# Platforms and Applications for Data Analytics

Eric Solano, Ph.D.

This year at Continental, I have deployed the most recent versions of the Telemetry Backbone (TBB) and the Modern Data Analytics Platform (MoDAP). Also, I developed an application for Tire Manufacturing Big Data Analysis.

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. The TBB uses Apache Cassandra as persistence layer without single point of failure serves as storage for the data provided. Cassandra scales linearly. Data is currently being ingested via stream (mainly from the ContiConnect Kafka Broker) or via Batch process from different batch engines. The TBB utilizes a message broker (Kafka) and provides dedicated topics per data source as ingestion zone for incoming telemetry data.

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. 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. MoDAP uses the Amazon Managed Workflows for Apache AirFlow (MWAA) service to orchestrate new data science project workflows.

The Tire Manufacturing Big Data Analysis application is currently in production in one of our plants. The tire manufacturing process involves several steps: Tire Design, Mixing, Calendering, Extrusion, Molding, Vulcanization or Curing and Inspection. All or most of these steps collect large amounts of data from equipment and sensors at the manufacturing plant and store the data in a system called DOPAC. This project used data sources from one of Continental plants located in Korbach, Germany. The dataset has very high dimensionality. Big data technology was required to analyze these datasets. The Apache Spark platform, the Hadoop Distributed File System (HDFS) and the Apache Mahout framework. An Amazon EMR cluster with 5 nodes was used. Each EMR cluster node is of type 'm5a.xlarge' (4 CPUs, 16 GB RAM).