research

# PULSE

# Modernizing Your Data Architecture to Unlock Business Value

By David Stodder

Sponsored by:

aws Qlik

tdwi
Transforming Data With Intelligence™

TDWI PULSE REPORT

Q3 2021

# Modernizing Your Data Architecture to Unlock Business Value

By David Stodder

## Table of Contents

## About the Author

**David Stodder** is senior director of TDWI Research for business intelligence. He focuses on providing research-based insights and best practices for organizations implementing BI, analytics, data discovery, data visualization, performance management, and related technologies and methods and has been a thought leader in the field for over two decades. Previously, he headed up his own independent firm and served as vice president and research director with Ventana Research. He was the founding chief editor of *Intelligent Enterprise* where he also served as editorial director for nine years. You can reach him by email (dstodder@tdwi.org), on Twitter (@dbstodder), and on LinkedIn (linkedin.com/in/davidstodder)

## About TDWI Research

TDWI Research provides industry-leading research and advice for data and analytics professionals worldwide. TDWI Research focuses on modern data management, analytics, and data science approaches and teams up with industry thought leaders and practitioners to deliver both broad and deep understanding of business and technical challenges surrounding the deployment and use of data and analytics. TDWI Research offers in-depth research reports, commentary, assessments, inquiry services, and topical conferences as well as strategic planning services to user and vendor organizations.

## About TDWI Pulse Reports

This series offers focused research and analysis of trending analytics, business intelligence, and data management issues facing organizations. The reports are designed to educate technical and business professionals and aid them in developing strategies for improvement. Research for the reports is conducted through surveys of professionals. To suggest a topic, please contact TDWI senior research directors Fern Halper (fhalper@tdwi.org), James Kobielus (jkobielus@tdwi.org), and David Stodder (dstodder@tdwi.org).

## Acknowledgments

TDWI would like to thank many people who contributed to this report, including our report sponsors, who reviewed outlines and report drafts. Finally, we would like to recognize TDWI's production team: James Powell, Lindsay Stares, and Michael Boyda.

## Sponsors

Amazon Web Services and Qlik sponsored the research and writing of this report.

# The Pulse: Evolving Data Applications Drive Data Architecture Modernization

Organizations have data everywhere—in legacy applications, software-as-a-service (SaaS) platforms, e-commerce systems, web servers, mainframes, and, too often, in numerous business intelligence (BI) and data management silos. Digital transformation of older business processes and applications is generating new data at a rapid pace, challenging organizations that use traditional means of collecting, storing, and processing data. Additionally, users want to access external data sources to gain new insights and have complete, 360-degree views of customer behavior, supply-chain networks, market trends, risks and threats, and governance concerns. External data sources include third-party providers, public and industry databases, data marketplaces, and data exchanges.

The data explosion puts pressure on organizations to modernize their data architectures so data application users unlock value at the speed of business. Diverse data has the potential to ignite innovation; organizations increasingly want to be data-driven, not merely data-aware. They want data applications that unleash a growing range of nontechnical users to explore larger data sets and investigate data interactively to understand real-time situations, uncover root causes, analyze data relationships, and determine the best course of action.

Data architectures that depend on legacy, one-size-fits-all BI and data warehousing systems are falling short as use cases push beyond simple, carefully managed, and predefined data consumption and involve more ad hoc interaction. Older technologies and practices frustrate data-savvy business users who depend on IT data analysts and developers for every new type of analysis or visualization—anything that goes beyond preplanned use cases such as managed reporting.

**Speed and agility are critical.** With data application workloads becoming numerous, varied, and complex, organizations need a modernized data architecture that delivers faster and more agile paths to insight. Traditional systems limit users to carefully prepared and proscribed selections of data sets that IT personnel have transformed to fit a fixed, predefined schema. Transformation and preparation processes are slow; more than half (57%) of organizations surveyed by TDWI say it takes more than a day and up to a full month to add new data to BI and analytics platforms, and 14% say it takes longer than one month.[1] Users find that accessing legacy data on mainframes and in enterprise applications is even slower and more difficult.

With data critical to modern applications, limitations on user agility and any increased latency in interacting with relevant data have significant business impacts. Today, businesses need an architecture that enables users to generate data insights for requirements that were not known during planning stages. This presents a major challenge when using legacy analytics technologies and practices.

