NASCIO 2019 State IT Recognition Awards
Category: Business Process Innovation

COMMONWEALTH of PENNSYLVANIA

Transportation Analytics Platform (TAP)

Tapping Big Data To Put PA Traffic Operations in the Fast Lane

Project Initiation: January, 2017
Project Completion: September, 2018
Project Sponsors:
Mike DeMatt, CIO, IE&D
John MacMillan, CIO, Commonwealth of PA

Submitted By:

Donald H. Kirschman | Director
Data Management & Analytics | Infrastructure and Economic Development
Commonwealth of PA
400 North Street, 5th Floor | Harrisburg PA 17120-0041
Phone: 717.705.1216

EXECUTIVE SUMMARY

The Pennsylvania Department of Transportation (PennDOT) relies on data from a multitude of sources to both manage and improve its operations, which span 40,000 roadway miles and 25,000 bridges. However, the department's existing data sets, analytics tools and technology infrastructure were limiting the discovery of new insights and opportunities to increase public safety, reduce costs and drive innovation. Can crowd-sourced data be merged with internal agency data to reduce traffic incidents and congestions? How can roadway assistance be dispatched to accident sites more efficiently? What can be done to optimize the agency's winter operations?

Our increasingly connected world is generating an endless torrent of data to be analyzed, which drives the need for vast amounts of compute and storage. Agency business customers are joining IT in the role of technology innovator, and they are exerting unprecedented demands for cutting-edge tools and technologies. The business is also more agile than ever, posing increasingly complex questions and demanding immediate results.

The Office of Administration's Infrastructure & Economic Development Delivery Center Analytics Team partnered with PennDOT's Traffic Operations and Roadway Maintenance staff to develop the Transportation Analytics Platform (TAP). TAP brings together core data sets, cloud analytics tools and an Agile product development approach. TAP leverages Microsoft Azure and Power BI to gather, manage, blend and visualize heterogeneous crowd-sourced and internal data sets and provide analytical and predictive reporting capabilities.

In addition to addressing specific business area needs, TAP brings a broader and more significant analytics capability that can benefit all lines of business in PennDOT, as well as other state agencies.

TAP has delivered the following results for the department:

- **Faster response to and clearing of traffic incidents:** The Traffic Alerts solution in TAP collects, analyzes and displays Waze incident and INRIX speed data in real time. Information is merged and visualized to provide first investigators with a detailed view of crowd-sourced incidents.
- **Reduced congestion and accidents:** TOA (Traffic Operations Analytics) blends traffic speed information from INRIX, crowd-sourced incident information from Waze, and PennDOT's incident and crash data to present a variety of metrics and visual analytics that help to improve congestion management, incident response and roadway safety.
- **More efficient winter operations:** The AVL Winter Materials solution collects detailed Automatic Vehicle Locator (AVL) data in real-time from PennDOT's plow trucks. The sensor data is integrated with weather and materials usage data to produce GIS visualizations and a dashboard for winter events, as well as detailed metrics and reporting for post event analysis.
- **Increased IT agility:** The Analytics Team can deliver solutions offering innovative technologies and capabilities and provide value to the business more quickly.
- **Better IT cost management:** TAP provides detailed cost metrics from the cloud provider that clearly show the technology costs for each solution, facilitating better and faster decisions about project costs versus business value.

Putting it all together, TAP has reduced risks and costs, while increasing capabilities and agility. Looking to the future, TAP has PennDOT well-positioned to handle the onslaught of data to come from IoT, connected vehicles, intelligent infrastructure and more.

