TDWI INSIGHT ACCELERATOR

Why the Data-Driven Marketer Needs a Cloud Data Warehouse

By David Loshin

MARKETING'S CHALLENGES

Marketing uses a lot of data. Increasingly collected from a wide range of customer touch points, this data is used to analyze prospect and customer behaviors, better understand an audience's needs, make real-time decisions, and create personalized messaging to influence future customer behaviors.

"Data-driven" marketing is rapidly becoming commonplace. According to industry analysts at Gartner, in 2016 two thirds (69 percent) of marketing leaders expected that the majority of their decisions would be driven by data by 2018[1]. In 2018, over three-quarters (76 percent) of marketers reported that the majority of their decisions are data-driven, with 81 percent expecting that to be so by 2020[2].

Clearly, the marketing function is the beneficiary of any organization's data intelligence operations. Effectively collecting, preparing, and enabling the analysis of a growing pool of information sources can lead to measurable benefits in terms of using predictive models to help with intent-driven marketing, omnichannel targeting, and sales acceleration. That does not mean that the benefits of a data-driven marketing strategy are free from challenges. Supporting the requirements of the marketing function requires attention to a number of technical challenges, including:

- **THE COMPLEX ANALYTICS PLATFORM.** Providing analytics capabilities to business users relies on a reporting and analytics platform, yet there is often a degree of complexity involved with the design, development, implementation, and continued management and maintenance of this platform.
- **DATA INGESTION AND INTEGRATION:** These platforms rely on data that originates from a variety of internal and external sources. Internally, these data sets are typically organized around business functions but are often marred by data redundancy and replication. External data sets often are acquired without knowledge of how they were created and are subject to data quality problems. Another summary point from Gartner's 2018 report noted that in most cases, "integrating and formatting data" consumed a large part of their team's time[3].
- **DATA VARIETY.** Marketing groups increasingly want to use a wide variety of "well-organized" data (from existing databases, for example) and more "free-formed" data sources (such as social media channel data or weblogs).
- **PLATFORM INTEROPERABILITY.** Because many marketing functions use hosted and cloud-based marketing automation tools and systems, a marketing analytics environment must interoperate with these other systems.

[1] "Key Findings from Gartner Marketing Analytics Survey 2018," Chris Pemberton, accessed via https://www.gartner.com/smarterwithgartner/key-findings-from-gartner-marketing-analytics-survey-2018/ November 9. 2018

[2] "Marketing Data and Analytics Survey 2018: Messy Data and Mismatched Resources Undermine Marketing Teams," Kune, Lizzy Foo, Virzi, Anna Maria, accessed via https://www.gartner.com/doc/3883171/marketing-data-analytics-survey- November 9, 2018

[3] Ibid.

- **ANALYTICS BREADTH.** Individuals will rely on different types of analyses to identify business opportunities, requiring a broad set of reporting and analytics functions.
- **REAL-TIME DECISIONINGS.** Transaction data must be consumed and analyzed in real time; decisions based on this analysis must be delivered to the appropriate targets to exploit the opportunities.
- **SPEED OF DELIVERY.** Increasingly, time is of the essence. If analyses can identify patterns for further investigation, these insights should be rapidly communicated to numerous data consumers.
- **DISTRIBUTED WORKFORCE.** As virtual organizations grow, the limitations of a centralized system are magnified; there is a growing need to satisfy the requirements of an increasingly geographically distributed workforce.

GOING BEYOND THE DATA WAREHOUSE

There must be a framework for acquiring, integrating, and analyzing data and *data warehouse* is the name of the system intended for the accumulation of data from a variety of sources and reorganization of that data to enable end-user reporting, analysis, and business intelligence. For example, a marketing data warehouse might combine information about prospects and customers from internal systems with marketing automation products and services to enable pattern analysis and predictive modeling to improve marketing activities.

Marketing needs a data warehouse, especially for combining data from many sources for analysis. A conventional data warehouse ecosystem bridges corporate transaction and operational systems by extracting data, bringing that data to a staging area for harmonization, standardization, and reorganization into a data model that is engineered for rapid responses to queries and creation of reports. In the past, this type of system might have been adequate to drive the creation of customer profiles and use them to make marketing decisions.

