

The Business Value of DataHub Cloud



Lynne Schneider

Research Director,
Data Collaboration and Monetization,
and Location and Geospatial Intelligence, IDC



Matthew Marden

Research Vice President,
Business Value Strategy Practice, IDC



Table of Contents



Click any title to navigate directly to that page.

Business Value Highlights	3
IDC Opinion	4
Situation Overview	5
DataHub Cloud Overview	5
Business Value of DataHub Cloud	6
Firmographics	6
Choice and Use of DataHub Cloud	7
The Business Value of DataHub Cloud	9
Data Use Benefits	10
Data Quality Benefits	15
Benefits to AI Operationalization	18
Business and Operational Benefits	19
Data Cost Benefits	21
Challenges and Opportunities	22
Conclusion	23
Appendix A: Methodology	24
Appendix B: Accessible Data Tables	25
About the IDC Analysts	27
Message from the Sponsor	28

Business Value Highlights

Click each [→](#) to jump to related content. Use "Return to Highlights" in the page header to get back here.

91% faster data searches



More than 3x higher data search success rate



58% less IT staff time required to support failed data searches



17% higher data engineering team productivity



82% faster to find/validate data sets



153% more data assets with complete metadata



75% more data sets with mapped data lineage



20% data governance team efficiency gains



119% more AI/ML models successfully moved to production



24% lower AI/ML project failure rate



58% faster to resolve data-related outages



IDC Opinion

Organizations using DataHub Cloud describe meaningful improvements in how data is managed and consumed across the enterprise. Interviewed customers point to reduced friction across engineering, analytics, and governance teams, along with greater confidence in data quality and lineage. These improvements are associated with faster time to insight and more consistent operationalization of analytics initiatives.

These reported benefits align with broader enterprise trends. As data estates expand in scale and complexity, and as regulatory requirements and AI-driven use cases intensify, organizations are prioritizing unified visibility and automated governance capabilities. In this context, platforms such as DataHub Cloud are positioned to support cross-functional data coordination and policy enforcement at the scale required for modern data and AI programs.

IDC conducted interviews with organizations using DataHub Cloud to discover, understand, govern, and monitor data assets. Interviewees described moving from fragmented, individual-dependent knowledge to a centralized, searchable platform with lineage and contextual metadata. As a result, employees can more easily find and validate data, reducing data team bottlenecks and shortening the time to build, troubleshoot, and operate data-driven solutions. Participants reported improved data context management, meaningful productivity gains for data engineers and analytics teams, greater confidence in data choices, and stronger governance and compliance capabilities.

Overall, interviewed customers described DataHub Cloud functioning as a robust context management platform that creates a unified foundation for enterprise data and AI management by comprehensively capturing organizationwide metadata. It connects data assets to their business meaning, quality metrics, ownership, and relationships to other systems, which include AI models, thereby enabling various stakeholders to discover, understand, maintain, and safely utilize trusted data assets. The result is accelerated insights, improved productivity, reliable data pipelines, reduced compliance risk, and more reliable AI deployments built on trustworthy data foundations.

Situation Overview

IDC observes that organizations are reassessing their data architectures as complexity increases across hybrid and multicloud environments. Growing data volumes, distributed ownership models, and expanding regulatory requirements are exposing the limits of fragmented governance approaches. Simultaneously, AI-driven initiatives are placing new operational demands on data teams, requiring reliable, well-documented, and policy-aligned data pipelines that can support production use cases.

In response, organizations are prioritizing architectures that provide unified metadata visibility, embedded governance controls, and greater automation across the data life cycle. AI capabilities are increasingly applied to core data management functions such as lineage capture, quality monitoring, and policy enforcement. The objective is to reduce manual effort while improving consistency, transparency, and trust in enterprise data.

Looking ahead, IDC expects closer integration across governance, metadata intelligence, and cross-cloud connectivity to become a defining characteristic of modern data platforms. As AI programs expand and compliance expectations evolve, organizations will place greater value on solutions that simplify data discovery and strengthen confidence in data assets. In this context, platforms such as DataHub Cloud, with an emphasis on metadata-driven automation and broad ecosystem integration, align with emerging enterprise priorities.

DataHub Cloud Overview

Organizations continue to face fragmented data environments, inconsistent governance practices, and increasing pressure to scale AI initiatives. Data teams spend significant time locating data, reconciling definitions, and resolving quality issues, limiting capacity for higher-value work. As cloud adoption expands and AI models move into production, enterprises require unified visibility across technical, business, and operational metadata to improve efficiency and support trusted decision-making.

DataHub Cloud is designed to provide a unified context management platform spanning discovery, observability, and governance. The platform supports cross-platform searches, natural language queries, and role-based metadata views intended to streamline data access. Real-time quality monitoring, automated lineage tracking, and anomaly detection help reduce operational risk and improve issue resolution before downstream impacts occur.

Organizations report measurable productivity and cost benefits. Data engineering and platform teams cite time savings through automated assertions, reduced manual governance effort, and more efficient change impact assessment. Customers also report identifying redundant or unused data, contributing to improved cloud cost management, including reductions of up to 25% in Snowflake storage costs. By strengthening metadata completeness and reinforcing trust in data assets, the platform supports the transition of AI initiatives from pilot phases to scaled production environments.

Business Value of DataHub Cloud

Firmographics

IDC interviewed organizations using DataHub Cloud to understand their experiences from both a quantitative and qualitative perspective. Study participants represented a mix of industries, including ecommerce, financial services, healthcare, IT services, and software, with operations across the United States, India, and the Netherlands. The sample reflects a mix of medium-sized organizations with a very large enterprise, with interviewed organizations employing an average of 126,380 people (median: 2,800) and having an average annual revenue of \$7.01 billion (median: \$2.80 billion), underscoring the scale and complexity of the data environments where they are using DataHub Cloud to support and leverage their businesses (**Table 1, below**).

