Traffic Operations Open Data Service

Skylar Knickerbocker
2022 Mid-Continent Transportation Research Symposium
Data Services

Continue to monitor cost and services then migrate as needed

Primary services are:

On-Premise Cluster

Cloud (Azure then AWS)
On-Premise Cluster

- 40 Node Hadoop cluster
- LSS as primary storage (cold storage)
- Presto as Query Engine
- 2 GPU Cluster
Primary On Premise Function

- Camera Snapshots (5-minute back to 2016)
- Deep Learning using Traffic Cameras (GPU Cluster)
- Crash/ATMS Video Download
Cloud (Azure then AWS)

Computation
- Two Linux Servers
- One Windows Server
- Glue and EMR (PySpark)
- Jupyter Notebook
- Athena (Query Engine)

Storage
- S3 (35 TB)
- MySQL Database
- PostgreSQL Database
- Compressed Format (parquet and ORC)
Establish Cloud Data Service

- Reactor Feeds
- AWS

Diagram:
- Third Party Data Servers
  - Data Consumption
    - Data Generation
      - Third Party Consumer
      - Third Party Consumer
      - Third Party Consumer
  - Data Archival
    - Hard Disk
      - Cell Phone
Establish Cloud Data Service

Serverless ETL for streaming data such as probe, sensor, etc.

Data Storage

Serverless interactive query service using SQL that executes queries in parallel

Data Visualization for providing access to data for end users
There is value to having the data in one location...and pushing this out for multiple uses.
Sensor Data

Real Time Feed
WZ Text Alerting
Work Zone Performance
LCPT History

Similar Identification Needs:
-LRS Information
-Work Zone Related
-Alerting (need higher data quality)
Product Services

Data Archival

Text Alert Module

Data Feeds
Archiving Data Sources

16 data sources in original service

• Traffic detector data (20-60 seconds)
• INRIX data (one minute)
• DMS inventory data (Daily)
• Work zone congestion events
• Waze data (one minute)
• ATMS data (real-time or daily updates)
Additional data sources

- Smart Arrow Boards/cTTCD
- RWIS
- Snowplow
- Truck Parking
- Weather Data (Mesonet, NCAR)
- Wejo/Connected Vehicle
- Camera Snapshots/Video
# Archiving Data Sources

<table>
<thead>
<tr>
<th>Data</th>
<th>Memory</th>
<th>Per month</th>
</tr>
</thead>
<tbody>
<tr>
<td>INRIX probe data</td>
<td>3.4 GB/day</td>
<td>102 GB</td>
</tr>
<tr>
<td>Wavetrionix sensor data</td>
<td>1 GB/day</td>
<td>30 GB</td>
</tr>
<tr>
<td>Waze data</td>
<td>330 MB/day</td>
<td>10 GB</td>
</tr>
<tr>
<td>Weather data</td>
<td>17.9 GB/day</td>
<td>537 GB</td>
</tr>
<tr>
<td>Wejo – Vehicle Movement</td>
<td>15 GB/day</td>
<td>450 GB</td>
</tr>
<tr>
<td>Wejo – Driver Event</td>
<td>0.2 GB/day</td>
<td>6 GB</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>22.6 GB/day</strong></td>
<td><strong>1.1 TB</strong></td>
</tr>
</tbody>
</table>
Data Feeds

- Unified Sensor Feed
- Real-time work zone slowdowns
- DMS location history
- LCPT Sensor Data
2.6 User Documentation
Online documentation can be found at http://reactorfeeds.org

2.7 Assumptions and Dependencies
As specified earlier, this application depends heavily on the third party servers to get the data. We assume that the developer has the access for all the external components. Below provides the list of external dependencies.

1. Wavetronix Data API
2. Inrix API
3. Waze Data API
4. DMS Inventory API
5. FTP accessibility for ATMS data.
6. External application for identifying the work zone traffic status.
7. Twilio SMS Gateway

4 Functional Requirements

The features are split into three major categories: data archival, work zone congestion alerts, and the data feeds. Each section includes the requirements that specify all the fundamental actions of the software system.

4.1 Data Archival

4.1.1 Traffic Detector Inventory Data

ID: FR1

TITLE: Download traffic detector inventory data.