Yet as we have already noted, there are technical challenges that have emerged since the maturation of this conventional data warehouse model that make it increasingly insufficient to meet today's (and tomorrow's) marketing analytics needs.

As an example, modern marketing functions cannot be limited to solely relying on structured data organized in tabular form that can be extracted from relational databases that originate within the corporate boundaries. Marketers have expanded their horizon to embrace software-as-a-service (SaaS) marketing tools such as Salesforce, Marketo, Eloqua, Google Analytics, and Hubspot, all of which provide data in structured form. These structured data sets can be integrated into a conventional data warehouse.

The challenge, however, comes when attempting to blend in data that is not in tabular form. Some data sources produce data in *semistructured* format, combining semantic tags or markers to define the content, which may contain some free-form text. Examples include feeds from social media channels such as Twitter or Facebook, weblogs, application logs, or content-experience or orchestration tools such as Uberflip or Engagio. According to the 2017 Salesforce "State of Marketing" report, "88 percent of high-performing marketing teams work in tandem with service to manage inquiries and service requests from social media channels[4]." To incorporate this semistructured data, much of which is produced by continuously operating environments (such as social media channels) or cloud-hosted tools, you need to go beyond the conventional data warehouse.

[4] Davis, Jon Suarez, "Salesforce Releases Fourth Annual 'State of Marketing' Report – Marketing Embraces the AI Revolution," accessed via https://www.salesforce.com/blog/2017/06/fourth-annual-state-of-marketing-report.html November 10, 2018

ADOPTING A CLOUD DATA WAREHOUSE

Addressing these challenges requires alternative reporting and analytics platforms that can adapt to the rapidly evolving universe of marketing tools, services, and data sources. Obviously, because many of these services are hosted in the cloud, the logical first approach would be to consider a cloud data warehouse. Simply put, a cloud data warehouse is an implementation of a conventional data warehouse architecture (including the core relational database system, computing instances for staging and integrating data, and bulk data transfer into the data warehouse).

There is a significant difference between an implementation of a conventional data warehouse architecture on cloud-hosted platforms and a data warehouse that has been designed to exploit the types of capabilities and services supported by the cloud provider. A conventional data warehouse implemented in the cloud is still bound by the same limitations–the complexity of the platform, the need for extraction, transformation, and loading, and an inability to ingest both structured and semistructured data. Furthermore, the interoperability challenges become even more acute with the need to not just span different systems, but also the need to cross the on-premises/cloud boundary.

A data warehouse engineered for the cloud does not necessarily conform to the traditional architecture. Instead, it can leverage key features of the underlying cloud provider's services, flexibility, and elasticity to provide a range of analytics capabilities with scalable performance at a manageable cost. Some examples of distinctive characteristics of a data warehouse engineered for the cloud include:

- **SERVICES-BASED.** A cloud data warehouse is provisioned as a service, so the client does not take on the burden of configuration, management, maintenance, or tuning of the environment, nor are there acquisition costs.
- **SCALABILITY.** The platform is scalable to accommodate today's growing data volumes and is easily capable of accommodating additional data sources and data streams.
- **SUPPORTS SEMI-STRUCTURED DATA.** Cloud data warehouses can exploit cloud services to parse and ingest both structured and semistructured data.
- **REAL-TIME.** A cloud data warehouse can be configured to support real-time ingestion, particularly from streaming data.
- **INTEROPERABILITY.** A data warehouse engineered for the cloud is designed to provide seamless integration with both on-premises systems and other cloud applications.
- **MULTI-TENANCY.** A single data warehouse instance can serve many departments and business use cases, as well as support data sharing among partners.

CLOUD DATA WAREHOUSE ADOPTION TRENDS

Recognizing the benefits, organizations are rapidly adopting cloud data warehouses. In a TDWI 2017 survey about advanced analytics, respondents were asked about the kinds of data management systems already in use and the types of technologies for which there were plans to adopt within the next 2 years. As can be seen in Figure 1, although only 23 percent of the group was currently using a cloud data warehouse, an additional 48 percent anticipated using one within two years, bringing the total to 71 percent should respondents stick to their plans. In addition, the use of a data warehouse in the cloud had the highest expected rate of growth among all the technologies presented[5].

[5] See the 2017 TDWI Best Practices Report, *Advanced Analytics: Moving Toward AI, Machine Learning, and Natural Language Processing*, online at tdwi.org/bpreports.

