CLOUDERA

WHITEPAPER

Solving the Data Disconnect

Building a consistent data architecture for analytics and AI

As AI and data demands accelerate, many organizations are discovering that their data is harder to use than it should be.

Hybrid IT infrastructures, once valued for flexibility, have created data silos that limit AI potential. Data unification is no longer optional.

Discover how an open, cloud-native data architecture enables access, governance, and analytics services, so you can bring AI to your data anywhere it lives.

CLOUDERA

Executive summary

As organizations look to turn analytics and AI into a real business advantage, most find themselves operating in hybrid IT environments that combine on-premises infrastructure with public and private clouds. While this reality is now common, it introduces a new challenge: data is distributed, but analytics and AI require access to it as a whole.

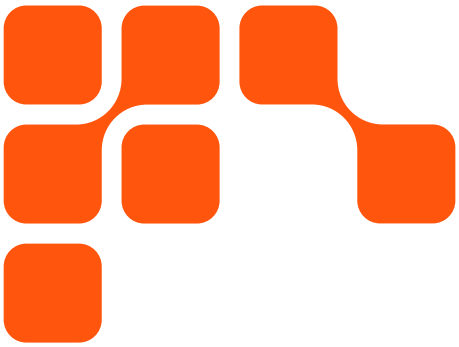
Simply operating across multiple environments isn't enough. Without consistent access, governance, and integration, enterprise data remains disconnected. This makes it difficult to apply AI at scale, manage risk, or move from experimentation to production.

In this white paper, you'll learn how organizations are addressing this disconnect with a common approach to access, governance, and data services across the full data estate—clouds, data centers, and edge environments. You'll see what it takes to support analytics and AI where they deliver the most value, while maintaining control and flexibility.

Key takeaways:

- See what sets a connected data architecture across hybrid environments apart from fragmented hybrid models, and why this distinction is critical for scaling AI
- Learn what it takes to support a modern, scalable data architecture that connects on-premises and multi-cloud systems
- Identify top priorities for successful implementation, including governance, integration, and operational efficiency
- Gain actionable insights for every stage of deployment, from early modernization to optimizing mature environments

Table of contents



Introduction

Cloud infrastructure marked a major turning point for IT, delivering unprecedented scalability and speed while transforming everything from software development to data security and analytics access.

Its top advantages include:

- Separation of compute and storage, so organizations can scale resources based on demand
- Managed infrastructure that reduces operational overhead, freeing teams to focus on innovation
- Access to a rich ecosystem of tools and services that accelerate insight and enable new business models

Yet, for all its benefits, the public cloud isn't a one-size-fits-all solution. Many organizations have learned—sometimes the hard way—that certain workloads simply don't belong there. Whether due to data sovereignty requirements, latency-sensitive applications, or regulatory constraints, some data must stay on premises.

As a result, most enterprises now operate across a mix of environments. This shift often happened reactively rather than strategically, resulting in patchwork architectures and siloed data.

As organizations move beyond traditional analytics toward AI, these silos become a larger obstacle. AI depends on access to data at scale. When data is spread across different systems, teams have to move it, duplicate it, or limit AI initiatives to the subset of what's available.

To move forward, organizations need a way to apply analytics and AI to their data wherever it lives, without increasing operational complexity or compromising governance.

The hidden costs of cloud-only strategies

For many organizations, migrating to the cloud has become a business imperative. The cloud offers clear advantages like greater agility and scalability, but how you get there matters just as much as the destination.

Too often, cloud migrations take longer, cost more, and deliver less than expected. Common migration approaches carry hidden trade-offs that become more pronounced as analytics and AI workloads grow.

These trade-offs are amplified for AI workloads. Moving large volumes of data introduces latency, egress costs, and governance challenges. In many cases, the most valuable data can't be moved without introducing unacceptable risk.

Organizations need an alternative that allows them to modernize incrementally, applying analytics and AI where data already resides while retaining the flexibility to adapt over time.

Approach	Lift-and-Shift	SaaS / PaaS Replacement
What It Is	Move existing systems to the cloud with minimal changes	Replace systems with managed services
Perceived Benefit	Simple, fast transition	Access to modern cloud-native features
Common Challenges	Operational mismatches, cloud-specific nuances, skills gaps	Loss of control, platform lock-in, data duplication
Cost Implications	Delays increase cost over time	Recurring compute and storage costs can grow quickly
Flexibility	Rigid and difficult to optimize post-migration	Usually a one-way move that's hard to reverse later
Risk Level	High risk of underperformance due to environment mismatch	High risk of vendor dependency and long-term inflexibility

Cloud power meets on-premises efficiency

Operating across environments doesn't have to mean operating inconsistently. When organizations establish shared access controls and governance policies across cloud and on-premises systems, they gain the ability to work with data the same way, regardless of location.

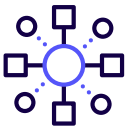
This enables several critical capabilities:

- **Consistent access and governance**
  Teams use the same security policies and controls everywhere, simplifying compliance and reducing risk.
- **Cloud-like capabilities on-premises**
  Features like separate compute and storage, elastic scaling, and self-service access extend cloud efficiency to on-premises systems.
- **Support for open data formats**
  Open standards reduce lock-in and make it easier to share and analyze data across tools and platforms.
- **Portability without rework**
  Workloads can shift between environments as requirements change, without rewriting applications or disrupting operations.

These capabilities allow organizations to apply analytics and AI across their full data estate instead of being limited by wherever data happens to be located.

Faster time to value

Teams can start using cloud-native services like object storage and open table formats right away. No need to wait for a full migration.



