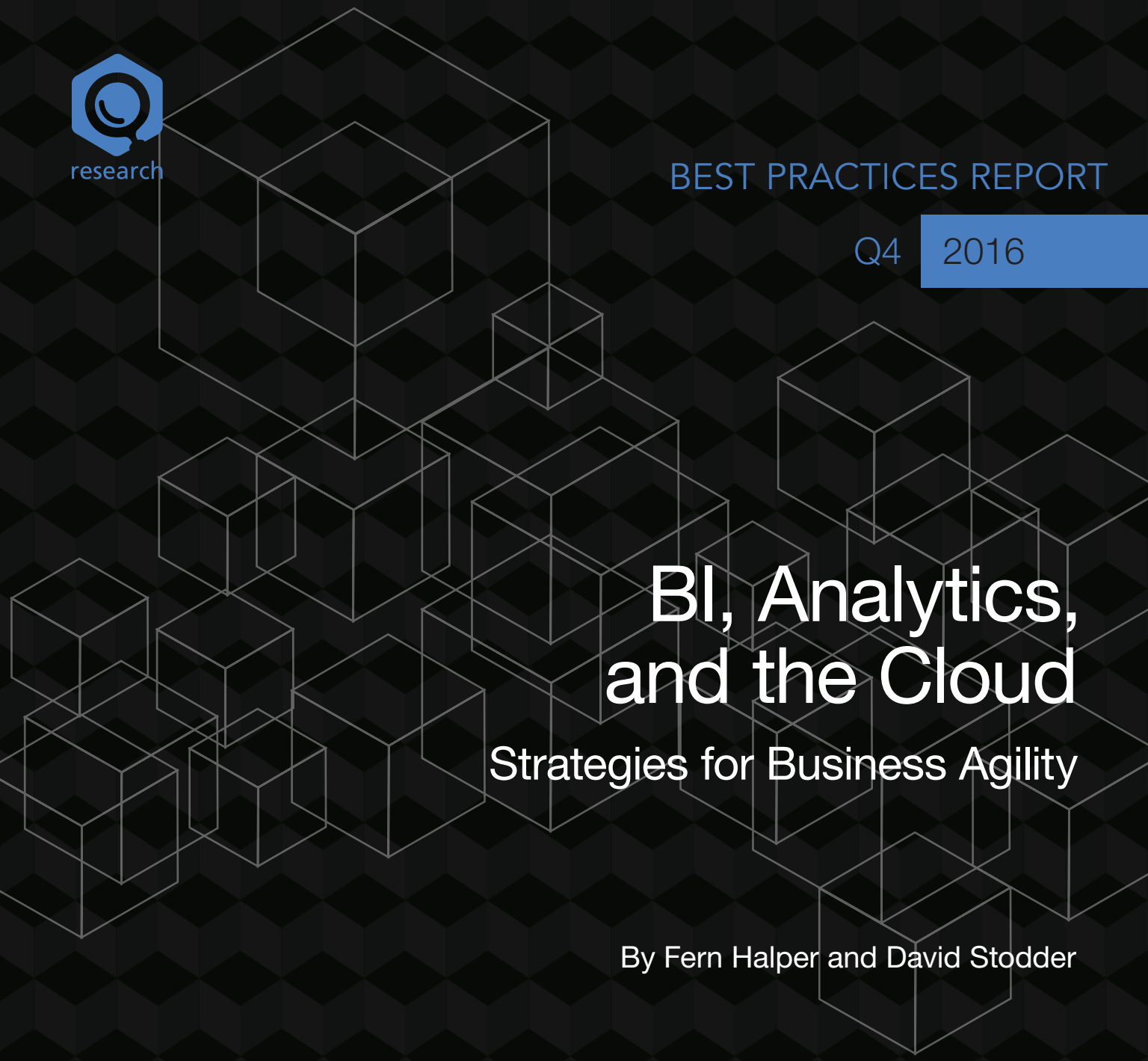


research

BEST PRACTICES REPORT

Q4 2016

BI, Analytics, and the Cloud

Strategies for Business Agility

By Fern Halper and David Stodder

tdwi
Advancing all things data.

Research Sponsors

Cloudera

IBM

Informatica

SAP

SAS

Snowflake Computing

Tableau Software

BI, Analytics, and the Cloud

Strategies for Business Agility

By Fern Halper and David Stodder

Table of Contents

About the Authors

FERN HALPER is vice president and senior director of TDWI Research for advanced analytics, focusing on predictive analytics, social media analysis, text analytics, cloud computing, and other "big data" analytics approaches. She has more than 20 years of experience in data and business analysis, and she has published numerous articles on data mining and information technology. Halper is coauthor of "Dummies" books on cloud computing, hybrid cloud, service-oriented architecture, service management, and *Big Data for Dummies*. She has been a partner at industry analyst firm Hurwitz & Associates and a lead analyst for Bell Labs. Her Ph.D. is from Texas A&M University. You can reach her at fhalper@tdwi.org, @fhalper on Twitter, and on LinkedIn at linkedin.com/pub/fern-halper/2/491/63.

DAVID STODDER is senior director of TDWI Research for business intelligence. He focuses on providing research-based insight and best practices for organizations implementing BI, analytics, performance management, data discovery, data visualization, and related technologies and methods. He is the author of TDWI Best Practices Reports and Checklist Reports on data discovery, data visualization, customer analytics in the age of social media, BI/DW agility, mobile BI, and information management. He has chaired TDWI conferences on BI agility and big data analytics. Stodder has provided thought leadership on BI, information management, and IT management for over two decades. He has served as vice president and research director with Ventana Research, and he was the founding chief editor of *Intelligent Enterprise,* where he served as editorial director for nine years. You can reach him at dstodder@tdwi.org, @dbstodder on Twitter, and on LinkedIn at linkedin.com/in/davidstodder.

About TDWI

TDWI, a division of 1105 Media, Inc., is the premier provider of in-depth, high-quality education and research in the business intelligence and data warehousing industry. TDWI is dedicated to educating business and information technology professionals about the best practices, strategies, techniques, and tools required to successfully design, build, maintain, and enhance business intelligence and data warehousing solutions. TDWI also fosters the advancement of business intelligence and data warehousing research and contributes to knowledge transfer and the professional development of its members. TDWI offers a worldwide membership program, six major educational conferences, topical educational seminars, role-based training, onsite courses, certification, solution provider partnerships, an awards program for best practices, live webinars, research publications, an in-depth research program, and a comprehensive website: tdwi.org.

About the TDWI Best Practices Reports Series

This series is designed to educate technical and business professionals about new business intelligence technologies, concepts, or approaches that address a significant problem or issue. Research for the reports is conducted via interviews with industry experts and leading-edge user companies, and is supplemented by surveys of business intelligence professionals.

To support the program, TDWI seeks vendors that collectively wish to evangelize a new approach to solving business intelligence problems or an emerging technology discipline. By banding together, sponsors can validate a new market niche and educate organizations about alternative solutions to critical business intelligence issues. To suggest a topic that meets these requirements, please contact TDWI senior research directors Fern Halper (fhalper@tdwi.org), Philip Russom (prussom@tdwi.org), and David Stodder (dstodder@tdwi.org).

Research Methodology and Demographics

Report Purpose. This report educates organizations in best practices and options for cloud business intelligence (BI) and analytics. This includes organizational strategies for the cloud as well as new platform options and other considerations. The report also examines how organizations are using cloud BI and analytics and gaining value from them.

Terminology. For the purposes of this report, *the cloud* includes several deployment models—public, private, and hybrid—as well as different delivery models, such as infrastructure-as-a-service (IaaS), platform-as-a-service (PaaS), and software-as-a-service (SaaS). We define these the first time each one is used in the report.

Survey Methodology. In June 2016, TDWI sent an invitation via email to the BI and data professionals in our database, asking them to complete an Internet-based survey. The invitation was also posted online and in publications from TDWI and other firms. The survey collected responses from 350 respondents. TDWI also added panel data from more than 50 additional respondents in order to obtain greater diversity of opinions. Survey branching limited the number of respondents for some questions. All responses are valuable and so are included in this report's data sample. This explains why the number of respondents varies per question.

Research Methods. In addition to the survey, TDWI conducted telephone interviews with technical users, business sponsors, and data management experts. TDWI also received briefings from vendors that offer products and services related to cloud BI and analytics.

Survey Demographics. The majority of survey respondents are business or IT executives (31%), followed by technical developers/architects (24%), data scientists (18%), and other business sponsors/users (15%).

The financial services and consulting industries (11% each) dominate the respondent population, followed by healthcare (9%), education (8%), software/Internet (8%), telecommunications (7%), and other industries. Most survey respondents reside in the U.S. (61%) or Europe (13%). Respondents come from enterprises of all sizes.

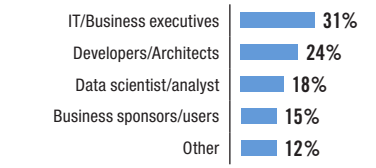
Acknowledgments

TDWI would like to thank many people who contributed to this report. First, we appreciate the many users who responded to our survey, especially those who agreed to our requests for phone interviews. Second, our report sponsors, who diligently reviewed outlines, survey questions, and report drafts. Finally, we would like to recognize Martin Pacino, TDWI industry research analyst, and TDWI's production team: James Powell, James Haley, Peter Considine, and Michael Boyda.

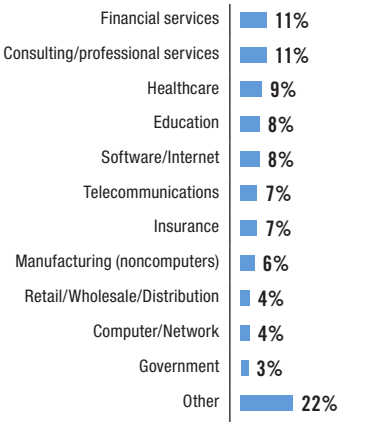
Sponsors

Cloudera, IBM, Informatica, SAP, SAS, Snowflake Computing, and Tableau Software sponsored the research and writing of this report.

Position

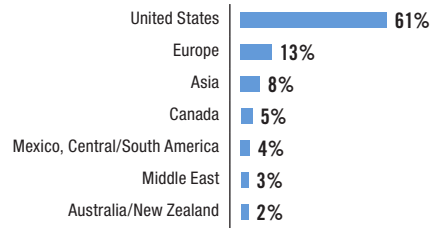


Industry

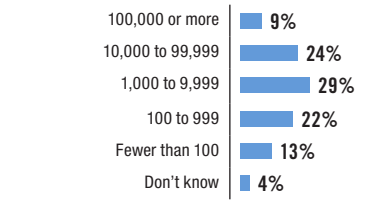


("Other" consists of multiple industries, each represented by less than 2% of respondents.)

