

Power your mission transformation with an enterprise data cloud



What does it mean to be insights-driven?

Is data the new bacon? The world's most valuable resource? The fuel that powers modern government? Whatever metaphor you choose, the point holds that the most successful government organizations power their missions with compelling insights derived from data. Being insight-driven isn't just the domain of the private sector. The government relies on data to increase internal departmental efficiencies, improve the way it engages with citizens, and increase mission effectiveness across a vast array of programs and services.

While putting data and analytics at the center of your mission may sound obvious, it's easier said than done.

Digital transformation often requires a change in an organizational culture that includes rethinking how people work with data, where it's stored, and the tools used to access it. In fact, it can be such a struggle that many organizations wind up stalling or abandoning their transformation strategies altogether.

This white paper will explore some of the challenges associated with being data-driven and how a new type of data platform—an enterprise data cloud—can address those issues.

Why government struggles with data

The most valuable and transformative government use cases—the ones powered by unique insights that can only be derived from a good data strategy—require multiple analytics frameworks, data science tools, and machine learning algorithms leveraging a diverse set of common data. It's in these workloads that innovative government organizations are able to unlock immense value from their data and achieve their mission goals. It's easier said than done, however, because the reality is that, traditionally, the government has not operated this way. Instead:

- ✓ Analytic workloads run independently in silos. Even newer cloud data warehouses and data science tools aren't designed to work together.
- ✓ Data is everywhere—in data centers, public clouds, and at the edge—and government organizations have no practical way to run analytics or apply machine learning algorithms across their data.
- ✓ Siloed analytics and data everywhere make a coherent approach to data privacy or IP protection nearly impossible. At best it forces onerous controls that limit productivity and increase costs.
- ✓ Traditional and even modern cloud-based machine learning and analytics tools require the use of closed-source and proprietary storage and algorithms, meaning government organizations have to essentially hand over their most valuable asset - their data - to extract meaningful insights.

Nearly 50% of government organizations are actively using cloud services.

*Understanding Cloud Adoption in Government, Gartner
4/11/18*

6 in 10 federal IT executives see cloud computing as a vital pathway to improve mission-critical services.

*Federal IT leaders report advances in cloud adoption for critical services, Fedscope
12/5/18*

State government CIOs ranked cloud services #2 top priority for strategies, management, and process solutions.

*Hybrid Cloud Addresses the Unique Concerns of Government, Statetech Magazine
7/16/19*



Introduction to an Enterprise Data Cloud

Being data-driven requires the ability to apply multiple analytics disciplines against data anywhere. To be able to process and stream real-time data from multiple endpoints at the edge while predicting key outcomes and applying machine learning on that same data set. To be able to take advantage of public cloud infrastructure for its agility, elasticity, and, increasingly, data gravity. And to be able to do all of this on an open platform where data security and governance are applied wherever the data lives and analytics run. This is what the industry calls an enterprise data cloud. Key characteristics of an enterprise data cloud include support for:

Multi-function analytics

Government agencies must be able to rapidly identify useful data, ingest and transform it, and deploy the analytic frameworks that make sense for their unique mission objectives. Addressing real-world government challenges generally requires the application of multiple analytic functions working together on the same data. For example, preparing for and responding to incidents like cyberattacks or fires requires the application of real-time data streaming from a wide array of information sources, followed by the development and deployment of machine learning algorithms to stakeholders, like citizens, to deliver critical information and make intelligent decisions (AI).

Hybrid and multi-cloud

To provide the flexibility to deploy use cases in the environment that best suits an organization's needs, an enterprise data cloud must operate with equivalent functionality on and off-premises, supporting major public and private clouds. Data has mass. It's expensive to collect and difficult to move, so it makes sense to analyze data where it naturally lives. That requires bringing the right compute resources to the data, whether that's in a public cloud object store or in an organization's data center.

In most government organizations, mission-critical data can be found in multiple locations. For example, an agency may choose to retain sensitive records that fall under compliance scope on-premises, while at the same time running proof-of-concept applications using anonymized datasets in the public cloud. Hybrid and multi-cloud architectures provide greater control, putting the agency in charge of where, when, and what type of analytic workloads to run.

Another important aspect of separating compute and storage is that it now provides government organizations the ability to choose between cost and performance. For example, a critical analysis may require additional compute resources from the public cloud, but for only a brief period of time. In this case, an emphasis is placed on speed rather than cost. Other analytic workloads—routine dashboards for example—can be cost-optimized by running on-premises or in a private cloud.

Unified security, governance, and metadata

In many government environments, a data warehouse or machine learning tool will often have its own proprietary data catalog and policies that secure data and control access. In order to connect the two or more systems in service of a critical use case, the data team must extract the output from one system and reset the schema, security, and access policies, so that they are in compliance with all other systems. This takes time, introduces risk, and consumes resources that would be better applied elsewhere.

The challenge is further amplified when organizations adopt hybrid infrastructures. An enterprise data cloud delivers a common fabric for data and metadata that enables multiple analytic functions to work in concert and leverage data from across public clouds and on-premises environments. This includes unified security and governance applied to data wherever it lives to ensure the organization is meeting regulatory constraints.

