THE DOWNLOAD

TECH TALKS BY HPCC SYSTEMS

Analyzing Telematics Data to Support the Connected Cars Industry

HPCC SYSTEMS®

Huafu Hu, Georgia State University

Georgia State University

Project Background

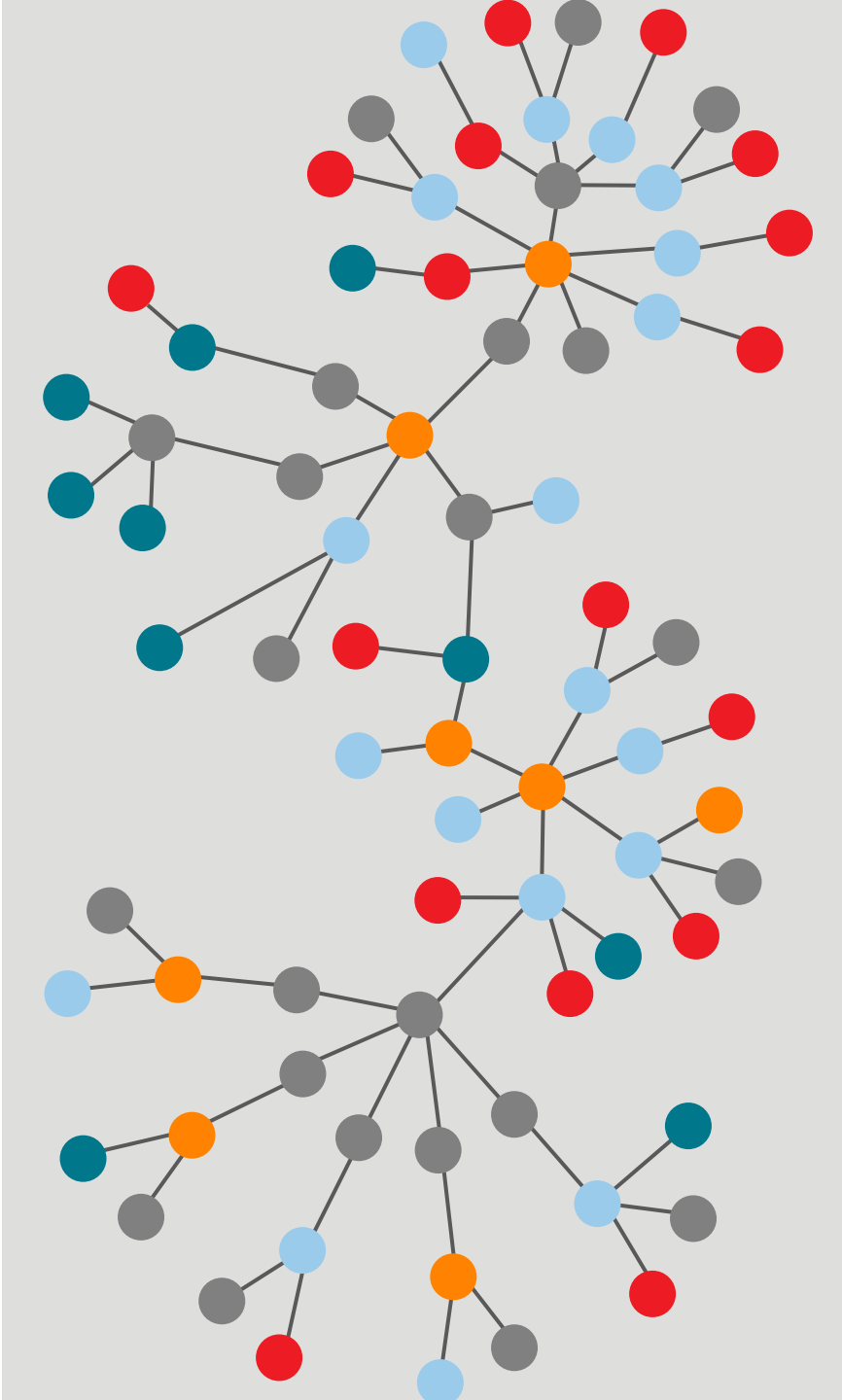
- Ten-years' Experienced Engineer
- Implemented Online Graduate Student Management System for Georgia State University and Contributes 100,000 lines of code in two years
- 12 Weeks time limited
- 1 million of trips

Quick poll:

What was the average auto insurance cost in the U.S. for the year 2016?

See poll on bottom of presentation screen

HPCC SYSTEMS®

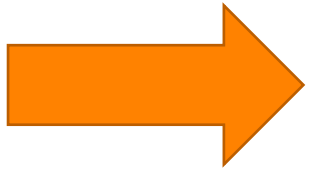


OUTLINE

- Introduction
- Why we establish such an initiation for the insurance industry
- How do we solve for the efficiency in the big data processing
- Result of our insight on big amount of trip data
- Conclusion

