Steps to Migrate and Modernize Traditional Data Warehouse Systems

By David Loshin

INTRODUCTION

Over the past three decades, the enterprise data warehouse (EDW) has become firmly ensconced as the fundamental resource supporting corporate business intelligence and operational analytics. Although a variety of different implementation strategies emerged early in the history of reporting, business intelligence, and analytics platform paradigms, the conventional on-premises data warehouse architecture has remained relatively static since it became the prototypical architecture. Today, most EDW deployments are engineered around extracting data from (mostly internal) source systems, accumulating data in a staging area for integration, and applying data standardizations and structural transformations, followed by loading into the target warehouse.

Some technical innovations have facilitated improved performance, most notably the evolution of high-performance data warehouse appliances and data warehouse automation tools that simplify the creation of the stereotypical EDW instance. However, because of radical growth in data volumes and increasing data diversity, many organizations have begun to reevaluate their commitment to the traditional on-premises EDW and its ability to scale performance in proportion to exploding data volumes. Many organizations have begun to re-envision their overall data warehouse strategy, seeking to take advantage of the many benefits of the cloud, and have started migrating their data warehousing capabilities to the cloud.

As organizations increasingly need to process and analyze exploding volumes of highly diverse data streaming at or near real time, cloud platform economics and increasingly sophisticated cloud-native technologies combine to provide on-demand, high-performance, scalable computing resources at a fraction of the cost of on-premises big iron. This TDWI Insight Accelerator considers what it means to transition to a modernized reporting and analytics environment and provides concrete steps for migrating and modernizing the enterprise reporting, business intelligence, and analytics ecosystem to the cloud.

MODERNIZE, DON'T JUST MIGRATE!

Migration implies moving an existing system implementation from one platform to another platform, and data warehouse migration (with its connotation of "lifting and shifting" existing system footprints) is a reasonable tactical first step. However, one must acknowledge that an on-premises EDW framework might not be designed to best take advantage of cloud resources and services. This is where modernization comes into play.

Modernization involves reevaluating business process requirements and considering technical dependencies. It affords an organization the opportunity to reexamine the degree to which its legacy systems continue to meet business requirements, to identify technology dependencies that can be eliminated using modern techniques, and to refactor or reengineer the reporting and analytics environment to align with emerging process demands.

The cloud provides a new strategic paradigm for reporting, business intelligence, and analytics with capabilities that far exceed the traditional relational, star-schema-structured data warehouse. TDWI research suggests that the cloud enables a more robust unified platform for modern analytics that encompasses the full analytics life cycle (see Figure 1).
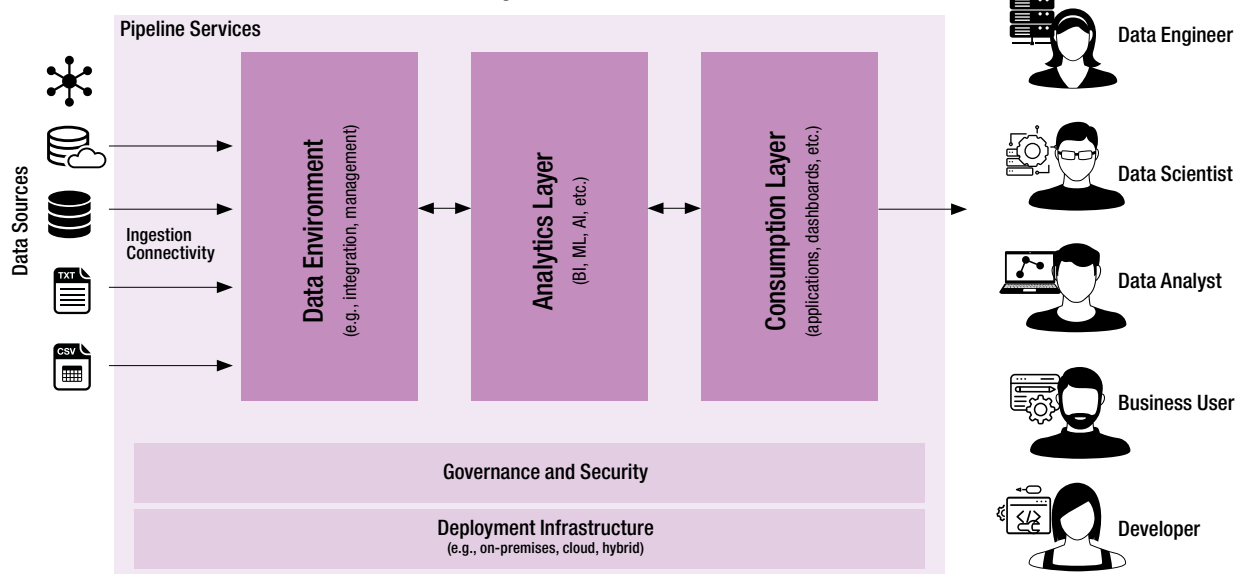
Unified Platform for Modern Analytics

