How to Manage the Crunch of Real-Time Data-Intensive Workloads

MARKET TRENDS REPORT

govloop   ⊲EROSPIKE

Executive Summary

The challenge federal agencies face with data isn't just one of constantly exploding volume. It's also one of velocity — the ability to ingest, analyze and draw value from that data in a relative blink of the eye. Too often, however, they're working with systems and tools designed for a smaller scale and more deliberate pace of operations. To effectively work with data in real time, they need a platform built to handle current and future data demands.

As in business and other sectors, data is the lifeblood of government agencies, whether they're dealing with threat assessments, troop deployments, disaster relief applications or park visitation permits. The Federal Data Strategy mandates that agencies leverage the full value of their data in aiding decision-making, while also ensuring security, privacy and confidentiality. If that weren't a tall enough order, they're also dealing with increasingly massive amounts of data — including data from the Internet of Things (IoT), mobile users and applications — in addition to the imperative to process that data in real time.

One way to meet that challenge is with a small-footprint, real-time data platform that can give them centralized control of their data within the limits of their budgets.

To learn more about the importance of managing data in real time, GovLoop partnered with Aerospike, which provides next-gen real-time NoSQL data solutions. In this report, we'll examine the particular challenges agencies face, the steps they can follow to gain control of their data and the benefits of a real-time platform.

By the Numbers

150 zettabytes

or 150 trillion gigabytes, of real-time data will need analysis by 2025.

73 zettabytes

will be the amount of data generated by IoT devices by 2025.

5 billion

networked devices/connections will exist in North America by 2023, up from 3 billion in 2018, along with 329 million mobile users, representing 88% of the regional population.

75%

of federal chief data officers (CDOs) say they have formed a diverse data governance body in line with the mandates of the Federal Data Strategy 2020 Action Plan.

56%

of survey respondents identified limited data access/sharing as an obstacle to using data to support their agency's mission, the most common obstacle cited. Limited Staff Skills or Workforce Hiring Challenges was next, at 51%.

80%

of public-sector leaders say implementing intelligent technologies would improve job satisfaction for current employees.

95%

of organizations say unstructured data is a problem.
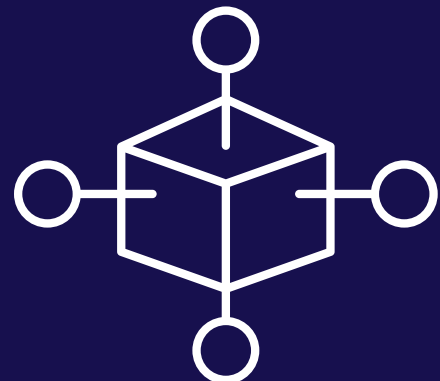
Data-Driven Decisions at the Speed of Events

Challenge: Too Much Data in Too Many Places

Agencies looking to access and analyze the data they collect face several built-in hurdles that block the path to meeting federal mandates for turning that data into actionable information while complying with security and privacy requirements.

Disparate technologies and architectures. The mounting volume of data is coming in multiple formats from a growing number of sources, including cloud applications and IoT sensors. Agencies' data stores often are siloed, incompatible and lack the automation required for easy retrieval.

Legacy systems were never designed with this kind of data in mind, said Srini Srinivasan, Chief Product Officer and Founder of Aerospike. "This nonlinearity is not something these products were designed to handle," Srinivasan said.

Proprietary systems and solutions. Legacy systems aren't equipped for real-time data sharing because of incompatible, proprietary formats and an inability to accommodate advanced technologies such as artificial intelligence (AI). This works against the goal of using data in a timely fashion. "Many of these decisions have to be made in some level of real-time service-level agreement," Srinivasan said. "There is typically somebody who is waiting for a response."

Organizational silos. Agencies have different structures and missions, which can slow or prevent information sharing, even with the latest technology tools.

Dispersed geographic locations. Agencies' reach, whether across the country or around the globe, throws other hurdles in the way. For example, emergency response or military missions could be in remote areas. And mobility further complicates the decision-making process since it may depend on the location of a user — or multiple users — on the move.

