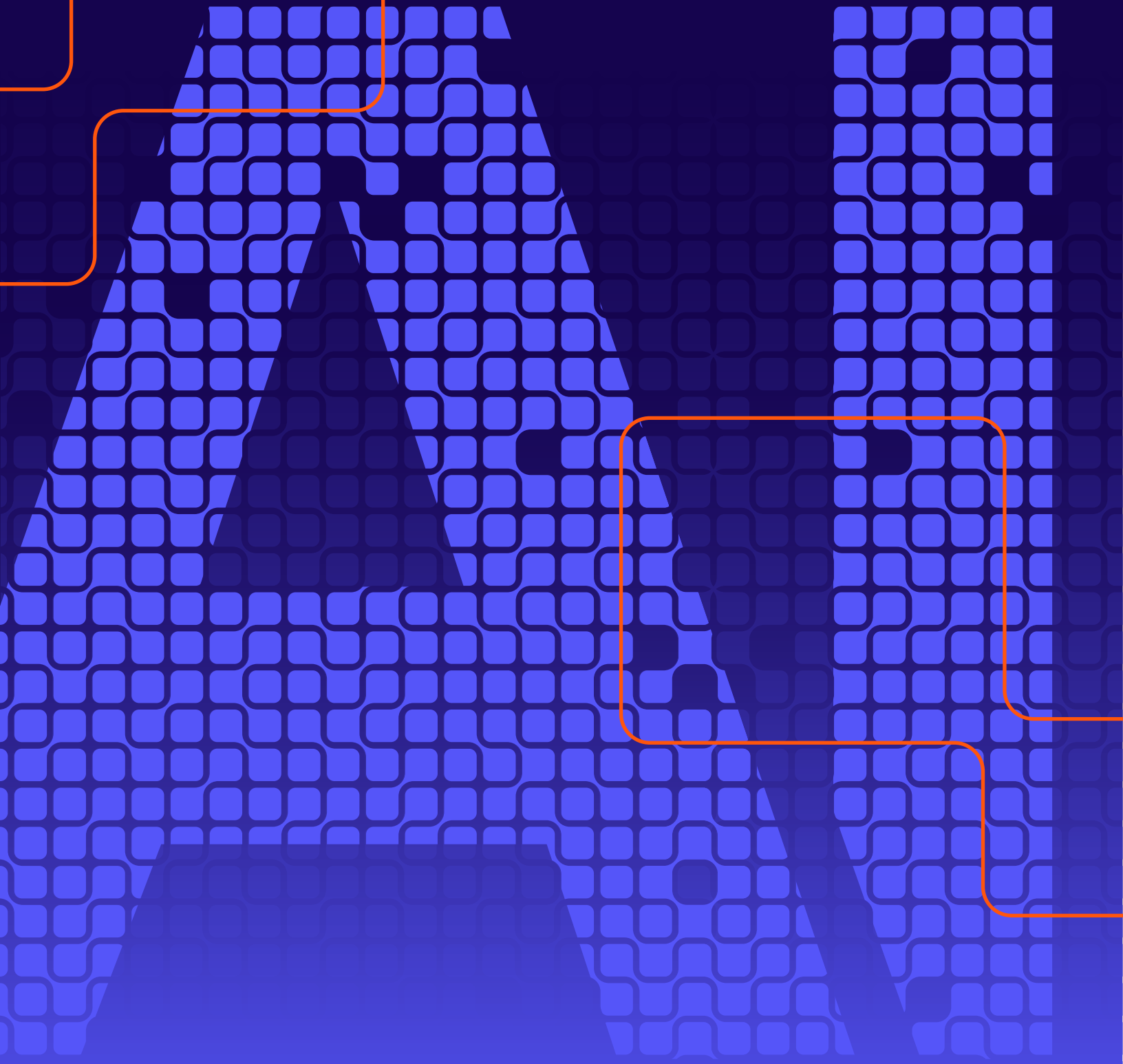
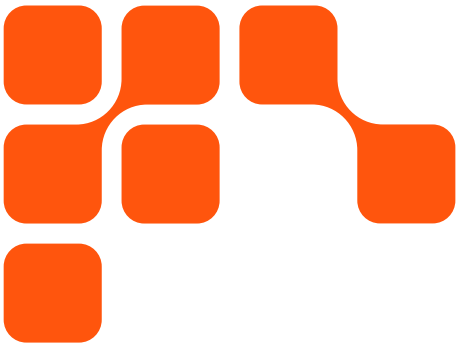
CLOUDERA

