ./lib/python2.7/site-packages (from seaborn) (1.16.2)
Requirement already satisfied: scipy>=0.14.0 in ./lib/python2.7/site-packages (from seaborn) (1.2.1)
Requirement already satisfied: pandas>=0.15.2 in ./lib/python2.7/site-packages (from seaborn) (0.24.2)
Requirement already satisfied: matplotlib>=1.4.3 in ./lib/python2.7/site-packages (from seaborn) (2.2.4)
Requirement already satisfied: pytz>=2011k in ./lib/python2.7/site-packages (from pandas>=0.15.2->seaborn) (2018.9)
Requirement already satisfied: python-dateutil>=2.5.0 in ./lib/python2.7/site-packages (from pandas>=0.15.2->seaborn) (2.8.0)
Requirement already satisfied: subprocess32 in ./lib/python2.7/site-packages (from matplotlib>=1.4.3->seaborn) (3.5.3)
Requirement already satisfied: cycler>=0.10 in ./lib/python2.7/site-packages (from matplotlib>=1.4.3->seaborn) (0.10.0)
Requirement already satisfied: six>=1.10 in ./lib/python2.7/site-packages (from matplotlib>=1.4.3->seaborn) (1.12.0)
Requirement already satisfied: backports.functools-lru-cache in ./lib/python2.7/site-packages (from matplotlib>=1.4.3->seaborn) (1.5)
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in ./lib/python2.7/site-packages (from matplotlib>=1.4.3->seaborn) (2.3.1)
Requirement already satisfied: kiwisolver>=1.0.1 in ./lib/python2.7/site-packages (from matplotlib>=1.4.3->seaborn) (1.0.1)
Requirement already satisfied: setuptools in ./lib/python2.7/site-packages (from kiwisolver>=1.0.1->matplotlib>=1.4.3->seaborn) (40.8.0)
Building wheels for collected packages: seaborn
Building wheel for seaborn (setup.py) ... done
Stored in directory: /home/hadoop/.cache/pip/wheels/fc/1c/74/c8f80a532c06a789599b8659b117ec7d7574cac4a06f7dabfe
Successfully built seaborn
Installing collected packages: seaborn
Successfully installed seaborn-0.9.0
[hadoop@instance-1 anaconda2]$
Type here to search
ENG
23:45
13-05-2019

FYI.


hadoop@instance-1:~/anaconda2
[hadoop@instance-1 anaconda2]$
[hadoop@instance-1 anaconda2]$ sudo yum install libXext libSM libXrender
Loaded plugins: fastestmirror
Loading mirror speeds from cached hostfile
* base: mirror.fileplanet.com
* epel: reflector.westga.edu
* extras: mirror.math.princeton.edu
* updates: reflector.westga.edu
Resolving Dependencies
--> Running transaction check
---> Package libSM.x86_64 0:1.2.2-2.el7 will be installed
--> Processing Dependency: libICE.so.6()(64bit) for package: libSM-1.2.2-2.el7.x86_64
---> Package libXext.x86_64 0:1.3.3-3.el7 will be installed
--> Processing Dependency: libX11 >= 1.5.99.902 for package: libXext-1.3.3-3.el7.x86_64
--> Processing Dependency: libX11.so.6()(64bit) for package: libXext-1.3.3-3.el7.x86_64
---> Package libXrender.x86_64 0:0.9.10-1.el7 will be installed
--> Running transaction check
---> Package libICE.x86_64 0:1.0.9-9.el7 will be installed
---> Package libX11.x86_64 0:1.6.5-2.el7 will be installed
--> Processing Dependency: libX11-common >= 1.6.5-2.el7 for package: libX11-1.6.5-2.el7.x86_64
--> Processing Dependency: libxcb.so.1()(64bit) for package: libX11-1.6.5-2.el7.x86_64
--> Running transaction check
---> Package libX11-common.noarch 0:1.6.5-2.el7 will be installed
---> Package libxcb.x86_64 0:1.13-1.el7 will be installed
--> Processing Dependency: libXau.so.6()(64bit) for package: libxcb-1.13-1.el7.x86_64
--> Running transaction check
---> Package libXau.x86_64 0:1.0.8-2.1.el7 will be installed
--> Finished Dependency Resolution
Dependencies Resolved
Package Arch Version Repository Size
Installing:
libSM x86_64 1.2.2-2.el7 base 39 k
libXext x86_64 1.3.3-3.el7 base 39 k
libXrender x86_64 0.9.10-1.el7 base 26 k
Installing for dependencies:
libICE x86_64 1.0.9-9.el7 base 66 k
libX11 x86_64 1.6.5-2.el7 base 606 k
libX11-common noarch 1.6.5-2.el7 base 164 k
libXau x86_64 1.0.8-2.1.el7 base 29 k
libxcb x86_64 1.13-1.el7 base 214 k
Type here to search
ENG
23:48
13-05-2019


