Chicago Taxi Trip Dashboard Implementation

# Set up EC2 Instance for this project

400GB

Download and run setup.script

./setup\_ucb\_complete\_plus\_postgres.sh /dev/xvdf

Mount the device when restart the instance

Mount /dev/xvdf /data

# Save all scripts in /data/taxi\_trip Directory

Under root folder

Git clone <https://github.com/CynthiaHu/W205_Summer.git>

[root@ip-172-31-16-56 ~]# cd W205\_Summer/"Course Project"/"1. Loading and Modeling"

# Run load\_data.sh to load data and put them into HDFS

# Run hive\_trip\_ddl.sql to create base Hive tables

# Run trip\_transforming.sql to transform datasets

# Run trip\_aggregation.sql to aggregate datasets for visualization and forecast

# Set up python and run trip\_forecast.py

Make sure Python 2.7 or higher version is used. If not, please use python\_setup.sh to install python3 and appropriate version of pip.

Next, use Python3 for PySpark and execute the python script

$ cd /data

$ su - w205

$ export PYSPARK\_PYTHON=python3

$ /data/spark15/bin/spark-submit /data/taxi\_trip/trip\_forecast.py

# Connect tableau to the Hive Server in EC2 instance

# Generate and Publish Dashboard