Whether in a planned or unplanned fashion, organizations are democratizing data access, visualization, and analytics, adding many more users. TDWI research reports that about a third (32%) of organizations surveyed say that 41% or more of their employees use dedicated tools or cloud-based services designed specifically for BI, analytics, and/or visual data discovery.[2] Although this percentage has risen over the years from around 20%, not all users have the skills, time, and interest to work with dedicated tools and services. Many prefer to stay within their business applications and use embedded functions.

[1] TDWI 2021 research survey about global data integration trends, see https://tdwi.org/articles/2021/05/26/diq-all-survey-lack-of-modern-data-integration-hindering-business-objectives.aspx.

[2] *TDWI Best Practices Report: Evolving from Traditional Business Intelligence to Modern Business Analytics*, online at tdwi.org/bpreports.

Thus, modern data architectures have to support both standalone data applications and embedded data visualization and analytics inside customer relationship management (CRM), sales force automation (SFA), enterprise resource planning (ERP), mobile apps, and specialized industry software solutions.

**Broad use cases demand a spectrum of data.** Whether embedded or standalone, today's data applications run the gamut from traditional BI reporting and dashboards to predictive analytics involving sophisticated models and machine learning. More advanced than nontechnical data consumers are the growing cadre of citizen data scientists—data-savvy business subject matter experts who want to incorporate self-service data exploration and analytics development into their workflow without having to rely on the technical teams to make data available to them (as with traditional legacy approaches).

To develop complete and coherent views of business situations, data-savvy users need access to all kinds of data: structured, semistructured, and unstructured. Disparate data management silos that support only one type of data or one data model make access slow because each silo has its own data access and interaction complexities. In recent years, organizations set up data lakes to support advanced data science; now, many organizations are responding to the needs of citizen data scientists by migrating data from disparate silos into single data lakes—a trend we will discuss in this Pulse Report.

**Architectures need modernization for data-hungry, AI-infused applications.** Today's data architectures must support growth in artificial intelligence (AI), especially the development and use of machine learning (ML). Embedded and standalone data applications are evolving to feature automated, AI-driven user recommendations. These applications apply AI to automatically address user requirements and learn usage patterns. Solutions can suggest data sets and visualizations helpful to users' analytics and ultimately to their business decisions.

Leading-edge data applications apply AI/ML to automatically discover data insights and prescribe recommended actions in the context of business processes or customer engagement.

Thus, use of AI/ML in data applications goes beyond enabling user access to bigger and more varied data sets, important as that is. Modernized data architectures are necessary to support AI/ML for proactive intelligence that increases speed and agility in decision making. Leading-edge data applications apply AI/ML to automatically discover data insights and prescribe recommended actions in the context of business processes or customer engagement. For example, AI-infused data applications could guide optimal responses to real-time maintenance situations or provide contact center agents and salespeople with personalized cross-sell or up-sell offers tailored to prospective customers.

About a third (32%) of organizations surveyed by TDWI say it is very important to automate decisions in operational or process systems; 37% say it is somewhat important.[3] AI-infused data applications are data-hungry; to produce recommendations and drive smart automation, modern data architectures need to supply them with seamless and continuous access to a wide

[3] *TDWI Best Practices Report: Faster Insights from Faster Data*, online at tdwi.org/bpreports.

range of data. In addition to supporting AI/ML use cases with vast amounts of data, modern data architectures enable successful ML pipelines by allowing data scientists to quickly detect patterns and anomalies in the data and visualize relationships between the data elements through BI tools for feature engineering.

**AI/ML plus a modern data architecture can improve daily decisions.** Faster analytics and contextual intelligence will lead to improved understanding of key performance indicators (KPIs) and reporting metrics, which many operational managers rely on for daily decisions. Across organizations, decision makers want sophisticated analytics that help them accurately forecast performance, explore causal factors, and identify correlations across data types. When KPIs are trending in an unanticipated direction, these insights help users discover why. In TDWI research, 68% of organizations surveyed say it is important to augment dashboards and KPIs with AI-driven insights; 60% want to improve forecasting with predictive modeling and AI/ML.[4]

**Unexpected situations highlight the importance of data agility.** Predictive and real-time insights are critical as conditions change in the face of unexpected events. Dealing with the unexpected has become today's preeminent business challenge as COVID-19-pandemic-related disruptions continue to impact supply chains, logistics, the workforce, customer demand, security, and general sentiment about population health and the economy. Corporate and brand reputations are vulnerable if organizations are not proactive and capable of adjusting to new circumstances. Users need the power of current, complete, and accurate data so they can ask new questions using ad hoc queries and investigate the full range of new and historical data sources.