Table 1
Firmographics of Interviewed Organizations

Firmographics	Average	Median
Number of employees	126,380	2,800
Number of IT staff	15,511	1,200
Number of business applications	1,631	350
Annual revenue	\$7.01B	\$2.80B

[Table 1 continued next page](#)

Table 1 continued

Firmographics	Average	Median
Countries	United States (3), India, Netherlands	
Industry verticals	eCommerce, Financial Services, Healthcare, IT Services, Software	

n = 5; Source: IDC Business Value In-Depth Interviews, February 2026

Choice and Use of DataHub Cloud

Interviewed organizations described choosing to adopt DataHub Cloud to address growing data volumes and complexity while improving their ability to discover, obtain, and operationalize data and to improve their ability to contextualize and use data. Study participants emphasized the importance of having an open, extensible data platform foundation, including the ability to integrate broadly across tools and react to changes through automation, alongside a desire to move away from the burden of self-hosting and infrastructure management. In addition, they cited the need to establish stronger governance, security, and compliance capabilities around their data environments and to build a more scalable data foundation that supports advanced analytics and emerging AI initiatives.

Interviewed DataHub customers spoke to their unique decision criteria:

Flexibility to react to required changes:

“The main driver of our choice of DataHub Cloud was flexibility. It’s based on an open source platform, and it has an actions framework that allows us to react to changes. We looked at other providers, but open source and the flexibility to react to change made DataHub Cloud the best option for us.”

Need for additional capabilities to handle growing data:

“Our data was growing, and we needed a way to catalog and search through it easily. ... When we compared the feature list, we decided to go with DataHub Cloud because the other option was very new and had very little support.”

Agility, scalability, and foundation for new technologies:

"We chose DataHub Cloud for agility, scalability, and cost efficiency, alongside the need to support emerging technologies like generative AI and advanced analytics. Centralized data management, governance, security, performance, and disaster recovery were also main reasons why we went for DataHub Cloud."

Security and moving away from hosting:

"Our company is very strict on security. ... With DataHub Cloud, the hosting is included, and we didn't have to self-host or manage all the underlying infrastructure with associated costs."

Study participants commonly described their use of DataHub Cloud as providing an enterprise-level data catalog and governance layer, with use cases that ranged from focused deployments to broad enablement for thousands of data consumers. On average, participants had 77 licensed users (median: 50) supporting approximately 3,080 data consumers (median: 250) (**Table 2, below**), indicating that DataHub Cloud often serves as a shared self-service capability for a larger audience than the direct administrator population. Interviewees integrated DataHub Cloud deployments with a wide variety of data platforms and applications (average: 387 integrations; median: 8) and commonly extended through custom entities (average: 2,107; median: 22) to support organization-specific data products and governance activities. Study participants also noted that DataHub Cloud supports data contributing to material portions of their revenue-generating activities (average: 29%; median: 31%).

Table 2
Use of DataHub Cloud by Interviewed Organizations

Use	Average	Median
Number of DataHub licenses	77	50
Number of data consumers	3,080	250
Number of integrated data platforms	387	8
Number of custom entities	2,107	22
Number of business applications	602	8
Percent of revenue	29%	31%

n = 5; Source: IDC Business Value In-Depth Interviews, February 2026

The Business Value of DataHub Cloud

Study participants described DataHub Cloud as providing a much-needed central source of truth for data context, including searchable metadata, ownership, and lineage, which reduces reliance on ad hoc knowledge sharing and accelerates self-service data discovery. By making it faster and easier for users to find the right data sets and understand how they are produced and used, interviewed organizations have reduced bottlenecks on data teams and improved the pace of analytics and data product delivery. Study participants also highlighted that broader visibility into usage and dependencies helped teams make changes with less risk and more confidence, reducing rework and duplication across pipelines, models, and reports.

In parallel, interviewed organizations reported stronger operational discipline with DataHub Cloud across data quality, governance, and issue resolution. DataHub Cloud has increased lineage coverage, improved metadata completeness, and clarified data ownership, enabling stakeholders to identify the impact of upstream changes and incidents more quickly. This visibility supports more consistent governance and compliance while improving operational continuity by shortening root-cause analysis during outages. Study participants also noted greater readiness to scale analytics and AI/ML initiatives through improved trust in data inputs and stronger monitoring and management of data products.

Study participants provided examples of these benefits achieved with DataHub Cloud:

Minimized data team bottlenecks:

"Before DataHub Cloud, it often took several days to find the right data because requests came to the data team. Now, people can go directly to DataHub Cloud and find data in minutes. This reduces the bottleneck on the data team and enables faster decision-making."

Much faster and more efficient data model creation:

"Researching a data model used to take about six months of effort. With DataHub Cloud, that research can be done in about a month. It cuts the research time significantly, and about 150 people save 20%–30% of their time to use for other work."

Understanding data use and making seamless changes:

"DataHub Catalog and Lineage help us understand data usage. Whenever we make changes to data, we use the Lineage feature to understand the impact to minimize disruption. We also use DataHub Cloud extensively when there are issues in the data pipeline to understand what broke and what was impacted."

Companywide analytics enabler and single source of truth for data:

"DataHub Cloud gives us enterprisewide analytics capabilities and a single focal point for business owners. It also supports regulatory compliance, which is critical in financial services. Having a single source of truth eliminates issues with distributed data and improves how the organization uses data overall."

