

Insider

Review: Kinetica analyzes billions of rows in real time

GPU database is not only hugely scalable, but integrates graph analysis, location intelligence, and machine learning with standard SQL



By **Martin Heller**

Contributing Editor, InfoWorld | April 29, 2019

AT A GLANCE

Kinetica 7

★★★★★

In 2009, the future founders of Kinetica came up empty when trying to find an existing database that could give the United States Army Intelligence and Security Command (INSCOM) at Fort Belvoir (Virginia) the ability to track millions of different signals in real time to evaluate national security threats. So they built a new database from the ground up, centered on massive parallelization combining the power of the GPU and CPU to explore and visualize data in space and time. By 2014 they were attracting other customers, and in 2016 they incorporated as Kinetica.



The current version of this database is the heart of Kinetica 7, now expanded in scope to be the Kinetica Active Analytics Platform. The platform combines historical and streaming data analytics, location intelligence, and machine learning in a high-performance, cloud-ready package.

As reference customers, Kinetica has, among others, Ovo, GSK, SoftBank, Telkomsel, Scotiabank, and Caesars. Ovo uses Kinetica for retail personalization. Telkomsel, the Indonesian wireless carrier, uses Kinetica for network and subscriber insights. Anadarko, recently acquired by Chevron, uses Kinetica to speed up oil basin analysis to the point where the company doesn't need to downsample its 90-billion-row survey data sets for 3D visualization and analysis.

Kinetica is often compared to other GPU databases, such as OmniSci, Brytlyt, SQream DB, and BlazingDB. According to the company, however, they usually compete with a much wider range of solutions, from bespoke SMACK (Spark, Mesos, Akka, Cassandra, and Kafka) stack solutions to the more traditional distributed data processing and data warehousing platforms.

Kinetica key features and architecture

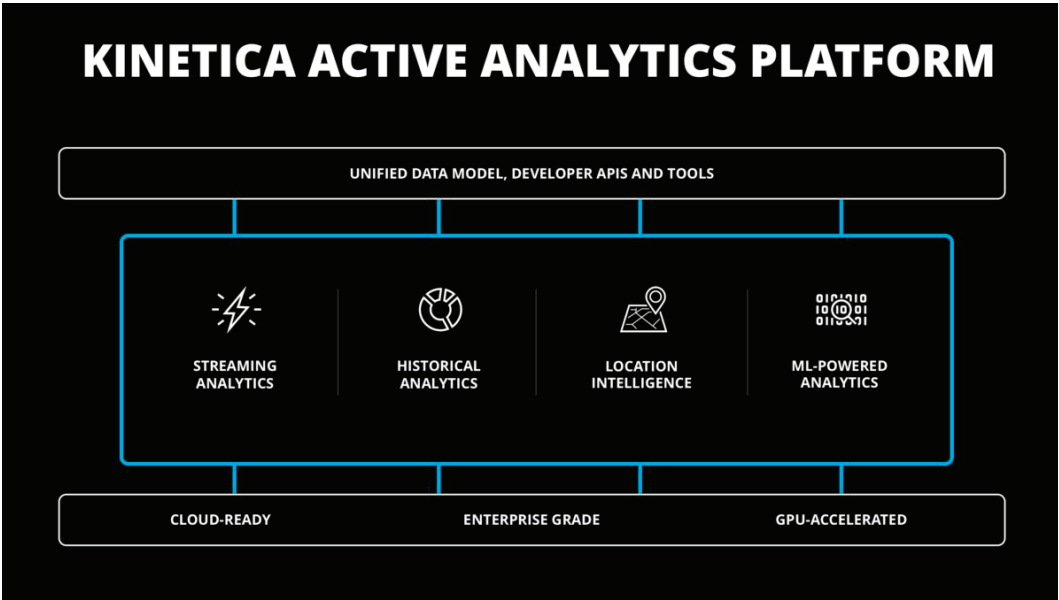
Kinetica combines its distributed, in-memory, GPU-accelerated database with streaming analytics, location intelligence, and machine learning. The database is vectorized, columnar, memory-first, and designed for analytical (OLAP) workloads, automatically distributing any workload across CPUs and GPUs. Kinetica uses SQL-92 for a query language, much like PostgreSQL and MySQL, and supports an extended range of capabilities including text search, time series analysis, location intelligence, and graph analytics.

