



ESG BRIEF

The State of Data Lakes

Date: June 2021 **Author:** Mike Leone, Senior Analyst; and Josh Clark, Research Associate

ABSTRACT: As organizations strive to utilize more data, data lakes are increasingly becoming an attractive option with limitless potential. Data lakes enable organizations to unite disparate data silos and make data more accessible across the business by serving as a centralized repository or collection of data, regardless of shape, speed, or size. Organizations can then leverage a data lake to feed other data-centric tools or utilize tools that sit on top of a data lake to work with the data in-place, such as query optimization solutions that can minimize data movement while enabling improved processing and analysis. And the economic advantages cannot be understated as organizations increasingly leverage cost-effective cloud storage and minimize operating costs through the consolidation of infrastructures silos.

Overview

ESG recently completed a survey on AI initiatives and the underlying modern infrastructure stack. The survey was completed by 325 IT professionals at midmarket (i.e., 500 to 999 employees) and enterprise (i.e., 1,000 or more employees) organizations across North America (United States and Canada).¹ As part of the survey, respondents were asked about their data lakes, including questions about environment locations, technology integrations, current challenges, and consideration of data lake solutions.

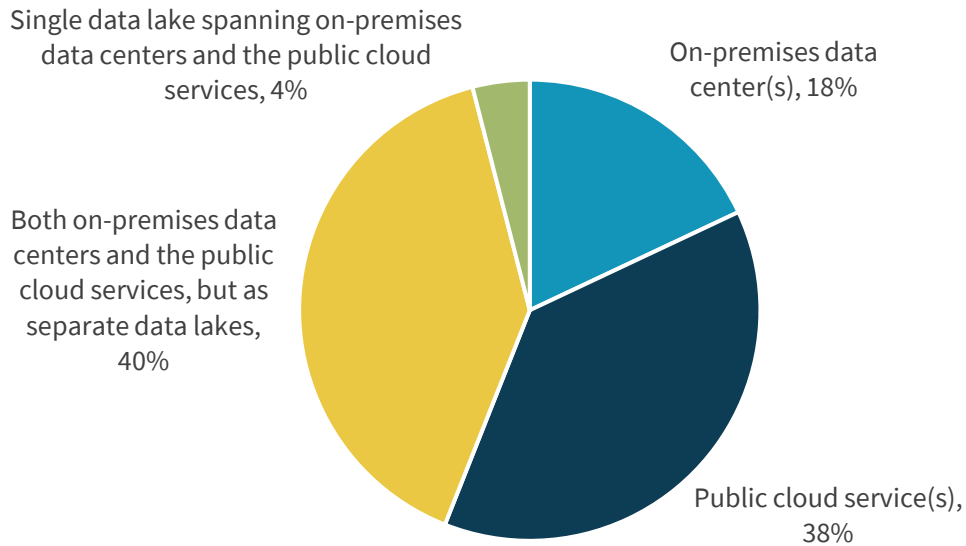
Data Lake Environments

As organizations embrace the cloud, data lakes are following suit. In fact, according to Figure 1, 78% of respondents said their data lake resides either on a public cloud service (38%) or in both on-premises data centers and public cloud services, but as separate data lakes (40%). Only 18% of organizations' data lakes reside only in on-premises data centers. One of the key components of a data lake is its ability to unify siloed repositories. While the data lake heritage began in on-premises environments with Hadoop, organizations continue to see the advantage of utilizing cloud environments to enable data lake success.

¹ Source: ESG Master Survey Results, [Supporting AI/ML Initiatives with a Modern Infrastructure Stack](#), May 2021.

Figure 1. Data Lake Environment Locations

Which of the following best describes the location(s) at which your organization’s data lake technology solution resides or will likely reside? In other words, where is it (or will it be) deployed? (Percent of respondents, N=258)



