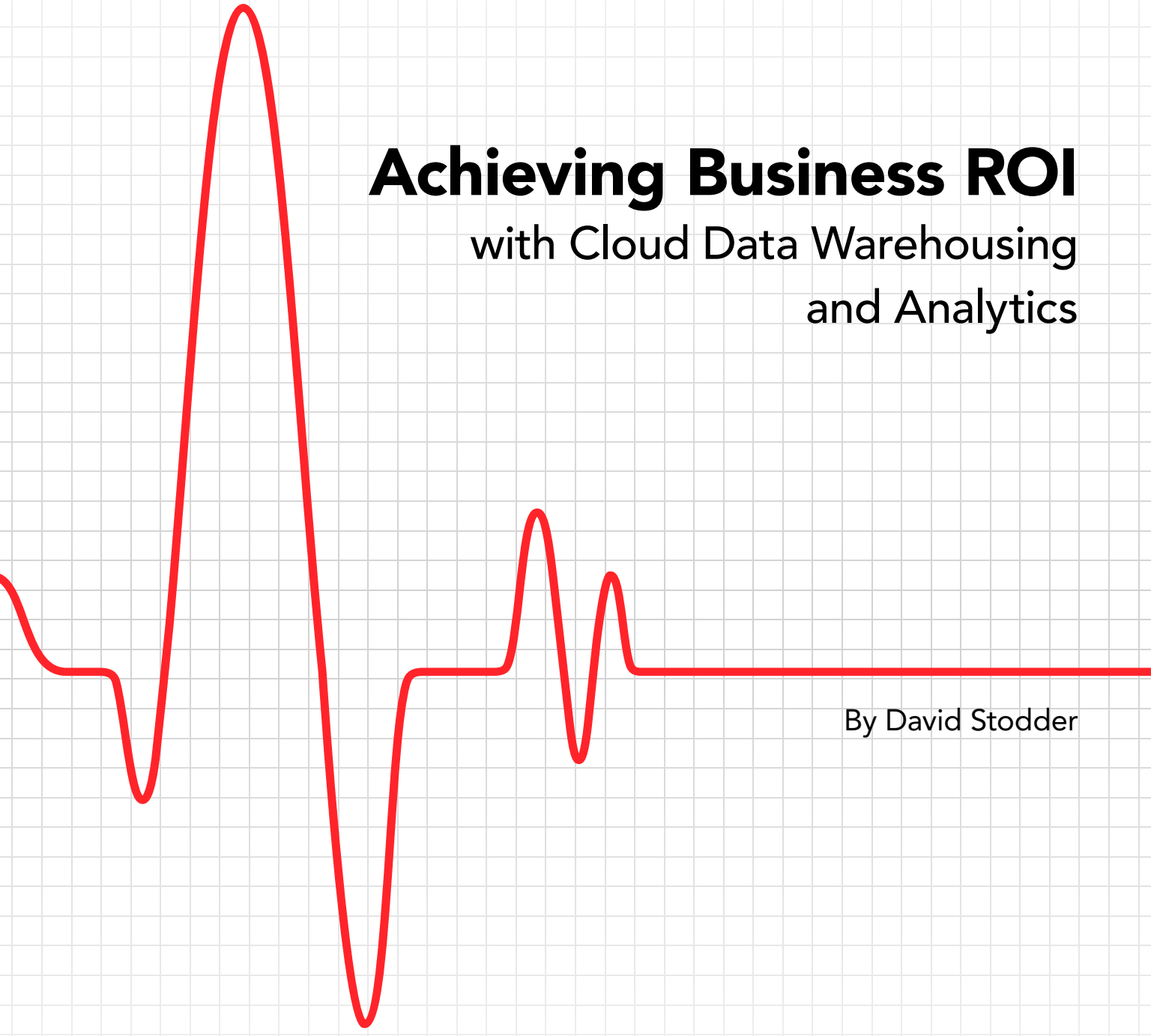
research

PULSE

Achieving Business ROI

with Cloud Data Warehousing and Analytics



By David Stodder

Sponsored by:

HEXAWARE

tdwi
Transforming Data With Intelligence™

Achieving Business ROI with Cloud Data Warehousing and Analytics

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 research and advice for data professionals worldwide. TDWI Research focuses exclusively on data management 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 data management 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.

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), Philip Russom (prussom@tdwi.org), and David Stodder (dstodder@tdwi.org).