Kinetica can operate on the entire data corpus by intelligently managing data across GPU memory, system memory, disk or SSD, HDFS, and cloud storage such as Amazon S3. According to the company, this ability to manage all storage tiers is unique to Kinetica among GPU databases.

With its distributed parallel ingest capabilities, Kinetica can perform high-speed ingestion on streaming data sets (with Kafka) and complex analytics on streaming and historical data simultaneously. You can train TensorFlow models against data directly in Kinetica, or import pre-trained TensorFlow or "black box" models to execute inferences via batch processing, stream processing, or public web service.

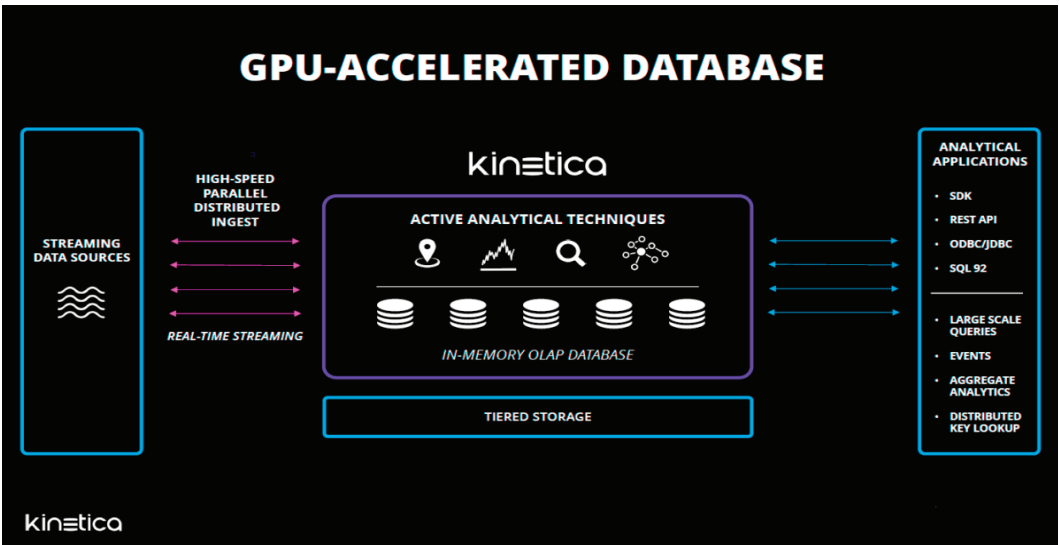
Kinetica has a robust and GPU-accelerated library of geospatial functions to perform on-demand filtering, aggregation, time series, spatial join, and geofence analysis. It can also display unlimited geometry, heatmaps, and contours, using server-side rendering technology (since client-side rendering of large data sets is very time-consuming).

You can use your relational data in a native graph context (by explicitly creating nodes, edges, and other graph objects from relational data) for understanding geospatial and non-geospatial relationships, and you can perform real-time route optimization and even social network analysis using Kinetica’s GPU-accelerated graph algorithms (using the `kinetica.solve_graph` function).



This diagram outlines the high-level functional components of the Kinetica Active Analytics Platform.

Kinetica



This diagram shows how the GPU-accelerated in-memory OLAP database ties to the other capabilities of the Kinetica platform.

Kinetica

Kinetica installation and configuration options

There are three methods for installing Kinetica. The preferred method is now KAgent, which automates the installation and configuration of Kinetica, Active Analytics Workbench (AAW) and Kubernetes, rings (high availability), and more. The two alternative methods are using Docker (for portable installations of Kinetica) and installing manually via command line using common Linux-based package managers such as yum and apt.

