What's the Big Deal about Big Data in Transportation?

Prepared for: Florida Department of Transportation, Data Symposium, October 23, 24, Renaissance Hotel, SeaWorld, Orlando, Florida

Raj Kanniyappan, Senior Solutions Architect, Teradata Aster

Bob McQueen, CEO, The 0Cash Company

Topics

- Big Data, Data Science and Data Analytics
- Trends in transportation
- The use of Big Data and Data Analytics in Transportation
- Summary

What is Big Data?

- "Big Data" describes a collection of data sets so large and complex that they become difficult to process using traditional data processing applications.
- The trend to larger data sets is due to the additional information derivable from analysis of a single large set of related data
- Compared to separate smaller sets with the same total amount of data, allowing correlations to be found to "spot business trends, prevent diseases, combat crime improve transportation."
- Describes the exponential growth and availability of data, both structured and unstructured in terms of:
  - Volume
  - Velocity
  - Variety
  - Variability

Challenges

- Complexity Analysis
- Capture
- Curation
- Search
- Sharing
- Storage
- Transfer
- Visualization
- Privacy

The Big Data Value Chain

Big Data may be as important to business and society as the Internet since more data leads to more accurate analyses

Big Data - What's New? What's Not New?

New

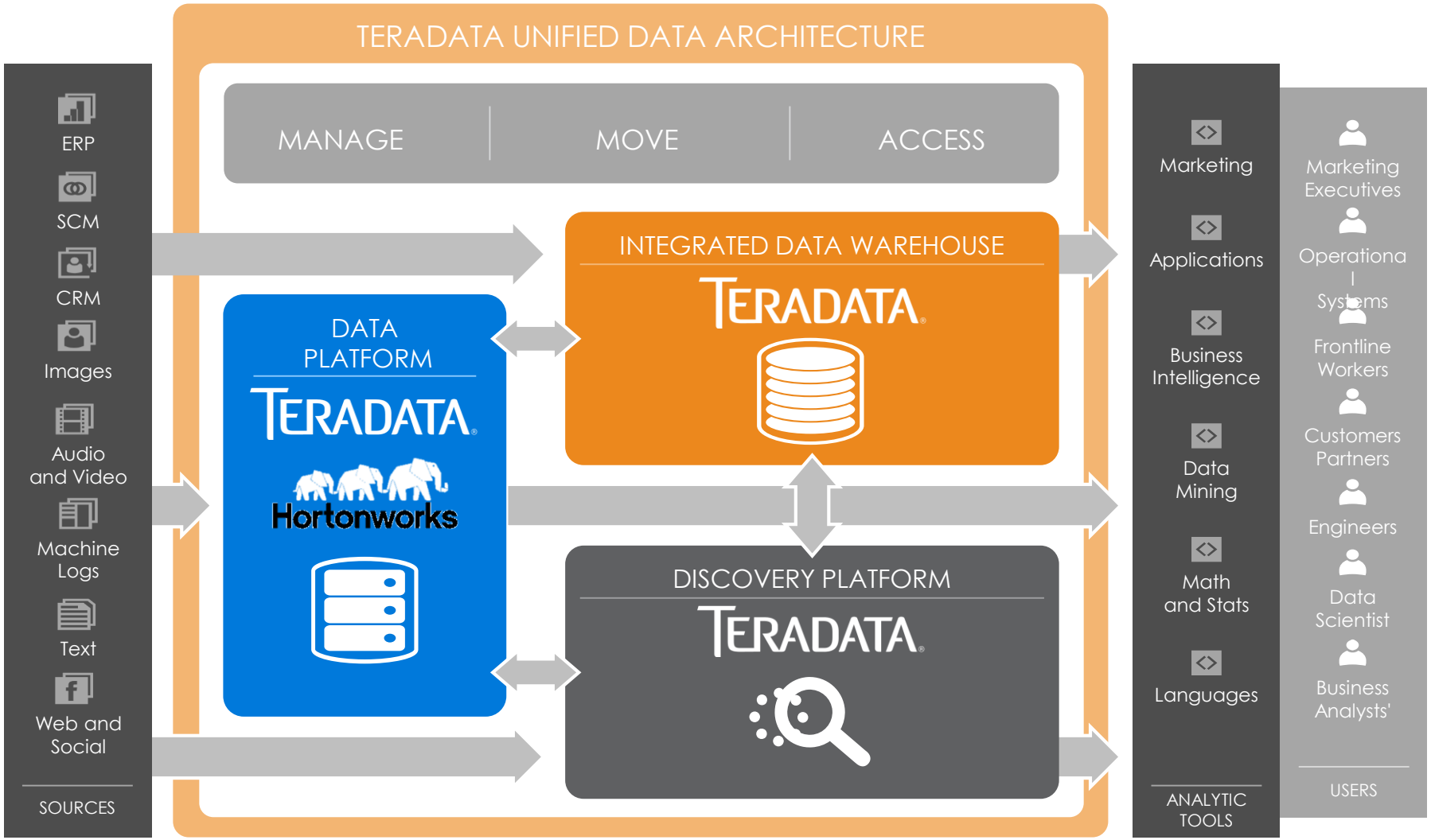
- Economics: increased amounts of data you can afford to capture
- Tools: uncover insights from new data types to quickly find the signal in the noise.
- Architecture Framework: a hybrid ecosystem that makes it easy to use both old and new tools on old and new data (i.e. use the right tool for the right job)
- New analytics on new, non-relational data types (coupled with existing relational data) open up new possibilities for unique insights