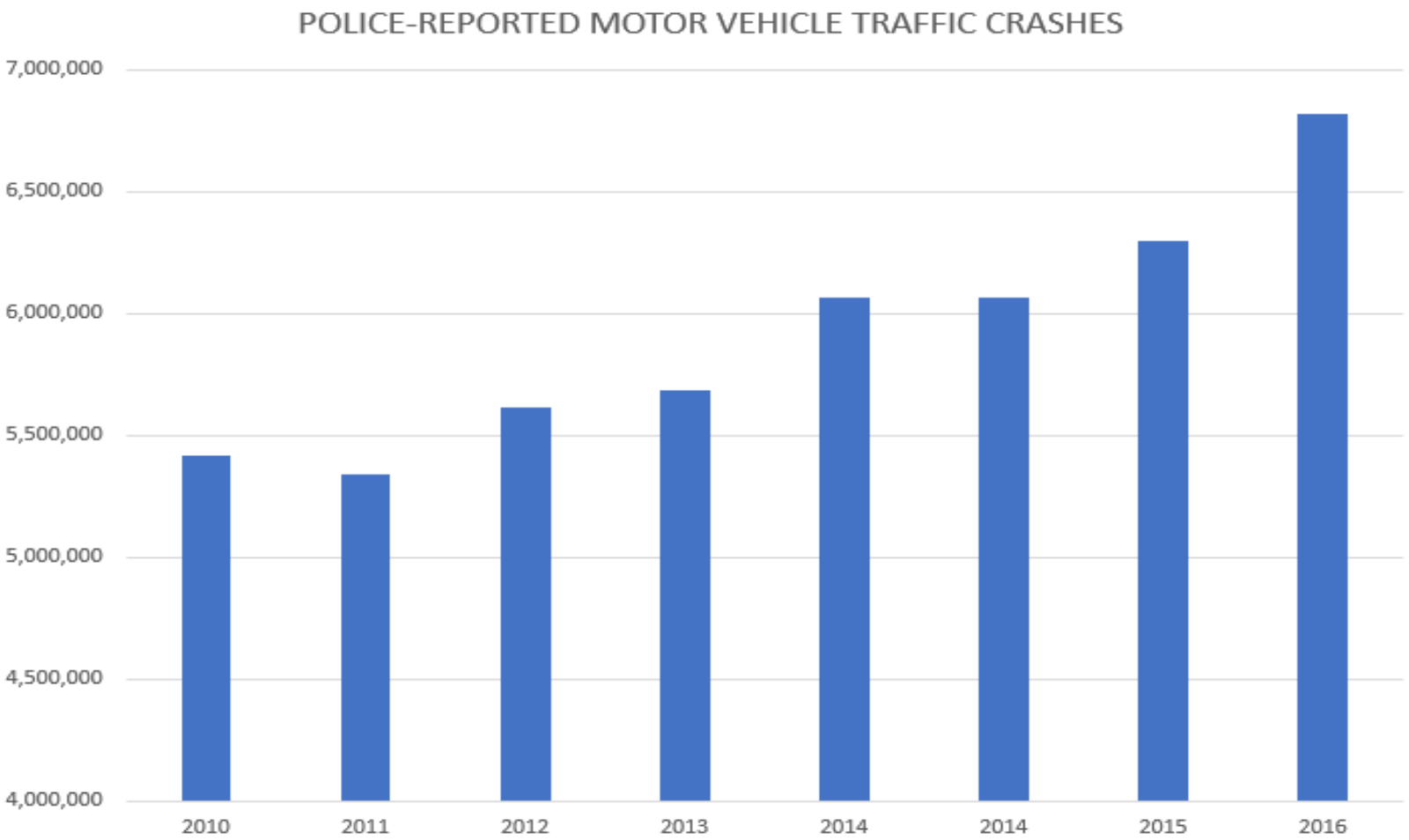
Introduction:

Source: US Department of Veterans Affairs Website



HPCC SYSTEMS®

Introduction: Accident increased year by year



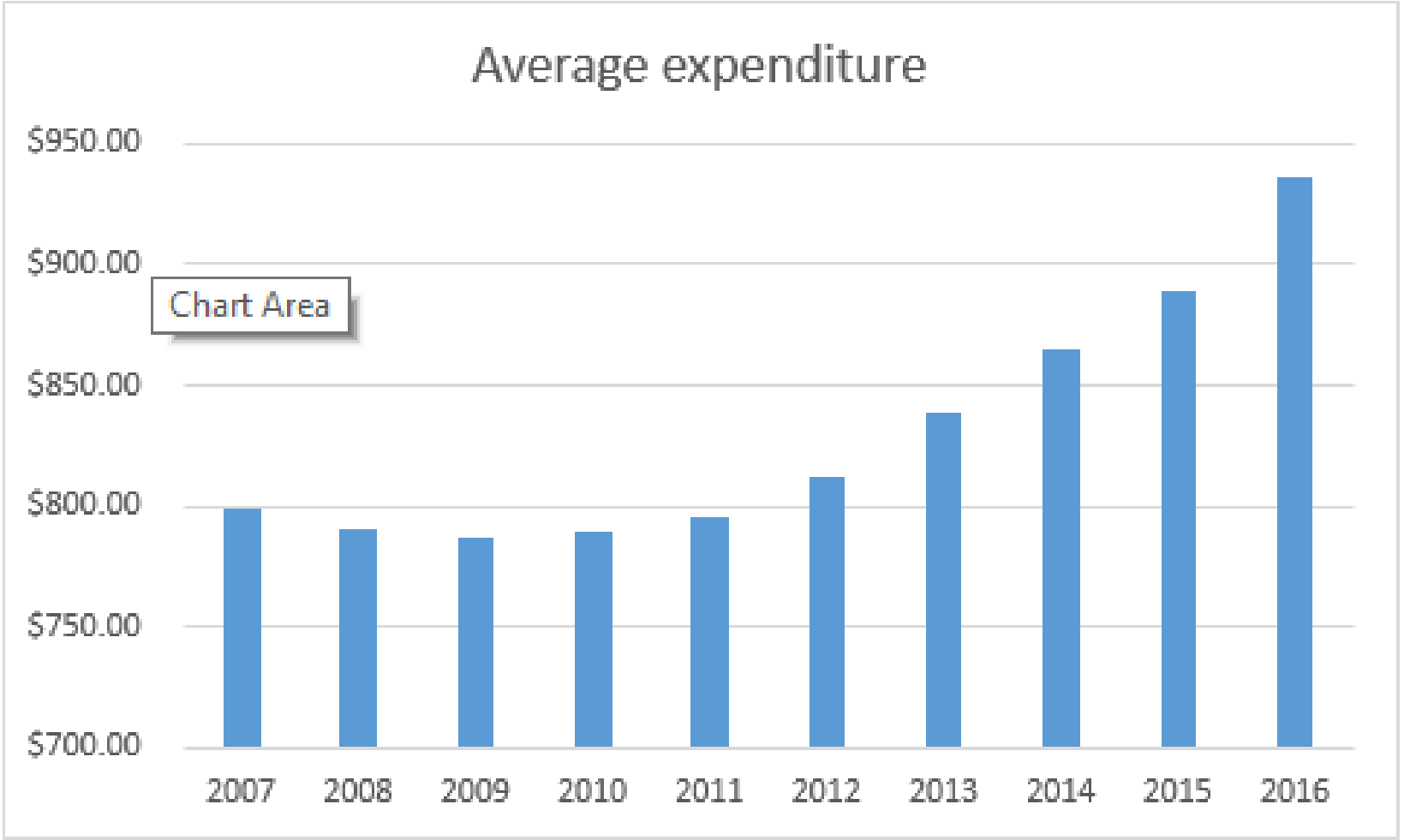
Source: National Highway Traffic Safety Administration's (NHTSA) website

- Traffic accidents occur every day in the United States. The left chart outlines how many car accidents per year in the US took place from 2010 to 2016

HPCC SYSTEMS®

Introduction: Average Auto Insurance from 2007 to 2016

- The diagram highlights that the mean insurance fee has increased by around 20% from 780$ at 2008 to 935$ at 2016.



Source: 2018 National Association of Insurance Commissioners (NAIC)

