

TITAN – Interactive, Web-Based Platform for Transportation Data Integration and Analytics

Yaw Okyere Adu-Gyamfi, PhD.

Carlos Sun, PhD, PE, JD.

Praveen Edara, PhD, PE.

Civil and Environmental Engineering.

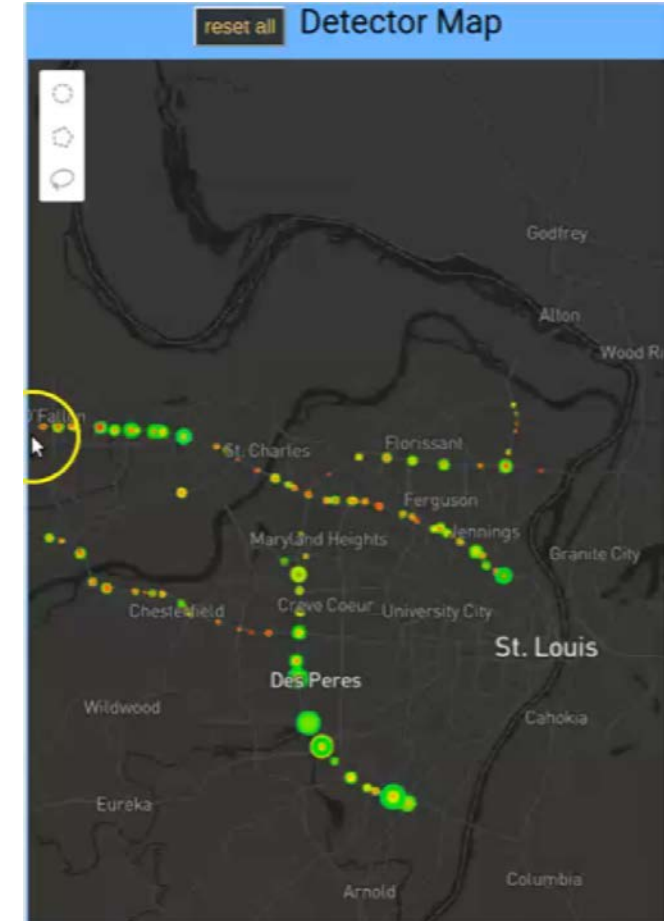
University of Missouri – Columbia

Presentation Outline

- Motivation
- Big Data Challenges
- Titan Platform Architecture and Solution
- Demo

How Did We Get Here?

Count - Measure- Alert



Download
Data

Prune, Filter,
Merge

Analyze,
Model

Visualize,
Communicate

How Did We Get Here?



