A strong data culture, unified open data platforms, and robust governance are essential foundations for successful, scalable AI adoption and innovation in modern organizations.

The Imperatives of AI-Ready Enterprises: Data Culture, Data Readiness, and a Unified Platform

July 2025

Written by: Marlanna Bozicevich, Research Analyst, Data, Analytics, AI, and Automation

Introduction

The rapid ascent of AI has fundamentally reshaped the priorities and strategies of organizations across every sector. The pace of AI advancement has accelerated so much that organizations must adapt with unprecedented speed or risk obsolescence. Yet, simply adopting AI technologies is insufficient. Organizations must deploy AI with rigor and discipline, ensuring that investments yield measurable returns and that the delivered value exceeds the costs of implementation and operation.

Central to the success of AI initiatives is robust data management. Data quality, accessibility, and trustworthiness have become board-level concerns because organizations recognize these as prerequisites for effective and responsible AI deployment. According to IDC's February 2025 *Future Enterprise Resiliency and Spending Survey,* data management was the second most important issue to address in an organization's AI strategy, closely behind responsible AI and ethics.

AT A GLANCE

KEY STATS

Data management is a top issue to address for organizations prioritizing AI strategy in 2025. In detail:

» 52% of organizations investing in AI data readiness indicated that data intelligence was the most important focus area, 14% above any other choice.

» 47% of analytic data repositories remain stored only on premises or in a private cloud.

Establishing Data Culture

This heightened focus on data management reflects a broader realization: The foundation of a successful AI and data strategy is not merely technological but also cultural. Historically, data was primarily managed by IT departments. However, as AI becomes ubiquitous and every team begins owning and managing its own data for AI applications and agents, establishing a strong data culture becomes paramount for organizational success. This cultural shift is especially critical in the era of large language models (LLMs), which draw not only from curated, structured data but also from unstructured information across the company.

Organizations must move from a technology-first mindset to a culture-first approach, prioritizing business goals, cross-functional collaboration, and continuous experimentation and learning. A robust data culture embeds data-driven decision-making at every level, enabling teams to understand their data, challenge assumptions, remove biases, resolve issues, and fill context gaps when building trustworthy AI solutions. It also ensures self-service access, intuitive interactions, real-time agility, and a predictive edge. Ultimately, data culture should drive technology decisions, not the other way around.

The business case for investing in data culture is clear. According to IDC's *Office of the CDO Survey,* the top outcomes from data initiatives include increased security and privacy (52%), higher-quality analytics (49%), greater data availability for AI (41%), regulatory compliance (35%), and improved productivity (35%). Organizationwide investment in data processes and best practices promotes data culture and yields numerous advantages. These benefits make data culture a strategic imperative, not a luxury. However, achieving these gains requires overcoming human and organizational barriers. IDC data shows top concerns among leadership for data and analytics teams to include the impacts of generative AI (GenAI), insight tracking and visualization, skills training, and change management. As GenAI automates tasks and introduces new capabilities, teams must adapt their skills and workflows. This underscores the need for a safe environment that encourages experimentation, learning from failure, and challenging the existing system.