Figure 1. The unified platform for modern analytics.[1]

This conceptual design informs the cloud analytics architecture by identifying opportunities for replacing legacy EDW components with cloud-native technologies. Consequently, a strategic approach to data warehouse modernization in the cloud will free the organization from identified technical dependencies and baggage inherent in the traditional data warehouse architecture and enable innovation in corporate business intelligence and advanced analytics.

A cloud-based data warehouse may incorporate a variety of data management frameworks for both structured and unstructured data assets that may be distributed across multiple cloud computing instances and even cloud service providers. That suggests that cloud data warehousing is less about conventional extraction, transformation, and loading data, and more about an enterprise platform data strategy with automated data onboarding and integration, orchestrated data pipelines, data engineering and preparation, analysis, self-service data delivery, and embedded analytical models to drive real-time decision-making.

BUSINESS DRIVERS FOR MODERNIZING WHEN MIGRATING

Recent TDWI research solicited survey respondents about the objectives that were most important in guiding the organization's strategy for cloud data integration and cloud data management (see Figure 2).

The responses and recommendations summarized in the report reflect a variety of business drivers inspiring modernization to the cloud, including:

Which of the following objectives are most important in guiding your organization's strategy for cloud data integration (DI) and management? Select up to five.
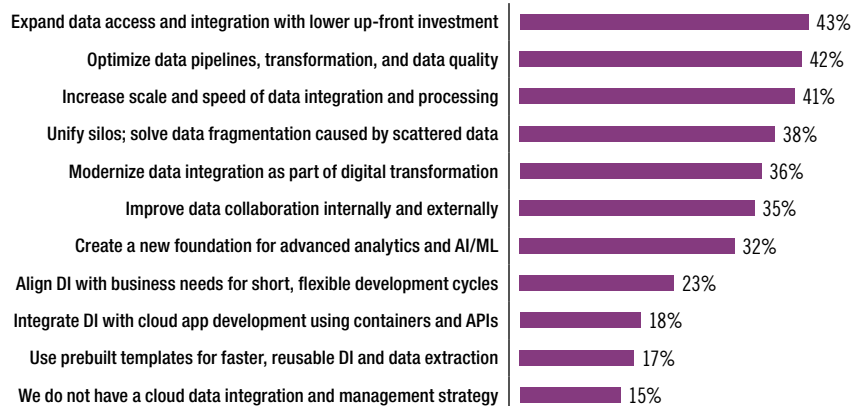
Objective	Percent
Expand data access and integration with lower up-front investment	43%
Optimize data pipelines, transformation, and data quality	42%
Increase scale and speed of data integration and processing	41%
Unify silos; solve data fragmentation caused by scattered data	38%
Modernize data integration as part of digital transformation	36%
Improve data collaboration internally and externally	35%
Create a new foundation for advanced analytics and AI/ML	32%
Align DI with business needs for short, flexible development cycles	23%
Integrate DI with cloud app development using containers and APIs	18%
Use prebuilt templates for faster, reusable DI and data extraction	17%
We do not have a cloud data integration and management strategy	15%

Figure 2. *Top objectives for cloud data integration and management.*[2]

- **LOWERING TOTAL COST OF OPERATIONS.** Organizations are progressing away from the on-premises model of capital hardware acquisition. The cloud promises economic benefits, but to understand these benefits requires considering the costs amortized across the entire data analytics life cycle and ecosystem. This perspective is intended to establish an economic strategy that demonstrates an overall lower total cost of ownership when compared to traditional on-premises data warehouse appliances.