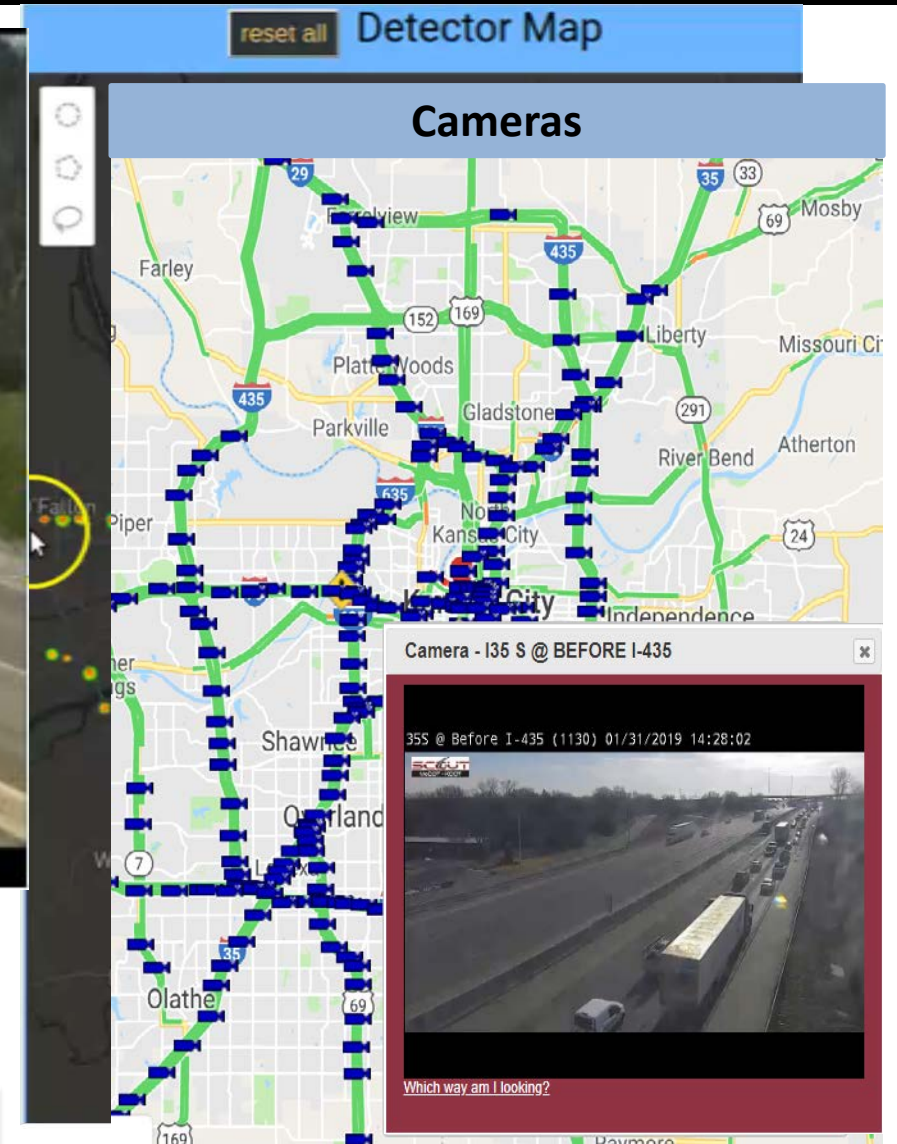
Limitation of sensors

Download
Data

Prune, Filter,
Merge

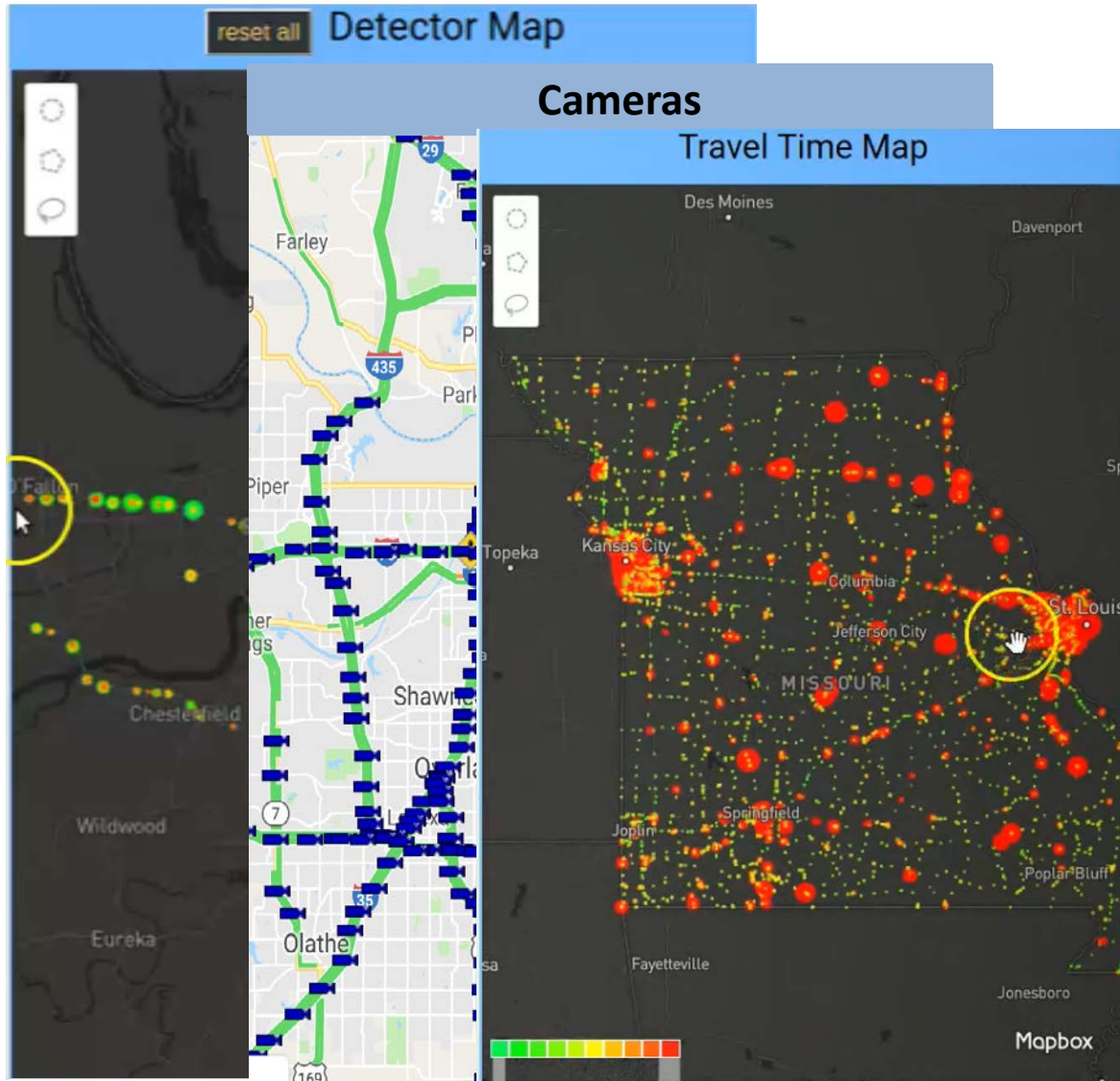
Analyze,
Model

Visualize,
Communicate



Eyes on the road

How Did We Get Here?



Increase Coverage

Cameras

Verify, Monitor,
Understand

Sensors

Monitor Selected
Roadways

Probe Data

Increase
Coverage –
Travel Times

How Did We Get Here?