Acknowledgments

TDWI would like to thank many people who contributed to this report. First, we appreciate the many professionals who responded to our surveys. Second, our report sponsor, who diligently reviewed outlines and report drafts. Finally, we would like to recognize TDWI's production team: James Powell, Peter Considine, Lindsay Stares, and Rod Gosser.

Sponsors

Hexaware sponsored the writing of this report.

The Pulse: Organizations Are Shifting to the Cloud for Business Advantage

Data management and analytics are in the midst of a sea change as organizations adopt cloud-based services to augment or replace existing on-premises systems. Early hesitancy about whether cloud-based systems are secure enough for these data "crown jewels" has fallen away. Although data security remains a top priority, more organizations are satisfied that security in the cloud is dependable and they can move ahead with migrations. Recent TDWI research finds that nearly two-thirds of organizations surveyed (64%) are planning to apply cloud data management to support analytics and about half are moving to the cloud specifically for data warehousing and reporting.[1]

Thus, the focus today is less on justifying the move to the cloud and more on how to get there faster and with less difficulty—and once in the cloud, how to manage data and analytics so organizations can realize greater business value and return on investment (ROI) than they could with strictly on-premises systems. The cloud can play a big role in freeing organizations from the physical and economic constraints imposed by on-premises systems as well as legacy practices that have been shaped by constraints on volume, data variety, speed, scalability, and compute power. Largely due to these constraints, older, traditional data warehouses and business intelligence (BI) systems have become rigid, expensive to augment, not scalable, and too slow for new business-critical workloads.

Cloud-based data platforms, storage, and analytics offer organizations opportunities to support new data-driven projects that may be too difficult, expensive, or slow to develop with existing systems. By augmenting existing systems with cloud-based systems, they can open up opportunities to explore new data, try new types of analytics, artificial intelligence (AI), and database technologies, and use the cloud's potential for quicker setup to meet immediate business needs. Some organizations are going beyond augmentation and are replacing ("re-platforming") on-premises systems entirely with new cloud-based systems.

TDWI research finds that an overwhelming percentage of organizations surveyed (86%) regard cloud data management as important to the success of their data strategy.[2] Top reasons cited by organizations surveyed include:

- Expanding the use of data with lower up-front investment
- Being able to afford newer ways of achieving scale and speed
- Modernizing data management to lay the foundation for cost-effective machine learning and deep learning analytics
- Having moved business applications and processes to the cloud, they need to move corresponding data processes there as well
- Responding to company strategies and IT mandates to move all systems to the cloud
- Fostering better collaboration among employees, external suppliers, partners, and customers

[1] See the *TDWI Best Practices Report: Cloud Data Management*, online at tdwi.org/bpreports
[2] Ibid.

New Options for a New Environment

Whether organizations are augmenting or replacing on-premises systems, for at least some period of time they will need to manage analytics, data warehousing, data integration, and other data management across both on-premises and cloud systems. Illustrating this point, 2019 TDWI research found that 64% of organizations surveyed are reliant on on-premises data warehouse systems while 36% are reliant on cloud-based data warehouses.[3] Adding to this "hybrid" reality of on-premises and cloud-based systems, many organizations report that they have cloud-based systems with more than one cloud provider, that is, they have a multicloud environment.

Multicloud will be common for data management because organizations have multiple options and may have different reasons for choosing each one. Data warehouse platforms include those provided by the dominant infrastructure-as-a-service (IaaS) providers: Amazon Web Services (Redshift), Microsoft Azure (SQL Data Warehouse), and Google Cloud (BigQuery). A significant number of independent vendors also offer cloud-based data warehousing or analytics services, including Cloudera, IBM, Oracle, Snowflake, and Teradata. Others offer data integration, transformation, pipelining, metadata management, and data preparation services that are important to cloud-based data warehousing, BI, and analytics. These solutions can help organizations save time and expense because they are set up and optimized for the cloud.

However, making the switch to (or incorporating) new platforms is never as simple as it seems. There are challenges regarding data migration, movement, and performance. Running workloads on data warehouse and analytics platforms in the cloud requires attention to issues that may not have been critical in the on-premises world but can affect user satisfaction. Finally, governance of the entire environment is essential to meeting regulatory and data quality requirements. Success in addressing new challenges is important to the ultimate goal of delivering business value and ROI. In this TDWI Pulse Report, we will discuss five key areas that demand focus.

From Raw to Refined: Extracting Business Value from Data

Data begins as a raw material that must be transformed and enriched to make it useful for most business requirements. Organizations want to draw insights from this raw material to run operations more efficiently and effectively, learn how to satisfy existing customers and land new ones, reduce fraud, innovate with product and service development, and more. Ongoing digital transformation continues to generate new types and sources of data, from online customer behavior logs to machine data collected from Internet of Things (IoT) devices. Different use cases will require different ways of preparing and enhancing this data so that it becomes meaningful and valuable.