Simply providing tools is no longer enough. A consistent data culture determines the architecture, tools, and governance needed for success. This requires committed leadership: Executives must not only endorse but also visibly use data and AI, communicate their strategic value, and foster a culture of experimentation and accountability. Effective leaders champion resources, manage risks, and ensure data-driven practices become ingrained habits.
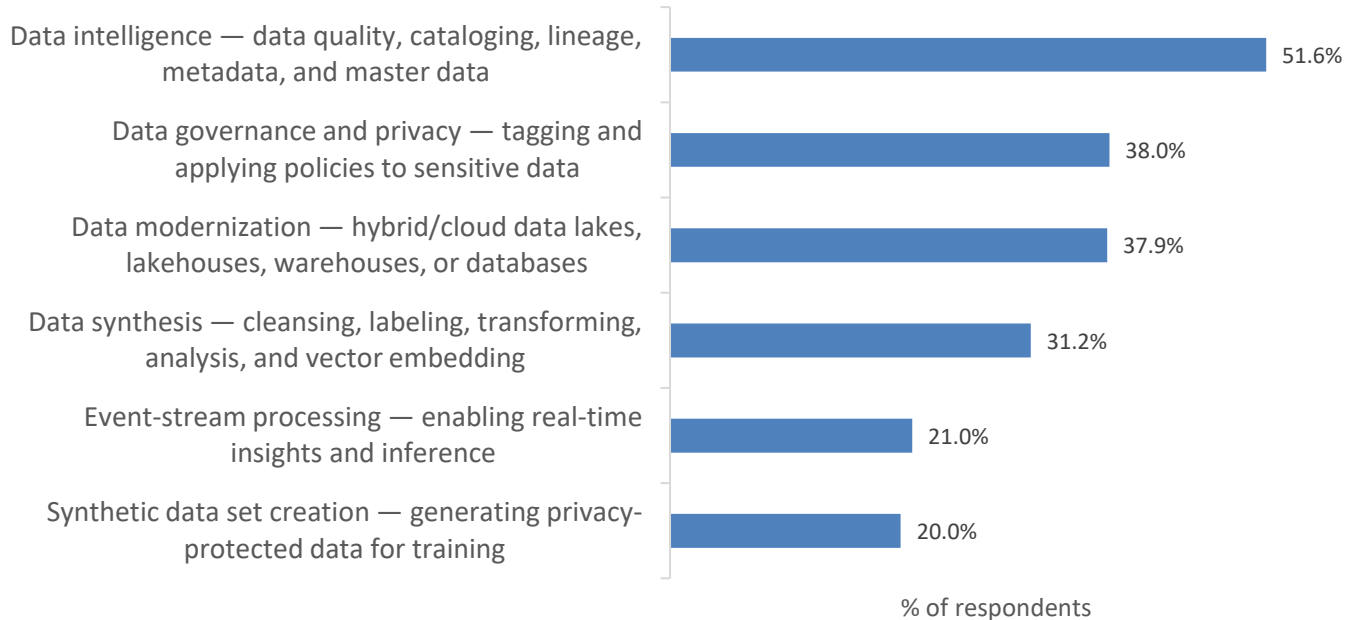
As AI becomes central to decision-making, organizations must also address ethical and transparency concerns, balancing innovation with compliance and risk mitigation. The rise of AI agents and self-service analytics makes data culture even more critical, ensuring every team can manage and leverage its data effectively for AI applications, learn from past mistakes, and align with business objectives.

AI Data Readiness

As organizations navigate these complex ethical landscapes and strive to foster a responsible data culture, it's crucial to also emphasize data readiness as another foundational element to successful AI implementation (see Figure 1). Data readiness means going beyond the work of data quality teams, implementing data preparation practices across data originators, owners, and consumers. Ensuring that data is accurate, accessible, and well governed must be a priority across the organization to support transparent and compliant AI-driven decision-making.

FIGURE 1: *The Most Important Focus Areas to Address in AI Data Readiness*

Q *You indicated that your organization is prioritizing AI data readiness in 2025. What are your organization's top 2 most important data-related areas of focus and investment?*

Focus Area	% of respondents
Data intelligence — data quality, cataloging, lineage, metadata, and master data	51.6%
Data governance and privacy — tagging and applying policies to sensitive data	38.0%
Data modernization — hybrid/cloud data lakes, lakehouses, warehouses, or databases	37.9%
Data synthesis — cleansing, labeling, transforming, analysis, and vector embedding	31.2%
Event-stream processing — enabling real-time insights and inference	21.0%
Synthetic data set creation — generating privacy-protected data for training	20.0%

% of respondents

n = 353

Source: IDC's Future Enterprise Resiliency and Spending Survey, Wave 1, February 2025

When preparing data for AI, organizations prioritize data intelligence, governance/privacy, and modernization. As seen in Figure 1, data intelligence is the leading focus (14% higher than any other area), underscoring the critical role of metadata in GenAI. Metadata provides the context, structure, and lineage that AI models need to interpret data accurately and produce relevant, trustworthy outputs. Without it providing underlying definitions, sourcing, and data relationships, AI systems risk misinterpretation and unreliable results.

These data priorities reflect a growing understanding that successful AI initiatives depend on trusted, well-managed, and accessible data. Data intelligence unlocks insights from complex data sets, governance ensures privacy and regulatory compliance, and modernization leverages advanced cloud-based technologies to manage and scale data for evolving business and AI needs.