Not New

Most big data Use Cases are variations on the same things organizations and government agencies have been doing for years

The Teradata Unified Data Architecture™

- Any User, Any Data, Any Analysis



UDA Platform Roles

INTEGRATED DATA WAREHOUSE

- Single view of your business
- Cross-functional analysis
- Shared source of relevant, consistent, integrated data
- Data Re-Use: load once, use many times
- Lowest total cost of ownership
- Fast, new applications time-to-market

DISCOVERY PLATFORM

- Interactive Data Discovery
  - Web clickstream, social
  - Set-top box analysis
  - CDRs, sensor logs, JSON
- Flexible, evolving schema
- Structured & multi-structured data
- Multi-genre analytics
  - Patented SQL-MapReduce
  - SQL
  - Statistics
  - Text
  - Graph
- 100+ packaged analytic functions

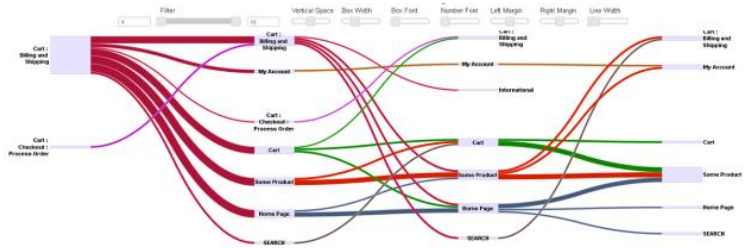
DATA PLATFORM

- Land/source operational data
  - Only one extract from source system
- History or long-term storage
  - Low cost storage
- Preprocess data
  - Sessionize data, remove XML tags
- Transformations
  - Structured and semi-structured data
- Simple math and batch processing
- $/Terabyte metric

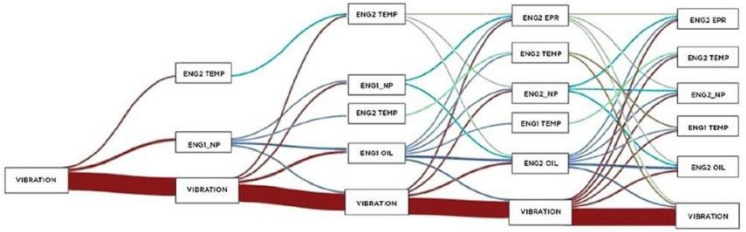
TERADATA.

Aster Discovery Platform

- Big Data Analytics Providing Competitive Advantage in Other Industries

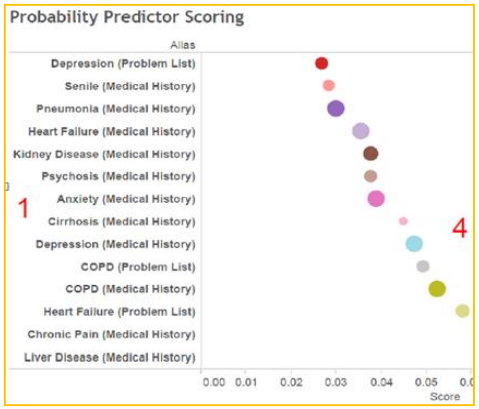


A bank uses Aster to perform cross-channel path analysis to detect fraud and to build a sequence of fraud score patterns to identify fraud better and faster

