



theCUBE  
research



Starburst

# The State of Data Management and Its Impact on AI Development

August 2024

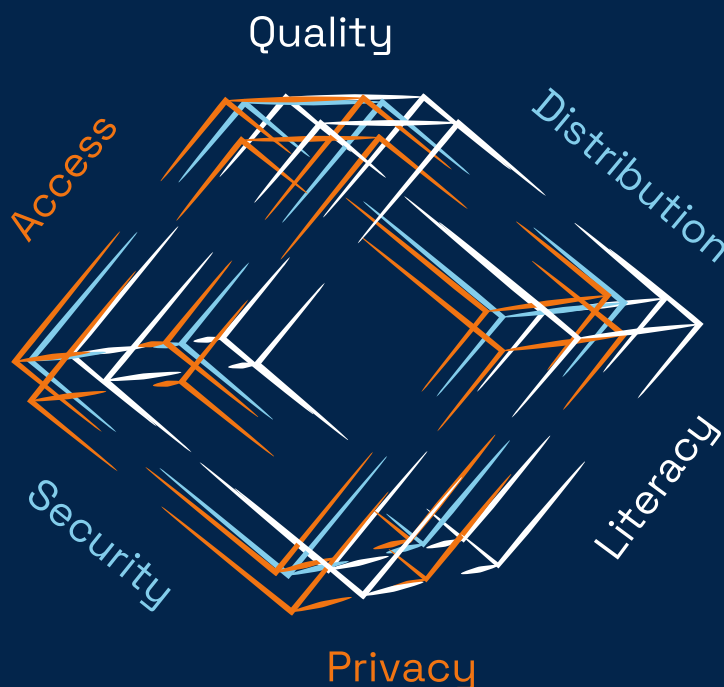
Shelly Kramer

Managing Director, Principal Analyst

# Introduction

Today's enterprises are eager to implement and leverage artificial intelligence (AI) for various applications, from operations to new product development. Effective data management is crucial for successful AI implementation. Companies must ensure data quality, distribution, access, privacy, security, and literacy. Inefficient data management can significantly hinder AI development, while well-planned data management can drive success.

This report provides insights into the state of data management practices and their impact on AI development, based on a survey conducted by theCUBE Research in partnership with Starburst. It highlights the challenges, strategies, and future trends that business leaders need to consider to maximize the potential of AI in their organizations.



# Respondents Profile

300

IT professionals with knowledge of enterprise data management techniques and AI development activities

## Roles

- 51% CTOs
- 25% Software Engineers or Developers
- 15% Data Engineers
- 9% Data Architects

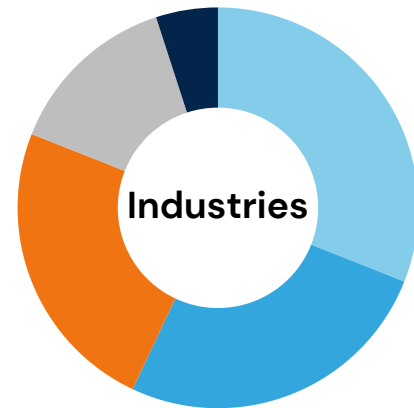


53% United States

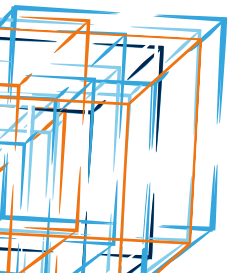
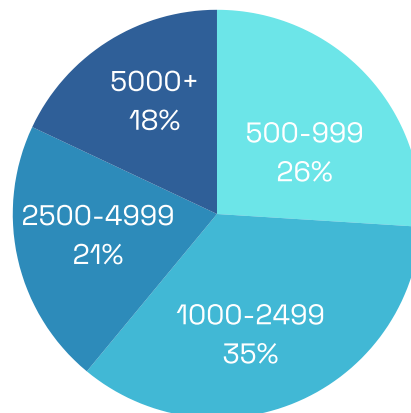


47% Western Europe

- Financial Services (31%)
- Manufacturing (26%)
- Retail and Consumer Goods (24%)
- Technology/Indep. Software Vendors (14%)
- Telecommunications (5%)



## Company Size (# of Employees)

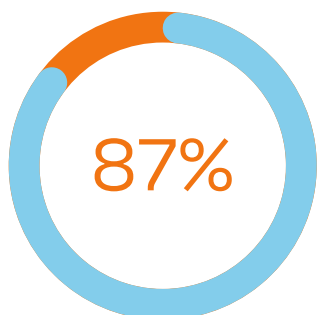


# Executive Summary

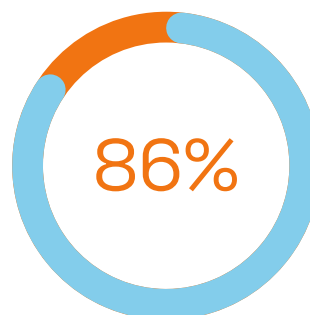
This report, based on a survey of 300 IT professionals, examines the state of data management practices and their impact on AI development across various enterprises.