Thus, never more than now, decision makers want in-context, data-driven insights that enable organizations to be smart, agile, and responsive. If legacy data architecture limits the range of potentially relevant data users can access, visualize, and analyze at speed, they are less able to unlock the value of that data and make informed decisions.

**Cloud computing is fundamental to modern data architecture.** Cloud computing offers scalability, flexibility, cost elasticity, and fast deployment. More than half (54%) of organizations surveyed by TDWI already have traditional data applications such as enterprise BI, reporting, and dashboards running in the cloud, and nearly as many (51%) have business-driven self-service BI and analytics there.[5] Trends such as SaaS adoption and the digital transformation of business applications are accelerating data gravity in the cloud. This means that organizations increasingly want modern data architectures centered in the cloud so data management is closer to the most important data.

Rather than be limited to monolithic on-premises systems with fixed storage, performance, and scale, modern data architectures use cloud data platforms to take advantage of looser coupling of storage and computation. This increases flexibility so organizations can cost-effectively store as much data as they need and use fast, low-latency networks and high-bandwidth communications to move data to scalable, high-powered processors as needed by workloads. The flexibility afforded by looser coupling enables support for the full range of users and data applications with continuous access to diverse data, including through AI/ML programs embedded in next-generation business applications.

[4] *TDWI Best Practices Report: Evolving from Traditional Business Intelligence to Modern Business Analytics*, online at tdwi.org/bpreports.
[5] Ibid.

# Opportunities, Challenges, and Recommendations

With the preceding discussion as context, we can now look more closely at key opportunities and challenges regarding data architecture modernization and examine how to unlock the business value of data.

## #1: Update Your Data Strategy to Take Advantage of the Cloud

Organizations need a smart data strategy to grow and sustain data architecture modernization and reap business value. A complete data strategy should cover the entire data life cycle from ingestion to use, sharing, and ultimate archiving or destruction. Data strategies must balance business-driven flexibility and self-service with the benefits of enterprise coordination and integration. However, the preeminent data strategy goal should be to streamline the path to unlocking business value.

Cloud computing has changed traditional data strategies. Serverless computing relieves organizations of processes for configuring, managing, and maintaining infrastructure. It accelerates time to value through faster deployment and quicker data pipeline development, reducing management overhead and total cost of ownership (TCO). Organizations are embracing the speed, scale, and flexibility offered by cloud data services and related BI, analytics, and data integration cloud-based solutions to run more, faster, and cheaper experiments and interaction with data.

Cloud data storage can provide a foundation for a balanced data strategy that includes data from legacy applications and systems.

However, without a balance between self-service and enterprise control, frustrated users will spin up systems in the cloud in a disorganized, piecemeal fashion. This results in complicated data fragmentation, quality, and governance issues and higher TCO. In addition, a rush to the cloud that does not consider how to improve access to data locked up in legacy applications and mainframes will leave behind valuable data.

Cloud data storage can provide a foundation for a balanced data strategy that includes data from legacy applications and systems. Many organizations focus first on migrating their on-premises data warehouses to the cloud, but by using object storage in the cloud effectively, organizations can gain the benefits of the cloud sooner and more cost-effectively.

Data strategies should start with using cloud data storage as the main ingestion and staging area. Then data can be loaded into the data warehouse or other target BI and analytics platforms, including those that support embedded functionality. As a resource available for all types of users and workloads, cloud data storage can increase access to data while reducing users' needs to build their own data silos.

Organizations can evolve their cloud data storage into a full-blown data lake with specialized zones, which this Pulse Report will describe later. Data strategies should prioritize tighter integration between the data lake and the data warehouse. With easier access to the data lake, citizen data scientists and business users can employ BI services for fast, interactive queries directly against the data lake, where query execution takes advantage of scalable, massively parallel processing (MPP).

For traditional dashboard and reporting workloads, organizations can process data in the cloud data warehouse. Data strategies can incorporate processes for moving “cooler” data (data not needed as frequently or continuously by dashboards and reports) from the data warehouse to cloud data storage to reduce costs.

## #2: Improve Access to Enterprise Application Data as You Modernize and Migrate to the Cloud

As organizations modernize their data architectures, they should not overlook problems in accessing data held in legacy applications and mainframe systems. Mainframes hold valuable data in older relational or nonrelational databases, flat files, and storage devices. In many organizations, these systems—as well as legacy applications on other platforms—continue to play an essential role in daily operations. They contain data that advanced data scientists and data-savvy business users would like to explore and investigate to spot trends, build predictive models, and uncover hidden data relationships and correlations. However, it is notoriously challenging to locate and access the right data in ERP systems such as SAP as well as in various CRM, call center, and supply-chain management systems.