- **COST-EFFECTIVE EXPANSION OF DATA ACCESSIBILITY.** Organizations are increasing their use of a variety of cloud service providers for computing and storage resources. Correspondingly, they want their developers and data consumers to easily access data across a multicloud footprint. Forty-three percent of respondents indicated that expanding data access and integration with lower up-front investment was an important objective.

- **THE NEED FOR SCALABILITY.** Exponential growth in data volumes, increasing speed of data delivery, and surging numbers of downstream data consumers all create pressures for ensuring that reporting and analytics systems continue to meet performance demands. Forty-one percent of survey respondents indicated that increasing "scale and speed of data integration and processing" was an important objective for cloud data integration.

- **INCREASING DATA DIVERSITY.** Traditional data warehouses are engineered and optimized to support structured data. Yet algorithmic analytics libraries and services increasingly support more diverse data asset types (e.g., semistructured data, unstructured text, audio, video). Modernized environments cannot be limited to only integrating structured data. TDWI's research suggests that "organizations investing in AI/ML and automated machine learning should focus on modernizing

data pipelines for speed and scalability and streamlining access to raw, semistructured, and unstructured data as well as prepared data aggregations," and 42 percent of the respondents indicated "optimizing data pipelines" as a key objective.[3]

More to the point: in earlier days, the data analyst modulated the boundary between the business consumer and the data warehouse. Today, there is a range of sophisticated data consumer personas (from data scientist to citizen data analyst), each of which has a degree of awareness and skill in data access, preparation, and analysis, and all must be empowered in a modernized environment. This suggests that supporting communities of increasingly sophisticated data consumers requires a modernized data environment.

STEPS TO MIGRATE AND MODERNIZE

The compelling argument for cloud adoption suggests that it is just a matter of time before most organizations will migrate at least some of their EDW capabilities to the cloud. Yet without a standard approach that can be observed across business units, there are likely to be impediments to a smooth migration. A strategic perspective on modernization takes a long view towards migrating data and applications to the cloud. However, an iterative approach may allow your organization to achieve early benefits and provide some latitude for architecting a longer-term framework that best leverages cloud-native resources and services.

Some steps supporting a migration and modernization process include:

- **REEVALUATING YOUR BUSINESS NEEDS.** Engage business consumers to understand current and future business intelligence and analytics

use cases. Gather information to identify the different data consumers' personas. Estimate the types of analytics applications, user volumes, and expected levels of performance. Ensure that the proper metrics and measures are defined and can be supported by cloud resources and services.

- **ASSESSING ORGANIZATIONAL READINESS TO MIGRATE.** Senior management's ardor for cloud adoption can be stymied by resistance to change. Engage the process and application owners to solicit their expectations, levels of risk tolerance, system complexity, and resistance to technology innovation. By specifying dimensions and how specific metrics are scored, you can prioritize the groups that are most ready to migrate and modernize their reporting and analytics applications.

- **ABANDONING THE ON-PREMISES APPLIANCE.** Cloud providers can furnish eminently scalable computing and storage resources along with specialized data warehouse, data lakehouse, and specialty analytics services. Because these services can be configured to perform at an equal or improved level over a traditional data warehouse architecture, you no longer need to manage an on-premises data warehouse appliance.

- **ADOPTING A FLEXIBLE ARCHITECTURE.** Recognize that rapid innovation in the cloud has expanded the palette of available resources, services, and tools to support enterprise reporting and BI. You no longer need to rely on traditional data warehouse architecture, especially when the same reporting and analytics outcomes can produce faster results more efficiently. Explore the inventory of available cloud-native services, determine the best ways to satisfy the business needs, and adopt a flexible approach to employing cloud-native services.