hadoop@instance-1:~/anaconda2
libX11-common noarch 1.6.5-2.el7 base 164 k
libXau x86_64 1.0.8-2.1.el7 base 29 k
libxcb x86_64 1.13-1.el7 base 214 k
Transaction Summary
Install 3 Packages (+5 Dependent packages)
Total download size: 1.2 M
Installed size: 3.9 M
Is this ok [y/d/N]: y
Downloading packages:
(1/8): libSM-1.2.2-2.el7.x86_64.rpm | 39 kB 00:00:00
(2/8): libICE-1.0.9-9.el7.x86_64.rpm | 66 kB 00:00:00
(3/8): libX11-common-1.6.5-2.el7.noarch.rpm | 164 kB 00:00:00
(4/8): libXau-1.0.8-2.1.el7.x86_64.rpm | 29 kB 00:00:00
(5/8): libX11-1.6.5-2.el7.x86_64.rpm | 606 kB 00:00:00
(6/8): libXext-1.3.3-3.el7.x86_64.rpm | 39 kB 00:00:00
(7/8): libXrender-0.9.10-1.el7.x86_64.rpm | 26 kB 00:00:00
(8/8): libxcb-1.13-1.el7.x86_64.rpm | 214 kB 00:00:00
Total 3.5 MB/s | 1.2 MB 00:00:00
Running transaction check
Running transaction test
Transaction test succeeded
Running transaction
Installing : libX11-common-1.6.5-2.el7.noarch 1/8
Installing : libXau-1.0.8-2.1.el7.x86_64 2/8
Installing : libxcb-1.13-1.el7.x86_64 3/8
Installing : libX11-1.6.5-2.el7.x86_64 4/8
Installing : libICE-1.0.9-9.el7.x86_64 5/8
Installing : libSM-1.2.2-2.el7.x86_64 6/8
Installing : libXext-1.3.3-3.el7.x86_64 7/8
Installing : libXrender-0.9.10-1.el7.x86_64 8/8
Verifying : libXext-1.3.3-3.el7.x86_64 1/8
Verifying : libICE-1.0.9-9.el7.x86_64 2/8
Verifying : libXau-1.0.8-2.1.el7.x86_64 3/8
Verifying : libSM-1.2.2-2.el7.x86_64 4/8
Verifying : libX11-1.6.5-2.el7.x86_64 5/8
Verifying : libX11-common-1.6.5-2.el7.noarch 6/8
Verifying : libxcb-1.13-1.el7.x86_64 7/8
Verifying : libXrender-0.9.10-1.el7.x86_64 8/8
Installed:
Type here to search
ENG
23:49
13-05-2019



How to run Jupyter Notebook

Use below command to run the jupyter notebook

bin/jupyter notebook --port 5566

[hadoop@instance-1 anaconda2]$ bin/jupyter notebook --port 5566

[I 18:25:03.344 NotebookApp] Writing notebook server cookie secret to /run/user/1000/jupyter/notebook_cookie_secret

[I 18:25:03.686 NotebookApp] JupyterLab extension loaded from /home/hadoop/anaconda2/lib/python2.7/site-packages/jupyterlab

[I 18:25:03.687 NotebookApp] JupyterLab application directory is /home/hadoop/anaconda2/share/jupyter/lab

[I 18:25:03.695 NotebookApp] Serving notebooks from local directory: /home/hadoop/anaconda2

[I 18:25:03.695 NotebookApp] The Jupyter Notebook is running at:

[I 18:25:03.695 NotebookApp] http://localhost:5566/?token=2debfc44803f84e67ed4f717a0cc0fbd15fc79b3fd5fc30d

[I 18:25:03.696 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).

[W 18:25:03.702 NotebookApp] No web browser found: could not locate runnable browser.

