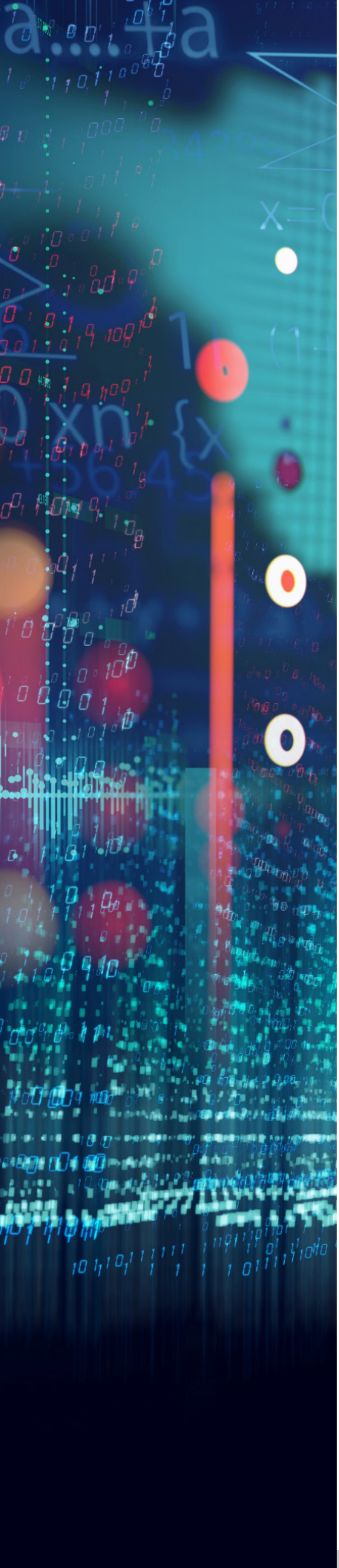
Power to the People

Addressing the Data Challenges of the Public Sector with Active Analytics

WHITE PAPER

early adopter research

kinetica



Contents

early adopter research

Introduction

In the US, the Federal Data Strategy from the Office of Management and Budget (OMB) seeks to move all federal agencies toward a single goal: "To harness data across the government to solve meaningful citizen challenges."

In the Fourth Industrial Revolution, the most valuable asset is data.

At the World Economic Forum in Davos in 2019, German Chancellor Angela Merkel stated that "data will be the raw material of the 21st century."

But refining data is not without challenges. According to Zvika Krieger, Head of Technology Policy and Partnerships at the World Economic Forum Centre for the Fourth Industrial Revolution, "The volume, speed, and complexity of data generated by the Fourth Industrial Revolution creates both challenges and opportunities. Computing power is constantly increasing, available data is rapidly expanding, and the number of connected devices is proliferating."

And with 5G on the horizon, the data available to government agencies will increase dramatically. Think billions of rows of data—streaming data, historical data, geospatial data, and more.

Government agencies are especially data-rich. Not only do most have decades of historical data, but they also have the broadest customer base possible: Every citizen is in effect a customer, a scale no business can match.

To capitalize on all of your data, you have to continuously capture it, analyze it, and instantly take action.

The challenge is not a lack of data. The challenge is whether government agencies can make use of all of it. The public sector has far more geolocation data than the private sector does. Being able to combine geolocation and geospatial data with copious public data sources, much of which is historical, offers near-limitless opportunities for government agencies.

In the US, the Federal Data Strategy from the Office of Management and Budget (OMB) seeks to move all federal agencies toward a single goal: "To harness data across the government to solve meaningful citizen challenges."[1]

1. https://www.nextgov.com/analytics-data/2019/06/16-things-omb-wants-agencies-accomplish-data-strategys-first-year/157481

Active analytics takes on billions of streaming and historical data points, visualized in three-dimensional space, orchestrated and analyzed in service of your mission.

Static analytics, downsampled datasets, and even dashboards are no longer enough. Legacy architectures, passive analytics (like reports and traditional BI), talent shortages, and limited budgets have not been conducive to public sector innovation. It's time for that to change.

Active analytics takes on billions of streaming and historical data points, visualized in three-dimensional space, orchestrated and analyzed in service of your mission.