DESC: Application should be able to access the Wavetronix Data API and download the inventory data using the HTTP GET or POST method. The system should convert the downloaded XML data into CSV format and archive that CSV file into the file system. The download should happen once a day.

PRIORITY: High

RAT: In order for the system to archive the inventory data.

DEP: None
Data Feeds

LCPT Sensor Data
Lane Closure Planning Tool

- Data needed for Lane Closure Planning Tool

- Automated feed which updates once a month
Work Zone Text Alerts

✔ Previously - Text Alert processing and distribution by InTrans

✔ Now – Data processing by InTrans with real-time feed. Distribution by Iowa DOT

✔ Has been leveraged by Iowa DOT for other alert distributions
Smart Arrow Boards

- Data Archived back to 2020
- Provided test bed for integration and real-time REST services
- REST services able to be imported into ATMS
- Data now directly updating work zones in ATMS
Alternate Data Flows

MySQL

✓ Utilize common database for sharing data

✓ Iowa DOT pushes snowplow and ATR data

✓ Iowa State pushes truck parking data
Camera Snapshots

- Camera images archived every 5 minutes
- FTP site for accessing images
- Dashboard for easier viewing of images
Crash/WZ Videos

- Crash video downloaded for each ATMS event nightly
- GIFS available for public
- Full video available with approved access
- Not intended as replacement for video download request
Data Services

Continue to monitor cost and services then migrate as needed.

Primary services are:

On-Premise Cluster

Cloud (Azure then AWS)
Know where data is

Know how to access data
<table>
<thead>
<tr>
<th>Data</th>
<th>Memory</th>
<th>Per month</th>
</tr>
</thead>
<tbody>
<tr>
<td>INRIX probe data</td>
<td>3.4 GB/day</td>
<td>102 GB</td>
</tr>
<tr>
<td>Wavetronix sensor data</td>
<td>1 GB/day</td>
<td>30 GB</td>
</tr>
<tr>
<td>Waze data</td>
<td>330 MB/day</td>
<td>10 GB</td>
</tr>
<tr>
<td>Work zone events</td>
<td>1 MB/day</td>
<td>0.03 GB</td>
</tr>
<tr>
<td>Weather data</td>
<td>17.9 GB/day</td>
<td>537 GB</td>
</tr>
<tr>
<td>Wejo – Vehicle Movement</td>
<td>15 GB/day</td>
<td>450 GB</td>
</tr>
<tr>
<td>Wejo – Driver Event</td>
<td>0.2 GB/day</td>
<td>6 GB</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>22.6 GB/day</strong></td>
<td><strong>1.1 TB</strong></td>
</tr>
</tbody>
</table>
Data Services

Continue to monitor cost and services then migrate as needed

Primary services are:

On-Premise Cluster

Cloud (Azure then AWS)
On-Premise Cluster

- 40 Node Hadoop cluster
- LSS as primary storage (cold storage)
- Presto as Query Engine
- 2 GPU Cluster
Primary On Premise Function

- Camera Snapshots (5-minute back to 2016)
- Deep Learning using Traffic Cameras (GPU Cluster)
- Crash/ATMS Video Download
Cloud (Azure then AWS)

Computation
• Two Linux Servers
• One Windows Server
• Glue and EMR (PySpark)
• Jupyter Notebook
• Athena (Query Engine)

Storage
• S3 (35 TB)
• MySQL Database
• PostgreSQL Database
• Compressed Format (parquet and ORC)
Cloud Service Cost/Speed

*Primary factor of speed and cost depends on how much data is stored and queried*

**Reduce Data Size**
- Limit data storage (example for weather data for only roadways)
  - 17.1 GB/Day => 10.1 GB/Day
- Convert data format to parquet
  - 10.1 GB/Day => 0.35 GB/Day

**Data Partitioning**
- Limit size of partition to minimize data scanned
  - Weather data by month/day/hour
  - INRIX data by month/day
Primary Data Flow

- **Serverless ETL for streaming data such as probe, sensor, etc.**
- **Data Storage**
- **Serverless interactive query service using SQL that executes queries in parallel**
- **Data Visualization for providing access to data for end users**
Data Utilizing S3/Athena

- Crash
- ATMS Events
- Probe Data
- RWIS
- Traffic Sensors
- Snowplow
- Truck Parking
- Weather Data
- Wejo/Connected Vehicle
- Work Zone Performance
Alternative Data Flows

MySQL
• Primarily used for sharing data with Iowa DOT (easiest to use)
  – 511/Work Zone, Smart Arrow Boards, snowplow, truck parking

REACTOR Feeds
• API for broader data sharing
  – Sensor Feed, Work Zone Alert Feed, Lane Closure Planning Tool, etc.