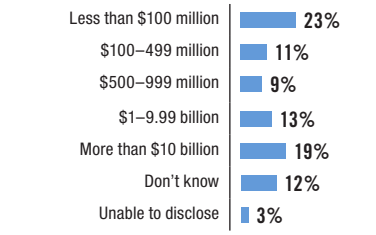
Geography



Number of Employees



Company Size by Revenue



Based on 255 respondents who completed every question in the survey.

Executive Summary

Organizations are making use of the cloud in numerous ways.

The cloud is becoming a mature platform for data management, integration, business intelligence (BI), and analytics. Although some organizations still will not make use of it (typically for security or compliance reasons), resistance is definitely diminishing. Companies understand that the cloud can provide flexibility and scalability for their BI and analytics projects. Of course, the cloud is not yet for every company, but organizations like the fact that it allows them to be more agile—they can quickly spin up (or down) instances of the cloud for BI and analytics and they like that.

We've found that organizations are making use of the cloud in numerous ways. Some are using it to expand the number of users who can perform visual analytics and discovery and to provide more sophisticated tools than spreadsheets. Others are using the cloud to help modernize their data warehouse for analytics. Still others are using it to experiment with analytics outside of the data warehouse, which often includes experimenting with open source technologies. They are also using the cloud for part of their more advanced analytics implementations, such as for big data analytics.

As expected, security still tops the list of barriers cited by respondents in this study by a margin of almost 2-to-1. Some organizations realize that security concerns are often really more about perceived security issues or, more likely, about losing control over data. This is often a political or cultural issue. Privacy also ranks high on the list of challenges associated with the cloud, along with security, isolation of a company's assets from those of others, identity management, and integration between on-premises data with their public or private cloud. These are all big considerations for organizations looking to move BI and analytics to the cloud. Integration of data residing on premises and off premises is an important aspect of cloud analytics because the majority of respondents in this study think that they will ultimately move to a hybrid architecture, where public and private cloud platforms are used along with on-premises systems. To do this effectively, they will need to deal with integrating data in an evolving ecosystem that includes the cloud and on-premises data warehouses or other data management environments.

Organizations are reasonably satisfied with the cloud across many areas.

Overall, it appears that respondents are more satisfied than not with cloud BI and analytics across many categories of variables. This includes their ability to load data from on-premises systems to cloud-based systems and perform data transformations there. However, there is significant room for improvement. One concern is the time it takes to perform these procedures. Additionally, although organizations are also reasonably satisfied with data integration, they face mounting challenges as data becomes more complex and users demand faster data refreshes. Cloud governance is also an issue.

This TDWI Best Practices Report examines organizations' experiences with and plans for cloud BI and analytics, including how satisfied organizations are with the cloud and why. It also looks at various challenges and how organizations are overcoming them, as well as some new cloud models. Finally, it looks at what organizations should consider when moving to the cloud.

An Introduction to BI, Analytics, and the Cloud

Defining BI and Analytics in the Cloud

Resistance to the cloud is definitely diminishing.

There is no doubt about it—the cloud (and often the public cloud) is being hyped as the go-to platform for BI and analytics because of its elasticity and scalability and the speed and ease with which organizations can spin up cloud-based systems and services compared to the time required to install and configure on-premises systems. As recently as three years ago, when TDWI asked respondents whether they were using or planning to use the cloud for analytics, about 25% said that they would never use the cloud. Resistance is definitely diminishing, though, as organizations start to understand the benefits of the cloud for a range of business applications, including BI and analytics. For example, in the survey for this Best Practices Report, only about 10% of respondents said they would never use the cloud for BI and analytics.

Of course, the cloud is not for every organization and the cloud means different things to different groups. There are many varied deployment and delivery models for the cloud. In this report, we talk about three different deployment models—*public, private,* and *hybrid.*

Public cloud: A multitenant resource available to the public or organizations on a pay-as-you-go, fee-per-usage basis, or as a free service. When people think about cloud BI and analytics, they are often thinking about software-as-a-service (SaaS) or BI-as-a-service, which is one possible delivery model. However, organizations might also utilize their own chosen BI, analytics, or data management software in a public cloud environment, utilizing an infrastructure-as-a-service (IaaS) delivery model where the cloud provider only offers hardware, networking, and storage services.

Private cloud: A set of computing resources typically located within the company's existing security and regulatory compliance infrastructure and processes. A private cloud serves only a single company but operates in a cloudlike manner in terms of hardware, storage, and network management. It may be owned and managed by the company it serves or by a third party, which may host it on premises at the client's location or in their own cloud environment.

Some vendors are providing platform-as-a-service (PaaS) delivery solutions for BI and analytics that can be hosted in the vendor's public cloud, in a private cloud, or on site in a customer's private cloud. PaaS provides a set of comprehensive services that enables organizations to handle the BI and analytics life cycle including data collection, storage, and analysis.

Hybrid cloud: A computing environment that includes the use of public and private clouds, often with one or more touchpoints between them. A hybrid cloud may also include some sort of integration process to enable the cloud and on-premises environments to work together. For instance, an organization might combine SaaS with a private cloud for BI, where SaaS provides customer relationship management (CRM) application data that the organization brings down from the public cloud and combines with on-premises transaction data; the combined data sets are then analyzed in a private cloud. Another example is an organization using multiple public cloud services. It might upload data to the cloud from its on-premises locations and combine it with data found in other cloud services utilizing a SaaS or PaaS analytics solution.

On-premises data center: This refers to a company's traditional data center environment. In this case, BI and analytics software is deployed in the data center or on desktops.

In addition to IaaS, PaaS, and SaaS, there are other "as a service" delivery options. These include older managed service provider (MSP) models that can run on cloud platforms. A current trend is toward analytics- or data-as-a-service models where vendors provide complete, managed solutions often aimed at specific vertical industry analytics or data needs. These are typically available in the public cloud and are accessed via application programming interfaces (APIs).

There are even cloud marketplaces that sell specific analytics functions as a service. For instance, a service provider or a software vendor might sell options such as churn-reduction-as-a-service, fraud-detection-as-a-service, or analytics-as-a-service tailored to financial planning applications. Vendors develop these similarly to packaged applications and they can be reused. Pieces of code may also be sold in this way for use in various analytics.

Finally, some vendors provide big-data-as-a-service solutions that enable companies to analyze semistructured or unstructured data that they cannot easily analyze using their existing on-premises systems.

Nearly 10% of respondents say they would never use the cloud for analytics.

In our survey, just over half of the respondents are using some sort of cloud deployment model for analytics—12% report using the public cloud for analytics, 19% are using a private cloud, and 19% are using a hybrid cloud. Thirty-one percent of respondents are planning to use the cloud in the next few years. Slightly less than 9% said they would never use the cloud. The rest (10%) did not know.

Why Are Organizations Using the Cloud for BI and Analytics?

Clearly, organizations are working past their concerns with the cloud, and as a result, cloud-based analytics is gaining ground. The cloud provides scalable storage, computing power, and elastic resources. As a large-scale distributed computing environment, it provides an important and flexible foundation for big data—it can scale as data grows, be implemented rapidly, and be spun up or down quickly.

Top cloud analytics drivers include scalability, flexibility, and cost.

We asked survey respondents what they felt were the top benefits for moving to the cloud for analytics (Figure 1).

Scalability, flexibility, and cost top the list. Not surprisingly, respondents noted scalability (51%) and flexibility (41%) as two top drivers for moving to the cloud. When asked about the cloud, survey respondents often comment that they like the ability to scale on demand and the flexibility and simplicity of the cloud. Companies can deploy servers as needed and access analytics software quickly—which is especially appealing to those accustomed to long waits when ordering hardware and software through their organization.

Cost is another driver for (37%) of respondents. Although there can be hidden costs associated with the cloud, some organizations feel that the faster time to implementation and avoidance of risk are overriding factors when it comes to cloud analytics.

Speed is also a factor. As part of being flexible, the speed of cloud implementations can provide the agility by getting an application up and running faster than using a traditional data center. According to one respondent, "The cloud provides a sandbox for rapid development." Respondents also cited faster response to business requests (22%) and quick access to feature functionality (20%). Some like that IT is less involved (17%), which may contribute to the view that this frees up IT resources for more value-added activities (12%). (As TDWI has seen in our qualitative research, when IT and the business collaborate on analytics, it is usually a more successful deployment.)

Users are paying attention to data, too. Respondents also cited data-related drivers for using cloud analytics. A considerable amount of data is now being generated in the cloud, which includes social media data, Internet of Things (IoT) data, and data from other cloud applications, to name just a few. Since it often makes sense to analyze data where it lives—what some in the industry refer to as "data gravity"—this may explain increasing acceptance of cloud BI and analytics. Seventeen percent of respondents cited big data as a driver for cloud analytics and 16% cited expanded access to data as a benefit.

What are your organization's top three drivers for cloud analytics? Please choose no more than three responses.

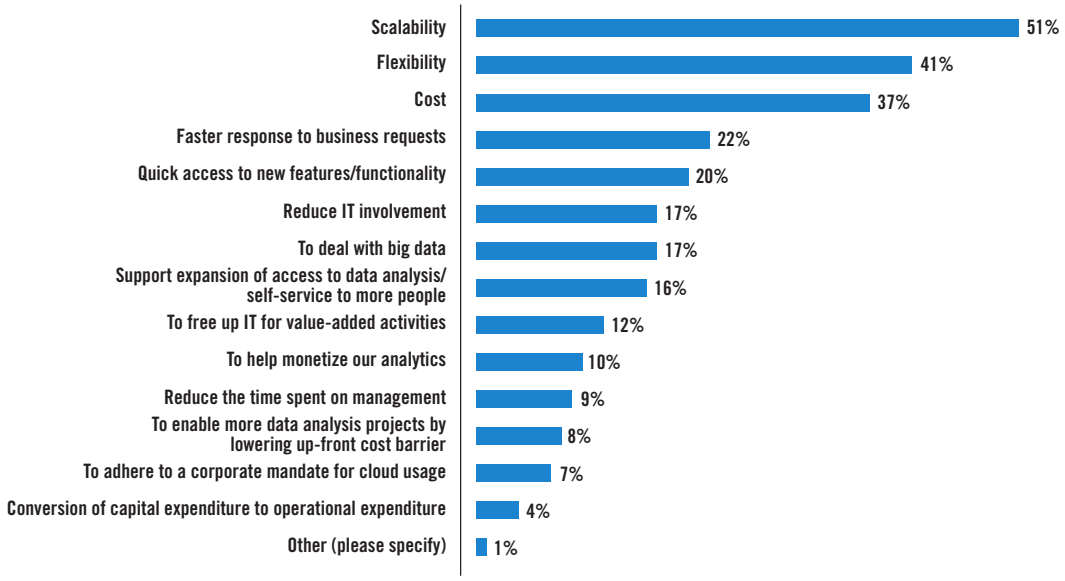


Figure 1. *Based on 240 responses.*

USER STORY SUGAR CANE PRODUCER GROWS WITH COST-EFFECTIVE AND CUSTOMIZABLE CLOUD ANALYTICS

TDWI recently spoke with Florida Crystals, a large sugar cane refiner that was an early enterprise cloud adopter. "We were early adopters of cloud technologies. I haven't had a data center since 2010, which is incredibly liberating," said Don Whittington, VP and CIO at Florida Crystals, adding that cloud adoption let them "really focus on the business … and that's key to our philosophy."