## Key Findings:

### Strong AI Adoption Intent Exists



Strong or very strong desire to implement AI in the next 12 months



Indicating significant or very significant progress in their AI initiatives

## Alignment of Data Management and AI Success

There's a clear correlation between this alignment and success in AI implementation.

# 90%

Report that their data management practices are somewhat or very aligned with their AI innovation goals.



## Primary Challenges Organizations Face in AI Evolution



**52%** Struggle with organizing structured data for machine learning.

**50%** Face difficulties in organizing unstructured data for retrieval augmented generation (RAG).

The most significant barriers to accessing high-quality data for AI projects are data privacy/security concerns (**28%**) and data volume (**25%**).

## Key Strategies for Improvement

**61%**

Are using agile methodologies for data project management.

**59%**

Are leveraging cloud-based platforms.

**52%**

Have adopted data governance and federated data access strategies.

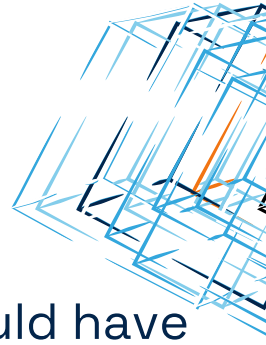
## The Importance of Real-Time Data



**62%**

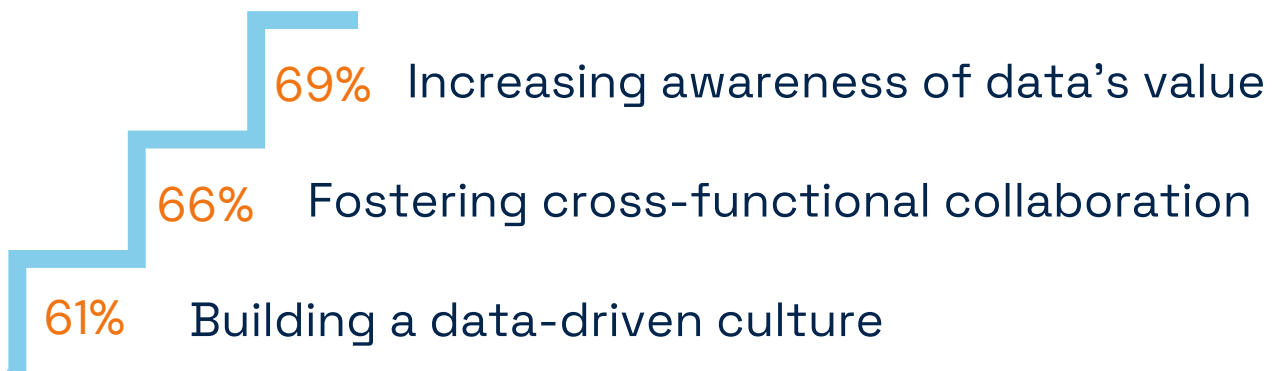
Cite real-time data access as requiring the most attention for AI success, with 65% considering it important or critical.

## Data Literacy Impact



Believe enhanced data literacy would have at least a moderate impact on AI project success, with 40% expecting improved data literacy delivering a significant impact.

## Leaders Must Lead the Way to Achieve AI Success



**AI implementation is deeply intertwined with robust data management practices. Organizations must prioritize real-time data access, data security, and privacy while fostering a data-driven culture to fully leverage AI's potential and maintain a competitive edge in the rapidly evolving digital landscape.**



# Quality Data Management is Driving Successful AI

## Growing Desire and Significant Progress in AI

Respondents across industries recognize the transformative potential of AI. An overwhelming majority of survey respondents report that their organizations have a strong desire to implement AI, with nearly 87% indicating either a "very strong" (47%) or "strong" (40%) intent. This enthusiasm is matched by tangible progress, with 86% reporting either significant strides (50%) or very significant (26%) progress in their AI initiatives.

