



Think you know hybrid?

A fresh take on cloud strategy

White
paper

Many organizations have taken a hybrid approach to cloud without full consideration of all the associated elements, including data, infrastructure, tools, and more.

If recent events have taught us anything, it's that businesses must be prepared for everything.

In 2020 a global pandemic shuttered businesses around the world, forced

employees to work from home offices, and accelerated cloud transitions as enterprises sought to make data and services accessible to a suddenly distributed workforce.

CIO

SPONSORED BY

CLOUDERA

In 2023 generative artificial intelligence (GenAI) burst onto the scene, capturing the imagination of business leaders everywhere and spurring a widespread redeployment of budgets to AI projects. Within two years, nine out of 10 global CEOs ranked AI as the most critical technology for future profitability and competitiveness.¹

More innovations are on the horizon. Order-of-magnitude performance improvements enabled by quantum computing draw closer to reality with each scientific breakthrough. Agentic AI promises to automate

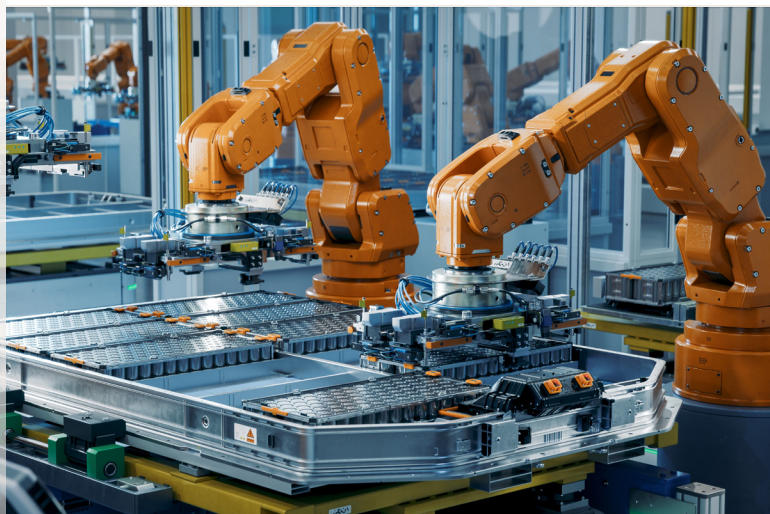
complex workflows that have long required humans. Robots are taking over mundane and dangerous tasks in warehouses and factories.

These and other developments will require organizations to pivot their IT infrastructure to adapt quickly. Inflexible legacy platforms will become a competitive liability.

These scenarios have one other thing in common: They require large amounts of data to deliver on their promise. Annual data volumes are expected to more than double between 2025 and 2028.² The ability to scale and manage rapidly growing data volumes is essential, as is the capacity to mine insights from it.

A compounding data challenge

Consistent data management across multiple platforms is a major challenge for many organizations. Isolated data silos spanning many



1 [Future-Ready Innovation: Strategies for 2025 and Beyond](#), NTT Data, January 16, 2025

2 [Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2023, with forecasts from 2024 to 2028](#), Statista, November 21, 2024

different data types and formats hamper unification efforts. Complexity increases as the quantity and variety of data types proliferate.

Legacy data platforms were designed to manage data in structured forms, such as business names, telephone numbers, account codes, and billing amounts. But unstructured and semistructured data – including messages, text documents, video, images, sensor readings, and streaming data – now comprises about 90% of the information that enterprises generate, and the volume is growing.³ These data types have enormous strategic value yet can't be neatly categorized into structured data models. They require different storage mechanisms on-premises and in the cloud.

Regulations and business demands add to this complexity, by increasingly dictating where data can be stored and for how long, limiting organizational flexibility. As a result, data at many organizations is now spread across transactional databases, cloud data

stores, local and cloud data lakes, and software-as-a-service (SaaS) applications. No one has a clear picture of what data the organization owns and where it's located.

The challenges go beyond managing infrastructure. Data analytics is essential to gaining insights about the business, but different analytical tools are required for each infrastructure variation the organization uses. The need for multiple administrative and analytical skills creates cost overhead and duplication of effort. What's more, the experience of using a tool on one infrastructure usually does not translate to another infrastructure, creating waste and the need for additional, often duplicative, user training.

The capabilities of a true hybrid option

These factors combine to make a compelling case for organizations to adopt a true hybrid platform for data, analytics, and AI.

A true hybrid platform:

- **Supports smooth data movement; a single set of tools; and consistent security and governance across on-premises, cloud, and hybrid cloud environments**
- **Enables data, analytics, and AI to move seamlessly across a consistent data layer to accommodate change and support the most optimal deployment scenarios**
- **Provides a consistent experience regardless of the underlying infrastructure**
- **Supports many popular tools and is based on open standards**
- **Offers unified data ingestion and integration from diverse sources, large-scale data transformation and normalization, automated feature extraction, real-time processing, consistent governance, and compatibility with popular machine learning frameworks**

Another significant benefit of a true hybrid platform: It uses cloud-native constructs such as separation of computing and storage, support for portable software containers, extreme scalability, and extension via application program interfaces (APIs). Streaming data ingestion – a market that Fortune Business Insights expects to grow 26% annually through 2032⁴ – addresses the urgent need for organizations to understand how their businesses operate in real time.