Florida Crystals has been growing and increasing its market share with cloud analytics playing an integral role, particularly to assess potential mergers and acquisitions and speed the transition after they are completed. "We've made acquisitions where we were able to demonstrate that we had superior information to the businesses we were acquiring," Whittington said. "Then we leveraged that further by gathering information about businesses we acquired rapidly and using the analytical capabilities that we had developed." Whittington noted that this streamlining occurs throughout the M&A transactional chain, allowing them to "analyze on the front end, get the piece in the middle, get the integration going, and then optimize the business going forward."

The ability to customize BI and analytics, which has been a concern for many companies considering cloud implementation, has come a long way, according to Whittington. "What we have found is that the cloud actually makes customization easier." Florida Crystals is using SAP BusinessObjects Cloud; like many SaaS and cloud-based BI and analytics providers, SAP continuously updates its cloud-based infrastructure, BI, and analytics products rather than waiting to do it all in one big release, which has been the custom for on-premises applications and systems. "We are getting software feature and function improvements or bug fixes on a continual basis. You don't have to wait for the next release."

To allocate the costs of cloud BI and analytics adoption, Florida Crystals employs an incremental system to manage costs. "We meter it and report back to business units. We allocate it; we don't bill it back. We can go back [to the business units] that are asking us, 'Why was this cost higher?' We can say, 'Well, here's the meter. If you want to reduce [cost], just use less of it. If it's adding value, you can provide it incrementally on a metered basis.'"

This incremental approach extends beyond just cost management and is a key deployment strategy that Whittington recommends for all companies. "Your cloud strategy can be incremental. You don't have to bet the business. You can move from that older push model to a pull model. Do it incrementally and look at the business value that came from that, and pull that along."

Use Cases and Examples

As described above, the use cases for the cloud are wide and varied. We asked experts and vendors how they see their customers using the cloud. Some of the common scenarios include:

Self-service visual analytics discovery. Most analytics vendors offer their software as a service, either in a public cloud or running on IaaS in public cloud environments. The price point has also come down enough that small and midsize businesses can make use of the same software as larger organizations. It is easy to get up and running in this way to use the cloud for a range of business analytics such as analyzing sales, understanding customer behavior, and creating customer segments.

A word of warning, though: some organizations get started on cloud analytics from the business side without up-front IT involvement, which can lead to the very security and governance issues our surveys show users are worried about (see the following section, "Those Who Don't Use the Cloud"). Therefore, TDWI recommends involving IT sooner rather than later so that BI and analytics can be properly governed and can take advantage of IT's broader enterprise perspective.

Integrated analytics and applications. As noted, eliminating delays and complexity in developing on-premises BI and analytics can help companies bring BI and analytics into their core business applications and processes more quickly. Adding easy-to-use, Web-based visual interfaces to simplify data analysis and interaction—or to develop and use predictive or other advanced statistical models—further increases the effectiveness. For example, adding cloud-based analytics to business and financial planning and forecasting can enable examination of complex cash flows with data drawn from a variety of sources rather than just one application—often a limitation of on-premises applications. With the availability of additional data sources, users can gain different perspectives on profitability and cost management by linking nonfinancial and financial models and applying cloud-based analytics to diverse data.

Marketing analytics in the cloud. One critical area of focus for many companies is marketing analytics. As more companies compete through data-driven marketing, sharpening efforts to attract and retain customers requires them to scale rapidly to handle large data volumes coming from multiple channels and sources. These sources include customer behavior data drawn from log files, clickstreams, Web tracking, and affiliate marketing and advertising sources. Cloud analytics often allows them to do this without the cost and delays usually associated with installing on-premises applications.

Companies also need the flexibility to do brief periods of intensive data analysis around a particular campaign and then bring the unneeded systems down to save costs. Cloud computing can support such flexibility requirements.

Modernization of data warehousing and data preparation. Data warehouse modernization is frequently referred to as *augmentation, automation, optimization,* or *modernization,* and it usually includes tools and platforms built to accommodate big data from new sources.[1] Often, organizations augment—but do not replace—their existing data warehouse when they add additional platforms and tools such as Apache Hadoop (with or without integrated MPP databases), appliances, newer kinds of cloud data warehouses, or other kinds of data management platforms. Some of these additions may be cloud based so that the cloud becomes part of the modern data and analytics ecosystem—what might be considered a hybrid cloud implementation. This, then, also includes the integration tools and platforms necessary to integrate various cloud and on-premises environments.

Data preparation practices and technologies are also undergoing modernization, enabling organizations to deploy smarter software (with embedded machine learning and other analytics) that makes collection, data quality assurance and remediation, profiling, transformation, cataloging, and other data procedures faster, easier, and capable of handling more diverse data. Some solutions provide tighter integration in the cloud between preparation and analytics so that users have less need to move between disconnected technologies to accomplish objectives.

[1] For more information on data warehouse modernization, see *TDWI Best Practices Report: Data Warehouse Modernization in the Age of Big Data Analytics* (2016), online at www.tdwi.org/bpreports.

Big data experimentation. Cloud-based implementations also provide organizations with ways to explore new methods for managing and analyzing the huge volumes of data they are now generating. For instance, an organization may want to try an Apache Hadoop distribution or Apache Spark, but because they have not used these technologies before, they don't want to invest in the necessary infrastructure just to find out if it might be valuable. Hadoop-in-the-cloud is a good way to understand whether Hadoop is a good fit in general for the organization before investing heavily in an on-premises implementation. The cloud also enables experimentation with different kinds of data management approaches, including data warehouses, data lakes, and open source technologies.

Analytics sandboxes. Some organizations use a public or private cloud as a sandbox to experiment with different algorithms or to build a proof of concept (POC). This can be done using any kind of data in user-controlled spaces (often for power users or business analysts) that are separate from production environments. For instance, a company might use an analytics sandbox to experiment with predictive analytics for understanding customer behavior and predicting certain kinds of actions.

Social media and sentiment analysis. Social media data can be helpful for understanding what people are saying about brands or specific products and for performing competitive analysis. Tweets, news reports, microblogs, and other online sources can provide insight into the voice of the customer. As social media data is generated in the public cloud, it can make sense to analyze there too rather than extract and load it to an on-premises system.

IoT and data streaming. IoT generates a lot of data, some of which is streaming. The cloud (public or private) is often a part of an IoT implementation. For example, a fleet operator might use sensors to collect data from their various trucks. Such data might include the temperature or number of vibrations per second of a particular part or parts. This data can then be collected by a gateway device and periodically sent to a data center—on premises or in the cloud—where data from all the trucks in the fleet is analyzed to determine what precipitates a part failure or when undue wear and tear is occurring. That information might then be encoded into a set of rules or a model and then pushed down to the gateway or other edge device and used to improve fleet maintenance and operational efficiency.

Those Who Don't Use the Cloud

Although resistance to the cloud is diminishing, there are still organizations that neither want to use the cloud nor feel they should use it for analytics. We asked those respondents who do not use the cloud or do not know if they use the cloud why they didn't.

Compliance, privacy, and security are reasons why organizations don't use the cloud.

A number of the respondents who said that they would not use the cloud for analytics often do make use of cloud applications for other purposes. These include (not shown) HR applications (18%), CRM (18%), financial applications (9%), specific vertical applications (5%), and supply chain (4%). Thirty-nine percent (not shown) of the respondents said that they used no cloud applications, although sometimes organizations actually do use cloud analytics when they think they do not. For instance, an organization might use the data from a cloud CRM system as part of its sales analysis but not use cloud-based tools for the sales analysis itself. In such a case, one might think that the cloud is separate from their on-premises data center when, in fact, data from the cloud is being sent on premises to be analyzed.

We asked, "If you are not using a public, private, or hybrid cloud for data warehousing or business analytics, why not?" The complete list of responses is shown in Figure 2.

Compliance, privacy, and security concerns. Unsurprisingly, many organizations that do not use the cloud do so because it does not meet their compliance requirements. This is supported by the fact that organizations in financial services and healthcare were more likely to cite this as a reason than in other industries. Similarly, responders from Europe, where privacy regulations are much more stringent, appeared more likely to cite these issues than those in the U.S.

The current data center suits their needs. For some organizations, the current environment is sufficient. Although these organizations may eventually grow their cloud analytics strategy—perhaps as they collect more data or as their concerns are addressed—or become more sophisticated analytically, they do not currently need a cloud solution.

If you are not using a public, private, or hybrid cloud for data warehousing or business analytics, why not? Please select your top three reasons.

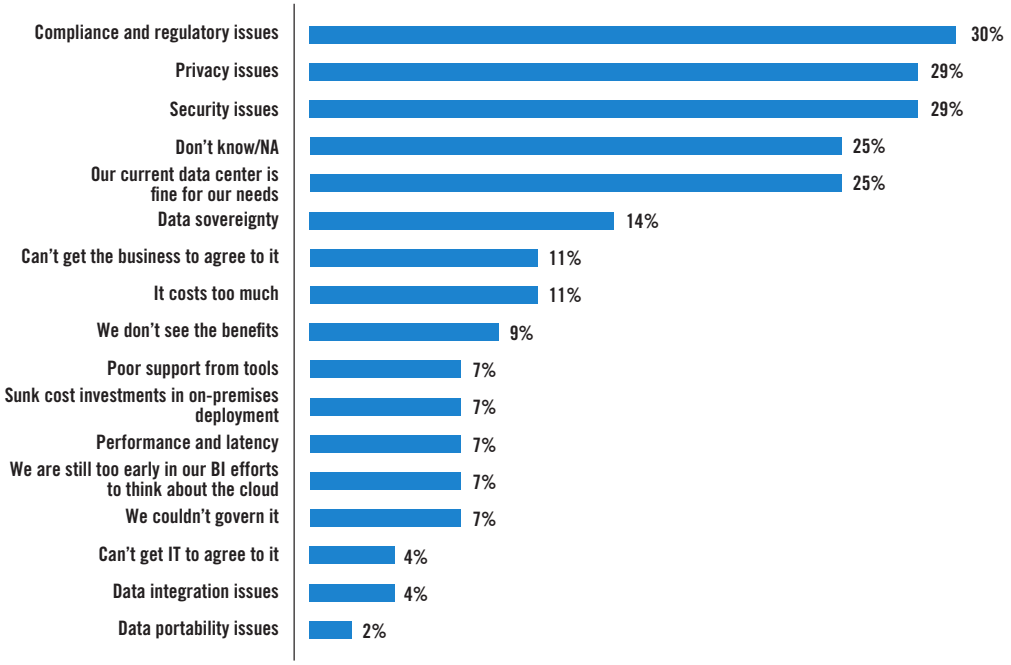


Figure 2. *Based on 56 responses.*

The State of BI, Analytics, and the Cloud

Cloud Analytics Adoption

As we have discussed, there are numerous deployment and delivery models for the cloud and multiple combinations of the two, as well as numerous kinds of analytics that can be performed in the cloud. We asked respondents about the specific kinds of BI and analytics they currently run on premises and in the cloud. The options included reporting; dashboards; and visual, predictive, and big data analytics.