Data Use Benefits

Interviewees reported that DataHub Cloud improves their organizations' day-to-day understanding and use of data by centralizing data discovery, context, and validation in a single interface. Searchable technical and business metadata, combined with enhanced lineage and usage information, helps their users identify relevant data products more quickly and reduces the need to rely on emails, Slack channels, or time-consuming consultations with data experts. Study participants also noted that standardized documentation and clearer ownership information with DataHub Cloud reduce uncertainty and debates about which data sets to use, enabling faster decisions and more consistent analytics.

Study participants explained that DataHub Cloud supports more efficient work by enabling data engineers and analytics teams to better understand data in context, reducing the time spent interpreting data, understanding dependencies, and troubleshooting pipelines. Improved lineage visibility and richer metadata help avoid duplicative work and accelerate model and report development by increasing confidence that data sets are accurate and fit for purpose. Interviewees also noted that extensibility and automation operationalize governance and quality signals, enabling scalable self-service access. Overall, DataHub Cloud provides a more complete, trusted data context layer, making it easier for non-specialists to find and use data without escalation.

DataHub Cloud customers interviewed for this study provided examples of these data-related contextual benefits:

Greater confidence in data:

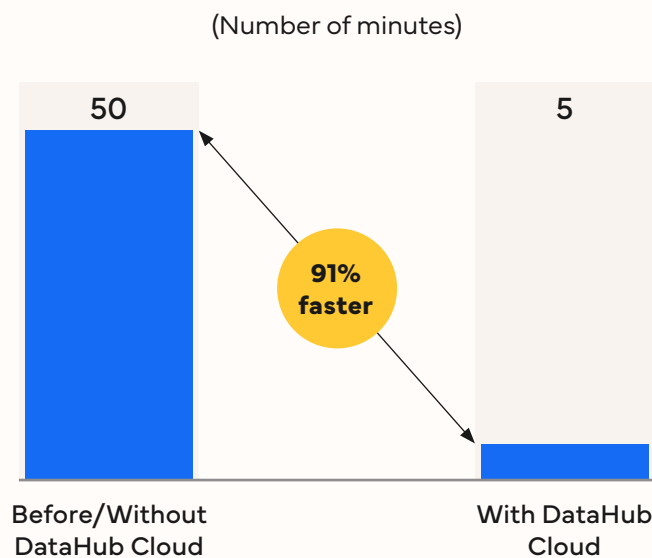
"Because we have DataHub Cloud, people are more confident that they are making the right choice and don't spend a lot of time debating about data. Every developer is saving at least two hours a week, which adds up to thousands of hours per month across the organization."

Much greater data lineage coverage:

"Before DataHub Cloud, we didn't have any data lineage, and now we have lineage for about 90% of data. This has been helpful in multiple cases this year where people were able to identify the use cases of their tables and proactively contact owners before making changes to avoid downstream issues."

A fundamental benefit of DataHub Cloud for study participants is significantly faster access to the data they need. Study participants reported that DataHub Cloud has reduced the average time required for an average data search from 50 minutes to 5 minutes, representing a 91% improvement on average, by providing centralized search, metadata, and lineage context (**Figure 1, below**). This accelerated discovery enables users to validate and use the right data sets sooner while reducing reliance on the data team for routine questions.

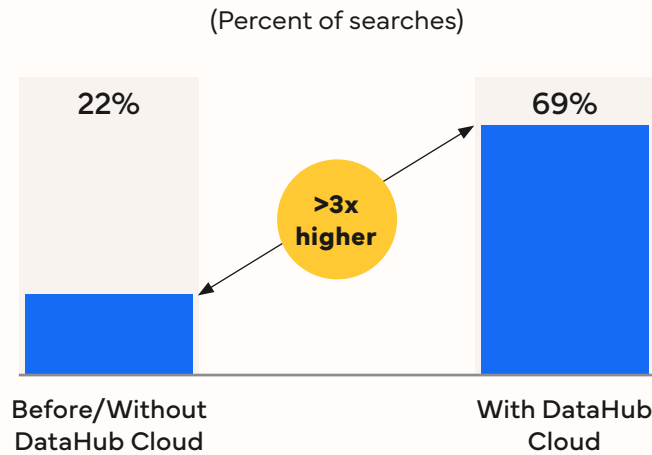
→ **Figure 1**
Impact on Data Search Time



n = 5; Source: IDC Business Value In-Depth Interviews, February 2026

DataHub Cloud has also significantly improved data search outcomes. Interviewees reported that their data search success rate increased from 22% to 69%, which is more than a threefold improvement, by enhancing keyword search, metadata visibility, and self-service access (**Figure 2, next page**). Higher success rates reduce rework and data team escalations, enabling stakeholders to move from questions to answers more quickly.

