

---

# Why a Successful Hybrid Cloud Strategy Requires an Enterprise Data Strategy

Five Strategic Considerations for Hybrid Cloud Success



HYBRID  
CLOUD  
SUCCESS

## Table of Contents

Introduction	3
Why have an enterprise data strategy?	3
The importance of a modern data architecture	4
1 Treat cloud as infrastructure, not data architecture	5
2 Prioritize open source	5
3 Balance business and IT needs	6
4 Keep sensitive data secure and compliant	7
5 Manage costs and optimize resources	7
Next steps	8

It's the data strategy that counts—and it should allow you to answer the following questions about your data regardless of its location, whether distributed across cloud and on-premises data lakes:

- \_ What data do you have?
- \_ Where does it reside?
- \_ How do you govern and secure it?
- \_ How do you access it?
- \_ How do you derive value from it?

### Introduction

Businesses today are moving to the cloud at unprecedented speeds. Nine out of ten companies will have some part of their applications or infrastructure in the cloud by 2019, and the rest expect to follow by 2021, according to [IDG's 2018 cloud computing research](#)<sup>1</sup>. But moving to the cloud is not an all-or-nothing strategy. In fact, most organizations—especially those with legacy on-premises applications—will settle on a hybrid cloud strategy, deploying applications, data, and infrastructure on a combination of on-premise and cloud resources.

Hybrid cloud brings substantial advantages in terms of rapid deployment and reduced infrastructure costs, but it also comes with a new set of data management challenges. As cloud environments multiply, new cloud data silos appear, some of which bypass IT altogether. Securing and governing data that lives across multiple clouds, each with their own architectures, is difficult. And cloud vendor lock-in can make it tough to migrate or export data.

Often, the result is that your cloud strategy, or lack thereof, severely challenges your ability to manage, access, secure, and derive quick insights from your data. And with [95% of C-level executives saying data is integral to their business strategy](#)<sup>2</sup>, now is the time to ensure your data strategy drives your hybrid cloud strategy<sup>2</sup>, not the other way around.

### Why Have an Enterprise Data Strategy?

Leading your IT organization through business transformation to a well-managed hybrid architecture requires new ways of thinking about your data strategy. In fact, you need a data strategy more than a cloud strategy. Your data is a strategic asset and needs to be treated as such. Cloud provides an immense opportunity for a scalable and robust delivery model, but it's the well-planned data strategy that lets you control costs and reduce risks while enforcing consistent security and governance across your enterprise data assets.

Your data strategy allows you to mitigate risks by focusing on data storage, management, and protection. It delivers security and governance, crucial for limiting fraud, preventing theft, and ensuring compliance—without which your business could face steep fines or not be permitted to operate. It also ensures the integrity or accuracy of the data flowing through your enterprise, making data known, discoverable, available, trusted, and compliant. Your data strategy should also support your business objectives—like increasing revenue, improving customer satisfaction, and driving profitability. Discovering and delivering data allows your lines of business to gain insights quickly, accelerate improvements to products or customer experiences, and understand how to gain cost efficiencies.

The importance of a modern data architecture

Your enterprise data strategy must aim to establish a hybrid architecture as a key requirement. An architecture that provides consistent data services and functionality, enabling you to share data, metadata, and workloads across bare-metal, private, and public cloud environments.

Modern Data Architecture



A modern data architecture allows you to

**Treat cloud as infrastructure.**

Cloud vendors offer critical infrastructure to power your cloud needs, but they don't provide the necessary foundational capabilities for establishing an enterprise-level hybrid architecture.

**Prioritize open source.**

Establishing an open architecture across on-premises and public cloud environments ensures application portability without any re-work.

**Balance business and IT needs.**

LOB practitioners want accelerated time to insights, so you'll need to deliver on-demand self-service access to data while allowing IT stakeholders to reduce risk and ensure compliance with enterprise standards for security, data governance, and operational reliability.

**Keep sensitive data secure and compliant.**

Define and enforce a consistent set of security and governance policies across hybrid cloud environments—including fine-grained access controls, data lineage, and audit logs.

**Manage costs and optimize resources.**

When moving data analytics workloads into the cloud, it's important to understand the key resource metrics for running your workloads, applications, and datasets and the associated cloud billing model and service costs. This is part of your ongoing efforts to plan and rationalize which workloads, applications, and datasets are suitable for the cloud and which need to stay on-premises.

This paper is designed to help Chief Technology Officers (CTOs), Chief Information Officers (CIOs), and Chief Data Officers (CDOs) understand the strategic and business considerations for managing data in a hybrid cloud environment. Because, ultimately, your enterprise data strategy will need to succeed in overcoming hybrid cloud challenges, delivering ongoing business value, and helping manage, secure, and govern data that's distributed across on-premises, multiple cloud, and edge environments.

**1. Treat cloud as infrastructure, not data architecture**

Many early cloud adopters started their journey with a single cloud vendor for good reasons—regional availability, pricing, service levels, and other factors—but now realize that their enterprise cloud strategy and their global business presence calls for multiple cloud providers or perhaps global availability. At the end of the day, cloud is providing you with access to compute and storage infrastructure much like a power utility is providing customers with access to power. Businesses should be able to take advantage of cloud compute and storage resources without fear of vendor lock-in or of a proprietary application layer hindering their progress in the hybrid cloud world.