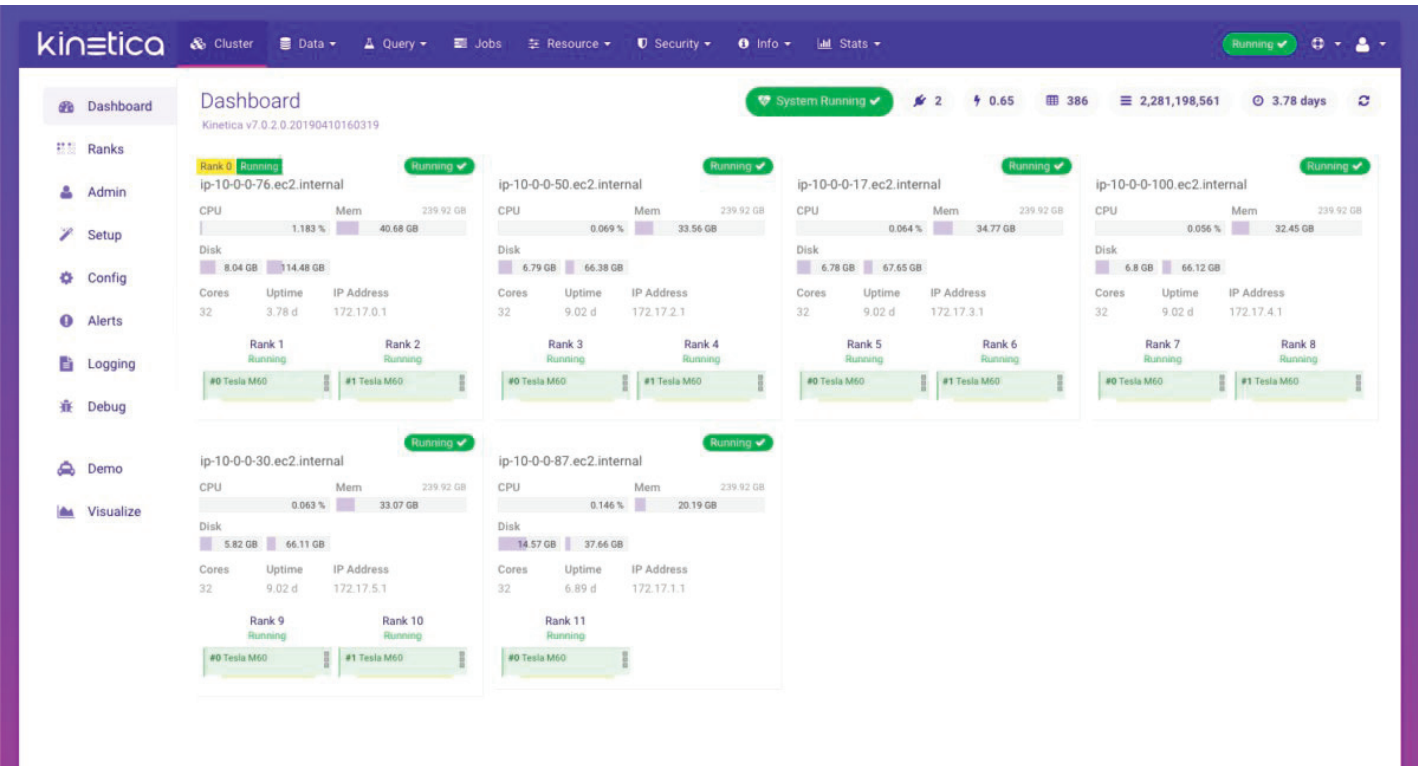
Resource management. Kinetica supports five storage tiers: VRAM, RAM, disk cache, persist, and cold storage. Any operations that make use of the GPU require the data on which they are operating to be located in the VRAM tier. Managing data in these five layers is a non-trivial problem.

Eviction is the forced movement of data from a higher tier to a lower tier in order to make room for other data to be moved into that higher tier. Each object in the system has a level of evictability that is dependent on the type of object it is and the available tiers below it into which it could be evicted. Eviction can be performed in response to a request, which can cause a lot of data movement, or proactively in the background based on high and low watermark levels and eviction priorities, which usually creates less data movement.

High availability. Kinetica HA eliminates the single point of failure in a standard Kinetica cluster and provides recovery from failure. It is implemented external to Kinetica to utilize multiple replicas of data and provides an eventually consistent data store. The

Kinetica HA solution consists of four components: a front-end load balancer, high-availability process managers, one or more Kinetica clusters, and a distributed messaging queue.

