



Why increasing demand for real-time data requires more powerful tools

While the Federal Data Strategy offers a framework for capitalizing on agency data, IT leaders still need robust integration tools to “master, not just manage” the data.

FedScoop Report

As the nation’s top government officials grapple with the continuing consequences of the COVID-19 pandemic, one factor in their decisions has never been clearer: The fundamental need for up-to-date and reliable data.

While the government’s continuing migration to cloud services has given agencies new-found capabilities, many officials have been reminded of how difficult it still seems to locate, share and analyze trustworthy data quickly in order to make critical daily decisions affecting millions of people.

What the public—and many of those same officials—also have witnessed first-hand are the

apparent shortcomings of agency decisions to sustain, rather than replace, outdated IT systems and data management practices.

[Stories](#) of federal agencies sharing data on spreadsheets and posting out-of-date information on websites suggest the need has only grown for deploying modern, enterprisewide data management solutions to deliver the right information to those in search of it.

“We’ve heard a number people in senior positions say publicly, ‘We’re not getting the data fast enough, or it’s not good enough. Our reports and predictions, and our public announcements, are based on data that’s days old,’”

says Michael Anderson, Chief Strategist, Public Sector, for Informatica, a global leader in enterprise data management.

“When you look at all the predictions, all the analytics going into decisions on whether to shut down [businesses], when and for how long, it all depends on having clean, timely data, run through a decision model or an AI tool. If an organization is not set up and prepared to do that before a crisis hits, they’ll run into some of the problems many are having now,” he says.

The hunt for authoritative data

It’s not that agencies don’t have sufficient information—they’re

already inundated with data. Nor, for the most part, do they lack sophisticated analytical tools; occasionally they even have the experts who know how to use them. Rather, it’s that so much of their data remains inadequately discovered and cataloged—making it difficult to know what you have and what condition it’s in, says Anderson. And increasingly, agency data is distributed across a multitude of IT systems and third-party cloud platforms, making it harder to find and assemble.

Congress and the Office of Management and Budget have been pushing federal agencies for much of the past decade to be more accountable stewards of the government’s data—with mandates to make it more open and accessible to the public, as well as to drive internal decision-making. OMB’s latest salvo came in December, with the [Federal Data Strategy 2020](#)



“Without [automated] cataloguing capability, it’s going to be very difficult to find and identify the right data to feed into artificial intelligence tools...and your analysts are going to struggle delivering meaningful results.”

Michael Anderson, Informatica

[Action plan](#), which has created new urgency for agencies to get on top of their data management practices.

Cloud experts like Susie Adams, chief technology officer for Microsoft Federal, have seen from personal experience how widely distribution pools of information make it challenging for federal agencies to rationalize their data in order to migrate to the cloud. The need to find and collect datasets also hinders the ability for agencies to take fuller advantage of high-powered cloud data analytic tools.

“When agencies start to investigate big data, artificial intelligence, machine learning and data analytics technologies to analyze very large data sets, one of the biggest challenges agencies have is that the datasets are distributed and stored in multiple disparate locations,” says Adams. Another big challenge she sees, “especially with large

agencies, is they really don’t really have a handle on what many of their authoritative data sources are.”

When an agency analyst or program manager has to prepare reports, she says, they typically have to pull data from multiple sources in multiple locations. Data might be stored in multiple clouds and on-premise. “You’re spending 80% of your time finding the data you need instead of making informed decisions based on the data itself.”

The key to mastering your data

Adams and Anderson, whose companies frequently work together to help government customers resolve enterprise data challenges, both agree that agencies need to establish two foundational competencies in order to “master, not just manage, their data.”

“One of the most important things, as part of an overall data management program that must be in place,” says Anderson, “especially when you move to the cloud—but even if you’re not moving to the cloud or, maybe, you’re still a year away—is having data governance in place.”

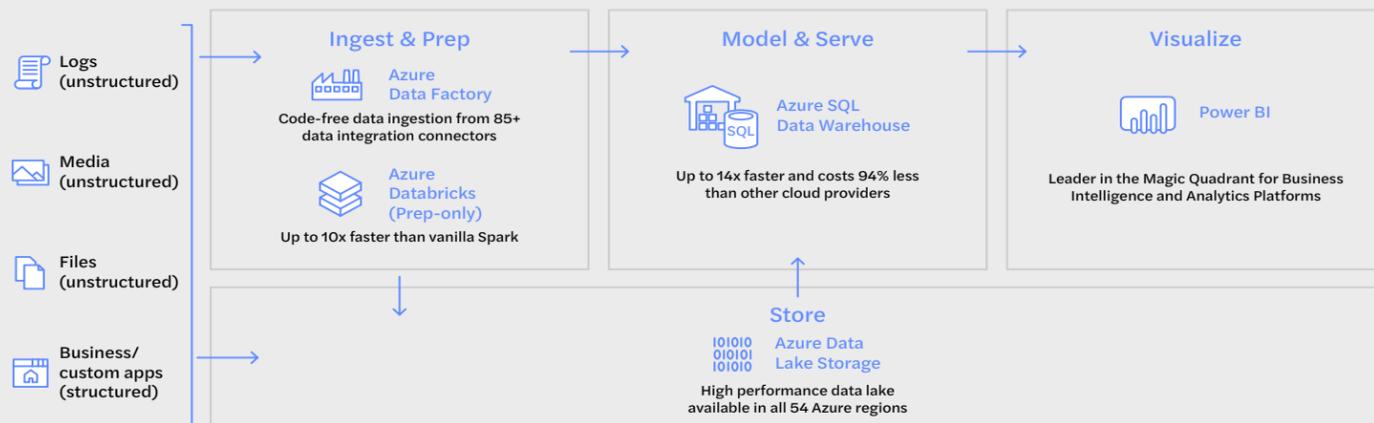
Anderson suggests agencies not only establish their own governance councils, similar to a