Government organizations need to be able to secure data and enforce the policies for who can access it and under what circumstances. They need to be able to track that access over time to be sure that what people do is permitted and answer questions about how and when data was used for auditing purposes.

Having a variety of analytic frameworks working together across different infrastructure and storage systems requires a consistent shared governance and security model; a single way to express a policy—applied across infrastructure and without regard to the analytic system—on who can touch the data and what they can do with it. That's the only way to be sure that those policies are being uniformly applied.

Open platform

Government organizations want to avoid vendor lock-in, and own their data and the insights they unlock. They don't want their analytic decisions to be limited based on their operating environments. Open platform means open source software, open compute architectures, and open data stores like Amazon S3 and Azure Data Lake Storage. It also means open integrations and open partner ecosystems. By avoiding lock-in, organizations can control their data and their future.

Use cases

Below are examples of companies and government organizations that have successfully applied enterprise data cloud principles to unlock valuable insights and power their mission transformations:

IoT and connected data

Octo Telematics provides telematics and data analytics solutions for the auto insurance industry. The company relies on data—11 billion data points generated by five million connected cars driving more than 170 billion combined miles—to change how insurers assess risk, deliver crash and claim services, and manage customer relationships.

The ability to flexibly use major cloud service providers such as AWS, Google Cloud Platform, and Microsoft Azure means Octo customers can support transient but compute-intensive projects, such as testing new pricing algorithms or risk model development, on a usage-based commercial basis.

As frequently occurs in a government context, Octo faced the significant task of leveraging vast amounts of data to accomplish its mission. By moving to an edge to AI data platform on hybrid cloud architecture, the company has significantly enhanced its goals and become driven by previously unavailable actionable intelligence.

Population data management

The **United States Census Bureau** serves as the nation's leading provider of quality data about its people and economy. The next census in 2020 will require counting an increasingly diverse and growing population of around 330 million people in more than 140 million housing units. To get an accurate count, the Census Bureau must build an accurate address list of every housing unit, maximize self-response to the census, and efficiently follow up with those who do not respond.

Adopting an edge to AI data platform has transformed how the Census Bureau processes the country's demographic and economic data using open source technology and high-performance cloud infrastructure. By establishing the Census Enterprise Data Lake vision, Census is able to enhance its big data processing capabilities to fulfill petabyte-scale data management and analytics, while satisfying security and privacy requirements and controlling costs.

The US Census Bureau's enterprise data cloud platform prevents unauthorized access to data, improves data visibility and regulatory compliance, ensures consistent security and governance policies across data and analytic engines, increases business agility and insight, and reduces manpower and operational costs.

Predictive and preventative care

Clearsense helps healthcare providers realize measurable value from their data. They have developed a secure, cloud-based healthcare data ecosystem that rapidly consumes data from an array of sources to provide a real-time view of any healthcare environment. The company's flagship product, Inception™, is designed specifically for the clinical, financial and operational needs of healthcare providers. Inception efficiently detects patient health deterioration, tracks key performance indicators to drive costs, and leverages data to improve operational efficiencies.

The edge to AI capabilities inherent in Inception™ led to the development of a clinical surveillance model that can predict whether a patient is likely to go into cardiac arrest within 4 to 12 hours and notify caregivers accordingly. The company is also applying predictive algorithms to reduce cases of sepsis.

Summary

These customers highlight the scope of what's possible for the insight-driven organization. The government must think first about the needs of its missions and then design the insight-driven use cases that power the mission strategy. Once use cases have been designed, you can start to think through your data and analytics needs and ultimately the environment in which your data management architecture will run. An enterprise data cloud helps the government extract meaningful insights from data and fuel innovation.

Cloudera offers the breadth of data analysis disciplines needed to solve the most demanding government use cases. Data disciplines from the Edge to AI—edge analytics, streaming analytics, data engineering, data warehousing, operational analytics, data science, and machine learning—all working together, securely, and operating across your choice of environments—data centers, multiple public clouds, and hybrid cloud.

Our current offerings—Cloudera Data Platform and Cloudera Data Flow—offer the key capabilities of an enterprise data cloud—hybrid and multi-public cloud, multi-function analytics, shared security and governance services, and an open-source platform with a choice of compute and storage. Cloudera Data Platform (CDP):

- ✓ Empowers customers to extend current Cloudera deployments with native cloud services on AWS and Azure
- ✓ Offers a full complement of open-source data management and multi-function analytics, with the agility, elasticity, and ease of use of a public cloud-like experience
- ✓ Provides a single control plane to manage infrastructure, data, and analytic workloads across hybrid and multi-cloud environments
- ✓ Extends SDX shared services to safeguard data privacy, regulatory compliance, and cybersecurity threats across all cloud environments
- ✓ Is 100 percent open source, supporting government objectives to avoid vendor lock-in and accelerate enterprise innovation

For additional detail on the enterprise data cloud and CDP availability, please visit www.cloudera.com.