

## Avoiding an ML Migraine!

by Laks Srinivasan

Here are some simple observations: setting up an ML stack/pipeline is incredibly hard, setting up a production ML stack/pipeline is even harder, and setting up an ML stack/pipeline that works across multi-cloud environments is a full-on headache.

That's not just the complaints of a technologist. These translate to wasted productivity of ML engineers. It takes them longer to build ML-powered Apps with a significant amount of manual effort to stitch the steps together and it is extremely error prone.

There is no one standard End-to-End process from data to ML system in production; we are still in the "artisan age". The typical end-to-end process includes: 1. Data Pre-Processing, 2. Feature Engineering, 3. Model Building 4. Model Deployment, and 5. Model Maintenance. ML model building is very small part of end-to-end effort. The majority of the work occurs before, in the data management step and after in serving up the model and monitoring and retraining the model. Most products in the market focus on making step 3 most-efficient and automated and so ML engineers are on their own when it comes to building an end-to-end ML pipeline.

An open source solution attempting to address this problem today is Kubeflow, which is all about building ML workflows. Kubeflow's stated main mission is "Make it Easy for Everyone to Develop, Deploy & Manage a Composable, Portable, & Distributed ML environment on Kubernetes".

It provides an integrated front-end for ML Pipeline Administration (Job management, Monitoring, Debugging, etc.), components for building and serving models, and utilities for data management – garbage collection, storage, data access control. With any early stage open source software, there is still a lot of manual efforts required when it comes to data management especially around data management through Kubernetes at the infrastructure level.

What's needed is a product that helps accelerate pipeline development. Arrikto's Rok fills this gap in Kubeflow today by providing higher-level and bigger lego blocks for building ML pipeline and Arrikto's Enterprise Kubeflow Service (EKF) packages Kubeflow with Rok into a nice package along with documentation, training and support. I've been impressed with the product's capabilities.

As a proof point, the Arrikto team wanted to demonstrate how to use MiniKF (which is a single node version of EKF) to build ML pipelines efficiently applying it to the Chicago Taxi dataset. The target for the ML pipeline is to predict transactions that will generate more than 120%+ revenue, i.e., a 20%+ tip. The dataset involved 100M trips, released by City of Chicago. The

Arrikto team built a baseline of ML pipeline for solving this using a pre-0.4 version of KF. It also, built the same pipeline with latest Version of MiniKF.

MiniKF enabled basic data management functions using UI and through higher-level functions: ability to create persistent volume from UI, the ability to compile versioning commands into the pipeline, and the ability to transfer a cloned dataset from notebook to pipeline. The Arrikto team saw a 50% reduction in steps to build this pipeline.

I believe MiniKF completes Kubeflow with critical higher-level functionality on data management for ML pipelines. A 50% reduction in steps will free up ML Engineers to spend their creative energies on solving more business problems and compel product managers to infuse ML in their products.

I highly encourage you to evaluate. The best way to understand how EKF improves your productivity is to run a [hands-on tutorial](#) with Mini Kubeflow (MiniKF). In about 1 hour, you can download and run a fully operational Kubeflow cluster on your laptop, and then build, train and deploy an ML pipeline for Chicago Taxi Tutorial, which is a popular TensorFlow TFX example.

About the author: Laks Srinivasan has 15+ years of executive management experience in global organizations that help enterprises harness AI and ML for concrete, measurable business results. Through dozens of real-world customer projects across a portfolio of verticals, Laks has a wealth of practical experience in deploying AI and ML at scale from concept to commercialization.

Previously, Laks was COO of Opera Solutions, an applied AI and ML solutions company with top placement in many ML competitions including the Netflix prize and the KDD Cup. Prior to Opera Solutions, Laks held various positions at FICO (NYSE: FICO), ExpenseAdvisor, Booz Allen Hamilton and Syntel, Inc (Atos Syntel since 2018).

Laks holds an MBA from Wharton in entrepreneurial management and finance, as well as a BS in electrical engineering from NIT, India.