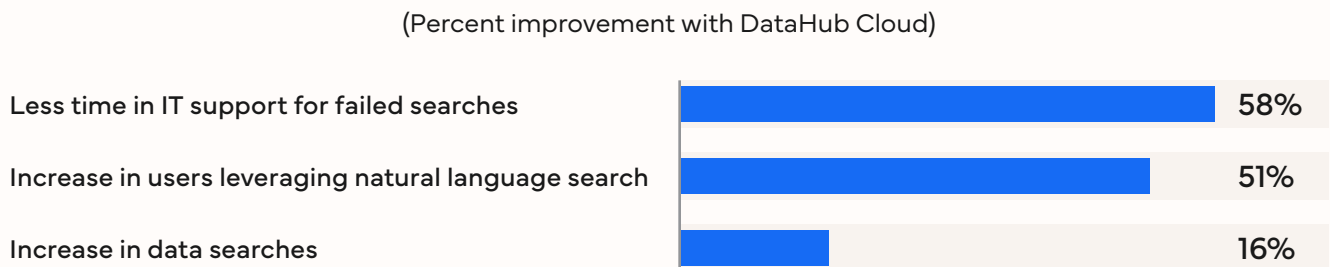
→ **Figure 2**
Impact on Data Search Success Rate



n = 5; Source: IDC Business Value In-Depth Interviews, February 2026

Study participants reported additional improvements beyond search time and discovery success, including gains in support efficiency. These benefits include a 16% average increase in the number of data searches conducted, a 51% rise in users leveraging natural language search capabilities, and a 58% reduction in IT time spent supporting failed searches (**Figure 3, below**). These gains indicate that DataHub Cloud not only accelerates individual searches but also scales self-service data usage across a broader user population, benefiting both data users and data support teams.

→ **Figure 3**
Impact on Data Discovery KPIs



n = 5; Source: IDC Business Value In-Depth Interviews, February 2026

Improved access to higher-quality data with DataHub Cloud has enabled teams that deliver value through data. For example, study participants reported that their data engineering teams are 17% more productive on average, representing a meaningful increase in the value these teams deliver.

Several interviewees described in detail the positive impact on their data engineering teams.

Significant productivity and utilization gains for data teams:

“Data engineers are about 30%–35% more productive, and their utilization improved from around 65% to about 80%. Analytics teams reduced delivery time from six or seven weeks to about three weeks because data is easier to access and trust.”

Efficiencies and productivity gains for data engineers and analytics teams:

“Data engineers save around 20%–25% of their time because they can understand the lineage and how the data is used. Analytics teams also saw about a 25% productivity improvement because they can find and validate data much faster.”

These gains reflect less time spent on research and duplication because teams leverage shared lineage and data context to build and maintain pipelines more efficiently (Table 3, below).

→ **Table 3**
Data Engineering Team Productivity Gains

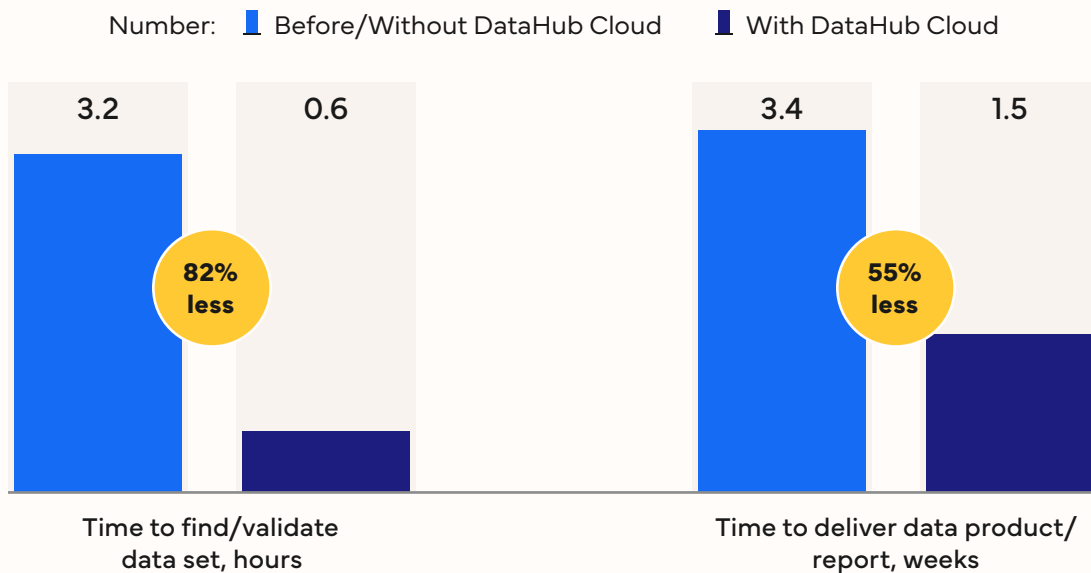
Gains	Before/Without DataHub Cloud	With DataHub Cloud	Difference	Benefit
Equivalent productivity level, FTEs	120	140	20	17%
Equivalent value of team productivity per organization	\$11.96M	\$13.98M	\$2.02M	17%

n = 5; Source: IDC Business Value In-Depth Interviews, February 2026

Study participants also attributed improvements in data analytics–related KPIs to their use of DataHub Cloud, including an 82% reduction in the time required to find and validate a data set and a 55% decrease in the time needed to deliver analytics outputs (Figure 4, below). By accelerating both discovery and delivery, DataHub Cloud enables analytics teams to respond more quickly to business needs while increasing confidence in their data choices.

→ **Figure 4**
Impact on Data Analytics KPIs

See the figure data in an [accessible table format](#).



n = 5; Source: IDC Business Value In-Depth Interviews, February 2026

These benefits contribute to average productivity gains of 18% for interviewed organizations’ data analytics teams (Table 4, next page). These productivity gains are driven by improved overall data ecosystems, but especially faster data set validation and reduced friction in developing reports and data products.

