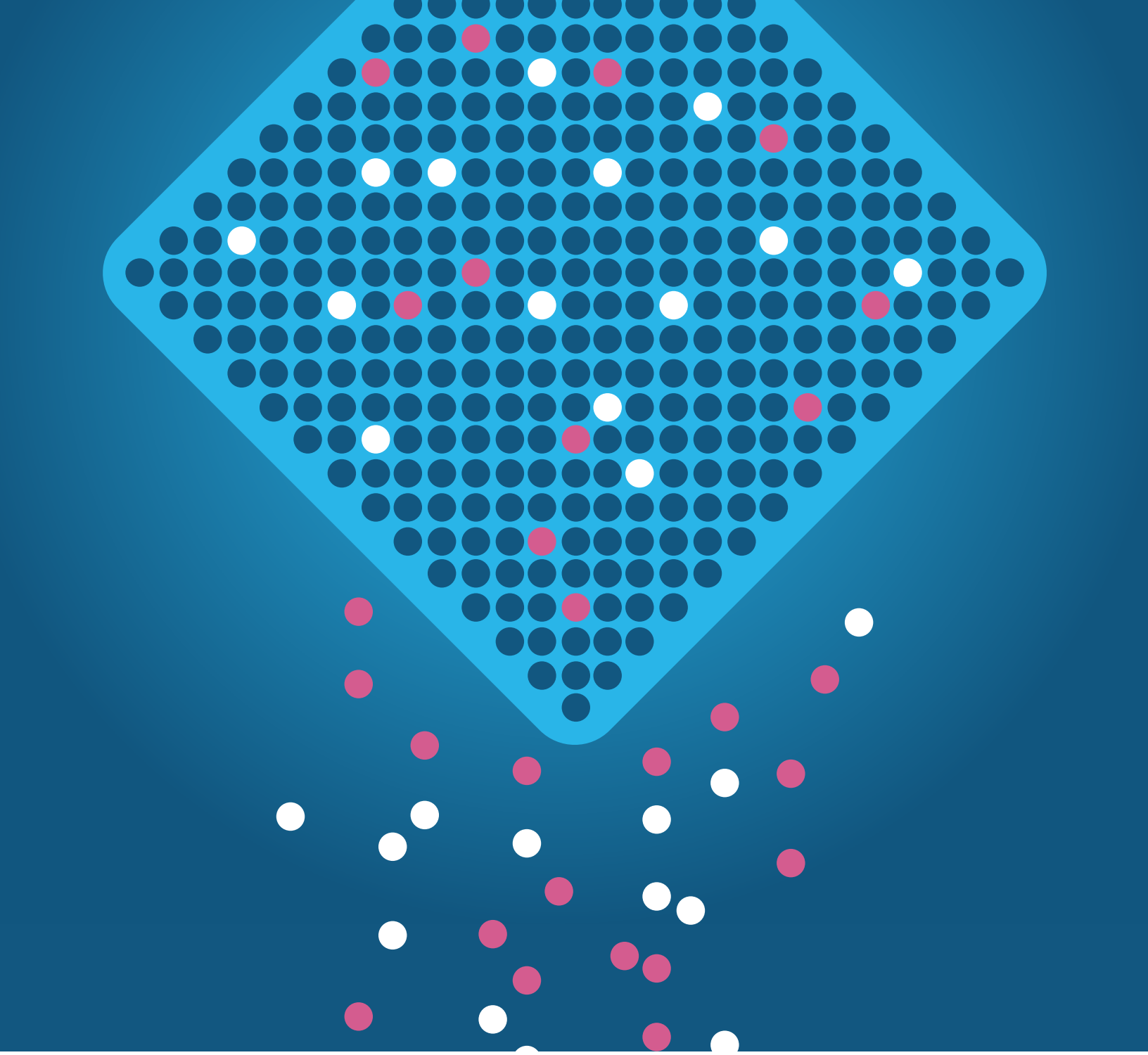


The Need for a Single Source of Data Truth

MARKET TRENDS REPORT

govloop snowflake fedresults

Introduction

This scenario is all too common in government: A doctor at the Department of Veterans Affairs (VA) wants to use data from the Department of Health and Human Services (HHS) to improve a treatment plan. But because that data is scattered across several systems, there's no easy way to efficiently search or access it, let alone analyze it. Now, the COVID-19 pandemic has brought to the forefront just how dangerous it can be when agencies can't efficiently collect, share and analyze data. Data-sharing has gone from an ideal to a necessity.

Artificial intelligence, machine learning, robotic process automation and other technologies can simplify the task for government employees of finding and acting on patterns. How to manage that data, however, remains a critical pain point for many agencies.

Traditional approaches to data management weren't built to handle the volume of data we see today. As a result, data ends up in siloed databases that prevent sharing, stymie analysis and make it challenging for employees to be data-driven.