We can see in Figure 1 that more than half (55%) of TDWI survey respondents regard utilizing data from mainframe/legacy applications as the most difficult, with 22% calling it “very difficult.” A significant percentage (43%) find utilizing data from on-premises business applications difficult. ERP systems such as SAP, which ideally (but not always) store all data used by multiple application modules in a single relational database, nonetheless have complex structures that can make data extraction to a data warehouse and transformation for user access and interaction difficult. Development costs due to complexity often limit the number and variety of reports, dashboards, and analytics workloads supported.

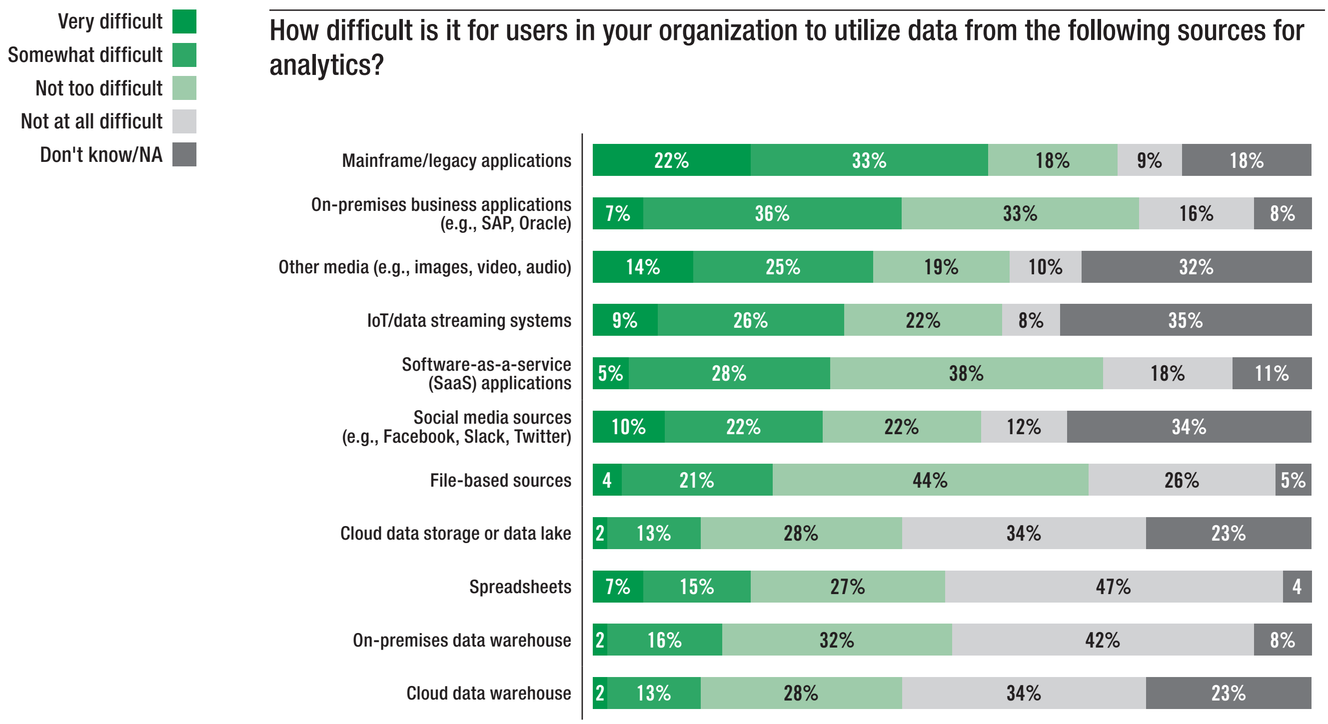


***Figure 1.*** *Based on answers from 244 respondents to 2021 TDWI survey about global integration trends. Ordered by highest combined "very difficult" and "somewhat difficult" responses.*

Many organizations today have a long-term strategy of migrating ERP, CRM, and other business applications to the cloud, particularly to SaaS systems. However, the size and complexity of these applications make full cloud migration slow and difficult. During this lengthy transition, users must navigate hybrid environments where some (if not the majority) of the data is managed in on-premises systems while active and newly generated data is managed in SaaS. Sourcing and transforming data for data warehouses is similarly time-consuming and difficult.

