Enhancing Machine Learning Models

With Real-Time Data Pipelines

Machine learning models are often times based on real-time calculations for the use cases they serve. Companies have a brief window of opportunity to make predictions that can have an immediate business impact. Real-time data pipelines can be used to enhance machine learning models.

InfinyOn

According to the Wikipedia definition, "Machine learning (ML) is the study of computer algorithms that can improve automatically through experience and by the use of data. It is seen as a part of artificial intelligence. Machine learning algorithms build a model based on sample data, known as training data, in order to make predictions or decisions without being explicitly programmed to do so." Essentially the model is only as good as the data that its fed.

Streaming data is an excellent way to feed machine learning models in order to make more accurate predictions. The machine learning model provides logic that assists the streaming data pipeline to uncover elements within the stream and potentially within historical data. Real-time scores based on the elements are generated and delivered to BI tools and applications so they can make their predictions or recommendations.
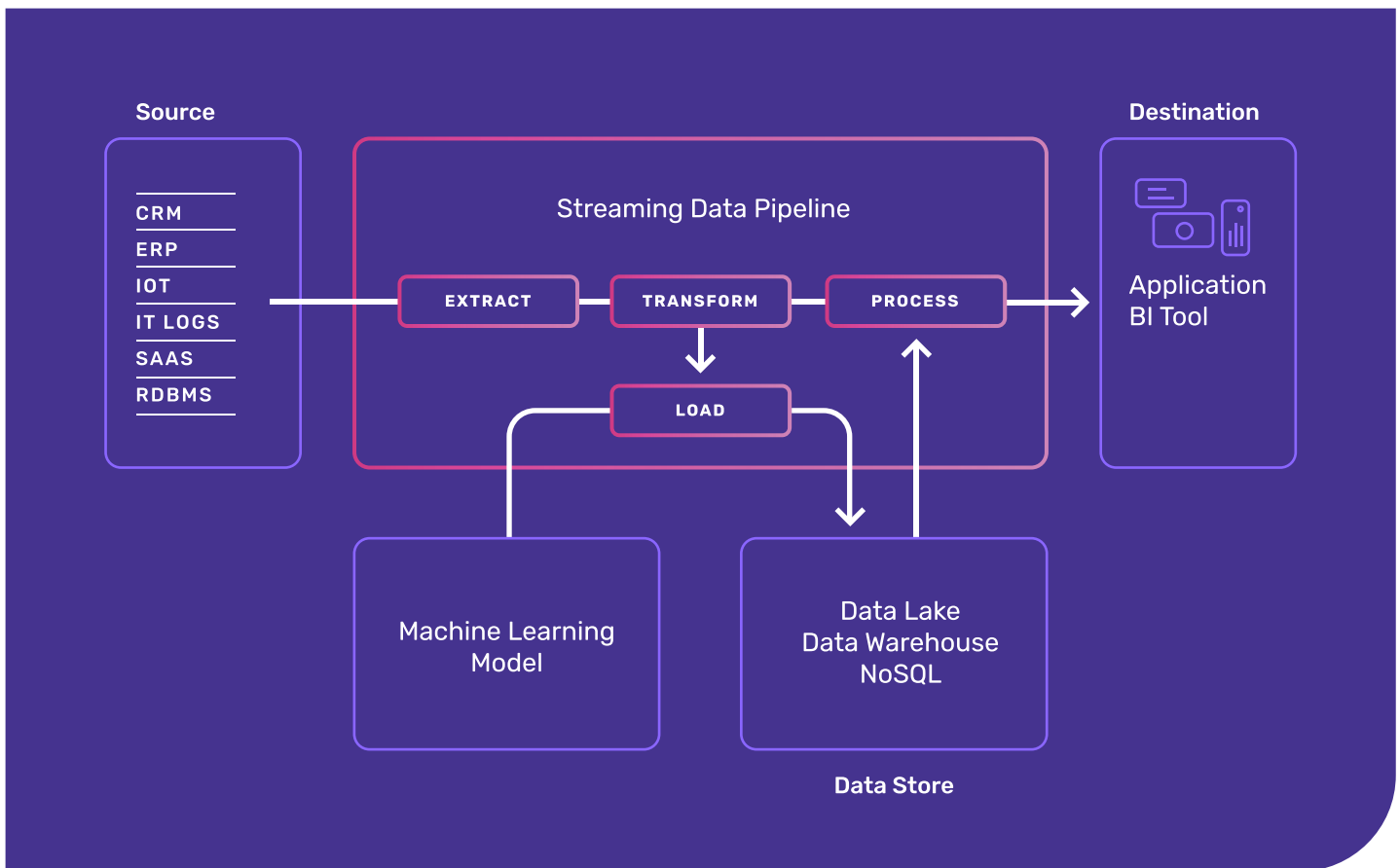
Streaming data is an excellent way to feed machine learning models

Diagram 1. Streaming ML Architecture

Data Sources

Modern data-driven enterprises typically have a vast set of data sources that often include Customer Relationship Management (CRM), Enterprise Resource Planning (ERP), IoT sensors, IT logs, SaaS Applications and Relational Database Management Systems (RDBMS). Source connectors can be used to stream data from these systems into the Fluvio data streaming platform. Apply custom programs with **SmartModules** to aggregate, correlate and transform data records in real-time as they move over the network.

Streaming Data Pipeline

Streaming ETL pipelines capture events in real-time and are far more efficient than batch processing. ETL data pipelines extract, transform and load data from operational sources into data warehouse targets.

Extract, Transform and Load (ETL)

APIs, Change Data Capture (CDC) or protocols such as MQ Telemetry Transport (MQTT) identify and capture events from various sources. The Fluvio event stream processing engine can capture and consolidate the events into streams that route to one or more destinations. Fluvio can also transform the streaming events to reformat data, join streams, or derive new values based on events. Finally, Fluvio will load the transformed data to the target such as Amazon S3, Databricks or Snowflake.

Process

The stream processing engine used in this example is Fluvio to support the streaming analytics powering the machine learning model. Fluvio can filter, evaluate and query the streaming events or look up historical elements within the data store.

Machine Learning Model

The machine learning model depicted in the diagram 1 leverages both historical and real-time streaming data. Data Engineers and Data Scientists can enhance their models and build more robust predictions and recommendations by using **Fluvio** open-source software or **InfinyOn Cloud** as a managed service.

> ## Companies have a brief window of opportunity to make predictions that can have an immediate business impact.

About InfinyOn

InfinyOn, a real-time data streaming company, has architected a programmable platform for data in motion built on Rust and enables continuous intelligence for connected apps. SmartModules enable enterprises to intelligently program their data pipelines as they flow between producers and consumers for real-time services. With InfinyOn Cloud, enterprises can quickly correlate events, apply business intelligence, and derive value from their data. To learn more, please visit infinyon.com.

InfinyOn