The data suggests that (at least for now) standardized reporting, standardized dashboards, and planning and forecasting are more likely to occur on premises than in the cloud by almost a 60-40 split (not shown). These most likely happen in the company's traditional data warehouse, which provides a good platform for repetitive, known kinds of analysis. The initial and often still-dominant use cases served by a data warehouse are those that demand data consistency, such as financial reporting or the provisioning of data for online analytical processing (OLAP) data cubes to allow drilling down and slicing and dicing of structured data.

On the other hand, our survey response data indicates that use of newer technologies for visual analytics is more likely to occur in a public, private, or hybrid cloud (not shown). Likewise, big data analytics (such as streaming analytics or analytics against IoT data) or analytics with newer data types (such as social media analytics) are also more likely to occur in the cloud (not shown). This follows the concept of data gravity described earlier, where data created in the cloud is analyzed in the cloud.

Over 40% of respondents believe that their analytics architecture will utilize a hybrid approach in the next three to five years.

However, the cloud is not necessarily an either-or proposition. In fact, when asked about where they see their analytics architecture in the next three to five years, many of the respondents (43%) believe that their analytics will utilize a hybrid ecosystem approach, where the cloud forms one component of a bigger architectural strategy (Figure 3). These modern, hybrid data ecosystems incorporate big data, analytics, and Hadoop with traditional enterprise data, the BI/DW technology stack, cloud, and numerous IT data platforms (both old and new) to support analytics. These are discussed in more detail later in the report.

Where do you see your analytics architecture in the next three to five years?

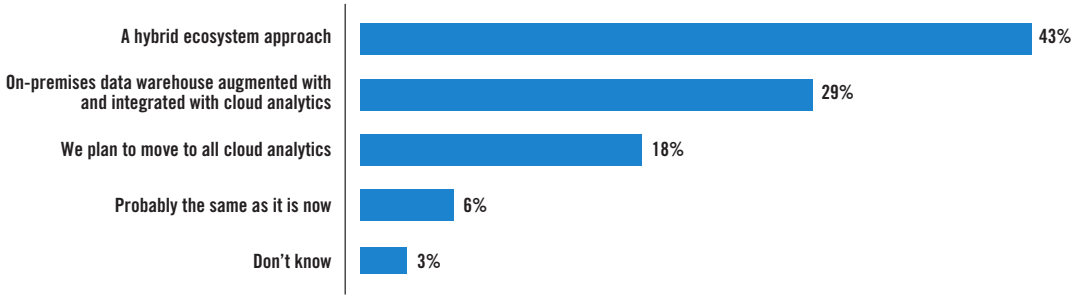


Figure 3. *Based on 217 responses.*

Plans for Cloud Analytics

As mentioned above, there are numerous use cases for cloud analytics. These run the gamut from visualization and big data analytics to data integration and more. We asked respondents about possible use cases for their organization today or in the near future (see Figure 4).

Self-service analytics is a popular use case for the cloud.

Self-service BI with visual analytics is the most common use case. Self-service BI and data discovery activities are at the top of the list when it comes to cloud analytics (58%). Self-service data discovery is about enabling users to do more on their own to analyze their data and answer their own business questions. Unlike reports that update standard views of data, users can explore without being limited to a preset list of metrics or questions. These tools typically have easy-to-use visual interfaces that enable flexible data exploration without having to write any code (even SQL). Many analytics vendors are making their visual discovery software accessible via public, private, or hybrid cloud.

These analytics might also be available in a SaaS delivery model. Here the organization deploys its analytics software in a public or private cloud environment and the vendor provides compute, storage, and networking services for that software to run on. Whether public or private, the cloud can be a quick, scalable, flexible way for organizations to get started in visual data discovery. However, native integration with cloud storage architectures and popular BI and visualization tools are critical for enabling true self-service.

Organizations are looking to do analytics experimentation in the cloud.

Users are also looking to experiment with their data. In addition to enabling visual analytics, respondents also cited using the cloud for analytics experimentation (50%) and for more advanced analytics in general (42%).[2] These use cases benefit most from the cloud's quick scalability and relatively easy access to different kinds of analytics methods and tools, such as text analytics for unstructured social media data.

For instance, if an insurance company is collecting telematics data from cars that is generated and transported over the Internet, it might choose to analyze the data in the cloud rather than move it on premises and have to deal with latency issues or data transport costs. Using readily available cloud-based tools, it might then build a model that uses this data to predict high-risk drivers. As another example, a retail store experimenting with a recommendation engine might want to use the cloud to experiment with a next-best-offer engine—especially if its Web and e-commerce servers are hosted externally. Lastly, a utility company that is collecting data from multiple sources that are cloud hosted (such as weather data and grid sensor data) to build a situational intelligence solution might consider experimenting and doing this in the cloud.

In all these cases, the cloud environment is eliminated or reduced once the experiment is completed and a solution found.

Collect data where it is created. Tied to the point above, respondents felt that the cloud is a good possibility for data ingestion (30%), data lakes (32%), and streaming data collection (32%). Organizations are thinking about newer kinds of data and about ingesting multiple data sources. The data lake—a low-cost, schema-on-read storage area for disparate data types including semistructured, unstructured, and real-time raw data—makes sense to put into the cloud because of the cloud's scalability and elasticity. Although many of these implementations are on premises today, that may change. Experimentation against a data lake or stream processing engines that work with real-time, cloud-generated machine data (such as that found in IoT use cases or in e-commerce) can also make sense to put into the cloud to take advantage of the flexibility in terms of how users can approach these analytics use cases.

[2] Here, "advanced analytics" includes algorithms and techniques for predictive analytics, optimization, and so on.

Which of the following are possible cloud analytics use cases for your organization today or in the near future? Select as many as apply.

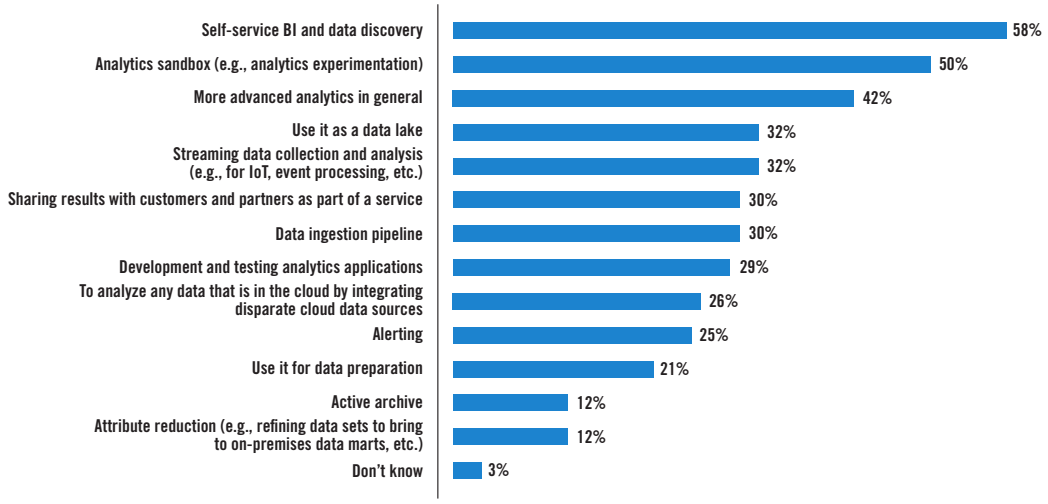


Figure 4. *Based on 231 responses.*

A Word About Analytics Application Development

As seen in Figure 4, development and testing of analytics applications is also a possible use case for the cloud (29%). Organizations are looking to develop analytics applications more often and want to do it quickly and iteratively. Some organizations use a platform-as-a-service (PaaS) to do this. PaaS provides the complete stack of services that the organization expects it will need to support development, such as middleware, operating systems, and security. As one respondent put it, "PaaS allows you to iterate as fast as you can." Once development is finished, organizations may then move into production with the new analytics application in the cloud or on premises.

Developers also make use of reusable Web services to speed development projects. Vendors are starting to provide what some refer to as *analytics services* or *microservices*—pieces of pretested analytics functionality developers can embed in their analytics applications via APIs. For example, in order to examine their sales performance across territories, an organization may want to see how sales quotas and sales performance relate to revenues per customer. Rather than set up a separate analytics application that would then need to be integrated, they can integrate the analysis within their existing sales or financial planning applications. Using the cloud, they can develop (or use prebuilt) analytics as services or microservices that connect to their applications through APIs, enabling them to embed analytics more easily into planning processes.

Developers are also interested in open source models for cloud development, such as OpenStack and Cloud Foundry. OpenStack is a set of software tools for building and managing cloud computing platforms for public and private clouds and is supported by a number of sponsors of this report.[3] Cloud Foundry is an open source PaaS environment; it is also supported by some of this report's sponsors.[4]

[3] For more information on OpenStack, see www.openstack.org.

[4] For more information on Cloud Foundry, see www.cloudfoundry.org.

USER STORY CLOUD SERVICES SOLUTIONS: A GREAT FIT FOR AN EMERGING MARKETING SERVICES COMPANY

89 Degrees is fast becoming a player in the customer engagement sphere by offering marketers a "mix of technology, data, analytics, and marketing strategy," David Simcik, director of consulting and campaign services, explained. The company offers "the 89 Degrees Cloud for Marketing, which at its core combines a hosted SAS server environment along with a proprietary marketing database that we've been refining for the past 20 years. We've taken our experiences working with B2C and B2B data across verticals such as retail, publishing, automotive, travel, pharma, etc., and have come up with a flexible approach to data—a blueprint, if you will—that works great with the SAS Customer Intelligence suite and SAS Visual Analytics applications." That includes understanding and predicting customer behavior as well as developing targeted campaigns.

According to Simcik, many companies lack the skilled personnel to cover all the necessary bases in cloud analytics. "Oftentimes, we find organizations with a single marketer or data scientist who is wearing a lot of different hats. 89 Degrees can come in to guide and implement more advanced forms of marketing and analytics that a client organization wouldn't necessarily have a lot of time to focus on." The cloud, along with the 89 Degrees service model, Simcik said, is leveling the competitive playing field for companies who may not have as many marketing or analytics resources at their disposal compared with some of their peers. "At the end of the day, especially for medium-size or smaller companies, going with the hosted solution and services is a no-brainer because of the efficiencies of working in the 89 Degrees hosted environment.... It allows an organization that doesn't have the necessary resources or skill sets to compete against larger players."