Big Data Problem

1500 CAMERAS

Cameras

Verify, Monitor,
Understand

5GB

Probe Data

Increase
Coverage

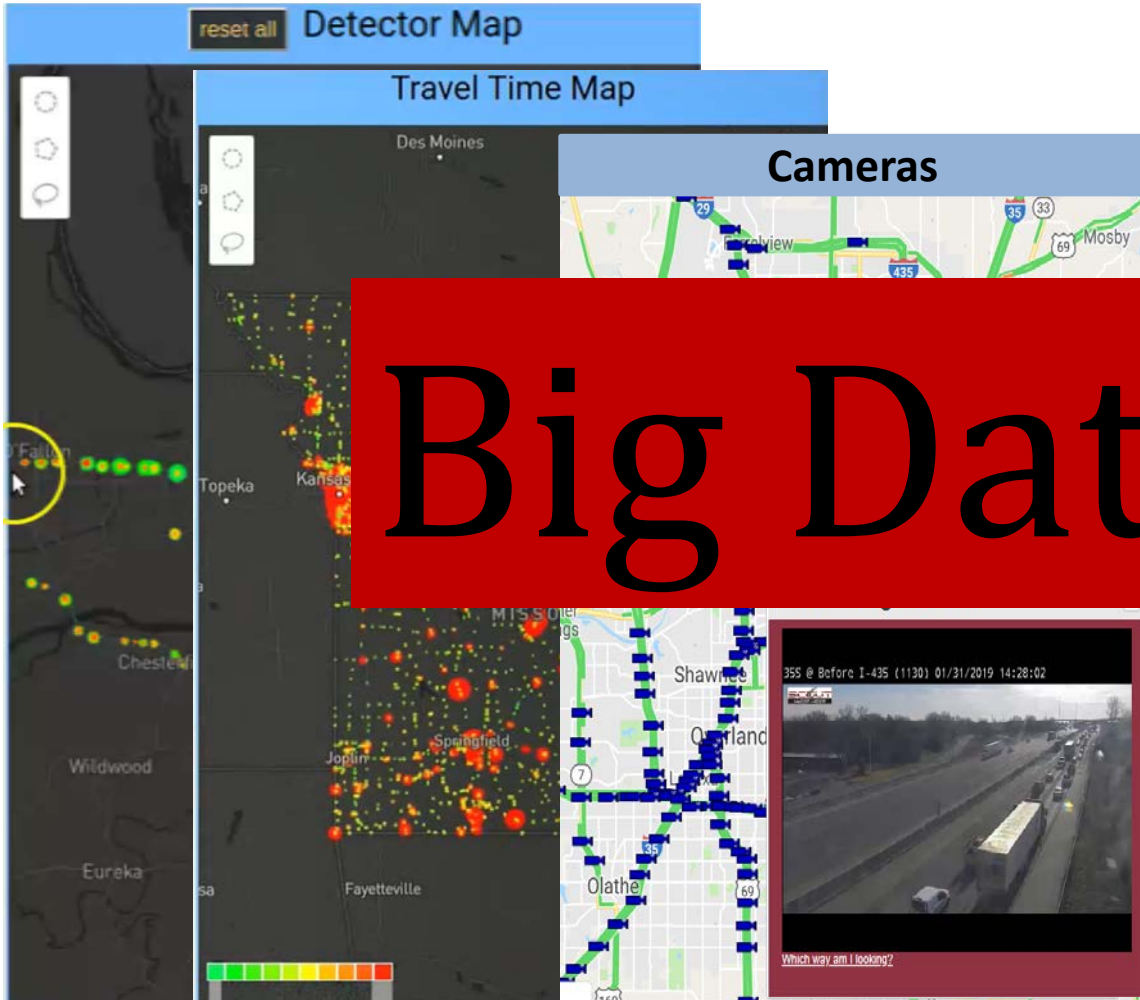
Monitor Selected
Roadways

~~Download
Data~~

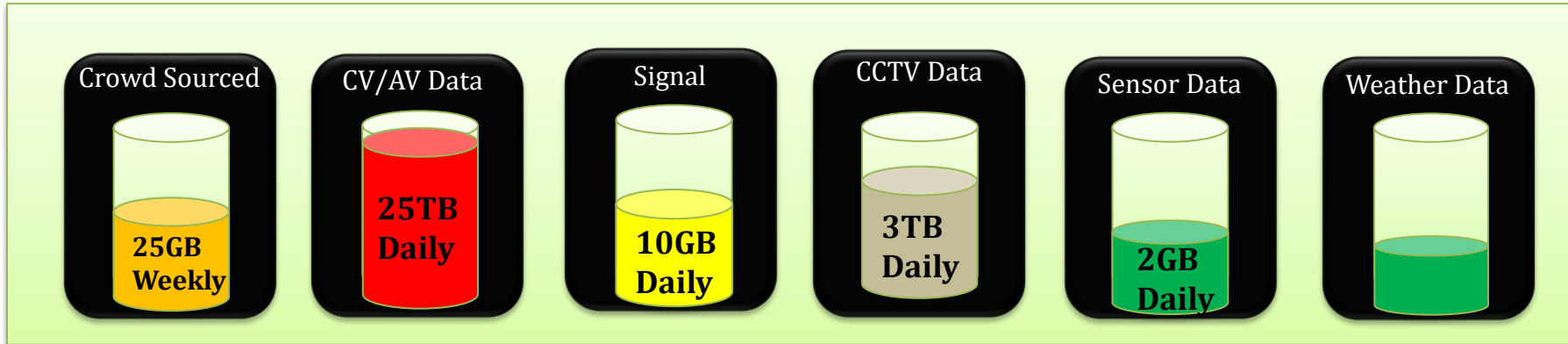
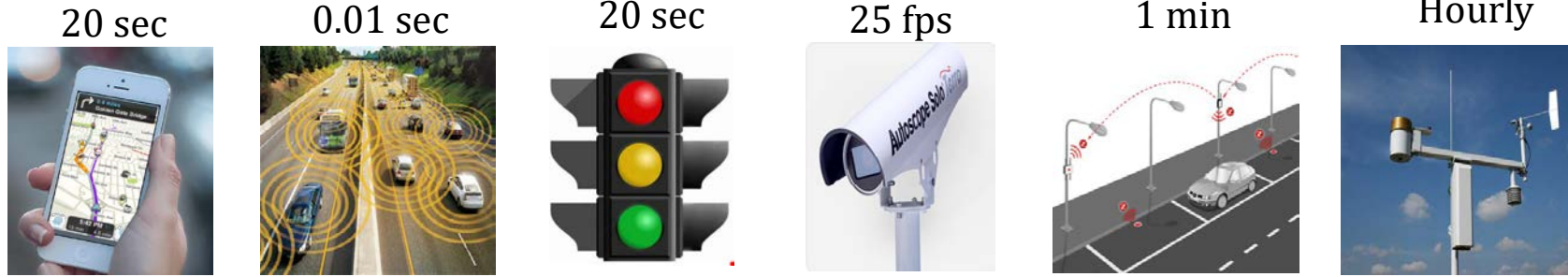
~~Prune, Filter,
Merge~~

~~Analyze,
Model~~

~~Visualize,
Communicate~~



Big Data Challenges - Storage



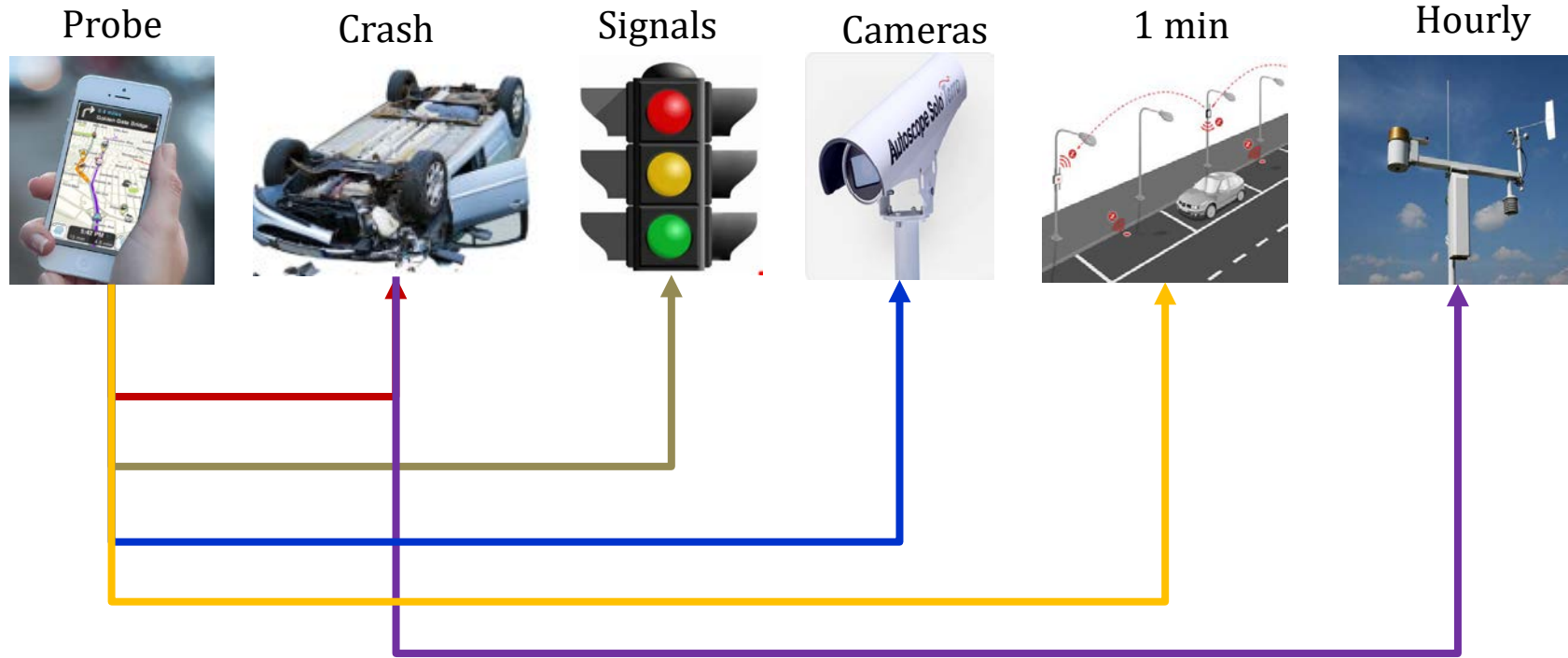
1. Volume - Velocity

- Quick storage, fast retrieval

2. Variety

- Structured Data – CSVs, relational databases
- Unstructured Data – Images, Video, Text.

