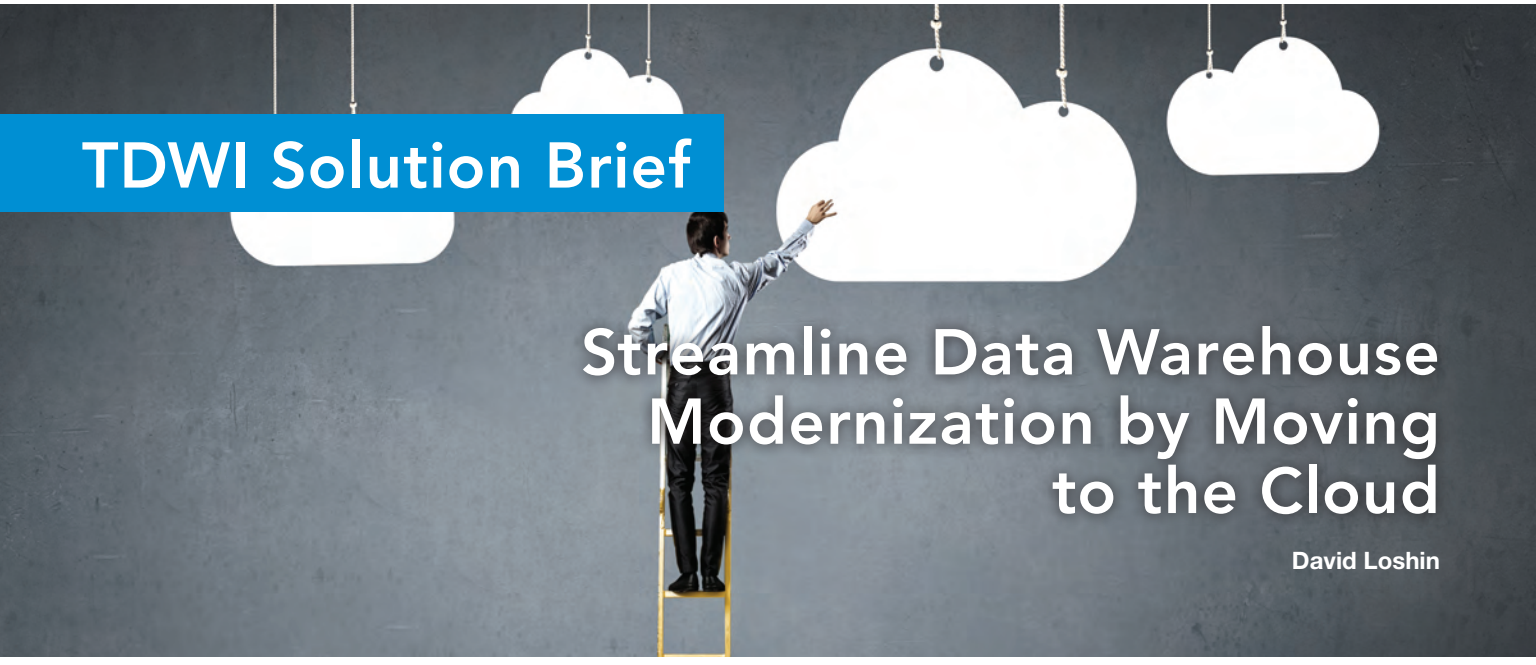
Google Cloud

January 2020



TDWI Solution Brief

Streamline Data Warehouse Modernization by Moving to the Cloud

David Loshin

In this Solution Brief, we explain why your organization needs to rethink its analytics strategy and modernize its data warehouse to support rapidly evolving analytics requirements. We explore how to streamline your migration from a legacy, on-premises data warehouse appliance to a cloud-based, modern data warehouse environment.

Trends, Challenges, and Netezza End of Life

The emergence of open source big data platforms such as Hadoop and Apache Spark has not only lowered the bar for many organizations desiring to adopt a reporting, business intelligence (BI), and comprehensive analytics program, in some ways it has upended existing enterprise analytics strategies.

Although the prototypical, monolithic, on-premises enterprise data warehouse may have satisfied organizational needs in the recent past, there are several trends and challenges motivating organizations to consider modernizing their enterprise data warehouse platforms, including:

- **Data democratization:** Increased demand for self-service access to data coupled with a growing number of savvy data analysts have created the "perfect storm" with surging business demand that exceeds the existing on-premises computational workload capacity.

