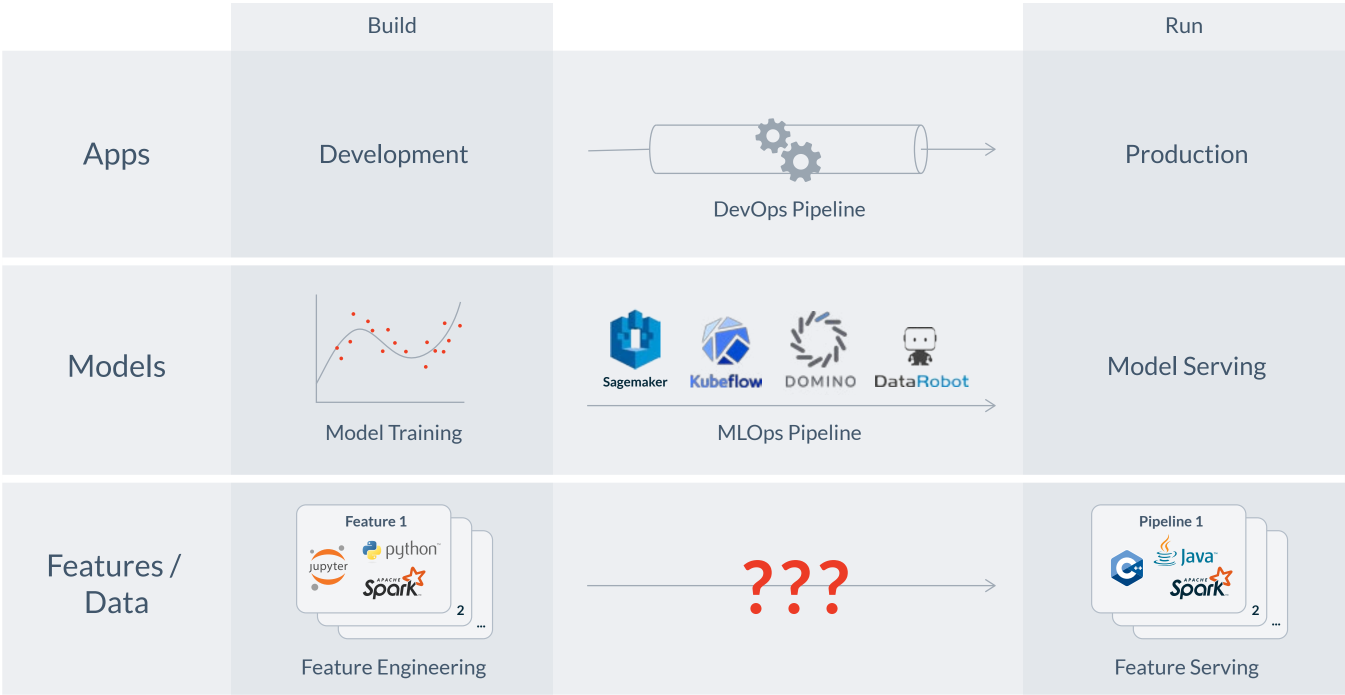
Tecton

The Data Platform for Machine Learning

"Data is the hardest part of machine learning and the most important to get right."
– Uber Michelangelo

Companies are increasingly investing in Machine Learning (ML) to deliver new customer experiences and re-invent business processes. Unfortunately, the majority of operational ML projects never make it to production. The most significant blocker is the lack of infrastructure and tooling required to build production-ready data for machine learning. Tecton was founded by the creators of Uber's Michelangelo ML Platform to fill this void. Based on the concept of the feature store, Tecton enables data scientists to build great ML features, serve them in production quickly, and do it at scale. By getting the data layer for ML right, companies can get better models to production faster and drive real business outcomes.

The Challenge: Tooling for managing ML data is almost non-existent



The most technically advanced companies are using ML in production to power mission-critical applications. "Operational ML" is the core driver of growth and innovation for technology giants such as Google, Amazon, and Uber, and is behind new customer experiences in every industry: think of recommendation engines, real-time pricing, personalized search, fraud detection, and demand forecasting.