### Primary Drivers for AI Adoption

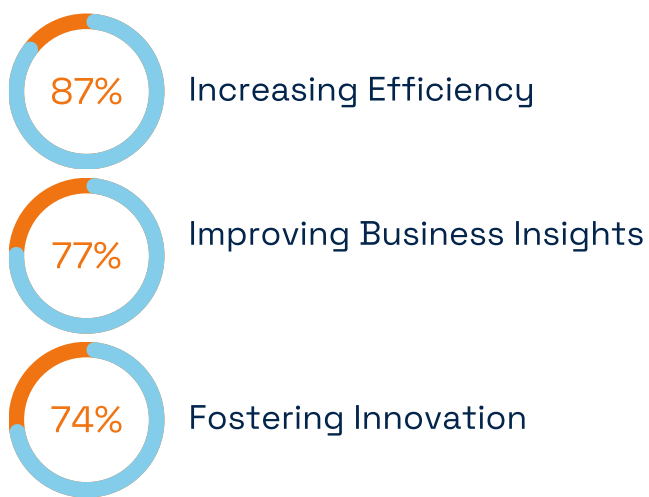
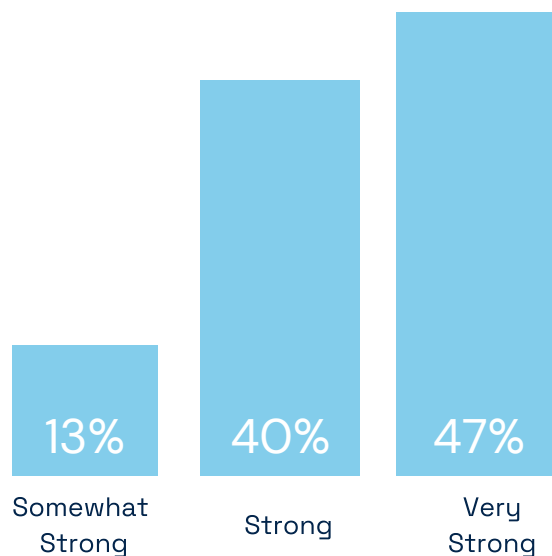


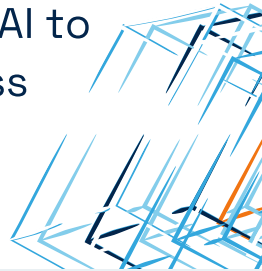
Figure 1: Rate your company's desire to implement AI in business operations over the next 12 months?



**Key Takeaway:** Progress in the adoption of AI is largely spurred by organizations wanting to take advantage of the potentially massive productivity gains AI can deliver, increasing efficiency, delivering best-in-class customer experiences, and driving innovation. Success in the enterprise AI journey will help organizations establish a competitive advantage in a relatively short period of time, which is part of its appeal.

# Alignment of Data Management and AI Success

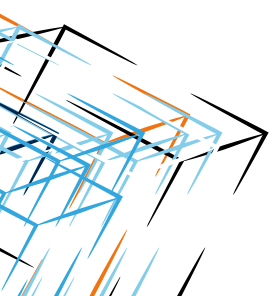
Successful AI implementation requires a seamless integration of data management practices with AI goals. Many organizations report positive alignment between their data management practices and their AI initiatives, with 53% of respondents indicating their practices are "somewhat aligned" and 37% stating they are "very aligned." This alignment enables companies to leverage AI to its fullest potential, creating new products, enhancing process efficiency, and generating deeper business insights.



# 90%

Rate their organization's alignment between data management practices and AI innovation goals as somewhat or very aligned.

That 37% report their data management efforts are "very aligned" with their AI development efforts reveals that approximately 6 in 10 respondents believe their companies have a distance to travel before they are set up optimally to capitalize on AI.





Notable also is the clear association between what respondents perceive as alignment of their organization’s data management practices and their perceived success with AI implementation. The table below shows this association. Note that the greater the alignment, the greater the success in implementing AI.

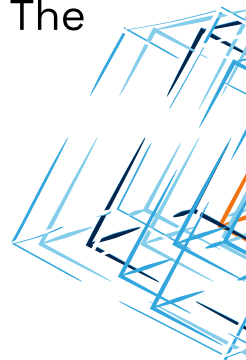


Table 1: Association between progress with AI initiatives and Alignment of data management practices with AI innovation goals.

		Rate the progress your company has made with its AI implementation goals.					
		None at all	Not very significant	Moderate	Significant	Very significant	Total
		(0)	(6)	(49)	(167)	(78)	
Rate the current alignment between your organization's data management practices and its AI innovation goals.	Very aligned	0%	0%	6%	28%	79%	37%
	Somewhat aligned	0%	17%	78%	63%	18%	53%
	Neutral	0%	50%	14%	5%	1%	6%
	Somewhat misaligned	0%	17%	2%	3%	0%	2%
	Very misaligned	0%	17%	0%	1%	0%	1%