THE DATA READINESS INDEX 2026

Understanding the Foundations for Successful AI





Introduction

The last few years have seen the full weight of enterprise innovation thrown into artificial intelligence (AI) initiatives–from generative tools and large language models (LLMs) to autonomous agents absorbing critical workflows and processes. AI use is scaling fast, and it's redefining the modern enterprise in the process.

But as the acceleration continues, many organizations are still searching for impact, wondering why projects aren't generating the intended ROI. The issue isn't purely a matter of technical complexity or flawed algorithms. The actual challenge is a matter of data readiness.

Data readiness is a critical step towards realizing AI's potential. It is not just about *having* the data but being able to take advantage of it in its entirety, anywhere it resides, to unlock unique analytical insights and develop AI capabilities that accelerate strategic goals. Organizations that treat data readiness as a core competency, rather than a technical afterthought, will be better positioned to innovate responsibly, scale efficiently, and derive sustained business value from AI.

To better understand the foundational elements of data readiness, Cloudera, in partnership with third-party research firm Researchscape, surveyed over 1,200 IT leaders across the AMER, EMEA, and APAC regions, and industries including: energy/utilities, financial services, public sector, healthcare, manufacturing, software/technology, and telecommunications.

The survey revealed a significant contrast. Organizations are ready to embrace new frameworks to improve data readiness and infrastructure capabilities. Most are even aligned on approach, with 85% believing their data strategy is clearly defined. However, a vast majority (79%) of respondents said that their data-backed initiatives are hindered because they cannot access 100% of the data needed across environments. Read on to explore what data readiness means and where common pitfalls have held data initiatives from reaching their fullest potential.

Table of Contents

Defining the Data Readiness Index

Data readiness, at its core, is the ability to fully govern, access, integrate, and trust data across all environments so it can power AI, analytics, and operational decision-making at scale. But what does it really mean to achieve readiness? Data readiness may hold different meanings for a major healthcare organization compared to an energy company or a manufacturer. For example, where 50% of energy industry leaders identified data literacy and training as the biggest barrier to using data effectively, healthcare organizations were more concerned with complicated access requirements and processes (45%). This is just one example, but it highlights the complexity in understanding what it means to build true data readiness.

The State of Infrastructure and Readiness

With the push for AI adoption at such a massive scale, enterprises are leaning further into cloud spending. Sixty-five (65%) percent of respondents said they were shifting their data infrastructure spend by increasing cloud spending.

65%

65% of respondents said they were shifting their data infrastructure spend by **increasing cloud spending.**

As IT leaders chase more powerful AI model performance, cloud environments offer advantages to overall performance. But it cannot be the only path forward. The data that trains AI is public, which, in a competitive market, brings limitations. What really adds value is when an organization can tap into its own data to feed into AI and analytics initiatives. That internal data is made up of information unique to the organization. Therefore, getting that data securely and accurately into AI models generates insights, automation, and predictions that competitors can't easily replicate.

A key indicator of data readiness is the extent to which users can independently discover, access, and use the data they need through self-service capabilities. It's not just a sign of data maturity; it's also a key component of effective governance, hinging on whether an organization has been able to gain control over 100% of its data. Survey respondents were asked how well their infrastructure supports self-service access for technical users. Less than one-third (31%) said they fully supported the capability.