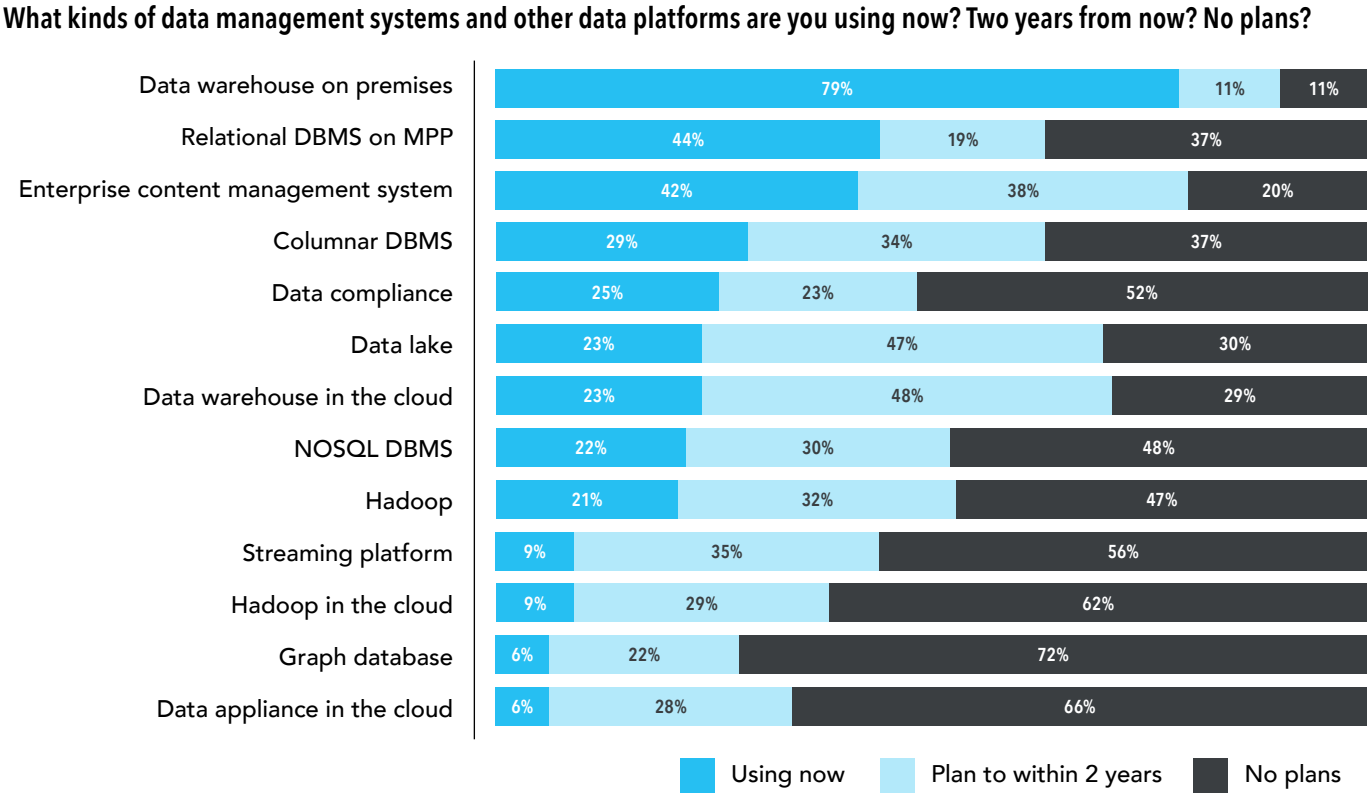

Figure 1: *Current and planned use of data management systems and platforms for reporting and analysis. (Totals may not equal 100% due to rounding.)*

Although on-premises, conventional systems based on relational database management systems (RDBMSs) remain a significant part of the enterprises for organizations that employ a multiplatform data architecture, the cloud is certainly gaining its adherents, as can be seen in Figure 2[6]. Half of respondents reported using on-premises data integration platforms and data warehouses; 29 percent indicated that they were using a data warehouse in the cloud, representing a steady migration.