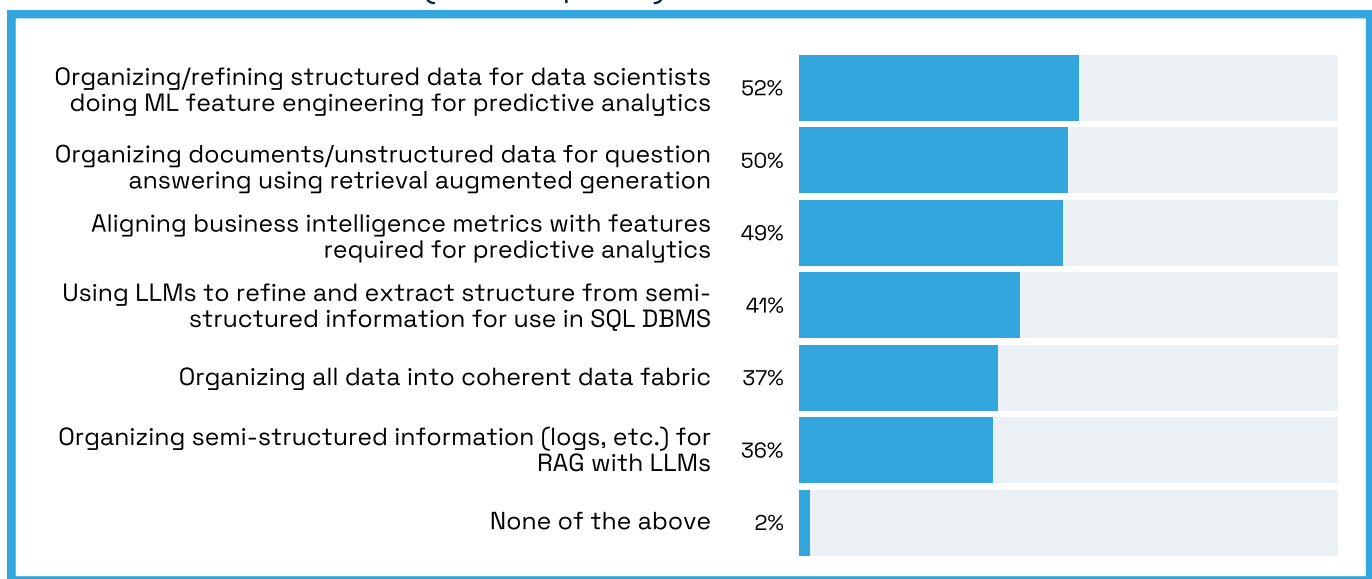
**Key Takeaway:** Success with AI is highly dependent on alignment between an organization’s data management practices and its overarching innovation goals. Data plays a vital role in an organization’s AI journey, and getting arms around both unstructured and structured data, and having best-in-class data management practices in place play an out-sized role in the ability to derive AI-powered, actionable business insights and benefit from corporate AI initiatives. When it comes to data, the adage: garbage in, garbage out could not be any more applicable in the age of AI. Success with AI depends on a solid foundation of data management practices — that should be your starting point.

# Privacy/Security Concerns and Data Volume Concerns are Limiting AI Development

## Organizing and Accessing High-Quality Data is Paramount

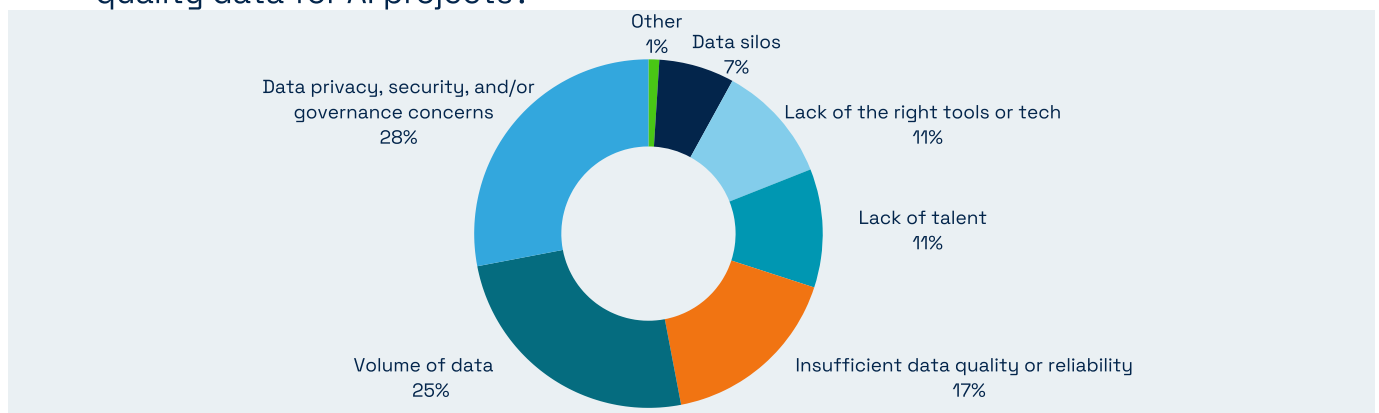
One of the most significant challenges is organizing and refining data for AI use. Over half of the respondents (52%) report difficulties in organizing structured data for machine learning feature engineering, while 50% struggle with organizing unstructured data for retrieval augmented generation (RAG).