CONCEPT

The Challenge

PennDOT traffic Operations and Roadway Maintenance engineers challenged the Analytics Team to help them further analyze the real world impacts of our traffic control and public safety measures. They challenged the team to go beyond internal sources of data. The availability of crowd-sourced and probe data aggregators offered new opportunities to incorporate additional data sets into complex analyses. Examples of data sets include:

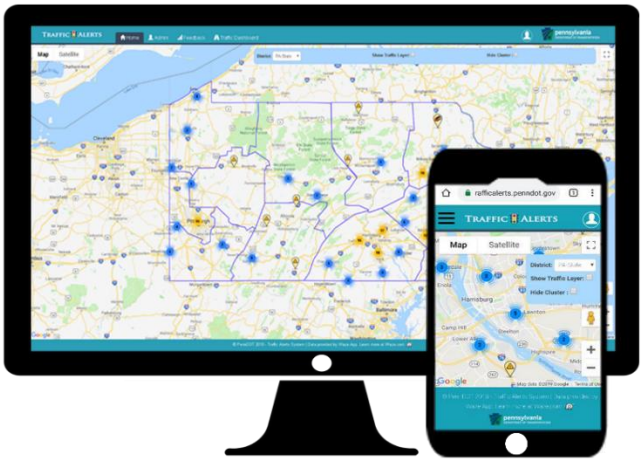
- Crowd-sourced traffic incidents from Google Waze
- Probe-data aggregators like INRIX that provide traffic speeds for every minute of the day across our entire road network
- IoT (sensor) data from PennDOT's plow trucks

The proposed solution needed to integrate and correlate this data in new ways. Fuzzy relationships combining events by geospatial and temporal proximity and handling massive amounts of data were needed. This required a platform able to support acquiring and analyzing information from the new landscape of data with the four Vs (volume, velocity, variety and veracity), and delivering visually powerful insights. The Analytics Team identified and defined the capabilities required to meet this challenge and then assessed the department's current capabilities. The assessment identified several key areas where improvement was needed.

- Utilization of crowd-sourced traffic incident and congestion data sets
- A data management platform capable of storing huge volumes of structured and unstructured data
- Tools, APIs (Application Program Interface) and solutions sufficient to analyze and visualize the "four V's" of Big Data
- The ability to quickly incorporate new capabilities like artificial intelligence and machine learning

SOLUTION

The Transportation Analytics Platform (TAP)



To overcome PennDOT's challenges with data, tools and agility, the Analytics Team looked to the cloud to design a modern analytics and data visualization platform. TAP encompasses core data sets, cloud-based PaaS and SaaS tools and an Agile product development methodology. The agile platform facilitates partnering between IT and the business in a way that was not possible previously. These boundary-pushing analytics projects require discovery, evolution and rapid iteration. TAP enables quick discovery and visualization of data sets, rapid prototyping and evaluation and adoption of emerging technologies. It allows the Analytics

Team and the business to experiment collaboratively, "fail fast" and then iterate to the optimal solutions. As a result, PennDOT can now:

- Acquire, store and process Big Data sets with hundreds of billions of rows
- Leverage a Massively Parallel Processing (MPP) database to perform Big Data analytics
- Scale up and scale down to cost-effectively meet highly elastic computing requirements
- Provide powerful, responsive and customizable visualizations

Paving the Way for Cloud Innovation

Very early on, the TAP requirements pointed to a cloud-centric solution. In developing TAP, the following steps were taken to pave the way for cloud innovation:

- Obtained buy-in from the state CIO, IT Bureau Directors and key business executives to define an "Opportunistic" cloud posture for the agency
- Worked with the IT business office to address procurement and budget concerns
- Worked with agency and enterprise IT security officers to assess risk and define evaluation criteria for the types of solutions and data that are suitable for cloud
- Leveraged the agency Enterprise Architecture & Service Management (EASM) program as a key governance body for making cloud-specific architecture decisions
- Drafted agency-level cloud policy and standards and other guiding documents
- Held executive, managerial and technical level presentations and workshops to promote transparency and share knowledge within the agency and across the commonwealth