tdwi
Transforming Data With Intelligence™

- **Data diversity:** The ability to ingest and analyze both structured and semistructured data sources has raised the stakes in terms of effectively using all types of enterprise data assets.
- **Data sprawl:** Enterprise data assets are strewn about disparate systems, creating fragmentation that confounds those seeking to include enterprise data assets from across the enterprise in their analyses.
- **System performance:** Existing monolithic platforms are unable to scale quickly enough to meet data volume growth, computation demands, and growing numbers of workloads. The rising data consumer demands lead to "performance outages" in which performance service-level agreements cannot be met.
- **Data security and reliability:** Data democratization motivates expanding data accessibility, but most systems' security profiles are inadequate to protect against unauthorized access.
- **The need for speed:** Legacy data warehouse platforms were not designed to process continuous data streams to enable real-time integrated analytics.
- **Cost:** The accelerated speed of innovation and increasing system demand call into question corporate capital investments in "big iron" hardware. The growing pressure for actionable insights requires capital-intensive investments and increasing resource overhead that result in spiraling costs.
- **The need for advanced analytics:** Modern organizations are no longer able to rely solely on descriptive analyses to influence business decisions; institutions that integrate advanced predictive and prescriptive analyses with algorithmic machine learning and artificial intelligence are clearly gaining competitive advantage.

However, many CIOs, CDOs, and CAOs are facing a more acute (and fundamentally more urgent) issue: legacy on-premises data warehouse appliances are rapidly nearing their end of life as appliance vendors such as IBM have announced they will no longer provide hardware or software support for their Netezza/PureData for Analytics product line. Organizations facing this dilemma have no choice—they must strategize to modernize their environments and do so immediately.

Fortunately, the imminent end of life of on-premises data warehouse appliances such as Netezza may be a blessing in disguise. It provides the impetus to consider migrating to a modern data warehouse in the cloud. In this Solution Brief, we will examine why organizations need to rethink their analytics strategy and modernize their data warehouse to support their rapidly evolving analytics requirements. We will then explore how data leaders can streamline their migration process when transitioning from a legacy, on-premises data warehouse appliance to a cloud-based, modern data warehouse environment such as Google BigQuery.

Rethinking the Analytics Strategy: The Modern Data Warehouse

Early decision-support systems were promoted as a way of segregating reporting workloads from transaction processing systems and ensure they meet their expected performance service levels. This isolated environment evolved into the conventional enterprise data warehouse—a standalone, on-premises, monolithic system

that is populated through the extraction of data from transactional and operational systems and the application of sequences of standardizations and transformations.

Although such a data warehouse supports data analysts' reporting and analysis needs, that monolithic system struggles to address emerging analytics expectations. Increasing data volumes and scattered data resources need to be incorporated into the environment in a way that does not overrun capacity and operating budgets. Data warehouse appliance computational power is inadequate to meet the growing business demand for analytics, especially with the growing reliance on algorithmic/AI/ML techniques. With a plethora of international and U.S. laws directing protection of sensitive data, these systems' security and protection mechanisms are easy targets.

Together, these conditions suggest there is a need to rethink the analytics strategy. The modern data warehouse can no longer be thought of as an application running on a relational database management system (RDBMS) installed on custom hardware. The modern data warehouse must be thought of as a *platform*: a cohesive, interoperating set of capabilities, blending the use of an enterprise data lake, a variety of analytics services, integration with streaming data sources in real time, and serverless computing, all encompassed within an elastic environment that automatically scales to meet performance needs and cost efficiency.

Requirements of a Modern Data Warehouse

It is critical to address these challenges while recognizing that the conventional data warehouse architecture is insufficient to meet the organization's expanding analytics use cases. The nearing end of life of systems such as Netezza only adds to the urgency. However, just "lifting and shifting" to an alternate RDBMS, even if it is deployed in the cloud, is not the same as implementing a *modernized analytics strategy*. Instead, consider these requirements of a modern data warehouse built for the cloud:

- **Data accessibility:** Don't let increased data volumes become a bottleneck to productivity. Real-time data ingestion and processing coupled with a data lake will reduce complexity and accelerate time to value for a broad array of data analytics consumers.
- **Scalability:** In a modern analytics platform, end users are not impeded by lackluster computational performance. Migrating to a cloud-based platform leverages elasticity and automated scaling to meet computational and storage scalability demands and does so transparently.
- **Reduced data sprawl:** Migrating data to a cloud-based data warehouse provides the ability to classify and organize the scope of data assets, which in turn raises data awareness and end-user enablement.
- **Data breadth:** Cloud services help virtualize access, enabling querying against external data sources while simultaneously supporting data ingestion from hundreds of sources. This breadth enriches the ability to derive analytics insights.

- **Cloud economics:** In a cloud environment, you pay for what you use when you need it instead of requiring a large capital investment in on-premises hardware. Fold in the reduction in operations and maintenance costs (resulting from the cloud host taking that responsibility) and this yields a lowered total cost of operations.
- **Data protection:** Cloud vendors have been diligent in providing layers of identity access management and additional mechanisms for guarding against unauthorized access to cloud-based data assets. Policy-based and role-based access management provide additional layers of protection.
- **Analytics services:** In a modernized analytics environment, data consumers are not limited to running descriptive reports from dimensional database models. Cloud vendors provide an array of value-added predictive analytics tools and out-of-the-box machine learning models that complement the conventional structure of a data warehouse. In turn, data warehouse vendors seamlessly integrate into the partner ecosystem, allowing analysts to use familiar 3rd party analytics tools.
- **Real-time analytics:** You can integrate your migrated analytics environment with real-time streaming data to analyze business events as they happen.
- **Integrated analytics:** You can integrate models that can identify patterns and opportunities in real time and automatically trigger prescriptive actions.
- **High availability:** Modernized environments are layered on top of fault-tolerant modules and redundancy to enable high availability of analytics capabilities.

