Enabling self-service and advanced analytics for TfNSW on Azure

Transport for NSW (TfNSW) manages one of the largest fleets of vehicles in Australia, including Buses, Ferries, Light Rail, Trains and Metro.

The real-time data generated by these vehicles provides analytical opportunities to improve transport services by measuring service performance and optimising routes,

however, the sheer volume of data presents data management challenges, which in turn makes it difficult to mine.

Data-Driven partnered with TfNSW to deliver their Data Foundation solution on Azure for the Operational Data Lake (ODL) platform to enable self-service and advanced analytics capabilities previously not possible.

Customer: Transport for NSW (TfNSW)

Website: https://www.transport.nsw.gov.au/

Customer Size: Corporate (10,000+ employees)

Country: Australia, NSW

Industry: Travel and Transportation

Products and Services: Microsoft Azure, Databricks

Data-Driven, TfNSW, & Azure data platform with Azure Databricks

data-driven

Customer challenges

Historical GTFS transport data has always been too large and costly to store efficiently and analyse. Every TfNSW vehicle sends its' location every 10 seconds which results in a huge stockpile of data containing valuable insights.

In the past, this data was not stored and only the last copy was published on the Transport Open Data Hub. This made insights from past vehicle trips almost impossible to obtain and ruled out the ability to predict trip delays or optimise trip routes. It also made it difficult to report on and analyse service performance and reduced the ability to improve customer services.

Partner Solution

The Data Foundation solution by Data-Driven was the perfect starting point for TfNSW as it was designed to be highly scalable and extensible whilst providing a platform for citizen data scientists and business users to perform advanced analytics.

This met TfNSW's vision to build the Operational Data Lake (ODL); a unified, next-generation data and analytics platform, leveraging native Azure Cloud services and Azure Databricks to enable the continuous collection/curation of diverse transport operational data sets and allowing self-service analytics and a platform for advanced analytics and machine learning.

Customer Benefits

The Operational Data Lake allows TfNSW to:

- ✓ Store infinite data in a cost-effective manner
- ✓ Access and analyse historical data for BI reporting, advanced analytics or ML use-cases
- ✓ Share data with internal and external users to remove data silos

- ✓ Allow data discovery with built-in monitoring
- ✓ Ensure data privacy and security, with robust governance & cost management controls
- ✓ Deliver insights to the organization in an automated, interactive and near real-time manner

"TfNSW needed a solution to capture real-time data for every vehicle in motion across the state. This solution just gives us that, so that we mine nuggets from this data later."
Sandeep Mathur
Program Manager, TfNSW
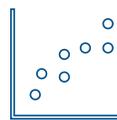
"We're now able to leverage machine learning and AI on the rich datasets to find more efficient ways to improve public transport services"
Daniel Yu
Senior Domain Architect, TfNSW

180 000 IoT Messages per day
5 TB a week of data ingested
4 datasets to explore

Microsoft Partner
Gold Data Analytics
Gold Data Platform
Silver Cloud Platform
Microsoft