Big Data Challenges - Data Integration



**Vendor - Driven
Network Level**

Answer New Questions

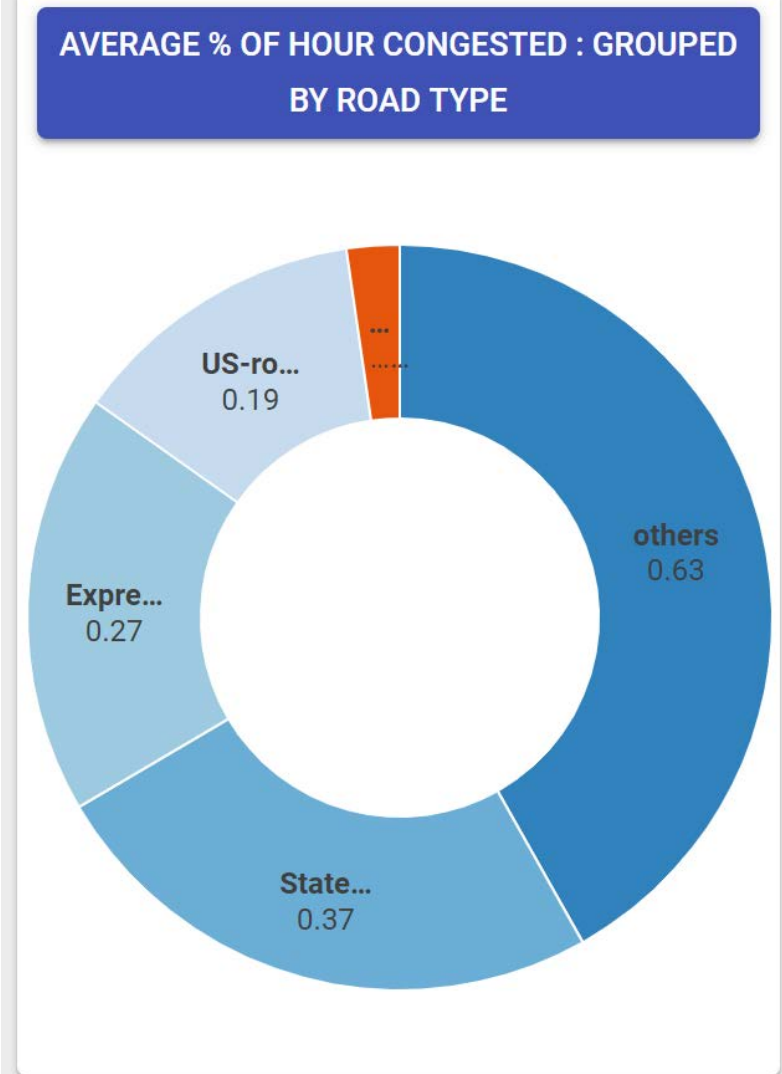
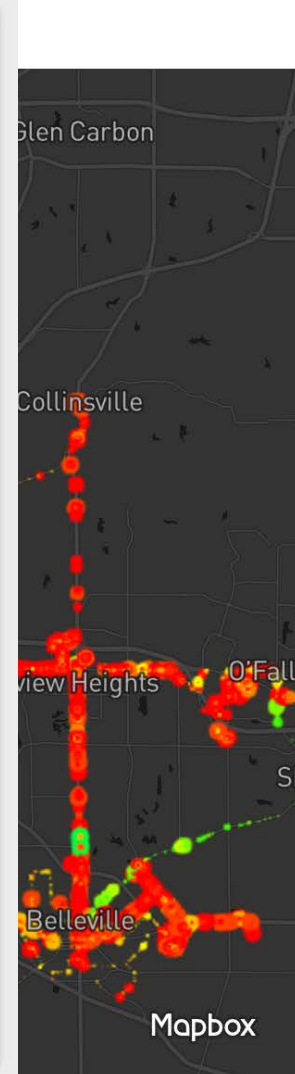
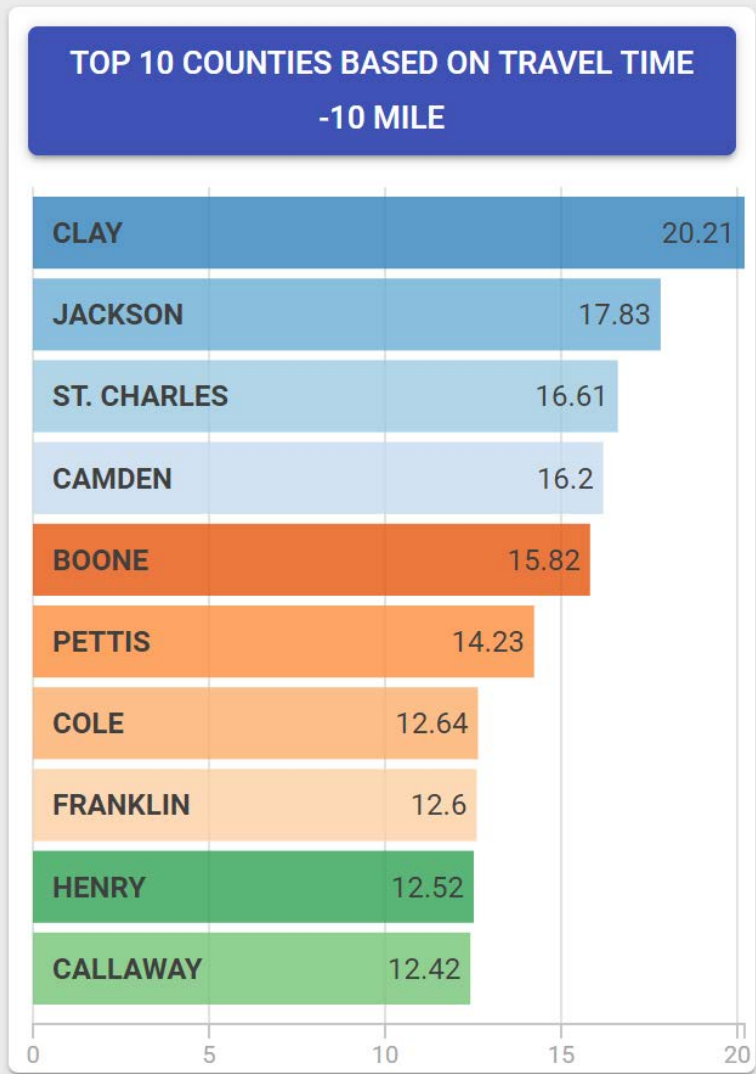
1. Integrate or Conflate Probe-Crash-Traffic Sensors-Cameras

- Explain congestion
- Impact of crashes on traffic flow

2. Integrate Probe – Signals – Weather – Crashes

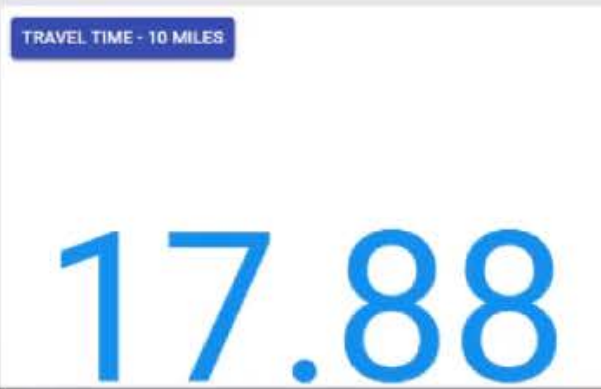
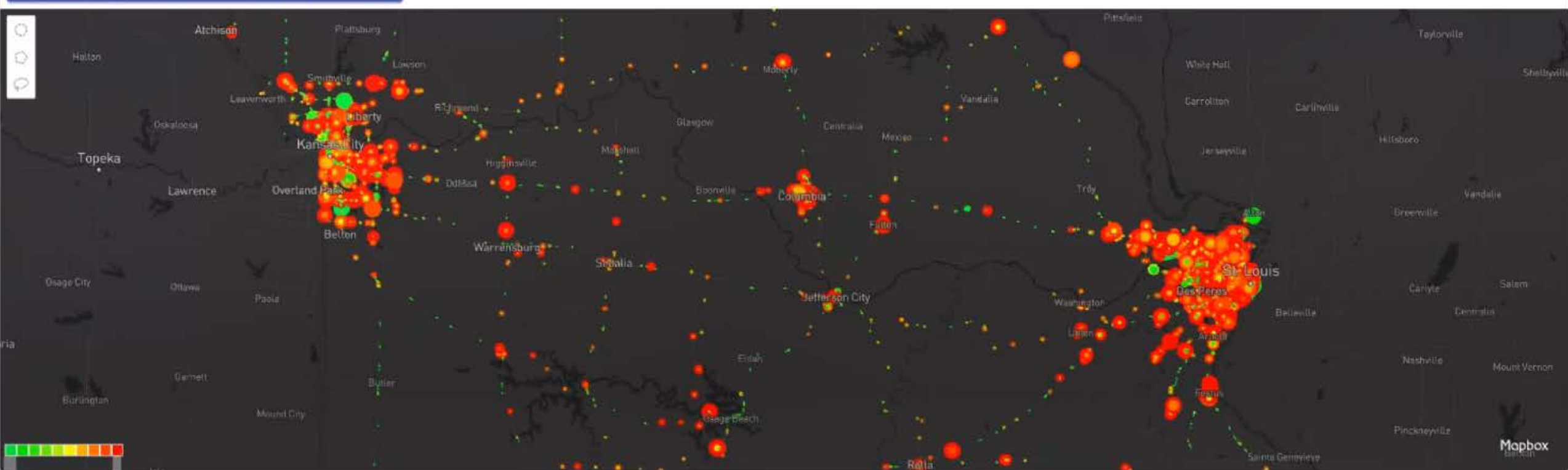
- Impact of weather events on traffic flow or crashes
- Separate controlled and uncontrolled delays

