Solution Showcase

Real-Time Data Analytics for Finance Applications

Kinetica's GPU-powered database enables real-time impact analysis for claims management and trading risk strategies.

Image courtesy of Kinetica

> "We built Kinetica from the ground up to provide real-time insights. Data is continuously ingested at speed and scale and is available for analysis—be that for complex joins and aggregations, entity graphing, vector similarity search, or real-time generative retrieval."
>
> **Nima Negahban**,
> CEO and Co-founder, Kinetica

As the volume of data in the financial services industry explodes, traders, bankers, and lenders increasingly need to leverage live data in real time. To identify and halt sophisticated financial crimes, organizations require solutions that can detect emerging patterns and adapt in real time as new threats unfold. Meanwhile, analyzing portfolio risk, credit exposure, and trade costs demands real-time insights into billions of daily market movements and transactions.

Kinetica's high-performance database addresses these challenges by processing massive real-time data streams, enabling businesses to gain immediate insights, detect anomalies, and make informed decisions to maximize revenue while ensuring compliance standards for financial trading, including MiFID II reporting requirements, disclosure of transparency data, and market surveillance.

Real-Time Financial Analytics

To monitor the impact of billions of market transactions per day on complex portfolios, Kinetica fuses streaming market data with static reference data for a full and constantly updated picture of positions with context on risk and exposure. A lockless architecture and advanced vectorized computing algorithms seamlessly re-aggregate data as it streams in, empowering teams to perform intricate analyses on the fly with the latest information.

kinetica

Kinetica is a high-performance database purpose-built for generative AI and real-time data. Leveraging the power of NVIDIA accelerated computing, Kinetica executes complex analytics on large volumes of both real-time and batch data.

Industry

- Trading
- Banking
- Insurance

Benefits

- Real-time data analytics
- Increased fraud detection
- Reduced false positives
- Reduced total cost of ownership

NVIDIA Hardware

- NVIDIA A100 GPUs
- NVIDIA A10 GPUs

NVIDIA Software

- NVIDIA® CUDA®
- NVIDIA cuVS
- NVIDIA NeMo™ Retriever
- NVIDIA RAPIDS™
- NVIDIA NIM™

NVIDIA Preferred Partner

Benefits of Kinetica's Real-Time Analytics Include:

Real-time updates: Vectorized processing enables faster ingestion of data and faster calculations, reducing the need for time-consuming indexes.

Lower TCO: Vectorized processing boosts power, making simpler data structures possible, which in turn leads to lower storage costs, more flexibility, and less time engineering data.

Time series analysis: Kinetica efficiently handles time series data and offers powerful tools for analysis, including context from historical data.

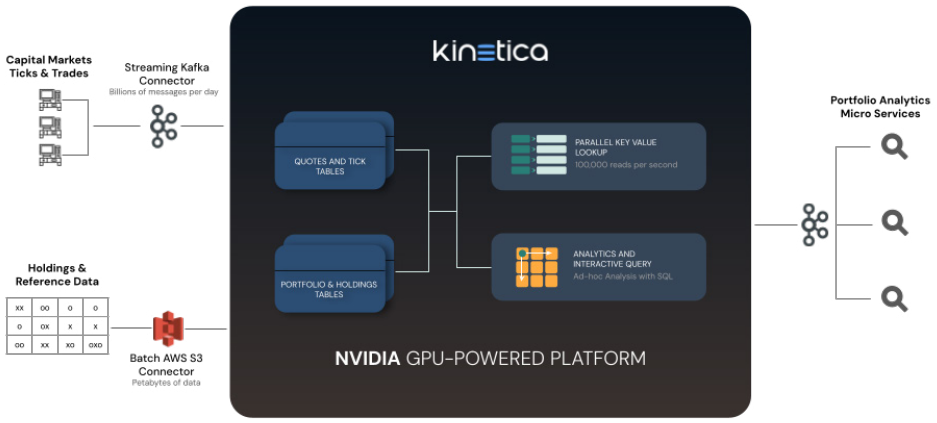

Figure 1: Continuous aggregation and analysis on demand using the up-to-the-moment data with Kinetica and NVIDIA accelerated computing.

Financial Institutions Leverage Kinetica for TCA Rollup Metrics

Financial analysts use transaction cost analysis (TCA) to measure the performance of each trade for metrics such as slippage, fill rate, and latency. TCA rollup metrics are used by institutions as summary statistics that aggregate trade-level data across a specified time period or across multiple trades.

Existing data analytics tools for distributed data processing, while suitable for a variety of tasks, are often less ideal for TCA rollup metrics. This is typically due to latency when processing extremely large volumes of data, the high cost of deploying such tools, and the inability to get actionable insights in front of traders in real time.

