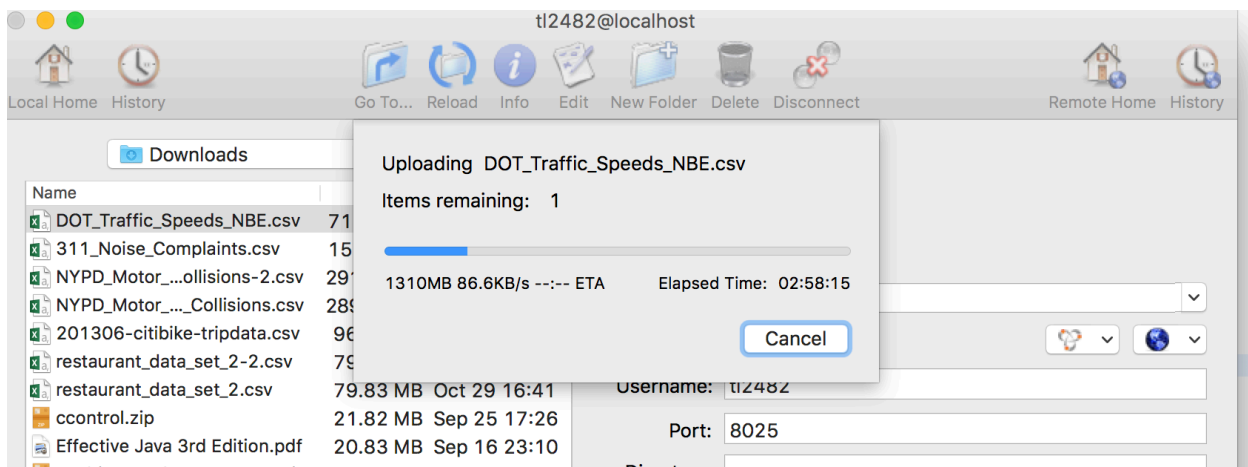


Homework 9: Code or Steps for Data Ingest

I used FUGU to transfer my data files from my local machine to dumbo. Fugu is a graphical frontend for the text-based Secure File Transfer Protocol (SFTP) client that ships with Mac OS X.

I followed following steps to transfer data using FUGU:

1. Start Fugu. Select SSH > new SSH tunnel
 - Create tunnel to: prince (or dumbo)
 - Service or port: 22
 - Local port: 8026 (8025 to dumbo)
 - Tunnel host: gw.hpc.nyu.edu
 - Username: <NetID>
 - Port:
2. In SFTP window
 - Connect to: localhost
 - Username: <NetID>
 - Port: 8026 (or 8025)
3.
 - Click connect and enter <NetID> password.
 - Drag and drop files to copy/paste to and from cluster.



I uploaded this data from dumbo to hdfs using following commands:

```
hdfs dfs -mkdir project_data
```

```
hdfs dfs -put restaurant_data_set_2-2.csv project_data/
```

```
hdfs dfs -put NYPD_Motor_Vehicle_Collisions-2.csv project_data/
```

hdfs dfs -put DOT_Traffic_Speeds_NBE.csv project_data/

```
*****
[tl2482@login-2-1 ~]$ cd rbda/
[tl2482@login-2-1 rbda]$ cd project_data/
[tl2482@login-2-1 project_data]$ ls -lrt
total 425408
-rw-r--r-- 1 tl2482 users 305281450 Nov  5 16:29 NYPD_Motor_Vehicle_Collisions-2.csv
-rw-r--r-- 1 tl2482 users  83803812 Nov  5 16:31 restaurant_data_set_2-2.csv
-rw-r--r-- 1 tl2482 users  60063744 Nov  6 10:36 DOT_Traffic_Speeds_NBE.csv
[tl2482@login-2-1 project_data]$
[tl2482@login-2-1 project_data]$
[tl2482@login-2-1 project_data]$
[tl2482@login-2-1 project_data]$ hdfs dfs -mkdir project_data
[tl2482@login-2-1 project_data]$ hdfs put restaurant_data_set_2-2.csv project_data/
Error: Could not find or load main class put
[tl2482@login-2-1 project_data]$ hdfs dfs -put restaurant_data_set_2-2.csv project_data/
[tl2482@login-2-1 project_data]$ hdfs dfs -put NYPD_Motor_Vehicle_Collisions-2.csv project_data/
[tl2482@login-2-1 project_data]$ ls -lrt
total 495168
-rw-r--r-- 1 tl2482 users 305281450 Nov  5 16:29 NYPD_Motor_Vehicle_Collisions-2.csv
-rw-r--r-- 1 tl2482 users  83803812 Nov  5 16:31 restaurant_data_set_2-2.csv
-rw-r--r-- 1 tl2482 users 117964800 Nov  6 10:38 DOT_Traffic_Speeds_NBE.csv
```