Big Data Challenges - Visualization



Big Data Challenges – Interactive Analytics

STATEWIDE TRAVEL TIME MAP: COLORED BY SPEED || SIZE BY TRAVEL TIME



TITAN Platform Architecture

INTERACTIVE DASHBOARDS

Integrate, Visualize, Interact and Analyze Big Datasets in the Web browser

DATA CENTER

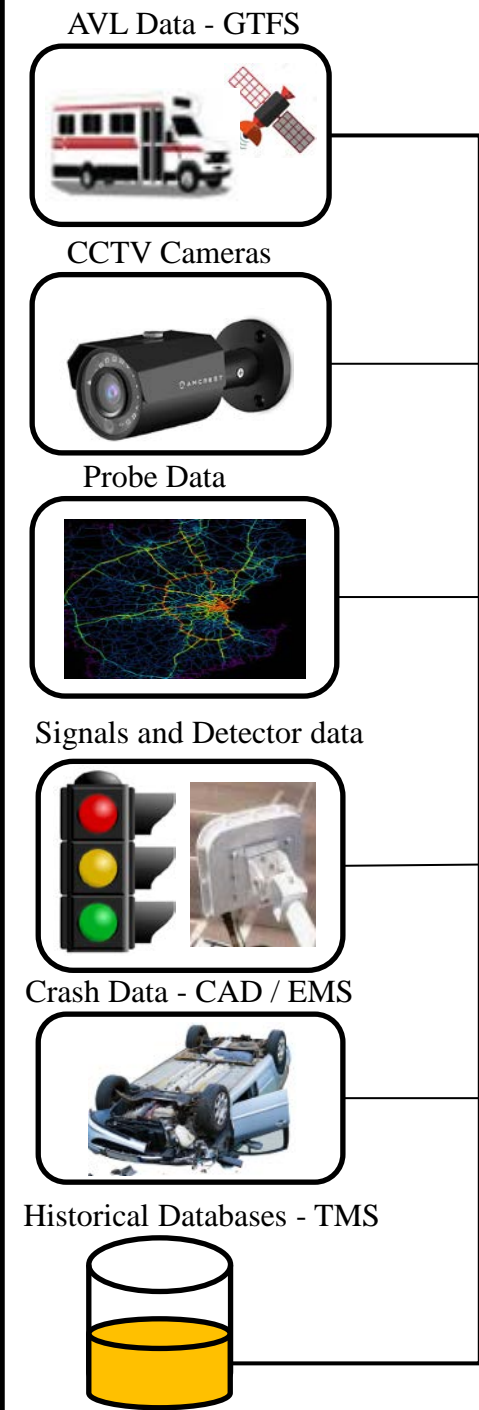
Upload – Query Datasets

PREDICTIVE ANALYTICS

Machine Learning || Video Analytics

Data Ingestion Engine

1. Automatically pull data from different sources simultaneously.
2. Store acquired data on the cloud

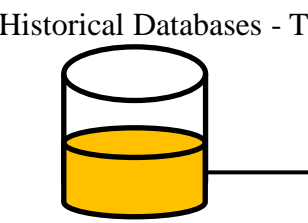
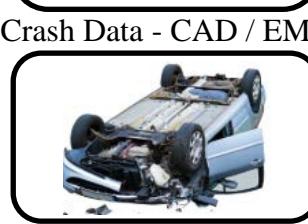
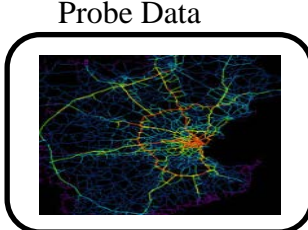
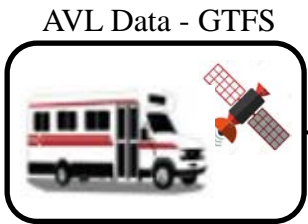


Cloud Environment

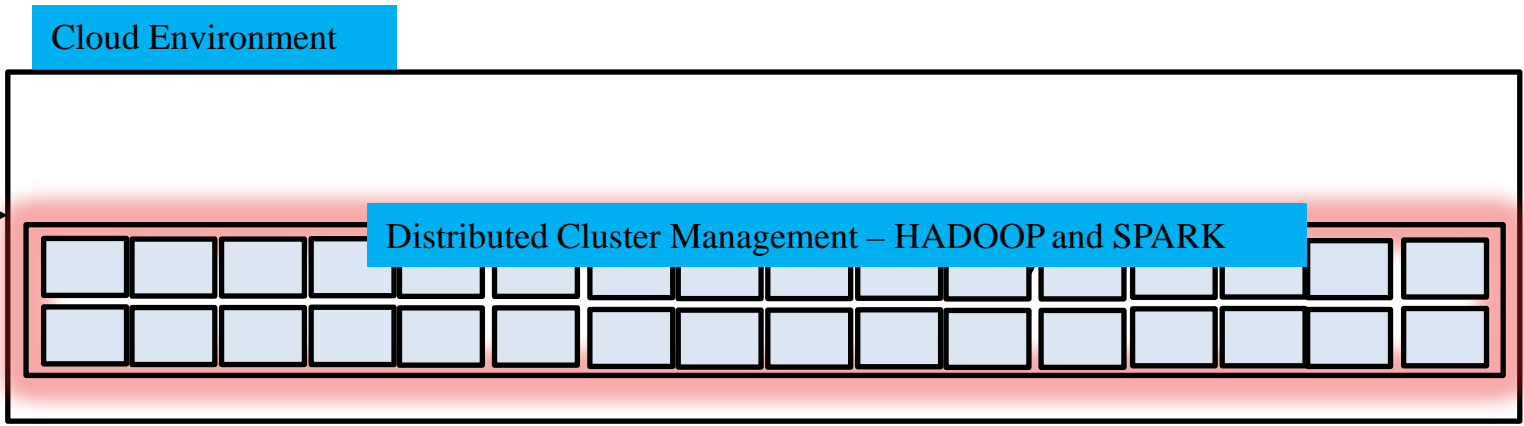
Data Ingestion Engine – Distributed Storage

Goal: Fast storage, download / querying of Big Data.

1. Store data on a cluster of machines
2. Big data softwares used to enable parallel processing of data – **Spark** and **Hadoop**.