Global banking groups and leading hedge funds are using Kinetica's real-time analytics database to overcome these challenges.

Due to its ability to handle complex data operations with low latency, Kinetica's native vectorized database is ideal for TCA rollup metrics that involve high cardinality joins and time series windowing functions.

While conventional distributed analytic platforms like Spark process data row by row and demand a great amount of computing resources, Kinetica's vectorized approach reduces the amount of compute and data engineering required, helping financial institutions control costs.

> "With Kinetica, finance customers can easily fuse their streaming data with static data for real-time multimodal analysis. This includes OLAP, time series, entity graphing, and real-time generative retrieval, leading to richer insights, better decision-making, and more responsive event handling."
>
> **Gwyn Isaac**,
> Head of Finance, Kinetica

At Citibank, for example, their migration to Kinetica reduced a 700-node Spark cluster to just 16 nodes on Kinetica, using NVIDIA A100 tensor core GPUs (32 nodes with active-active high availability), delivering superior results, including:

> Loading more than 5 billion records a day
> 1-second data latency SLA
> 2-second query latency SLA

With Kinetica, banks and traders can achieve fast and accurate rollup calculations to quickly identify trends and patterns across trade desks, as well as gain company-wide visibility of TCA across the enterprise.

Property and Casualty Insurance Company Leverages Kinetica for Real-time Claims Management

A leading property and casualty insurance corporation is using Kinetica's solution to revolutionize real-time claims management, enhancing both the policyholder's experience and the insurer's bottom line.

Kinetica combines real-time weather data from sources like the National Weather Service with policy data from their client data warehouse, merging it with building footprint information using advanced geo-joins and temporal joins. This integration allows for the invocation of containerized machine learning models to predict damage, severity, and liability on the fly.

The billions of data points are then visualized in real time on a highly interactive map, providing detailed insights without the need for data aggregation. Kinetica's generative AI capabilities enable analysts and end users to ask natural language questions of their constantly updated data, offering a real-time operational picture of their entire insured property portfolio. To achieve this, Kinetica's client uses A10 GPUs to power the interactive analysis on massive scale maps.

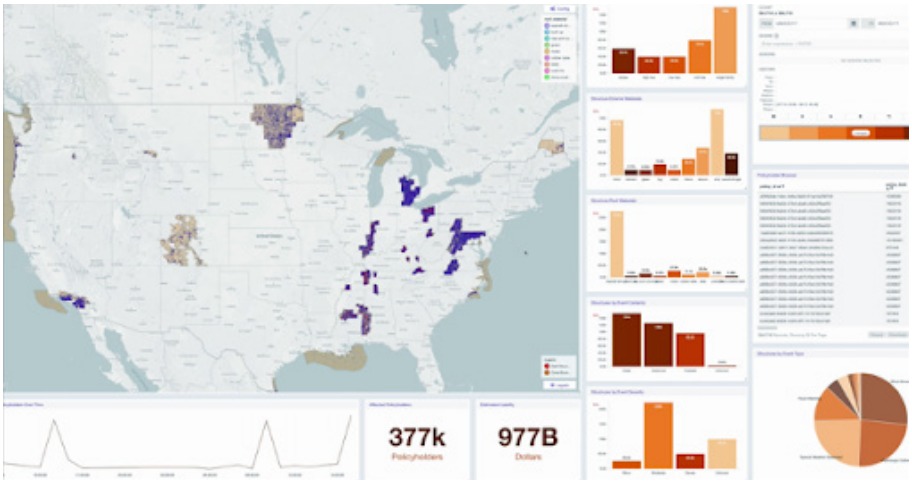

Figure 2: Real-time assessment of severe weather events on an entire nationwide property portfolio with Kinetica and NVIDIA accelerated computing.

This real-time approach **accelerates the claims process** by integrating real-time data sources with spatial and actuarial datasets, reducing the time from the first notice of loss (FNOL) to claim settlement. It also **improves claims adjuster efficiency** by directing them to the most impacted areas and providing predictive insights ahead of weather events. Additionally, Kinetica **enhances the ability to detect claim fraud** by using machine learning models with real-time data to identify discrepancies between reported claims and predicted damages.

Powered by NVIDIA

Kinetica was built from the ground up to harness the parallel processing capabilities of NVIDIA accelerated computing, including software solutions such as NVIDIA cuVS, NVIDIA NIM, and NVIDIA NeMo Retriever. By storing data in fixed-size blocks known as vectors, Kinetica can execute query operations in parallel across thousands of CUDA cores. This approach enables the query engine to process multiple data elements simultaneously, leading to faster query execution, enhanced performance, and a reduced total cost of ownership. Kinetica's integration with CUDA and NVIDIA GPUs ensures optimal performance for the complex real-time analytics demands of today's financial services industry.