Digging a level deeper into the inner workings of governance practices, surveyed IT leaders were asked about their use of data catalogs and lineage tools to support data discovery and governance. And of those respondents, an overwhelming 91% said they leverage both business and technical catalogs, including automated lineage. However, 47% said they only use data catalogs and basic lineage for select teams or use cases, suggesting standardization across the enterprise is still limited in some instances.

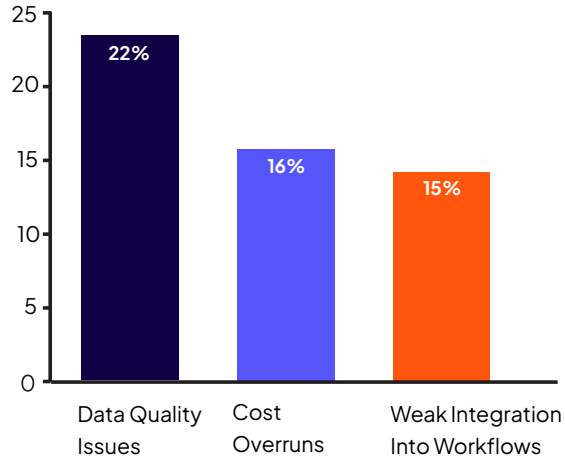
Lingering Roadblocks to Readiness

In Cloudera's 2025 survey, The Evolution of AI: The State of Enterprise AI and Data Architecture, their results showed that organizations were overwhelmingly embracing AI for critical business operations. In fact, that survey found that 96% of respondents had integrated AI into their core business processes.

But AI integration itself is not a guarantee of a meaningful return on that investment. And this new 2026 survey indicates a number of common issues that make ROI fall short. Respondents cited data quality (22%), cost overruns (16%), and weak integration into workflows (15%) as the top contributors to lackluster ROI. With its direct correlation to analytics and AI outputs, data quality has become a major concern for enterprise organizations. In fact, a recent Cloudera study, conducted by the Harvard Business Review Analytic Services, highlighted this challenge, finding that 73% of respondents felt their organization should prioritize AI data quality more.

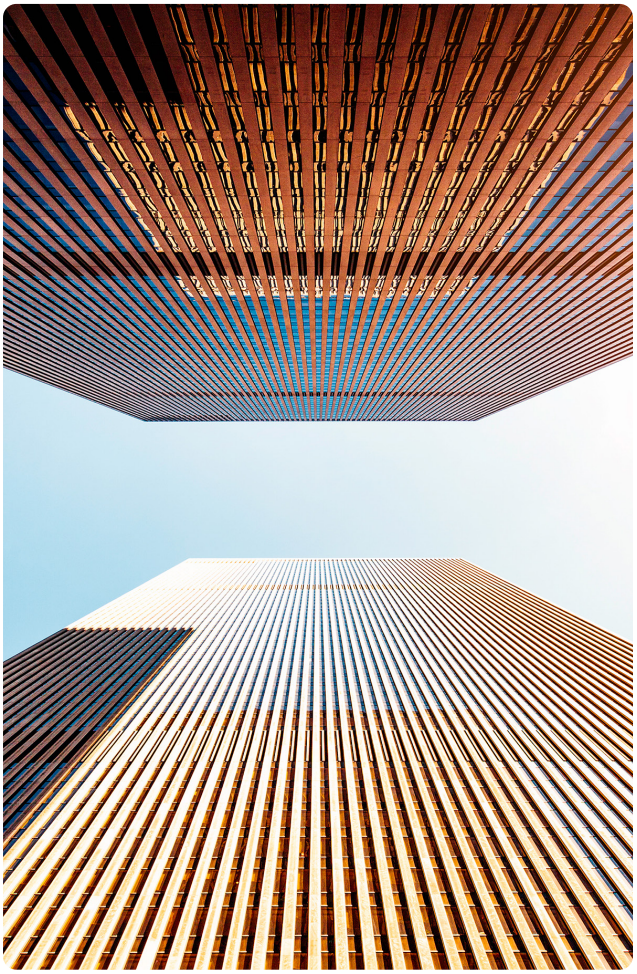
The path to success with AI is filled with obstacles that can keep it from achieving long-term success. The 2025 survey found that data integration (37%) ranked as the top technical limitation to supporting AI. And the results of Cloudera's newest survey show exactly how those limitations manifest within data infrastructures. It's clear that infrastructure performance is still suffering from performance deficiencies. Seventy-three percent of respondents said that infrastructure performance had hindered operational initiatives. And of those, 32% said this was often the case.