Today's data landscape is distributed, dynamic, diverse, and often "dark" or unexplored. Overcoming these challenges requires frameworks that break down silos, improve accessibility, and foster a strong data culture. Investing in these areas creates an environment that is secure, compliant, agile, and ready for innovation, ensuring AI delivers meaningful business value.

To address these needs, the data management software market is moving toward unified platforms that cover the entire data life cycle, from storage and integration to intelligence and access. These solutions help organizations unify and govern data across hybrid cloud and multicloud environments, supporting AI readiness, compliance, and ongoing

innovation. Ultimately, realizing AI's full potential depends on the synergy between technology, data culture, and leadership.

Definitions

» **Data intelligence** leverages business, technical, relational, and operational metadata to provide transparency of data profiles, classification, quality, location, lineage, and context, enabling people, processes, and technology with trustworthy and reliable data. Data intelligence software collects, organizes, classifies, and maintains this collection of metadata that constitutes data intelligence. Data intelligence represents one of the key usage patterns that provide organizations with answers to the five Ws of data: who, what, when, where, and why, plus how. Data intelligence has also added an R to the set of questions, understanding how data is related.

» **Data platform software** is a comprehensive solution to manage the entire data life cycle, from collection to analysis, providing a unified environment for handling vast amounts of diverse data. Comprising five foundational layers — dedicated data storage, streaming and batch data ingestion, data transformation, data intelligence, and user access — it offers scalable storage solutions, centralizes data from various sources, prepares it for analysis, ensures quality and governance, and supports powerful querying and analysis tools. Some organizations may also refer to this type of product as a data fabric.

Benefits

Unified, metadata-driven data management platforms provide organizations with essential advantages in the AI age by delivering rich contextual information (such as data source, lineage, quality, and usage) that enhances data understanding and trust. These frameworks enable users and agents to efficiently locate, interpret, and analyze data, accelerating innovation and supporting faster, more accurate decision-making. As AI and automation advance, metadata serves as the foundation for intelligent data management, guiding AI agents and ensuring trusted data. By offering detailed context, automated controls, and robust lineage tracking, unified data intelligence frameworks help protect sensitive information, enforce regulatory compliance, and minimize the risk of bias or error in AI outcomes.

The move toward an open data architecture further enhances these benefits. Open platforms and table formats, such as Apache Iceberg, minimize vendor lock-in and enable organizations to adopt plug-and-play analytics and AI engines. This fosters innovation by allowing decentralized self-service analytics while maintaining centralized control and governance. Such flexibility is vital for organizations aiming to balance agility with oversight.

Data platforms also enable the creation, maintenance, and consumption of data products. Data products are packaged data and AI assets that connect data producers with consumers, allowing both humans and AI agents to seamlessly discover, access, and interact with trusted data sources. By facilitating communication and collaboration between agents through shared data products, organizations foster a dynamic ecosystem where insights are exchanged, workflows are triggered, and data is continuously enriched. This interconnected approach simplifies data use across personas, drives data culture, and provides the curated, protected, and integrated assets necessary to fuel analytics and AI initiatives. By making data products accessible and ensuring proper ownership and business value, companies can support scalable AI initiatives and maximize the return on their data investments, all by leveraging the benefits of a unified data platform and open architectures.

Trends

Several major trends are reshaping data management and AI, with significant impacts for both users and technology providers. One prominent trend is the increasing emphasis on data control, governance, and compliance, especially as organizations democratize data access in regulated industries. The rise of the data control plane, which unifies governance, compliance, security, and visibility across distributed data estates, is increasingly vital as open table formats and ecosystems proliferate.

Hybrid and multicloud data estates are now the default architecture for modern enterprises. Currently, 53% of analytic data repositories are stored in public cloud, hybrid cloud, or multicloud environments while 47% remain on premises or in private clouds, according to IDC's *Office of the CDO Survey*. This distributed landscape is driven by regulatory demands, data sovereignty, cost optimization, and the need for flexible compute resources, all further complicated by macroeconomic and geopolitical factors. Managing distributed data and workloads across these varied environments requires unified oversight and seamless visibility.

To address this complexity, the data management market is converging on unified, AI-enabled platforms that streamline operations and reduce technical debt. Centralized, governed data estates enable rapid experimentation and AI deployment by providing a secure, well-managed foundation. Meanwhile, AI-driven automation is transforming integration, transformation, and quality assurance, accelerating time to insight and lowering operational overhead. This supports self-service analytics and decentralized access, empowering more users and fostering a data-driven culture.