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

Thus, having flexibility and speed in how your organization collects, transforms, analyzes, and collaborates on data is essential to success. Cloud platforms can help organizations increase flexibility in data warehousing environments, including the cycle of data integration, preparation, quality, and data pipeline development activities needed to meet different demands. An organization introducing a new product or service, for example, could begin analyzing customer data much sooner, including near-real-time data, because it would not have to wait to set up systems on premises.

Scalable and elastic platforms in the cloud can help organizations address growing and diverse workloads involving analytics, data visualization, and AI techniques such as machine learning. With cloud computing, organizations have more options for getting the right amount of speed, scale, breadth, and support for numerous and concurrent data-driven workloads—challenges that organizations of every size today face as they collect and refine data. Data science workloads for developing predictive models and machine learning algorithms have different requirements than those for business users seeking another layer of data interaction to analyze metrics in a dashboard. Cloud platforms can give organizations the flexibility to serve both types of workloads in different ways.

Is your organization planning to augment or replace its existing BI, analytics, and data warehousing systems with any of the following systems or cloud-based services, solely or in combination? (Please select all that apply.)

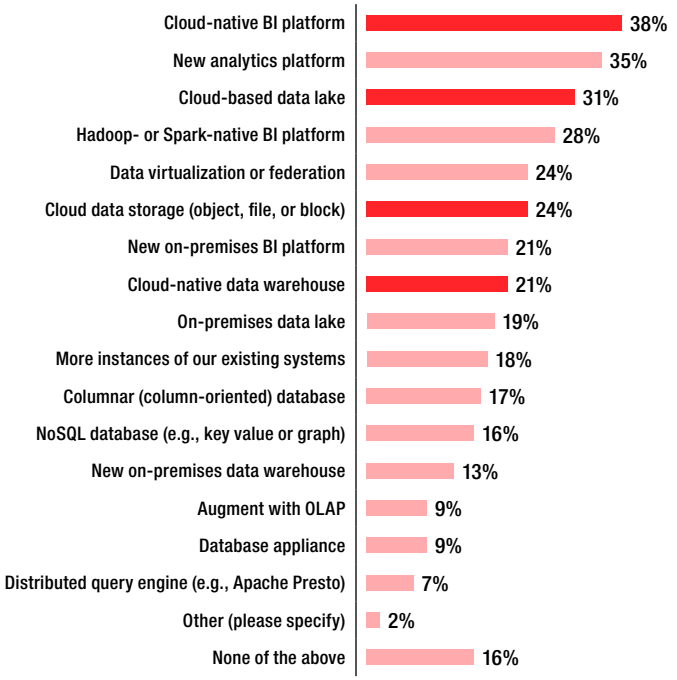


Figure 1. *Based on answers from 232 respondents. From* TDWI Best Practices Report: BI and Analytics in the Age of AI and Big Data. *Highlighted bars denote cloud technologies.*

Cloud computing is having an impact on every phase of the data refinement life cycle that provisions users, applications, and automated processes with data. Cloud platforms now have a data gravity of their own, pulling ever more data toward them from cloud-native software-as-a-service (SaaS) applications as well as other applications and processes developed by user organizations themselves.

It is therefore easy to see why TDWI and other industry research groups see strong growth ahead for cloud-based data warehouses, data lakes, and data integration and preparation systems, as well as BI and analytics services in the cloud. This can be seen in Figure 1, which shows intentions for augmenting or replacing existing BI, analytics, and data warehousing systems. Cloud-based solutions figure prominently in surveyed organizations' plans with higher percentages shown than for on-premises alternatives.

The Cloud's Impact on BI and Analytics

Cloud computing is reshaping BI and analytics platforms. BI is about enabling users to understand data, look at how data relates to other data, and use visualization to see and communicate insights. BI has traditionally focused on structured, relational data and models; however, some BI solutions are able to integrate text and other semi- and unstructured data into user environments such as dashboards. BI solutions and platforms cover "the last mile," so to speak, in the flow of actions performed on the data to take it from a raw, often dirty state to a cleaned, transformed, and enriched state, particularly for nontechnical users' needs. Some solutions and platforms incorporate more preparation and data cataloging (or metadata management) activities than others. Some also optimize queries for performance or work with the underlying database management system to ensure good performance.