Figure 2: What are the top three challenges you face in integrating data management with AI initiatives? (Select up to 3)



Beyond these foundational issues of data management, other challenges lurk that can confound success with AI implementation. When asked about the most significant barrier to high quality data for AI projects, the top two responses were data privacy, security and/or governance concerns (28%) and Volume of data (25%).

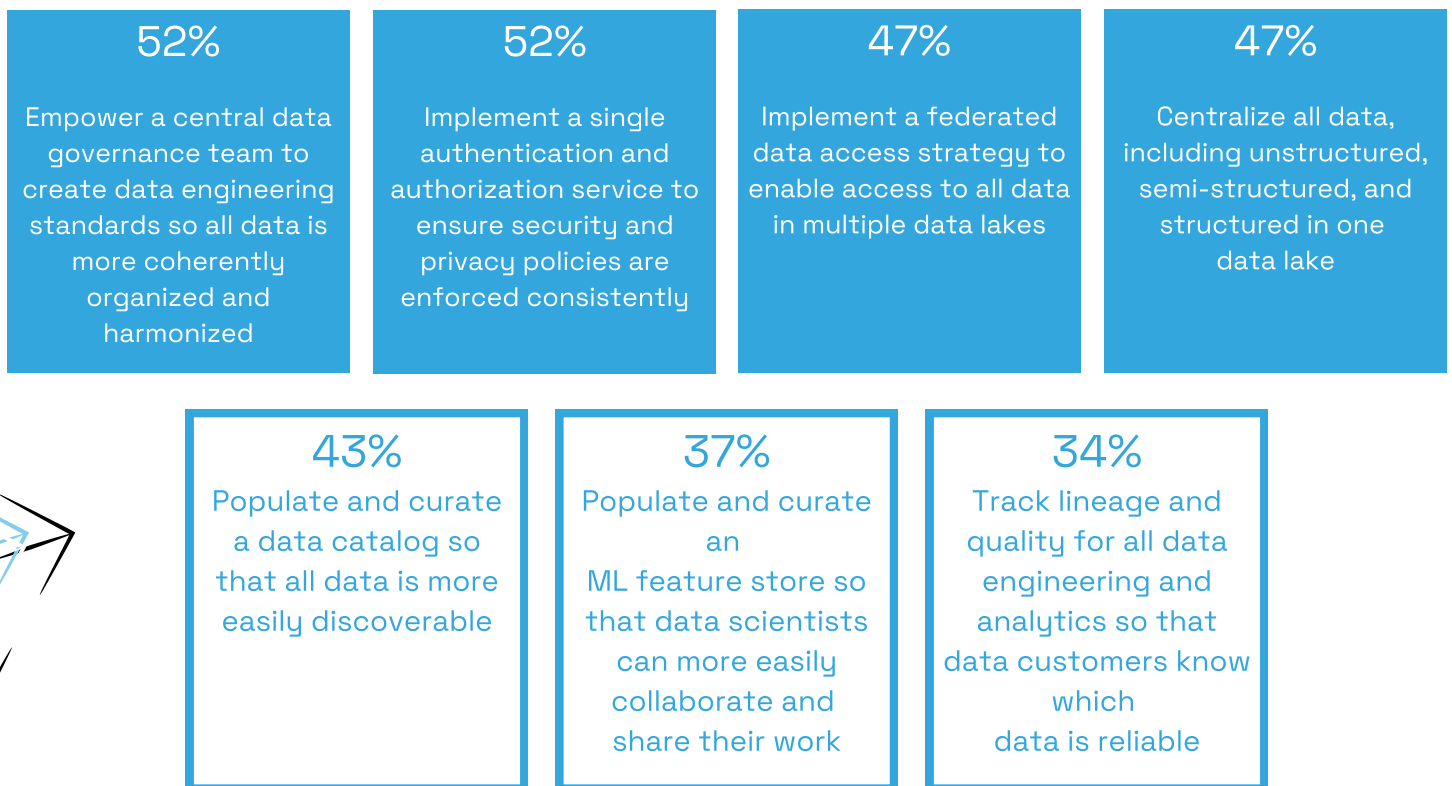
Figure 3: In your experience, what has been the most significant barrier to accessing high-quality data for AI projects?



# Enterprises are Addressing Data Management Challenges with Advanced Data Governance, Universal Data Access, and Real-Time Data Processing

To meet challenges around data accessibility, respondents report a series of strategies. Survey respondents were asked about a variety of potential strategies they are currently using to improve data accessibility for AI success (Figure 4). These strategies are discussed below.