Barriers to Adoption

Leading barriers to cloud analytics adoption still include security and privacy.

Despite the many potential benefits of the cloud for analytics—the ability to move quickly, the fact that patches and maintenance are handled by providers, and access to analytics features and functionality—there are also barriers to cloud adoption (see Figure 5), which we discuss in more detail here.

Security still tops the list of cloud challenges. As mentioned previously, security is still the most common barrier cited by respondents (59%). Whether that concern is real or perceived, it is not necessarily a bad thing. Cloud analytics providers (at least the large ones) have made it their business to focus on security and obtain security certifications to address it. In many cases, their security is now stronger than that of their customers. It is naturally important to ensure the platform and tools meet security and governance requirements across the business. However, as one respondent put it, "It's entirely possible that our security concerns are out of date. The cloud risks can be properly mitigated. There may be some exceptions, some cases where certain data is kept in house. However, it shouldn't be a blanket 'nothing allowed.'"

That said, it is still up to the organization to do its homework on whether its cloud provider meets its security requirements. It is also important for the cloud provider to ensure that its customers have visibility into its environment and that they are notified about any security incidents.

Ultimately, the organization is responsible for its data. Because the cloud distributes computing, it increases the number of connection points, and that adds a layer of vulnerability. However, organizations should already be making the effort to have the right set of controls in place, whether in the cloud or on premises.[5]

[5] For more information on cloud security, consult resources such as the Cloud Security Alliance (www.cloudsecurityalliance.org) or ISACA (www.isaca.org).

Privacy is also a big concern. Privacy goes hand in hand with security concerns. In the survey, 39% of respondents cited this as a top concern. We have spoken with many organizations that are uncomfortable storing or processing personally identifiable information (PII) outside of their on-premises data centers or on-premises private clouds due to concerns about PII being accessed or used without the organization's awareness or permission. There are geographic considerations, too. For instance, new EU data protection rulings will replace the current data protection directive, with a general regulation designed to give citizens more control over their own private data and to standardize the law across the EU. Many vendors are expanding their data centers (used to host as-a-service tools) into new regions in order to deal with these laws.

Organizational issues rank high. Lack of skilled personnel (28%) and trust/culture (25%) are also cited by respondents as challenges with cloud computing. The culture and trust issues may well be part of those same security and privacy issues. Some organizations still feel— rightly or wrongly—that they will lose control over their data if they put it in a cloud, especially a public cloud. As a best practice, organizations should consult with vendors and service providers to understand how much control they will have over their cloud-based data and its management. Increasingly, cloud vendors and service providers are giving organizations more control and transparent views into how their data is stored and managed.

Integration is critical to make sure that data from multiple sources can be analyzed.

Data integration is important. Data in a cloud ecosystem typically comes from many sources. Therefore, integration is critical to ensure that it can be analyzed, especially if some of the data is in the public cloud and some of it is on premises or in a private cloud. Along with concerns about the metadata management necessary to support integration, respondents spoke about performance issues related to integrating and then analyzing data that is stored in different locations. Organizations should consider whether they will need to query and analyze data from on-premises sources without having to first load it into a cloud data store. If so, they will need to evaluate whether their chosen solutions require uploading the data into a cloud data warehouse before users can query it. In either case, technologies to enable organizations to push queries to both cloud-based and on-premises sources and bring only the query results back into the cloud are maturing, which will improve performance and reduce costs by not having to move and load so much data.

Until then, what are users doing to overcome these challenges? Some businesses fortunately do not face these challenges because they were built more recently with a cloud mandate from the start. For those trying to bring the cloud into their company, though, we've heard that education is critical. Many organizations are trying to educate executives about the benefits and risks surrounding the cloud. Some start small and show the stakeholders what they can deliver. As one respondent explained, "Prove you can walk before you can run and start with non-PHI/PII data."[6] Others admitted that they are still trying to get their organization to adopt the cloud.

[6] PHI, or protected health information, is a type of PII generated in healthcare scenarios by a variety of patient activities, including payments. As such, organizations must protect PHI and use it in accordance with strict regulatory privacy rules.

What are/were the biggest barriers to adoption of cloud in your organization? Please select up to three responses.

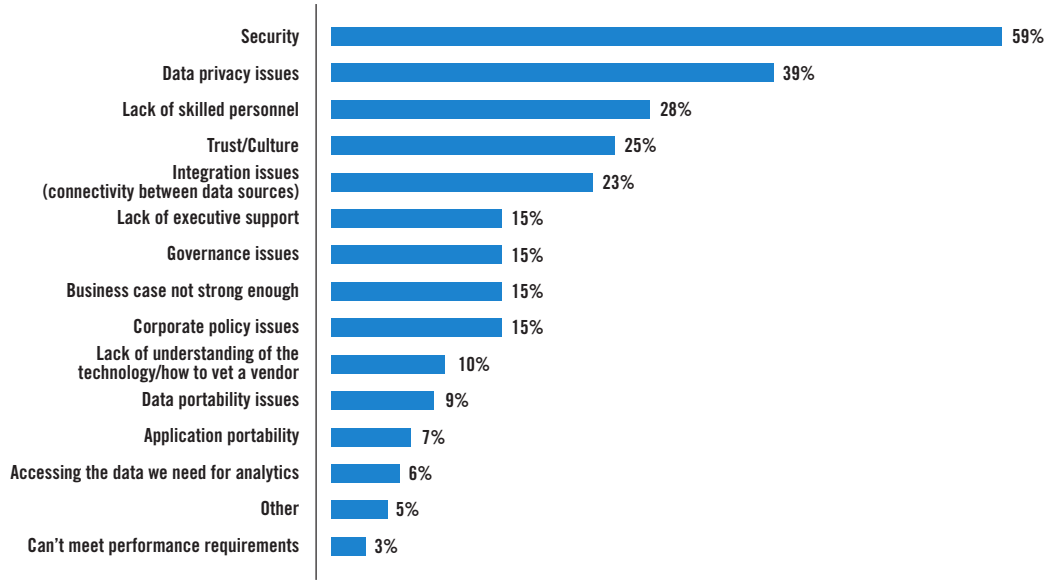


Figure 5. *Based on 239 responses.*

Organizational and Technical Considerations for Adopting Cloud BI and Analytics

As we saw in the discussion of Figure 1, organizations are pursuing a range of objectives with cloud computing for BI and analytics, with the top drivers being scalability, flexibility, and lower costs. Many firms see the cloud as a place to start from scratch with experimental systems to support a new business venture or model. In other words, they do not know the exact system requirements up front and therefore need the very qualities of flexibility and low cost that the cloud provides.

However, even if they do not know their precise requirements, they still need to follow best practices across organizational, technical, and other dimensions. Critical issues in these areas do not magically go away when organizations adopt the cloud for BI and analytics. This section will discuss some of the key considerations and look at research findings associated with them.

Considerations for Choosing a Cloud, PaaS, or SaaS Provider

As in all other areas of cloud BI and analytics, security is the top consideration for choosing a public cloud provider.

It's still about security. As with all other questions about cloud BI and analytics, security is also the most important factor in choosing a public cloud provider, including related concerns of isolation of assets from other customers' assets and identity management (Figure 6).

Organizations are (or should be) using the same level of scrutiny with PaaS, IaaS, and SaaS provider choices as they do with on-premises solutions. They have traditionally been nervous about moving their data off premises where it may be exposed to unknown security risks, especially data subject to stringent regulations and governmental policies (such as PHI and financial data). Thus, ahead of all other considerations, organizations are counting on cloud services providers to protect their data as well as related applications, infrastructure systems, and virtualization software dedicated to supporting their BI and analytics.

Along with security, organizations need to inspect identity management (IM) procedures to ensure that role-based access policies for cloud BI and analytics services are clear and kept up to date. As APIs and connectors for integrating and accessing data in the cloud multiply, IM can get complicated, often because organizations will grant users access rights haphazardly and not document them.

However, most cloud, PaaS, IaaS, and SaaS providers are aware of concerns about security, demand for isolation of assets, and identity management. Cloud providers' security management policies will typically cover deterrence, prevention, detection, and correction—often in a stronger fashion than what organizations are doing for their on-premises systems because it is core to their business. That said, organizations should examine providers' security and related management procedures so that they are satisfied that they can safely put sensitive data on the platforms.

Disaster recovery and service-level agreements are often overlooked considerations. While disaster recovery and service-level agreements did not rank among the top factors in selection of public cloud service providers by research participants, it is a best practice to devote attention to them. On-premises technologies and procedures for protecting data and dealing with disruptions due to systems or database failures generally do not carry over into the cloud realm. Cloud providers have improved regarding availability, fault tolerance, and disaster recovery in recent years; these are essential qualities in their competitiveness.

However, significant outages still occur, and some cloud providers may be more vulnerable as they spread their growing services across globally distributed data centers. Organizations should look at cloud-based disaster recovery options to ensure that providers offer continuity and high availability, automatic failover and replication, and can perform related steps for recovery. Service-level agreements should be specific and offer legal assurances against damage to business reputations.

Flexibility is key. Throughout our research, respondents identified flexibility as important for cloud-based BI and analytics; that is, the ability to set up systems to meet immediate business needs and take them down when resources are no longer needed. Close to 40% stated this was a very important consideration when choosing a cloud provider—understandably so, as this is one of the main promised benefits that makes the cloud distinct from on-premises systems, which can become an ongoing financial burden if they are not being continuously used at an optimal level. Organizations should make sure that cloud pricing models are flexible so that they can scale up and down easily, depending on BI and analytics needs.

Organizations should make sure that pricing models are flexible so that they can scale up and down easily.

Flexibility is also critical as more users adopt self-service visual analytics and discovery, which will have personalized data requirements not satisfied by "one size fits all" reporting and data provisioning. These users will demand that the array of functionality in their BI and analytics tools be supported by data management and provisioning flexibility as well.

Integration between cloud and on-premises is also on users' minds. Respondents are aware of the importance of integration between cloud and on-premises sources, with 78% stating that this was either important or very important. Integrating data from different on-premises sources alone can be challenging, so the last thing organizations need is additional complexity from users needing data from business applications located on a cloud platform.

Organizations can use a variety of techniques to build unified views of the data depending on user requirements. These include expanding cloud-based data warehouses or other repositories to incorporate this data, implementing data federation or virtualization software, using data preparation tools that can "blend" data from numerous sources in a data mart or in-memory cache for specific users, and using commercial or custom-coded point connectors to access specific data sources. Some BI, analytics, and data preparation solutions provide capabilities for tracking data lineage, which enables users to understand where their data came from, whether it is fully trusted data, and what transformation procedures were performed on the data they are analyzing.

People are thinking about portability, but not much more than anything else. We often hear anecdotally that companies are concerned about the potential for lock-in with a single cloud provider. To address this concern, companies will set up multiple cloud provider relationships and seek assurances that they can move data and analytics from one to another as necessary. In response to this customer concern, some BI and analytics solution providers have designed their platforms to be portable so that if a customer wishes to change providers or move data and analytics from one to another, they can do so easily.