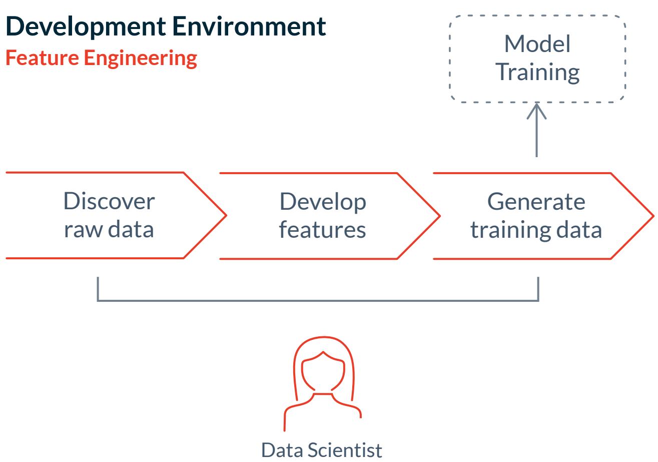
But for all its promise, operational ML is largely an unsolved problem outside of the few companies that have built proprietary systems. Teams looking to put together a machine learning platform can turn to emerging MLOps solutions, such as AWS SageMaker for model serving and MLflow for model maintenance. Unfortunately, such solutions focus on the models, but don't address the hardest part of ML: the data.

An ML application is fundamentally just a data application. To make predictions, it relies on features: intelligent, live signals about a product's customers, users, market, performance, etc. Feature data represents the business' most curated data, the result of long and expensive processes of cleaning and massaging. But feature data is also very hard to get right. Building features relies on a complicated process that requires repetitive reimplementation between data science and data engineering organizations.

Development Environment

Feature Engineering

In development, data scientists create new features that are used for model training. Feature engineering happens in the data scientist's development environment, typically using interactive notebooks and Python code. Without centralized tooling to manage features and make them reusable across an organization, data scientists operate in silos and re-implement duplicate work within their models.



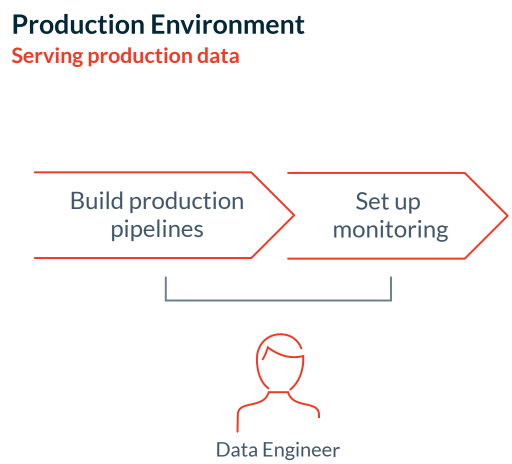
Feature Engineering challenges

- Lack of collaboration reduces productivity and feature re-use
- Feature code is not managed with software engineering best practices
- Training data is difficult to generate and prone to data leakage
- Feature transformations are complex when combining batch, streaming and real-time data

Production Environment

Serving production data

In production, the code generated by data scientists is not production-ready, so data scientists are left relying on engineering teams to productionize their feature pipelines when they deploy their models. This typically means "throwing it over the wall" for a production-ready re-implementation, usually in a different, production language like Java or C++. This entire process often takes more than 6 months just to get the first version of the model in production, and is then repeated for each iteration.