This rapid adoption rate is not surprising considering the potential benefits of the cloud platform in general and cloud data warehouses in particular. In a 2017 TDWI research survey of 161 respondents, the most frequently identified drivers for adopting a cloud data warehouse by those individuals whose organization was in the process of transitioning to a cloud data warehouse are (in descending order of importance) economic factors (flexibility in the cost models associated with cloud computing), the ability to take advantage of the features and services the cloud host provided, and the ability to achieve greater performance. For individuals in an organization that had already moved to the cloud, the first two drivers were the same, but the third most frequently noted driver was that their exiting on-premises data warehouse was reaching its limits. Other important drivers can be seen in Figure 3[7].

[6] See the 2018 TDWI Best Practices Report, *Multiplatform Data Architectures*, online at tdwi.org/bpreports.

[7] Personal communication from Fern Halper, TDWI senior director of research

For the MDA that you use most, what types of data and compute platforms are involved? Select all that apply.

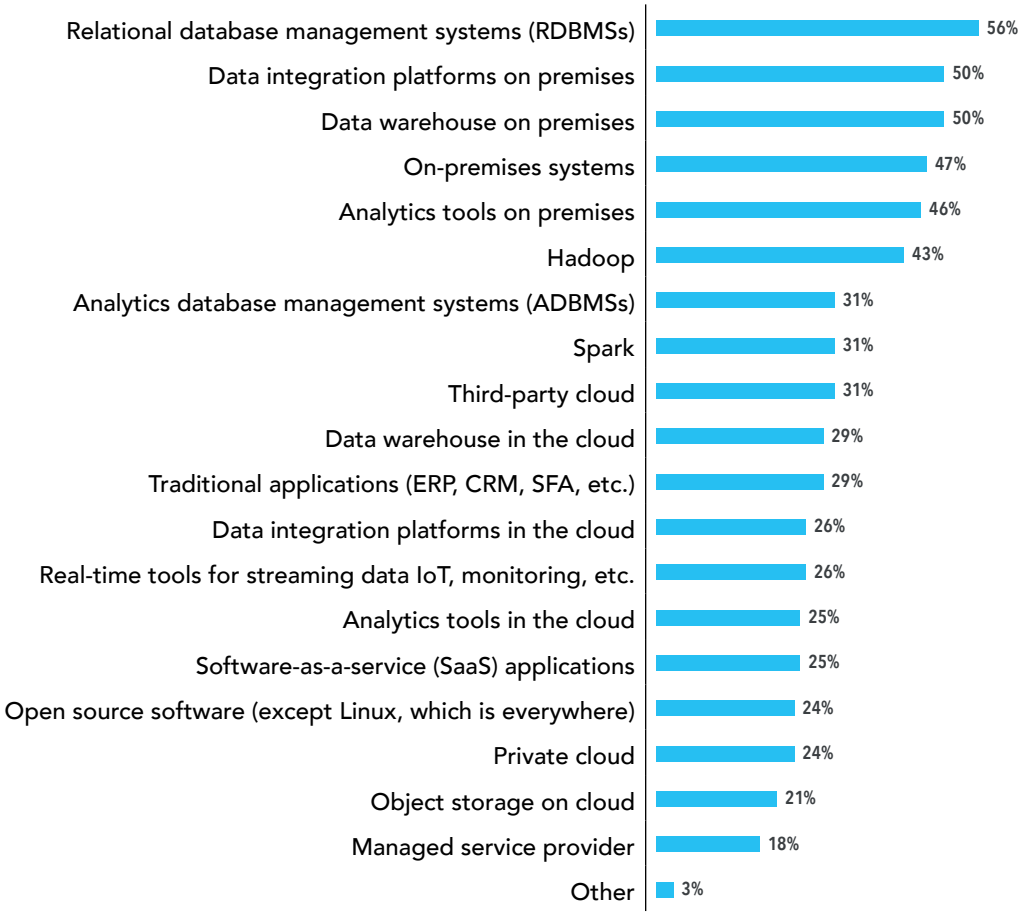

Figure 2: *The types of data and compute platforms used in a multiplatform data architecture.*

BENEFITS OF CLOUD DATA WAREHOUSE FOR MARKETING

Adopting a cloud data warehouse will address some of the key challenges outlined earlier in this report:

- **SIMPLICITY IN IMPLEMENTATION AND INTEGRATION.** A data warehouse engineered for the cloud reduces the complexity of designing, developing, and implementing a platform for reporting and analytics. With the right ecosystem, a cloud data warehouse will work with supporting tools to streamline data integration, especially when it comes to ingesting streaming data.
- **DATA VARIETY.** Cloud data warehouses can ingest both structured and semistructured data.
- **ENHANCED ANALYTICS.** A data warehouse in the cloud enables enhanced analytics capabilities by virtue of its accessibility to additional cloud services, its virtual proximity to other cloud-based systems and services generally used by marketing groups, and by its ability to automatically scale its available resources to meet computational needs.
- **REAL-TIME DECISION MAKING AND RAPID RESPONSE.** Similarly, computational scalability coupled with immediate access to streaming data allows a cloud data warehouse to analyze data in real

What are the top 3 drivers for using a CDW in your organization?

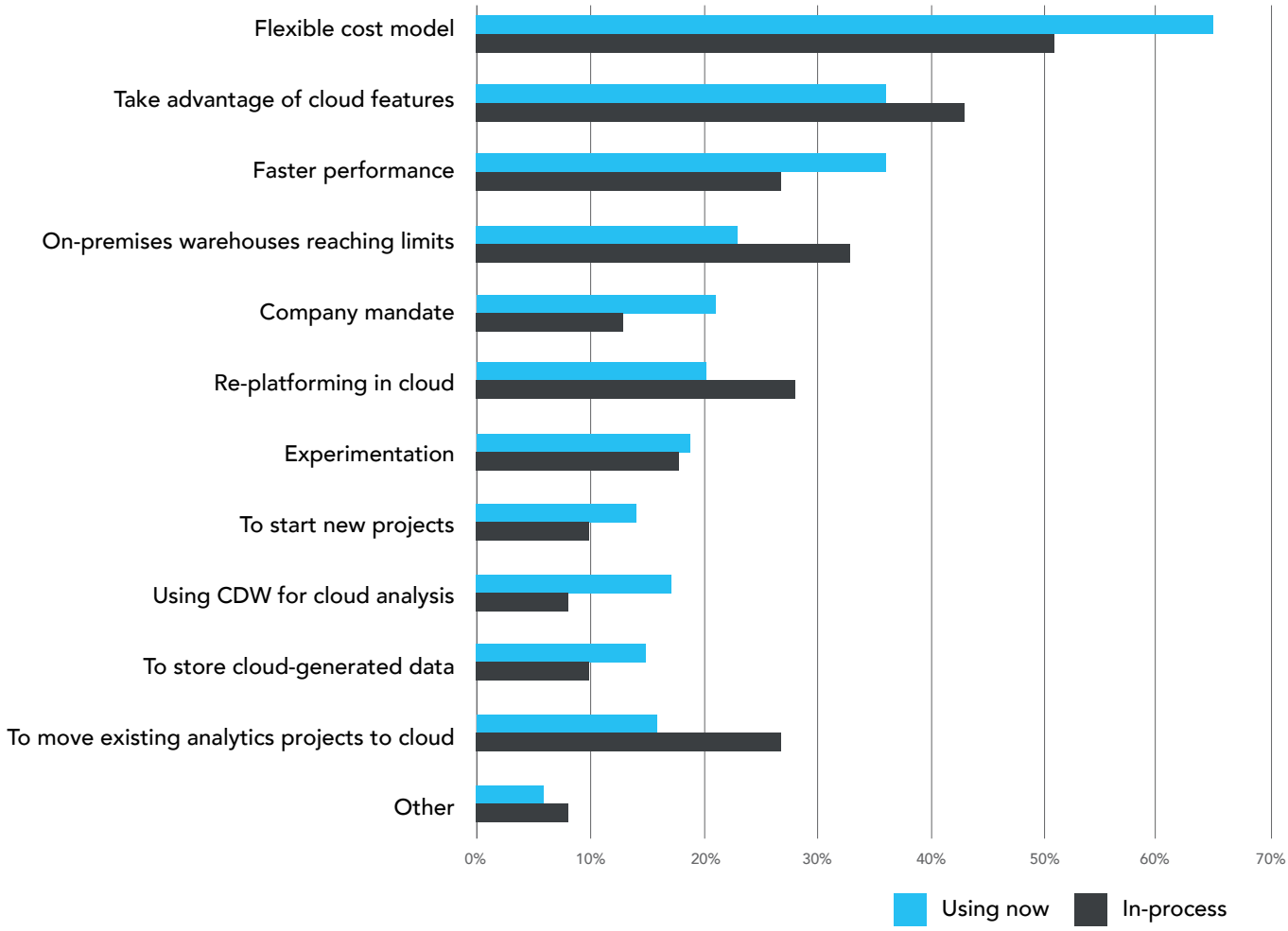

Figure 3: *Top drivers for using a cloud data warehouse.*

time and provide analytical results that can be immediately communicated to the right data consumers.