HPCC SYSTEMS®

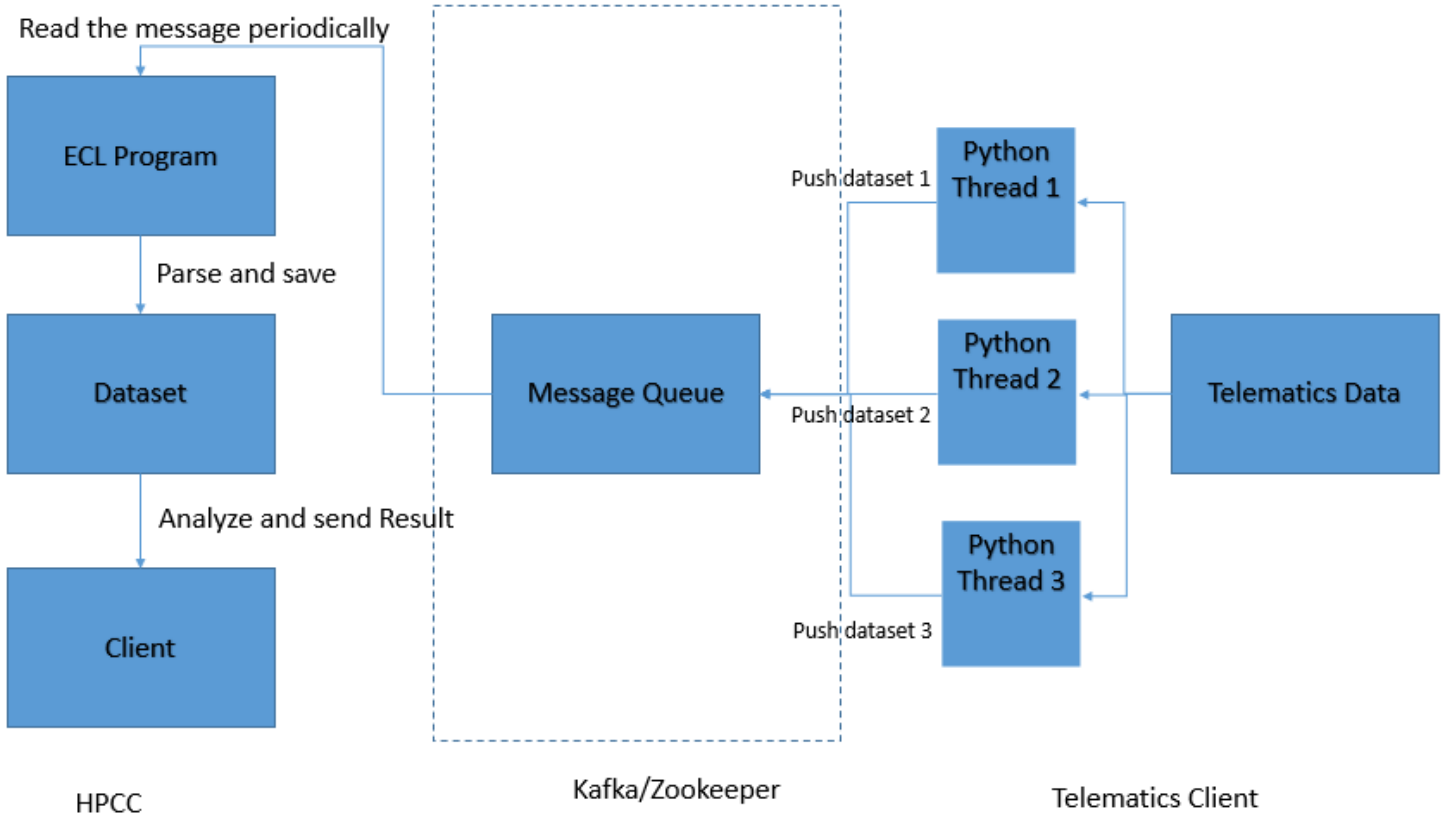
Challenge to the problem

- Huge data
- Precise the data
- Leverage the pattern of the data

HPCC SYSTEMS®

Solution

- The project consists of three modules, Simulated Telematics Demo, Kafka Message Queue, HPCC Data Analysis Demo.



Data Flow for Telematics and HPCC

HPCC SYSTEMS®

ETL Process: Extra, Transform, and Load Data on HPCC System

- Understand the meaning of each column in the dataset
- Find the key column(s) for this dataset
- Locate the top columns through previous steps and start to statics these properties by grouping the key column(s)
- Based on this statics, it will much easy to seek out the vital attributes

Device ID: Car ID

deviceid	cnt
0	29.57
5	1.16
7	2.12
9	40.93
10	98.18
12	95.93
14	0.27
16	23.1

Cnt: Total driving time(Hour)

Cnt Trip: Total strip per car

deviceid	cnttrip	avgtriptime
0	87	20.394833333333334
5	6	11.561166666666667
7	4	31.8625
9	157	15.643333333333333
10	431	13.667833333333334
12	278	20.7035
14	4	4.0875
16	43	32.2365

Cnt: Total driving time(Minute)

RPM: Round Per Minute(Engine)

deviceid	tripid	gps_speed	rpm	id	accelerationgps	accelerationrpm
0	1	24.2612	1010.75	1	0	0
0	1	23.15	815.5	2	-1.1112	-195.25
0	1	18.7052	862.25	3	-4.4448	46.75
0	1	16.4828	817	4	-2.2224	-45.25
0	1	17.4088	804.25	5	0.926	-12.75
0	1	15.0012	831.5	6	-2.4076	27.25
0	1	11.6676	878.25	7	-3.3336	46.75
0	1	10.0008	818	8	-1.6668	-60.25
0	1	12.2232	783.25	9	2.2224	-34.75

AccelerationGPS: Acceleration based on gps speed
AccelerationRPM: Acceleration based on RPM

ETL Process(Cont.)

- Hard acceleration(left graph) or hard braking(right graph) is a driver event when more force than normal is applied to the vehicle's accelerator or brake system. It can be an indicator of aggressive or unsafe driving behavior. At the very least, this driving habit is wasteful and uneconomic.
- There is no official definition to the threshold of hard acceleration or hard brake. Some articles refer this can range from 1 mph to as high as 15 mph. The most common threshold to record a hard brake or hard acceleration is a sudden decrease or increase the 18-wheeler's speed by 7 mph(mile per hour) or more within one second(10.2 feet per square second).

11

Demo

- The clients (it can simulate hundreds/thousands of senders by multiple threads in one cluster) with google cloud instances in this system could export these real-time data to JSON messages and delivery them to Kafka

```
thread:Thread-3 send index:666 data:"tripjson": {"Row":[{"tripID":413,"deviceID":1210,
thread:Thread-2 send index:668 data:"tripjson": {"Row":[{"tripID":415,"deviceID":1110,
thread:Thread-3 send index:667 data:"tripjson": {"Row":[{"tripID":414,"deviceID":1210,
thread:Thread-2 send index:669 data:"tripjson": {"Row":[{"tripID":416,"deviceID":1110,
thread:Thread-3 send index:668 data:"tripjson": {"Row":[{"tripID":415,"deviceID":1210,
thread:Thread-2 send index:670 data:"tripjson": {"Row":[{"tripID":417,"deviceID":1110,
thread:Thread-3 send index:669 data:"tripjson": {"Row":[{"tripID":416,"deviceID":1210,
thread:Thread-2 send index:671 data:"tripjson": {"Row":[{"tripID":418,"deviceID":1110,
thread:Thread-3 send index:670 data:"tripjson": {"Row":[{"tripID":417,"deviceID":1210,
thread:Thread-2 send index:672 data:"tripjson": {"Row":[{"tripID":419,"deviceID":1110,
thread:Thread-3 send index:671 data:"tripjson": {"Row":[{"tripID":418,"deviceID":1210,
thread:Thread-2 send index:673 data:"tripjson": {"Row":[{"tripID":420,"deviceID":1110,
thread:Thread-3 send index:672 data:"tripjson": {"Row":[{"tripID":419,"deviceID":1210,
thread:Thread-2 send index:674 data:"tripjson": {"Row":[{"tripID":421,"deviceID":1110,
thread:Thread-3 send index:673 data:"tripjson": {"Row":[{"tripID":420,"deviceID":1210,
thread:Thread-2 send index:675 data:"tripjson": {"Row":[{"tripID":422,"deviceID":1110,
thread:Thread-3 send index:674 data:"tripjson": {"Row":[{"tripID":421,"deviceID":1210,
thread:Thread-2 send index:676 data:"tripjson": {"Row":[{"tripID":423,"deviceID":1110,
thread:Thread-3 send index:675 data:"tripjson": {"Row":[{"tripID":422,"deviceID":1210,
thread:Thread-2 send index:677 data:"tripjson": {"Row":[{"tripID":424,"deviceID":1110,
thread:Thread-3 send index:676 data:"tripjson": {"Row":[{"tripID":423,"deviceID":1210,
```

```
[DEFAULT]
broker = 172.31.42.181:9092
telematics_json_directory = /mnt/disks/sdb/dataprepare
telematics_json_start = 11
telematics_json_end = 12
threadnumber = 2
sleeptime = 0.3
~
```