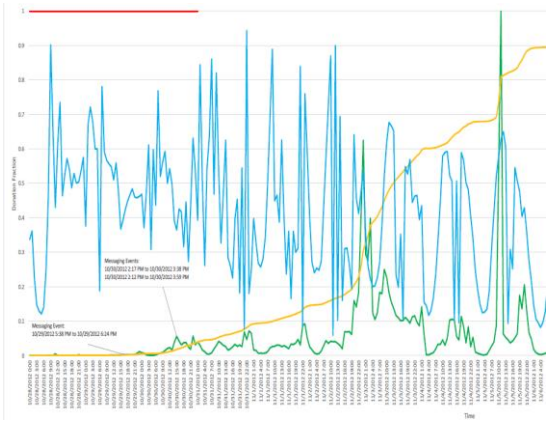


Teradata Aster *n*Path Viz sankey displayed vibration as a leading indicator and engine faults that followed the vibration event.

An operator with multiple aircraft fleets uses Aster to visually understand the key events and conditions that cause engine part failure, which has led to more accurate part failure predictions, enhanced preventative maintenance action, improved safety and lower operating costs



A large not-for-profit health care system uses Aster to "see" patterns of diagnosis/treatment that can drive better patient outcomes and help focus attention/assets on biggest cost drivers and issues



A humanitarian organization that provides emergency assistance and disaster relief uses Aster to analyze tweets to understand the nature and location of emergencies, better target volunteers, optimize blood/$ donations

A state DOT is evaluating Aster to analyze traffic patterns and to measure speed variability on major roads and arterials to improve traffic flows.

The "Internet of the Automobile"

- Connecting Big Data, Apps, Devices, and the Car to the Driver

Ford Fusion Energi plug-in hybrid
"has more than 145 actuators, 4,716 signals, and 74 sensors to monitor the perimeter around the car as well as the car's functions and driver responses.

These sensors produce more than 25 gigabytes of data hourly from more than 70 on-board computers that analyze it in real-time."

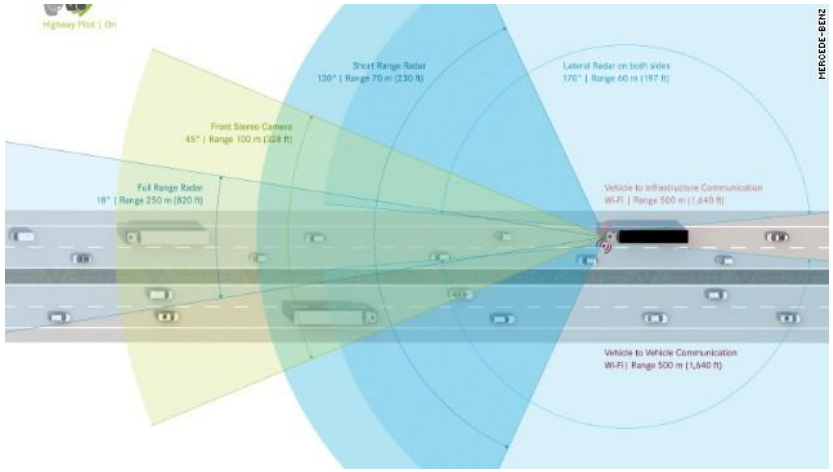
2 Zetabytes per year across the USA

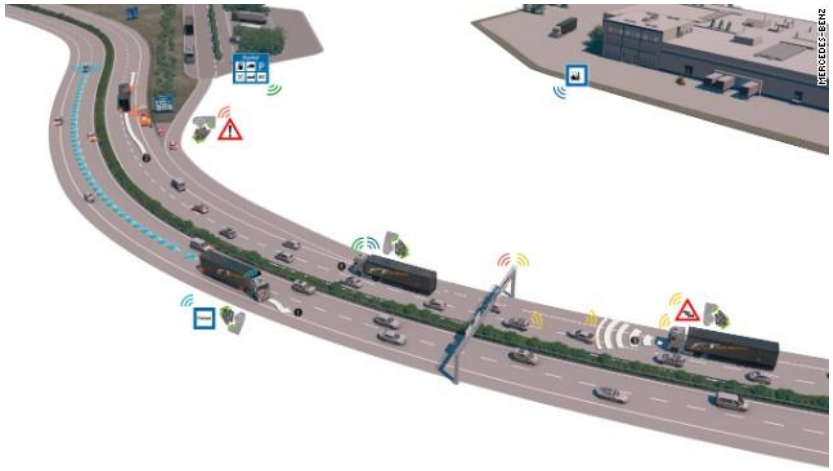
Google's autonomous vehicle: 1GB/Sec

The Future of Freight

"Future Truck 2025" is envisioned to communicate with other vehicles and connect to growing sources of online information as Big Data balloons on the road