Acquiring Core Data Sets

TAP will be used to deliver numerous specific analytics solutions to the business; however, a core set of new Big Data sets (traffic volume/speed, incidents, weather, etc.) was needed to support them. Recognizing this, the PennDOT Traffic and Roadway Divisions led the way with innovative and aggressive data acquisition. This drove demands for the Analytics Team to derive value from this high-potential data. There were several approaches to acquiring data:

- Procured subscriptions to INRIX, a leading data aggregator for roadway traffic, which provided average traffic speeds for every minute of the day across PA's road network broken down by granular road segments of less than 1 mile in length.
- Signed a data sharing agreement with Waze via the Connected Citizens Partnership (CCP), which enabled free access to the Waze real-time API for its core crowd-sourced data.
- Acquired PennDOT's sensor data, such as plow truck AVL and weather station readings.
- Conducted a joint effort between business and IT to identify all internal systems from which relevant data sets could be collected, such as incident, crash, and road maintenance work.

Addressing Data Volume, Variety, Velocity and Veracity

The largest core data set (INRIX traffic speed data) consists of hundreds of billions of records, or 10+ terabytes of data, per year. Not only did the team need to store and load this data, but it had to do complex analysis to derive insights, such as congestion events and average traffic speed grouped by weather conditions and by road segments. The on-premise relational database environment was not designed for this. At best, it would have required a major capital investment and significant re-architecture for a project with a considerable risk of failure. With TAP, the team provisioned and activated an MPP database within minutes and did a proof-of-concept to determine if it could handle the workload. With the concept proven, this massive weekly workflow was operationalized by

configuring and building an automated process to turn on the environment, scale it to the optimal tier, process the data and then deactivate the environment.

Analyzing and Correlating Big Data

Data blending, i.e. integrating data via data correlations, was possibly the biggest challenge and greatest innovation. It required combining disparate data sets with no natural keys, requiring relative and fuzzy correlations of data such as geospatial, time, structured and unstructured, big and small, etc. For example, crashes were correlated to traffic congestion by geospatial proximity and time boundaries. This effort required deep data analysis and subject matter expertise, as well as learning and applying modern technology concepts and tools. Lastly, analytical data models were created and populated to enable the business domain analytic solutions.

Data Visualization

In selecting a data visualization tool for TAP, the Analytics Team conducted a rapid, intensive, side-by-side comparison of industry leaders Tableau and Qlik, along with the relative newcomer, Microsoft Power BI. Testing with IT and business users found the functional capabilities of these tools to be very similar. However, Power BI stood out for the pace of innovation (with monthly releases), relative ease-of-use, ease of integration with our existing IT environment and the economical subscription-based pricing model. The choice of Power BI has proven to be the correct one for TAP. Our proof is the positive feedback from the business community.

Agile Product Delivery

Prior to TAP, PennDOT's DW/BI development lifecycle was largely a waterfall methodology. For modern analytics, a more agile approach is needed. Key elements of TAP's agile methodology include:

- Early and frequent prototyping with continuous refinement based on feedback from business users
- Data profiling and discovery conducted throughout the project
- Production releases in weeks, not months
- A cloud-hosted, modern Application Lifecycle Management (ALM) tool and a Continuous Integration/Continuous Delivery (CI/CD) pipeline

SIGNIFICANCE

TAP is a key component in PennDOT's digital transformation toolkit. From strategic decision-making to optimizing operational efficiency, TAP is having a direct impact on optimizing infrastructure and transportation spending and improving public safety in Pennsylvania. How can we improve efficiency in winter maintenance with a budget of $225 million annually? How can we reduce the number of crashes in Pennsylvania (totaling 128,188 in 2017, resulting in 1,137 fatalities and 80,612 injuries)? TAP is playing a vital role in answering these questions and many others.

More broadly, TAP defines an analytics paradigm of carefully curated, self-service data sets and models, where business users can leverage powerful computing capabilities and perform "what-if" and other types of specialized analytics without compromising IT policies and with less IT support. This capability has applications across PennDOT and other agencies.