HPCC SYSTEMS®

Demo

- Received Kafka message on HPCC periodically, parsed and save the dataset to a super file

levin::kafkamessagetable2019-10-24t13:53:01.571.csv	hpccdemo	levin::kafkamessageta...		hthor__myecla...	286,318	66.08 MB	1 2019-10-24 13:53:18
levin::kafkamessagetable2019-10-24t13:52:01.570.csv	hpccdemo	levin::kafkamessageta...		hthor__myecla...	167,883	38.75 MB	1 2019-10-24 13:52:11
levin::kafkamessagetable2019-10-24t13:51:01.588.csv	hpccdemo	levin::kafkamessageta...		hthor__myecla...	192,356	44.39 MB	1 2019-10-24 13:51:11
levin::kafkamessagetable2019-10-24t13:50:01.579.csv	hpccdemo	levin::kafkamessageta...		hthor__myecla...	143,846	33.20 MB	1 2019-10-24 13:50:10
levin::kafkamessagetable2019-10-24t13:49:01.527.csv	hpccdemo	levin::kafkamessageta...		hthor__myecla...	194,395	44.86 MB	1 2019-10-24 13:49:14
levin::kafkamessagetable2019-10-24t13:48:01.500.csv	hpccdemo	levin::kafkamessageta...		hthor__myecla...	35,674	8.23 MB	1 2019-10-24 13:48:05

Analysis Result

- Figure 1 and 2 compares eight drivers in these two angle. The X-axis indicates the car ID. The Y-axis in figure 1 means The average hard acceleration. The Y-axis in figure 2 means the average hard brake per hour. From these two graph, it's simple to determine driver 1110 and 1116 are the worse driver in this list.

