Forrester Consulting Thought Leadership Checklist
Commissioned By Clouder, Intel and HPE
AUGUST 2023

FORRESTER®

The Data Decision-Maker's Checklist For Improving ML/AI Operations With An End-To-End Lakehouse

Organizations can experience great benefits from adopting an end-to-end data lakehouse, including reduced efforts in data integration, a smoother pathway to ML/AI, accelerated time to value for advanced analytics, and improved governance and productivity through enhanced collaborations. However, a critical question arises for IT and business decision-makers: How can they justify investing in yet another data product when their technology stack already contains numerous tools, sometimes eight or more tools per step in the data lifecycle? That's where an end-to-end data lakehouse comes in. Here are the steps enterprises can take to benefit from an end-to-end data lakehouse:

1. **Work to align end-to-end data lakehouse adoption with technology upgrades and business growth objectives.**

Data decision-makers can leverage technology upgrades, digital transformation, or data migrations as an opportunity to introduce an end-to-end data lakehouse. While planning for infrastructure or software upgrades, emphasize the compatibility and integration benefits of a data lakehouse architecture. Showcase how it can leverage the latest technologies, such as cloud-based storage, distributed computing, and real-time processing, to optimize

data management and analysis capabilities. It is important to emphasize that solving technological challenges is not the sole value proposition. Businesses are seeking to foster a data-driven culture, enhance operational efficiency, improve customer experience, and ultimately drive revenue growth. By aligning the adoption of an end-to-end data lakehouse with both technology upgrades and overarching business objectives, organizations can realize the full potential of their data assets.

2. Begin to associate an end-to-end lakehouse with breaking down data silos and the productivity of data scientists.

Organizations face the challenge of fragmented data across different systems and applications. Data scientists spend countless hours integrating these multiple data sources, products, and platforms to gain a holistic view of operations, customer behavior, or market trends. A key benefit of the end-to-end lakehouse is seamless data integration and promoting interoperability, effectively breaking down data silos. It not only addresses the complex nature of data integration but also boosts productivity. In fact, a staggering 75% of respondents have acknowledged that they could save more than 4 hours each day if the various stages of the data lifecycle were integrated into a single platform.

3. Consider a phased adoption approach.

Our research indicates that obstacles in lakehouse adoption include sunk costs in existing technologies (54%), a focus on immediate needs (55%), and a lack of budget (43%). A phased approach to implementing a data lakehouse along with a cost-benefit analysis to emphasize ROI could address these challenges. Acknowledge that a full-scale implementation may not be feasible or desirable for every organization. Instead, prioritize specific use cases or departments where the benefits are most significant.

4. Be sure to prove your organization's ROI with metrics.

Quantitatively calculate ROI through reduced time to insights and improved decision-making and qualitatively emphasize enhanced data intelligence to lead to revenue growth, cost optimization, and streamlined collaboration.

This incremental approach demonstrates the value of the lakehouse concept, leading to broader adoption across the organization.

75%

of decision-makers have acknowledged that they could save more than 4 hours each day if the various stages of the data lifecycle were integrated into a single platform.

5. Accelerate your organization's advanced analytics and machine learning use cases.

Advanced AI-powered analytics, predictive modeling, and machine learning algorithms require access to large volumes of data for training models or analyzing complex patterns. Additionally, robust data governance is crucial to ensure compliance and maintain high data quality, which is essential for validating the accuracy of results. A data lakehouse enables data governance by offering data lineage tracking, data access controls, and centralized data storage and management. It consolidates data from various sources, simplifying data quality issue identification and resolution. With scalability and flexibility, the data lakehouse supports advanced analytics, empowering data scientists and analysts to leverage diverse sources for sophisticated modeling and analysis.

6. Reduce tool proliferation in different stages of the data lifecycle.

Respondents indicated using more than eight tools for each stage of data lifecycle. This fragmented tool landscape can lead to inefficiencies, increased complexity, and potential data inconsistencies. Moreover, different tools may have their own learning curves, requiring additional training for users. By implementing a

comprehensive data lakehouse, organizations can streamline their data management processes, consolidate tools (eliminating need for separate data warehouses and data marts), and ensure a unified approach to data handling throughout its lifecycle. This consolidation not only simplifies operations but also enhances data consistency and reduces the risk of errors and inconsistencies across various stages of the data lifecycle.

> Using too many tools with varying levels of integration leads to difficulty meeting compliance/governance standards (63%) and increased vulnerability to cyberthreats (61%).

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.

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

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"

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.