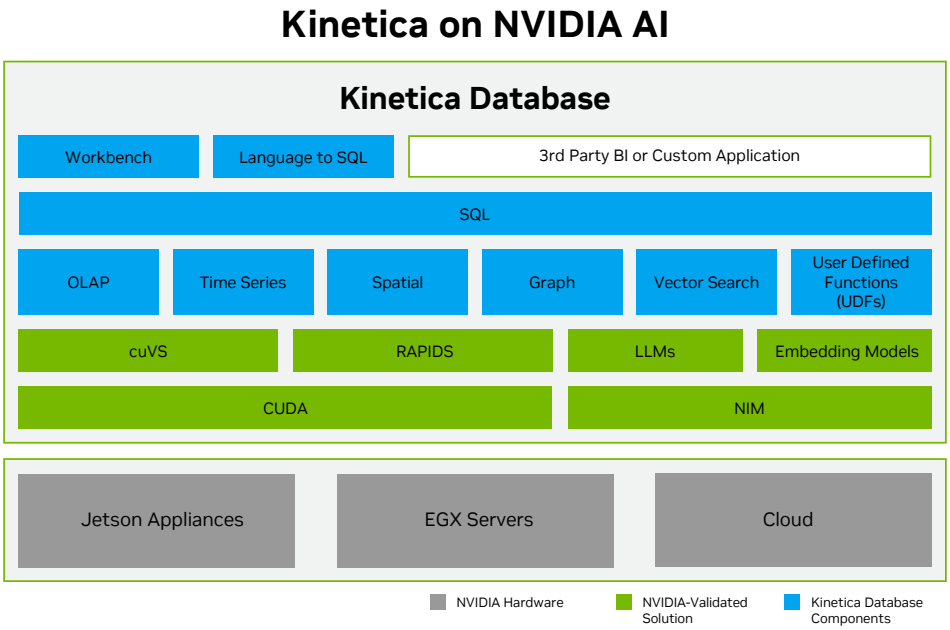

Figure 3: Kinetica's vectorized computing algorithms are built on NVIDIA GPU-accelerated software.

For generative applications, Kinetica can orchestrate embedding generation within the database through SQL. Developers can leverage NIM and NeMo Retriever microservices, such as a NeMo Retriever embedding NIM (embed-qa-4), by defining a remote model in SQL and using the GENERATE_EMBEDDINGS table function during streaming or batch data ingestion.

Kinetica uses NVIDIA cuVS and its approximate nearest neighbor algorithms (ANN) to deliver real-time vector similarity search for production AI applications. By leveraging NVIDIA cuVS in conjunction with its own internal data management systems, Kinetica empowers businesses to achieve lightning-fast, real-time vector search, enabling them to glean actionable insights from their data with remarkable speed and efficiency.

In addition to Kinetica's built-in capabilities for vector, spatial, time series, and graph analysis, developers can extend Kinetica with custom algorithms and machine learning models using user-defined functions (UDFs). Kinetica UDFs can leverage NVIDIA RAPIDS libraries, including cuDF for DataFrame operations, cuML for machine learning, and cuGraph for graph analytics. By minimizing data movement and utilizing GPU acceleration, these custom UDFs can provide tremendous performance. Since they can be invoked through SQL, they can easily be integrated and combined with Kinetica's standard analytic functions.

Unleash the Power of Real-time Data in Finance

Kinetica's high-performance database delivers unparalleled real-time insights, enabling financial institutions to detect and prevent fraud, manage risk, and optimize trading strategies with precision and speed. By leveraging NVIDIA's accelerated computing platform, Kinetica ensures that these critical operations are not only faster and more efficient but also cost-effective, empowering traders and bankers to stay ahead in a data-driven market.

Ready to Get Started?

To learn more about NVIDIA solutions for the financial service industry, visit: nvidia.com/finance

To learn about detecting and preventing fraud with Kinetica, visit: kinetica.com/solutions/financial-crime-fraud/

© 2025 NVIDIA Corporation and affiliates. All rights reserved. NVIDIA, the NVIDIA logo, ConnectX, DGX, HGX, Hopper, NGC, NVIDIA-Certified Systems, and NVLink are trademarks and/or registered trademarks of NVIDIA Corporation and affiliates in the U.S. and other countries. Other company and product names may be trademarks of the respective owners with which they are associated. 3842801. MAY25

NVIDIA