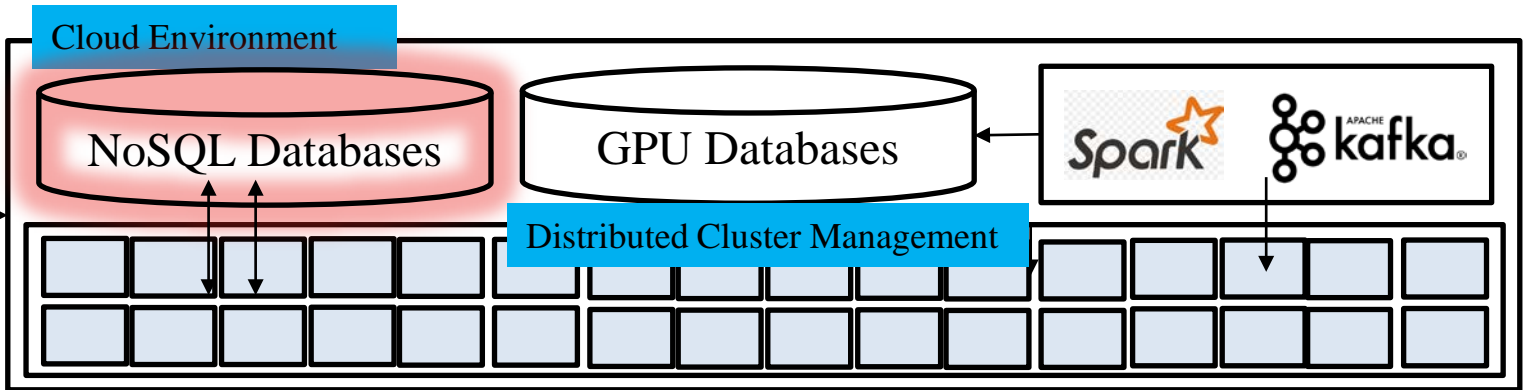
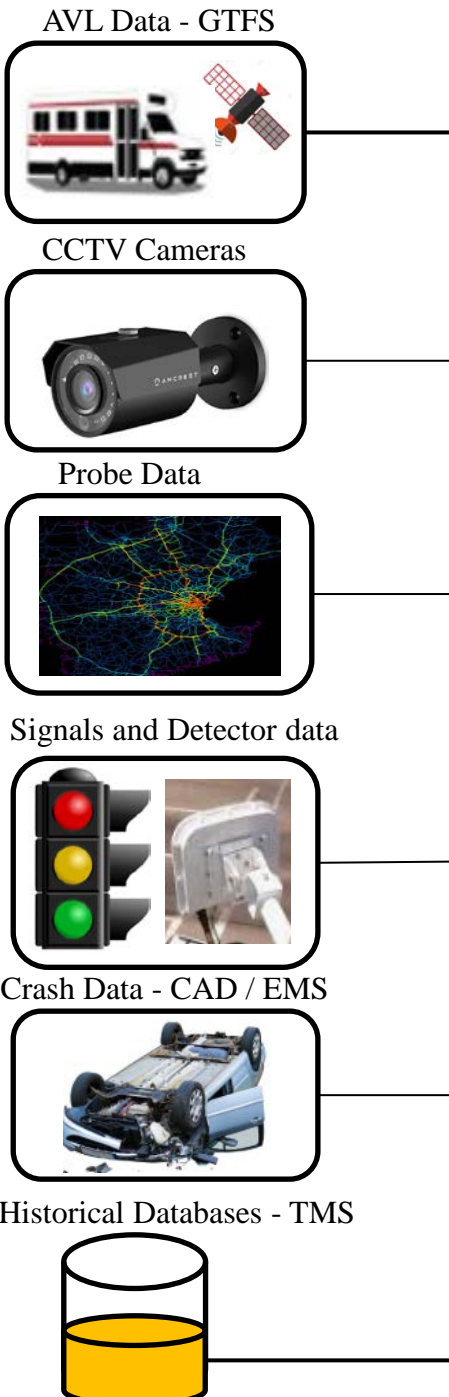
Data Ingestion Engine



Back – End and Database Design

Providing Access

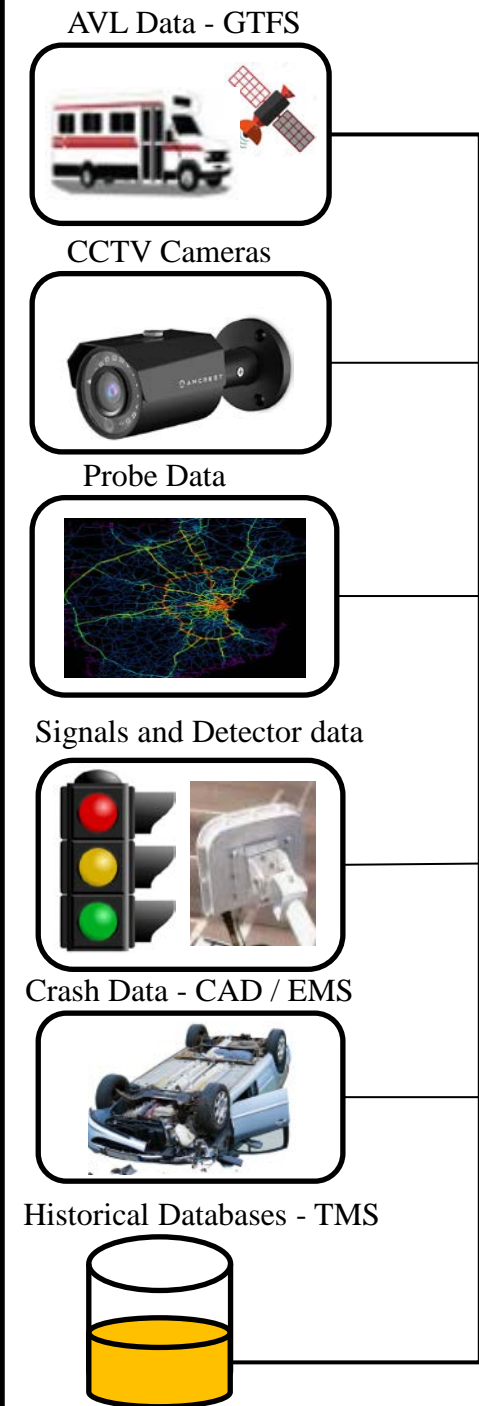
- Real-time Querying
- In-memory Databases – Cassandra, MongoDB (25GB or less).
- GPU Databases – Fast processing of medium-sized datasets – less than 100GB
- Types of Data Stored
 - Video, Images, JSON, XML, Text, Tables (CSV)



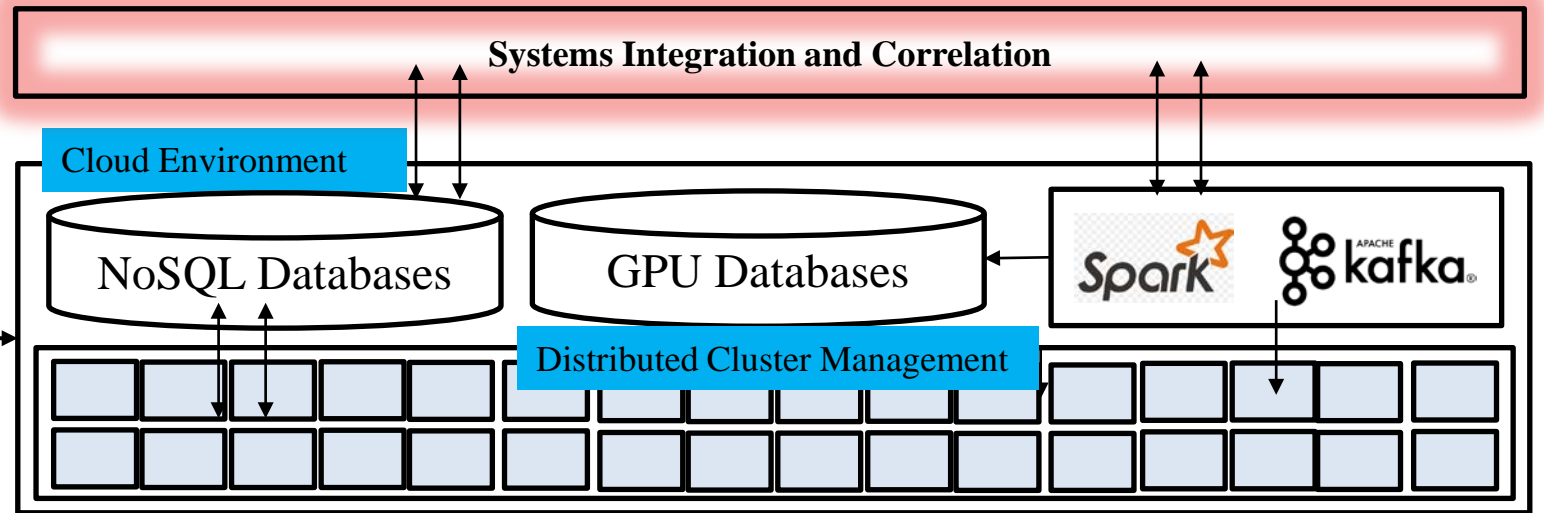
Back - End and Database Design

Data Integration or Conflation

- **Integrate different datasets in space and time**
 - Probe – Crash – CCTV
 - Probe – Sensor – Crash – Weather
 - Probe – Workzones - Weather
- **Automated with manual checks**