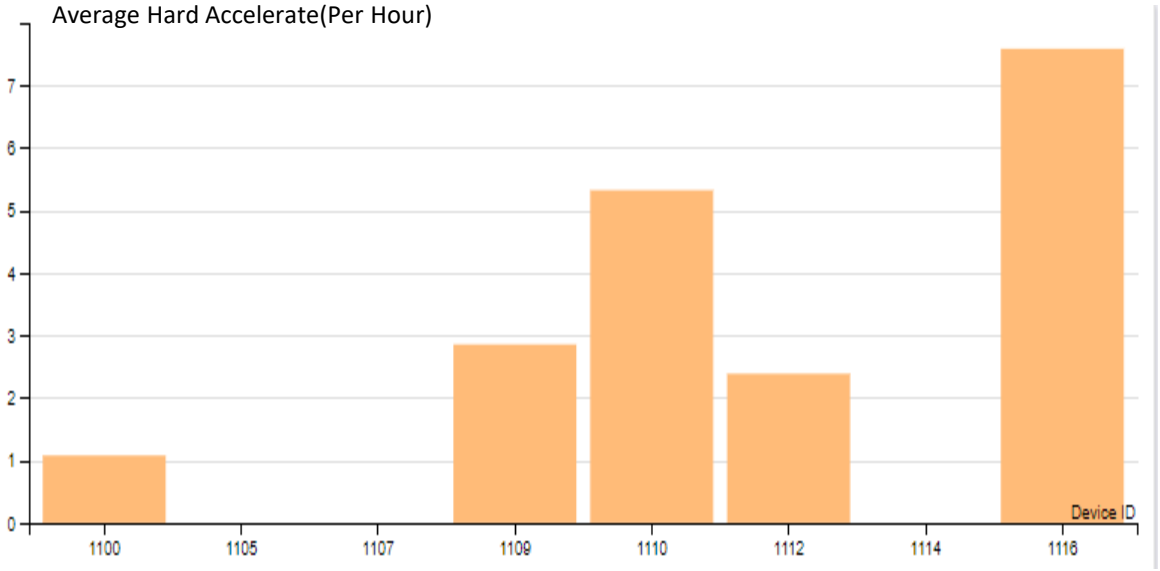


Figure 1: Vehicle and Average Hard Accelerate

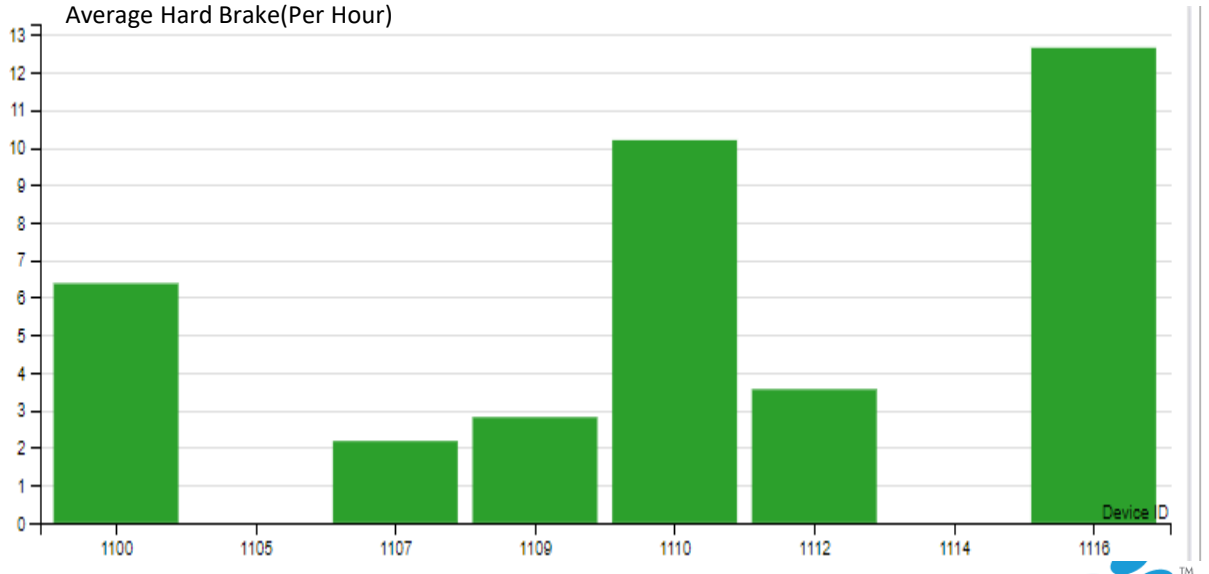


Figure 2: Vehicle and Average Hard Brake

HPCC SYSTEMS®

Analysis Result(Cont.)

- Figure 3 and 4 compares eight drivers in these two angle, total trips and total driving hours. The X-axis indicates the car ID. The Y-axis in figure 3 means The total trips for each driver. The Y-axis in figure 4 means the total driving hours.