The importance of this approach to analytics is paramount given the challenges that government agencies face today. For example, when asked about lessons learned in the wake of Hurricane Maria's devastation, Puerto Rico's CIO, Luis Arocho, offered the National Association of State CIOs (NASCIO) this perspective: "Put your critical infrastructure in a geospatial database"[2]—a key component of active analytics.

But as vital as disaster preparedness is, CIOs across government departments should be turning to active analytics because it has the power to revolutionize every type of government service delivery. Smart analytical applications will transform the way citizens interact with governments.

What could you achieve if all the data available to your agency, however siloed and incompatible in its current state, could be brought together and analyzed in real time? Think current systems (on premises and in the cloud) and publicly available data sources (from real-time feeds or API calls): What if you could bring all that data into multidimensional spatial views, ask questions of all of it, and get answers on the fly? Given the barriers and delays around data in many organizations across both the public and private sectors, it sounds too good to be true. But this is the promise of active analytics.

We live in a time when governments use drones to better understand environmental issues, cities install infrastructure in preparation for autonomous vehicles, and governments use machine learning to uncover trends to help guide planning initiatives. How do we prepare for this new world? And how do we overcome the roadblocks that stand in our way?

2. https://www.nextgov.com/analytics-data/2018/04/tech-lessons-learned-devastation-hurricane-maria/147724/

In many areas, there's a marked shift toward government agencies both using shared services and offering one-stop-shop websites to replace the litany of sites citizens and government employees alike must navigate for identification cards, building permits, health services, property tax bills, and more.

The Data Challenges for the Public Sector

Agencies face many challenges in turning data into decision-ready information. Some of these challenges are unique to the public sector.

Massive Data Scale and Scope

Foremost among those challenges is the scale of the data. Governments serve every citizen. The "customer base" is the entire population, and you can't choose who you do business with. Government is bound to serve all its citizens, each of whom generates more data each day, ranging from census facts to utility information, tax details to social welfare data. And that's just citizens! Every government department also works with troves of data generated through the research that serves their particular mission. The scale and scope is well beyond that of the typical business use case.

In the UK for instance, government agencies must integrate geospatial, historical, and real-time data, along with machine learning workloads to perform predictive analytics. As a prime example, every home in the UK can get a free energy monitor. Data includes real-time energy usage data (streaming data), energy prices across different suppliers (static data), renewable energy availability (geospatial boundaries), and supplier cost over time (machine learning predictive analytics models), among others.

Fragmentation and Silos

Government data is often siloed across individual departments. For instance, in many cities, permitting data is separate from utilities information, which is separate from transportation data. The public sector needs tools to break down these silos to operate more effectively.

And that wealth of location data mentioned earlier? Too often it is siloed in GIS systems, serving important but limited use cases.

In many areas, there's a marked shift toward government agencies both using shared services and offering one-stop-shop websites to replace the litany of sites citizens and government employees alike must navigate for identification cards, building permits, health services, property tax bills, and more. Using a centralized data platform ensures that these initiatives, as well as future efforts, will not suffer from the fragmentation and silos that citizens (and employees) find so frustrating.

Increased use of AI and machine learning makes transparency even more important in the face of questions about potential programmed biases, AI ethics, and the rationale for decisions that we outsource in part to machines.

Resource Constraints

Government agencies face continual cost pressure, tasked with providing more for less. Agencies don't have the luxury of a data center refresh or hiring more technical staff. They must use the data they have, the systems they have, and the employees they have to fulfill their mission. Any new technology must leverage existing systems and skillsets. Rip and replace is not a viable strategy, and technology must be adopted carefully to ensure value from existing data and platforms is retained. A wealth of historical data must be leveraged, not left behind, which leads us to...

Accountability

Because government agencies are publicly funded, they are required to provide transparency into resource usage and decision-making. In addition to public accountability, many agency heads require decisions, actions, and allocations to be trackable and traceable, as well. Increased use of AI and machine learning makes transparency even more important in the face of questions about potential programmed biases, AI ethics, and the rationale for decisions that we outsource in part to machines.

While the challenges are myriad, the abundance of data holds promise. Data has the power to transform the way government serves its citizens and to shape the society of the future. That is, if we can analyze it.