**Replicating data from legacy systems into the cloud.** To provide complete access to data sooner, organizations should consider replicating data from on-premises enterprise applications and mainframes to cloud data storage. This enables organizations to consolidate data more quickly for a data lake or for staging data for transformation needed by a cloud data warehouse. Organizations can use data replication tools to choose how to keep data in sync between sources and the cloud data warehouse or lake.

A replication strategy allows organizations to unlock value from legacy sources without having to wait for the full cloud migration of applications, mainframe systems, and even on-premises legacy data warehouses. The data is available sooner for business-critical analytics such as predictive modeling and AI/ML—sooner than it would be if the data were not replicated and access were still directed to each on-premises mainframe system and legacy application. Business sponsors will appreciate the value of having data in scalable cloud platforms; as a result, they will be more receptive to strategies for accelerating complete data and application migration and modernization.

**Using ELT to make more data available faster.** Traditional extract, transform, and load (ETL) routines involve extracting data to staging areas for transformation before the data is loaded into a targeted cloud data warehouse. TDWI research finds that 38% of organizations surveyed regard data loading, movement, and integration steps associated with ETL as among their biggest challenges in trying to augment or replace on-premises systems with new cloud-based services.[6]

Organizations should evaluate whether a better strategy would be to use data pipelines to load data rapidly into a cloud data lake as the target platform and then perform transformations after loading. This approach, known as ELT, has the advantage of being able to use powerful and scalable cloud-based MPP database engines for faster and more workload-specific data transformation and cleansing.

To conclude, replication and pipeline extraction into cloud data storage enable organizations to make data from legacy applications and mainframes available sooner. For example, complete migration of each ERP module could be ongoing without standing in the way of business users and data scientists who need to access this valuable data today. Depending on business requirements, organizations could decide to focus replication and extraction first on active transactional and master data, which might be the most valuable to users, and then follow with replication and migration of older historical data.

## #3 Integrate Cloud Data Lakes and Data Warehouses for Complete User Access

Executives and managers in lines of business and departmental functions want to gain comprehensive situational awareness and insight into critical trends and predictive patterns through analysis of large volumes of diverse data—data that might be structured, semistructured, or unstructured. With real-time data from logs, devices, and sensors streaming into data systems, organizations need modern tools and data platforms to support powerful operational analytics that answer complex business questions. Tighter integration between the cloud data warehouse and data lake would help organizations overcome data silo problems that thwart faster insights for business users and advanced data scientists.

Traditionally, business users and analysts access carefully prepared data in a data warehouse. Unsatisfied with just reports and limited data access, many users will employ self-service tools to blend extracts of new data in “shadow” data silos that exist outside the data warehouse. Data silos over time contain duplicate, incomplete, and inconsistent data, which complicates data quality and governance and increases latency. TDWI research finds that most users desperately want single, complete views of all relevant data, but silos force them to go from one data platform to another, adding delays and frustration.

Historically, data lakes have been the province of data scientists and programmers who know how to work with raw data and set up custom data pipelines and transformation routines. Data lakes have been essential for analytics-driven NoSQL workloads—for example, those that deliver advanced personalization. These workloads touch a variety of customer behavioral data, demographics, location intelligence, and contextual data.

Today, organizations can easily set up cloud data lakes using object storage on cloud platforms. This offers a faster path to launching data science projects involving exploratory, predictive, and real-time analytics. Data scientists can develop and test AI/ML models and algorithms that depend on access to hundreds of terabytes (if not petabytes) of data contained in the data lake and sourced from log files, social media, mobile device data, and more.

[6] Previously unpublished survey data collected for the 2020 *TDWI Best Practices Report: Evolving from Traditional Business Intelligence to Modern Business Analytics*.

**Data lakes set up by experienced organizations are not typically one giant body of data.** Often, organizations will partition data lakes into zones. A raw zone collects data from sources in raw form, which is valuable for data scientists but lacks the cleansing and preparation required for standard business users. Raw zones contain the most data, far more than other zones that feature carefully selected, structured, and curated data.

The prepared zones typically contain data ingested from more structured business application sources, which data engineers have transformed to adhere to a canonical data model and schema so business users or data scientists can interact with it immediately for visualizations, calculations, and more. This is how organizations can use cloud data lakes to support some traditional BI and analytics data interaction.

A lake can serve as a staging area for data transformation and cleansing. It could also serve as an operational data warehouse that collects near or true real-time data from online transaction processing (OLTP) systems, e-commerce, and business applications. Finally, some data lakes will have additional zones dedicated to the requirements of nontechnical data consumers who just need regular extracts of the data or are simply viewing feeds in dashboards.