Cloud-based or SaaS BI solutions essentially expand and streamline what on-premises BI platforms can do. Using the scalable and elastic compute and processing platforms available in the cloud, BI users no longer have to limit the amount of data they can access to what is in the BI server, data mart, or data warehouse. Setting up the tools is obviously easier and faster because the solutions are in the cloud, addressing another common complaint about BI taking too long to develop. However, organizations still need to determine much about the data, including what the sources are, how much and what kind of transformation and preparation should be done, what level of performance and query speed are necessary to accomplish objectives, and what governance is needed.

Analytics platforms also benefit from a far less limited environment with cloud computing. BI platforms traditionally focus on reporting, dashboards, performance metrics, and limited ad hoc querying related to set business questions. Analytics platforms complement BI; they are oriented more toward deeper and broader data exploration. Analytics workloads typically involve development and testing of predictive models, exploring new data for patterns and trends, and finding unexpected correlations and data relationships across sources.

Data lakes and now cloud-based object storage do not impose a rigid data model, which makes them useful for advanced analytics that need to examine a variety of data, not just structured, relational data (although many analytics workloads do benefit from the cleansed and structured data provided by a data warehouse). Today, analytics extends to AI and machine learning to dramatically expand the scale, speed, and depth of exploration and analysis.

Cloud computing is playing a key role allowing analytics to push past the scale and speed limits of on-premises computation and processing systems through use of massively parallel, highly scalable computing in the cloud. Along with reducing restrictions in the volume and variety of data organizations can analyze, cloud analytics platforms can enable users to run analytics workloads when needed for business objectives rather than have to wait for off-hours batch cycles, which is often the case with busy on-premises systems. With less limits on computation and data storage, organizations can even consider applying cloud-based real-time analytics on real-time data streams.

In this next section, we examine five areas that organizations need to address to ensure business value and ROI as they increase the role of the cloud in data life cycles.

Critical Focus One: Improve Data Migration, Integration, and Transformation

The business value of data is largely determined by choices made in what data to source and what to integrate and prepare through data quality procedures, enrichment, and transformation. For example, to get a complete customer perspective and analyze multichannel marketing campaigns, decision makers need well-integrated data that is transformed appropriately so that sensible insights can be drawn from it. Leaders in finance need not only integrated data to support perhaps hundreds of performance metrics but also confidence that calculations remain valid as they view different data sets. Thus, as organizations shift to the cloud, they must invest in continuous improvement of technologies and practices for data migration, integration, and transformation.

Big data sources such as customer behavior log files, social media, IoT, and event data from across diverse channels offer potential for sharper insights if organizations can properly collect and massage the data to fit purposes such as running predictive models, machine learning algorithms, and operational dashboards. Older technologies and practices set up for enterprise reporting and online analytical processing (OLAP) cubes can fall short with increased data volume and variety. Real-time data processing is also becoming more important; many organizations today regard it as a competitive advantage if they can dramatically reduce latency between when data is recorded and when it is ready for analytics and other data consumption.

Recommendations

INCREASE AUTOMATION AND STANDARDIZATION. Software automation is critical to meeting speed and scale demands as well as establishing better standards for how data is migrated, collected, profiled, and transformed. Many organizations still engage in manual hand-coding of extract, transform, and load (ETL) routines, even though modern solutions are available that can automate steps. Cloud-based automation can reduce the cost and time required for data integration and migration. Leading tools today embed AI capabilities that enable smarter profiling, integration, and transformation. Smart, automated data profiling, for example, increases the speed and consistency of statistical methods used to analyze characteristics such as the presence of null values, deviations, and anomalies in data coming from new sources. Profiling is often time-consuming and can require considerable human intervention. The combination of AI and automation helps data engineers and analysts find, collect, evaluate, and address issues in the data faster.

BROADEN DATA MIGRATION AND MOVEMENT STRATEGIES. Simple lift-and-shift migration of data from on-premises to cloud platforms can work, but in many cases, moving data is far more complicated—and it can be slow depending on whether networks can handle heavy data traffic.

Data models and interfaces between data platforms, applications, and BI and OLAP systems must also be checked to ensure full functionality. Organizations should therefore broaden their approach to include options such as data virtualization—which uses metadata and a middleware layer to enable access to multiple data sources without moving the data to a centralized location—and change data capture (CDC) solutions, which can reduce extractions to only the changed data based on preset conditions.

