Key responsibilities and Performance indicators

Eric Solano, Ph.D.

At Continental I oversee the big data repository for telematics data, the Telematics Backbone (TBB). TBB's main purpose is to support Data Scientists with a solid and easily accessible layer of all available and enriched telemetry data we collect at Continental from passenger vehicles and trucks. We install telematics devices on vehicles to collect vehicle data such as speed, GPS, acceleration, weight, etc., in addition to the data we obtain from the tire-mounted sensors: e.g. tire temperature and pressure. My title at Continental is Internet of Things Data Scientist. With the emergence and adoption of the Internet of Things at Continental, large volumes of data are generated today. Many types of hardware are delivering information about their state and how they are being used, and we are using this data for completely new products and services such as "Predictive Maintenance".

My key responsibilities with the TBB are to facilitate these use cases. I must have data readily available to analysts trying to get specific information. I need to make sure that data is in one single location, data can be easily understood, fetching different datasets is technically non challenging, and known data flows have been removed and data has been aligned automatically. Through the TBB, I provide a single location allowing data scientists to find the data they are seeking without the need to focus on different delivery strategies; a platform to easily grasp the structural information of all datasets without the need to request that information from stakeholders; and a tool to easily consume the data in a technical way without the need to learn new APIs and interfaces for each individual dataset.

Data Scientists are the main consumers of the TBB. To comply with their specific needs to work on non-aggregated and non-selective data, I focus on delivering raw data to them via the TBB, in some cases enriched with other datasets that can help them detect relevant business information. In addition to data science users, the TBB can also work as a consolidated access layer to telemetry data for analytical applications. Examples are: "Vehicle Location Tools" that require information about selected vehicles or "Customer Care Tools" that provide predictive maintenance services and need information about different loggers and their measurements.

Key performance indicators of my job as the TBB manager are the increase of use cases consuming data from the TBB in the last 5 years from a handful of projects to dozens of projects; the increase in the volume of data stored in the last 3-4 years, the savings in hundreds of hours in data collection and processing thanks to automation, the generation of dozen of new TBB data driven ideas to improve production processes, the industrialization of a handful of TBB data supported analytical applications, etc.

Additional key responsibilities include software development to support data science projects and building machine learning models for proof of concepts and productive uses. I also spend at most 25% of my time mentoring students and new hires. For the last couple of years, I was part of an expert team that developed the Modern Data Analytics Platform (MoDAP). MoDAP is a

composition of cloud services from Amazon Web Services to enable developers to build and schedule data pipelines, deploy machine learning models and store data from different sources. It abstracts the complexity of using AWS services.

I developed MoDAP using Infrastructure as Code (IaC) and AWS Cloud Development Kit (CDK). The AWS Cloud Development Kit (AWS CDK) is an open-source software development framework developed by Amazon Web Services (AWS) for defining and provisioning cloud infrastructure resources using familiar programming languages.

The key performance indicators of MoDAP after it was developed and industrialized include: a constant increase of users to up to thirty users. Users are not required to have knowledge about AWS. Everything is abstracted from the users. Another key performance indicator is substantial savings on AWS account fees. Due to the scaling capabilities only required resources are used. This leads to high-cost efficiency. Another performance indicator is Reduced Time to Market: Creating Data Engineering Pipelines is now achieved in minutes by our data scientists and trainees.