When ROI falls short, what is the most common reason?



Why is infrastructure performance such a frequent issue? Many organizations, while quick to embrace new technologies, are letting data silos stick around, hiding valuable insights and generating incomplete datasets. More than one-third (34%) of respondents said siloed data was a top issue preventing them from collaborating, sharing, managing, and using data effectively. The top responses to this question are also common problems for data teams, centering around complicated access requirements and processes (47%), limited visibility into where data resides (44%), insufficient training and data literacy (41%), and cultural resistance to data sharing (34%).

Looking at what is allowing data silos to persist, one issue that stood out with respondents was just how integrated data is within enterprise systems. Most said their data sources were somewhat integrated across environments, but gaps remain. Just 30% of IT leaders said their data sources were fully integrated, while 52% said they are mostly integrated. It's a positive move in the right direction, but this gap suggests many enterprises are still not fully prepared to enable AI projects at scale.



The Data Governance Challenge

As the name implies, the success of data-driven initiatives hinges on one thing—data. AI models are trained on huge repositories of data, much of it public. But what really makes that AI valuable is tapping into organizational data for training. That means that IT teams need to have total visibility and access into 100% of their data, wherever it lives. The survey findings make clear that a majority of respondents' (79%) data-backed initiatives are hindered because they cannot access 100% of the data needed across environments.

Top reasons for companies to execute on data governance in 2026:

- Data analytics, visualization, reporting
- Agentic AI
- Streaming and data collection

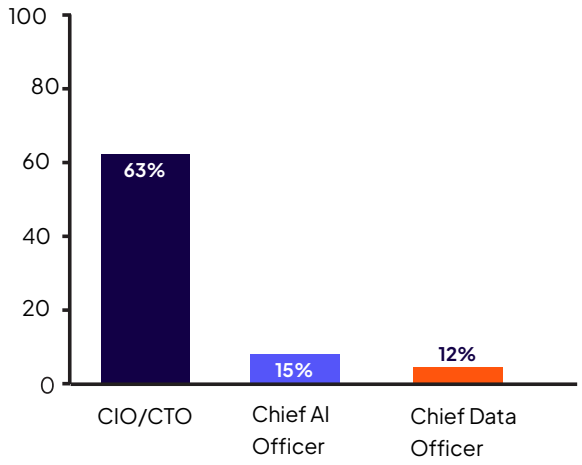
Despite frequent hindrances, respondents are confident in their organization's data. Cloudera's survey found that 84% of respondents felt confident in the accuracy, completeness, and alignment of their organization's data. But in reality, this confidence is at odds with current data governance realities. Less than one in five (18%) respondents said their data was fully governed. And while 71% say most of their data is governed, data-backed initiatives depend on control over the entirety of an organization's data. Leaving datasets ungoverned or unaccounted generates outputs that aren't fully optimized.

A key part of ensuring data quality and maintaining effective governance is the ability to tap into data, regardless of format. Global enterprises across industries, that must include the management of unstructured data. Any governance framework needs to account for all forms of data – structured, unstructured, semi-structured – to ensure accessibility and usability with data-backed initiatives. Of survey respondents, nearly one in five (23%) said all (8%) or almost all (15%) of their data is unstructured.