SUPPORT FLEXIBILITY IN DATA COLLECTION AND TRANSFORMATION. A source of conflict between users and administrators of traditional, centralized, IT-managed data warehouses is that these systems are too rigid and inflexible. TDWI research finds that the inability to interact with new data quickly enough is a source of user frustration and a major reason why users set up their own data platforms, leading to disparate data silos. As organizations set up data warehouses in the cloud, they need to prioritize making it easier and faster to incorporate and transform new data. IT can also give users sanctioned alternatives such as easy creation of cloud-based analytics sandboxes and data marts that are derived from the central, cloud-based data warehouse for self-service exploration of new data.

As organizations set up data warehouses in the cloud, they need to prioritize making it easier and faster to incorporate and transform new data.

Critical Focus Two: Once in the Cloud, Improve Data Warehouse and Analytics Platform Management

Spinning up cloud data platforms and moving and migrating data into them is big, but it's only the first step toward getting business value from data in the cloud. The second step is managing the data environment once data is in the cloud. Of course, not all cloud arrangements are the same; in some cases, organizations want a bare metal infrastructure cloud service based on dedicated servers rather than shared resources—this lets them tailor every aspect of how the data is managed. Other organizations will opt for multitenant sharing of server resources with less (or no) hands-on configuration. Organizations must decide what arrangement is best based on cost, needs for infrastructure flexibility, security concerns, and other data management considerations. Many will use a combination of approaches.

Data management priorities will be shaped by the type of cloud service organizations choose, but some priorities remain constant no matter what. Organizations still need adept and agile data management to promote efficiency throughout data life cycles and ensure that users receive the required levels of data availability, performance, and governance. For every type of cloud platform, organizations will need to monitor factors that affect costs, including the cost of data access over networks. Overall, as organizations shift to the cloud, they should modernize data management with the latest technologies for better automation and monitoring.

Leading solutions offer capabilities based on embedded use of AI and machine learning to enable smarter automated data management.

Recommendations

USE ADVANCES IN DATA MANAGEMENT TO SUPPORT GROWTH IN ANALYTICS WORKLOADS. TDWI research finds that enabling analytics is the top priority for most organizations as they set up cloud data services. To align with dynamic business needs, organizations need to reduce delays in scaling cloud data platforms such as data warehouses or data lakes so that critical analytics can be operationalized when needed. Organizations should take advantage of automation capabilities from cloud platforms and third-party solution providers for auto-scaling and automated rebalancing so that organizations gain the cloud's elasticity benefits. Leading solutions offer capabilities based on embedded use of AI and machine learning to enable smarter automated data management, inform administrators about potential issues, and provide recommended options for dealing with those issues.

FOCUS ON CONTINUOUS OPTIMIZATION OF DATA PREPARATION, PIPELINES, AND TRANSFORMATIONS. More than half of organizations surveyed by TDWI (57%) regard improvements in data preparation, pipelines, and transformation as "absolute must-haves."[4] Users of cloud-based data platforms, including customer- or partner-facing applications that monetize data, depend on trusted, quality results from these processes. As with other aspects of data management, organizations should evaluate how smart automation can support continuous optimization and improvement. Organizations should examine how they can modernize these areas to make them faster, more flexible, and reusable to accommodate the variety of analytics use cases and volume of workloads.

USE THE RIGHT DATA PLATFORM FOR THE JOB. Analytics, reporting, AI, and real-time notification workloads all have different data requirements. Organizations may be frustrated if they try to use a data warehouse for all of them. OLAP and dashboard reporting requirements are different from those for machine learning and analytics. Analyzing semi- and unstructured customer behavior data from web logs and social media sources or data streaming from IoT sensors may be more appropriately handled using a data lake or specialized analytics platform. Analytics projects are often compute-intensive with lots of data interaction. Organizations may need to adopt a multicloud data management strategy to ensure that they have a cloud platform provider that can support requirements with the right type of system. Data strategies should embrace a holistic approach that is not limited to one type of data platform; strategies should articulate how organizations will integrate data flows between cloud providers' platforms.

[4] See the *TDWI Best Practices Report: Cloud Data Management*, online at tdwi.org/bpreports

If your organization were to implement cloud data management, what would its leading benefits be? Select seven or fewer.

Figure 2. *Based on 605 responses from 98 respondents. 6.2 responses on average. From* TDWI Best Practices Report: Cloud Data Management.

Critical Focus Three: Reduce Expenses and Increase Scale

Organizations should create a cloud-based enterprise data warehouse (EDW) to support users' reporting and data analysis needs. It should provide a central repository of quality, integrated data drawn from multiple, typically operational sources. The advent of the cloud gives organizations an opportunity to address longstanding challenges that have limited the value of traditional, on-premises enterprise data warehouses.