Use Cases Across the Public Sector Landscape

To get a glimpse of how transformative active analytics could be for the public sector, let's look at a few use cases.

Smart City Initiatives

Worldwide, many cities are adopting technologies that will help them operate more efficiently and effectively. For "smart city" policies to thrive, however, data—from energy to transit to communication to healthcare—has to be collected and centralized, rather than siloed and distributed as it is now. This is necessary to drive smart-city planning and the creation of new digital government services. For instance, if a city can combine transportation and development data, planners can make faster,

Being able to operationalize AI and machine learning in real time enables the public sector to take all the information generated by sensors, cameras, satellites, and other connected devices to surface anomalies and address them.

better-informed improvements to public transportation, traffic flows, and public safety. Cities will then need to be able to respond to streaming data in real time in order to operate these smart-city initiatives: cars, bikes, buses, and pedestrians, for example, all interact, and must receive the right information in real time to avoid traffic jams as well as accidents.

Fraud Detection

The public sector faces many of the same challenges with fraud as private businesses do. For example, proper tax collection is vital to ensuring government as a whole has the financial resources it needs to serve the public. The right active analytics data tools can improve the ability of the public sector to audit and hold taxpayers accountable, as they can provide transparent models to identify anomalous behavior and systemic patterns of tax evasion or other types of fraud (such as providers overcharging public healthcare systems).

As of mid-2019, governments around the world have gleaned more than $1.28 billion in fines and back taxes as a result of the Panama Papers investigation's exposure of the offshore finance industry. With 11.5 million documents comprising 2.6 TB, the Panama Papers illustrate the scale of external data that can be used to fuel active analytics.[3]

Intelligence and Cybersecurity

The right analytics architecture vastly improves the ability of the public sector to fend off attacks, both online and in the real world. By applying machine learning and AI, analytic tools wade through the mountains of data governments have available and find needle-in-the-haystack threats. Experienced analysts then decide how to respond. Being able to operationalize AI and machine learning in real time enables the public sector to take all the information generated by sensors, cameras, satellites, and other connected devices to surface anomalies and address them.

Many government agencies are also facing the retirement of knowledge workers. The loss of these workers is especially critical when it comes to intelligence and cybersecurity. Critical domain knowledge must be captured and codified in ML models that fuel new intelligence products that fill in that gap.

3. https://www.icij.org/investigations/panama-papers/panama-papers-helps-recover-more-than-1-2-billion-around-the-world/

Predictive analytics can identify weak spots in a city or country's infrastructure so that it can be repaired or rebuilt. If a natural disaster is imminent, relevant parties can foresee and prepare to address the problems that will likely result.

This may be possible:

With sensors everywhere, a proliferation of satellite data, and an explosion of IoT devices, those tasked with public safety can establish what amounts to a "persistent stare." ML can sift through all that video, audio, and sensor data and uncover important anomalies for humans to take a look at.

Millions of machine eyes never get tired of analyzing data while knowledge workers who have traditionally monitored those images and videos get tired, miss patterns they are not looking for,[4] and retire—taking their domain knowledge with them. Surfacing anomalies via machine learning uses computers for what they are best at (examining thousands of patterns) and offers human experts more interesting work investigating the cause of those anomalies. The learning from these expert investigations can then be captured and fed back into the machine learning algorithms, capturing human knowledge to separate potentially important events (drone activity) from visually distinct but more mundane events (trees with colored lights near the end of the year).

Predictive Maintenance and Logistics

The public sector can use analytics to fuel predictive maintenance on the wide range of assets under its purview—from buses and trains to snowplows and street sweepers, air conditioners and computers to security systems and electrical equipment in government buildings, recreation facilities, and schools.

Predictive maintenance of this kind is incredibly data-intensive, and especially dependent on geolocation and geospatial data. For example, a highway construction project should take into account not just the road, but the streetlamps, underground pipelines, electric cables, bus routes, and phone lines impacted by the work. Predictive analytics can identify weak spots in a city or country's infrastructure so that it can be repaired or rebuilt. If a natural disaster is imminent, relevant parties can foresee and prepare to address the problems that will likely result.

Abundant data is available to fuel digital transformation and use cases across defense, public health, urban planning, and other sectors. But in order to make use of that data, government agencies cannot rely on their legacy infrastructure or attempt to augment it piecemeal with open source software.