Today, agencies are moving toward a new approach: creating a single source of data truth by consolidating all their data in a single cloud data platform that serves as the system of record.

To better understand this trend toward a single source of data truth, GovLoop partnered with Snowflake Cloud Data Platform. In this report, we explore the challenges of legacy systems and share how an effective cloud data platform enables governments to efficiently share, access and analyze data.

By the Numbers

The Federal Data Strategy mandate for agencies to hire a Chief Data Officer and the creation of a Federal CDO Council lends weight to the importance of making data central to how agencies operate within their departments and with other agencies. The Centers for Disease Control and Prevention is just the latest agency to announce it is actively seeking a CDO.

10 million

the number of data resources on Data.gov

22%

how much of their IT budgets federal agencies spend on cloud

Source: Gartner

20%

the increase in investments using cloud services at the General Services Administration (GSA) and Social Security Administration (SSA) between fiscal years 2016 and 2019

Source: Government Accountability Office (GAO)

2.5 quintillion

bytes of data are created per day.

Source: Forbes

"If the Federal Government does not maintain its role as a preeminent supplier and sophisticated and ethical user of data, it will no longer be able to fulfill its civic duty to the public."

Source: Federal Data Strategy

51

the age of a Treasury Department system that contains taxpayer information.

Source: GAO

$60.87 million

how much the Technology Modernization Fund awarded to four projects that plan to migrate or deploy systems to cloud services, as of March 2019

Source: GAO

THE CHALLENGE
Legacy Systems Hinder Sharing & Analyzing of Data

The President's Management Agenda provided agencies with a clear goal: Leverage data as a strategic asset. In theory, they are in great positions to do so. They have growing volumes of data available to use and develop actionable intelligence to drive decision-making.

Most federal agencies, however, store much of this data in legacy systems, some of which are decades old. These systems create information silos that make it difficult for agencies to share and analyze their wealth of data and create a "single source of data truth." Without a single source of data truth, misinformation can spread, which can hurt public trust and hinder agencies' abilities to perform critical missions.

That's not to say agencies don't try. Inspired and resourceful data experts will find tools and processes to work around those barriers and get their work done. But that's far from an ideal scenario.

"If you spend all of your time developing the tools to answer questions, the motivation to keep asking questions becomes lesser and lesser," said Nicholas Speece, Chief Federal Technologist at Snowflake.

The data environment itself has also grown more complex. About 10 years ago, the emergence of the Internet of Things (IoT) and new datasets generated by sensors led to the development of new platforms such as Hadoop and Google Big Query. These platforms could analyze a lot of data simultaneously but couldn't scale. Additionally, they were tough to develop applications for because they used specific languages and required developers to have targeted skills. Security is also a concern for agencies as they look to secure their data without it becoming an obstacle to data-sharing.

THE SOLUTION: A PURPOSE-BUILT CLOUD DATA PLATFORM

Whether it's different departments within an agency, different agencies or even different state and local governments, data is currently not always easily shareable or quickly accessible to those who need it most.

The key to addressing this challenge is to move data to a cloud data platform, where all data from across the organization can be accessed, shared, analyzed and protected in one location – a single version of truth.

For that to happen, agencies need to take a different approach to data management. In a traditional legacy environment, storage, compute power and data pipelines are interwoven, which creates silos, whether on premises or in the cloud.

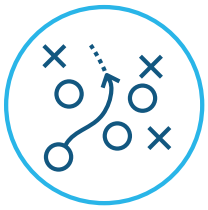
"Those three things should be completely independent and decoupled, yet still tightly integrated," Speece said. "They're the three biggest components of getting data into the cloud. If I'm going to take data from seven silos in my organization, I don't want to just create seven new silos in a cloud server."

In a modern cloud data platform, storage, compute and services layers logically integrate but scale infinitely and independent of one another. That decoupling has important financial ramifications. Traditionally, buying more compute means getting more storage because they're connected. A cloud data platform lets agencies efficiently build an environment that can scale a single component without affecting others.

A cloud data platform serves the middle ground between data lakes — where raw data is stored in bulk — and data warehouses that put data through a transformation process, normalizing it so that it's easily and readily consumable. In a data warehouse, agencies can combine semi-structured data such as Extensible Markup Language (XML) and JavaScript Object Notation (JSON) with structured data in one place to enable ad hoc queries separating storage from compute.

Cloud-ready data platforms can scale, access more sources and incorporate better analytics, which leads to better and quicker decision-making. This also breaks down both intra-agency and interagency silos because data stored in the cloud can move among agencies. Finally, security is improved and simplified in a cloud environment.

BEST PRACTICES
Enabling a Single Source of Data Truth



Start with a strategy.

What data do you have in your current environment, and what data do you want to move? How is it going to get into the cloud data platform, and what will it do once it's there? How can the data best be shared? Five to 10 years ago, the common migration approach was "lift and shift" — taking the data as-is and moving it to its new location — but that's just moving data, not a true migration, Speece said. "Get that foundation built and then build the applications on top of it. It's all workflow- and use case-based," he said.

