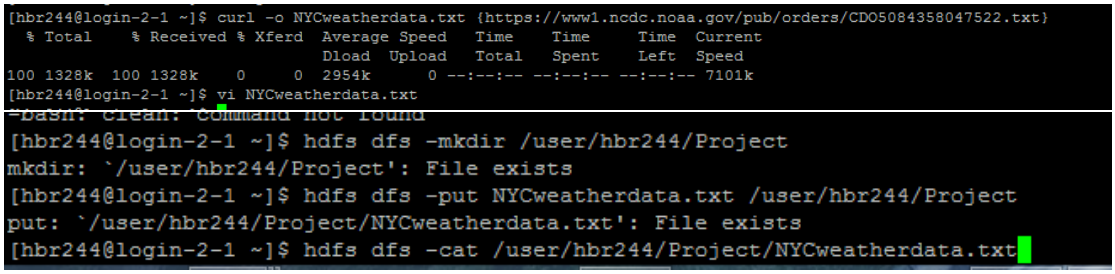
DATA INGESTING APPROACH

I tried using two ways, one was the curl command, the other one is the transferring it from local machine to Hadoop node.

Step1-

I used the curl command to ingest data initially,



Step-2-

Downloaded the csv version and transferred it to Hadoop using WinSCP. This way the entire data is now comma separated and easy to operate on.

To transfer using WinSCP

1) Connect to VPN
2) Open winscp
3) Login using your nyu creds
4) Drag and drop the file from your local to PUTTY