

Accelerate Machine Learning And The Path To AI With An End-To-End Data Lakehouse

In May of 2023, Cloudera, Intel, and HPE commissioned Forrester Consulting to evaluate how organizations are choosing the technologies that support the storage, management, and analysis of their proprietary data. Forrester conducted an online survey with 840 data practitioners and decision-makers at organizations in Australia, New Zealand, the UK, and the US to explore this topic. We found that while most organizations have begun their journey towards data environment transformation, they must prioritize the hybridization of their teams and the centralization of data lifecycle steps to reap benefits in productivity and insight generation. In fact, 75% of respondents acknowledged that they could save four or more hours each day if the steps of the data lifecycle were integrated into a single platform.

KEY FINDINGS

Data science teams — and data — are becoming more distributed across organizations. Data science teams are moving toward becoming more decentralized, frequently by adopting hybrid models of reporting into the business. At the same time, data decision-makers are adopting point solutions that meet immediate needs rather than making larger, strategic purchases that centralize their data and data management functions.

Environments with too many tools are costing practitioners valuable productivity time in their workday. Employees aren't frustrated with the performance of individual solutions, but rather the large number of tools used, difficulty activating machine learning models, and lack of effective integration.

End-to-end lakehouses reduce the complexity of the data environment and improve the employee experience. Adopting an end-to-end lakehouse that consolidates data tool functionality not only eases the stress of managing the full data lifecycle, but also provides data ownership clarity and saves valuable company resources.

Data Science Teams Are Becoming More Distributed, While Data Must Become More Consolidated

With an increase in distributed data across cloud environments, constantly evolving compliance and governance standards, and new threats to data security via an increasingly hybrid workforce, enterprises are responding by changing how they manage their data. The ability to perform the steps of the data lifecycle within a single platform will become increasingly critical to the success of enterprises that want to provide an excellent customer experience for customers; consolidation of data tools into a single platform — or at least an environment with less than eight tools — can reduce the time it takes to perform core data functions and reduce time to value and time to customer satisfaction.

- Data science teams are moving toward a hybrid or decentralized reporting model, but their environments aren't enabled for efficient data work.** Fifty-five percent of respondents said they report into a hybrid model with another 17% report into a strictly decentralized model. With this move towards hybridized decentralization, any move to simplify or consolidate data-related functions will be a boon for productivity. Yet, data decision-makers are still purchasing tools with immediate needs in mind rather than long-term, strategic ones. In fact, 59% of respondents indicated that they select tools based on immediate needs. To effectively enable employees, data decision-makers will need to adopt a more holistic approach to their data lifecycle (see Figure 1).

Data scientists use eight or more tools to complete each step of the data lifecycle.

Figure 1
Number Of Tools Used For Each Step In The Data Lifecycle

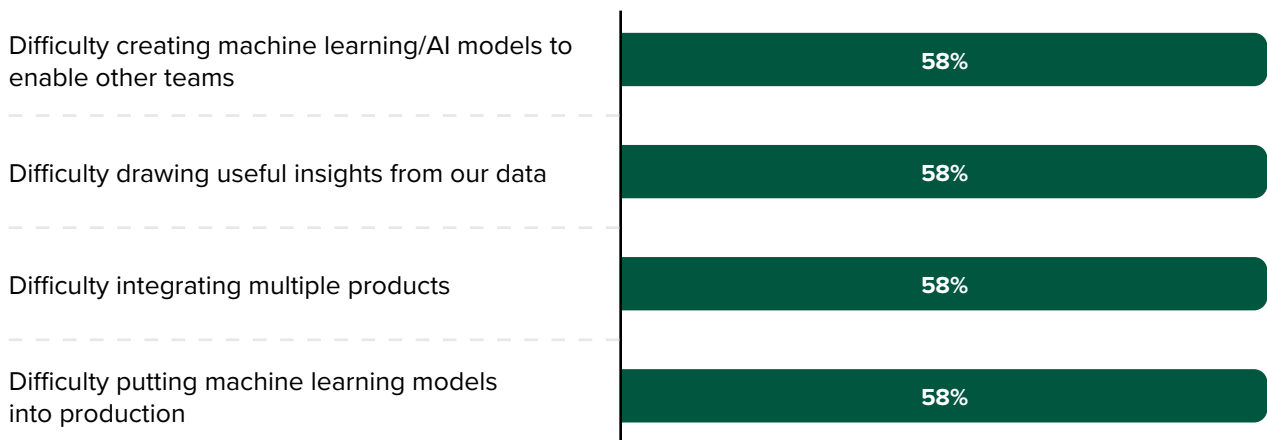
STEP IN DATA LIFECYCLE	MEAN NUMBER OF TOOLS
Ingesting/streaming	8.3
Data preparation/engineering	9.2
Analyzing	9.2
Predicting for machine learning	9.2
Publishing into the business	10.4

Base: 840 global practitioners and decision-makers in development and data science
 Source: A commissioned study conducted by Forrester Consulting on behalf of Clouder, Intel and HPE, May 2023

