The Impact of Multi-Cloud Environments on Data Governance

By Paul Davis
NetApp CTO

When it comes to cloud adoption in the federal government, we're far beyond the decade-old question of, "Will they or won't they embrace the cloud?" The fact is, today's federal agencies haven't just embraced a single cloud solution or cloud service. They haven't just migrated one single workload or application into the cloud. Today's federal agencies have embraced multiple cloud services and solutions.

Between Software as a Service (SaaS) solutions that agencies are procuring licenses for (e.g., customer resource management [CRM], Payroll, HR, etc.) and leveraging in the cloud, to the multiple, disparate AWS, Google Cloud, and Azure instances that are being used to develop new applications or store data, federal agencies have gone "all in" on cloud solutions.

Agencies have embraced the many benefits of multi-cloud infrastructure that include scalability and cost. But this infrastructure also comes with ramifications and challenges – including impacts on data governance.

What is data governance? My associate at NetApp, Jim Cosby, recently shared an excellent and succinct definition that I think is quite accurate. According to Jim:

"Data Governance can mean different things to different agencies and companies, but it is a set of steps and processes on how to manage data assets for any enterprise which will help set forth the rules and authorities around how to manage the data for that agency or company."

Establishing and enforcing data governance is often one of the responsibilities of an agency's Chief Data Officer (CDO). And the job of the CDO is becoming increasingly more complex due to a number of different factors, one of which is the evolution of multi-cloud environments.

Managing massive amounts of data, clouds, and mandates

Paul Davis

NetApp CTO

Modern CDOs are facing three converging trends that make their job increasingly difficult. First, the sheer amount of data being generated by government agencies is increasing exponentially. Government IoT and modernization initiatives – and the adoption of "network-enabled everything" – have agencies aggregating data of incredible quantities.

This data being generated by digital platforms, IoT devices, connected sensors, and the massive ecosystem of end-user devices has to be stored, managed, and analyzed if agencies are going to get any viable benefit from it. And in today's government IT landscape, that data storage invariably is happening in a multi-cloud environment.

"...the job of the CDO isn't just about creating governance that works for the agency, getting policies on the books, and encouraging the adoption of tools and solutions that enable compliance with data policies. They also need to get agencies bought into the importance of data governance and build a culture of compliance with data policies."

These multi-cloud ecosystems are the second trend impacting the job of the CDO, and they didn't evolve by accident. There are a number of reasons why agencies have found themselves embracing multiple different cloud services and solutions from a wide ecosystem of providers.

Not all clouds are good at the same things or are good for particular use cases or workloads. One cloud provider might offer a particularly powerful or useful AI/ML tool that makes it the right choice for an AI/ML program or solution, but not the best choice for hosting back-office productivity applications.

Then, there's the reality of how government agencies do business. Agency procurement and acquisition often begins with bids being solicited for a particular service or solution, and the lowest cost bid is often chosen. This creates an environment that fosters multi-cloud environments because different cloud providers might wind up being the low-cost or preferred vendor for a particular project or program.

"...the job of the CDO is becoming increasingly difficult thanks to the confluence of three major trends – the increase in data, the increase in data policy, and the evolution of multi-cloud environments. But they have the tools available to them to make data governance effective and enforceable if they can get their agencies on board."

But regardless of why these multi-cloud environments exist in such frequency across federal agencies, they're now the standard. And these multi-cloud environments are adding complexity to data storage and management – with data spread out across multiple disparate clouds from multiple disparate cloud providers.

Finally, there are government data and privacy regulations and policies that are constantly expanding in scope and increasing in number. For example, federal agencies that deal with private healthcare data find themselves bound by the data requirements and mandates laid out by HIPAA. As new data and privacy legislation is introduced and adopted, the rules around what can be stored, where it can be stored, and how long it can be stored all need to be addressed.

When you add these three trends together, you wind up with a very challenging scenario for agency CDOs. They have mountains of data coming in that will be stored and accessed from across a complex, multi-cloud environment, and that data has an increasing number of rules and requirements that the agency needs to be in compliance with.

And this is the challenge that CDOs today face when it comes to establishing and enforcing data governance – more complexity, more data, and more rules that all need to be accounted for. So, what tools does the modern CDO need at their disposal to ensure that they can establish and enforce governance?

Picking the right tools for the tough job

As we've established, the CDO is on the hook for the creation of data governance. That process requires taking policies and mandates from other organizations and aligning them into a set of policies that can be applied to their own organization. But it's a process that is also a balancing act.

Governance that is too weak might not meet stringent data privacy standards and requirements. Or, far worse, it can be susceptible to data leakage and cybersecurity challenges. This is especially concerning in complex multi-cloud environments where the attack surface is larger and more complex to secure.

However, if data governance policies are too onerous to implement, there could be other ramifications. While the data might not leak or be compromised, it might not be able to be used as required by the agency.

This is where balance becomes essential – CDOs need to develop data governance that is strict enough to meet compliance requirements and product data, but not so restrictive that it undermines or negates the agency's ability to leverage its data to accomplish its mission.

To establish balanced governance and enforce those policies, there are a few capabilities that they need, and a few different solutions that can deliver them:

- **Data tagging for easy classification:**

Managing data becomes much easier if it can be tagged. By tagging content, agencies get the ability to more easily categorize, aggregate, and manage data. For example, all data with private health information can be tagged "Protected Health Information (PHI)." This would make it easier in the future to identify where that data is being stored, how it's being used, and which data governance policies are being applied to it.

- **Automation:**

Once data is tagged, it's incredibly useful for CDOs to be able to create policies that can be implemented as rules or algorithms that will identify the tags and apply particular rules to any data with those tags. Using the same example, all data with the "PHI" tag could be automatically stored in a particular location or be unable to be used in a particular way based on HIPAA requirements.

- **Easy data movement within and between storage devices and clouds:**

Data governance policies about where data can be stored, how it has to be secured, and how many different locations data can exist would be incredibly difficult to enforce if data couldn't be easily identified and then moved seamlessly, as needed. Solutions that can enable government agencies to comply with CDO policies as

> **"Agencies have embraced the many benefits of multi-cloud infrastructure that include scalability and cost. But this infrastructure also comes with ramifications and challenges – including impacts on data governance."**

easily as “dragging and dropping” data and workloads between storage locations – including between different cloud solutions – are essential for making compliance with data governance policies a less heavy lift.

Luckily, there are a number of tools available to government agencies today that are capable of delivering these capabilities.

NetApp’s BlueXP, ONTAP, and StorageGRID can give agencies the ability to tag data for easy identification and management. They can provide a single, centralized control pane in which data can quickly be located and moved seamlessly between clouds – even cloud instances from disparate providers. They can also enable automation so that enforcing data governance policies is less of a heavy lift on IT organizations.

But the job of the CDO isn’t just about creating governance that works for the agency, getting policies on the books, and encouraging the adoption of tools and solutions that enable compliance with data policies. They also need to get agencies bought into the importance of data governance and build a culture of compliance with data policies.

The capability for the government to have effective, enforceable data governance exists today. Agencies just need to make it a priority. It’s incumbent on CDOs to educate the organizations as to why they should care about data governance and be evangelists within their organizations to make them want to participate.

Yes, the job of the CDO is becoming increasingly difficult thanks to the confluence of three major trends – the increase in data, the increase in data policy, and the evolution of multi-cloud environments. But they have the tools available to them to make data governance effective and enforceable if they can get their agencies on board.