TFX: A TensorFlow-Based Production-Scale Machine Learning Platform

Problem:

Creating and maintaining a platform for reliably producing and deploying machine learning models can be particularly challenging when data changes over time and fresh models require to be produced continuously. – previous orchestration is often done ad hos using glue code and custom scripts developed by individuals, leading duplicated effort and fragile systems with high technical debt.

Solution:

TensorFlow Extended (TFX) – integrating the aforementioned components.

1. Standardize the components
2. Simplify platform configuration
3. Reduce time to production
4. Provide platform stability that minimizes disruptions
5. Limit the technical debt arising from implementation

Covering existing ML algorithms, giving users the ability to switch between ML algorithms

Challenges:

1. One platform for many different learning tasks- difference in data representation, storage infrastructure and ML tasks
2. Continuous training and serving – models need to be updated based on new data

**Support interaction between data visitation and warm-starting (subsets) options, along with dynamic option for data visualization**

1. Human-in-the-loop – ML platform needs simple user interface for the need to deploy and platform monitoring

**One common configuration that is passed between components, enabling global garbage collection policies, unified debugging and status signals**

1. Production-level reliability and scalability – platform need to be resilient to disruptions from inconsistent data and scalable to high data volume

**Model validation is critical, coupled with data validation to detected bad training data. It is vital to the reliability and robustness of ML platform. The best practices in many production pipelined is included in this platform.**