- **DISTRIBUTED WORKFORCE.** Because the cloud data warehouse is not physically situated within one of the organization's data centers, access can be easily expanded to a distributed team. Replication and change-data-capture methods allow mirroring of the data warehouse at different cloud regions with minimal data latency, expanding access to a broad geographic distribution.

RECOMMENDATIONS FOR IMPLEMENTING CLOUD DATA WAREHOUSES

Some challenges are not immediately addressed by a cloud data warehouse. In the same TDWI research survey, among those who had a cloud data warehouse in place, data governance, data integration, and loading data remain issues.

Several best practices can help you assess the potential benefits of adopting a cloud data warehouse and develop a transition plan in a way that can finesse the potential challenges of adoption:

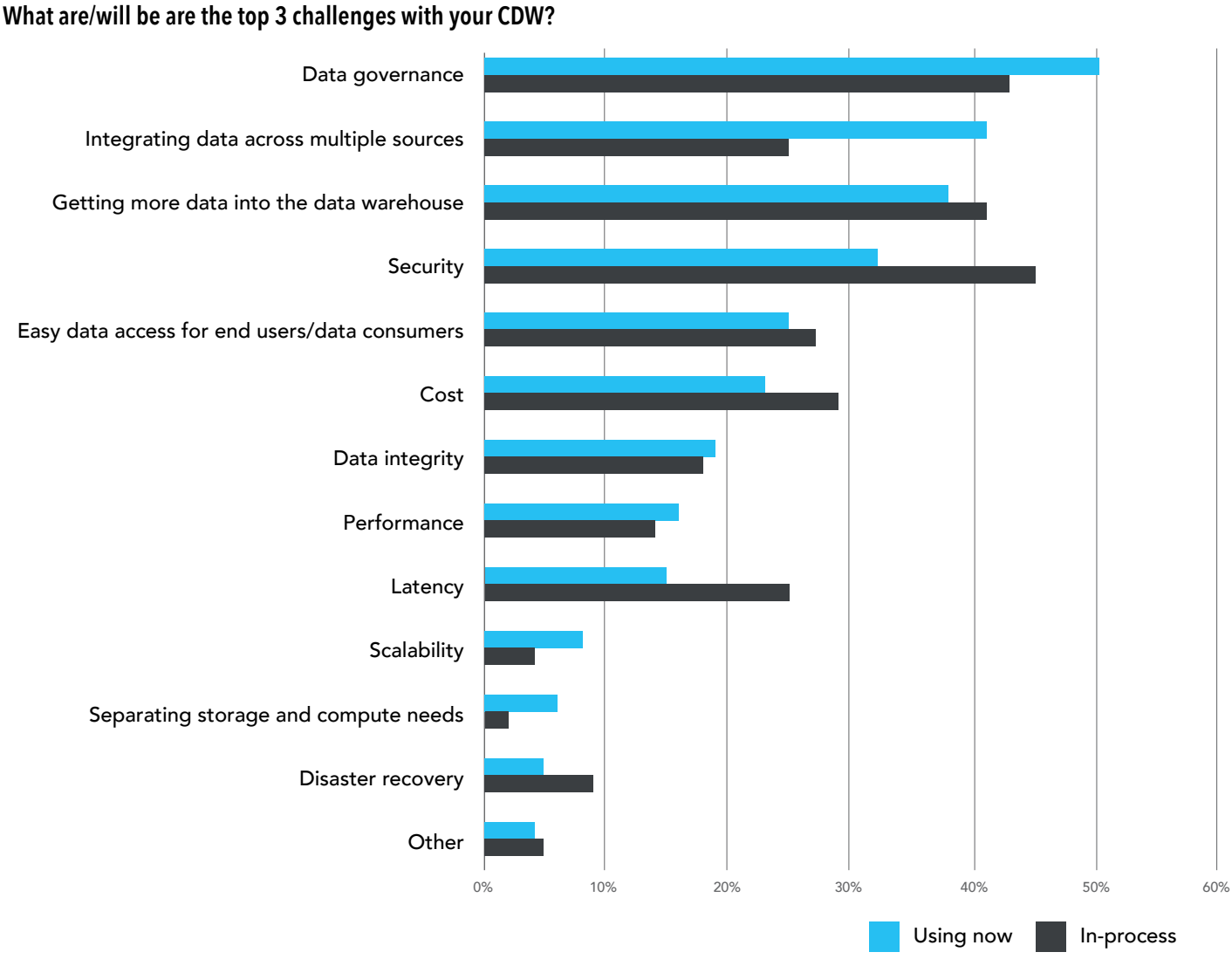

Figure 4: *Top challenges of a cloud data warehouse.*

- **COMPARATIVE COST ANALYSIS.** Consider your ongoing costs and establish a baseline for maintaining on-premises data warehouses. Develop a model that allocates those ongoing costs across the spectrum of on-premises reporting and analytics systems and develop a cost model that can be used for comparison to a cloud data warehouse.
- **PERFORMANCE NEEDS ASSESSMENT.** Assess your current requirements for both computational performance and for storage, especially when you expect to radically increase data volumes. Determine your need for scalability and consider your periodic computational demands in relation to complexity of analysis as well as the number of concurrent users at different times of the day. This will frame your need for automated scalability and elasticity because these demands will factor into the comparative cloud cost model. Assess the user base and its growth expectations and review your current data volumes and anticipated data volumes to assess upgrade costs. This will help you develop a comparative performance analysis model.

- **INTEROPERABILITY "SCAN AND PLAN."** Consider what systems need to remain on premises and what cloud-based services and SaaS applications already run in the cloud. Identify the systems that need integration and develop an interoperability plan. Determine whether you expect to be incorporating an increasing number of real-time data streams and map out a plan for streaming data integration.