There are additional considerations when selecting a cloud data warehouse vendor that will not only provide a modernized platform environment but also streamline your migration from obsolescing systems such as Netezza. Keep these topics in mind:

- **Simplified migration:** Cloud vendors provide services to help assess an existing data warehouse landscape and prepare for migration.
- **Simplified management:** Reduce the operational management and maintenance burden because the cloud vendor manages the platform for you.
- **Leveraging analytics services:** Cloud vendors provide an array of value-added services, particularly in the areas of machine learning and artificial intelligence.
- **Simplified data pipelines:** Managed cloud platforms provide services for the end-to-end data workflows, allowing for orchestration of data ingestion, transformation, and machine learning pipelines.
- **Shifting the focus:** Adopting a fully managed cloud platform environment eliminates much of the operational overhead of system management, allowing your analysts to focus their attention on analysis and not of mechanics of data warehouse population.

Preparing for Migration and Managing the Migration Process

Once the decision is made to migrate to a modern cloud-based data warehouse/analytics platform, it is important to properly plan and manage the migration process. Consider these steps:

- **Understand the operational context:** Data democratization and the emergence of the citizen data analyst dramatically expand the pool of actors engaged with the migrated data warehouse environment. Before migrating, solicit information about the operational context: data analysts, the analytics consumers, the types of reports and analyses being performed, along with documentation of operational procedures.
- **Perform a data survey:** Assess the data landscape to identify, classify, and categorize (for usage scenarios and for sensitivity) all data assets used for reporting and analysis.
- **Profile use cases and workloads:** Identify and profile the individual workloads. Review operational tasks and corresponding scheduling, understand resource usage within the existing environment, determine whether there are tasks that can be executed in parallel, and look for dependencies across use cases.
- **Define success measures and SLAs:** Define success criteria, service-level agreements, and corresponding measures. This will facilitate the proper prioritization and staging and enable continuous monitoring of progress during the migration process.
- **Prioritize use cases:** Use the success measures to evaluate and determine the use cases that are of highest criticality and value and prioritize accordingly.
- **Assemble a migration plan:** Specify the completion criteria; perform time, cost, and resource estimates; conduct a proof of concept or a pilot to identify any specific issues; then map out a migration plan that, for each use case, includes data migration, application migration, migrating workloads and data pipelines, verification and validation, and implementation.
- **Execute the migration plan:** Given a prioritized list of use cases and an iterative procedure for use-case migration, develop a more detailed migration plan aligned with your organization's development and deployment methodology.

At this point, your organization will be positioned to engage cloud platform vendors to evaluate suitability of their offerings to meet the needs of the modernized data warehouse platform.

TDWI Analyst's Perspective: Migrating to Google BigQuery

After reviewing the Google Cloud Platform (GCP) offering for a modernized data warehouse platform, it is clear that (similar to other large-scale cloud providers) the fundamental capabilities satisfy the requirements of a modern data warehouse.

cloud.google.com

BigQuery, Google Cloud's serverless, highly scalable and blazing-fast enterprise data warehouse, is designed to help organizations make informed decisions quickly so they can transform their business and stay competitive. BigQuery provides unified stream and batch processing to support integrated real-time analytics and it delivers out-of-the-box machine learning models to accelerate predictive analytics. With BigQuery, organizations have a highly reliable data warehouse that delivers a 99.9% uptime SLA, and because there is no infrastructure to set up or manage, organizations can jumpstart data analysis cost-effectively and accelerate their digital transformation journey with ease.

For security and data protection, Google Cloud offers fine-grained identity and access management controls, provides services for classifying and masking sensitive data, always-on data encryption, along with a data catalog for metadata management. Finally, Google Cloud streamlines data warehouse migrations with a complete migration guide, architecture advisory guidance and workshops, and resources and partner tools.

- Migration offer: g.co/cloud/dw-migration-offer
- Technical data warehouse migration guide: g.co/cloud/dw-migration-guide

About the Author

David Loshin, president of Knowledge Integrity, Inc. (www.knowledge-integrity.com), is a recognized thought leader, TDWI affiliate analyst, and expert consultant in the areas of data management and business intelligence. David is a prolific author on topics related to business intelligence best practices. He has written 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, online seminars, and sponsored websites and channels including TechTarget and The Bloor Group. His best-selling book, *Master Data Management*, has been endorsed by many data management industry leaders. David is also the Program Director for the Master of Information Management (MIM) program at University of Maryland's College of Information Studies.

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

About TDWI Research

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