

# Simplify your Data and Analytics Architecture for Faster Insights

Accelerate AI/ML Initiatives with Automated Data Ingestion



VENTANA  
RESEARCH

Sponsored by Qlik & Databricks



# Data and Analytics Requirements are Changing

Organizations face increasing competition and compressed time frames that require intelligent use of all available data. Reports of yesterday's sales and operational figures must be accompanied by analyses of a variety of structured and unstructured data, including streams of sensor data, social media posts, and data from many other internal and external sources.

Organizations also need to perform sophisticated artificial intelligence and machine learning (AI/ML) analyses on the data they collect and they must be able to do so in real time. Maintaining separate systems for each of these requirements makes it difficult to be agile and responsive in today's fast-moving markets. New data and analytics architectures are available that can **support all these requirements with a common platform.**

**Takeaway:** To be competitive, organizations require a modern data and analytics architecture.

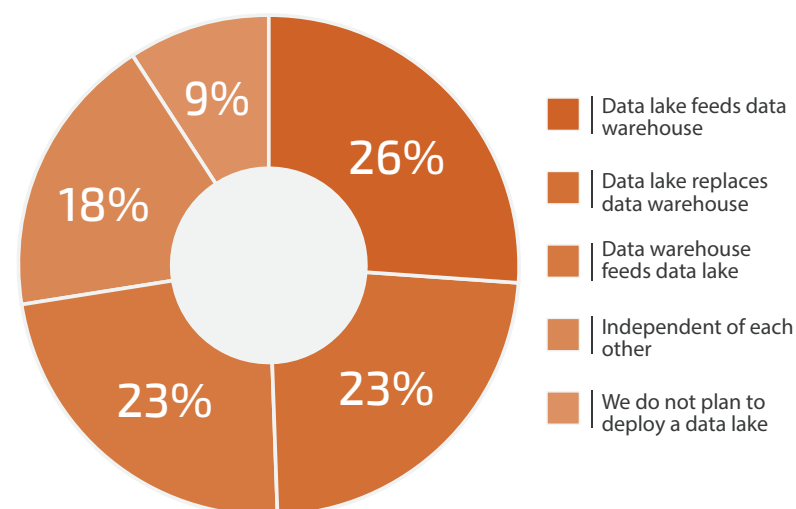
# Combine Data Warehouses and Data Lakes

Line-of-business functions have traditionally relied on data warehouses to provide information organized in a business-oriented format, ready for analysis. Recently, data lakes have emerged as a way to manage massive amounts of data, including unstructured data. Organizations need both these capabilities and they are often used together. In fact, nearly three-quarters **(72%)** of our research participants report that their **data lakes are associated with their data warehouses**. In one-quarter of the cases **(23%)**, the **data lake serves as a superset of these capabilities** where the data warehouse is built within the data lake, giving rise to the term “lakehouse.” This combination streamlines integration to reduce complexities, minimize the risk of inconsistencies and minimize any delay in accessing the information.

**Takeaway:** The emerging lakehouse architecture supports the entire data and analytic needs of your organization by combining data warehouses with data lakes.

## Data Lakes and Data Warehouses

A variety of approaches are popular



# Organizations Need Data at Scale

Regardless of their size, organizations need to capture and analyze massive amounts of information. Every event with every customer. Every website click. Every reading from every sensor. Every customer service interaction. Every log file. Publicly available data. Third-party market and demographic data. Years of history. All the raw data, not just aggregates.

A data platform that combines data warehouse and data lake capabilities provides this scale. Our research shows wide use of scalable data lakes across both customer-facing functions (marketing, customer service, sales) and internal functions (finance, research and development, supply chain and operations). With all this data it's no surprise that **the most common data preparation system requirement is the ability to process large volumes of data** (reported by **53%** of our research participants).

