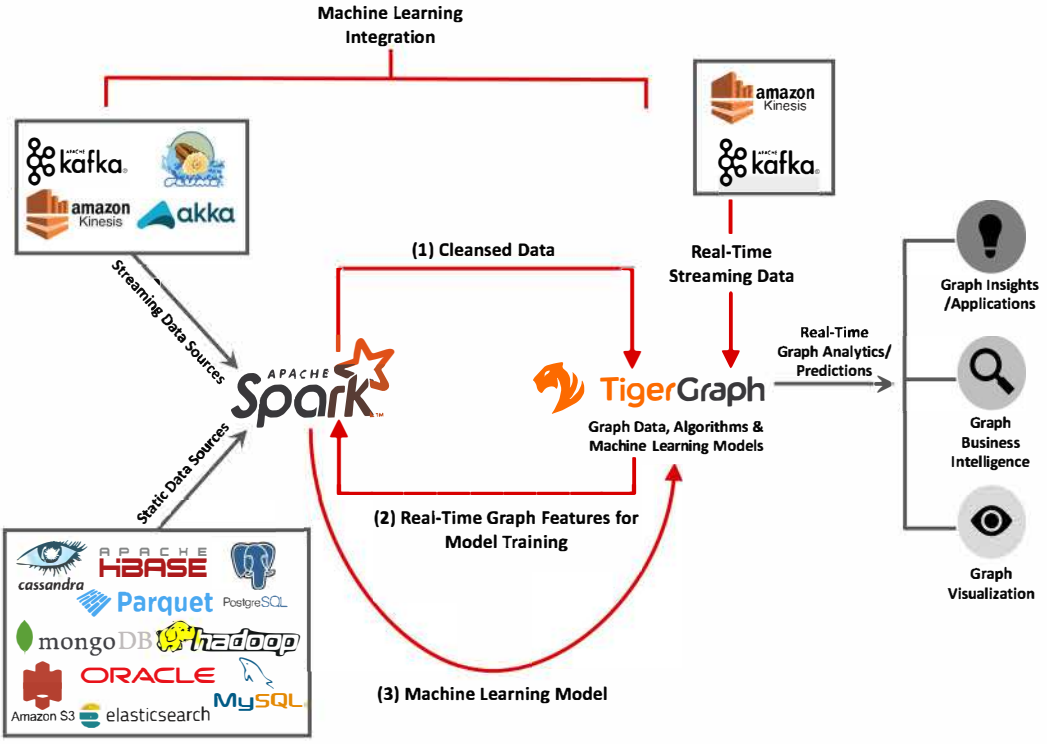
Spark+AI+TigerGraph: A Happy Marriage

TigerGraph is the perfect complement to the power of Spark technology. Bringing scalable real-time deep link graph analytics in both OLTP and OLAP fashion, TigerGraph augments Spark with graph analytics and predictions for common Machine Learning use cases.

ETL with Spark in Your TigerGraph Data Pipeline

"Where wicked fast scalable ETL meets wicked fast scalable graph analytics for today's ever growing big datasets"

If you are already using Spark in your data pipeline to clean static and streaming data, TigerGraph's high-speed loading and data integration helps you ingest your cleansed data sources in real-time (milliseconds). With Spark as part of the ETL process and TigerGraph as the trusted system of record and graph analytics computation engine, you'll accelerate the pace to the business value for your connected data. Say "No" to sleepy data in Data Lakes!



With built-in ETL, parallel loading, and OLTP capability, TigerGraph can also independently import data from other data sources directly and integrate data silos into a system of record for transactional data. What's more, TigerGraph can be an excellent helper for advanced ETL process requiring entity resolution and master data management, supporting Spark's data cleansing process by untangling the relationship among the data sources.

TigerGraph + Spark Machine Learning Integration

"Unlock your data's value in record time with real-time ingestion and real-time graph feature extraction via TigerGraph"

Make TigerGraph and Spark part of your graph-enhanced Machine Learning pipeline: Use TigerGraph to extract (graph and non-graph) features and to apply the model on streaming data; use Spark to train and tune machine learning models at scale. Together, the integration provides deeper real-time insights for machine learning.