Table 4
Data Analytics Team Productivity Gains

Gains	Before/Without DataHub Cloud	With DataHub Cloud	Difference	Benefit
Equivalent productivity level, FTEs	74	87	13	18%
Equivalent value of team productivity per organization	\$5.18M	\$6.09M	\$914,400	18%

n = 5; Source: IDC Business Value In-Depth Interviews, February 2026

Data Quality Benefits

Study participants observed that DataHub Cloud has improved data quality by making quality signals, ownership, and lineage more visible and actionable across data assets. By consolidating metadata and mapping dependencies, teams can more quickly identify the source of issues, understand downstream impacts, and prioritize remediation based on business criticality. Several organizations also noted that richer context supports more consistent data documentation practices, improving how data is curated and consumed over time.

In addition, DataHub Cloud supports greater confidence in data for analytics and AI/ML by reducing gaps in lineage coverage and improving the completeness of key metadata fields. Interviewed organizations described using the DataHub platform to ingest or surface monitoring and quality status information alongside data sets, making it easier for users to determine whether data is reliable and up to date. This visibility helps prevent data incidents from propagating unnoticed and strengthens governance processes that depend on clear ownership and traceability.

Interviewees shared examples illustrating these data quality benefits:

Fewer data quality issues, better data completeness and timeliness:

“With DataHub Cloud, we’ve reduced the number of data quality issues by around 60%. Data completeness and timeliness issues have also improved significantly because data lineage and metadata are much better defined.”

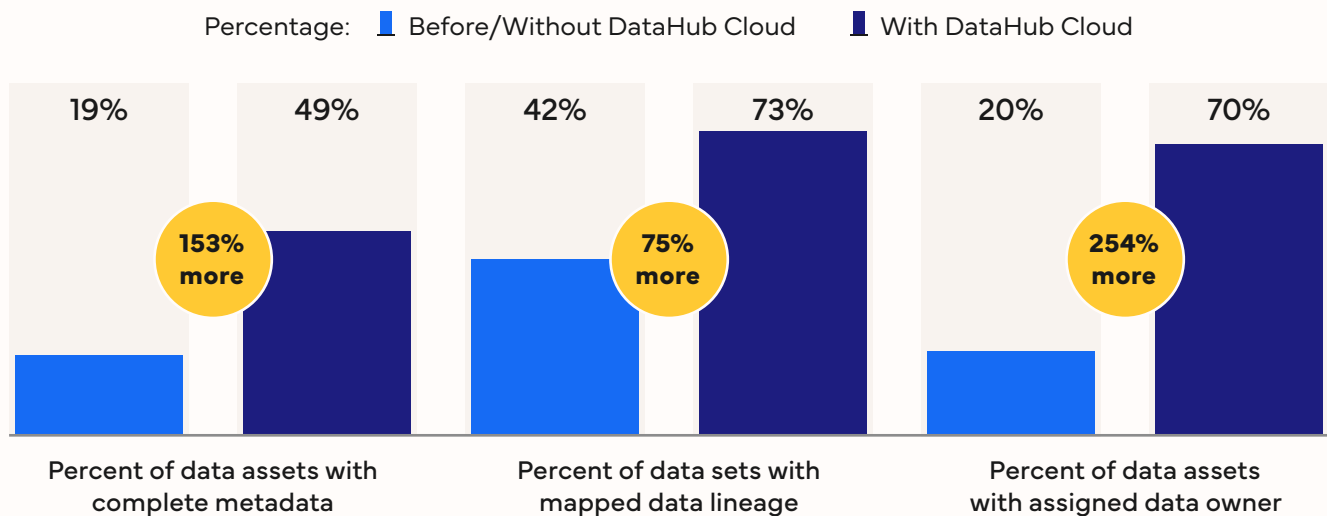
Better data quality due to visibility:

“Our data quality is higher with DataHub Cloud. We can now ingest the data quality status for our assets more easily, which makes it much clearer whether the data is reliable. That visibility into data quality is very important for how people use the data.”

Study participants reported achieving notable gains in the share of data assets with complete metadata (153% increase) and mapped lineage coverage (75% increase) with DataHub Cloud (**Figure 5, below**) while also increasing the share of assets with an assigned owner by more than 3.5 times. These improvements strengthen trust and accountability by making it easier to understand data provenance, ownership, and expected quality.

→ **Figure 5**
Impact on Data Quality KPIs

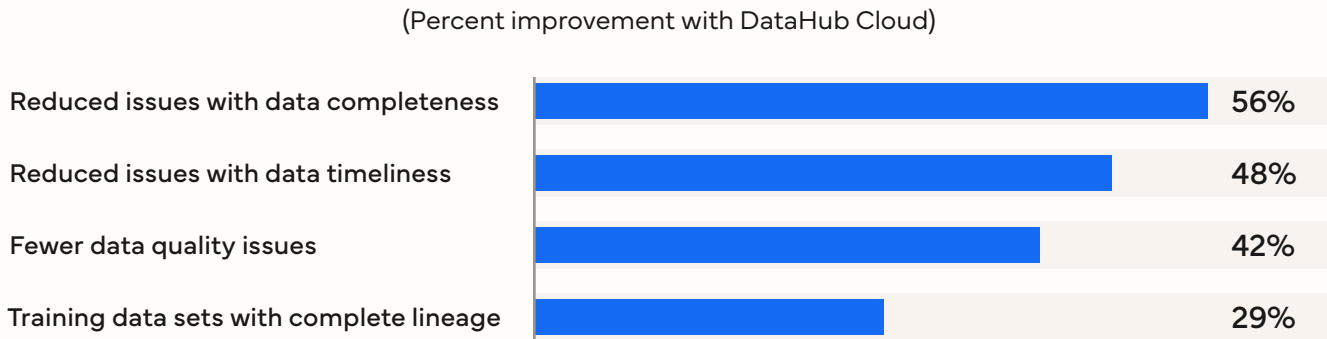
See the figure data in an [accessible table format](#).



n = 5; Source: IDC Business Value In-Depth Interviews, February 2026

Study participants reported measurable reductions in friction stemming from data quality-related issues, including 42% fewer data quality issues overall and 48% fewer issues related to data timeliness. They also cited 56% fewer data completeness issues and a 29% improvement in training data sets with complete lineage, supporting more reliable analytics and AI/ML outcomes (**Figure 6, next page**).