PostgreSQL
• Used for integrating with Iowa DOT linear referencing system (RAMS)
  – Covered in more detail later
Know how data can be integrated
Iowa States Approach: Linear Referencing System (LRS)
Why the LRS?

Integration through database functions rather than spatially

Provide access to additional data integrated with LRS

Currently over 138 features including:
  Number of Lanes
  Traffic Information
  City Name
  Urban/Rural
  Median Type
Roadway Asset Management System (RAMS)

ArcGIS REST Services Directory

Home > services > lrs (MapServer) > IOWA_LRS_NETWORK

JSON

Layer: IOWA_LRS_NETWORK (ID: 0)
Name: IOWA_LRS_NETWORK

Iowa LRS often referred to as RAMS

Utilizes ESRI Roads and Highways

Integration uses REST API to get Route and Measure
RAMS Challenge

Limitation on data processed through REST API (1500 records/request)

Evaluated alternative options:

<table>
<thead>
<tr>
<th>Method</th>
<th>Time to process 13,334 points with 50 meter tolerance</th>
<th>Throughput</th>
</tr>
</thead>
<tbody>
<tr>
<td>ArcPy</td>
<td>933 seconds (15.5 minutes)</td>
<td>~ 14/sec</td>
</tr>
<tr>
<td>PostgreSQL with geography</td>
<td>1220 seconds (20.3 minutes)</td>
<td>~ 11/sec</td>
</tr>
<tr>
<td>PostgreSQL with geometry</td>
<td>36 seconds</td>
<td>~ 370/sec</td>
</tr>
</tbody>
</table>
RAMS and PostgreSQL

Iowa LRS network in RAMS downloaded daily and uploaded to PostgreSQL

Allows local processes for conflating data to LRS

Developed process for identifying and extracting dominant route as needed
RAMS Integration Use Cases

Output:

Arrow Board
RouteID: M015546105E
Measure: 0.79

Connected Cone
RouteID: M015546105E
Measure: 1.01

Actual Work Zone
RouteID: M015546105E
Measure: 0.79

From Measure: 0.61
To Measure: 1.28
From Measure: 0.70
To Measure: 1.10
From Measure: 0.79
To Measure: 1.01
RAMS Integration Use Cases

Associating devices to work zones

RouteID are equal
AND
measure is > FromMeasure
AND
Measure is < ToMeasure

<table>
<thead>
<tr>
<th>Planned Data</th>
<th>Connected Devices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work Zone ID</td>
<td>Route ID</td>
</tr>
<tr>
<td>1</td>
<td>S001910035N</td>
</tr>
<tr>
<td>2</td>
<td>S001910035N</td>
</tr>
<tr>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Arrow Board</td>
<td>S001910035N</td>
</tr>
<tr>
<td>Connected Cone</td>
<td>S001910035N</td>
</tr>
<tr>
<td>Arrow Board</td>
<td>S001910035N</td>
</tr>
<tr>
<td>Connected Cone</td>
<td>S001910035N</td>
</tr>
<tr>
<td>Arrow Board</td>
<td>S001910035N</td>
</tr>
<tr>
<td>Connected Cone</td>
<td>S001910035N</td>
</tr>
</tbody>
</table>
RAMS Integration Use Cases

Waze Event Clustering

I-80 EB

I-235 EB

Adding
Road name and direction

Adding Time and incident type

Iowa State University
Institute for Transportation

IOWA DOT
RAMS Integration Use Cases

Probe Data Integration

Used to extract volume estimates for delay cost
  • Probe -> RAMS -> AADT

Used to integrate with Crash Data (which was also related to RAMS)
  • Identify congestion due to crash
  • Identify secondary crashes
  • Identify delay due to crash
Other Future Work

- LSS Phase-out
- Serverless Architecture
- Documentation
- Convert on-premise cluster to test bed
- Explore extending AWS network to InTrans/DOT