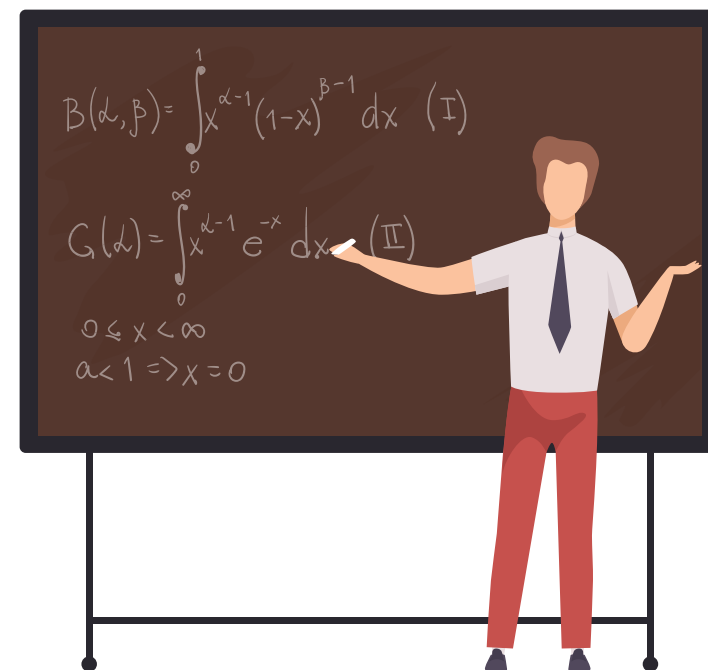
**Takeaway:** Streamline data architectures to manage the large volumes of data organizations require to be competitive today.



# Organizations Want Advanced Analytics

Data warehouses collect information that provides insights into the past. Data lakes, in conjunction with data warehouses, collect the information and provide the infrastructure needed for AI/ML-based predictive analyses that help anticipate what might happen in the future. This combination also supports processing unstructured information such as calls, emails and social media using text analytics. Our research shows approximately one-half **(51%) of organizations are working with unstructured data** in their AI/ML-based analyses, for instance to determine customer sentiment or perform advanced competitor analysis.

A common data and analytics architecture that supports AI/ML analyses addresses one of the most common challenges participants in our research face: deploying results of their AI/ML analyses into operations.

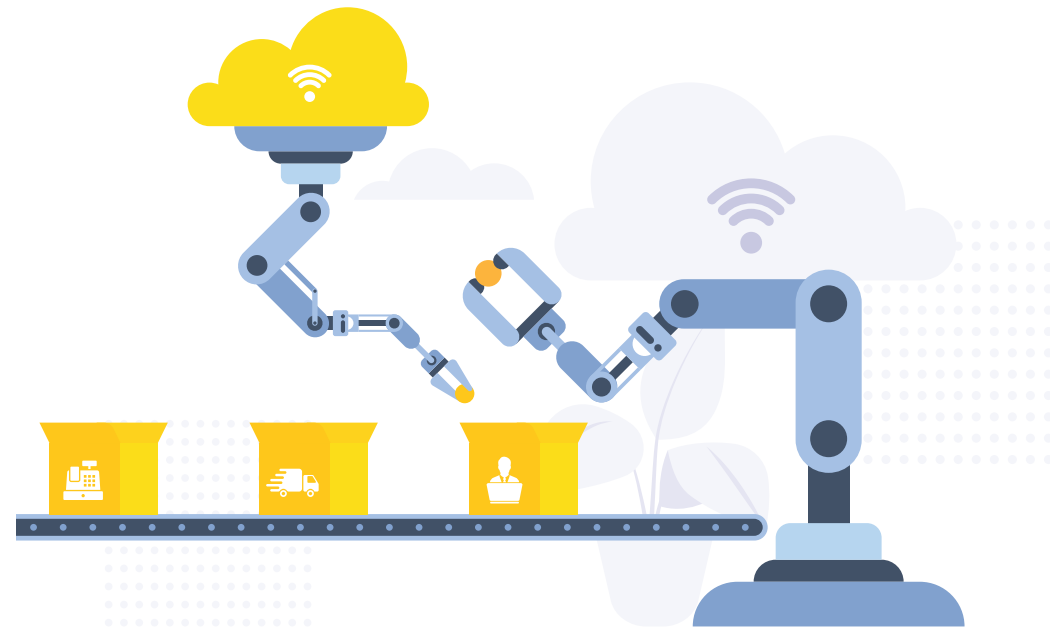


**Takeaway:** AI/ML-driven analytics is a must for modern data architectures to maximize the value of data.

# Robust Pipelines for Data Access and Prep

Our research shows that **accessing and preparing data to make it ready for analytics is one of the most time-consuming activities** for more than one-half (55%) of the participants. And two-thirds (67%) indicate **data is not accessible**. The combination of data lakes and data warehouses in a single platform can make this process easier since the data is brought together and is readily available for data engineers to transform into data warehouse structures.

Once data preparation pipelines are created, they need to be automated and scheduled so updated information is constantly available to the organization. Participants in our research identify this automation as one of the top three most important system-level capabilities they require.