Data Preparation and Integration (TigerGraph/Spark)
As discussed in the previous section, both TigerGraph and Spark are excellent tools for data preparation. Graphs are particular good for merging datasets and for entity resolution.

Unsupervised Learning (TigerGraph)
TigerGraph's real-time deep-link analytics and open-source graph algorithm library are the natural choice for unsupervised learning approaches, such as finding clusters, frequent patterns, similarities, rankings, and outliers.

TigerGraph

Spark+AI+TigerGraph: A Happy Marriage

Feature Extraction for Supervised Learning (TigerGraph/Spark)
Machine learning is about seeing the connections between things, and graph is a natural medium for expressing connections. Unsupervised learning can be an end in itself, or its output can be used as additional features to enrich your training data. You can also write pattern-matching queries to measure targeted features. For example, China Mobile uses TigerGraph to map out the network of all phone calls over the past two months and to extract 118 graph features for each of the hundreds of millions of phone numbers to train its anti-fraud system.

Model Training (Spark)
Spark's distributed and parallel processing and its rich set of machine learning libraries makes it a great choice for taking extracted features to train a model based on those features.

Validate and Apply Model (TigerGraph)
Using the China Mobile example again, when a new phone call comes in, they extract the 118 features in real-time for the current caller in the current call graph, to categorize the call as one of three types of unwanted calls. Each new phone call modifies the graph in real time.

Visualize and Explore Internconnectd Data
TigerGraph's intuitive visual SDK, GraphStudio, empowers data scientists and business users to explore and analyze realtionships among interconnected business objects.

Graph Algorithms by TigerGraph - The Most for Your Money

TigerGraph is the best-in-class choice for performant, scalable, and cost-effective graph algorithms, including full-graph, deep-link, and iterative algorithms like PageRank and Community detection. The reason? TigerGraph is built from the ground up for this workload. Native graph storage architecture, compressed data to minimize footprint and cost, and massively parallel processing (MPP) mean crazy-fast and distributed graph operations. Now add an open-source graph algorithm library written in GSQL, TigerGraph's high-level, Turing-complete graph query language, and you have the best platform for graph algorithms and analytics. You can run graph algorithms on Spark for convenience, but the mismatch between DataFrames and graph architecture will always limit your performance.

TIGERGRAPH BENEFITS

Deep Link Analytics - Uncover relationships and make correlations that were unattainable before. TigerGraph easily traverses 3 to 10+ hops per query.

Fast Loading and Real-Time Updates - Loads 50 to 150 GB/hour/machine online. Handles 100M+ update/machine/hour.

Fast Processing - Whether it's high volume transactions or complex, multi-hop analytics, TigerGraph delivers real-time results. Traverse 100M+ edges/sec.

Distributed - Scales and grows with you. Distribute your data and processing across multiple nodes in a cluster.

Battle-Tested - In production for more than 4 years in mission-critical applications at leading companies.

CONTACT

TigerGraph
3 Twin Dolphin Drive, Suite 225
Redwood City, California 94065
United States

www.tigergraph.com

About TigerGraph

TigerGraph is the only scalable graph database for the enterprise. Based on the industry's first Native and Parallel Graph technology, TigerGraph unleashes the power of interconnected data, offering organizations deeper insights and better outcomes. TigerGraph fulfills the true promise and benefits of the graph platform by tackling the toughest data challenges in real time, no matter how large or complex the dataset. TigerGraph's proven technology supports applications such as IoT, AI and machine learning to make sense of ever-changing big data, and is used by customers including China Mobile, Intuit, VISA, Wish and Zillow. The company is headquartered in Redwood City, CA, and is backed by investors including the founders of Yahoo!, Walmart Labs, DCVC and Morado Ventures. Follow TigerGraph on Twitter at @TigerGraphDB or visit www.tigergraph.com.