However, the number of survey participants that said portability of vendor licenses—which, of course, is just one aspect of portability—was an important factor landed fairly in the middle of the range, indicating it's important but not more so than other factors. Licensing can vary among cloud-based application solution providers, not to mention cloud platform providers themselves; licensing could be per user, per enterprise (so that the number of users is unlimited), per number of concurrent users, by the level of processor activity, or by some other activity-related pricing. Organizations should examine how pricing is determined and what factors can lead to a rise in costs so that the bill does not surprise them.

How important are the following factors in your selection of a public cloud provider for your data warehousing or analytics?

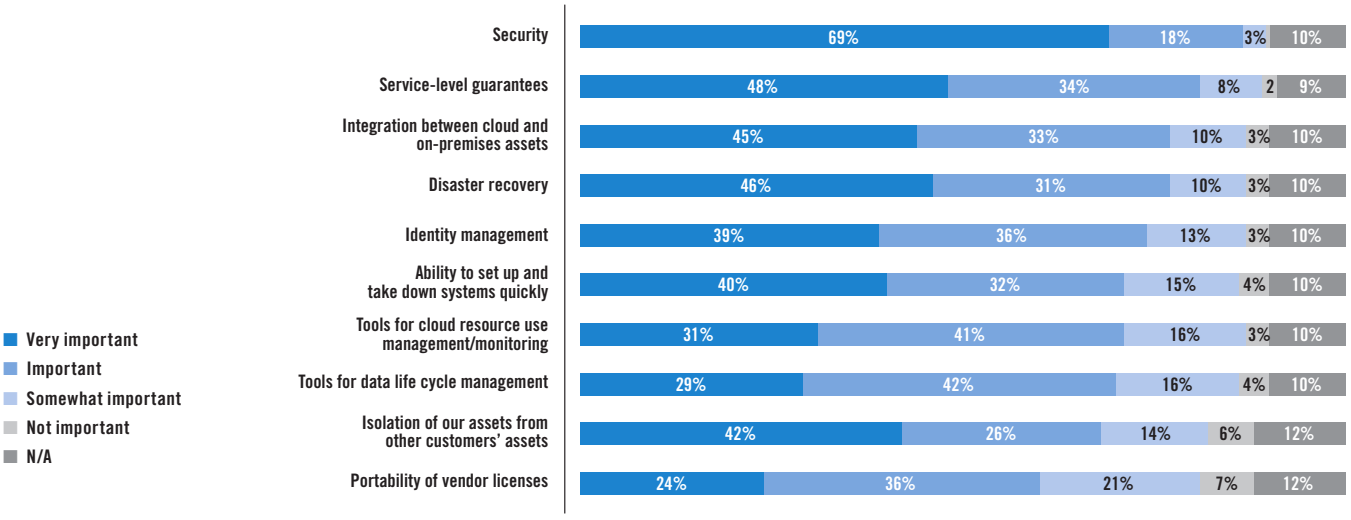


Figure 6: *Based on 231 responses. Ordered by highest total of "very important" and "important" combined.*

USER STORY BAZAARVOICE RIDES THE DATA TSUNAMI WITH CLOUD SCALABILITY

For many companies, data does not flow evenly into analytics platforms and the instant scalability that the cloud offers is critical to success. BazaarVoice—a company that offers consumers instant, authenticated product reviews, as well as brands a forum for engaging customers on a deep and comprehensive level—is one such case. With 700 million unique visitors per month, it relies on the volume afforded by cloud hosting, and with the fluctuations in seasonal shopping patterns, the scalability of the cloud is paramount.

"It's staggering how much data flows into our system on Black Friday, which is our biggest day of the year," John Steinmetz, senior product manager for analytics, told TDWI in a recent conversation. "We can see a 10x multiplier on the amount of data that flows into our system." In addition to this data, BazaarVoice ingests a stream of real-time data that it uses for diagnostics to monitor the health of its systems and for developing new features and functionality for its platforms. BazaarVoice uses Cloudera CDH, which includes Cloudera's Hadoop platform distribution and related technologies and systems, and Amazon Web Services.

When asked about best practices for cloud analytics, one that Steinmetz cited was redundancy in scaling. "The biggest suggestion I have is to fully manage your redundant systems, because when you're using a cloud-based service, there's always a failure somewhere in the chain. You need to make sure you're managing your auto-scaling groups and constantly revising those processes to make sure they are set up for your current state."

Like many companies whose business and analytics models are beholden to an irregular flow of traffic, the scalability of BazaarVoice's cloud platform is the primary reason for its deployment. "I would say [flexibility] is secondary. The primary reason would be scalability. We had so many clients coming on board and so much data [that] we had to have a system that would scale accordingly. We didn't want to be rebuilding something every six to eight months."

> With the its data platforms and analytics in the cloud, BazaarVoice ensures that they will stay on the proper footing to not only handle their current data load but to also alleviate the pains that customers could feel from growth. "This year we are expecting around a 15%–20% increase in [events per second], so scalability and best practices are essential to us staying profitable and give us the ability to service our clients at the same rate."

Cloud Computing and Data Governance

Data governance practices are important in the cloud.

One danger of the easy availability of cloud computing is that it could exacerbate data chaos by increasing the number of data silos. Organizations are already having difficulty getting their arms around their "Wild West" of disparate data silos for governance and general data management purposes, much less to provide single views of the information spread across them. Some organizations we interviewed prohibit—or at least strongly frown upon—use of public cloud solutions out of concern that it will reduce their control of the data. Of course, the danger of an outright ban is that users may go around the restrictions and set up cloud-based "shadow IT" systems on their own. It is often a better practice for IT to work with users to come up with a mutually acceptable solution, which may include use of governed data on cloud platforms.

Data governance is an organizational effort involving the creation of plans, policies, rules, and accountability for overseeing the data, both at rest and as it moves and is used throughout the organization. As such, data governance practices are important to ensure that data assets are protected and that quality and consistency are valued and maintained. Data security is often a primary governance concern, especially regarding the protection of sensitive data and adherence to data privacy regulations.

Organizations are still relatively immature when it comes to governing cloud analytics deployments.

Figure 2 showed the significance of compliance and regulatory concerns in preventing organizations from using the cloud. We asked research participants how they are organized to provide governance for cloud analytics deployments (Figure 7). That about a third (34%) answered that "they are currently trying to figure that out" and 8% said "we don't govern our cloud deployments" shows that many organizations are still immature in governing cloud analytics deployments.

A little over a third (37%) said they are using their data and analytics governance team to provide governance for cloud analytics deployments. This frequently means that organizations are only allowing previously governed data to be used in the cloud—that is, data that has been approved for use by those responsible for governance (usually IT). Analytics developed in the cloud is therefore derived from governed data. Supplying only governed data to cloud resources and ensuring governance continuity within and outside the cloud is a best practice that most organizations should follow. This way, data use in the cloud is covered by the same security and governance policies and procedures as data in on-premises systems.

In practice, however, many organizations have gaps in their data governance. For example, data security may differ from one data store to another and may even be altogether inadequate in less carefully managed sources such as data lakes. Organizations may neglect to properly govern data throughout its life cycle and can lose track of how it is being used, transformed, shared, or stored. As organizations increase their use of cloud computing resources, they should consider the impact on current governance policies and procedures and whether these may need to be updated.

How are you organized to provide governance for cloud analytics deployments?

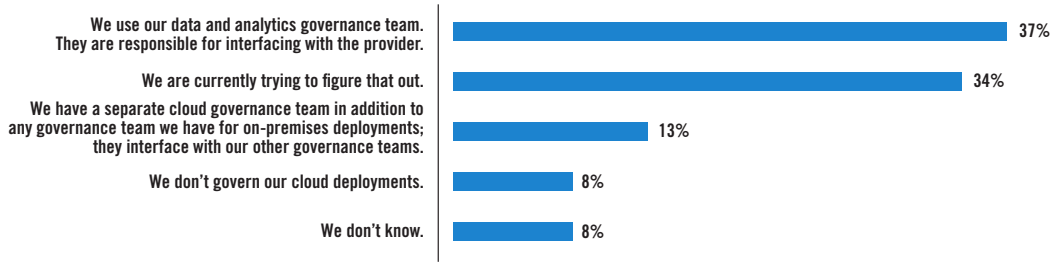


Figure 7. *Based on 212 responses.*

Data Management and the Cloud

As organizations move their on-premises data to public or private cloud platforms or generate significant new data there with cloud-based business applications, they need to adjust how they think about data management. Of course, cloud often becomes relevant in organizations precisely *because* of strains on data management, so they will have already begun to rethink it. A common issue is that organizations have so much invested in existing on-premises data management platforms that they cannot afford either the time or the cost to expand those platforms to handle new applications and new business user requirements.

Rather than go through a long and expensive journey to add new instances of their existing database management systems (DBMSs) or ETL systems, they choose to deploy Hadoop ecosystem technologies in the cloud. This is viewed as a more cost-effective way to spin up analytics sandboxes for users to experiment with data, perform custom transformations, and develop new systems designed to meet immediate—and possibly only momentary—business needs. Some will intend to move data management back to on-premises systems once these cloud-based BI and analytics applications are developed, tested, and ready for production.

In other cases, organizations are looking at how they can reduce data management and storage costs overall and see cloud platforms as a way to offload some of these tasks to less expensive platforms. Today, new organizations, particularly small and midsize firms, often launch with their data management in the cloud rather than on premises to reduce costs, increase flexibility, and bring data management closer to their most prevalent data sources, which are also in the cloud.

Spreadsheets are a popular data source for cloud analytics.

Spreadsheets are the most common data sources for cloud analytics. Not unexpectedly, the largest percentage of research participants (60%) indicated that they need to connect to spreadsheet applications for their cloud analytics (Figure 8). Spreadsheets are ubiquitous; TDWI research finds that these are the most commonly needed sources for BI and analytics. However, the ease of setting up and accessing cloud-based BI and analytics could give users an alternative to the habit of pulling data down into a spreadsheet to perform analysis, thereby further democratizing use of BI and analytics in organizations. About 58% of respondents also indicated that on-premises databases or warehouses formed the main source of data to be integrated.

Social media and SaaS applications are other prime sources.

Social media data is important. A common source for cloud analytics is social media data (48%). The explosion of interest in analyzing social media behavior challenged many organizations because they did not know how or where to manage and store the data. Social media functions in organizations, which may or may not be part of marketing organizations, often have little budget for on-premises systems, particularly of the size and complexity needed for managing social media data. The cloud quickly became a less expensive alternative for storing mass amounts ("data lakes") of unstructured social media data. Social networks themselves, of course, set the tone by using the cloud for storing and backing up their own and users' social media and analyzing it there. This is another factor driving organizations that want to access the networks' social data. Because the networks themselves use cloud-based applications and databases, it is easier to access and load data, as well as do analytics, in the cloud.