Figure 6
Impact on Other Data Quality KPIs



n = 5; Source: IDC Business Value In-Depth Interviews, February 2026

Importantly, study participants linked the centralized data ownership and enhanced lineage visibility that DataHub Cloud enables to helping their governance teams operate more consistently at scale without proportional increases in effort. One interviewed DataHub customer explained: *“One of the key outcomes of using DataHub Cloud has been introducing data governance into our company. We didn’t really have anything before that monitored all our data sources or accumulated signals to tell us whether everything in a data source was in good shape or if there were problems to address.”*

On average, study participants reported that their data governance teams are 20% more efficient with DataHub Cloud ([Table 5, below](#)).

→ **Table 5**
Data Governance Team Efficiencies

Efficiencies	Before/Without DataHub Cloud	With DataHub Cloud	Difference	Benefit
Equivalent productivity level, FTEs	70	56	14	20%
Equivalent value of team productivity per organization	\$4.91M	\$3.94M	\$976,700	20%

n = 5; Source: IDC Business Value In-Depth Interviews, February 2026

Benefits to AI Operationalization

For study participants using DataHub Cloud to support their AI and ML initiatives, benefits include stronger model outcomes driven by improved visibility into upstream data quality, lineage, and ownership. These factors directly influence model reliability and production readiness. Expanded lineage coverage and clearer metadata help ML teams understand how data flows into training data sets and models, enabling earlier identification of upstream issues and reducing downstream performance degradation. By consolidating data and AI context within a centralized platform, organizations improve collaboration between data engineering and ML teams and advance initiatives more consistently from experimentation to production.

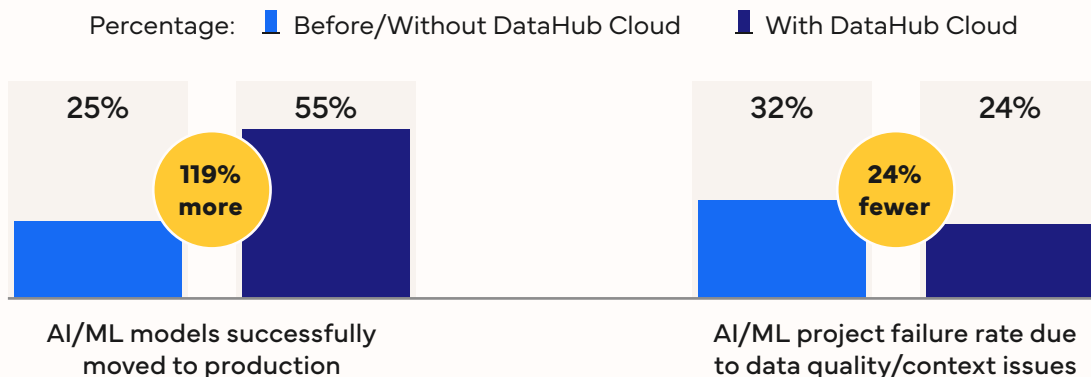
One interviewed DataHub customer commented:

“DataHub Cloud is helping our machine learning teams understand how upstream data quality issues impact their model performance. We now understand how most data incidents affect downstream data, and that has been helpful in multiple cases this year.”

As **Figure 7 (below)** shows, study participants increased the number of AI/ML models successfully moved into production by 119% and reduced AI/ML project failure rates by 24%. Given that AI and ML initiatives are often central to business models, improving success rates with DataHub Cloud has the potential to generate substantial value through higher revenue and lower operational costs, beyond the efficiencies gained in developing and delivering these initiatives.

→ **Figure 7**
Impact on AI Project Outcomes

See the figure data in an [accessible table format](#).



n = 5; Source: IDC Business Value In-Depth Interviews, February 2026

Business and Operational Benefits

Interviewed organizations described broader business and operational benefits achieved with DataHub Cloud through reduced data friction and improved visibility into how data products are built and used. By enabling self-service discovery and faster validation of trusted data, stakeholders make decisions faster, and teams spend less time waiting on data specialists. One study participant explained: *“There is less support needed from the data team with DataHub Cloud because people can use self-service for their questions. This removes bottlenecks, saves time for the data team, and allows stakeholders to make decisions faster without waiting for answers.”* Interviewees also noted that a centralized catalog and lineage context provide a consistent foundation for enterprisewide analytics.

Study participants highlighted operational continuity improvements with DataHub Cloud. It helps teams understand changes, issues, and downstream impact, enabling faster root-cause analysis and more targeted communication to affected stakeholders. This is especially important where compliance obligations and customer-facing SLAs depend on reliable data products and demonstrable control over dataflows.

One interviewed DataHub customer noted:

“Whenever we have an outage or some kind of incident, DataHub Cloud comes in handy for debugging issues. We use it to understand how things broke, what other things are impacted, and how to minimize disruption to everyone’s workflow.”