4. The phenomenon of overlooking elements hidden in plain sight is illustrated in attentional research related to the Invisible Gorilla study, in which radiologists were shown images that include a gorilla. More than 80% of radiologists missed the gorilla.

The Kinetica Active Analytics Platform tackles the inherent challenges of analyzing public sector data at scale, combining location intelligence, streaming and historical data analytics, and AI and machine learning on a cloud-ready GPU-accelerated in-memory database.

Many government agencies have made significant investments in Hadoop and other architectures that don't offer the capabilities or speed to operationalize today's data. The public sector must find tools that significantly accelerate insights from legacy systems. The key is to find ways to unlock the power of data, continuously combining and analyzing the latest data along with historical data in order to respond to today's needs and plan well for future initiatives.

Power to the People: The Kinetica Active Analytics Platform

Kinetica offers an analytics architecture that is ideal for the public sector to fully harness the power of all the data at its disposal, both now and in the future.

In a single solution, Kinetica Active Analytics Platform tackles the inherent challenges of analyzing public sector data at scale, combining location intelligence, streaming and historical data analytics, and AI and machine learning on a cloud-ready GPU-accelerated in-memory database.

Let's examine how this unique combination of capabilities works together.

Location Intelligence

It is generally estimated that 80% of data has a spatial component.[5] In the public sector, geospatial data and the ability to visualize it, blend and enrich it, and analyze it is mission-critical. Kinetica is designed from the ground up to deliver interactive geospatial analysis at unprecedented scale, combining streaming and historical location-enabled data on demand and empowering developers and business analysts to build smart applications that effectively leverage geospatial data.

Too often, GIS platforms have been siloed, relegating spatial analysis to specialists within an agency, and their relatively narrow but certainly important use cases. The Kinetica Active Analytics Platform makes it possible for everyone to use location intelligence, democratizing geolocation data.

5. The origins of this widely cited figure are difficult to identify, but experts stand behind it, putting the percentage even higher. Dr. Tim Foresman, former chief environmental scientist for the UN, stated on StackExchange, "I used to quote the 80-90% spatial data wisdom to my classes at UMBC in the early 1990s. Then I changed pedagogy and challenged the students to identify non-spatial data. Not easy."

Decision-ready intelligence must be informed by real-time streaming data (what's happening now) in the context of historical data analysis (what we have seen over time).

To make it easy to incorporate location intelligence into applications and visualizations, Kinetica offers API access to Web Map Service (WMS), an open standard that allows geolocation data to be fed into leading mapping providers and projects.

With Kinetica, location visualization and awareness are part of the platform, not an additional development effort. Rendering geospatial visualizations leverages the same GPUs that process billions of rows of data.

Streaming and Historical Analytics

Decision-ready intelligence must be informed by real-time streaming data (what's happening now) in the context of historical data analysis (what we have seen over time).

Kinetica analyzes billions of rows of historical data along with live feeds in real time, providing important context. Complex open source solutions like Spark are labor and processor-intensive.

While specialized and expensive engineering resources are required to deploy the so-called SMACK (Spark, Mesos, Akka, Cassandra, and Kafka) stack, Kinetica leverages existing skill sets like SQL and can massively simplify the architecture and reduce the time to get return on investment (ROI).

AI and Machine Learning

The Fourth Industrial Revolution is fueled by AI and ML. Data-rich government agencies and open data sources hold tremendous potential for machine learning, but the challenges of finding experts and auditing have left many AI initiatives stalled in the pilot phase.

The Kinetica Active Analytics Platform makes it easier to operationalize AI and integrate it into intelligent applications. Kinetica offers the traceability needed to track, govern, and audit analytics and ML workloads. When questions arise about why a particular action was taken, agencies need immediate answers that offer transparency into what data was used and why the AI recommended that action.

No matter what your agency's mission is, deep learning on troves of historical data reveals patterns and anomalies that warrant further investigation. Given the scarcity of AI professionals, the ability to easily uncover patterns in data and experiment with the latest TensorFlow modules will accelerate adoption of AI.