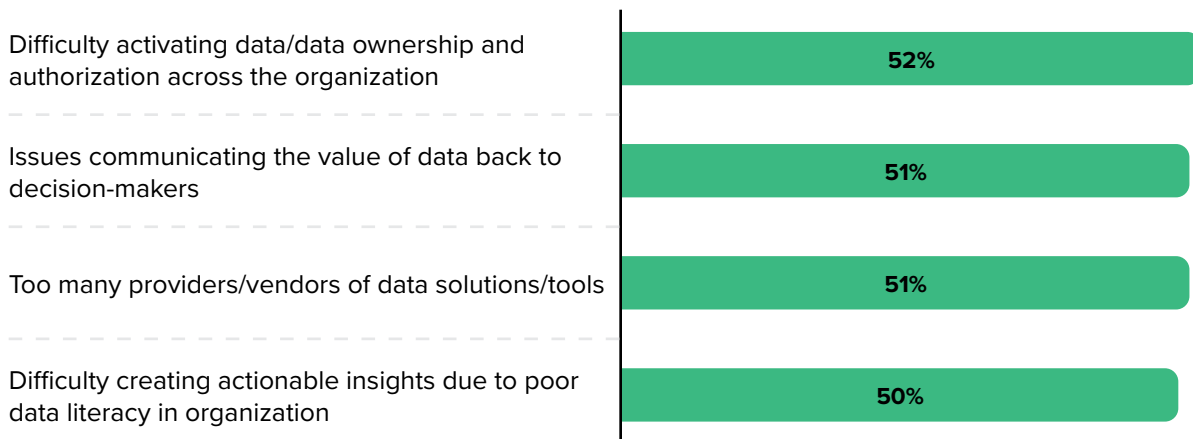
- Data science teams are stuck switching between too many tools to perform core job functions.** The highest total ranked technical challenges were related to enabling other teams with ML models (58%), difficulty drawing useful insights from data (58%), difficulty integrating multiple products (58%), and difficulty putting machine learning models into production (58%). Additionally, 51% of respondents faced difficulty activating their data and understanding data ownership across their organization (see Figure 2). These challenges also naturally lead to revenue implications; respondents noted increased time to value (56%), as well as increased technical debt (54%) eating into their profits.

Figure 2

Technical Challenges Relating To Data Projects



Organizational Challenges Relating To Data Projects



Base: 840 global practitioners and decision-makers in development and data science

Note: Showing only top 4 challenges

Source: A commissioned study conducted by Forrester Consulting on behalf of Cloudera, Intel and HPE, May 2023

Reduce Complexity, Increase Profitability With An End-To-End ML

This action of adopting more and more point solutions to meet immediate business needs can only subsist for so long. As data practitioners get bogged down from switching between platforms to perform individual tasks, they will continue to be less productive over time. Consolidating the organization's data lifecycle — especially when it comes to machine learning tasks — into a single end-to-end lakehouse will increase productivity, reduce time spent context switching, and increase employee satisfaction.

- **An end-to-end lakehouse solution does more than just provide additional data management capabilities.** An end-to-end lakehouse can level up data quality to improve collaboration across teams (57%), improve the ability to communicate ML results to the business (58%), improve the ability to communicate the value of data back to decision-makers (56%), and provide clear data ownership across the organization (54%). This addresses issues with data activation, as well as making machine learning work for the business. Respondents also noted looking for improved security from their new data solutions (44%).
- **A lakehouse that improves collaboration, productivity, and data security is good for the bottom line.** Respondents noted their businesses experienced positive benefits after investing in an end-to-end lakehouse, including scaling data management and insight practices alongside business growth (53%), improved attribution of success to data efforts (50%), better predictive analytics capabilities via ML (49%), improved collaboration across the organization (48%), and unified security across applications/data sources (48%). An end-to-end lakehouse can improve employee productivity, collaboration across the organization, and the ability to attribute credit for success where it is due, all while keeping data safe. While data decision-makers are currently focused on meeting the immediate needs of the business, sometimes a strategic approach can meet current and future needs (see Figure 3).

Figure 3

Expected Business Effects Of Having An End-to-End Data Lakehouse



Base: 840 global practitioners and decision-makers in development and data science
Source: A commissioned study conducted by Forrester Consulting on behalf of Cloudera, Intel and HPE, May 2023

METHODOLOGY

Clouder, Intel and HPE commissioned this study to understand the evolving adoption patterns of data storage and management solutions at enterprise organizations.

To achieve these objectives, Forrester conducted an online survey with 840 practitioners and decision-makers at organizations in Australia, New Zealand, the UK, and the US.

SUMMARY OF RESULTS FROM THE CLOUDERA, INTEL AND HPE-COMMISSIONED THOUGHT LEADERSHIP STUDY, “INCREASE EFFICIENCY IN THE DATA LIFECYCLE FROM INGESTION THROUGH AI VIA AN END-TO-END LAKEHOUSE”

To read the full results of this study, please refer to the Thought Leadership Paper commissioned by Clouder, Intel and HPE titled, “Increase Efficiency In The Data Lifecycle From Ingestion Through AI Via An End-to-End Lakehouse.”

Project Team:

Madeline Harrell,
Market Impact Consultant

Kate Pesa,
Associate Market Impact Consultant

Contributing Research:

Forrester’s [Technology Architecture & Delivery](#) research group

ABOUT FORRESTER CONSULTING

Forrester Consulting provides independent and objective research-based consulting to help leaders succeed in their organizations. Ranging in scope from a short strategy session to custom projects, Forrester’s Consulting services connect you directly with research analysts who apply expert insight to your specific business challenges. For more information, visit forrester.com/consulting.

© Forrester Research, Inc. All rights reserved. Unauthorized reproduction is strictly prohibited. Information is based on best available resources. Opinions reflect judgment at the time and are subject to change. Forrester®, Technographics®, Forrester Wave, RoleView, TechRadar, and Total Economic Impact are trademarks of Forrester Research, Inc. All other trademarks are the property of their respective companies. For additional information, go to forrester.com. [E-57144]