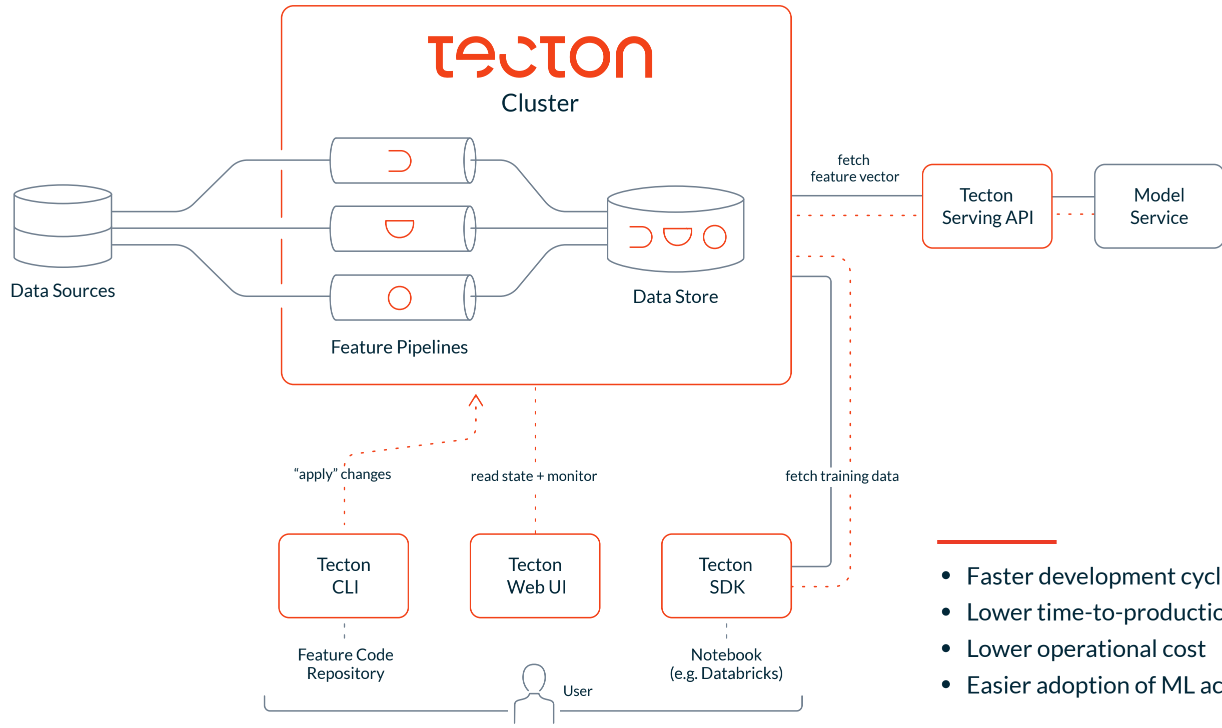


Production data challenges

- Data scientists depend on others to deploy and maintain production pipelines
- Long lead times increase time-to-market for new models
- Training-serving skew reduces model accuracy
- Data quality issues are hard to monitor and debug

Tecton: The data platform for machine learning

Tecton was founded by the team that created the Uber Michelangelo feature store, which enabled Uber to scale to tens of thousands of ML models in production in just a few years. Based on our learnings from Michelangelo, we built Tecton to bring DevOps-like agility and governance to the feature engineering process.



- Faster development cycles
- Lower time-to-production
- Lower operational cost
- Easier adoption of ML across teams

Tecton is an enterprise-grade feature store that manages the complete lifecycle of ML features from development to production:

Feature Pipelines: Tecton connects to enterprise data sources including Streaming Platforms, Data Warehouses, Data Lakes, and Databases. Using feature transformations expressed in Python, SQL, or PySpark, Tecton orchestrates the continual calculation of fresh feature values.

Data Store: Tecton's Data Store provides a single source of truth for curated feature data. Historical feature values are kept in an offline Data Store for batch retrieval. Fresh feature values are maintained in the online Feature Store for low-latency retrieval.

Serving API: Tecton provisions model-specific API endpoints on-demand and automatically. Those endpoints scale elastically and are designed to serve feature vectors at sub-100ms to power online predictions.