As the core of a single, unified data architecture, a data lakehouse can support all types of workloads, at any latency, and at any scale.

**A data lakehouse reduces silos, provides easier access, and supports new use cases.** With data zones set up for curated data, data lakes sometimes overlap with the purposes of a data warehouse. Thus, it makes sense to bring the data lake and data warehouse into a more tightly integrated architecture; this is sometimes called a data lakehouse.

As the core of a single, unified data architecture, a data lakehouse can support all types of workloads, at any latency, and at any scale. Tightly integrating the cloud data lake and data warehouse within a single architecture allows all types of users to take advantage of robust, scalable processing to speed data profiling, transformation, and other pipeline preparation steps that normally require separate staging areas. Organizations can eliminate delays caused by extracting data to separate staging areas and then loading it into target data warehouses.

The consolidation of data silos enabled by a data lakehouse allows organizations to centralize data management and monitor performance and governance across both the data lake and data warehouse. Organizations could use the single, unified data architecture to support classic data warehouse workloads for BI reporting, dashboards, and online analytical processing (OLAP) but on scalable and elastic cloud storage and processing. The data lake would continue to provide data scientists with access to raw, semistructured, and unstructured data for advanced analytics and AI, but expand their reach into the data warehouse.

**A data lakehouse is more than a "one-size-fits-all" data architecture.** A data lakehouse can deliver benefits beyond just the efficiency and increased data access provided through integrating a data lake and data warehouse. It offers a single architecture that can scale up or down to handle traditional BI workloads, data science, NoSQL workloads, and, importantly, new data-driven consumer applications such as mobile apps, self-service kiosks, and chatbot interactions. They could also include services offered to partners and customers through data marketplaces and exchanges as part of a data monetization strategy. A data lakehouse eliminates the silo problems that prevent organizations from having an agile data architecture that responds to the requirements of each type of workload.

## #4: Modernize Data Architecture to Handle Real-Time Data Use Cases

Often, the closer organizations get to real-time data, the more valuable that data is. A decided majority (77%) of organizations surveyed by TDWI indicate that near or true real-time data, dashboards, and analytics are important to their firm's success, including 30% who say they are very important.[7] Streaming data can give organizations new insights into solving problems, beyond what they gather from historical data. For example, many organizations want to tap Internet of Things (IoT) sensor data from machines and other equipment; to interpret this data correctly, they need numerous sources of data (such as geospatial data) and sophisticated analytics models and algorithms.

Organizations seeking the benefits of real-time data and analytics should evaluate the use of lambda and kappa architectures.

Taking advantage of real-time data involves special considerations from a data architecture perspective. Some organizations establish a lambda architecture, which enables efficient processing of huge amounts of data at low latency using a hybrid of batch and stream processing. Using a lambda architecture allows organizations to handle high throughput with low latency while ensuring transaction consistency.

The lambda architecture provides organizations with a balance of speed and reliability. The architecture's speed layer deals with real-time data and creating real-time views. The batch layer provides views of consistent, historical data residing in fault-tolerant distributed storage. The combination enables users to gain a complete, if not quite fully integrated, view of batch and real-time data. Having a unified cloud data lakehouse is helpful to support a lambda architecture and for improving integrated user access and views.

For near or true real-time data streaming, many organizations are taking advantage of open source Apache Kafka, a popular framework for building streaming data pipelines and messaging engines. This has led organizations to adopt a kappa architecture, which enables users to tap data streams in a Kafka messaging engine without having to first land them in a data lake. Kappa uses Kafka to treat all data as streams; the stream processing engine in a kappa architecture reads data stored immediately after it appears in the Kafka messaging engine. In this way, it provides a single technology stack for both real-time and batch processing.

Although both lambda and kappa architectures have benefits that deliver value, kappa addresses some weaknesses and complexities in the standard lambda architecture. Kappa is good for data architectures that do not need a batch layer. The stream processing engine can function as the sole data transformation engine, formatting data as needed for analytics and AI/ML. Kappa enables organizations to operationalize AI/ML entirely on real-time data streams. Additionally, organizations can store formatted data in a database for business users' BI needs or for

[7] *TDWI Best Practices Report: Faster Insights from Faster Data*, online at tdwi.org/bpreports.

data-driven consumer applications to query. An integrated technology stack provided by a cloud data lakehouse can be this database and in other ways support a kappa architecture.