Administration. You can administer Kinetica with the graphical GAdmin tool, the Linux command-line service command, or KAgent. The screenshot below shows a GAdmin dashboard for a 6-node cluster.



You can administrate Kinetica using a GUI. Here we see the status of all the nodes in a cluster. The ec2 domain names are a hint that the cluster runs on AWS. Note the Tesla M60 GPUs.

Kinetica demos

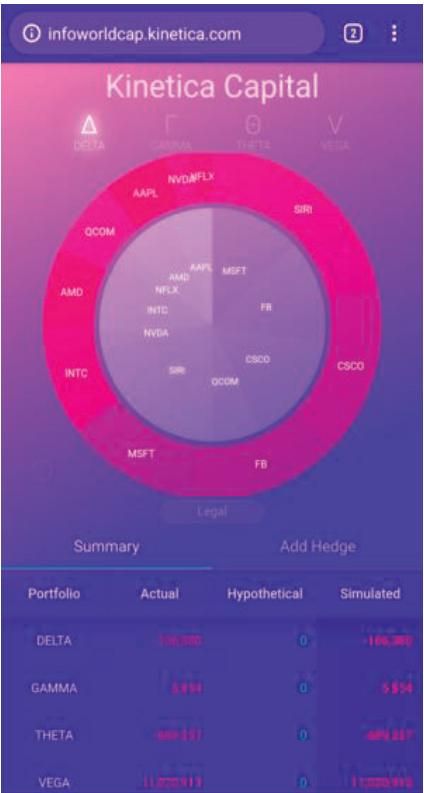
In addition to GAdmin and KAgent, Kinetica offers a web-based visualization tool, Reveal, and the Active Analytics Workbench (AAW), which is for integrating machine learning models and algorithms.

The six-node cluster shown in the screenshot above is the one I used to explore several Kinetica demos. The cluster is comprised of g3.8xlarge instances that each contain two Nvidia Tesla M60 GPUs and 32 Intel Xeon E5 2686 v4 CPUs. Each instance has 244 GiB of RAM and 16 GiB of VRAM per GPU. This setup could be scaled down, up, and out to accommodate any use case. After I finished my tests, the database contained 413 tables and 2.2 billion records.

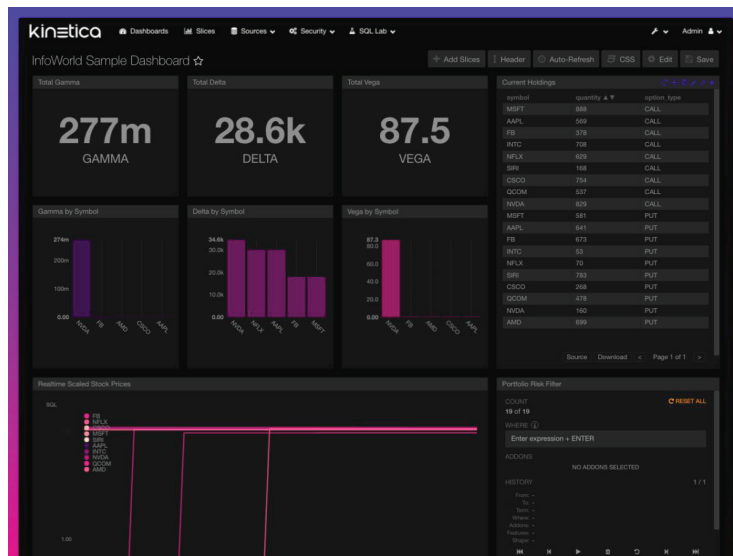
The demos I explored were for financial risk forecasting using options, insurance risk for floods in Texas, network security assessment based on traffic inspection, and taxi rides in NYC. In the process I noticed that, unlike OmniSci’s demos (see my review), which all used single flattened tables (for speed), the Kinetica demos often used multiple tables, views, and analytic dashboards.

Financial risk forecasting with options

This application is essentially a proof of concept of real-time financial risk management with Kinetica. A React mobile app and two web dashboards allow a risk manager to see all the “greeks” (factors in measuring risk) for his or her portfolio and add hedges. Behind the scenes, transactions stream into the database and a Black Scholes machine learning risk model updates continuously on the live data. By contrast, traditional risk management involves copying the transaction data to a separate cluster that runs risk models nightly.