[3] Ibid.

- **PLANNING TO ITERATE.** Even if the traditional architecture may not be optimally deployed in the cloud, lifting and shifting an existing data warehouse may provide an easy pathway to transition. At the same time, iterative refinement of the enterprise strategy may benefit your organization by exploring ways the cloud data warehouse can continue to meet current requirements while being architected to satisfy future business opportunities. Plan to continually review refinement opportunities, especially as new cloud services emerge.

- **IMPLEMENTING DATA SELF-SERVICE.** Different data consumer communities have increased expectations for data accessibility and integration with their favorite end-user tools. Empower data analysts with a combination of cloud-native reporting and analytics tools and self-service data access.

- **MODERNIZING THE ANALYTICS LIFE CYCLE.** Cloud service providers provide a full array of services supporting the end-to-end data analytics life cycle. Instead of solely focusing on the data warehouse perspective, modernize your organization's approach to data ingestion, preparation, engineering, integration, and analytics.

Tactical recommendations can inform data warehouse architects about the challenges of migration and modernization and influence the design of a flexible and extensible cloud-based data warehouse. By gathering intelligence about enterprise business use cases, process requirements, and performance expectations, you can evaluate business function applications, identify candidate application and data migration opportunities, and prioritize tactical steps to migrate and modernize

corporate data reporting, business intelligence, and analytics capabilities.

Choose the right tools and technologies that modernize data warehouse design, contribute to lowered costs, and simplify future application development. Cloud data warehouse and data lakehouse architectures are the foundation of a future-proof architecture that allows reporting and analytics applications to be rapidly designed, developed, and integrated and to conform to expected levels of performance within a predefined managed budget.

ABOUT OUR SPONSORS

Amazon Web Services provides a comprehensive and broadly adopted cloud offering. Over the past 15 years, 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, Internet of Things, mobile, security, hybrid, virtual and augmented reality, media, and application development, deployment, and management from 81 Availability Zones within 25 geographic regions, with announced plans for 21 more Availability Zones and seven more AWS Regions in Australia, India, Indonesia, Israel, Spain, Switzerland, and the United Arab Emirates. Millions of customers—including the fastest growing start-ups, largest enterprises, and leading government agencies—trust AWS to power their infrastructure, become more agile, and lower costs. To learn more about AWS, visit aws.amazon.com.

Quantiphi, an Advanced AWS Consulting Partner, is an award-winning applied AI and big data software and services company driven by the desire to solve transformational problems at the heart of our customer's business. Our unique approach combines deep industry experience, disciplined cloud and data-engineering practices, and cutting-edge machine-learning research to achieve quantifiable business impact.

Quantiphi has delivered breakthrough results for industries such as healthcare, insurance, financial services, education, media and entertainment, and retail and manufacturing to name a few. Over the years, Quantiphi has helped multiple organizations embark on enterprise-wide data lake and data warehouse migration and modernization journeys to organize their data and streamline their data pipelines to help reduce time to generate actionable insights. These implementations have been completed while ensuring a strong data governance and security framework without compromising on flexibility to allow customers to scale at ease. Learn more about how Quantiphi delivers value to customers at www.quantiphi.com.

ABOUT THE AUTHOR

David Loshin, president of Knowledge Integrity, Inc., (www.knowledge-integrity.com), is a recognized thought leader and expert consultant in the areas of data management and business intelligence. David is a prolific author regarding business intelligence best practices as the author of numerous books and papers on data management, including *Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools, Techniques, NoSQL, and Graph and The Practitioner's Guide to Data Quality Improvement*, with additional content provided at www.dataqualitybook.com. David is a frequently invited speaker at conferences, web seminars, and sponsored websites and channels. David is also the program director for the Master of Information Management program at the University of Maryland's College of Information Studies.

David can be reached at loshin@knowledge-integrity.com.

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.