One challenge has been cost. Fortunately, storing data in the cloud is significantly less expensive; organizations can control costs by taking advantage of the cloud's elasticity to avoid overprovisioning storage and compute resources. They can align use of cloud-based resources on a daily or even hourly basis to fit business and IT needs. Cloud data management services are offering increasingly sophisticated auto-provisioning and auto-scaling so that organizations no longer have to set up their own capabilities and tooling to provision cloud resources. Beyond gaining greater IT efficiency, organizations can increase business ROI through better alignment of compute and storage resources with business needs.

The second biggest challenge has been scalability. According to TDWI research, the leading benefit sought from cloud data management is scalability for data storage and integration workloads followed by automatic and elastic resource management (see Figure 2). The inability to scale on-premises EDWs to support the growing number and variety of analytics-oriented workloads holds organizations back from getting full business value. For example, scalability limitations can prevent organizations from running smarter and more efficient marketing campaigns because they cannot analyze patterns and trends in high-volume data drawn from multiple channels or from different time intervals.

Taking advantage of elasticity requires a mindset shift away from traditional strategies for on-premises systems.

Recommendations

DETERMINE THE EXTENT AND PACE OF EDW MIGRATION TO THE CLOUD. Not all organizations have the same plans regarding how the cloud fits into their EDW strategy. Some plan a limited migration whereas others are doing a complete, end-to-end migration that includes all related data integration and preparation systems, data marts, and other satellite systems. Organizations should examine their current and planned workloads, determine priorities, and set short- and long-term goals. Some key questions to answer include requirements for performance and data freshness, how many concurrent workloads need to be supported, and which current and new data sets will be needed.

MIGRATE IN STAGES DEFINED BY BUSINESS NEEDS. Rather than migrate in one big move, organizations might choose to migrate their EDW environment to the cloud in stages coordinated with defined business needs. Then, the migration can produce quicker wins. If the organization is familiar with agile development methods, the migrations could be defined as iterations. Business stakeholders could test each iteration of the migrated portion of the EDW and provide feedback.

TAKE ADVANTAGE OF ELASTIC SCALABILITY AND FLEXIBILITY. Trends in cloud computing toward separation of computation and data storage are increasing flexibility. Organizations can choose computation frameworks, such as Apache Spark, that best fit their workloads and use object storage from Amazon, Google, Microsoft, or another cloud provider for scalability and easier replication of the data across multiple clouds for computation.

Taking advantage of elasticity requires a mindset shift away from traditional strategies for on-premises systems that have been based on preset capacity limits, resource scarcity, and rigid data architectures. Organizations should ensure that no matter what degree of cloud EDW migration is planned, their strategic mindset takes into consideration the cloud's potential for scalability and elasticity.

Critical Focus Four: Establish Governance and Quality Across All Cloud Data

Governance rules and policies set out how an organization protects sensitive data assets and meets regulatory requirements such as the European Union's General Data Protection Regulation (GDPR). TDWI research finds that many, though not all, organizations integrate data governance with data security. For those that do, governance can include access and authentication management as well as other security considerations.

To improve data quality, many organizations combine governance with a broader idea of data stewardship. Stewardship is about applying best practices and technologies for data quality; stewards can be mentors who advise users about finding, working with, and sharing trusted and relevant data in their analytics, reports, and dashboards.

Data in the cloud has historically been a challenge for organizations to govern. Outside of IT's oversight, business functions and data scientists have spun up cloud data storage or database instances. These data silos then become part of IT's data management and governance challenges, potentially exposing sensitive data, such as customers' personally identifiable information (PII). With GDPR and other regulations requiring organizations to show that they are protecting PII, governance of all platforms is a high priority. Organizations cannot expand further with cloud data management unless they are certain that sensitive data is protected. TDWI research finds that organizations are less confident in their governance of cloud-based data (see Figure 3).

How confident is your organization in its ability to apply governance rules and policies to the following BI, analytics, and data management and integration systems?