Kinetica automatically distributes workloads across CPUs and GPUs for optimal results and uses industry-standard SQL to process and analyze billions of rows of data in a matter of milliseconds.

Cloud-Ready

Modernization and agility are baseline requirements for government agencies facing increasing pressure to deliver cloud-based services.[6] With its cloud-ready architecture, Kinetica offers ease and flexibility of deployment on premises or in the cloud. Cloud deployments offer the high availability, elasticity, and scalability required for mission-critical applications, from defense to constituent engagement. High availability can be configured in minutes, and incremental backup and restore offers instantaneous recovery from any unexpected changes. As demand changes, deployment is easily scaled up or down at the push of a button or via an API call.

GPU-Accelerated Database

The Kinetica Active Analytics Platform is designed for the scale and scope of public sector data today and in the future. The platform is built on a distributed, in-memory, GPU-accelerated database that utilizes a powerful combination of CPUs and GPUs to analyze massive, complex datasets with millisecond response times.

Kinetica automatically distributes workloads across CPUs and GPUs for optimal results and uses industry-standard SQL to process and analyze billions of rows of data in a matter of milliseconds.

Designed for enterprise scale, Kinetica can operate on the entire data corpus by intelligently managing data across GPU memory, system memory, disk/SSD, HDFS, and cloud storage like S3 for optimal performance.

Smart Applications

An active analytics solution empowers agencies to build smart applications that offer compelling value and user experiences for stakeholders and citizens alike.

The power of an integrated platform is immediately evident to developers, who can create intelligent applications that draw on the Kinetica Active Analytics Platform without the need to string together multiple

technologies. APIs, REST calls, and SQL enable developers to focus on the user experience and complete their analysis in Kinetica without having to shuffle data to other platforms.

6. In the US, for example, the Federal Cloud Computing Strategy shifted from Cloud First to Cloud Smart, mandating specific actions and timelines.

Kinetica has the prime advantage that it can perform the high-impact, demanding analytics work needed by the public sector in a single, affordable, high-performance package.

Kinetica also offers APIs and technical integrations with a wide variety of data tools and platforms that make it easy for developers to build smart analytical applications.

Conclusion

Kinetica has the prime advantage that it can perform the high-impact, demanding analytics work needed by the public sector in a single, affordable, high-performance package.

From its inception, Kinetica was built to meet the needs of the public sector, from delivering high-resolution geospatial analytics, to consuming real-time streaming data, to offering high-speed analytics that can be driven by machine learning and AI. The alternative is to cobble together a hodgepodge of different products, leading to greater costs and complexity.

Your mission today may not yet require every facet of the Kinetica platform, but with the advent of 5G, the explosion of IoT devices, more real-time data feeds, and the increasing importance of location awareness and analytics, the Kinetica Active Analytics Platform offers an architecture for your current needs and a vision for a future that will be here before you know it.

How can you solve your toughest problems? Ask yourself:

Digital Transformation Questions

- What if you could visualize and analyze geospatial data? What if your apps were location-aware, like the best consumer apps?
- What if you could analyze all of your historical data along with streaming data, rather than downsampling?
- What if you could replace static reports with intelligent analytical apps? What if those apps were tailored to meet particular needs?
- What if machine learning and AI surfaced anomalies for people to investigate? What if professionals enjoyed the challenge of solving real problems instead of wasting time on false positives and repetitive work?

Kinetica was built to meet the needs of the public sector, from delivering high-resolution geospatial analytics, to consuming real-time streaming data, to offering high-speed analytics that can be driven by machine learning and AI.

- What if you could easily migrate applications from on premises to the cloud to meet demand? What if you could scale up your services on demand?
- How could a geodatabase of critical infrastructure help you?

Use Case Questions

- What if you could coordinate an evacuation without clogging the highways?
- What if you could close, replace, or repair that dam or electrical grid proactively, before a problem arises?
- How could you use real-time feeds on the spread of disease to bring supplies where they are needed most?
- What if you could safely adopt autonomous vehicles ahead of other municipalities thanks to the ability to apply machine learning to streaming data, historical data, and location data?
- What challenges could you solve?

This paper was written by Early Adopter Research and sponsored by Kinetica

Learn more about the Kinetica Active Analytics Platform or request a live demo.

Connect with us

early adopter research