[C 18:25:03.703 NotebookApp]

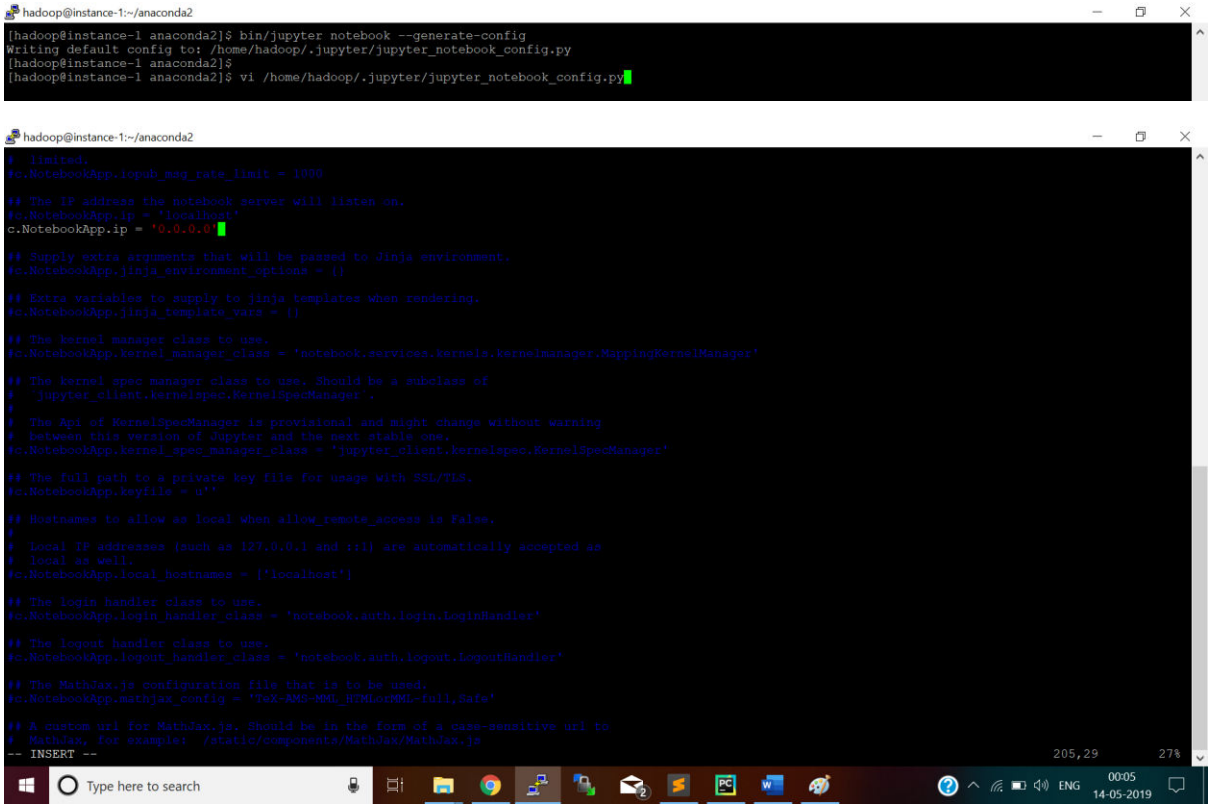
To access the notebook, open this file in a browser:

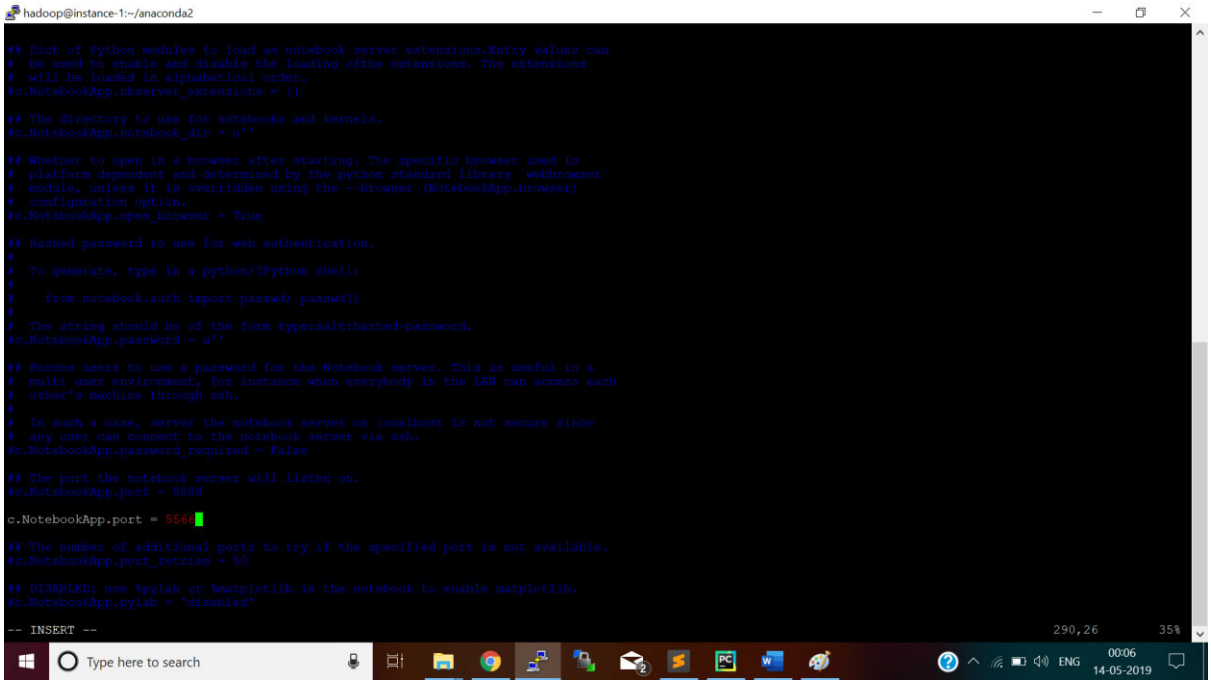
file:///run/user/1000/jupyter/nbserver-15017-open.html

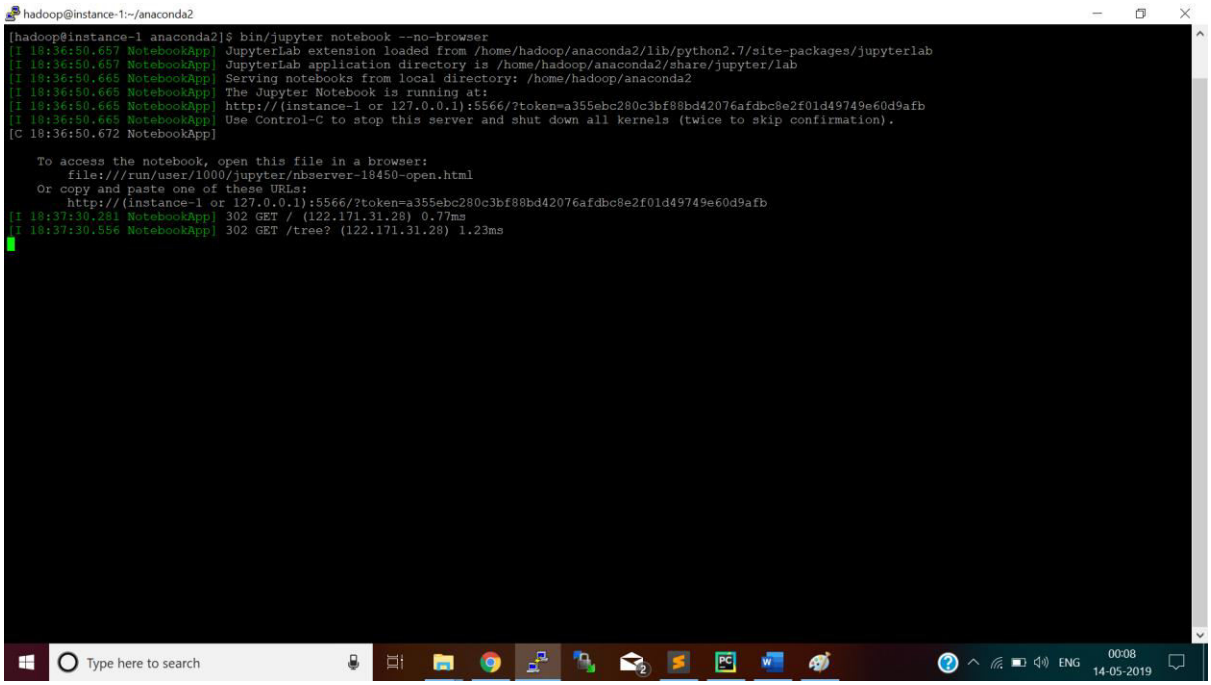
Or copy and paste one of these URLs:

http://localhost:5566/?token=2debfc44803f84e67ed4f717a0cc0fbd15fc79b3fd5fc30d

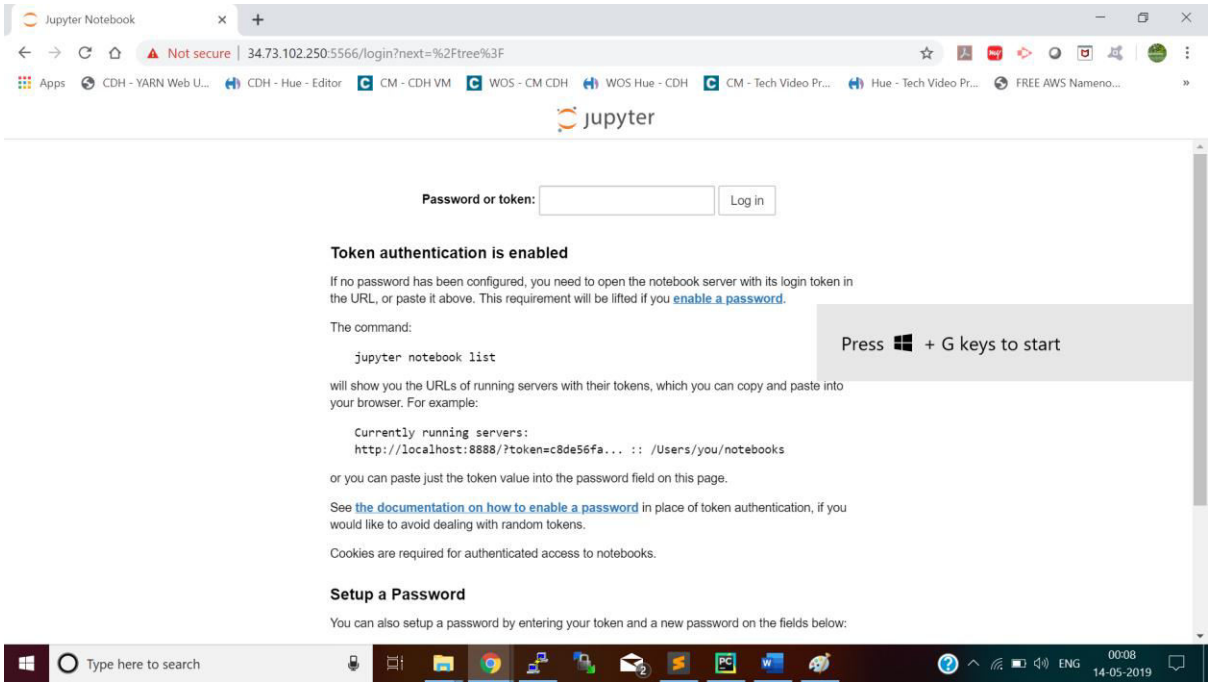
But this run only in localhost mode, so we need to configure remote access







Url: http://34.73.102.250:5566



Use the token to login Jupyter

Or

Create a login password using the following command

[hadoop@instance-1 ~]$ cd anaconda2/

[hadoop@instance-1 anaconda2]$

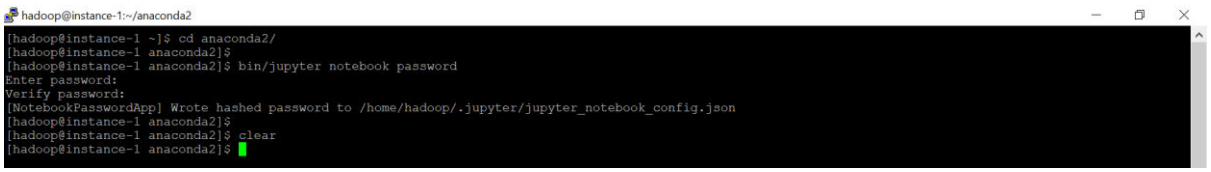
[hadoop@instance-1 anaconda2]$ bin/jupyter notebook password