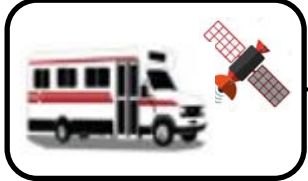
Data Ingestion Engine



Back - End and
Database Design

APP CENTER

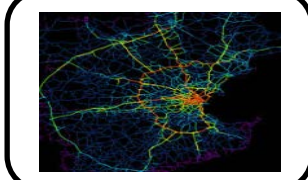
AVL Data - GTFS



CCTV Cameras



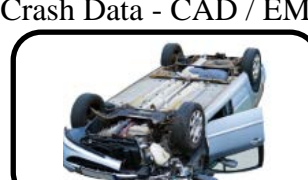
Probe Data



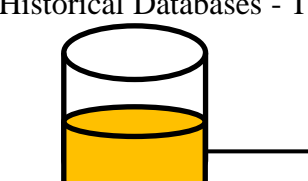
Signals and Detector data



Crash Data - CAD / EMS



Historical Databases - TMS



Data Ingestion Engine

APP CENTER

Predictive Analytics Performance Assessment Dashboards - Reports Query Data

Systems Integration and Correlation

Cloud Environment

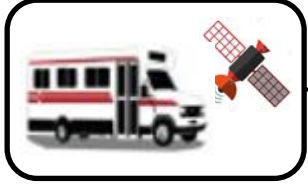
NoSQL Databases GPU Databases Spark kafka

Distributed Cluster Management

Back - End and Database Design

Data Management, Security and Governance

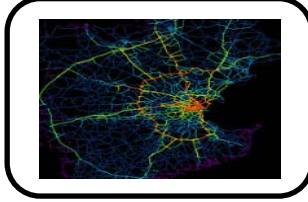
AVL Data - GTFS



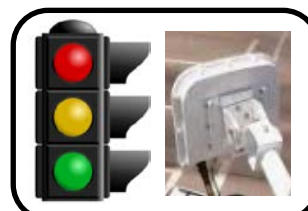
CCTV Cameras



Probe Data



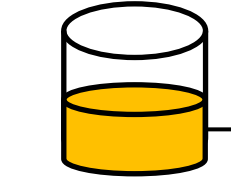
Signals and Detector data



Crash Data - CAD / EMS



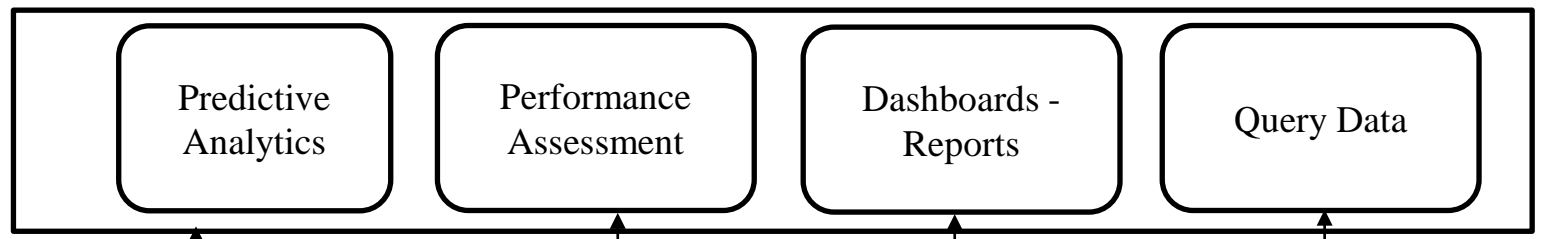
Historical Databases - TMS



Data Ingestion Engine

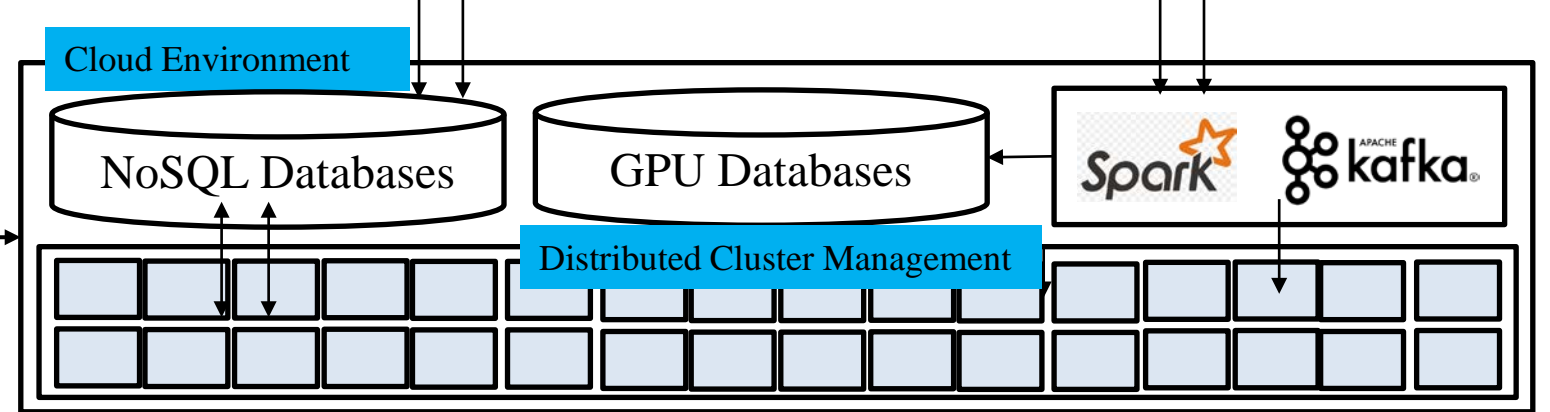


Interactive FRONT-END



APPCENTER

Systems Integration and Correlation



Back - End and Database Design

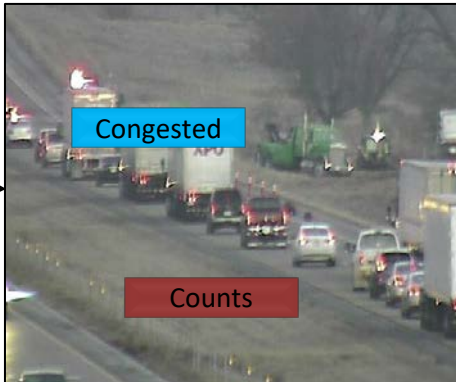
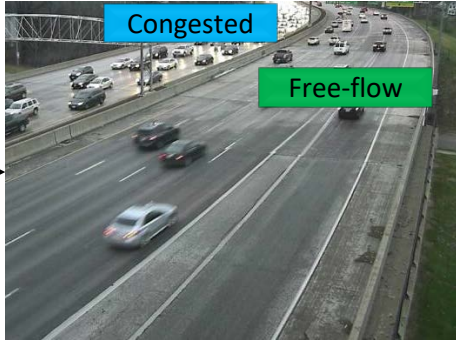
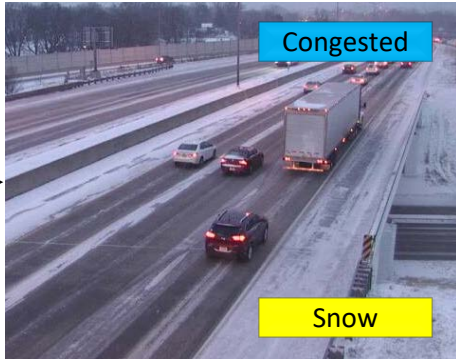
Data Management, Security and Governance

