CLOUDERA
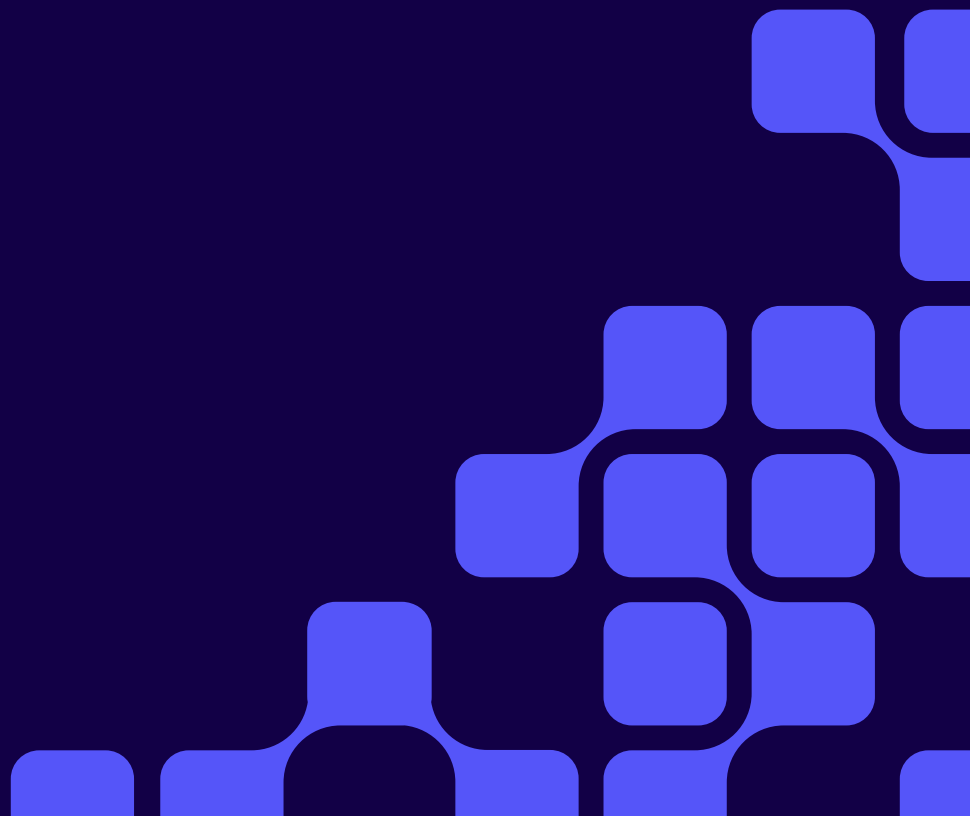
Data Strategy Drives Cloud Success

Government needs a sound data strategy to make the most of cloud

The great potential of the cloud lies in its ability to manage vast tracts of data while its scale, speed and flexibility support data-driven processes. But to take full advantage of a cloud-driven ecosystem, the government must have more than just a cloud strategy — a game plan for bringing processes and applications forward into a new hardware environment. Agencies also need a data strategy, a deliberate and well-conceived set of policies aimed at making data consistent, clean and readily available across the full spectrum of hardware implementations, both on-prem and in multiple cloud environments.

What is a Data Strategy?

At the highest level, a data strategy is a guiding document or set of principles that lay out the significance of data and recognizes its unique attributes. The data strategy addresses security, governance and the longevity of data. When applied to system architecture, the data strategy applies equally to infrastructure, applications and visualization layers.

The data strategy is essential in helping a federal agency ensure that it has a unique and distinctive plan for how it will manage ownership and control of data when engaging with, for example, a third-party cloud provider.

The rules established here will do more than just safeguard data integrity today: They will ensure that integrity remains intact even in the face of rapidly changing technologies. With the rise of the Internet of Things, for example, agencies may encounter a dramatic escalation in data speeds and feeds. A sound strategy ensures data remains inviolate across present and emerging technology infrastructures.

A sound data strategy
offers not just greater control but enhanced flexibility in a rapidly changing technology landscape.

Cloud Migration

The cloud is merely hardware, a commoditized infrastructure capable of rapid expansion and contraction in terms of both compute and storage resources. Moreover, because data drives the cloud, any cloud migration requires an effective data strategy.

The chief risk in moving forward without a data strategy is that an agency becomes beholden to whatever strategy a given cloud provider has defined. When agencies cede control in this way, they lose flexibility. Imagine an agency that opted to embrace Azure or AWS, but did so without a well-defined strategy to govern secure management of its data. Should a future cloud offering bring something better to the table, that agency could find it difficult to extract its data from the initial vendor's systems.