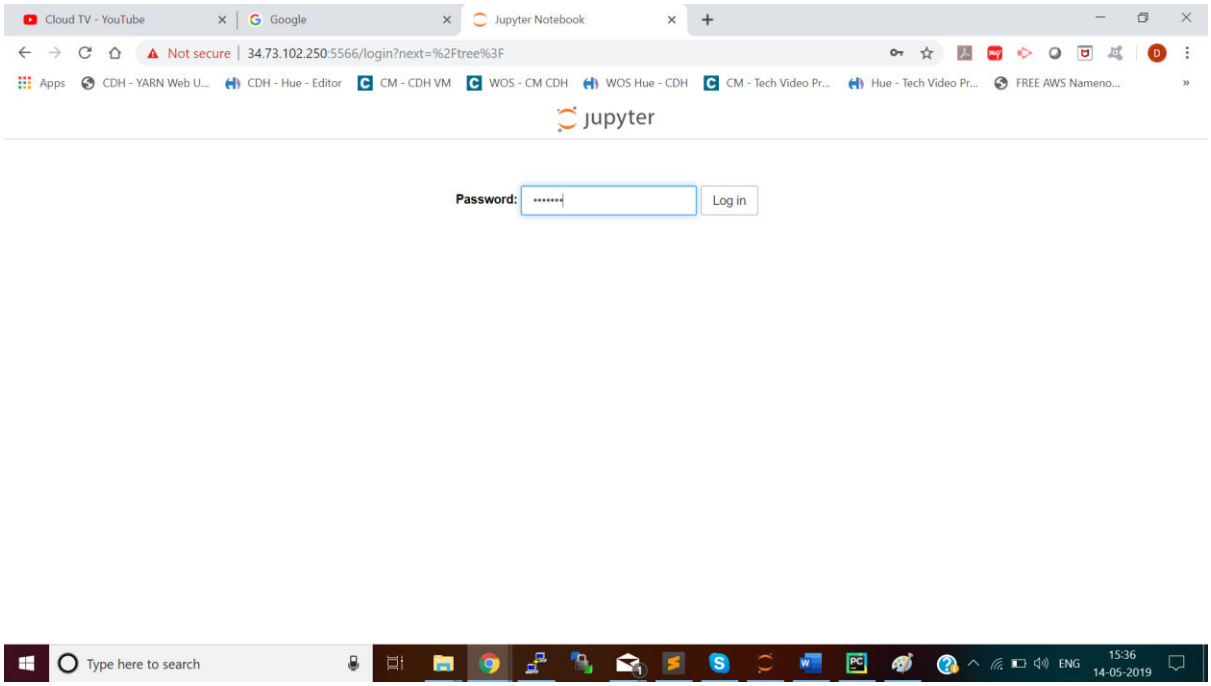
Enter password:

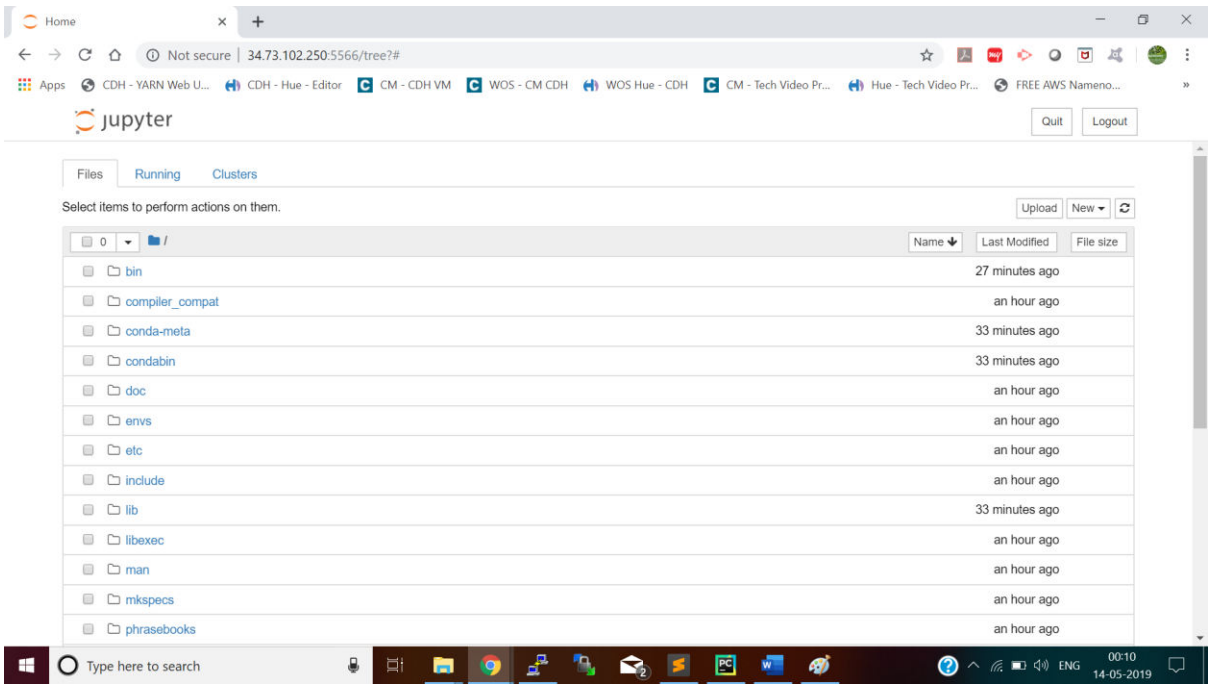
Verify password:

[NotebookPasswordApp] Wrote hashed password to /home/hadoop/.jupyter/jupyter_notebook_config.json

[hadoop@instance-1 anaconda2]$







Configure Spark in Jupyter

Open a below file and add the content

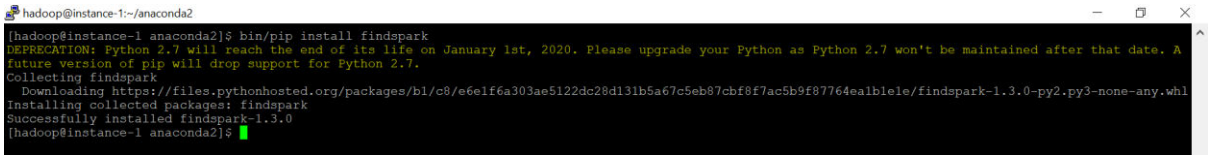
vi ~/.bashrc

export SPARK_HOME=/opt/cloudera/parcels/CDH-6.2.0-1.cdh6.2.0.p0.967373/lib/spark

export PATH=$PATH:$SPARK_HOME/bin

Install findspark python package

bin/pip install findspark



Start the Jupyter and login using password, then create a new notebook and add below lines of code

#Setting Spark installation location to jupyter

import os

import sys

os.environ["SPARK_HOME"] = "/opt/cloudera/parcels/CDH-6.2.0-1.cdh6.2.0.p0.967373/lib/spark"

os.environ["PYLIB"] = os.environ["SPARK_HOME"] + "/python/lib"

In below two lines, use /usr/bin/python2.7 if you want to use Python 2

os.environ["PYSPARK_PYTHON"] = "/home/hadoop/anaconda2/bin/python"

os.environ["PYSPARK_DRIVER_PYTHON"] = "/home/hadoop/anaconda2/bin/python"

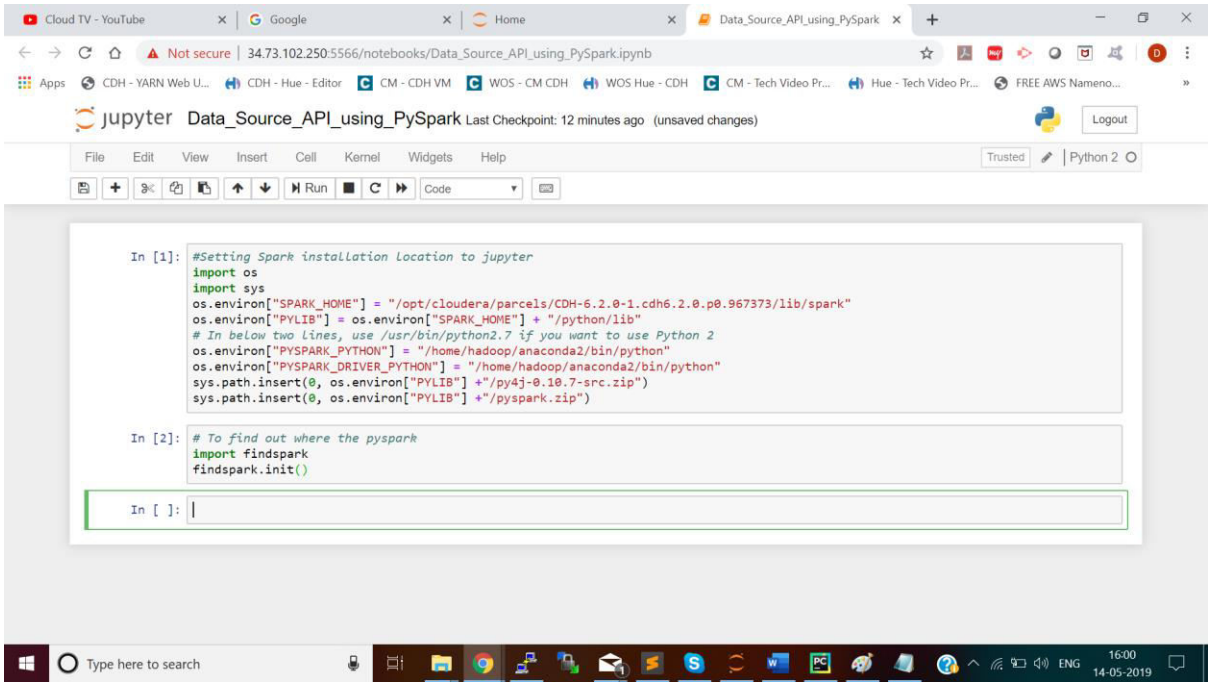
sys.path.insert(0, os.environ["PYLIB"] +"/py4j-0.10.7-src.zip")