Solution: Tools for Putting Data into Action

With the right platform, agencies can ingest, analyze and make use of data in real time, and do it at any scale and with constant uptime. Among the tools that can make this possible:

- A flash-optimized storage layer, treating flash as memory, which provides accelerated input/output operations, higher performance and small-footprint scalability.

- Storage indices in dynamic random-access memory, with data on optimized solid-state drives, which provides predictable performance at any scale.

- A multithreaded, massively parallel design for scaling up or out, adding that capacity while using as few nodes as possible.

- Self-healing clusters, providing a single hop to data along with superior uptime, availability and reliability.

These elements can create a hybrid architecture that can produce five key outcomes:

Ingesting and analyzing big data. A distributed system that incorporates flash storage with the latest hardware technology, the cloud and extreme-scale applications can bring immediate access to data, no matter the source or location.

Centralizing systems in real time. A platform that combines systems and scales on the fly can put control of the data into a central panel. It also can incorporate all of an agency's historical data, putting everything at your fingertips.

Integrating modern applications with legacy systems. A flexible system that can incorporate input from different software utilities, such as Hadoop or Spark, can also make use of machine learning (ML) and AI to, for example, track a cyberthreat in real time.

Delivering new services in near-real time. A real-time platform can deliver new services highly efficiently. For example, it could allow IT teams to implement in an hour a new service that otherwise would take a day to deploy.

Lower total cost of ownership (TCO). An edge-to-core platform that delivers five-nines uptime and predictable high performance at any scale while maintaining a small network footprint can also substantially lower the TCO, enabling the lowest possible spending on capital and operational expenditures.

What to Look for in a Real-Time Data Platform

Affordability. The stakes are high for agencies in meeting the mandates for using their data. But the cost doesn't have to be. It is now possible to affordably inject a real-time component into agency architectures without having to re-platform. A real-time platform can reduce the nodes required for an enterprise-scale system while reducing TCO.
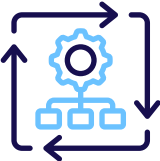
A premium on performance. Agencies need a real-time data foundation that matches extreme-scale applications with consistently high performance and availability, both now and in the future, as data demands grow and new technologies arrive. Among the features they should look for:

- AI and ML
- Personalized public services
- Edge computing at scale

- Operational and financial compliance
- High-performance computing
- 5G architecture

Reliability at scale. Any system must be able to scale up, out and down when required, but it needs to retain uptime no matter how large the architecture gets. A provider's infrastructure must keep running regardless of the demand.
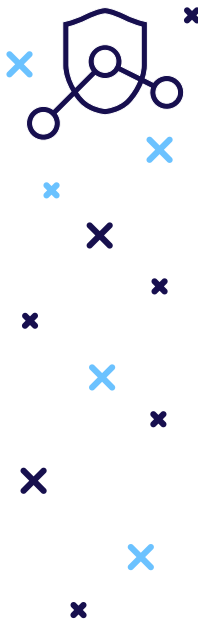
Security and privacy. Security is at the heart of everything government agencies do. In the current cloud and mobile computing environment, a platform that can combine data from across the enterprise in real time is essential. "We cannot have these systems isolated anymore, because everybody's on a mobile device," Srinivasan said. An attack can move from location to location, and threat assessments must be able to move with them.

Agencies should look for a platform that adheres to the most effective methods of enterprise security and privacy, including:

- **Authentication** performed at local and Lightweight Directory Access Protocol, and Kerberos and access control list levels

- **Authorization** that supports policy by adhering to predefined roles and privileges

- **Encryption** of data in motion and at rest, whether in a client-server setup or in clusters of nodes

- **An audit trail** that is both granular and configurable

- **Granular whitelisting** at the domain and user levels to support strong security

- **Supporting public-key infrastructure** by tightly controlling access to tokens, passwords, certificates and encryption keys for protecting highly sensitive and secret data

- **Use of advanced technologies,** such as AI and ML, high-performance computing, and a 5G architecture to enable real-time data collection and analysis

Open to innovation. A benefit of a real-time data platform is the opportunities it presents. It can give agency IT pros ideas about how to do things that they hadn't thought of before, Srinivasan said.