Organizations seeking the benefits of real-time data and analytics should evaluate lambda and kappa architectures. Each one can help organizations unlock business value depending on requirements. With its batch layer, the lambda architecture ensures greater transaction consistency; however, batch cycles can add some latency and the lambda architecture can necessitate more complex programming and data modeling. The kappa architecture positions organizations to get full value out of Kafka and real-time data streams for AI/ML. However, the lack of a batch layer may require adding other components and specialized programming to provide the batch layer if needed.

Unifying the data lake, data warehouse, and other data silos in the cloud into a data lakehouse would benefit implementation of either architecture. The combination of architecture and cloud data lakehouse would give organizations a modern data strategy for unlocking the value of real-time data streams and supporting new use cases.

## #5: Establish a Unified Metadata Repository for Easier Location of Trusted Data

Diverse users and workloads are demanding access to varied and voluminous data. This situation increases the need for accurate, well-managed metadata. A complete and up-to-date repository of metadata, such as in a data catalog, makes it easier to establish consistent data and table definitions and discover schemas, mappings, transformations, and other critical information about data. A centralized metadata repository helps organizations reduce confusion about the data's quality and consistency. These are key objectives of a modern data architecture; organizations should therefore prioritize establishment of a metadata repository such as a data catalog.

TDWI research finds that most organizations are looking for improvement. Only 9% are very satisfied with how well their data catalogs, metadata, and semantic layers support self-service BI, analytics, AI/ML, and related data preparation, transformation, and integration; 49% need a major upgrade.[8] Fortunately, technology trends toward incorporation of AI-driven automation in solutions enable organizations to move beyond the intensive manual work traditionally necessary to keep metadata complete, accurate, and up to date. Today, AI-driven cloud-based services can automatically crawl for metadata contained in business applications as well as data lakes, data warehouses, and BI/analytics platforms.

Another TDWI Pulse Report in this series will address the role of data catalogs and metadata repositories in governance. Here, we highlight two important benefits for unlocking business value:

- **Data findability.** TDWI research finds that the most common priority behind setting up a data catalog or other metadata repository is to make it easier for users to search for and find data. At the point of data use, users can learn information about the data, including its consistency, age, and any governance constraints. Metadata repositories help organizations associate data definitions with descriptive, higher-level meanings for the data expressed in business language, which might be contained in a business glossary or master data management system. Services that track data lineage can inform users about the data's origins and its current location.

[8] *TDWI Best Practices Report: Evolving from Traditional Business Intelligence to Modern Business Analytics*, online at tdwi.org/bpreports.

- **Data quality and trust.** As workloads reach out to data from varied sources, data quality becomes challenging. Two-thirds (66%) of organizations surveyed by TDWI indicate that improving data quality and trust in data is the number one step that would improve their organization's BI and analytics.[9] A metadata repository helps organizations improve data quality by providing a central resource of consistent definitions and related information about data sets. With traditional manual methods, it is difficult to scale metadata repositories. Modern, AI-driven services provide intelligent automation for complete and accurate data quality profiling, assessment, remediation, and monitoring.

## #6: Drive Business Value by Modernizing Architecture to Enhance Self-Service Analytics

In recent years, there has been a strong trend toward business-led adoption of self-service BI, analytics solutions, and cloud-based data platforms. TDWI research finds that more than half (54%) of organizations surveyed anticipate that the business side's role will continue to increase in the selection, funding, and management of such solutions.[10] This makes enhancing and modernizing data architecture to support expansion in business-driven self-service data interaction a priority.

Self-service solutions integrated with easily set up cloud data platforms enable business users and data scientists to respond to immediate, even unexpected, needs.

Traditional on-premises IT solutions would require considerable time to develop and configure BI and analytics systems. Self-service solutions integrated with easily set up cloud data platforms enable business users and data scientists to respond to immediate, even unexpected, needs for data visualization, analytics, predictive model development, and new data exploration. Ease of use is a critical success factor for these solutions.

However, with self-service, users need not work with a blank slate. Organizations can implement solutions that make it possible to set up predefined, parameterized reports and dashboards that users personalize to their needs. Some solutions enable organizations to develop predefined templates, OLAP cubes, and analytics zones (i.e., "sandboxes") in cloud data lakes. Automation and AI-driven recommendation features help users derive faster value from large and complex data sets.

Here are three important considerations for modernizing data architecture to enhance self-service analytics:

- **Provide access to a fuller range of data.** The design of most traditional data warehouses has been to support repeatable and auditable managed reporting. These frequently rather rigid systems are less suited for today's demands, which include significant ad hoc querying; flexible exploration of new, sometimes real-time data; and complex analytics workloads that demand high-performance processing of granular and unstructured data. Organizations should evaluate whether a unified cloud data lakehouse could provide

[9] Ibid.
[10] Ibid.

a more complete data service to handle the range of use cases and workloads demanded by data-savvy business users.

