TECTON + **kafka**

Automate the transformation of topics into features for production machine learning

Solution Overview

Together, Tecton and Kakfa provide a simple and fast path to building and serving production-grade features from data streams to support a broad range of machine learning applications, including fraud detection, recommendation systems, dynamic pricing, search, and much more.

Available to all Confluent customers and those managing Kafka themselves, Tecton acts as the central hub for ML features, allowing data teams to define features as code using PySpark, Python, or SQL and then automating production-grade ML data pipelines across batch, streaming and real-time sources to generate accurate training datasets and serve freshly computed features online for real-time inference.

Key Benefits

Power
Build more powerful models by easily incorporating batch, streaming, and real-time data.
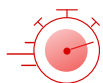
Speed
Deliver value from real-time models in minutes rather than months.

Flexibility
Continuously improve and iterate on production ML models across teams and use cases.

Key Challenges for Production ML

Tecton solves the many data and engineering hurdles that keep development time painfully high and, in many cases, prevent applications from ever reaching turning data streams into machine learning features in production at all, including:

- Training-serving skew
- Point-in-time correctness
- Late arriving and out-of-order data
- Productionizing notebooks
- Real-time transformations
- Melding batch + real-time data
- Latency constraints
- Data scientist and data engineering siloed workflows
- Limited discovery and re-use of features across teams
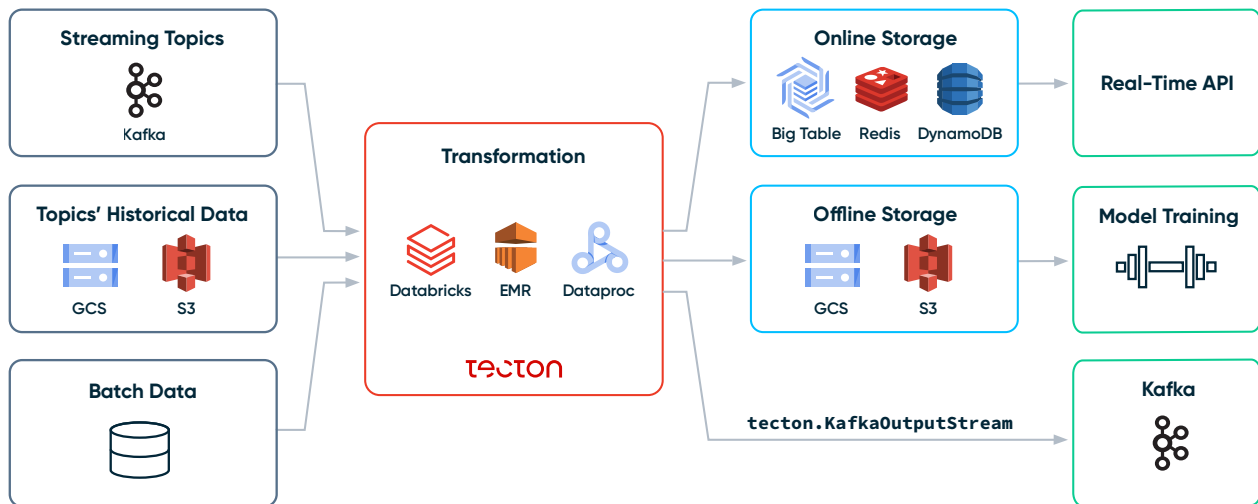
How it Works

Tecton enables data teams to easily build, test, and deploy feature pipelines across batch, streaming, and real-time data pipelines with minimal code and automated orchestration.

Kafka topics are added as streaming data sources to define features. These topics are registered with an associated batch source representing the stream's historical event log, an accurate historical record of the stream, to enable materialization to an offline store and bootstrap an online store during a backfill.

Tecton executes materialization jobs of stream Feature Views with Spark on your connected data platform to perform transformations, aggregations, and deserializations necessary to produce features from your Kafka streams. Tecton manages watermarks for each Feature View that allows you to fine-tune the acceptance or rejection of late arriving or out-of-order data for each topic.

Tecton automatically creates, runs and monitors these materialization jobs and places the resulting feature views in an offline store for training data generation and an online store for real-time inference. Feature Views can be sent to Kafka as well, via a KafkaOutputStream that defines the Kafka topics the records will be appended to.

> "
> *We have events coming on Kafka as well as batch jobs that are feeding into Tecton. For us, Tecton is a unified layer of historical and real-time features. If you are looking for a particular data set, Tecton makes it available, both historically in order to do analytic tasks on top of constrained models, estimate rule impact, and then also makes it available in real-time, so that your systems can use it in order to make low latency decisions.*
>
> Hendrik Brackmann
> VP Data
>
> **tide.**

Get started today!

Request a demo or a free trial at **tecton.ai**.

tecton