Study participants emphasized that faster diagnosis and clearer impact assessment → **reduce disruptions, limit unplanned downtime, and maintain continuity. On average, they reduced data-related outages by 48% and resolved them 58% faster with DataHub Cloud.**

Beyond operational resilience, DataHub Cloud also strengthens organizations’ ability to manage compliance obligations more efficiently. Interviewees explained that centralized metadata, ownership assignment, and lineage visibility provide a clearer record of how data is sourced, transformed, and consumed across the enterprise, which is especially important in regulated industries. By establishing a single source of truth for data context and monitoring signals, organizations can demonstrate stronger control over dataflows, support audit processes more effectively, and maintain certifications that are critical to customer trust.

Interviewed organizations spoke in detail about these compliance-related benefits:

Enables compliance certifications important to customers:

“If we can’t monitor our data products, that introduces a lot of security risk, which would directly impact our ability to keep our compliance certifications. Those certifications are important to our customers, and DataHub Cloud plays a key role in supporting them.”

Drives industry-specific analytics and required compliance capabilities:

“DataHub Cloud gives us enterprisewide analytics capabilities and supports compliance with regulations, which is critical in financial services. It provides a single focal point for business owners and helps eliminate issues caused by distributed and duplicated data across the organization.”

**Table 6
Compliance Team Efficiencies**

Efficiencies	Before/Without DataHub Cloud	With DataHub Cloud	Difference	Benefit
Equivalent team members required, FTEs	78	72	6	8%
Equivalent value of team time per organization	\$5.45M	\$5.02M	\$431,200	8%

n = 5; Source: IDC Business Value In-Depth Interviews, February 2026

From a broader business operations perspective, study participants reported that DataHub Cloud improves the effectiveness and productivity of their data pipeline development teams by reducing research time, preventing duplication, and clarifying dependencies before changes are made. With searchable lineage and standardized metadata, data engineers can more quickly understand how data sets are structured and used, enabling them to design and modify pipelines with greater confidence and fewer downstream disruptions. Interviewees described significant time savings in researching existing data models and validating assumptions, allowing teams to redirect their efforts toward innovation and higher-value development work. This increased efficiency accelerates analytics and product delivery timelines while strengthening overall business agility by ensuring data infrastructure evolves alongside changing operational and customer requirements. On average, participants attributed 13% productivity gains in their data pipeline development teams to DataHub Cloud (**Table 7, next page**).

Table 7
Data Pipeline Development Team Productivity Gains

Gains	Before/Without DataHub Cloud	With DataHub Cloud	Difference	Benefit
Equivalent productivity level, FTEs	263	299	35	13%
Equivalent value of team productivity per organization	\$26.34M	\$29.86M	\$3.52M	13%

n = 5; Source: IDC Business Value In-Depth Interviews, February 2026

Data Cost Benefits

In addition to the data- and business-related benefits already discussed, study participants reported that DataHub Cloud has helped reduce and optimize data and storage costs by increasing visibility into what data is being used, by whom, and for what purpose.

By improving discovery and lineage context, they can better identify redundant data sets and pipelines, rationalize storage and processing footprints, and prioritize retention for high-value assets. This visibility also supports governance-driven life-cycle management, such as identifying stale or orphaned data products that can be retired. Study participants provided concise statements about these types of data-related savings, such as:

"DataHub Cloud has helped us save on data storage costs by around 20%–25%, which is about \$250,000–\$300,000 per year," and *"DataHub Cloud has helped us identify data that was not used anymore, about 10%, which is saving us around \$40,000 per year."*

On average, interviewed organizations require 8% less storage due to these efficiencies.

Furthermore, organizations described leveraging DataHub Cloud to reduce the operational burden associated with maintaining data platforms and tooling. Greater self-service capability reduces the need for manual support and investigation, while clearer ownership and usage of information helps teams coordinate changes and avoid costly downstream remediation.

Challenges and Opportunities

Organizations evaluating DataHub Cloud should weigh both its capabilities and the operational changes required to realize value. For enterprises with limited governance maturity or largely manual documentation practices, adoption may require clearer ownership models, defined processes, and executive sponsorship. Without these foundations, organizations may not be able to fully leverage the offering's automation and metadata intelligence features.

In competitive replacement scenarios, buyers must also assess migration complexity, integration effort, and user retraining. Many prospects already operate incumbent catalog or governance tools, making differentiation dependent on demonstrable gains in productivity, data quality oversight, or cloud cost management rather than incremental feature expansion.

Cross-functional alignment can present an additional hurdle. Because DataHub Cloud spans discovery, observability, and governance, purchasing decisions often involve multiple stakeholders across data engineering, governance, and business teams. Building a cohesive business case tied to measurable operational or financial outcomes is therefore critical.

At the same time, market conditions create opportunities. Enterprises advancing cloud modernization, addressing redundant storage, or preparing AI initiatives for production may find increased value in platforms that unify metadata and automate governance workflows. Organizations facing decentralized data growth or post-merger integration challenges may be particularly receptive to a consolidated context management approach.

Conclusion

As data estates expand across hybrid and multicloud environments and AI initiatives move from experimentation to production, organizations face increasing pressure to improve visibility, governance, and operational discipline across their data assets. Fragmented documentation, limited lineage coverage, and inconsistent ownership models can slow analytics delivery, increase operational risk, and constrain the realized value of data and AI investments.

To help address these challenges, DataHub offers DataHub Cloud, a context management platform designed to unify discovery, observability, and governance across enterprise data environments. By centralizing technical and business metadata, mapping lineage, surfacing ownership and quality signals, and enabling self-service search, the platform aims to reduce friction in how data is found, validated, monitored, and operationalized.

Across the interviewed organizations, DataHub Cloud was associated with 91% faster data searches, more than three times higher search success rates, 17%–18% productivity gains for data engineering and analytics teams, 119% more AI/ML models moved to production, 24% lower AI/ML project failure rates, and up to a 25% reduction in Snowflake storage costs.