SaaS data also needs to be integrated. Another common data source among participants is data from other software-as-a-service (SaaS) sources (52%), which would include services such as Salesforce .com for CRM and Netsuite for business management. For many organizations, it has been a challenge to access data from SaaS applications and integrate it with other internal and external data sources. Metadata and data definitions do not line up easily and data quality can vary, requiring master data management, data preparation, and data integration software to resolve.

Frequently, an organization's developers will write custom point-to-point integration routines that can then become difficult-to-maintain "spaghetti software." Organizations should evaluate automated data integration and connectivity software that can standardize integration and begin to replace spaghetti code.

Research participants indicated that other prominent sources are on-premises application data (43%), geospatial and third-party demographic data (32%), other text data (36%), and mobile data (36%).

What kind of data sources do you or would you need to connect to for your cloud analytics efforts? Please select as many as apply.

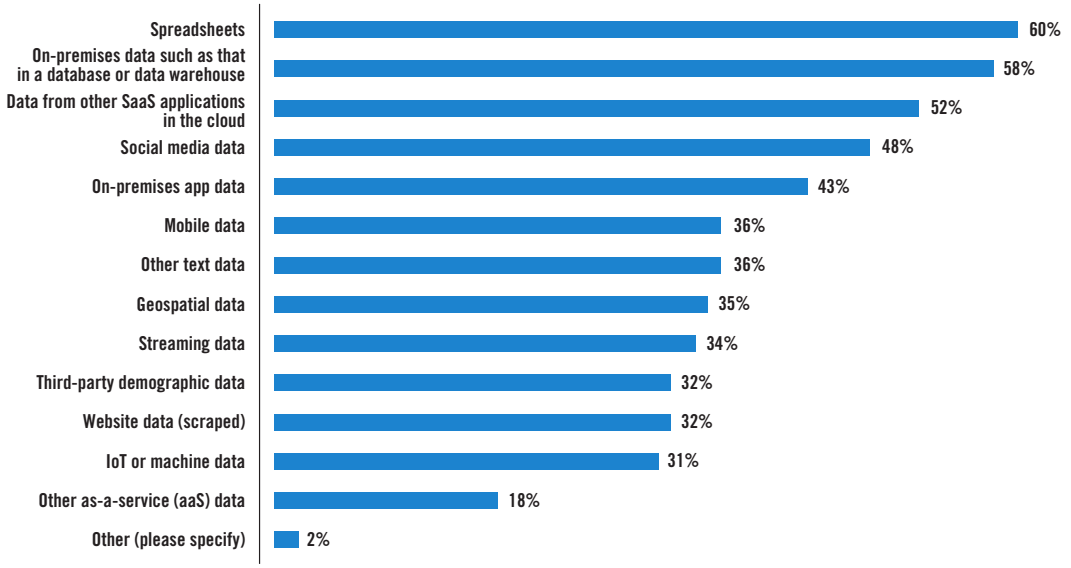


Figure 8. *Based on 222 responses.*

USER STORY EXPEDIA VALUES THE VERSATILITY AND SPEED OF CLOUD-BASED ANALYTICS

Analysts project that the online travel industry will top $830 billion in revenue in 2017.* With roughly 40% of the market share, Expedia conducts a staggering number of online transactions, each a critical touchpoint in consumer engagement in an industry laden with competition and customer migration. In order to monitor the success rate of these transactions, Expedia's Global Consumer Payments group turned to a private cloud analytics framework hosted by Amazon Web Services (AWS) with Tableau Online for visualization and reporting.

"We can move quickly and don't have to go through a procurement process," Expedia senior business analyst Tad Buhman said. "We can spin up new instances and drop them when we're done with them, compared to having to procure all of that hardware and get it all configured and installed. All of that would take months. Now we can do it in minutes." Buhman said Expedia lets Tableau manage hardware and software as part of the managed service it provides. Expedia personnel have access to dashboards for visibility into the performance of Tableau and the AWS platform.

Expedia has historically housed its large volume of transactional data on premises but is increasingly moving to a cloud-based platform. "Everything we look at, we bring up to the cloud in AWS and combine it using Hadoop and MapReduce," Buhman reports. "There are very few instances where we'll need to join some on-premises data, such as spreadsheet, with data in our cloud data mart. Those are one-off scenarios, although one of the reasons we went with Tableau Online is because it was easier to allow users to continue to work with spreadsheets for their analysis. We can administer security on a user level so that they have rights to certain reports and projects and create projects on the fly. This made it easier for users' spreadsheet-based analysis to move into the cloud. They still have the ability to do one-off analyses and reports, and if they have something they want to save, they can promote it up to a production report."

Like most organizations, Expedia's Global Consumer Payments group had some growing pains. "A lot of people knew the technology we had been using well and, of course, had never used the cloud-based technology. We definitely had some challenges with performance when we first started out, but I think it was because we didn't yet understand the technology," Buhman said. When asked how this transition can be done smoothly, he recommends a strong foundation of connectivity and security. "Make sure your connectivity and security are set up before you start worrying about getting data set up in the cloud. You don't really know how if it's going to run until you get the data loaded, and then it can be a scramble to get the performance you need."

* Marketwired.com, available at http://www.marketwired.com/press-release/global-online-travel-and-tourism-sales-to-reach-us830-billion-in-2017-1885494.htm.

Data Management Infrastructure Needs

Supporting the data needs of business users working with BI and analytics puts stress on particular areas of data management. One is delivering on the promise of elasticity, which is one of the most hyped benefits of cloud computing. Elasticity enables organizations to spin up and bring down systems as needed, particularly to support seasonal marketing campaigns and the need to analyze sudden spikes or shortfalls in business performance, fleeting market opportunities, and more. Cloud computing allows organizations to respond to heightened demand without having to build on-premises systems that are configured (expensively) for peak requirements.

Organizations should learn the characteristics of their chosen cloud DBMS's parallel implementation.

An important underlying technology for achieving elasticity is parallel processing. This enables data management systems to structure tasks to execute at the same time. Parallel systems can scale out or shrink as needed by adding or removing nodes. This way organizations only pay for what they need. Parallel databases should execute most analytics queries faster than other types of databases. Organizations should learn the characteristics of their chosen cloud DBMS's parallel implementation, including how nodes share resources and how tasks are synchronized. In addition, organizations should determine if it would be difficult to run applications that may not have been developed originally to take advantage of parallelism on the systems.

Because organizations usually will be paying for use of bandwidth for queries as well as storage, a second important consideration is that cloud databases must be efficient in their use of resources while providing good performance for complex analytics queries. Cloud data management systems should have adequate performance and optimization tools to overcome unevenness and inconsistency in the performance of individual nodes. Parallel databases that can scale compute power independently from storage resources can better handle performance inconsistencies and high-concurrency demands. Cloud data management systems also need to be resilient and available to handle the required number of concurrent users.

Cloud-based data management infrastructure is becoming more popular.

Organizations use a variety of types of data management infrastructure. We asked research participants what kind of data management infrastructure they utilize for analytics in the cloud currently and plan to use in the future (Figure 9). Selections included both on-premises and cloud-based systems. Regarding what they are using currently, the largest percentage (68%) are using a database on premises, with 37% using a database in the cloud. Similarly, 60% are using a data warehouse on premises, more than double the percentage using a data warehouse designed for the cloud (26%).

These results are not unexpected given the greater maturity of on-premises databases and data warehouses. However, 43% plan to use a database in the cloud within three years and 45% plan to use a data warehouse designed for the cloud within that period, which indicates that many organizations have a strategy to increase their cloud-based installations.

Currently, just under a quarter (23%) of survey participants are using Hadoop in the cloud and 29% are using Hadoop on premises. TDWI Research has found in previous studies that although Hadoop use is less than that of traditional systems for BI and analytics, its presence is growing. Here, we can see that the percentage of current Hadoop use in the cloud is not far behind on-premises use, and 36% plan to use Hadoop in the cloud within three years compared to just 23% who plan to use Hadoop on premises within three years. It appears that the strongest growth in Hadoop for BI and analytics will occur in the cloud.

Organizations need to integrate technologies together into a hybrid architecture.

Of course, most organizations that have existing on-premises databases, data warehouse, analytics platforms, and Hadoop systems will still have them even as they increase adoption of cloud-based technologies. This means that the development of hybrid systems that employ technologies that can integrate on-premises and cloud systems into a single, comprehensive architecture will be important. BI and analytics projects will need seamless access to data in both types of systems; users will want to blend data from both for their analysis and data visualizations. A comprehensive, hybrid architecture will also help organizations choose the right platform to fit each project's business requirements, scalability and performance expectations, data security policies, and other considerations.

What kind of data management infrastructure do you utilize for analytics in the cloud?

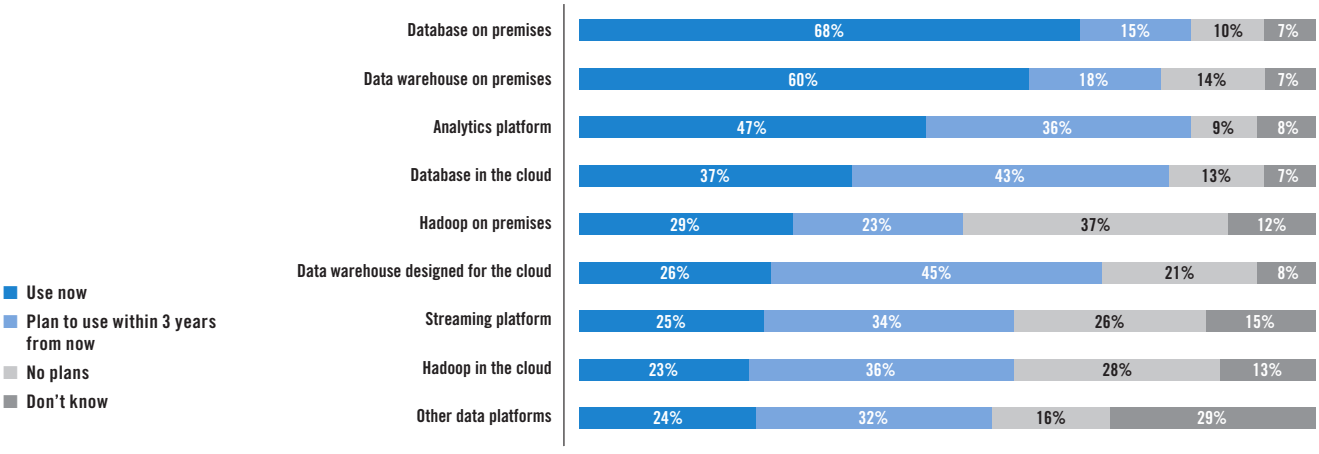


Figure 9. *Based on 217 responses. Ordered by highest "use now."*

USER STORY CLOUD DATA WAREHOUSING ENABLES ASK.COM TO MOVE BEYOND BEING REACTIVE

One of the common pitfalls of on-premises BI and analytics in a big data age is the amount of time and resources that must be devoted to physical-level database administration tasks, such as replacing disks, managing clusters, and so on. In many cases, BI teams are struggling to perform the strategic analysis that is needed by the business; they are forced to focus on less-business-value-adding database performance management and data delivery. Ask.com, owned by InterActiveCorp (IAC), was in this frustrating situation before its move to a cloud-based topology. We spoke to Keith Lavery, until recently senior director of BI, data, and analytics at Ask.com. "We didn't have a lot of time to spend honing in on analytics. We were very reactive on the BI and analytics side. We didn't have the presence or the time to go out and question the business, dive deeper into analytics, and help them solve the problems that were instrumental to the company's viability. We were busy running our own miniature IT shop, which is where we started looking when evaluating how we could be more successful."