Looking ahead, the need for Big Data analytics is going to continue to grow with innovations like connected and autonomous vehicles. As a part of the TAP initiative, PennDOT is actively retraining and retooling its IT and business staff to take advantage of emerging technologies. TAP solutions have

started to incorporate artificial intelligence and machine learning to uncover business intelligence in new ways. With nimble yet robust and scalable technical foundation and an Agile product delivery model, TAP is poised to meet PennDOT's analytics needs well into the future.

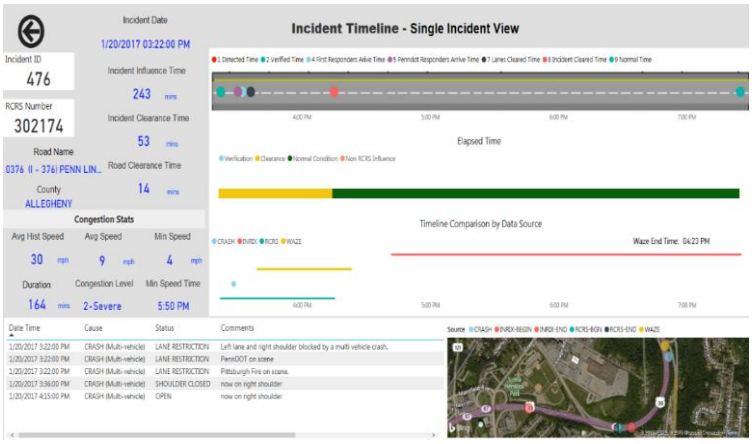
IMPACT

Responding to and Clearing Traffic Incidents Faster

The Traffic Alerts Dashboard is used by Traffic Management Center (TMC) staff to aid them as they work to respond to and clear traffic incidents. The solution consumes and merges Google Waze incident data, INRIX traffic speed data and PennDOT's weather station data and visualizes it in a live, interactive dashboard. The dashboard also provides access to geospatial maps to zoom in and visually track the life of an incident, and, where cameras are available, access real-time video. Additionally, TMC staff have the capability to flag the reported incidents as a True or False report. Going forward, this information lays the foundation for a machine learning model to give the TMC reliability scoring to reduce dispatches for false alarms. Since the implementation of this solution, TMCs across the state have improved their response times and efficiency in dispatching roadway assistance.

Reducing Congestion and Accidents

A strategic imperative for PennDOT's Bureau of Maintenance and Operations (BOMO) is to improve traffic operations planning and practices. The Transportation Operations Analytics (TOA) solution introduced metrics that allow BOMO to improve project and fiscal decision-making and to compare operational performance among Traffic Management Centers (TMC's) across the state. TOA integrates crowd-sourced traffic incident and speed data from Google Waze and INRIX with incident, accident and road closure data from PennDOT's own internal systems. Visualizations, dashboards and interactive reports provide new insights into traffic incident performance and the effectiveness of traffic operations practices and traffic control measures.

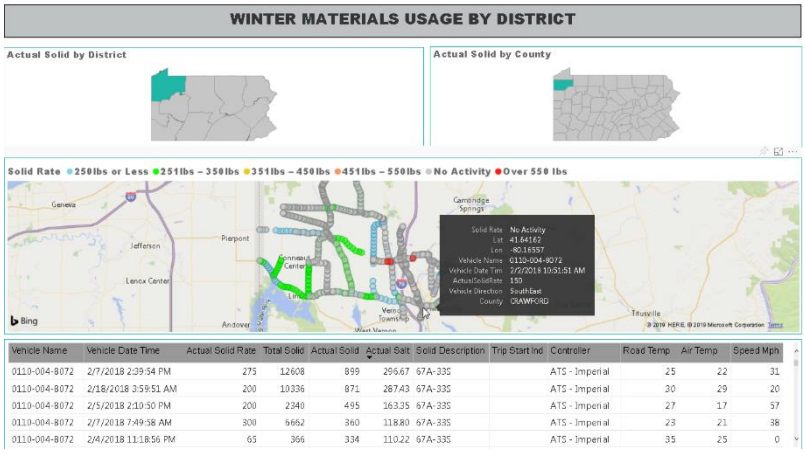


- The data insights led to the adjustment of TMC operating hours to better align with congestion and accident frequencies.