A sound data strategy thus offers not just greater control but enhanced flexibility in a rapidly changing technology landscape.

In this sense, flexibility equals portability. Imagine another agency that is vertically integrated with a cloud provider, but that has not defined its own parameters around data control and management. It could well lose portability in both onboarding and offboarding, as well as the ready interoperability of data between on-prem and cloud systems. That is too big a risk to run.

The problem is compounded when accounting for the need to have consistent data across multiple hardware platforms. Applications increasingly reside both on-prem and in multiple clouds, so agencies will need consistent data across this hybridized landscape, with deployments interoperating seamlessly. Standardization — achieved through consistent data policy — will reduce complexity and help ensure data security.

CLOUDERA

Next steps

For agencies looking to migrate their systems to the cloud, formulation of a data strategy begins with defining the use case. To set parameters around data access and manipulation, it makes sense to first consider the business aims of the application. How does data drive the process today, and how might it be leveraged in support of a cloud-based iteration?

Cloud will bring greater speed and scale to data-driven applications, so the data strategy should reflect that anticipated expansion. The strategy should likewise consider the potential roadblocks. Where does data get jammed today, and where might those problems worsen as data use scales up?

The strategy must also account for emerging sources of data, taking into consideration not just data at rest but also continuous IoT feeds and data produced by various publishers across the network. This encompasses emerging data from voice, video, imagery and telemetry.

Finally, agencies will want to partner with cloud providers whose philosophy and capabilities align with their newly defined data strategies. Cloudera's commitment to open source, for example, ensures easy integration on a platform that encompasses a variety of tools, apps and infrastructure, with open, backward compatible APIs. Likewise, Cloudera's true hybrid platform allows users and operators alike to manage data in any cloud environment, in alignment with any given data strategy.

An open-source philosophy and an enterprise-class, data-driven cloud architecture promote a seamless migration without the risk of vendor lock-in. A solid data strategy in turn can help ensure government agencies get the most out of their cloud investment, meeting the need for security, consistency and continuity today and tomorrow.

Why Cloudera?

1. **Hybrid and Multi-Cloud**
   Control, analyze and experiment with data wherever it lives. Easily and securely move data and metadata between on-premises file systems and cloud object stores.

2. **Analytics from Edge to AI**
   Solve demanding business use cases. Apply real-time stream processing, data warehousing, data science and iterative machine learning across shared data, securely, at scale on data anywhere from the Edge to AI.

3. **Security and Governance**
   Simplify data privacy and compliance for diverse enterprise data. Use a common security model, role and attribute based access policies and sophisticated schema, lineage and provenance controls on any cloud.

4. **100% Open**
   We empower customers with the freedom to choose — open source, open compute, open storage, open architecture and open clouds. Open for developers, partners, and open for business. No lock-in. Ever.

CLOUD=RA

About Cloudera

Cloudera is the only true hybrid platform for data, analytics, and AI. With 100× more data under management than other cloud-only vendors, Cloudera empowers global enterprises to transform data of all types, on any public or private cloud, into valuable, trusted insights. Our open data lakehouse delivers scalable and secure data management with portable cloud-native analytics, enabling customers to bring GenAI models to their data while maintaining privacy and ensuring responsible, reliable AI deployments. The world's largest brands in financial services, insurance, media, manufacturing, and government rely on Cloudera to be able to use their data to solve the impossible — today and in the future.

To learn more, visit **Cloudera.com** and follow us on **LinkedIn** and **X**. Cloudera and associated marks are trademarks or registered trademarks of Cloudera, Inc. All other company and product names may be trademarks of their respective owners.
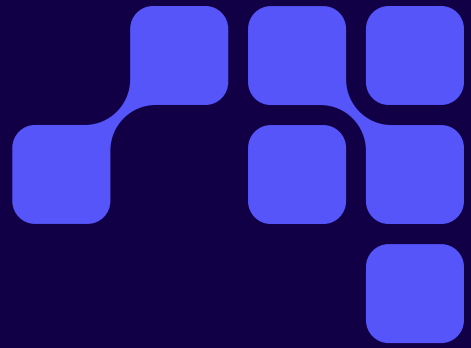
CLOUDERA   Cloudera, Inc. | 5470 Great America Pkwy, Santa Clara, CA 95054 USA | cloudera.com