Big Data in Transit and Rail
Big Data Overview

• What is Big Data?

Any large volume of data
Structured or Unstructured
Coined in early 2000’s
Hadoop is Big Data Framework

• The 4 V’s of Big Data

Volume – the Scale of the Data
Velocity – Analysis of Streaming Data
Variety – Different Forms of Data
Veracity – Quality of Data
Today, data scientists max out at yottabytes, but soon, brontobytes will measure the volume of sensor data generated by the Internet of Things.
Big Data Velocity

- Certain information value decays over time. Incident or equipment failure data today is of less use tomorrow.
- Rail vehicle sensors generate massive log data in real time
- Ridership behavior can also be captured in real-time AND stored for trend analysis
Cybersecurity Challenges of Big Data
The “Elephant” in the Room

• Hadoop is an open source, java-based programming framework developed for storing and processing huge data sets

• As with most open source software, Hadoop was not written with security in mind

• Hadoop is frequently used with multiple vendor products which are also security challenged.

• The Cost of a Data Breach with BIG DATA is not quantifiable currently but assuredly will be quite large.

• The Bigger the Data, the Bigger the Risk.

• September 2017: Recent security breach at Equifax exposes the enterprise to lawsuits estimated to be in the billions of dollars.
Use Case – Electronic Fare Payment Systems

• Hop Fastpass (Portland, OR/Vancouver, WA)
  – **Account-based**: real-time processing of payment transactions and system performance information
  – **Open Architecture**: key system interfaces based on published Application Programming Interfaces (APIs)
  – **Information Protection**: separate data repositories for customer PII and transit use data

• Types of Data Available
  – **Ridership**: by individual (anonymous), fare category, agency, type of service, date/time, geolocation
  – **Sales Channel**: web, mobile, retail outlet, transit store, vending machine
  – **Customer Data**: collected via surveys linked to anonymous Hop customer accounts (age, income, first/last mile, # in household, etc.)
  – **Operational Data**: equipment failure (response and analysis)
Use Case – Electronic Fare Payment Systems
How Might Data Be Used?

• Service Planning & Customer Service
  – Identify ridership patterns not accurately captured by rider surveys or APCs such as inter- and intra-agency transfers
  – Real time information for enhanced customer service and service planning (i.e. crowd sourcing via mobile)
  – Asset management provides proactive device maintenance
  – Accurate and simplified NTD Reporting

• Third-party Integrations
  – Incentivize Transit Use: gamifying (use transit XX-times earns a discount at a local retailer)
  – Bike Share: enhanced customer convenience and to influence first/last mile mode choice
  – City Parking: enhanced customer convenience and to influence mode choice
  – Congestion Management: by knowing and influencing customer behavior
Account-based Architecture
All Systems Feed Into Data Warehouse
Other Big Data Opportunities in Transit Rail Vehicle Sensors - The Value of NOW

1. HVAC management
2. Temperature
3. Passenger Information System
4. Diagnostics, crew HMI management
5. Lighting management
6. Water tanks, toilets
7. CCTV system management
8. Battery charge monitoring
9. Door control
10. Emergency communications
11. Event recorder, legal recording unit
12. Train-to-wayside communication
13. Pantograph control
14. Remote input/output module
15. Speed measurement
16. Lateral vibration
17. Brakes
18. Traction
Other Big Data Opportunities in Transit
How Might Data Be Used?

• **Safety and Security** – CCTV alerts enable operations and security response

• **Operator Report Card** – for training and incident investigation

• **Real time Maintenance and Failure Alerts** – for timely predictive maintenance and improved service

• **Accurate Vehicle Location** – enhanced customer service and agency performance monitoring
Trends For Big Data in Transportation

• The value of NOW means processing in real-time provides actionable information not just historical data for trend analysis

• Convergence of IoT, big data, and cloud services provided by Amazon AWS, Google Cloud, and Microsoft Azure is enabling transit agencies to utilize cloud service dashboards

• Agencies with recent big data Initiatives per 2016 APTA TCRP study include:
  
  MBTA, San Joaquin RTD, LA County MTA, TriMet, Port Authority of Allegheny County, Capital Metro, Utah Transit Authority, York Region Transit, Yuma County, and WMATA

• The list of transit agency big data projects will continue to grow
Brin Owen, VP Payment Systems
CH2M San Francisco Office - 415-728-0677
Brin.owen@ch2m.com

Raja Kadiyala, VP Intelligent Systems
CH2M Oakland Office - 510-587-7676
Raja.Kadiyala@ch2m.com

Susan Howard Intelligent Systems Cybersecurity Analyst
CH2M Portland Office - 503-872-4702
Susan.Howard@ch2m.com
Thank You