- TMCs were provided with maps showing the locations where high congestion crashes tended to occur and where additional situational awareness might be warranted.
- INRIX is going to highlight Pennsylvania's work in a case study in the upcoming months; according to BOMO, no other state has taken analysis of their data to this level.

More Efficient Winter Operations

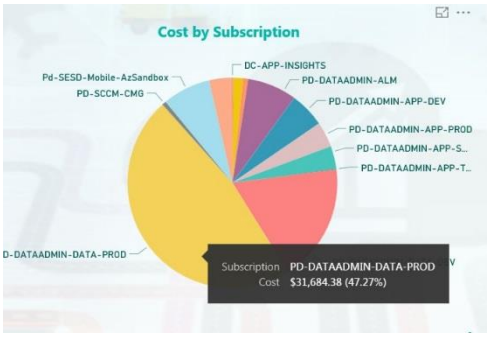
In 2014, PennDOT began outfitting its snowplow fleet with on-board AVL devices. These devices provide detailed telemetry data in real-time, including vehicle location, speed, direction and route. Additionally, the devices provide continuous data about the plow depth and angle and the application rate of roadway treatment materials. The AVL Winter Materials analytics solution collects and integrates AVL and winter materials usage data in near real-time. Data is presented on interactive maps with powerful zoom, pan and drill capabilities. Comprehensive metrics and reporting, including detailed material usage calculations, are available to managers within hours for after-action analysis.



Vehicle Name	Vehicle Date Time	Actual Solid Rate	Total Solid	Actual Solid	Actual Salt	Solid Description	Trip Start Ind	Controller	Road Temp	Air Temp	Speed Mph
0110-004-8072	2/7/2018 2:39:54 PM	275	12608	899	296.67	67A-33S		ATS - Imperial	25	22	31
0110-004-8072	2/18/2018 3:59:51 AM	200	10336	871	287.43	67A-33S		ATS - Imperial	30	29	20
0110-004-8072	2/5/2018 2:10:50 PM	200	2340	495	163.35	67A-33S		ATS - Imperial	27	17	57
0110-004-8072	2/7/2018 7:49:58 AM	300	6662	360	118.80	67A-33S		ATS - Imperial	23	21	38
0110-004-8072	2/4/2018 11:18:56 PM	65	366	334	110.22	67A-33S		ATS - Imperial	35	25	0

- Near real-time data from over 1,400 plow trucks collected and visualized
- Elimination of error-prone and labor-intensive paper-based data collection process
- Detailed, actionable data available to managers within hours instead of days
- Savings on winter materials expected "in the millions of dollars," according to department leadership

IT Cost Transparency and Management



Microsoft Azure exposes very detailed cloud service cost and utilization data via API. By leveraging these API's and Power BI, the Analytics Team has incorporated a set of interactive cost and usage visualizations into TAP. TAP solution stakeholders can see the cost of TAP cloud services across any number of dimensions, including time, solution, environment (e.g., Development, Test, Production), service category and more. With developers and infrastructure teams more aware of costs, they proactively optimize provisioning and scaling of services. Coupled with the on-demand pricing, this means that TAP provides superior analytical solutions to the business customer at a lower cost. For example:

- On-demand use of an MPP SQL Data Warehouse engine that would cost more than $1 million to set up on-premise totaled less than $6,000 for all of 2018.
- TAP has smart cost management features like cost projection, anomaly detection and alerts to help spot and stop budget overruns and unnecessary services and costs