

Steps to Transfer Data

yellow_trip Dataset

1. Copy data files from local to Dumbo:

```
scp yellow_tripdata_2015-01-06.csv rrm404@dumbo.es.its.nyu.edu:/scratch/rrm404
```

2. Compile the Code to Transfer the large file to HDFS Cluster

```
javac -classpath `yarn classpath` -d TransferDataToHDFS.java
```

```
jar cvf TransferData.jar *.class
```

3. Run the Hadoop jar to Transfer data to a directory on HDFS

```
hadoop jar TransferData.jar TransferDataToHDFS yellow_tripdata_2015-01-06.csv
```