Reduced risk

Organizations can keep things running smoothly and stay in control by having continuous access to data before, during, and after the transition.

Skills and readiness

Employees can build confidence and cloud skills in a familiar setup, making adoption easier and reducing the usual steep learning curve.

Flexibility for the future

A platform that enables portability across environments helps avoid vendor lock-in and rework—now and later down the road.

Consolidating systems on an open, portable data platform provides a practical, low-risk approach to modernization. Organizations can adopt cloud capabilities incrementally, maintain control over their data, and give teams the flexibility to evolve architectures as requirements change—without forcing disruptive migrations.

Flexible deployment powers complex workflows

Analytics and AI workflows rarely run in a single location. Different stages of a workload have different requirements for performance, cost, and scalability. This is especially true for complex workflows like machine learning (ML).

Cloudera helps teams run each stage—on-premises, across clouds, or at the edge—wherever it fits best. That way, teams get the speed and control they need to innovate without losing visibility or driving up costs.

ML workflow in action

Consider an ML pipeline:

Data preparation	Often best handled close to existing pipelines, where infrastructure is predictable and pipelines are already in place
Model development and training	Compute-heavy and happens in bursts, making dynamic cloud resources a better fit
Inference and real-time processing	Frequently needs to run close to operational systems to reduce latency and support faster decisions

Cost efficiency

Avoid overbuilding on-premises infrastructure by scaling compute in the cloud when needed. Steady workloads stay on-premises, while variable ones use the cloud's flexibility, making better use of resources and cutting overall costs.

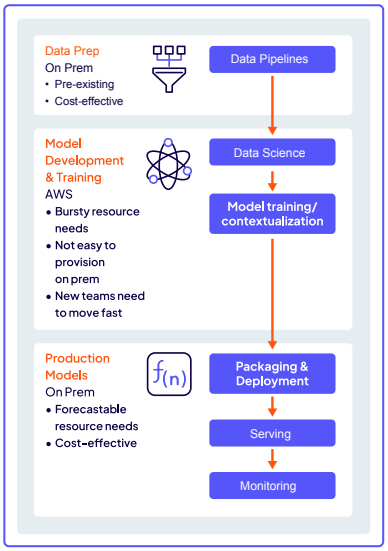
Performance optimization

Running each part of a workflow on the right infrastructure helps boost performance and cut down on delays. It makes processes like model training more efficient, speeds up insights, and gets the most from both cloud and on-premises resources.

Operational flexibility

Teams adjust resources dynamically based on workload needs. They can scale up in the cloud during busy times or move heavy processing where it makes sense—keeping performance steady and avoiding disruptions.

This flexibility is essential as AI initiatives move from experimentation into production.



CLOUDERA

Continuous optimal deployment for hybrid data

As data volumes grow and AI use cases evolve, static deployment decisions quickly become limiting. Organizations need the ability to adjust where workloads run as performance needs, cost pressures, and regulatory requirements change.

With a common operating layer, teams can build workflows once and deploy them where they deliver the most value.

This enables:

- **Ongoing performance optimization**
  Workloads remain aligned with the most efficient infrastructure.
- **Adaptability without disruption**
  Changes in regulation, data growth, or AI requirements do not require refactoring or downtime.
- **Operational resilience**
  Resources can be reallocated to maintain continuity and meet new demands.

As AI systems become more dynamic, supporting real-time inference and continuous improvement, this adaptability becomes a functional requirement.

Built for what's now—and what's next

Modern analytics and AI depend on access to enterprise data in all its forms and locations. By enabling secure access to data wherever it resides, organizations can run AI and analytics where they deliver the most value, without unnecessary movement or duplication. This reduces risk, controls cost, and accelerates the transition from insight to action.

Whether supporting advanced analytics today or preparing for future AI initiatives, the ability to work with 100% of your data across clouds, data centers, and the edge delivers a lasting competitive advantage.

The business case for Cloudera

By bridging cloud and on-premises environments with a single, cloud-native platform, Cloudera helps organizations innovate faster, stay secure, and maximize ROI from data and AI investments.

$100M+
annual revenue gains through real-time ML

95%
boost in fraud detection accuracy

75%
fewer repetitive tasks with connected data systems

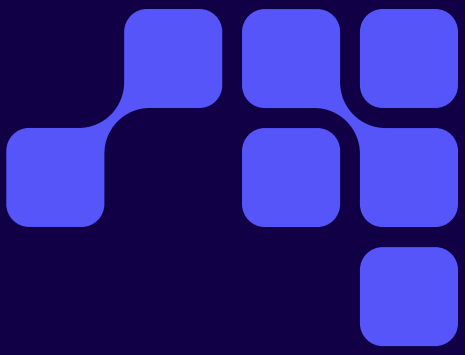
Ready to grow, secure, and connect your data?

We'd love to hear from you.

Learn more

Get in touch

About Cloudera



Cloudera is the only data and AI platform company that large organizations trust to bring AI to their data anywhere it lives. Unlike other providers, Cloudera delivers a consistent cloud experience that converges public clouds, data centers, and the edge, leveraging a proven open-source foundation. As the pioneer in big data, Cloudera empowers businesses to apply AI and assert control over 100% of their data, in all forms, delivering unified security, governance, and real-time and predictive insights. The world's largest organizations across all industries rely on Cloudera to transform decision-making and ultimately boost bottom lines, safeguard against threats, and save lives.

To learn more, visit Cloudera.com and follow us on LinkedIn and X.

CLOUDERA

Cloudera, Inc. | 6220 America, Center Drive, 5th Floor, San Jose, California 95002 USA | cloudera.com