- Special cameras, wireless LAN, and multiple radar systems watch the road, the sides of the road, and cars and trucks behind the vehicle
- [Future Truck's] computerized controls will make it more fuel efficient
- …Two cameras and a sensor under [the trucker's] seat will monitor his/her activities

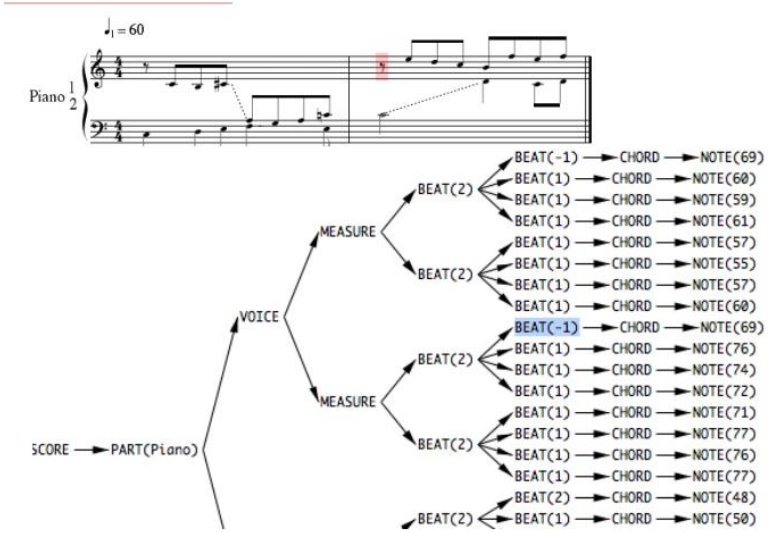




Source: "Truck of the future aims to drive itself," CNN, Jul 6, 2014
http://www.cnn.com/2014/07/04/tech/mercedes-future-truck/index.html?c=tech

