Getting the most from data with an enterprise-wide strategy

FedScoop Report

Agencies won’t get the most efficiencies from their data if it isn’t accessible across the lines of business. This need underlines the importance of governance tooling and data strategies.

Cloud services offer the promise of new efficiencies in IT infrastructure and operations. But in order for agencies to realize greater value from their modernization efforts, leaders need to build a data strategy that is both enduring and flexible.

For organizations to transform the way they consume and leverage data, an enterprise-wide data strategy needs to focus on evolving the processes and tools to address multiple strategic objectives across lines of business.

It’s not just about building cloud capabilities but creating an environment that facilitates data mobility and seamless governance, with security policies that move with the data no matter where it resides, explains Henry Sowell, CIO of Cloudera Government Solutions. “When leaders focus on building a data strategy, the policy pieces they put together help them take full advantage of the benefits of the cloud,” he says.

Paving the way to an enduring data strategy

The Office of Management and Budget’s Federal Data Strategy 2020 Action Plan sets the stage for how agencies establish processes, build capacity, and align efforts to leverage data to better serve the mission. While the action plan lays out some fundamentals, experts believe that until agencies also take steps to standardize and share their data, they will be hard-pressed to realize the potential payoff.

The challenge becomes how to prioritize the work of executing the Federal Data Strategy Action Plan, given the needs of agencies, and to determine which technology tools are best suited to get them there.

Sowell sees six factors to paving the way to an enduring data strategy:

1. Establish a data strategy around mission goals.
2. Hire or retrain people to manage data and guide the data strategy.
3. Build a multi-cloud environment to ensure data mobility across the organization.
4. Integrate tools to be able to ingest, transform, query, optimize, analyze and do predictive work with data.
5. Ensure universal security and governance policies through the lifecycle of the data.
6. Use tools that are open source and future-ready to integrate artificial intelligence and machine learning capabilities down the road.

A platform built to deliver a consistent, shared data experience and standardize user experience across functions can facilitate these strategic goals while making it faster and easier for program leaders to analyze data, says Henry Sowell.

U.S. Census Bureau’s big data experience

Unifying data management and analytic functions across one platform greatly simplifies an agency’s ability to process and create added value from the data troves, Sowell says, pointing to efforts at the U.S. Census Bureau as an example.

The Census Bureau recognized that its vast collections of data would require a different type of management solution when it started its digital overhaul for the 2020 census. When the agency modernized operations to move online, it expected to receive petabytes of data that would need to be stored, analyzed and secured.

Census Bureau leadership established the Census Enterprise Data Lake (EDL) initiative to provide coordinated capabilities to process, manage and analyze the flood of new data expected in the 2020 census, while also satisfying security and privacy requirements. The EDL approach reduces total operating costs and helps protect the quality and integrity of data.
costs and drives efficiencies throughout the agency's business operations.

The Census Bureau chose Cloudera's HDP (Hortonworks Data Platform) to help mine, process, and extract data and generate insights that inform a wide range of decisions at all levels of government.

In addition to developing a census questionnaire that simplifies the collection of data, efforts were made on the backend, analytics tools, that helped the agency to easily clean its data, by cross-checking with existing administrative records. That helps the EDL reduce redundant data collection, yielding better data to support the bureau's mission.

Creating a consistent data experience

Sowell explains that behind the scenes is a powerful engine — Cloudera’s Shared Data Experience (SDX) — that facilitates the large-scale exchange of information for faster analysis.

Built across the suite of solutions on Cloudera’s Cloud Data Platform, SDX ensures that data has consistent security and governance across any cloud or on-premise environment.

SDX allows data managers to set security and governance policies once and apply them across all data and workloads. Those policies stay with the data even as it moves across all supported infrastructure.

Enterprisewide layers of data security with technologies such as Cloudera’s Shared Data Experience provide the tools to make an agency’s enterprise data cloud secure by design, explains Sowell.

Secure by design means the data will never leave the public cloud account. The architecture of the public cloud works across two virtual networks. The first is the control plane on the Cloudera network, which is used to create virtual operating environments, provision clusters, and manage experiences. The second is the public cloud vendor’s virtual network where all the data lives.

This architecture can give data scientists access to citizen data, without compromising personally identifiable information (PII). Or it can give engineers an elastic workspace for development, without compromising the data for the rest of the organization.

The Cloud Data Platform allows users to efficiently find, curate and share data, enabling access to trusted data and analytics on two planes.

One platform for better data insights

And as agencies progress with their data strategies, they need a solution that provides flexibility to host data in a multi-cloud environment and can adapt to whatever future changes lay on the horizon for enterprise IT.

An integrated enterprise data platform, like Cloud Data Platform and SDX, is built to manage data assets securely over their lifetime, with full data mobility, both on-premise and in the cloud. It can ingest, transform, query, optimize and use predictive analytics, all with a foundation of a shared data experience for universal security and governance policies that IT and data leaders require.

All these solutions are open source, which gives the government greater long-term flexibility so, regardless of how IT environments change, agencies can continue to shape their modernization journey with no risk of lock-in to a single solution.

“Cloudera works with over 2,400 partners, so agencies don’t have to get rid of their current investments to work with Cloudera,” says Sowell. “They can continue to integrate their existing investments into our open source platform, without risk of vendor lock-in.”

Data capture, data movement, data engineering, data science, and data storage are all present on the Cloud Data Platform. So, instead of single-point solutions to perform these functions, agencies can use a single platform, effectively lowering operational costs and ensuring uniform security and governance policies.

Learn more about creating an enterprisewide data strategy for better insights and security.

This report was produced by FedScoop and underwritten by Cloudera.