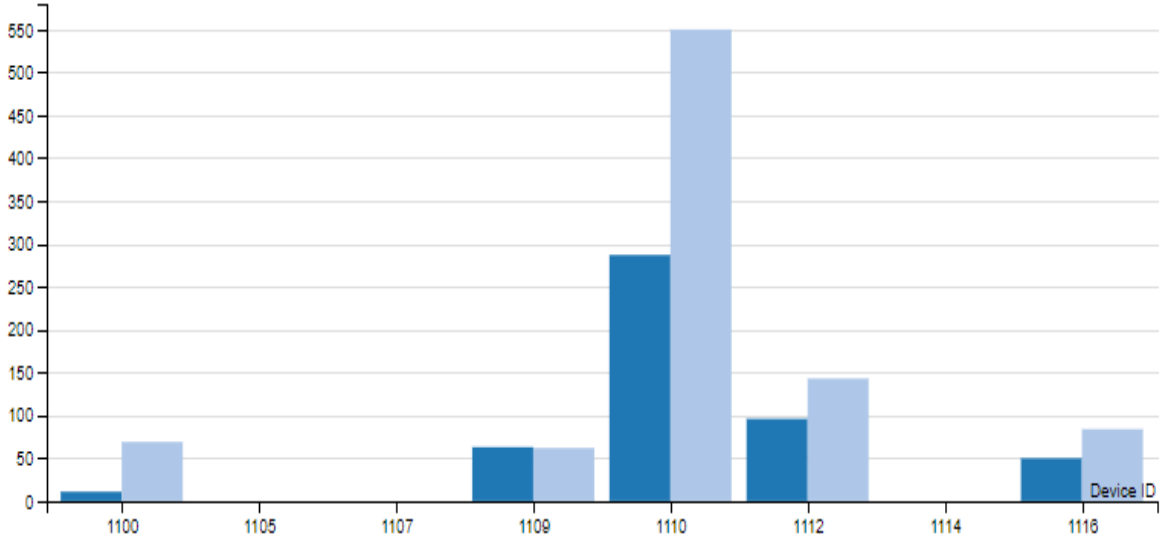


Figure 3: Vehicle and Total Trips

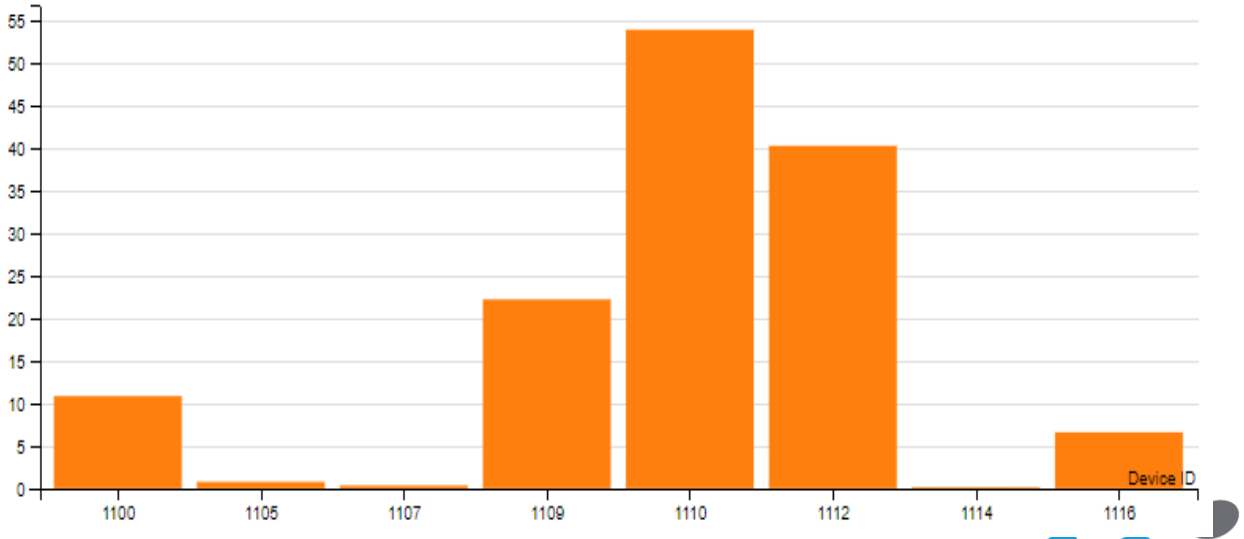
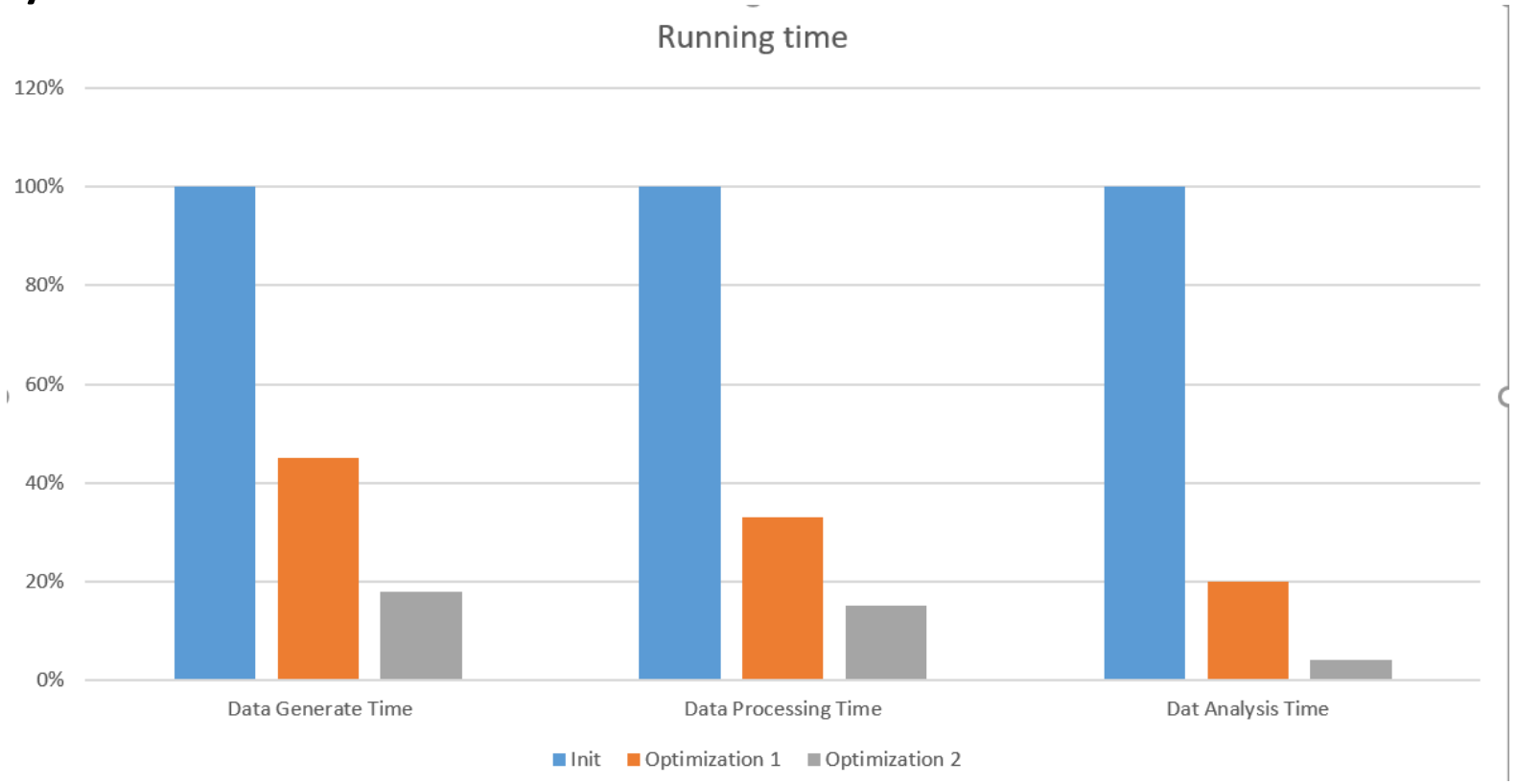


Figure 2: Vehicle and Total Driving Hours

HPCC SYSTEMS®

Performance Improvement

- For the dataset of 1 million lines, after the optimization of the program, the running time of data generation , data processing and data analysis has decreased by 80% to 95%.