- **ENGAGE CLOUD DATA WAREHOUSE PROVIDERS.** Identify which vendors have a data warehouse engineered for the cloud. Review each provider's offerings and make sure that the architecture addresses the aforementioned challenges. Use the comparative cost and performance analysis models to compare the benefits and economics of adopting a cloud data warehouse solution and to select a cloud data warehouse provider.
- **DEVELOP THE GOVERNANCE PLAN.** Consider how your data integration challenges can be addressed using the cloud data warehouse solution. Identify the data consumer roles and appropriate data access rights, and work with the provider to develop a security and data protection plan. Document governance policies and request guidance in enforcing those policies from the vendor.
- **MODERNIZE YOUR DATA WAREHOUSE.** Review how migration to the cloud enables capabilities not currently offered by your existing system. Develop and execute a data migration plan and reengineer your data acquisition, ingestion, and integration processes to adopt the cloud provider's tools. At that point you should be ready to deploy the cloud data warehouse.

ABOUT OUR SPONSOR

www.snowflake.com

Snowflake started with a clear vision: make modern data warehousing effective, affordable, and accessible to all data users. Snowflake enables the data-driven enterprise with instant elasticity, secure data sharing, and per-second pricing across multiple clouds. Because traditional on-premises and cloud solutions struggle at this, Snowflake developed a new product with a new built-for-the-cloud architecture that combines the power of data warehousing, the flexibility of big data platforms, and the elasticity of the cloud at a fraction of the cost of traditional solutions.

Snowflake: Your data, no limits. Find out more at snowflake.com.

ABOUT THE AUTHOR

DAVID LOSHIN, president of Knowledge Integrity, Inc, (www.knowledge-integrity.com), is a recognized thought leader, TDWI instructor, 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 frequent invited speaker at conferences, Web 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.

ABOUT TDWI RESEARCH

TDWI Research provides research and advice for BI professionals worldwide. TDWI Research focuses exclusively on analytics and data management issues and teams up with industry practitioners to deliver both broad and deep understanding of the business and technical issues surrounding the deployment of business intelligence and data management solutions.

TDWI Research offers reports, commentary, and inquiry services via a worldwide membership program and provides custom research, benchmarking, and strategic planning services to user and vendor organizations.