Figure 4: What strategies have you employed to improve data accessibility for AI projects? (Select all that apply)

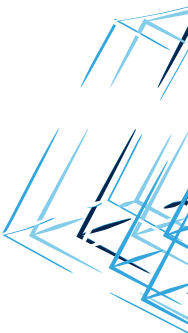


# One-half of respondent organizations have adopted data governance and federated data access strategies

Empowering a central data governance team to establish data engineering standards is a crucial step – 52% of respondents report that their organizations are using this as a key strategy. This approach ensures that data is coherently organized and harmonized, facilitating easier access and higher data quality.



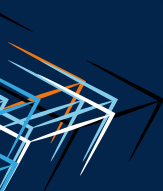
Report that their organizations are implementing single authentication and authorization services further enhances data security and privacy compliance.



As suggested by survey respondents, data silos are a major hurdle. In many organizations, data is dispersed across various systems, often across clouds and on-premises, making it difficult to fully integrate and utilize data for AI projects. Addressing this issue involves implementing strategies like centralizing data in a single data lake (47%) or employing hybrid data federation techniques (47%) to enable seamless access across multiple sources.



**Key Takeaway:** Data is a primary concern across all organizations as they work to embrace and incorporate AI into business operations. Starting with the premise that data is foundational to successful AI initiatives is key, and adopting an approach that prioritizes data governance, eliminates silos across the organization, and commits to centralizing data so that it's accessible to those who need it will play an outsized role in overall success with AI initiatives.

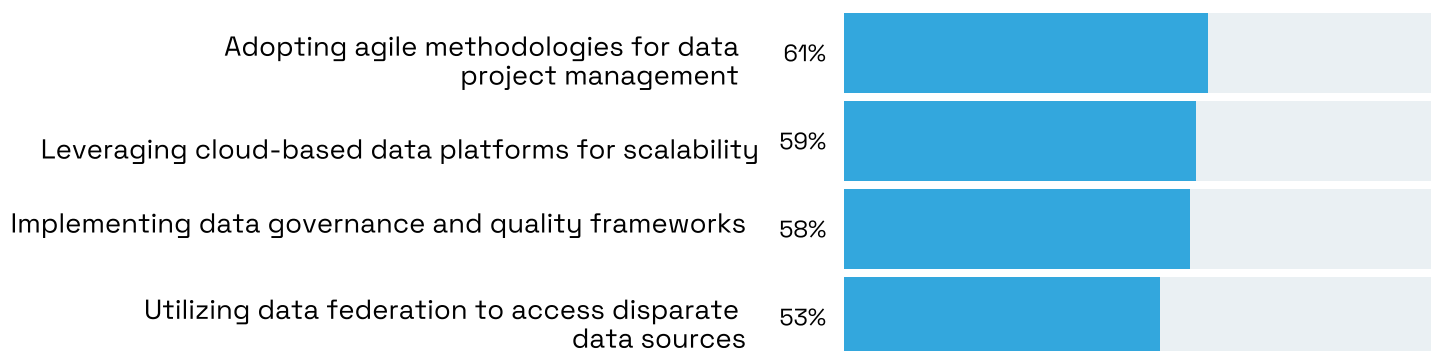


# Leveraging Agile, Cloud and Data Federation Strategies form a Core Approach to Integrating Data Management with AI

Leveraging cloud-based platforms (59%) and utilizing data federation strategies (53%) are effective ways to scale AI initiatives. These technologies enable organizations to manage large volumes of data efficiently and ensure that data is readily accessible for AI applications. Moreover, implementing robust data quality and observability tools (58%) helps maintain high standards of data reliability.

Also apparent is that the landscape of data management and AI is constantly evolving. Organizations must stay agile, adopting flexible data infrastructure practices that can quickly adapt to new AI technologies and methodologies. This agility allows businesses to remain competitive and responsive to changing market demands. 61% of the respondents' organizations have found that adopting agile methodologies for data project management is an effective strategy.

Figure 5: What strategies have been most effective in your organization for integrating data management practices with AI capabilities?





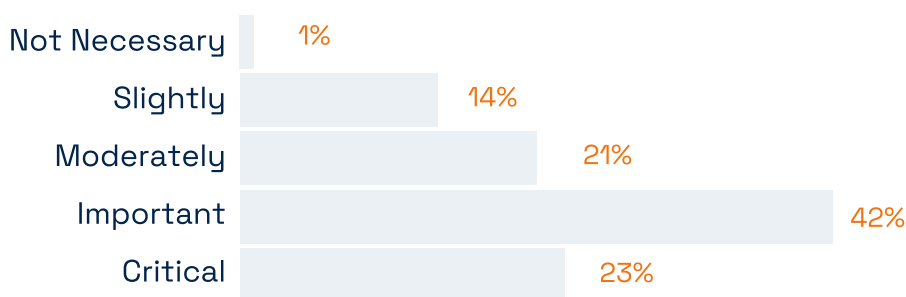
# Real-Time Data Access and Quality Emerges at Critical to AI Success


Real-time data access is a data management area cited by 62% of respondents as requiring the most attention critical for the success of AI initiatives (Figure 6). Ensuring that data is accurate, consistent, and readily available enhances the reliability and performance of AI models. When asked how critical real time access to data is to the success of their organizations' AI initiatives, 65% of respondents report it is important or critical (Figure 7).