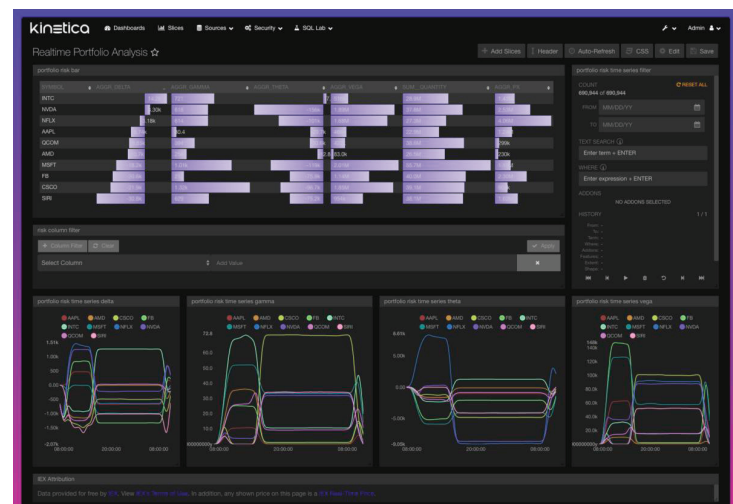


This screenshot from my Android phone displays a continuous real-time mobile risk forecast dashboard based on the “greeks” of an option portfolio. The front-end is a custom React web app, and the back-end of this app runs in Kinetica.



IDG

The screenshot above shows a Kinetica Reveal web dashboard for the continuous real-time risk forecast. This dashboard shows three of the “greeks” and the contributions from the individual symbols in the portfolio, a table of holdings, and a workspace for applying filters for exploration.

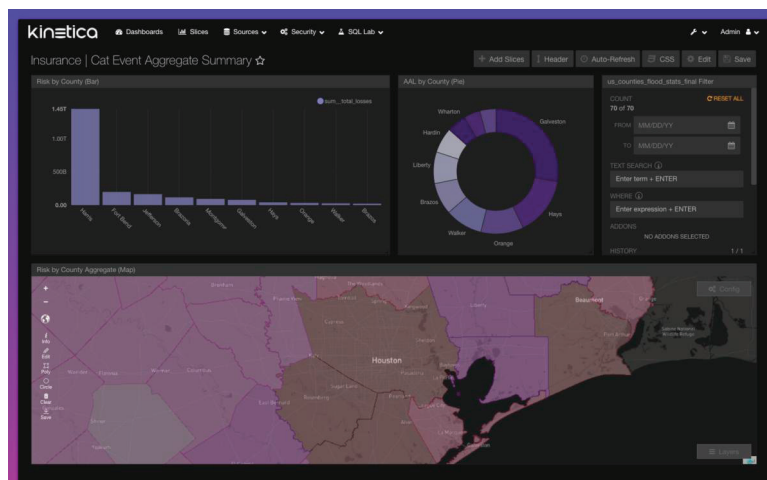


IDG

The screenshot above shows another Kinetica Reveal dashboard for the same real-time portfolio risk management scenario. This one shows a real-time bar chart that aggregates the greeks for each symbol, line charts showing the greeks over time for each symbol, and filters for the time series and aggregate displays.

