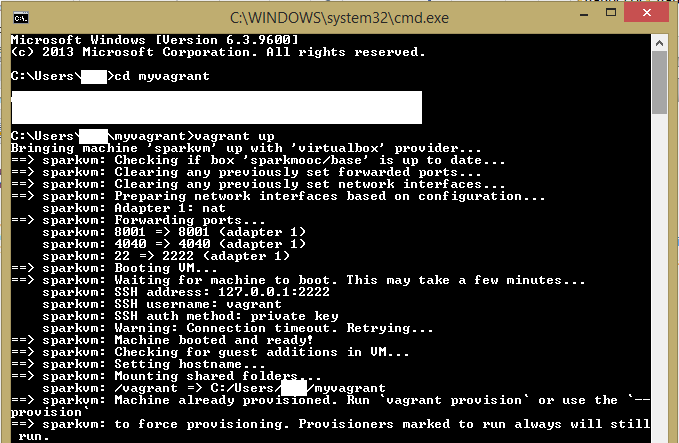
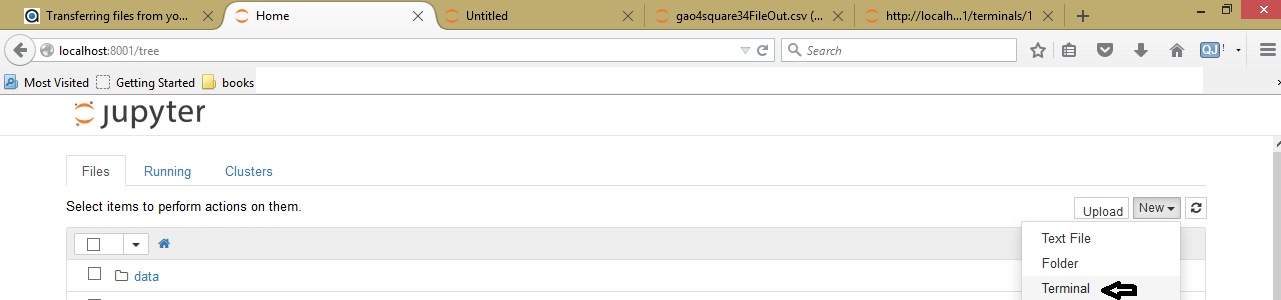
How to upload data files to Vagrant

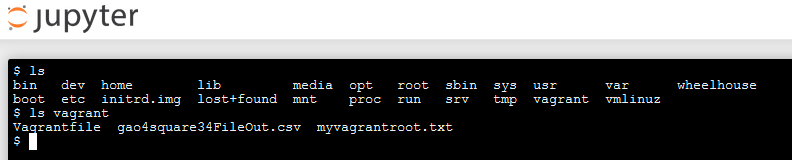
*Tags: Spark, VirtualBox, Vagrant, Jupyter*

Usually the upload feature in the Jupyter window fails when data files are large. I will demonstrate a method that uses shared folder. You should know the shared folder’s name in your hard drive. You can get it from the output when you start vagrant:

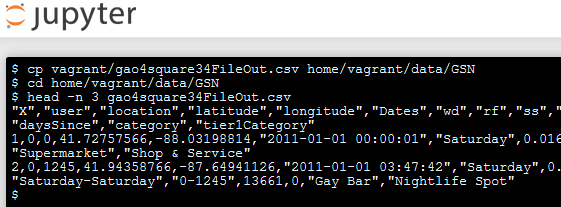


Typically, the shared folder is c:/Users/USERNAME/myvagrant. Copy the file containing your dataset in this directory. The shared folder is merely a folder in your virtual box that is running a linux OS. You can explore the shared folder using a terminal in the Jupyter.

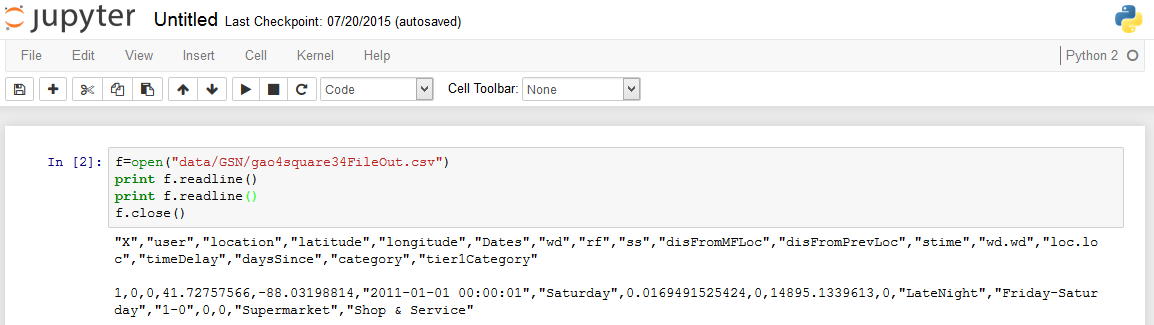




As you can see, I have uploaded the file gao4square34FileOut.csv which over 300MB. Now I have to copy this file into /home/vagrant folder so that I can see the dataset from Jupyter.



Some of the output is cut due to the space constraint. Now let’s write a python program to read the first two lines of the file.



The dataset is now in vagrant and we can read it.

Another way of uploading dataset is to use scp command. Format of the command is following:

scp fileYouWantToUpload vagrant@127.0.0.1:2222

You are likely to get an error saying bad file number. This is due to the fact that SSH is typically blocked on the specified port. You can resolve the issue by editing SSH configuration file. You can follow this [post](http://stackoverflow.com/questions/7144811/git-ssh-error-connect-to-host-bad-file-number).

Thanks for reading!