Figure 6: In your experience, which areas of data management require the most attention to better support AI capabilities? (Select all that apply)



Figure 7: How critical is real-time data access in the success of your AI initiatives?



 Key takeaways: Agility and scalability are crucial when developing AI initiatives. Planning to manage large volumes of data and implementing tools to maintain data quality and reliability. Flexible data infrastructure that can adapt to new AI technologies and methodologies allows businesses to stay competitive in an evolving landscape. Real-time access to accurate data is critical to the success of AI initiatives.

# Emergent Trends and Recommendations for Aligning Data Management with AI Initiatives

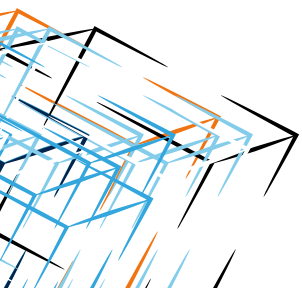
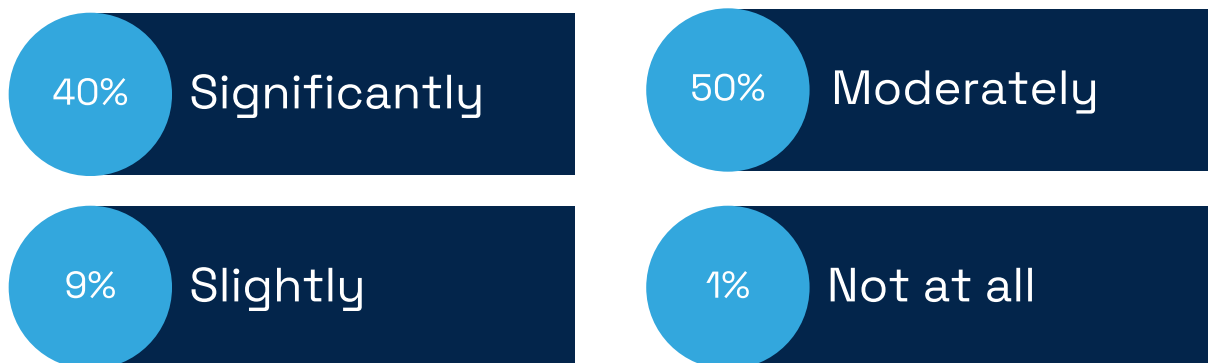
## Ethical Considerations and Federated Learning

As AI continues to grow in importance, data management practices will increasingly need to address ethical concerns, ensuring responsible use of data. Federated learning, which allows AI to train on decentralized data while enhancing privacy, is seen as a key trend by 17% of respondents. Business leaders should consider these trends to stay ahead in the AI development race.

## Building a Data-Driven Culture

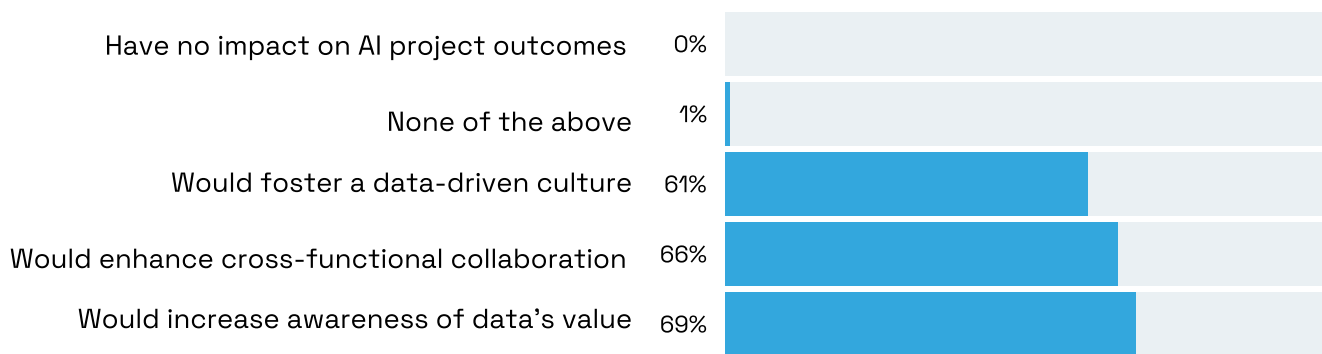
Enhanced data literacy across the organization is essential for maximizing AI's impact. 90% of respondents indicate that greater data literacy would have at least a moderate impact on the success of AI projects, 40% reporting it would impact the success significantly.