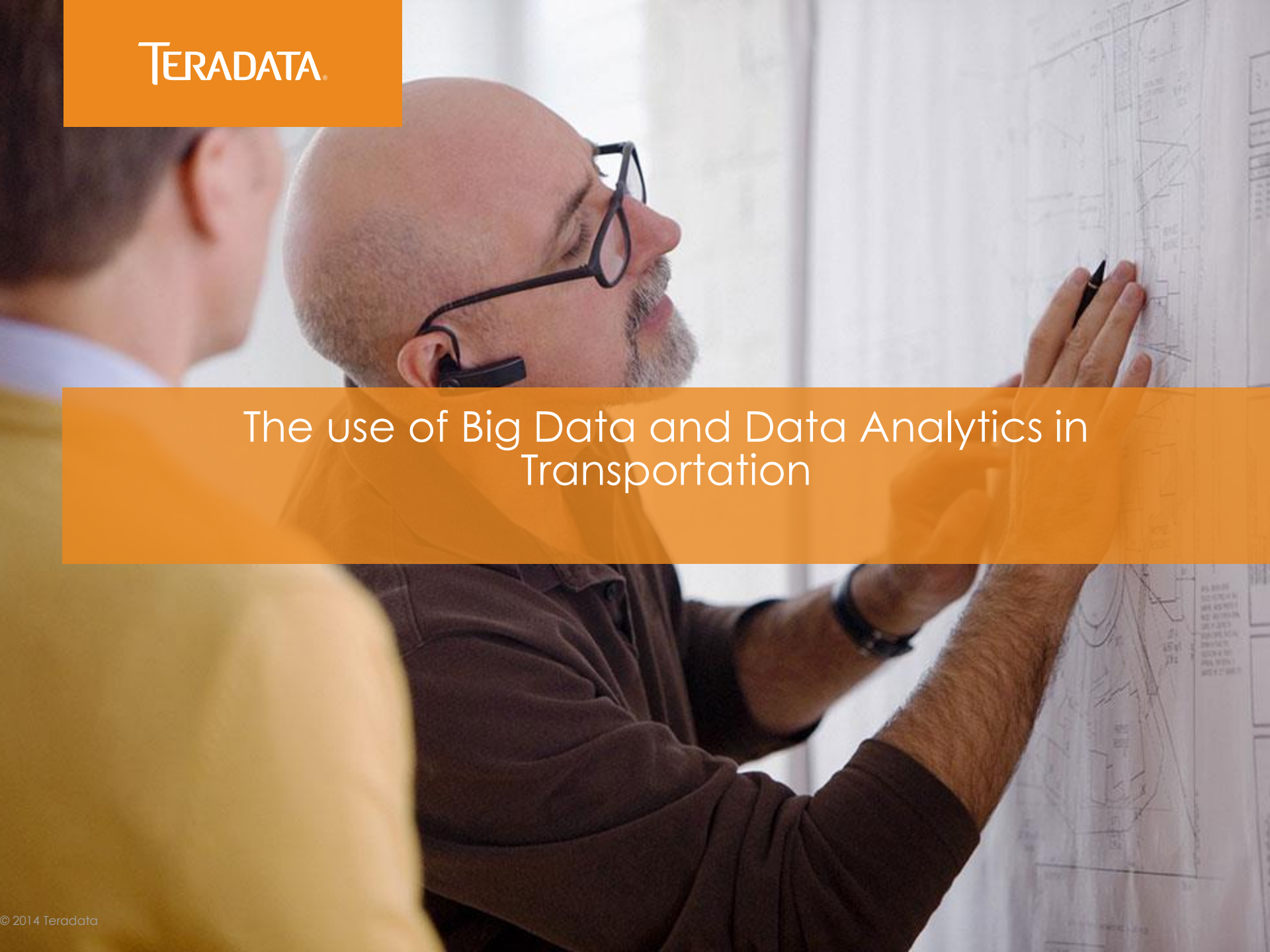
Bits and Bytes to Music

Data



Information

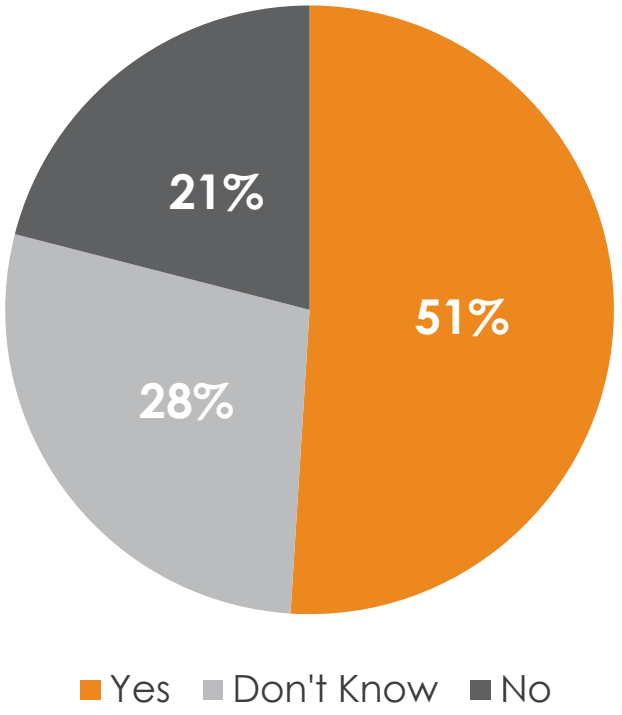
Can you hear the music in your data?



The use of Big Data and Data Analytics in Transportation

Why Big Data and Transportation Data Analytics?

- We are moving from data poor to data rich
- Transitioning from data averse to data hungry
- Advances in sensors, telecommunications and the connected vehicle are driving a new wave of data
- Many of the challenges have already been addressed outside of transportation
- We are entering a "results-driven" era in transportation
- If you can't measure it you can't manage it and if you are just measuring it, you are still not managing it
- Need for decision quality information
- Movement towards management of transportation as a system
- States will spend approximately $128.4B on transportation in 2014
- 2014 Estimated IT Spend: $9.6B