Demo - TITAN

- <https://mizzou-titan.com>

Video Analytics

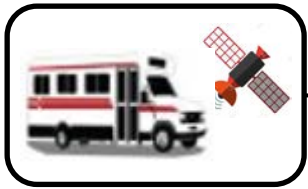
TITAN



```
{  
  "articles": [  
    {"largeImageURL": "1.jpg",  
     "tags": ["snow", "congested"],  
     "camera_id": "I-80/32",  
     "id": 11212,  
     "counts": 18,  
     "time": "am",  
     "duration": 15,  
     "severity": "medium"}, |  
    {"largeImageURL": "2.jpg",  
     "tags": ["congested", "dry"],  
     "camera_id": "I-30_AMES",  
     "id": 3432132,  
     "counts": 5,  
     "time": "12:15pm",  
     "duration": 2,  
     "severity": "low" },  
    {"largeImageURL": "3.jpg",  
     "tags": ["medium flow"],  
     "camera_id": "US50123",  
     "id": 876978,  
     "counts": 12,  
     "time": "pm",  
     "duration": 45,  
     "severity": "high" }  
  ]  
}
```



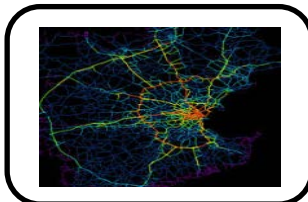
AVL Data - GTFS



CCTV Cameras



Probe Data



Signals and Detector data



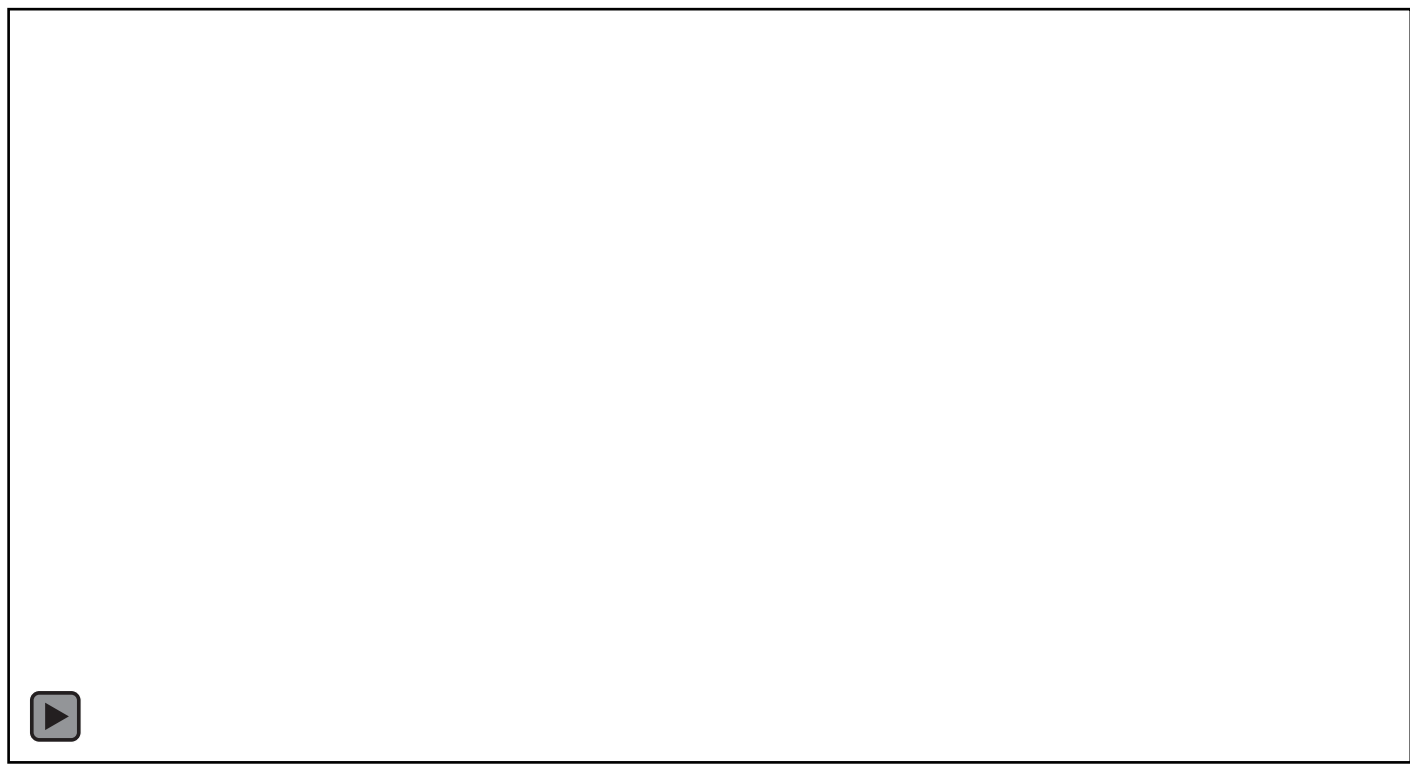
Crash Data - CAD / EMS



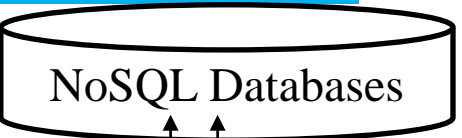
Historical Databases - TMS



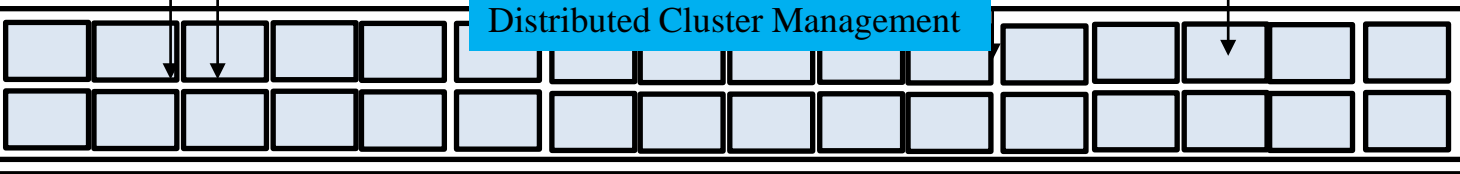
Data Ingestion Engine



Cloud Environment



Distributed Cluster Management



Back - End and Database Design

Coming Soon ...

- Additional Data Sources
 - Bridge, Pavement, Geotechnical assets, etc.
- Predictive Analytics
 - Infrastructure for running machine learning models
 - Integrated Data for training models

Cloud vs In-House

- Costs
- Administration
- Security
- flexibility

Summary

- **TITAN uses big data to**
 - Provide an efficient way to **store** and **retrieve** large datasets
 - **Integrate** and **link** different sources of transportation datasets in a spatio-temporal manner
 - Enable users to **interact**, **visualize** and **analyze** big data in the web-browser at fast speeds.