Figure 8: To what extent could enhanced data literacy across the organization impact the success of AI projects?



Further, respondents indicate that Increasing awareness of data's value (69%), fostering cross-functional collaboration (66%), and building a data-driven culture (61%) are pivotal strategies. By prioritizing data quality, governance, and scalability, organizations can create a robust foundation for AI innovation.

Figure 9: Respondents indicated that enhanced data literacy across the organization could impact the success of AI projects moderately or significantly. Which of the reasons listed below do you believe may account for the potential for enhanced success? (Select all that apply)

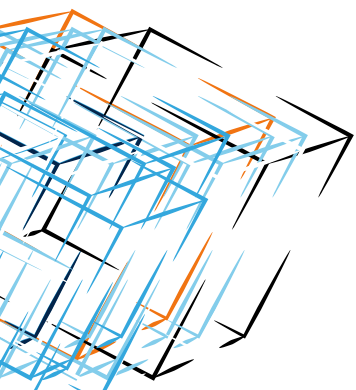
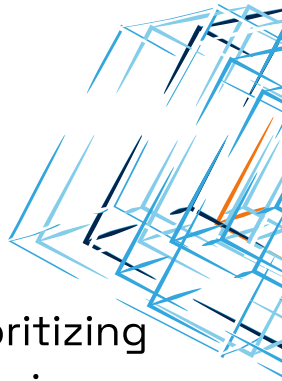


Key takeaways: Building a foundation of data literacy throughout an organization and creating and fostering a data-driven culture are the keys to success with any corporate-wide AI initiatives. Success requires buy-in from top leaders, and relies on communicating the importance of a data-driven culture throughout the organization on a regular and ongoing basis. Every worker is in the data business, just as every employee is in the customer service business, and every employee is in the security business — and success with AI requires a data-first mindset. For best results, treat this effort, this building of a data-centric culture, as an ongoing communication and training exercise; it should not be treated as a one-and-done effort.

# Conclusion and Recommendations: Survey Reveals Critical Role of Real-Time Data Access in AI Success

The journey toward effective AI implementation is deeply intertwined with robust data management practices. By prioritizing data accuracy, embracing scalable architectures, and enhancing data security, business leaders can unlock the full potential of AI. This will drive efficiency, foster innovation, and provide valuable business insights, positioning organizations for long-term success in an increasingly data-driven world.

Perhaps the most critical features of a future data management strategy will focus on real-time access to data, which requires attendant strategies for protecting data security and privacy. Data management approaches, platforms, and tools that address these needs simultaneously will enable companies to fully engage with the AI revolution ahead of companies that have not addressed the multiple challenges or are unaware of their impact.



# About theCUBE Research

TheCUBE Research is an independent research, analysis, and advisory firm in the enterprise and emerging tech sectors, tracking disruptive innovation, industry trends, and creating industry categories. theCUBE Research is where tech executives and business leaders go to gain real-time, transformative insights and expert advice and guidance designed to help grow share of voice and share of market.

For additional information, visit [theCUBE Research](#).

# About Starburst

Starburst, the Open Hybrid Lakehouse, is the leading end-to-end data platform to securely access, analyze, and share data for analytics and AI across hybrid, on-premises, and multi-cloud environments. As the leaders in Trino, a modern open-source SQL engine, Starburst empowers the most data-intensive and security-conscious organizations like Comcast, Halliburton, Vectra, EMIS Health, and 7 of the top 10 global banks to democratize data access, enhance analytics performance, and improve architecture optionality. With the Open Hybrid Lakehouse from Starburst, enterprises globally can easily discover and use all their data to power AI and other business-critical applications like anti-money laundering and fraud analytics, next best products, customer 360, log analytics, and ESG reporting.

For additional information, please visit [Starburst](#).

## DISCLAIMER:

The information presented in this document is for informational purposes only and may contain technical inaccuracies, omissions, and typographical errors. theCUBE Research disclaims all warranties as to the accuracy, completeness, or adequacy of such information and shall have no liability for errors, omissions, or inadequacies in such information. This document consists of the opinions of theCUBE Research and should not be construed as statements of fact. The opinions expressed herein are subject to change without notice.

TheCUBE Research provides forecasts and forward-looking statements as directional indicators and not as precise predictions of future events. While our forecasts and forward-looking statements represent our current judgment on what the future holds, they are subject to risks and uncertainties that could cause actual results to differ materially. You are cautioned not to place undue reliance on these forecasts and forward-looking statements, which reflect our opinions only as of the date of publication for this document. Please keep in mind that we are not obligating ourselves to revise or publicly release the results of any revision to these forecasts and forward-looking statements in light of new information or future events.