- **Integrate self-service BI and analytics with data catalogs.** As noted earlier, data catalogs can make it easier to locate data sets and understand their quality and lineage. With a data catalog, users move faster to discover and explore data relationships and examine variations across data sets, including by examining metadata and other information about semi- and unstructured data in the cloud data lake. Just over half (51%) of organizations surveyed by TDWI say that managing metadata in a data catalog is one of the most important steps they could take to increase success with BI and analytics.[11]

- **Increase data literacy.** Having the right tools and data architecture is critical to enabling people to increase their skills and experience. However, it takes improvement in data literacy to realize full value from data. Data literacy addresses human aspects of how people interact with data. The primary goal is to raise individuals' proficiency in understanding what data means and their ability to communicate and share analytics insights, typically through visualizations such as charts and graphs.

  A second important data literacy goal is to increase recognition of people's responsibility and accountability for how they collect, integrate, prepare, and protect data. This is crucial to governance, adherence to data privacy regulations, and overall success in building trust in data and analytics through internal quality standards.

  Organizations should set up data literacy training for users. They should organize data stewardship programs that share best practices between experienced and novice users. Training plus stewardship will accelerate data literacy so more people gain value from data sooner.

# A Final Word

With change being the only constant amid today's uncertainties, unlocking business value from all data is paramount. People need the means to explore, analyze, visualize, and share data insights easily and creatively so they can address changing circumstances and make informed decisions. It is important for organizations to modernize by using integrated technologies that derive value from new sources such as real-time streaming data without adding complexity. The six topics discussed in this Pulse Report are essential for using data to improve business resilience, agility, speed, and competitiveness.

Using cloud data storage and analytics as the foundation, organizations can reap value from all current and new data sooner. They do not need to wait for complete cloud migration of complex business applications and mainframe systems. Organizations can use the cloud to make critical data available sooner so users have complete access and are not blind to unexpected situations, trends, and patterns. With a modern data architecture, organizations are better able to realize data-driven opportunities for improving operational performance, enhancing customer relationships, and accelerating product and service innovation.

[11] Ibid.

# About Our Sponsors

For over 15 years, Amazon Web Services has been a comprehensive and broadly adopted cloud platform. AWS has been continually expanding its services to support virtually any cloud workload, and it now has more than 200 fully featured services for compute, storage, databases, networking, analytics, machine learning and artificial intelligence (AI), Internet of Things (IoT), mobile, security, hybrid, virtual and augmented reality (VR and AR), media, and application development, deployment, and management from 80 Availability Zones (AZs) within 25 geographic regions, with announced plans for 18 more Availability Zones and six more AWS Regions in Australia, India, Indonesia, Spain, Switzerland, and the United Arab Emirates. Millions of customers—including fast-growing start-ups, large enterprises, and major government agencies—trust AWS to power their infrastructure, become more agile, and lower costs. To learn more about AWS, visit aws.amazon.com.

Qlik's vision is a data-literate world, where everyone can use data and analytics to improve decision making and solve their most challenging problems. Qlik provides an end-to-end, real-time data integration and analytics cloud platform to close the gaps between data, insights, and action. Qlik's Data Integration platform uniquely enables and simplifies continuous data delivery from a wide variety of enterprise data sources including SAP, legacy mainframe systems, databases, data warehouses, and more to AWS by facilitating real-time data ingestion and data pipeline automation. Qlik's SaaS-based data analytics platform empowers customers to drive actionable insights from all their data through analysis of massive amounts of data in near real time with unique augmented analytics, associative indexing, and cognitive engine technologies.

Over 38,000 customers around the world leverage Qlik to transform their data into value in real time.

research

TDWI Research provides industry-leading research and advice for data and analytics professionals worldwide TDWI Research focuses on modern data management, analytics, and data science approaches and teams up with industry thought leaders and practitioners to deliver both broad and deep understanding of business and technical challenges surrounding the deployment and use of data and analytics. TDWI Research offers in-depth research reports, commentary, assessments, inquiry services, and topical conferences as well as strategic planning services to user and vendor organizations.

tdwi

**Transforming Data With Intelligence™**

A Division of 1105 Media
6300 Canoga Avenue, Suite 1150
Woodland Hills, CA 91367

E info@tdwi.org