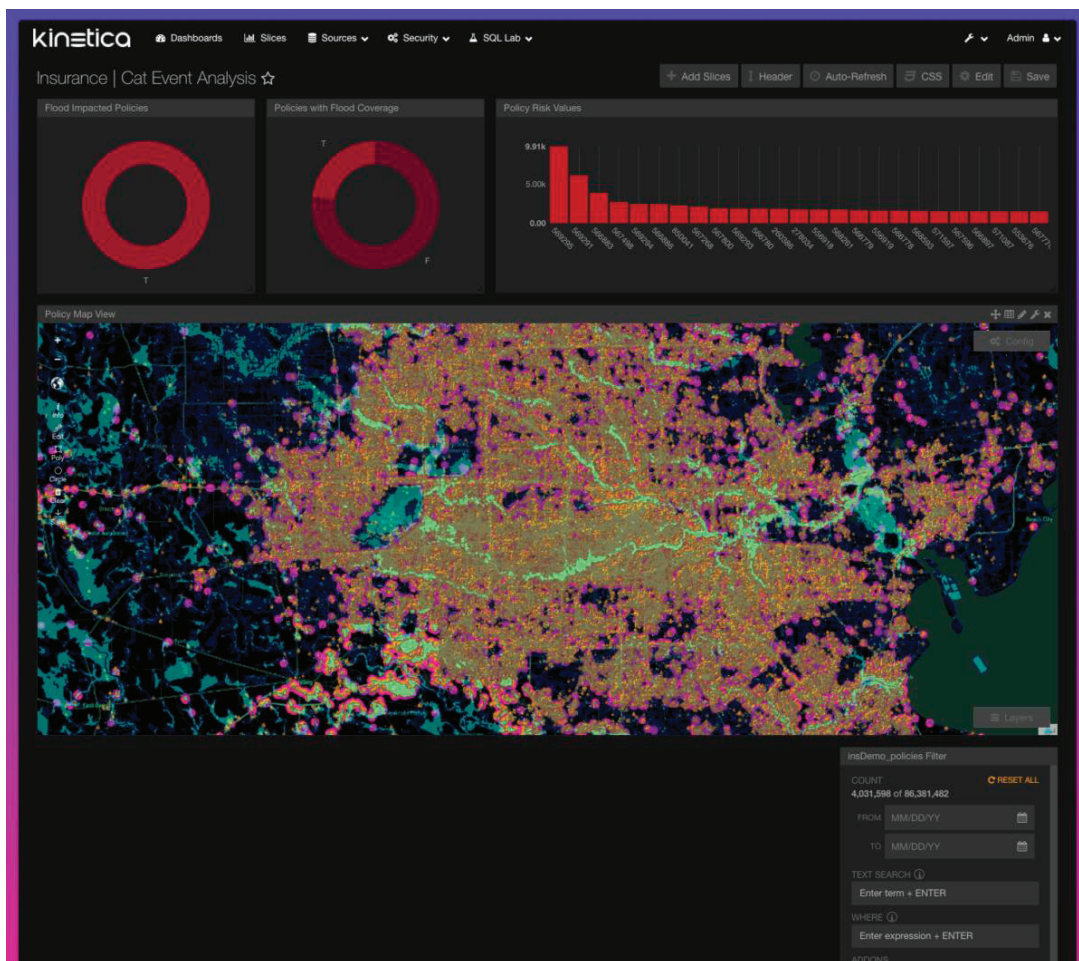
Insurance risk for catastrophic floods in Texas

The goal of this application is to assess an insurance company’s risk exposure to catastrophic floods in Texas from a table of policy holders and the Hurricane Harvey flood zones. The application does heavy geospatial computations in SQL along with statistical computations.



This dashboard shows the aggregate risks for catastrophic floods in Texas broken out by county, for flood insurance purposes.