Move data and build prototypes.

Once you have a strategy in place, start moving data in a phased approach that prioritizes what you need to get from the data. Ultimately, a plan should get as much data into a single source of truth as possible because the more versions you have, the more confusion you'll have, Speece said. Then, build prototypes to make quick wins that will foster buy-in.

Trust in security.

Cloud can increase data security, according to the Federal Cloud Computing Strategy, especially when agencies move or add controls to the data layer itself. A cloud data platform should include a multilayered security strategy that addresses encryption, access control, data storage and physical infrastructure in addition to monitoring and alerts. To ensure that such security mechanisms are part of the foundation of the data platform, government agencies should work with trusted vendors, such as those certified through the Federal Risk and Authorization Management Program (FedRAMP).

Get leadership support and general buy-in.

This type of effort requires a cultural change, not just a technological one. To get buy-in from leaders, use a cost model that shows how the agency can do more for less money, such as how employees agencywide could be more productive and derive greater satisfaction from their jobs through data use. To get employees on board, show them how having access to more data will make it easier for them to do their jobs. "Once you start showing people the value of the data and the value of giving that data back to the people who create it, I think that's how you start the cultural change," Speece said.

> "Once you start showing people the value of the data and the value of giving that data back to the people who create it, I think that's how you start the cultural change."
>
> *- Nicolas Speece, Chief Federal Technologist, Snowflake*

CASE STUDY

Case Study: Sharing Data During COVID-19

As COVID-19 spread from China's Hubei Province westward across the globe, health and government officials and the public at large became laser-focused on one thing: data – specifically publicly available and shareable data.

The dataset that serves as a single source of truth about COVID-19 resides on the Snowflake Data Exchange and is being made available at no cost to any organization worldwide that wants to use it. Because it's in the cloud, data can be loaded from multiple sources into one dataset that users can then access from anywhere.

Gathered by Starschema, a Hungarian data company, the dataset is composed of data points from sources such as the World Health Organization, the COVID Tracking Project, state and government organizations and Johns Hopkins University's Center for Systems Science and Engineering. Additionally, the dataset has details on demographics for individual geographical units and health care provisions, such as the availability of intensive-care beds at hospitals and the location of hundreds of thousands of health care providers and pharmacies worldwide.

"There has not been a time in history when this much data was available about any particular outbreak or epidemic," said Chris von Csefalvay, Vice President of Special Projects at Starschema.

On their own, each dataset may be of limited use, but together, they enable more intricate analysis. For example, case counts have a singular meaning, but mapped against the number of staffed ICU beds by county, users can see where there is a high risk that the health care systems will be overwhelmed.

"When we use data in decision-making about critical situations and crises like the COVID-19 outbreak, it's absolutely indispensable that we have solid, quality data unified along the same identifiers so that we can bring multiple resources in and collate the value of each individual data source into one particular dataset," von Csefalvay said.

HOW SNOWFLAKE HELPS

Snowflake offers a FedRAMP-authorized cloud data platform. It is cloud-agnostic and works across multiple environments that operate in modern IT environments. Specifically, Snowflake provides a fully relational, SQL data warehouse that lets users make the most of skills and insights from other clouds or on-premise systems, but in the cloud, opening them to more analytical capabilities and datasets. This also paves the way for agencies to incorporate emerging technologies such as automation, machine learning and IoT into their analytics.

"We're the first foundational component to any cloud approach," Speece said.

Snowflake is FedRAMP-authorized on both AWS and Azure Government cloud. It equips organizations with a single, integrated cloud data platform. The platform is instant, secure and governs access to agencies' entire network of data. It offers a core architecture to enable many types of data workloads, including a single platform for developing modern data applications.

Learn more at snowflake.com/federal.

Conclusion

The need for modern, data-driven operations and services is not new. Agencies have been working for some time to implement AI, cloud, machine learning and RPA to optimize their data. The challenge has been in breaking down traditional silos that make sharing data within and among agencies difficult and, as the COVID-19 pandemic has shown, detrimental. Recent events have shown that data-sharing is a necessity, not an ideal.

Agencies have also struggled with the record number of data sources and bringing them together in a useful, meaningful way. The most effective approach to managing all that data and creating a single source of truth is for agencies to consolidate it in a cloud data platform that becomes the system of record for everyone.

ABOUT SNOWFLAKE

Snowflake Cloud Data Platform shatters the barriers that prevent organizations from unleashing the true value from their data. Snowflake equips agencies with a single, integrated platform that offers the only data warehouse built for any cloud; instant, secure, and governed access to their entire network of data.

Find out more at snowflake.com/federal.

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.

1152 15th St. NW Suite 800
Washington, DC 20005

P: (202) 407-7421 | F: (202) 407-7501

www.govloop.com
@GovLoop