The 'Golden Path' to preparing your agency's data to operate in the cloud

There are a lot of processes and steps to preparing your data to function cost effectively in the cloud. Microsoft Federal CTO Susie Adams offers best practices on what's helped agencies be more successful in making the move:

- 1 Discover and rationalize your data.** Determine which applications produce and collect structured and unstructured data across your agency, using modern data cataloging tools, to create a clear and holistic picture of the data and applications you have, how they're used and where they are located.
- 2 Determine the targets.** Then figure out which applications and data make sense to actually move to, adapt for, or operate in the cloud. "Don't just lift and shift."
- 3 Prepare your data.** Make sure the source of the data is authoritative, that the data is clean and properly tagged, and ultimately is the right data to be used in the cloud now and in the future—especially when migrating very large data sets.
- 4 Determine how to ingest the data.** Resolve how best to provide end users with access to the data and applications—deciding for instance, whether the data should reside in a SQL data warehouse or an Azure data lake—and then determine how to ingest or move the data.
- 5 Model and serve the data.** Once the data is in the cloud, test different approaches to model, serve and visualize the data, to create new insights and value.

Best -end-to-end ecosystem to turn your data into actionable insights



Source: Microsoft

governmentwide data governance council initiated by OMB, but also to acquire the capability to implement policies and decisions. "A data governance council is important because they are the ones to establish how data is used over its lifetime across the organization. The council needs to establish who are the stewards of the data, who secures or protects it, and who authorizes who has access to it," he says. "Further, the council then must have a mechanism to implement, measure and automate compliance with the data governance council decisions."

Another important role of governance is to establish the ground rules for defining data, includes maintaining systems and processes to ensure data quality. That includes having comprehensive data glossaries that standardize the formatting and meaning of data.

The second foundational component is having "proven, robust and automated cataloging tools to properly identify, tag and process your data at scale," says Adams. "Once your data has been properly cataloged, getting it migrated and then standing

it up in the cloud, can be pretty straightforward."

"Federal government data is stored on many different vendors' resources," she explains. "It could be in an Oracle database, it could be in Hadoop, it could be in a legacy system. It could be structured or unstructured. So how do you actually get this data into, for example, a SQL data warehouse or an Azure data lake, so that you can have it in a single data source—all while making sure you're actually migrating the right data?"

Anderson likens the challenge to "going into the Library of Congress not having an automated way to find a series of books you want. If you don't catalog the books in a comprehensive way—that takes advantage of embedded artificial intelligence and that will help you put in a data query and identify related datasets—you'll likely overlook all kinds of meaningful information."

Supplying the demand for AI-ready data

The power of enterprise data cataloging goes beyond knowing what you have, says Anderson. It's becoming essential as agencies begin to focus on ingesting their data to train artificial intelligence tools and develop advanced analytics.

"Without that cataloging capability, it's going to be very difficult to find and identify the right data to feed into artificial intelligence tools," he says. As importantly, "if you feed in the wrong data, or it's not complete data—or you haven't found it the massive amounts of unstructured data in your data lake—your analysts are going to struggle, or spend too much time, delivering meaningful results."

Finally, he adds, agencies need to know "how does the data flow through an organization? And you need to be able to do that on a consistent basis. It's not just a



"When agencies start to investigate big data, artificial intelligence and data analytics technologies to analyze very large data sets, one of the biggest challenges is that [agency] datasets are distributed and stored in multiple locations."

Susie Adams, Microsoft



one-and-done event, because there's new data being created or collected every day.

One of the reasons all of the leading cloud providers, including Microsoft, work with Informatica is the comprehensive array of data management tools that Informatica offers. Informatica's experience working with large government enterprises for over two decades has also helped the company keep innovating. No fewer than five of Informatica's enterprise management solutions are recognized as "Leaders" in [Gartner's Magic Quadrant Reports](#)."

From Adam's perspective, Informatica's "Extract, Transform and Load (ETL) capability gives us speed—it can really accelerate data migration from on-premises to the cloud, by discovering and

automatically cataloging and classifying data," she says.

Adams insists that the upfront work is a crucial step, if not the most important one. "Without going about it in an intelligent way and planning through the process, you're just storing everything in the cloud, paying for consumption you're not using. Microsoft really doesn't want that, because in the end it doesn't help our customers."

Making the effort to properly catalog your data also helps give top agency leaders greater assurance about the quality of the data they're relying on, adds Anderson. "When you think of the Department of Homeland Security, the Centers for Disease Control and Protection, or the U.S. Food and Drug Administration, and all the things they monitor, providing the right data to decision makers is crucial. So is having the right tools to deliver the best outcome."

[Find out more](#) on how Informatica can help your agency master and not just manage its data.

This report was produced by FedScoop and underwritten by Informatica.