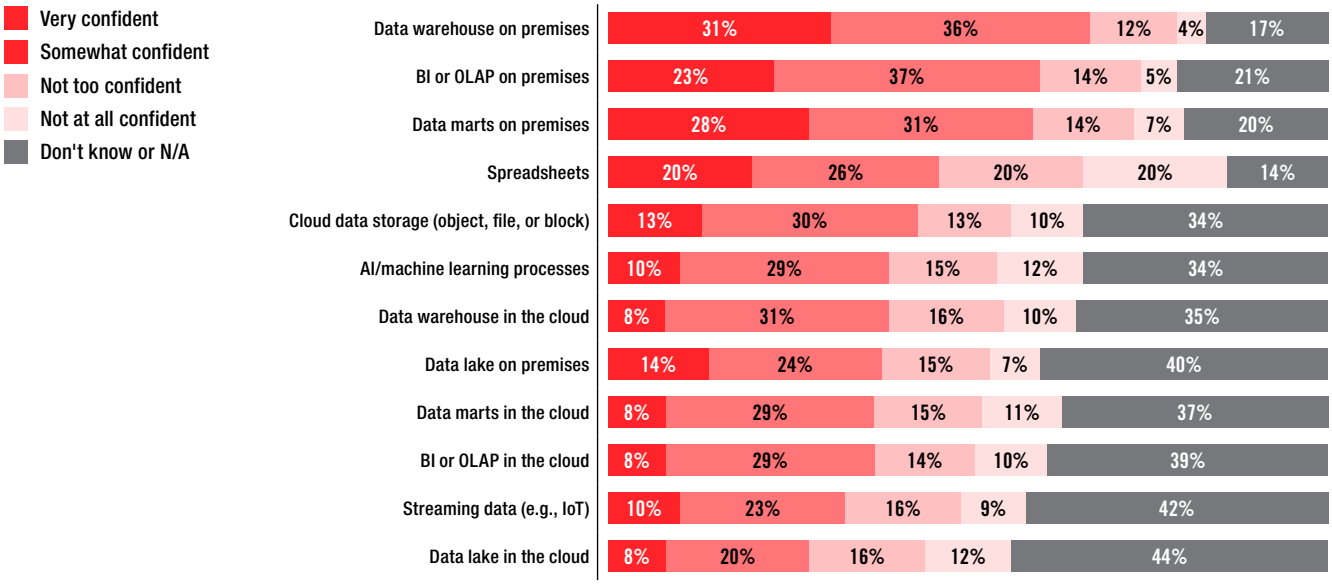


Figure 3. *Based on answers from 198 respondents. Ordered by highest combined "very confident" and "somewhat confident" responses. From* TDWI Best Practices Report: BI and Analytics in the Age of AI and Big Data.

Establishing a data warehouse can help organizations govern data more effectively. Often, organizations will choose to put their most sensitive data in a secure data warehouse where procedures for controlling access are in place. As Figure 3 shows, organizations have more confidence in their ability to apply governance rules and policies to data warehouses on premises. Organizations must now extend their governance expertise to cloud-based data platforms.

Recommendations

APPLY DATA GOVERNANCE RULES TO DATA IN THE CLOUD. To improve governance, organizations should consider consolidating cloud-based data silos that have been spun up over time into a cloud-based enterprise data warehouse. An alternative to consolidation may be data virtualization; users would only be able to access protected data through the data virtualization layer that is integrated with security and authentication procedures. Although consolidation or virtualization for governance is necessary for adherence to rules and policies, it can offer the additional business benefit of guiding users to trusted data.

UNIFY GOVERNANCE ACROSS HYBRID MULTICLOUD DATA ARCHITECTURE. As much as possible, organizations should apply consistent rules and policies wherever their data resides—across on-premises systems and on multiple cloud providers' platforms.

JOIN GOVERNANCE WITH DATA STEWARDSHIP TO IMPROVE DATA QUALITY. Organizations need to take steps to fight the spread of poor-quality data that is being shared among users both inside and outside the firm. Data stewardship can increase attention to the importance of data quality; stewards can help users make sure data quality is a priority in the development of data pipelines and transformations. Advanced solutions in the market use machine learning to improve data quality management across sources that are high in volume, variety, and velocity.

Critical Focus Five: Manage Metadata for Data Location, Access, and Governance

As organizations amass data in the cloud across multiple providers' platforms, it can be difficult for users, administrators, and data engineers to locate data that is valuable and relevant to pertinent business questions. Metadata is a key resource of information about the data, such as what it means, who created or acquired it, and its lineage. Metadata management across platforms is important for easing data integration, access, sharing, and analysis of data relationships. Data catalogs and glossaries are services that contain records of the metadata, the data's location, and other key information about schemas and data models.

TDWI research finds that the majority of organizations would like to have solutions for coordinating data meaning across sources and resolving inconsistencies, including for data in the cloud. Historically, metadata and data catalog creation and management has involved considerable manual work. Software solutions are now available to automate steps and apply machine learning, search, natural language processing, and other AI technologies to discover, integrate, and manage metadata as well as find errors that need to be corrected.