IDG

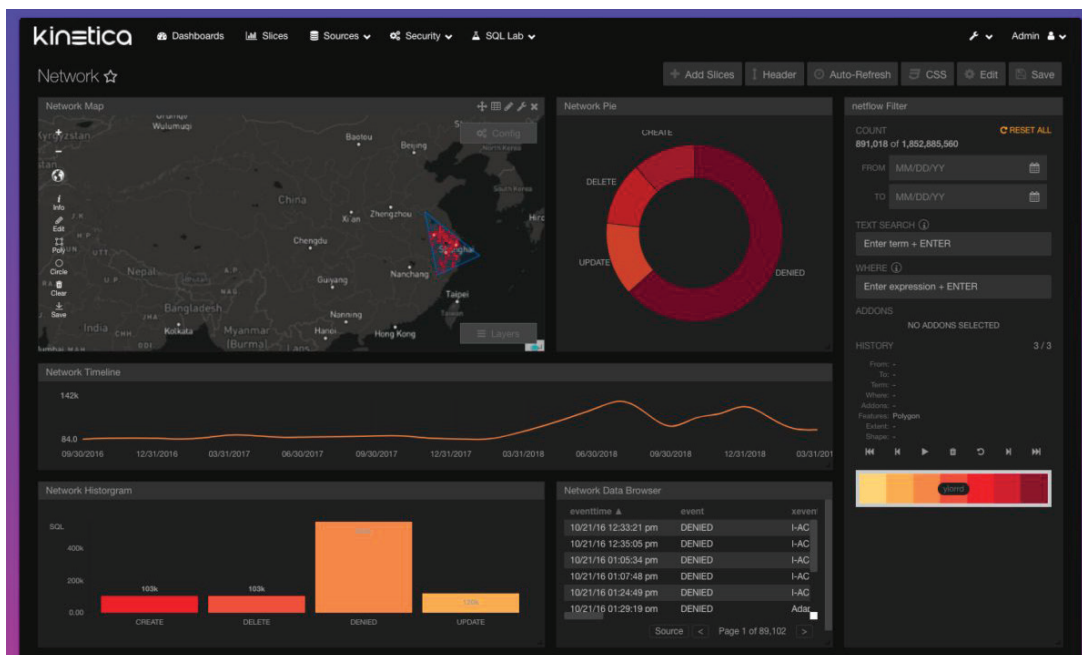


IDG

This dashboard shows the analysis of the insurance impact of a catastrophic hurricane in the Houston area. I have zoomed in and restricted the display to policies impacted by floods.

Network security assessment

This application is designed to help a network security officer protect a network from intrusions. The underlying Kinetica table combines about 1.8 billion historical network requests with a real-time feed.

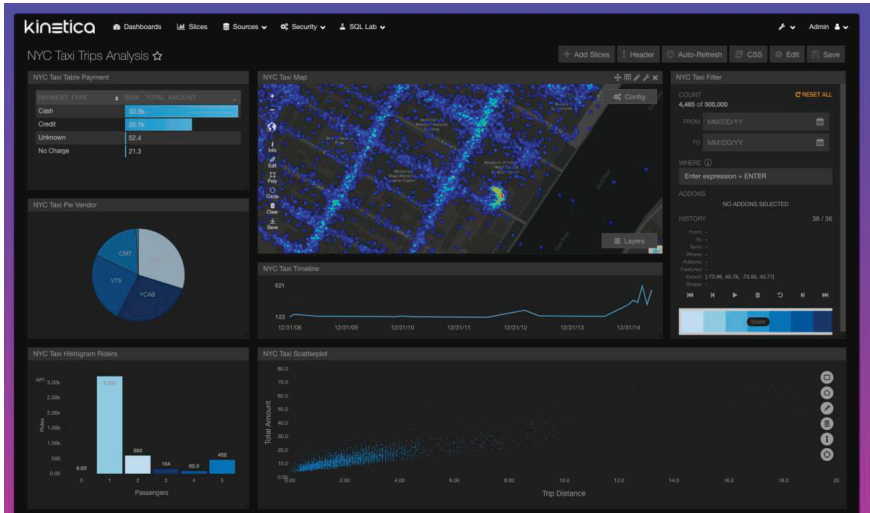


IDG

The dashboard above shows network metadata zoomed in on the Shanghai area, which originated a high percentage of packets that were denied, meaning that the firewall considered them malicious. Worldwide, denied packets were a very small percentage of the total 1.8 billion records.

NYC taxi rides

The New York City taxi ride database is something I also looked at in OmniSci. Kinetica provides it as a data set that you can load; that took about a minute. Initially it took longer to update all of the charts after each map zoom operation in Kinetica than I remembered from OmniSci; then I changed a setting so that Kinetica would not plot data outside of the zoomed map

