Real Time Big Data Project - Input Data & Terminal Record

This is an explanation file that explains each data source used in my code part.

Tweet Analytics
blockCount.txt : block count result
phoenix_complete_feature_clean.csv : complete feature table with NULL fileds replaced with 0, and duplicated fields deleted
phoenix_complete_feature.csv : complete feature table
phoenix_full_feature.csv : selected feature from the complete feature table. used for machine learning
tweets_phoenix_clean_combine : combined tweets data collected for 4 weeks
tweets_phoenix_clean_combine_blockNum_senti.tsv : tweets data with block num labeled and sentiment analysis added

Yelp Dataset
yelp_feature : yelp feature table. For detailed information, see Jinglin's data explanation.

GTFS Dataset
For detailed explanation, see Xiao Xi's document.

Terminal Record
This is the running record from the terminal, which shows successful Hadoop MapReduce job running / Hive query.