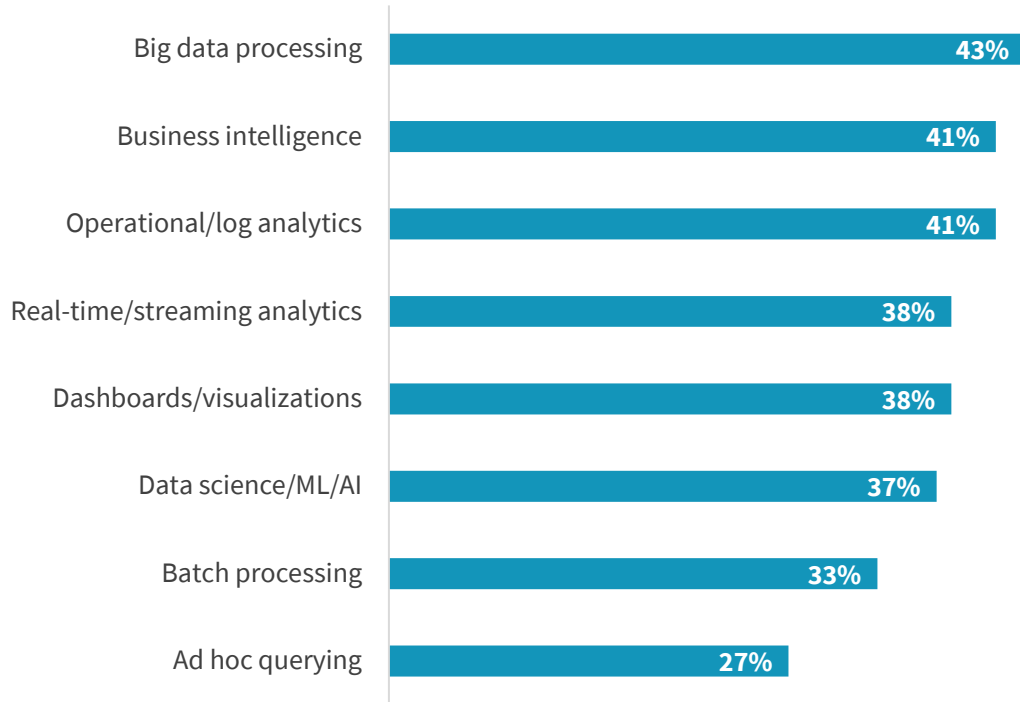
Source: Enterprise Strategy Group

Technology Interacting with Data Lakes

With all data in a unified repository, data lakes can enable greater access to data by more stakeholder across business units. Understanding business units across the enterprise leverage data in different ways through different tools, it is important to understand how organizations currently or plan to integrate those tools with a data lake. As shown in Figure 2, big data processing (43%), business intelligence (41%), and operational/log analytics (41%) are the three most likely technologies to interact with an organization’s data lake. But more importantly, the fact that there is not one technology that overwhelmingly interacts with a data lake more than another speaks to the wide range of use cases that can take advantage of a foundational data lake.

Figure 2. Tools Interacting with Data Lakes

Which of the following interacts – or will likely interact – with your organization’s data lake? (Percent of respondents, N=258, multiple responses accepted)