HPCC SYSTEMS®

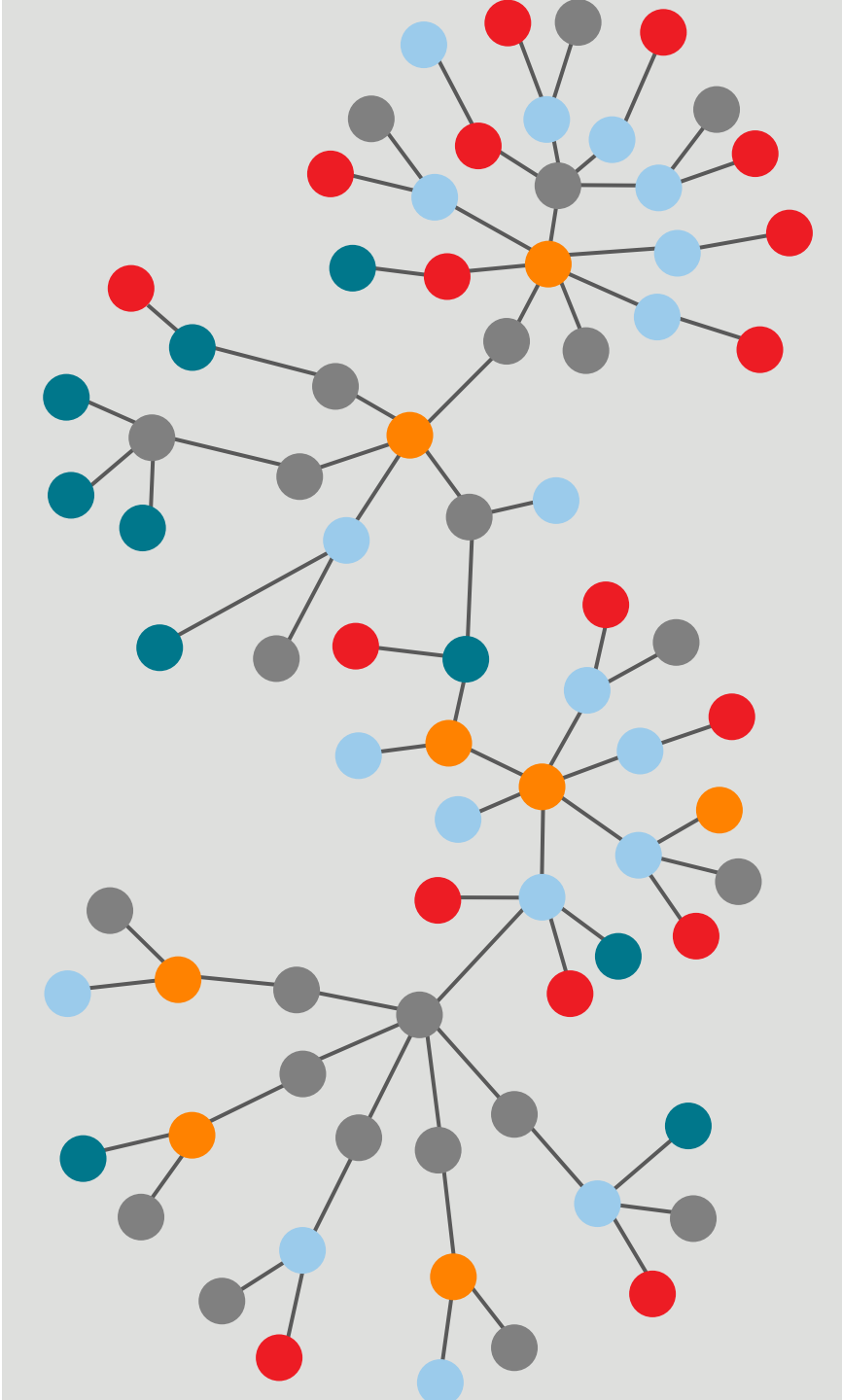
Conclusion

- Addressed the customized pricing problem in the vehicle insurance industry based on the ***customized driving habit*****.**
- Implemented a data ***processing pipeline*** with HPCC Systems
- Provide ***deep insight*** into the driving habit to model the risk of human behaviors
- Provide a ***comprehensive one-stop solution*** for insurance industry to implement an intelligent pricing model for each customer

Quick poll:
Do you think this model using HPCC Systems may help decrease your insurance costs?

See poll on bottom of presentation screen

HPCC SYSTEMS®



Questions?

Huafu Hu
Georgia State University
Master Student and Team Leader of OGMS
hhu4@student.gsu.edu

HPCC SYSTEMS®