IDC's Business Value research indicates that interviewed organizations also achieved qualitative benefits. Beyond measurable KPIs, interviewees emphasized increased confidence in data choices, reduced debate over data set validity, improved cross-team collaboration, and stronger support for compliance requirements. ●

Appendix A: Methodology

IDC utilized its standard Business Value methodology for this project, gathering data from organizations currently using DataHub Cloud to discover, understand, govern, and monitor their data assets as the foundation for the model. IDC collected quantitative benefit information during the interviews using a before-and-after assessment of the impact of using DataHub Cloud. For this study, IDC interviewed five organizations that are using DataHub Cloud.

IDC bases its financial calculations on several assumptions, which include: Multiplying time values by burdened salary (salary + 28% for benefits and overhead) quantifies efficiency and productivity savings. For this study, IDC used its standard salary assumptions of an average fully loaded IT salary of \$100,000 per year and an average fully loaded non-IT salary of \$70,000 per year. IDC assumes that employees work 1,880 hours per year (47 weeks x 40 hours).

Note: All numbers in this document may not be exact due to rounding.

Appendix B: Accessible Data Tables

This appendix provides an accessible version of the data for any complex figures in this document. Click "Return to figure" to get back to the original figure.

Figure 4 accessible data
Impact on Data Analytics KPIs

Impact	Time to Find/Validate Data Set, Hours	Time to Deliver Data Product/Report, Weeks
Before/Without DataHub Cloud	3.2	3.4
With DataHub Cloud	0.6	1.5
Difference	82% less	55% less

n = 5; Source: IDC Business Value In-Depth Interviews, February 2026

[Return to figure](#)

Figure 5 accessible data
Impact on Data Quality KPIs

Impact	Percent of Data Assets with Complete Metadata	Percent of Data Sets with Mapped Data Lineage	Percent of Data Assets with Assigned Data Owner
Before/Without DataHub Cloud	19%	42%	20%
With DataHub Cloud	49%	73%	70%
Difference	153% more	75% more	254% more

n = 5; Source: IDC Business Value In-Depth Interviews, February 2026

[Return to figure](#)

Appendix B: Accessible Data Tables (continued)

Figure 7 accessible data
Impact on AI Project Outcomes

Impact	AI/ML Models Successfully Moved to Production	AI/ML Project Failure Rate Due to Data Quality/Context Issues
Before/Without DataHub Cloud	25%	32%
With DataHub Cloud	55%	24%
Difference	119% more	24% fewer

n = 5; Source: IDC Business Value In-Depth Interviews, February 2026

[Return to figure](#)

About the IDC Analysts



Lynne Schneider

Research Director, Data Collaboration and Monetization,
and Location and Geospatial Intelligence, IDC

Lynne Schneider is a research director leading IDC's Data Collaboration and Monetization, and Location and Geospatial Intelligence market research and advisory practices. Schneider's core research coverage in DaaS includes data sourcing and delivery services of traditional and emerging data providers and evolving data aggregation and dissemination platforms. The breadth of coverage includes services that enable an organization to externally monetize data generated as part of the organization's ongoing operations, value-added information derived from this data, and the marketplace for combining data with other solutions. This research analyzes the supply-and-demand-side business and technology trends of this emerging category. The Location and Geospatial Intelligence coverage includes data providers, location and geospatial analytics and visualization, mapping, navigation, geospatial developer tools and platforms, and other supporting technologies.

[More about Lynne Schneider →](#)



Matthew Marden

Research Vice President, Business Value Strategy Practice, IDC

Matthew Marden is responsible for carrying out custom business value research engagements and consulting projects for clients in several technology areas, focusing on determining the return on investment of their use of enterprise technologies. Marden's research often analyzes how organizations are leveraging investment in digital technology solutions and initiatives to create value through efficiencies and business enablement.

[More about Matthew Marden →](#)

Message from the Sponsor



DataHub Cloud is an enterprise context platform that transforms enterprise data into trusted context, enabling intelligent decision making by humans and AI agents.

Built on the #1 open-source metadata platform trusted by 3,000+ organizations including Netflix, Visa, and Apple, DataHub Cloud connects 100+ data sources with real-time lineage, automated quality monitoring, and AI-powered discovery at machine scale. Its open, extensible architecture adapts to any data stack, enabling enterprises to scale governance and operationalize AI initiatives with confidence. With Ask DataHub conversational discovery and the proprietary AI Fitness Score, teams ensure data is model-ready before AI projects reach production. Backed by a 99.5% uptime SLA and SOC II compliance, DataHub Cloud meets the security and reliability standards enterprises demand.

For more information, visit www.datahub.com

IDC Custom Solutions

IDC Custom Solutions produced this publication. The opinion, analysis, and research results presented herein are drawn from more detailed research and analysis that IDC independently conducted and published, unless specific vendor sponsorship is noted. IDC Custom Solutions makes IDC content available in a wide range of formats for distribution by various companies.

This IDC material is licensed for external use and in no way does the use or publication of IDC research indicate IDC's endorsement of the sponsor's or licensee's products or strategies.



[idc.com](https://www.idc.com)

[in @idc](https://www.linkedin.com/company/idc)

[X @idc](https://twitter.com/idc)

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications, and consumer technology markets. With more than 1,300 analysts worldwide, IDC offers global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries. IDC's analysis and insight help IT professionals, business executives, and the investment community make fact-based technology decisions and achieve their key business objectives.

©2026 IDC. Reproduction is forbidden unless authorized. All rights reserved. [CCPA](#)