Gaining Business Advantage from Cloud Data Lakes
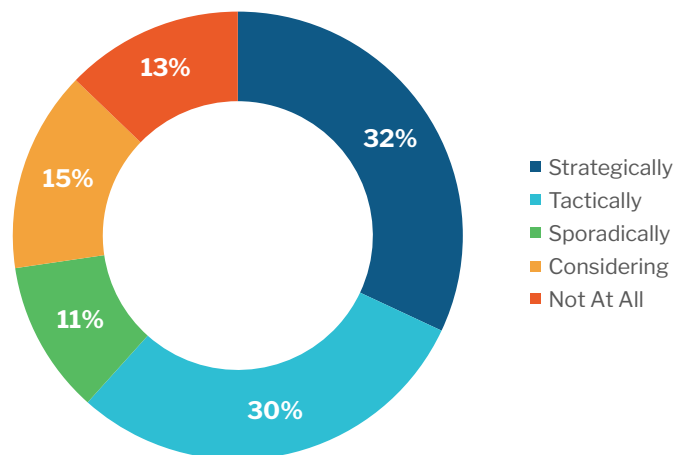
The 451 Take

It is now 10 years since the term 'data lake' was coined by the then CTO of Pentaho, James Dixon, to describe a data management environment used to store raw data from multiple sources (including unstructured data) that could be accessed by multiple users for multiple purposes. In the decade that has followed, the concept of the data lake has evolved to encompass a larger ecosystem of tools and capabilities to manage the entire data processing and analytics lifecycle. In doing so, it has emerged as a key component in enterprise data management strategies.

The data lake is designed to deliver flexibility and business agility while enabling enterprises to enjoy economic advantages through the use of object storage (including cloud storage services) as the underlying data store. By enabling multiple analytics projects to be performed against data in a single environment, the data lake also reduces data migration and movement complexity, particularly in enabling analytic queries to be applied to both structured and semi-structured or unstructured data. As a result, the data lake enables the democratization of data through self-service access while diminishing the regulatory and compliance headaches of using multiple data platforms and reducing overall time to insight.

Generating business value from data lake projects can be easier said than done, however. Key challenges remain, including the need to align technologically driven proof-of-concept projects with longer-term business goals, integration with existing data and analytics infrastructure, and the operationalization of multiple workloads, along with data quality, data curation and data governance requirements.

Digital Leaders' Adoption of Data Lakes

Source: 451 Research's Voice of the Enterprise: Data & Analytics 1H 2019

- **Strategically** 32%
- **Tactically** 30%
- **Sporadically** 11%
- **Considering** 15%
- **Not At All** 13%

Despite these ongoing challenges, it is clear that data lakes have the potential to drive transformation efforts. Data from 451 Research's Voice of the Enterprise: Data & Analytics study indicates that data lake adoption is growing, particularly among enterprises that are more advanced in terms of initiatives to drive digital transformation. Nearly three-quarters of digital leaders (those that have a formal digital transformation strategy and are actively digitizing business processes and technologies) are using data lakes today, almost a third (32%) of which have adopted data lakes strategically.

Business Impact

INNOVATION ACCELERATION. The data lake has the potential to accelerate innovation by enabling enterprises to combine data from multiple sources (relational and non-relational databases, as well as stream processing engines) in multiple formats (structured and un/semi-structured) to facilitate the development of new products and services, improve customer engagement and increase competitive differentiation.

ALIGNMENT WITH BUSINESS GOALS. The potential success of any data lake initiative depends on the ability to align technical functionality with business goals, as well as in-house skills and existing infrastructure and data processing/analytics approaches. Cloud data lakes enable enterprises to start with business objectives and take advantage of a data platform that is inherently adaptable to accomplish the business goal.

AGILITY AND FLEXIBILITY. The data lake has the potential to help organizations evolve in response to rapidly changing business conditions and regulatory requirements. In particular, cloud services based on a combination of decoupled storage and compute, automation and metadata management can evolve rapidly in response to changing business requirements and technological innovation.

GOVERNANCE AND OPERATIONALIZATION. In order to make the concept a functional reality, enterprises need to invest in the governance and operationalization of data lake environments. Technologies such as data catalogs and universal semantic layers combined with modern data governance processes and culture can provide the guard rails that act as enablers for self-service data access and analytics.

QUALITY, SECURITY AND PRIVACY. As with any data-related project, it is important to consider data quality, data security and data privacy concerns. This is especially true for the data lake given that it is designed to enable analytics agility, while the underlying data storage does not necessarily come equipped with data quality and privacy functionality.

Looking Ahead

451 Research anticipates greater adoption of cloud data lake projects as enterprises adapt to a greater volume of applications and data residing in the cloud. While much of the early adoption of data lakes was driven by experiments to prove the technology, we foresee a shift toward more strategic business-driven projects that enable business change. These strategic projects involve the integration of data management, security and metadata management functionality from the outset, enabling enterprises to lower the gap between proof of concept and proof of value. Additionally, while many initial tactical deployments were do-it-yourself IT projects, enterprises should increasingly be looking for partners to accelerate the deployment of strategic data lakes that can drive faster innovation and growth through projects that enable the business to respond to competitive threats, improve efficiency and drive new business value.

HITACHI Inspire the Next

Cloud Data Lake Modernization solution from Hitachi Vantara delivers data and analytical model management, governance and self-service capabilities. Our secure, production-ready data platform can be quickly and efficiently deployed in hybrid, cloud or multicloud environments. Combining our experience with industry blueprints and automation-based accelerators, we have completed hundreds of cloud transformation programs, helping enterprises to increase business agility and improve the resilience of their IT landscapes. Click here to learn more.