Analysis of Parking Violations in New York City Metro Area

Team Members

Justin Nichols: jnich56@lsu.edu

Bryce Lee: blee57@lsu.edu

Dataset

The dataset contains parking violation data around the NYC metro area. We plan to use information regarding the issue date, registration state, violation code, plate type, vehicle make, and vehicle body type.

Link to data: https://www.kaggle.com/new-york-city/nyc-parking-tickets

The dataset contains 4 separate files.

1. Parking_Violations_Issued_-_Fiscal_Year_2014__August_2013___June_2014_.csv (1.87 GB)
2. Parking_Violations_Issued_-_Fiscal_Year_2015.csv (2.86 GB)
3. Parking_Violations_Issued_-_Fiscal_Year_2016.csv (2.15 GB)
4. Parking_Violations_Issued_-_Fiscal_Year_2017.csv (2.09 GB)

Big Data Frameworks

1. Apache Hadoop MapReduce
2. Apache Hue
3. Cloudera VM pseudo-distributed mode

Current Status

Currently, our team has brainstormed how to implement the MapReduce problem we have on hand as well as sifting through the data to see what attributes are useful to use. We have also created and set up our environment in the virtual machine. Along with these initial steps, we have started programming our big data application. As of the midterm report milestone, we have successfully set up the driver code, completed the development for both the mapper and the partitioner, and started the reducer.

Plans

From now until the project due date, we plan to complete the development of the reducer and test and refine the code we have currently written. We also plan to analyze our generated outputs and make inferences on the parking violation data. Finally, we plan to unify everything we have learned and translate it into a coherent final report and presentation.

Division of Roles

We have worked simultaneously on the project, equally committing the same amount of work and time. We have utilized pair programming to help streamline our development process. From finding the data, to writing the code, we have both contributed equally and have kept an open collaborative environment to allow ideas to bounce off one another.

Appendix

File Name	Number of Lines
ParkingViolationsDriver.java	78
ParkingViolationsPartitioner.java	35
ParkingViolationsMapper.java	73
ParkingViolationsReducer.java	23