Cloud vendors offer critical infrastructure to power your cloud needs, but they don't provide the necessary foundational capabilities required to establish an enterprise-level hybrid data architecture. Most organizations today are managing a very complex application and data ecosystem that demands a full suite of enterprise features such as security and governance, operational controls, application portability, and enterprise support.

Your cloud vendors are simply not focused on enabling a hybrid data architecture because they don't have any footprint on-premises (let alone with other cloud vendors), and they're naturally motivated to get you locked into their ecosystem of compute, storage, and applications. Therefore, your data strategy for a hybrid cloud requires additional technology that addresses your global data management requirements while fully leveraging your investment in infrastructure

**2. Prioritize open source**

As you embark on a journey to define an enterprise data strategy, think of your enterprise data assets at a global scale—spanning multiple on-premises, cloud, and multi-cloud environments. To truly grasp the potential of your data and turn it into a strategic asset, you're going to need enterprise data technology that is open source and that unifies, secures, and governs all siloed enterprise data assets regardless of where the data resides.

Remaining true to open source will help you:

- \_ alleviate vendor lock-in concerns,
- \_ benefit from the rapid pace of open source software community innovation,
- \_ take advantage of the open source ecosystem partnerships,
- \_ and ensure that your business success is not tied to any proprietary technology.

Not all vendors who provide open source software deliver intelligent support that proactively analyzes your big data environments both on-premises and in the cloud to maximize performance and reduce risk. Providers should not only be able to respond within specified SLAs but also deliver security and fix patches in a timely fashion and feed them back into the open source code.

Your enterprise support vendor must be a true partner that communicates early and often, and keeps your business interests aligned with the open software community vision and product roadmap. Innovating in the open source community is done by consensus, and voting determines whether to include code modifications. Look for a vendor that has gained favorable votes from the community and isn't simply an open source software packager or service provider.

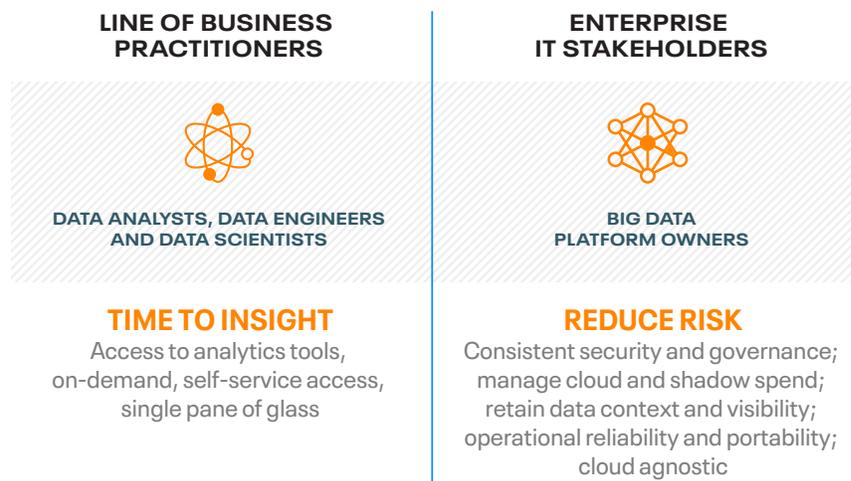
### 3. Balance business and IT needs

As you start thinking about an enterprise data strategy, it's imperative to balance the needs of your line-of-business (LOB) practitioners with enterprise IT standards. Your data strategy must accelerate LOB practitioners' time to insights with on-demand self-service access to data and a broad set of analytics tools. On the other hand, your enterprise IT stakeholders still need to retain control of the enterprise data technology, so they can reduce risk and ensure compliance with enterprise IT standards for security, data governance, and operational reliability.

Developing an enterprise data strategy that gives your LOB practitioners self-service access to the tools they need without creating shadow IT or duplicating data and analytics silos is not a trivial task. Data leaks and breaches can cripple any business, so enterprise IT must have a long-term vision for the enterprise data strategy, one that simplifies on-boarding and operations without hindering progress.

In the rush to cloud and the agility, simplicity, and efficiencies it delivers, it is paramount you understand and rationalize your data assets and workloads for cloud readiness based on security, sensitivity, ownership, and other concerns. This is especially important since you've probably invested years into implementing, optimizing, and operationalizing systems of record in on-premises-only environments.

## Balancing Enterprise Requirements for Cloud Data Strategy



#### 4. Keep sensitive data secure and compliant

Determining which workloads should remain on-premises and which should move to the cloud also requires a thorough understanding of your data assets. For instance, to reduce the risk of data breaches, you may have decided to keep sensitive data such as personally identifiable information (PII), payment card industry (PCI) data, HIPAA-protected health information (PHI), and other regulatory types of data on-premises while moving other less-sensitive data and workloads into the cloud.