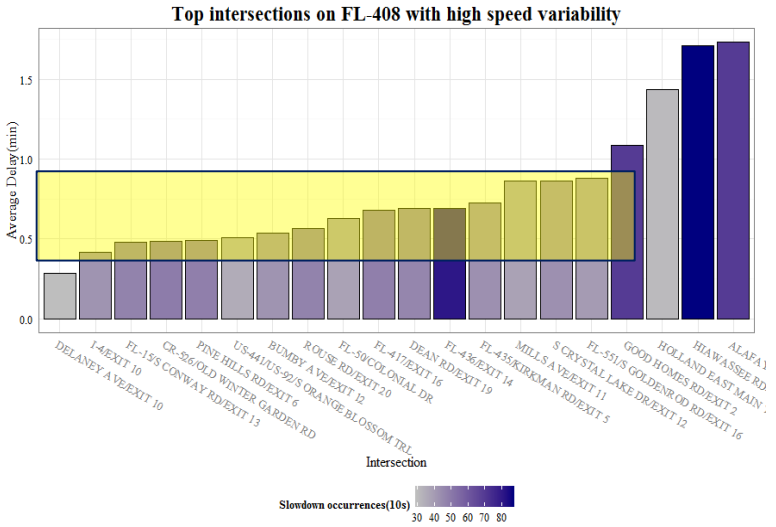
My agency or department will have an increased focus on ITS within the next two years

General Questions

- How effective are our transportation investments?
- Are travelers getting a fair deal?
- Are our ITS investment programs results driven?
- Can we determine the status of our transportation networks and services at any given time?
- Do we understand the mechanisms and patterns that underlie transportation in our jurisdiction?
- Are we collecting data once and using it many times, or duplicating data collection efforts and submerging data in silos?
- Does our entire organization have easy access to information required to support work processes?
- Is our data collection and acquisition optimized?
- What is the current value proposition in terms of safety, efficiency and customer service?
- Do we understand public perception of transportation service delivery at any given time?

Specific Questions

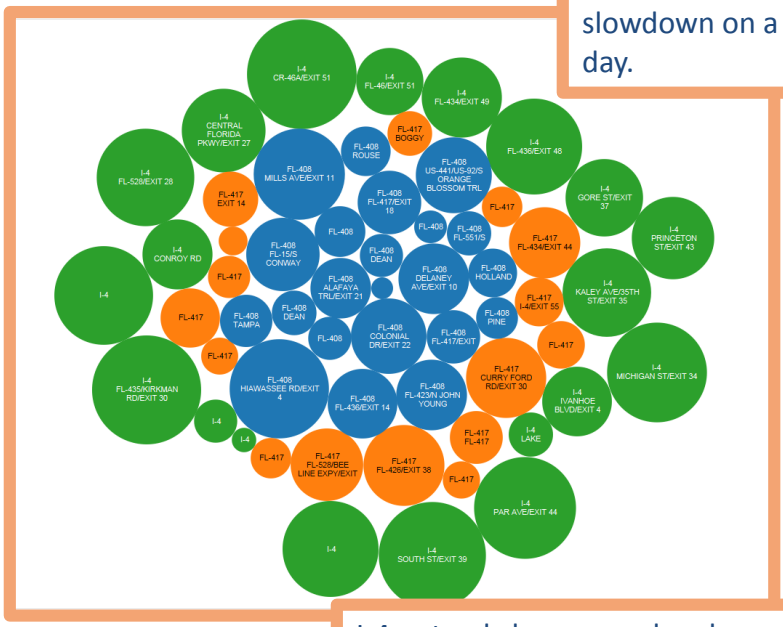
- What are the causal mechanisms for pedestrian fatalities?
- How many slow downs are there every day and how long is the recovery period?
- What is the trip time and trip time reliability by various modes?
- When we plan and design, how well do we capture previous experience?
- Is there a relationship between maintenance quality levels and safety?
- Can we measure the basic performance measures consistently and comprehensively?



TERADATA.

Example Applications

- Trip time reliability analysis
- Connected vehicle data analysis
- Freight door to door delivery times
- Pedestrian safety
- Public perception of transportation services



The size of the bubble represents the amount of time it takes for each one of these intersections to recover from a slowdown on a given day.

I-4 not only has more slowdowns, but also has longer recovery times than SR408/SR417

TERADATA.


TERADATA
Summary (Wrap up)

Summary

- Big Data and analytics techniques are proven and available now
- We need to get ready to be data hungry
- There are trends and patterns to be discovered through analysis of multiple data sets
- We can monetize Big Data for transportation in terms of safety, efficiency and customer service in planning, design and operational management activities
- Harnessing the power of Big Data, Data Science and Transportation Data Analytics we can improve planning, design and service delivery for transportation