Source: Enterprise Strategy Group

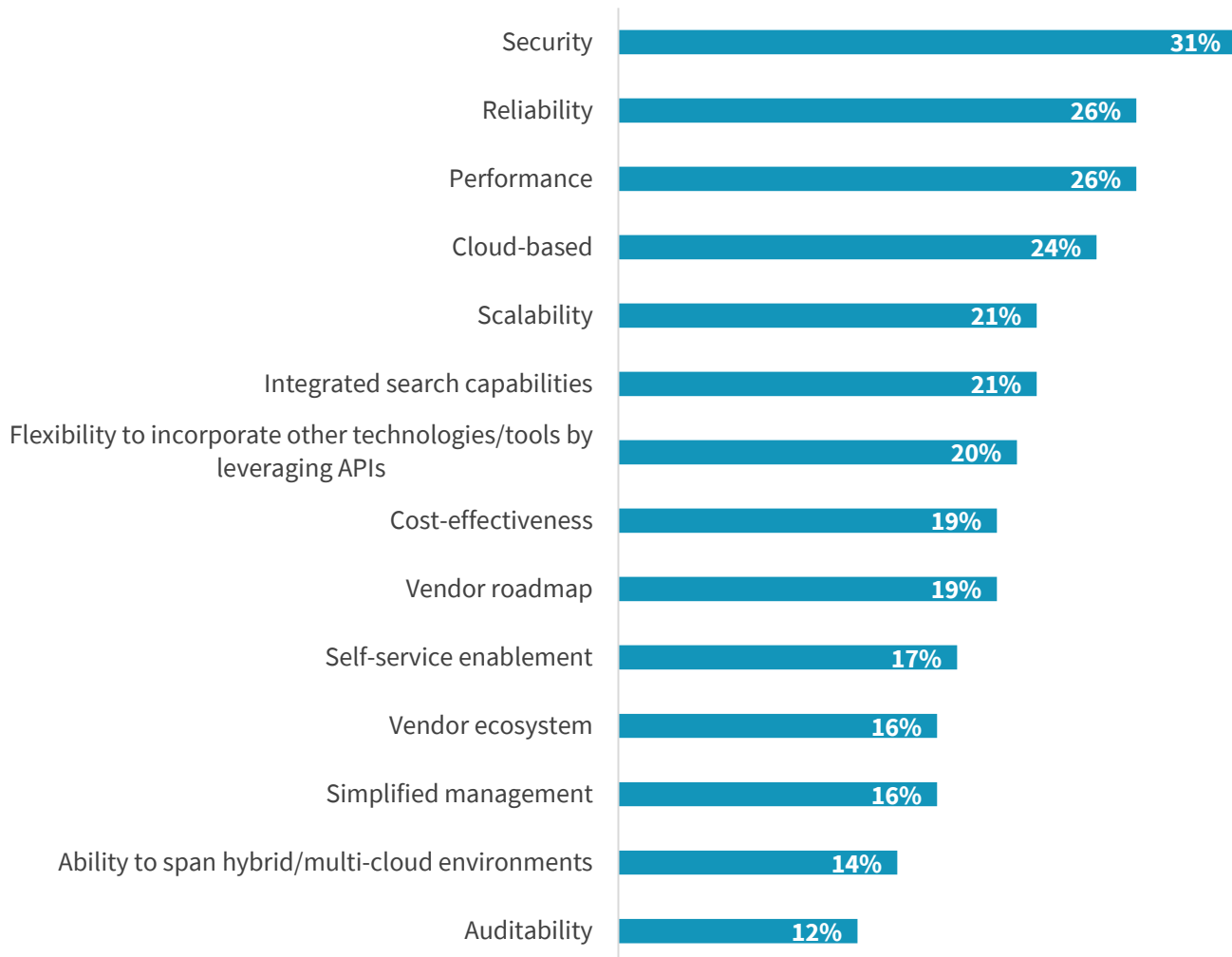
Data Lake Challenges and Considerations

At the heart of a data lake is the underlying storage infrastructure. As data continues to grow, and more end-users want more access to it, strain on the storage infrastructure is forcing organizations to reevaluate how they manage data. In fact, of all the data lake environment challenges organizations experience today, ESG research shows the greatest challenge is the management, optimization, and automation of data placement. Other top challenges experienced by at least 20% of organizations include data security/governance (23%), making data available to a range of BI/analytics tools (22%), longer deployment/provisioning times (22%), ensuring data protection (21%), and data migration (21%).

Though several challenges exist when it comes to the underlying infrastructure supporting a data lake, organizations continue to look for data lake solutions and approaches that can address those challenges head-on, understanding the value an organization can achieve when implementing a data lake the right way. According to Figure 3, security (31%), reliability (26%), performance (26%), and being cloud-based (24%) are the most important considerations when it comes to a data lake technology solution.

Figure 3. Security, Reliability, and Performance Top the List of Data Lake Considerations

What are the most important considerations when it comes to a data lake technology solution? (Percent of respondents, N=258, three responses accepted)



Source: Enterprise Strategy Group

The Bigger Truth

As data-driven initiatives continue to be prioritized across organizations, they are focusing on technologies that can enable wider access to growing data sets by more stakeholders without breaking the bank. Data lakes increasingly fit the bill. And while challenges remain, technology vendors have prioritized their data lake offerings to simplify adoption, consolidate silos (both data and operational), and improve time to value. A data lake that is implemented correctly with the right technology integrations will provide organizations with a cost-effective approach to improve data-centricity. All personnel throughout the business will gain more reliable access to the right data, enabling them to ask more questions, collaborate on ideas, and walk away with useful and actionable insights.

All trademark names are property of their respective companies. Information contained in this publication has been obtained by sources The Enterprise Strategy Group (ESG) considers to be reliable but is not warranted by ESG. This publication may contain opinions of ESG, which are subject to change. This publication is copyrighted by The Enterprise Strategy Group, Inc. Any reproduction or redistribution of this publication, in whole or in part, whether in hard-copy format, electronically, or otherwise to persons not authorized to receive it, without the express consent of The Enterprise Strategy Group, Inc., is in violation of U.S. copyright law and will be subject to an action for civil damages and, if applicable, criminal prosecution. Should you have any questions, please contact ESG Client Relations at 508.482.0188.



Enterprise Strategy Group is an IT analyst, research, validation, and strategy firm that provides market intelligence and actionable insight to the global IT community.



www.esg-global.com



contact@esg-global.com



508.482.0188