Responsibility, Accountability & Organizational Alignment

Even as organizations report some degree of inconsistency in terms of data governance, accessibility, and usage, many respondents signaled a strong awareness of data strategy, responsibility, and structure within the broader enterprise. Asked who in their organization was ultimately responsible for enterprise-wide data readiness for AI, a majority of respondents (63%) put the onus on CIOs and CTOs, with the next closest responses being Chief AI Officers (15%) and Chief Data Officers (12%). Respondents also felt that senior leadership has a solid understanding of what's needed to enable successful AI and data initiatives. Eighty-nine percent said that their organization's senior leadership understands and prioritizes the necessary data infrastructure to scale AI.

Who is ultimately accountable for enterprise-wide data readiness for AI?



Enterprises also show a similar level of confidence when it comes to their business strategies, approach to data, and willingness to adapt. A majority of respondents (85%) believe their data strategy is clearly defined and tied to the broader business objectives of their organization. Of respondents, 38% said it was extremely well-defined in their organization, and 47% said it was very well-defined in their organization.

As AI reshapes the enterprise landscape, most respondents have also signaled a recognition and willingness to adapt critical data infrastructure functions, like governance, to this new reality. Every respondent surveyed said their organization was at least somewhat willing to adapt existing frameworks

to ensure data readiness. And the breakdown of these responses reveals even more confidence, with 54% saying their organization was extremely willing to do so and 41% saying their organization was very willing.

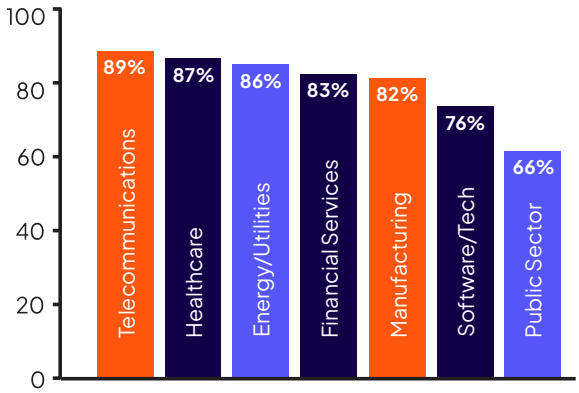
These findings show significant organizational alignment. Respondents feel confident in their leadership's direction and understanding of data infrastructure. All of the respondents even signal a willingness to adapt as needed if it means ensuring data readiness. But even with organizational buy-in, these enterprises show signs of a disconnect, where many of the same issues that have plagued AI initiatives manifest in readiness as well. And some organizational struggles still exist, as some barriers to effective data use exist, including siloed data (34%), cultural resistance to data sharing (34%), and lack of executive sponsorship (28%).

Data Readiness: A Verticalized View

Data readiness must be a priority for every business, regardless of industry. But some industries find themselves ahead in areas, while others lag behind the curve and have a wider gap to close. Looking at how much visibility organizations have into where their data resides, 54% of respondents in telecommunications said it was 'extremely true' that they had complete visibility into where 100% of their data resides. Comparatively, just 30% of respondents in financial services and 31% of respondents in the public sector felt the same.

Data access capabilities showed a similar divergence between industries. Fifty-one percent of respondents in telecommunications said they could access 100% of their organization's data, anytime, regardless of format. Again, this was most starkly contrasted by financial services at 24% and public sector respondents at just 16%. Differences like these highlight the complexities that can arise in highly regulated sectors or sectors that are subject to increased compliance demands, making the need for effective, adaptable data governance practices even more important.

"I have complete visibility into where 100% of my organization's data resides" (extremely true + very true)