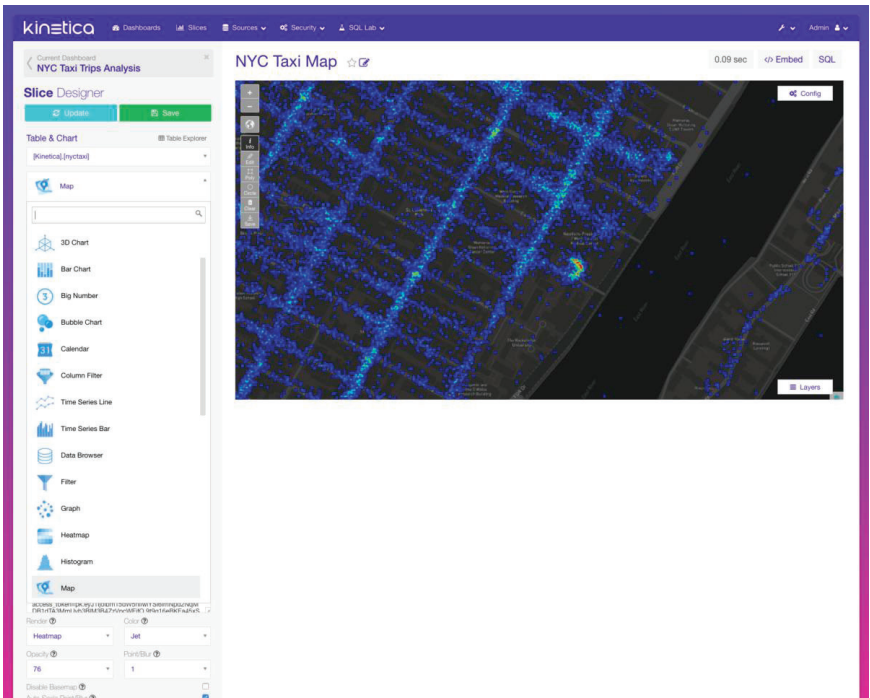


The dashboard above shows about 4,500 of the 500,000 NYC taxi trips in the data set from 2009 through 2014, zoomed into an area around Presbyterian Hospital on the upper east side. You can clearly see the high density of drop-offs at the hospital's curved drive.

on the other graphs, and the response time dropped to the sub-second range.

Kinetica slices and dashboards

Individual graphics in Kinetica Reveal are called slices. The slices are organized into dashboards.



Kinetica's slice designer can be used to create a wide variety of charts, maps, and tables.

The slice designer is quite similar to the designers you'll find in OmniSci and a number of BI products, such as Tableau. I didn't test the graph analysis portion of Kinetica, but I like the way it's designed. If graph databases are only a small part of what you need to do with your data, then reusing stored rows from the relational tables as edges and nodes makes perfect sense. Using GPUs to accelerate graph algorithms also makes perfect sense.

Seeing how Kinetica integrates machine learning with its GPU database, real-time analysis, and geographical information makes me understand where OmniSci wants to go—but Kinetica is already there. Also, seeing how Kinetica manages its storage tiers makes me understand why Kinetica usually competes with big data and data warehouse systems.

Overall, Kinetica is very impressive. It does what it claims, leaping tall databases with a single... I mean, analyzing databases with billions of historical rows and live feeds in real time. I wish I had a feel for the subscription cost, but that's proprietary, as is often the case with systems of this scale.

Cost: Kinetica charges an annual subscription based on the number of in-memory terabytes; it does not charge for data storage in other tiers. The subscription license enables you to run Kinetica anywhere—on-premises or in the cloud. Subscription costs are completely predictable. 30-day free trial available.

Platform: RHEL, CentOS, Ubuntu, Suse, or Debian Linux server with at least eight CPU cores and 8 GB of RAM; Nvidia K40 or above GPUs; on-premises, in the cloud, or at the edge on a Jetson TX2 embedded device. Kinetica also runs on Docker, with or without GPUs.

Martin Heller is a contributing editor and reviewer for InfoWorld. Formerly a web and Windows programming consultant, he developed databases, software, and websites from 1986 to 2010. More recently, he has served as VP of technology and education at Alpha Software and chairman and CEO at Tubifi.

AT A GLANCE

Kinetica 7

[Learn More](#)

on Kinetica

Kinetica can analyze databases with billions of historical rows and live feeds in real time, and integrates graph analysis, location intelligence, and machine learning with standard SQL.

Pros

- Analyzes billions of rows in real time
- Performs location intelligence and visualization
- Integrates deep learning and machine learning, including training TensorFlow models in-place
- Can deploy on-premises or in the cloud

Cons

- Requires Nvidia GPUs for speed
- Pricing is proprietary, but annual subscription is based on RAM