The benefits of a true hybrid platform

A wide range of business imperatives can be met with this solution:

Choice. It frees organizations from vendor lock-in, by enabling them to choose where and how they deploy infrastructure, data, and workloads. An adaptable platform based on open standards that supports the most popular administrative and analytical tools future-proofs them against technology changes while enabling them to integrate innovations smoothly.

Cost reduction. Managing multiple incompatible systems, tools, and applications for storing and analyzing data requires specialized skills and tooling. A true hybrid platform reduces administration overhead, saves on software licensing fees, and enables the same tools to be used across an organization's data landscape. It creates the foundation for the contextualization of generative, multimodal, and agentic AI models. Operational data from deployed models is fed back into the system to detect and correct drift and bias.

For example, Asian bank DBS achieved an 80% reduction in operating costs through customer service and operational improvements. Adopting a hybrid data architecture enabled the company to scale more economically and manage larger data volumes. Risk managers can now better detect fraudulent transactions, and operational staff can predict customer flow, ATM load, and call



center volumes with greater precision, resulting in fewer losses and lower costs. DBS has deployed more than 800 AI models across 350 use cases and expects their economic impact to exceed \$1 billion.⁵

Improved performance. True hybrid infrastructure enables resources to be allocated for optimum efficiency. For example, on-premises infrastructure can be used for predictable workloads such as transaction processing, and workloads can burst into the cloud during peak activity periods. Dynamic workload balancing optimizes location and performance. Hot data can be kept in low-latency storage for fast access while cold data is offloaded to cost-effective cloud storage.



White paper

Workloads can be shifted as internal needs and business conditions change. For example, data scientists training AI models need a lot of data and processing capacity for short periods. Allocating permanent infrastructure to serve those needs is costly and wasteful. A true hybrid platform enables resources to be flexibly allocated to meet those short-term needs and then refactored for other uses. Observability-guided workload management powered by machine learning can optimize resources and their costs.

Greater business agility. The need to respond quickly to changing business conditions has never been more pronounced. Having a consistent

way of dealing with all data stores and the workloads that operate on them across all types of infrastructure simplifies IT's job and gives the business better access to data and analytics. More use cases can be implemented in less time to deliver faster value.

Healthcare analytics firm [IQVIA](#) consolidated data spread across 250 data

warehouses to give users a high-performance self-service analytics platform for developing predictive algorithms. As a result, IQVIA quadrupled its ability to predict if a patient is eligible for a clinical trial before becoming symptomatic.

Enhanced security and governance.

As data becomes more distributed, security risks and complexity multiply. Administering multiple access policies across on-premises and cloud environments creates significant management overhead. This complexity is further compounded in multinational organizations, where data sovereignty laws impose strict regulations on how data is stored,

processed, and transferred across borders. Managing compliance on a region-by-region basis can be an administrative nightmare.

A true hybrid platform provides a single set of governance tools that encompass lineage, auditability, and discoverability of data assets. Data can be located and governed by applicable regional and industry regulations.

A multinational oil and gas [company](#) used Cloudera on Amazon Web Services (AWS) with a Cloudera data warehouse and data engineering tools to automate streaming data ingestion from 130,000 PCs around the globe using Apache Kafka. The company balanced the processing of on-premises infrastructure with AWS Cloud. The project reduced cybersecurity threat detection time by 90%, cut log volume by 60%, and saved \$2 million in licensing costs over five years.

Key considerations

When considering a unified platform for data, analytics, and AI, consider these factors:

- **A distributed cloud model spreads services across data centers, clouds, and edge computing sites operating on a single platform.**
- **The experience is consistent across all types of infrastructure.**
- **Administration and operational control are centralized, even when cloud services are in different locations.**
- **Data, analytics, AI models, and workloads can be moved between on-premises data centers and clouds without refactoring, helping lower the total cost of ownership.**
- **The platform supports each stage of the data life cycle, including ingestion, transformation, warehousing, and machine learning.**
- **A unified and centralized platform enables organizations to apply and enforce one set of policies and access controls spanning metadata management, encryption, control, and governance across all environments.**

Be ready for whatever comes next

Today's rapidly shifting business and technology landscape allows little leeway for long-term planning. A more practical strategy is for businesses to leave their options open.

What we do know is that data volumes and types will continue to proliferate. Organizations can't afford to tie themselves to proprietary dead ends. Flexibility may be the only sustainable source of competitive advantage.

A true hybrid platform provides a seamless, scalable solution that enables unified data ingestion, transformation, governance, and analytics across on-premises and cloud environments. This yields cost-saving and performance benefits and, more importantly, enhances organizational agility without sacrificing security and compliance.

Explore a true hybrid option. | Visit [Cloudera](#) for more information.