As these trends accelerate, metadata's role as a strategic asset will only grow, guiding both people and AI through increasingly complex data landscapes.

Considering Cloudera

Cloudera has evolved from a Hadoop pioneer into a comprehensive hybrid data platform, enabling organizations to manage, analyze, and operationalize data and AI across any cloud or on-premises environment. Formed through the 2017 merger with Hortonworks, Cloudera offers a unified platform that spans the entire data life cycle, from edge ingestion to AI-driven insights, addressing the growing complexity of enterprise data needs. Its true hybrid platform approach, launched in 2019 and refined by 2022, supports seamless workload portability and unified governance across legacy datacenters and major public clouds, reflecting Cloudera's commitment to innovation and flexibility.

The Cloudera platform's architecture centers around three core pillars, as shown in Figure 2.

FIGURE 2: *Cloudera at a Glance*
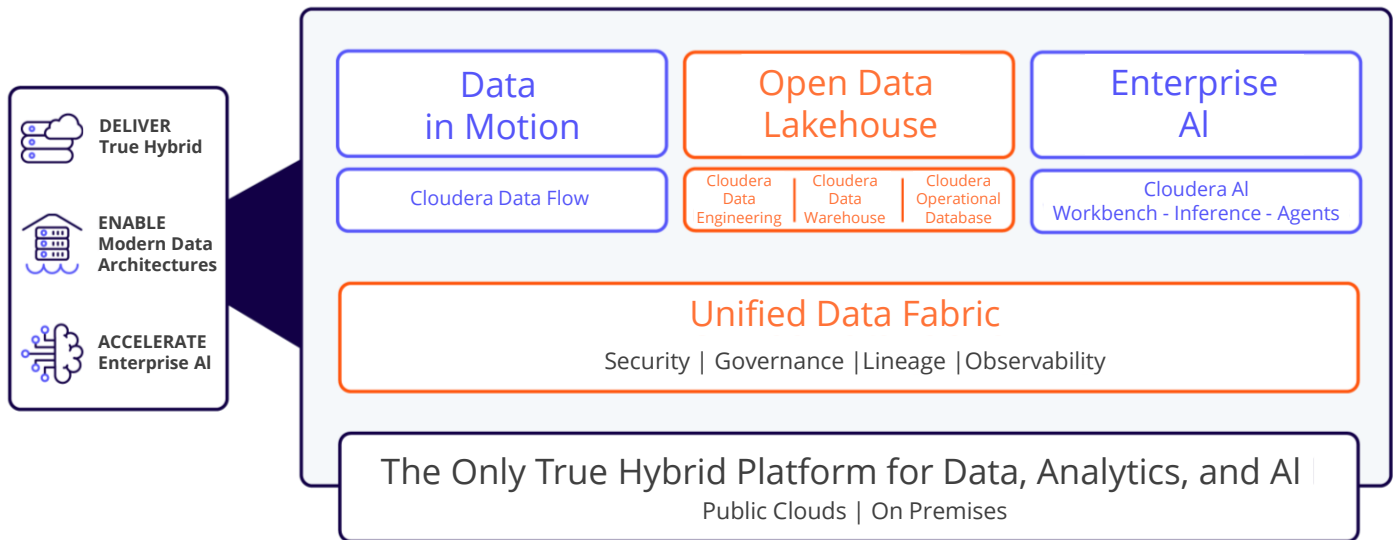
Source: Cloudera, 2025

At the heart of the platform is Cloudera's data-in-motion architecture, which integrates technologies such as Apache NiFi, Apache Flink, and Apache Kafka via Cloudera DataFlow. This enables real-time data ingestion, processing, and analytics both at the edge and in the cloud, supporting use cases such as fraud detection, IoT analytics, and customer retention. An example of this is delivering the ability to run AI models at the edge using MiNiFi to analyze and detect cyberthreats where they happen and take immediate actions to prevent them. This significantly reduces threat detection time, preventing business disruption and losses.