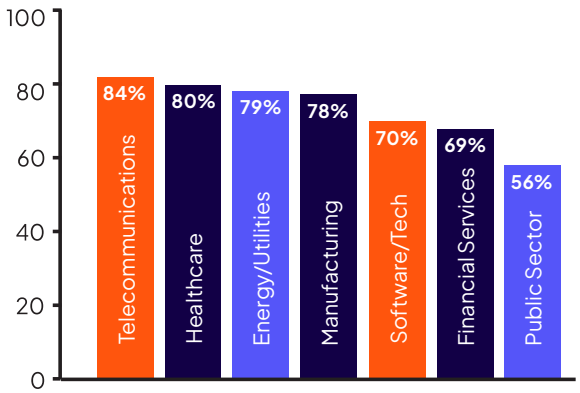


While respondents from telecommunications organizations reported higher confidence in visibility and accessibility, that success hasn't totally translated into operational success. Among industries, telecommunications respondents said infrastructure performance always hinders operational initiatives

at 60%. The next closest industry was software and technology at 39%, followed by financial services (38%), healthcare (28%), public sector (23%), energy and utilities (19%), and manufacturing (16%).

What exactly is hindering those operations? Data governance remains a challenge for many of the industries that Cloudera surveyed. Asked how much of their data is fully governed, government and public sector organizations scored the highest, with 72% saying all or almost all. The remaining industries were relatively equal, with healthcare at 66%, energy and utilities at 65%, software and technology at 63%, financial services at 61%, manufacturing at 58%, and telecommunications at 57%. But where a gap starts to emerge is with the respondents who said all of their data is fully governed. One-third (33%) of respondents in telecommunications said their data was fully governed, followed by software and technology with 26%. The rest of the industries surveyed were all at 20% or less.

"I can access 100% of my organization's data at any time, regardless of format or where it exists" (extremely true + very true)



When asked about the most common reason ROI fell short, each industry revealed a different priority. Earlier, the findings showed that on the whole, organizations were most concerned with data quality issues. But looking individually, energy and utility organizations cited cost overruns (25%) as their top challenge, while software/technology (30%), public sector (24%), and telecommunications (22%) organizations placed the most blame on data quality issues. Healthcare, manufacturing, and financial services organizations all cited weak integration into workflows, with 20% of respondents from each group, respectively.

Building a Foundation for Data Success

The landscape for data infrastructure, and what it takes to successfully power AI and analytics workloads, is rapidly changing. In the face of that change, surveyed respondents are optimistic that they can tackle whatever the future has in store. Eighty-five percent of respondents said they were confident that their current data infrastructure could support their organization's strategic priorities for the next 2–3 years. Of those, 36% said they were extremely confident in that assertion.

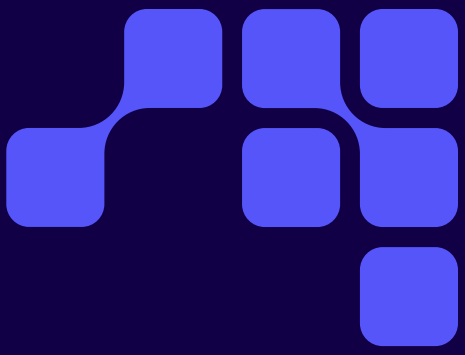
But even as confidence is high, enterprise IT leaders are frequently finding operations hindered by infrastructure challenges and ROI hampered by longstanding complications like data silos. The key to bridging this gap is data readiness. Ensuring data infrastructure is built on a strong foundation of governance, data accessibility, visibility, and quality is key to gaining control over 100% of organizational data, wherever it resides. Moving forward, a shift is underway. The secret to fueling AI and analytics initiatives is becoming less tethered to experimentation and is instead growing deeply intertwined with the maturity of the data infrastructure.

Learn more about how Cloudera is helping enterprise organizations ensure their data is prepared for an AI-driven future.

Methodology

The survey, commissioned by Cloudera and fielded by Researchscape, examines the views of 1,270 IT leaders based across the AMER, EMEA, and APAC regions who work at companies with more than 1,000 employees. The survey was fielded from January 22, 2026, to March 3, 2026. The results of this survey have been weighted to be representative of the overall GDP of surveyed countries.

About Cloudera



Cloudera is the only hybrid 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, on-prem 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, improving security, governance, and real-time and predictive insights. The world's largest brands 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 Dr, 5th Fl, San Jose, CA 95002 USA | cloudera.com