Use Case:
Staying Ahead of
Cyber Threats

Implementing a real-time data platform can allow an agency to use its data effectively in ways it can't now, whether it involves processing requests or dealing with more urgent matters.

Threat assessments are one example. In building a threat or risk assessment, security teams generate a model based on the current threat landscape (ransomware, phishing tactics and so on), the assets and data the agency has, and incidents in the cyber world in, say, the past 30 days. They'll then match actions as they come in against the model. A real-time platform often processes the requests in less than 30 milliseconds, which can effectively cut off the threat before it can do any damage.

But the threat landscape isn't always so simple, of course. Attackers refine their methods, finding new ways to get in and stay hidden. A 30-day timeframe for building threat models isn't feasible. Realistically, a security team should ingest data, generate new models and improve existing models in real time. Threat assessments must be able to evolve as quickly as the attack cycles.

A real-time data platform using technologies such as AI and ML can close that loop between building a threat assessment and applying it to specific requests as they happen. And that ability will be critical to keeping systems and data safe as attacks get increasingly more sophisticated.

HOW AEROSPIKE HELPS

Federal agencies and major corporations — including real-time businesses such as PayPal, Wayfair and Yahoo — rely on Aerospike and its patented Hybrid Memory Architecture™ as the foundation for current and future management of their data. The company's global infrastructure, its embrace of the latest technologies and its distributed, real-time data architecture deliver a reliable, secure, high-performing data platform that can help agencies accelerate IT modernization goals and realize their enterprise data strategies — and produce savings in the process.

"We're not replacing apples with apples," Srinivasan said. Aerospike can increase the number of transactions per second a thousandfold while costing considerably less than it would with traditional systems. "That's the trick," he said. "It is the growth of this data load in real time that basically transforms the enterprises, including agencies."

Its comprehensive real-time data management and highly secure approach also can help agencies efficiently meet their security and compliance mandates.

Learn more: aerospike.com / federal@ aerospike.com

Conclusion

The term "data-intensive" no longer applies only to certain scientific and military agencies that handle millions of transactions a day. Practically every agency in government is becoming data-intensive.

But the growing influx of data, and legacy systems' inability to handle neither the volume of data nor the speed with which it needs to be turned into useful information, threatens to leave agencies mired in their own data, rather than using it to their best advantage.

The answer is in a real-time data platform that can provide a secure, next-generation, highly scalable edge-to-core data foundation that merges data from across the enterprise and uses advanced technologies to process and analyze it in real time. It gives agencies the ability to improve performance through better decision-making.

ABOUT AEROSPIKE

The Aerospike Real-time Data Platform enables organizations to act instantly across billions of transactions while reducing server footprint by up to 80%. The Aerospike multi-cloud platform powers real-time applications with predictable sub-millisecond performance up to petabyte scale with five-nines uptime with globally distributed, strongly consistent data. Applications built on the Aerospike Real-time Data Platform fight fraud, provide recommendations that dramatically increase shopping cart size, enable global digital payments, and deliver hyper-personalized user experiences to tens of millions of customers. Customers such as Airtel, Experian, Nielsen, PayPal, Snap, Wayfair and Yahoo rely on Aerospike as their data foundation for the future. Headquartered in Mountain View, California, the company also has offices in London, Bangalore and Tel Aviv.

ABOUT GOVLOOP

GovLoop's mission is to "connect government to improve government." We aim to inspire public-sector professionals by serving as the knowledge network for government. GovLoop connects more than 300,000 members, fostering cross-government collaboration, solving common problems and advancing government careers. GovLoop is headquartered in Washington, D.C., with a team of dedicated professionals who share a commitment to connect and improve government.

For more information about this report, please reach out to info@govloop.com.