

Advanced In-Database Analytics, on the GPU, with Kinetica

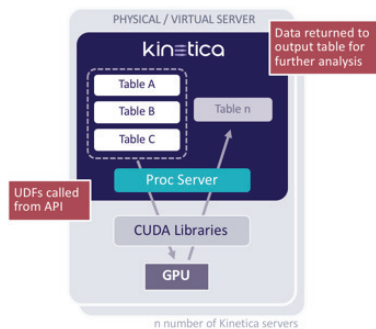
Summary

User-defined functions (UDFs) enable GPU-accelerated data science logic and custom code to power advanced business analytics, on a single database platform. This is the first time such In-database processing functionality has been made available on a database that fully utilizes the parallel compute power of the GPU on a distributed platform. In-database processing in Kinetica creates a highly flexible means of doing advanced compute-to-grid analytics.

This industry-first functionality stands to help democratize data science. Until now, organizations have typically needed to extract data to specialized environments to take advantage of GPU acceleration for data science workloads, such as machine learning and deep learning. Kinetica now makes it possible for sophisticated data science workloads to be deployed and made available on the same database as is used for business analytics.

How It Works

User-defined functions and the associated orchestration API enable data to be processed with custom code that can draw on the power of distributed GPUs. UDFs have direct access to CUDA APIs, and can take full advantage of the distributed architecture of Kinetica. Because Kinetica is designed from the ground up to take full advantage of the GPU, users have an advanced mechanism for distributed computation.



- UDFs are able to receive filtered data, do arbitrary computations, and then save output to a separate table. Such computations might include linear interpolation, anomaly detection, clustering, regressions, or risk simulations such as Monte Carlo analysis. The brute-force parallel compute power of the GPU delivers fast response which makes it more suitable for interactive analytics and experimentation.
- GPUs are particularly well suited for the types of vector and matrix operations found in machine learning and deep learning systems. With in-database processing, custom functions will be able to call machine learning/artificial intelligence libraries such as TensorFlow, BiDMach, Caffe, Torch and others to work directly on data within Kinetica.
- The orchestration API is available with Fuzzy Logix integration offers in-database and GPU-based analytics on comprehensive and growing libraries of over 600 mathematical, statistical, simulation, data mining, time series and financial models.
- TensorFlow bundled with Kinetica offers a unified solution for data preparation, model training, and model deployment into production.

Why This Matters

As businesses move from OLAP to predictive analytics and machine learning, more and more operations stand to benefit from the parallel processing power of the GPU.

For organizations requiring low latency response to sophisticated analytics on large datasets, Kinetica's in-database functionality will offer more customization in addition to significantly faster operations. Pre-existing custom code that operates in separate systems can often be quickly reconfigured to run in Kinetica. UDFs open up a world of options for automating processes and performing business calculations within the analytics platform.

WHO BENEFITS

In-database processing on Kinetica appeals to customers that want to make advanced and predictive analytics available to line of business users. For example, fraud detection, trade decisioning, risk management, trend and pattern recognition are all use cases that can benefit.

Vehicle Design and Testing

Vehicle manufacturers and sports teams now collect millions of data points from wind tunnel runs and testing. These are used in 1000s of calculations per run that typically take hours to process. UDFs will provide them the ability to perform on-demand calculations and adjust parameters in real time.

Pricing and Risk Calculations

Financial institutions are merging more and more sources of data and need to run ever more sophisticated risk management algorithms and return results pre-trade. UDFs will enable a next generation risk management platform that also enables real-time drill-down analytics and on-demand risk modeling.

Genomics Signalling

Pharma companies utilize genomics data to identify new signals and to accurately predict drug targets. Natural Language Processing (NLP) can be used to extract specific features and simulation algorithms can be called to locate genome signals and accelerate drug development.

MOST BENEFIT FROM GPU				
Simple Reporting	Standard Analytics	Real-time Analytics	Machine Learning	Deep Learning
List defaults from customers in the last 3 years.	What is the default rate for customers over a certain age, by region? by income?	What is the risk-profile of this customer up to and including the transactions he made 10 seconds ago?	Given location, buying history, demographic, past-history, past-purchases, what is the likelihood this customer will default?	Deduce from unspecified signals across a wide range of datasets the likelihood this customer will default?

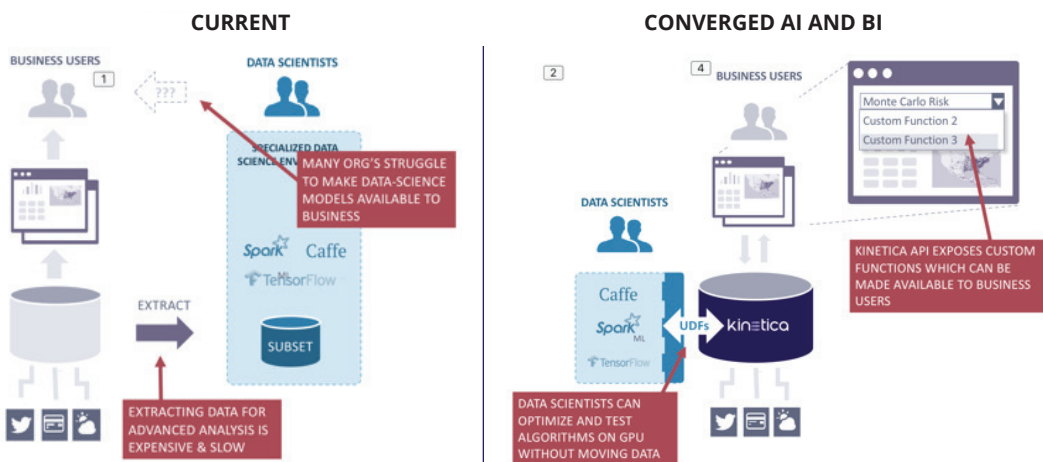
With predictive analytics and more advanced data science workloads, organizations have often resorted to exporting data to specialized high-performance computing (HPC) systems staffed by data scientists who can work on these more complex problems. These groups demand GPU-acceleration for the performance benefits. Because deep learning algorithms run several times faster on a GPU compared to a CPU, learning times can be reduced from weeks to hours, or hours to minutes. This performance boost is critical because such algorithms need to be tweaked, then trained, and the time to iterate adds up.

Moving data to such specialized systems is cumbersome and slow, and it's common for data science to become disconnected from the business with no easy way to return models into production use.

Bringing Data Science Modeling and Business Analytics Together on the Same Platform

With in-database processing on Kinetica, BI and AI workloads can run together on the same GPU-accelerated platform. Doing so eliminates the time and effort required to transform data and move it back and forth between a database and a separate data science system.

This empowers business users to do more sophisticated analysis without resorting to code. Data science teams can develop and test gold-standard simulations and algorithms while making them directly available on the systems used by end users. Foreseeably, in addition to query, reporting and analytics, users could also call a Monte Carlo simulation, or other custom algorithms, straight from their BI tools. In this way, Kinetica stands to democratize data science.



kinetica

For more information on Kinetica and GPU-accelerated databases, visit kinetica.com

Kinetica and the Kinetica logo are trademarks of Kinetica and its subsidiaries in the United States and other countries. Other marks and brands may be claimed as the property of others. The product plans, specifications, and descriptions herein are provided for information only and subject to change without notice, and are provided without warranty of any kind, express or implied. Copyright © 2017 Kinetica