sys.path.insert(0, os.environ["PYLIB"] +"/pyspark.zip")

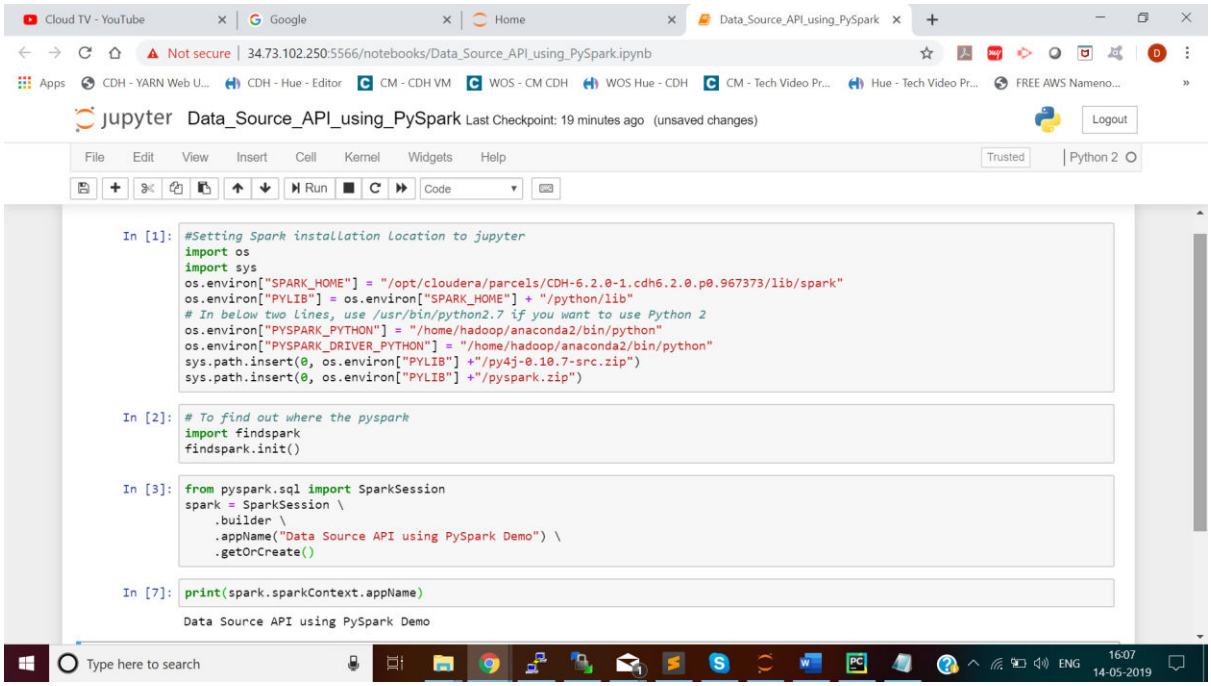
To find out where the pyspark

import findspark

findspark.init()




```
from pyspark.sql import SparkSession
spark = SparkSession \
    .builder \
    .appName("Data Source API using PySpark Demo") \
    .getOrCreate()
print(spark.sparkContext.appName)
```



Spark is successfully configured in the Python Jupyter.