But the trend is to move more and more sensitive data to the cloud, so it becomes important to have a unified security model. To reduce your risk of data exposure and unauthorized access, you need to have consistent security and governance controls across on-premises and the cloud that allow you to apply fine-grained security policies with full end-to-end data provenance and lineage as well as an audit trail to track who has accessed the data. After all, in the case of unauthorized access, you need to be able to immediately identify which data assets were exposed, what information was potentially leaked, and identify the perpetrators by analyzing a system-wide access audit log.

#### 5. Manage costs and optimize resources

Understanding cloud service costs can be daunting considering that every service uses a different pricing model consisting of multiple components. For instance, some services are priced based on consumption of compute resources for virtual machines and storage capacity. Depending on the type of storage medium used, storage can be priced based on the actual amount of space consumed or on provisioned storage regardless of actual consumption. Other services may be priced based on data being scanned on a per query basis, number of API requests, number of bytes transferred over a network, and many other factors that are often difficult to estimate up front. If your data and analytics environment is sitting idle or if your jobs, queries, and pipelines aren't properly designed and tuned, you might, in fact, be paying significant overhead due to improper consumption of cloud resources while getting very little use or business value from that consumption.

Your data pipelines may incorporate multiple steps, with each step requiring a different technology. For example, an end-to-end data processing pipeline may require tools such as Apache Kafka for streaming data, Apache NiFi for managing data flows, Apache Spark for data science and machine learning (ML), Apache Hadoop for storing the data using a Hadoop distributed file system (HDFS), and Apache Hive for running SQL queries to answer business intelligence questions from your data.

Implementing this entire pipeline with disparate cloud-native services is challenging from both operational and budgetary perspectives. While some cloud vendors may offer native services that provide similar functionality, there are three major challenges:

1. Cloud providers are mostly packagers of open source software and typically lack the open source talent needed to support your data workloads in production let alone contribute fixes and updates back into the community

## About Cloudera

At Cloudera, we believe that data can make what is impossible today, possible tomorrow. We empower people to transform complex data into clear and actionable insights. Cloudera delivers an enterprise data cloud for any data, anywhere, from the Edge to AI. Powered by the relentless innovation of the open source community, Cloudera advances digital transformation for the world's largest enterprises.

[Learn more at cloudera.com](https://www.cloudera.com)

## Connect with Cloudera

About Cloudera: [cloudera.com/more/about.html](https://www.cloudera.com/more/about.html)

Read our VISION blog: [vision.cloudera.com/](https://www.cloudera.com/vision) and Engineering blog: [blog.cloudera.com/](https://www.cloudera.com/blog)

Follow us on Twitter: [twitter.com/cloudera](https://twitter.com/cloudera)

Visit us on Facebook: [facebook.com/cloudera](https://www.facebook.com/cloudera)

See us on YouTube: [youtube.com/user/clouderahadoop](https://www.youtube.com/user/clouderahadoop)

Join the Cloudera Community: [community.cloudera.com](https://www.cloudera.com/community)

Read about our customers' successes: [cloudera.com/more/customers.html](https://www.cloudera.com/more/customers.html)

2. It's up to you to orchestrate and integrate multiple cloud-native services into an end-to-end data processing pipeline, without any enterprise-level support from the vendor
3. Cloud vendor billing models for each service may be completely different, making the overall cost nigh on impossible to estimate up front

When moving data analytics workloads to the cloud, it's important to understand the key resource metrics for running your workload and the associated cloud billing model and service costs. One of the key cost control strategies is to keep long-running workloads on-premises while moving ephemeral workloads that need resource elasticity to the cloud. Having the right operational controls in place for optimizing use of cloud resources for your analytics workloads can save you significant expense.

## Next Steps

Today's data architectures need to support hybrid cloud environments, but not as an afterthought. Rather, your data strategy should drive the cloud strategy so they are aligned and reinforce one another. Combining hybrid cloud with diverse analytic capabilities on a single platform is also known as an enterprise data cloud. An enterprise data cloud represents a new data management architecture that includes:

- \_ Multi-function analytics including data processing and analysis at the edge to data warehousing, real-time operations and machine learning
- \_ Unified security, governance, metadata, and control to drive your enterprise data strategy
- \_ Open frameworks for integration with third-party tools
- \_ And the flexibility to deploy use cases on public, private, and hybrid clouds leveraging any type of data wherever that data lives.

For more information and a framework for ensuring that happens, download "[Power your Business Transformation with an Enterprise Data Cloud.](#)"<sup>3</sup>

## Sources

<sup>1</sup> [https://resources.idg.com/download/executive-summary/cloud-computing-2018?utm\\_campaign=Cloud%20Computing%20Survey%202018&utm\\_source=Forbes%20-%20Louis](https://resources.idg.com/download/executive-summary/cloud-computing-2018?utm_campaign=Cloud%20Computing%20Survey%202018&utm_source=Forbes%20-%20Louis)

<sup>2</sup> <https://www.edq.com/globalassets/white-papers/2018-global-data-management-benchmark-report.pdf>

<sup>3</sup> <https://www.cloudera.com/content/dam/www/marketing/resources/whitepapers/power-your-business-transformation-with-edc.pdf.landing.html>