Apache Spark, Iceberg, and other open source engines anchor the company's open data lakehouse, unifying data engineering, warehousing, and operational analytics. Iceberg support provides a high-performance, open table format that enables partition evolution and time travel, allowing organizations to optimize analytics without costly data rewrites. Enterprise AI rounds out the architecture, offering a flexible environment for building and running AI applications (from high code to no code), integrating with partners such as NVIDIA and Crew AI, and supporting model development and inference at scale. For instance, the newly launched AI studios for RAG, fine-tuning, and building agents enable an organization to leverage LLMs, feed them enterprise data, and leverage them to build agents to optimize any task. The unified platform makes it easy to deploy these agents and use proprietary data safely.

A key differentiator is Cloudera's unified data fabric, which delivers security, governance, lineage, and observability across all data assets, regardless of location. The acquisition of Cloudera Octopai Data Lineage (formerly Octopai) has enhanced end-to-end data lineage, enabling the recovery of lost metadata and providing a comprehensive view of dataflows inside and outside Cloudera. Such a unified fabric is critical for organizations seeking to maintain compliance, ensure data quality, and enable responsible AI across increasingly complex data estates.

Cloudera's platform is purpose built for large enterprises and is trusted by many Global 2000 organizations in banking, insurance, telecommunications, manufacturing, and the public sector. Customers use Cloudera for mission-critical

workloads, such as real-time fraud detection, digital underwriting, network analytics, IoT edge analytics, and rapid healthcare data aggregation.

Strategically, Cloudera is focused on delivering a true hybrid experience, enabling modern data architectures, such as data lakehouses, data meshes, and data fabrics, and accelerating enterprise AI adoption. Cloudera offers the flexibility to seamlessly move workloads from a datacenter to the cloud and vice versa, adapting to the dynamic environment that is core to large-scale enterprises. Its open ecosystem approach, supporting a wide range of partners and technologies, ensures customers are not locked into proprietary formats or cloud providers.

Challenges

The platform's depth of capabilities may introduce operational complexity, requiring specialized skills for setup, management, and troubleshooting, especially in hybrid cloud and multicloud deployments. In addition, on-premises deployments, while offering control and security, can be expensive and complex to upgrade, with version control and workload management still areas for improvement. Service deletion may require substantial manual cleanup, adding to operational overheads. Cost is another consideration. While Cloudera's capabilities are robust, the platform can be more expensive than some alternatives, particularly for smaller organizations or those with limited budgets. Although features such as workload isolation and resource optimization help manage expenses, and some costs can be offset by the efficiencies gained from a unified integrated platform, cost remains a barrier for some. Finally, Cloudera operates in a fiercely competitive market, contending with cloud-native hyperscalers and established data platforms, all of which are rapidly advancing their own hybrid, AI, and data lakehouse offerings.

Conclusion

Data and AI strategy are heavily connected. The foundation for successful AI initiatives is a robust data management framework with a data-driven culture. Data culture and technologies must work synchronously to keep up with the latest AI innovations. Data platforms have become essential for organizations to unify today's modern data landscape and extract valuable insights in real time across the organization. Data culture will further enable data platform adoption, enhancing self-service data access, intuitive data interactions, and the agility organizations need across technical data personas and C-suite business executives.

> Data culture and technologies must work synchronously to keep up with the latest AI innovations.

About the Analyst

Marlanna Bozicevich, *Research Analyst, Data, Analytics, AI, and Automation*

Marlanna Bozicevich is a research analyst on the Data, Analytics, AI, and Automation team. Mrs. Bozicevich's research concentrates on data platforms, including data warehouses, lakes, and lakehouses, focusing on data storage and management. She examines core technologies for implementing analytic data platforms across various environments, data modeling techniques, and emerging open table structures. Her research closely aligns with the data control plane that helps govern and manage data, all in preparation for analytics that provide data-driven solutions to impact business operations.

MESSAGE FROM THE SPONSOR

To successfully navigate the complexities of AI adoption, data leaders need to prioritize robust data management. Cloudera empowers organizations to build a strong data foundation and culture so they can accelerate AI initiatives. Our hybrid data platform enhances security, governance, and lineage across diverse data assets, systems, and environments. Powered by technologies like Apache Spark and Iceberg, Cloudera centralizes and unifies data management, engineering, warehousing, and operational analytics. This enables organizations to ensure data quality, accessibility, and trustworthiness—prerequisites for effective and responsible AI deployment. For more information on Cloudera, visit our website at cloudera.com.

○ **IDC** Custom Solutions

The content in this paper was adapted from existing IDC research published on www.idc.com.