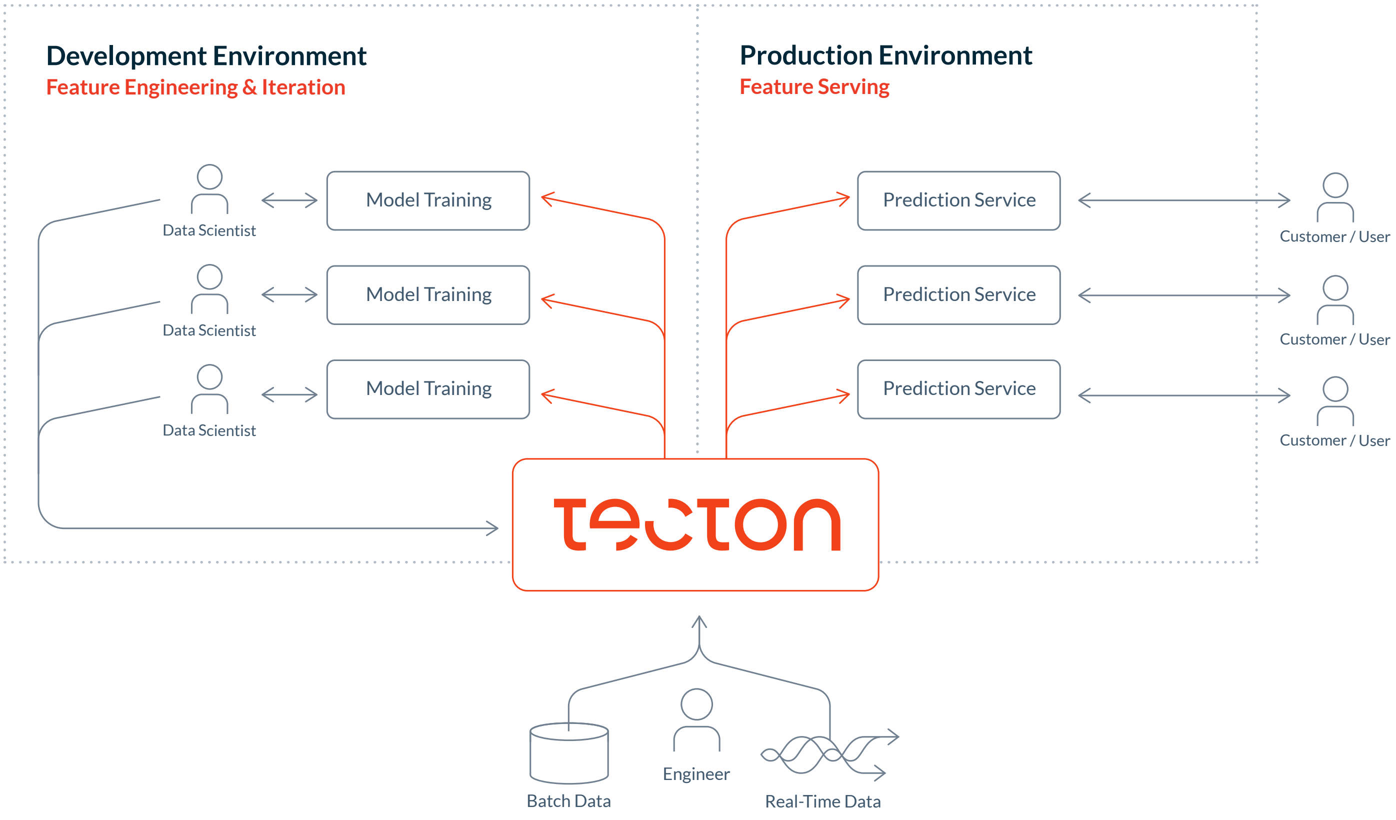
Tecton SDK: Tecton provides a Python SDK that data scientists can use in their interactive notebook environments to retrieve training data from the feature store. Tecton supports continuous (row-level) time travel, providing accurate feature values and eliminating data leakage concerns.

Feature Code Repository: Feature transformation code and feature configuration are managed in a version-control Git repo.

Tecton Web UI: The web UI enables teams to search, manage and monitor features.

Build a library of great features. Serve them in production. Do it at scale.

Tecton enables data science teams to develop high-quality features using DevOps-like engineering principles. It puts data scientists in control of their data from development to production, without relying on data engineering teams to re-implement production pipelines.



Share, discover and re-use features: Instead of re-implementing duplicate features across models, data scientists can search and discover existing features to maximize re-use across models.

Manage features as code: Data scientists manage their features in a Git version-control repository to track changes and coordinate work between team members. Teams can build a library of well-maintained, consistent feature transformations that facilitate collaboration.

Build accurate training data sets using time travel: Data scientists can easily retrieve accurate training data from the Tecton feature store directly into their notebook environments. Training data is provided with continuous (row-level) time travel that prevents data leakage.

Create high-quality features using batch and real-time data: Tecton enables teams to build high-quality features using complementary data sources such as data warehouses and event streaming platforms. Batch data sources are most useful for large-scale aggregations with low freshness requirements, while real-time sources are most appropriate for simple transformations with high freshness requirements.

Deploy and serve features in production instantly: Data scientists can instantly serve features in production by simply applying a configuration change, accelerating time-to-market by months and reducing the load on data engineering teams.

Eliminate training/serving skew: The Tecton Data Store provides a single source of data for training, batch inference, and online serving, eliminating training-serving skew and increasing the accuracy of online predictions.

Monitor production features: ML teams can monitor features for breakages, drift, and serving latency.

Tecton complements emerging MLOps platforms such as AWS SageMaker and MLflow. Together, they provide a complete stack for operational ML that brings DevOps agility to both the modeling and data layers. This new stack makes world-class machine learning accessible to every company. ML teams can get new models to production faster and iterate quickly. Line of business owners can accelerate the time-to-market for new applications to drive real business outcomes and increase the ROI on their ML investments.