**Takeaway:** Provide an architecture that supports data engineering and automated data pipelines on a single platform.

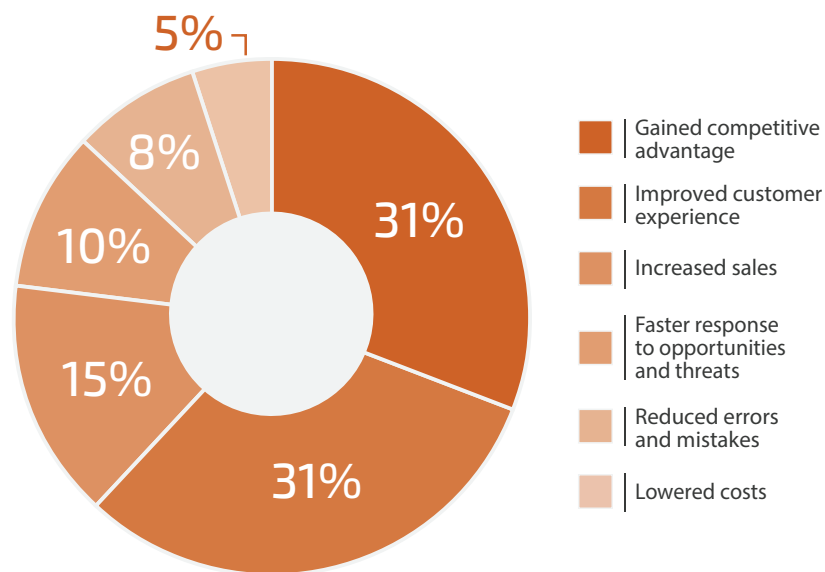
# Support AI/ML to Unlock Critical Insights

Artificial Intelligence and Machine Learning (AI/ML) are very rewarding for organizations. Most often, organizations report they achieved a competitive advantage and improved customer experience as a result of their ML efforts. They also report improving the bottom line due to increased sales and lower costs, and the ability to respond faster to opportunities and threats in the market. These benefits are significant enough that nearly three-quarters **(72%) plan to increase their use of AI/ML.**

To support their AI/ML efforts, organizations need to manage large amounts of detailed information not typically stored in a data warehouse. They also need additional data preparation capabilities to support data pipelines that feed AI/ML models and they need a data platform that goes beyond SQL to enable execution of the AI/ML models.

## Benefits of Machine Learning

Competitive advantage and customer experience top list



**Takeaway:** Data platforms should support data science workloads in addition to business intelligence.

# Responding in the Moment

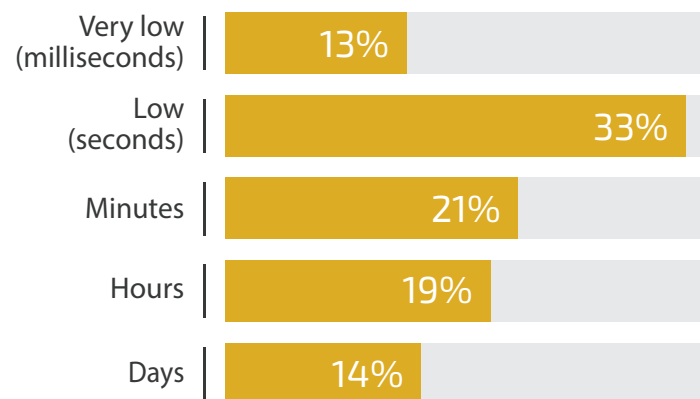
Data streams into organizations continuously. Therefore, data pipelines must be able to collect and process this information in real time. Nearly one-half (**47%**) of participants in our research report **it is essential to process some of their data in seconds or sub-seconds**. They must also be able to process their AI/ML models on this streaming information quickly enough that they are able to respond before the opportunity has passed. Nearly four in 10 (**38%**) have **automated outcomes to be able to respond in the moment**.

By combining the best qualities of data warehouses and data lakes organizations can provide a single approach for all major data workloads and support use cases from streaming analytics to BI, data science, and AI/ML.

**Takeaway:** Real-time insights require models applied to both complete and current data and are made easier by an architecture that combines these capabilities.

## Latency Requirements for Event Processing

Nearly one-half consider seconds essential





# Modern Information Architectures

Organizations are moving their data and analytic workloads to the cloud: nearly nine in 10 organizations **(86%) expect the majority of their data to be in the cloud.** Cloud deployments accelerate time to value, reduce complexity and increase agility.

