Data: The Foundation for Flexible Innovation in Government

Data Silos in Government

Data has become a strategic tool for public sector organizations, providing insight across every segment of government from federal agencies to K–12 school districts. By granting unprecedented access to actionable information both immediate and long term, data can:

- Generate a greater understanding of operations
- Improve citizen services and information delivery
- Identify more cost-effective ways to meet mission requirements

To begin accessing this information, it is necessary for agencies to effectively consolidate and operationalize their data. This includes discovering information repositories and embedding past-, present-, and future-looking analytics into their end users' workflow in order to impact the metrics that matter. However, if this data is fragmented across sources, then relevant data is left out of workers' analyses, latency occurs in applications, and organizations don't get the information they need to make informed decisions. This, in turn, reduces the value of data investments and creates inefficiencies at the operational level.

As growing volumes of data continue to be operationalized across the public sector, pressure is put on existing, often legacy, infrastructure, causing challenges to arise. These challenges have compelled organizations to rethink how they approach data management by revamping their architecture to support this new way of business. By modernizing their architecture, organizations can adopt new data storage and management architectures for big data designed to meet the challenging analytical needs of government today. These modern data platforms are capable of centralizing and managing massive data sets—integrating structured and unstructured data—to deliver advanced analytics and machine learning capabilities for diverse use cases simultaneously. By implementing this new architecture, these forward-looking organizations can lay the foundation needed to support the current and future needs of government.

One solution is to bring all data together in a modern platform, which provides greater visibility into information across all areas of the organization, not siloed views. This big data platform can be used for integrating disparate data from multiple sources so that analysis and reporting can be carried out even while government operations are occurring live. This is an important first step that public sector organizations should take as they look to manage the increasing volumes of data being generated from active sources that range from embedded sensors in roads, video images from police body cameras, student test results, and citizen requests for benefits.

Limitations and Challenges

Countless challenges arise across the public sector if a proper data platform is not in place, including disconnected data stores, limited data collection, data latency, and the need to frequently delete or archive data. This problem will only get worse given the growing volume of new incoming data being driven by everything from the Internet of Things to compliance mandates. But each of these challenges first must be examined and understood in order to find the right solution.

Limited Data Access — Governments and other organizations are demanding new data sources while maintaining longer retention to fuel daily decisions and mission success. Since unstructured data must be reformatted to fit into a relational schema before it can be loaded into the system, it requires an extra data engineering step that slows ingestion, creates latency, and can sometimes eliminate elements of the data that might become important down the road. Meanwhile, increasing data volumes are continuing to put strains on existing infrastructure.

Processing Inefficiencies — The data processing infrastructure that was implemented to run public sector operations over the past decades is having trouble keeping pace with today's digital landscape. Across every sector of education and government, daily interactions and transactions are moving online, and business services are becoming increasingly automated. Volumes of multi-structured, machine-generated data from a variety of sources have skyrocketed, and smart organizations want to capture and make use of it all. After all, a comprehensive view of the landscape can help agencies inform policy and improve the quality of citizen services. As data volumes grow and the complexity of data types and sources increases, data engineering workloads take longer to run, and the time available for reporting and analysis is reduced.

Security and Data Access — The increase in unstructured data being ingested can also translate to more sensitive data being stored—which in turn makes it more difficult to secure and govern in compliance with regulations and best practices.

Data Archived and Deleted — As strains are put on traditional systems, and IT departments try to meet SLAs, potentially valuable data is archived or even deleted in order to free up capacity for optimal performance. This data costs the organization resources to collect and process and analyze, but simply archiving it significantly reduces the returns. With this historic data made unavailable, it cannot be used even in key analytics that can be crucial—now or in the future—for mission-critical decisions.

These challenges are leading organizations across the public sector to modernize their architecture by creating a single place to store all data for processing, leveraging new data platforms that are far more efficient in meeting their new requirements. With the right architecture in place, public sector organizations not only gain short-term insight, but also unlock future data innovations that require efficient processing of massive data volumes.

New Foundation Fuels Change

Cloudera Enterprise is the fastest, easiest, and most secure data platform available for machine learning and advanced analytics. Built on the latest open source technologies, it unlocks value by allowing customers to capture, store, process, and analyze any data type at massive volumes—eliminating the need to archive data—while allowing for quick, familiar, and safe data access to end users and applications. These value-add capabilities free up resources by allowing key players to focus on using data to unlock important advances, instead of letting data storage costs, management overhead, and processing times limit innovation. In addition, it allows for better collaboration across departments by providing better access to data that was once siloed or deleted by organizations using an outdated architecture.

Visualize Data Across Your Organization — Users want quick access to new information of all kinds while using existing tools. By modernizing with Cloudera Enterprise, organizations can better ingest, process, and store any volume or type of data from multiple sources in full fidelity. Data discovery is streamlined, enabling valuable new use cases to be quickly defined.

Optimized Data Processing — ETL data engineering workloads that previously ran on storage systems can migrate to the Cloudera Enterprise Data Hub (EDH), where they run in parallel in order to process any volume of data at speed. Optimizing the placement of these workloads frees capacity on expensive data warehouses, allowing them to focus processing power on mission-critical OLAP, reporting, and other applications. Agencies that have leveraged an EDH have seen 50x increase in data processing capability while going months without ETL failures.

Data Longevity — An EDH offers one secure place to cost-effectively store all data, in any format, any volume, keeping it available for as long as needed. This allows organizations to naturally process and store data without having to worry about archiving it, while providing storage for replay when needed. Organizations can then deliver historic data on demand to satisfy internal and external analytic and mission data needs.

Automated Secure Archive — With various data sources and types, security is key for an EDH as more users access this information and more sensitive and classified data is stored. Only solutions with security built into the core can provide the security and governance needed for the public sector. Cloudera's EDH is the only solution with integrated encryption, key management, and unified authentication. Additionally, this integration allows for data governance—allowing for comprehensive auditing and data lineages to trace data over time.

Public Sector organizations that are leveraging a big data platform, such as Cloudera Enterprise, have already seen significant returns by modernizing to free up their existing data warehouse and analytic database systems and eliminating data silos. The result is an increase in the amount and type of data they collect, and the ability to store everything in active storage with no need to archive it. This has changed the way they view data processing by providing them the flexibility and scalability that traditional systems have struggled with, and prepared their organization to use advanced analytics on all data, not just a sample. They can then take advantage of everything from improved processing of sensor data to artificial intelligence and machine learning.

Where Improved Insight Can Lead

As government and education organizations modernize, vast opportunities present themselves that open new avenues of citizen service and drive change in how agencies can better detect waste, fraud, and abuse of scarce government resources. Public sector organizations are beginning to discover the power of advanced analytics to pinpoint ways in which programs can be improved and how actions they take can be measured in real time, which can enhance public trust and satisfaction. The world in which government operates is quickly shifting, and enterprise architects and IT professionals alike are sitting in the center of this change. When making system decisions, these professionals must not only take into account the immediate needs of their organizations, but also understand the future needs of the government. This paradigm shift requires a modern platform that is built for a world of large volumes of diverse data and new applications.

About Cloudera

Cloudera delivers the modern platform for machine learning and advanced analytics built on the latest open source technologies. The world's leading organizations trust Cloudera to help solve their most challenging business problems by efficiently capturing, storing, processing, and analyzing vast amounts of data. Learn more at cloudera.com.