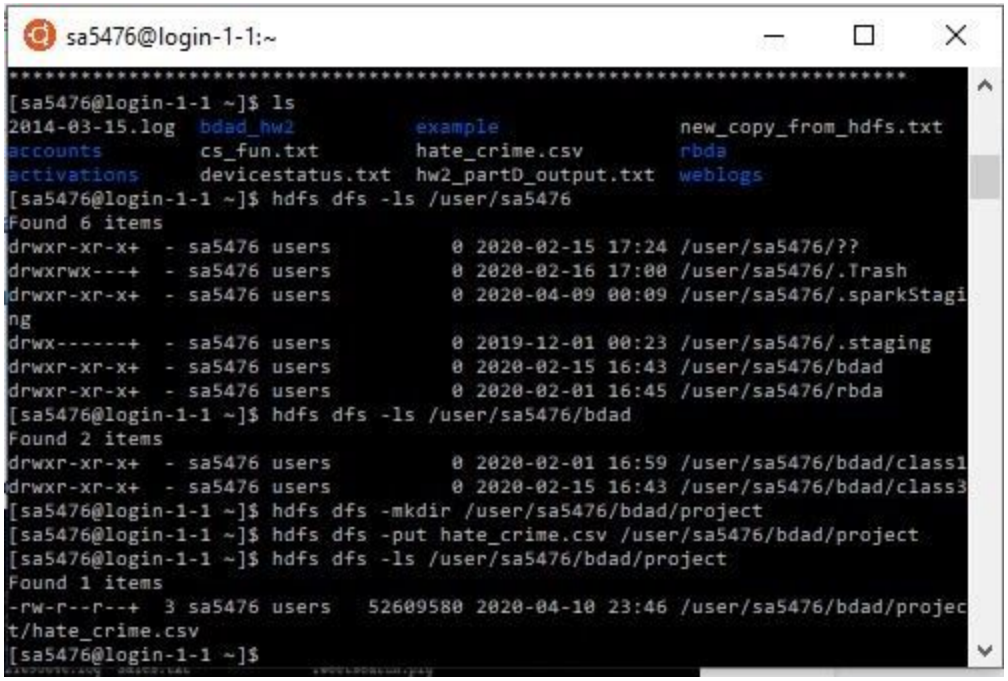
Data Ingestion

Steps to ingest data:

1. Downloaded WinSCP Software from NYU Data Services Lab

   - https://guides.nyu.edu/software
   - https://winscp.net/eng/download.php

2. Install and Open installed WinSCP Application
3. Use IP address of dumbo to login to WinSCP
4. Drag and drop data from local to dumbo.
5. Check in the dumbo terminal for the required folder using -ls command.



6. Folder “hate_crime.csv” is present.
7. Use the following commands to make folder project_01 in hdfs and ingest this data.:

   hdfs dfs -mkdir /user/sa5476/bdad/project
   hdfs dfs -put hate_crime.csv /user/sa5476/bdad/project
   hdfs dfs -cat /user/sa5476/bdad/project/hate_crime.csv

8. Snippet of the dataset.