But organizations also require flexibility to avoid being locked into a particular vendor or a particular cloud provider. Therefore, modern information architectures must support open standards for data storage, including object storage with standard file formats. They must support open standards for processing including not only SQL, but also the variety of frameworks required by data scientists like R, Python and other languages.



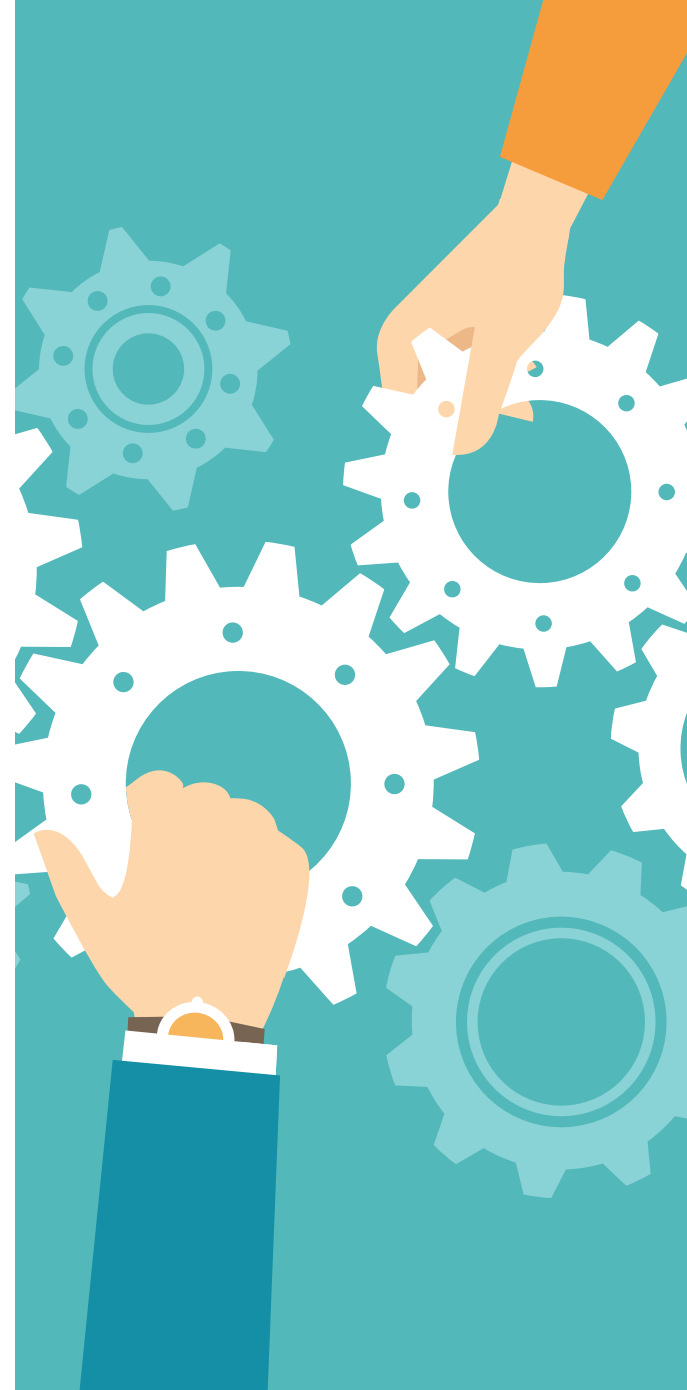
**Takeaway:** Use multi-cloud architectures based on open standards to ensure flexibility and agility.

# Combine Data Warehouses and Data Lakes to Accelerate Insights

Today's information architectures require many pieces, but disjointed data and analytics systems mean disjointed operations. Data engineers, data analysts and data scientists need to be working together on common systems. Data lakes should be integrated with data warehouses. Data pipelines should be automated for real-time processing. AI/ML analyses need to be made available throughout the organization. The fewer moving parts an organization has, the more nimble and responsive its processes will be. Modern data infrastructures based on open standards and delivered in multi-cloud environments bring all these elements together into a common architecture to accelerate insights and drive optimal performance in your organization.

Sponsored by   databricks

Ventana Research Dynamic Insights and Benchmark Research reports can be found at [www.ventanaresearch.com](https://www.ventanaresearch.com).



© 2021 Ventana Research