An important part of many EDWs is their data catalog or similar repository for metadata management. However, the scope of EDW metadata in traditional, on-premises systems has

historically been just the EDW itself; applications such as BI tools will coordinate their metadata with the EDW metadata. Ideally, a cloud-based EDW should have a more easily extensible metadata repository that enables users to examine and analyze metadata, data meanings, and data relationships as they may exist in both current and new data, including in sources outside the EDW, such as a data lake.

Recommendations

USE METADATA TO RESOLVE DATA INCONSISTENCIES ACROSS SOURCES. As users integrate diverse data for single views of customers or other objects of interest, there will be inconsistencies and irregularities. Ignoring them, however, will lead to visualizations and analytics full of erroneous information and misleading conclusions. Organizations should employ modern tools to improve definitions and semantic understanding of data relationships. The tools can help administrators and data engineers account for discrepancies and collaborate with users (as subject matter experts) to remedy them where possible.

USE METADATA MANAGEMENT TO IMPROVE GOVERNANCE AND TRUST. Good metadata makes it easier to understand and track data location and lineage, which are essential to governance and protecting sensitive data. Metadata is also important to users' trust in the data. When there is less confusion about how the data is defined and greater confidence that inconsistencies will be found and addressed, users are more confident in basing their analytics and actions on the data, which is key to realizing ROI.

A Final Point: Accountability and Leadership

It can be a sizeable undertaking to shift an organization's center of data management and analytics gravity to the cloud. Keep in mind, however, that people and organizational issues are just as important as technical and process issues, if not more so. Rather than make changes in a piecemeal, disorganized fashion, organizations should consider assigning accountability for the cloud transition to a chief data officer (CDO) or other executive with similar authority. One key focus of the CDO role would be to coordinate between business and IT stakeholders to ensure good communication and continuity as the organization shifts to the cloud. The CDO role can also make sure that all stakeholders are aware of relevant governance rules and policies and see to it that reporting and transparency requirements are met. Additionally, the CDO can encourage stakeholders to fix mistakes of the past rather than just carry over bad habits from older systems and processes. These should be improved and modernized to fit a cloud-centric environment.

The CDO role can also help business and IT find the right balance between self-service analytics and the centralized EDW. Users need to understand the value of a centralized EDW and why it is worthwhile to give up some autonomy to gain better data quality, data pipeline and preparation efficiency, and scalable and reliable performance. At the same time, IT needs to fully understand the levels of flexibility and independence that users need to succeed with analytics. A CDO can foster communication between business and IT. Setting up a center of excellence with stakeholder participation is a good way of facilitating regular communication.

Finally, organizations should expect failures during the transition. Stakeholders may find that data models are not quite right, interfaces need adjustment, important data may be missing, and so on. Organizations should have contingency plans that include rollback to former on-premises systems

until the transition is fully successful. Preparing for failures ahead of time will reduce user disruptions during the transition. This will give users confidence as they move forward to take advantage of new cloud data warehouse and analytics platform capabilities.

About Our Sponsor

Hexaware Technologies is one of the fastest-growing IT service providers, with a focus on helping enterprises "*Automate Everything, Cloudify Everything* and *Transform Customer Experiences.*"

As part of its mission to make enterprises agile, nimble, and easily scalable, Hexaware has built a platform called Amaze™ for Data & AI that helps enterprises automate the transformation of their data warehouse and analytics ecosystems to the cloud (including Amazon Web Services, Microsoft Azure, and Google Cloud Platform).

HOW CAN AMAZE™ FOR DATA & AI HELP ENTERPRISES THROUGH AUTOMATION?

- Over 60% reduction in time, cost, and efforts in the transformation to cloud services
- Swift and comprehensive assessment of current landscape (ETL, DB, reporting) through metadata
- Avoid business disruption through Hexaware's foolproof cloud transformation methodology

HOW IS AMAZE™ FOR DATA & AI DIFFERENT FROM OTHER SOLUTIONS IN THE MARKET?

- End-to-end integrated platform, from assessment and cloud recommendation to transformation
- Prebuilt connectors for 25+ platforms and 50+ assessment reports
- AI/ML libraries for multiple use cases spanning diverse industry segments

Hexaware has proven, industry-specific use cases and success stories in transformation to cloud, helping enterprises move from legacy databases and massively parallel processing (MPP) appliances to cloud data and analytics landscapes.

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.

Transforming Data With Intelligence™

555 S. Renton Village Place, Ste. 700
Renton, WA 98057-3295

T 425.277.9126
F 425.687.2842
E info@tdwi.org

tdwi.org