The changes with cloud implementation generally and adoption of Snowflake Computing's cloud-based data warehouse service were significant. "Especially with the JavaScript Object Notation (JSON) support, it became easier for the BI team to own and manipulate changes on the fly," he continued. "It provided us this simple way of having multiple environments, including nonproduction environments and the ability to spin those up to do development work. Then, when we were done, we could spin them down." In an industry rife with tough competition, Ask.com had a "fast, cheap, and flexible way to have a controlled release process without backups and refreshes and pinging data across the environment."

While most of the focus of cloud-based big data platforms is on how well they can deliver high-volume computing performance for massive data volumes, Ask.com discovered other unforeseen benefits. "We also have to do lots of drib-drab small things, such as moving data in from various pipelines. We run our pipelines in real time, so the actual data we copy into the databases is small. By allowing us to stream that data along in a smaller compute footprint, we're not forced into a 'one size fits all' workload." What's more, the company could "now afford to keep the long tail of data in one place," which makes access easier for the data science team by eliminating the time it takes to clean and process data before running algorithms. With the new cloud-based data warehouse process, "the data is cleaned and ready to go in this long tail." Through close interoperability between the data warehouse and Amazon S3 storage, data scientists can spin up additional clusters as needed.

Cloud Satisfaction and Value for BI and Analytics

We have been citing benefits, drivers, and considerations for using the cloud for BI and analytics throughout this report. However, we also wanted to explore how satisfied organizations are currently with various aspects of their cloud deployments. Additionally, we wanted to examine in more detail the value the cloud provides.

Satisfaction with Integration Between On-Premises and Cloud-Based Systems

Satisfaction did not differ dramatically among different aspects of cloud deployments.

For most organizations, realizing a fully integrated and cohesive hybrid data architecture that brings together on-premises and cloud-based systems is still a work in progress (although it is still a consideration as mentioned in the previous section). However, they must still meet pressing needs today for at least some level of integration. We asked research participants about their level of satisfaction with a variety of factors regarding integration between on-premises and cloud-based data warehousing, Hadoop, or analytics systems (Figure 10). Satisfaction levels did not differ dramatically among the factors, but respondents indicated the most satisfaction with integration for identity management (e.g., with LDAP/Active Directory) and integration of internal and external data. For both factors, the combined "satisfied" and "very satisfied" percentage was a shade under half of respondents (49%).

With security a high priority, it makes sense that identity management, which covers users' rights and restrictions for accessing and working with resources over networks, would be a chief concern and therefore the most mature in terms of satisfaction. Many organizations are experienced with using LDAP and Active Directory for their internal networks and have a leg up on extending identity management protections to include access to cloud-based systems.

Integration of internal and external data can be difficult, but it appears that using cloud-based systems allows nearly half of respondents to report this same level of satisfaction, with an additional 20% somewhat satisfied. Nearly as many show satisfaction with integrated access from front-end tools and with ETL and data loading (both roughly 46% when combining "very satisfied" and "satisfied").

Although both areas show room for improvement, the research finds that organizations are making progress in enabling users to access both on-premises and cloud-based systems from their BI and analytics tools. Respondents are equally satisfied with the performance of ETL and data loading across both types of systems, with 29% somewhat satisfied and only 6% not at all satisfied. Increasing data complexity is putting pressure on data integration, transformation, and preparation to be more agile. Enterprises often regard traditional ETL as lacking flexibility, but our research shows that at least for the requirements of nearly half of respondents, ETL and data loading are satisfactorily meeting needs.

How satisfied is your organization with the following factors that have to do with integration between on-premises and cloud-based data warehousing, Hadoop, or analytics systems?

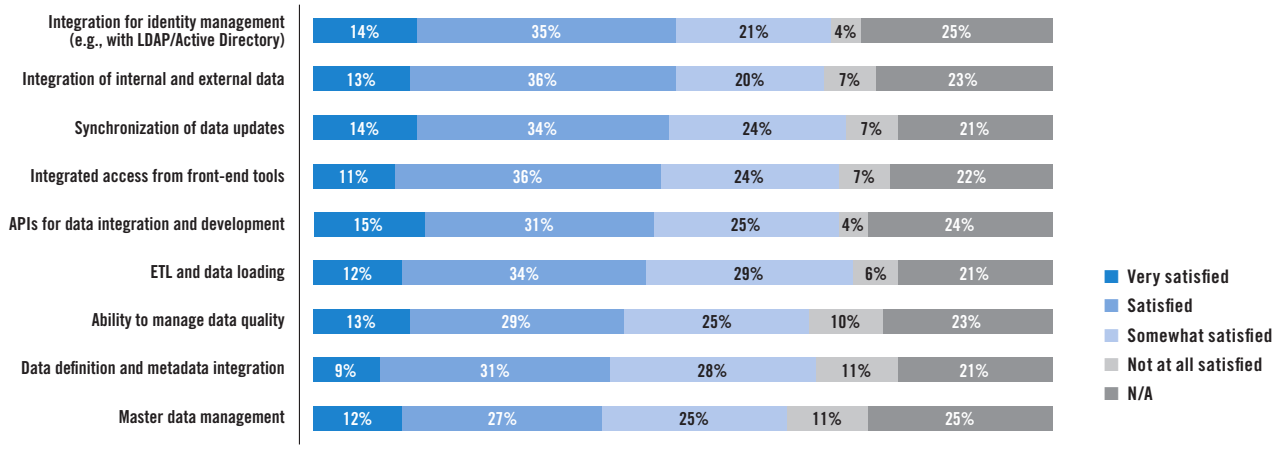


Figure 10. *Based on 217 responses. Ordered by highest total of "very satisfied" and "satisfied" combined.*

USER STORY PROLOGIS UTILIZES CLOUD FOR MULTIPLE BUSINESS FUNCTIONS

Recently TDWI spoke with Cathleen Davidson and John Deahl of Prologis, the world's largest owner of industrial real estate and recent adopter of Informatica Cloud for data analytics, customer management, and HR. On the customer side, the company finds the cloud an indispensable tool for its particular dynamic of offerings. "We build and buy properties that are typically 100,000 square feet and above. Our business is about providing turnkey logistics for properties that our clients lease. We are also a REIT (real estate investment trust) so we are kind of a hybrid organization in that we both own properties outright and also with our fund investors," Deahl reports. Sensing an invariable, oncoming shift in the competitive landscape with regard to data analytics, CRM, and people management, Prologis strategically "took on an initiative to embrace new cloud technologies as much as possible," and its use is integral across all departments. "We have different cloud solutions for major applications. We have Workday for HR, our Operations platform is SaaS, and our CRM platform is Salesforce."

From a BI standpoint, the cloud has allowed a streamlining of data distribution. "Data is a strategic asset and we need to govern it, manage it, improve performance, and get information out to the field so that they have the ability to make better decisions," Deahl continued. The success of this initiative caused Prologis to look inward at just how the cloud might be able to assist with HR functions. According to Davidson, "Our latest project is a real-time feed from Workday to Microsoft Active Directory. We're pretty excited about this because it is our first step toward doing real-time integration with Informatica Cloud."

Using a cloud-hosted enterprise data integration platform has meant unprecedented flexibility to Deahl's team. "I like the flexibility you have in the cloud. We have taken advantage of changing our connectors on the fly. When we realized we were no longer going to use a particular solution as a target, I freed up that slot and we went to another connector for another solution. This has allowed us to do some very innovative development."

Respondents are relatively satisfied with security during data loading.

Our research finds further evidence of respondents' relative satisfaction with security in answers to our question about loading data from on-premises to cloud-based systems (Figure 11). Over half (56%) are either very satisfied or satisfied with security during data loading processes. ETL and data integration technologies are fairly mature, enabling organizations to accomplish this step in a secure fashion. Respondents are also reasonably satisfied with costs associated with data loading; 55% are either very satisfied or satisfied.

Respondents are less satisfied about the time it takes to load data from on-premises to cloud-based systems, though. Under half (45%) are either very satisfied or satisfied, and 9% are not at all satisfied. In interviews, we've found that many organizations struggle with slow load times, especially with large volumes of data. This can become an obstacle to using the cloud for BI and analytics if users are unable to access refreshed data as a result.

How satisfied is your organization with the following factors regarding loading data from on-premises to cloud-based systems?

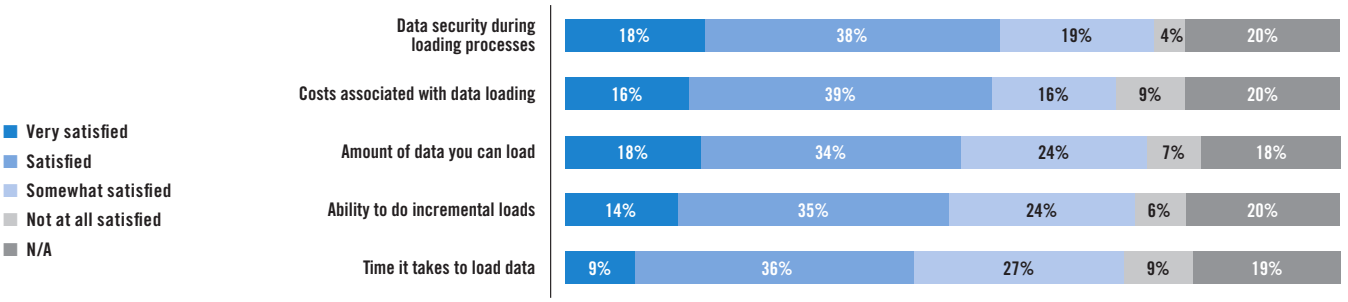


Figure 11. *Based on 217 responses. Ordered by highest total of "very satisfied" and "satisfied" combined.*

Helping to Focus on the Business

Throughout the history of BI and analytics, many organizations have been highly frustrated that it has taken such considerable IT effort and expertise to deliver BI reporting or nearly any kind of analytics capabilities beyond simple spreadsheets. "We don't want to be in the IT business," is a common refrain of organizations that just want to focus on using data effectively within their domain.

Business-side demand for more control over BI and analytics and less IT involvement is driving the strong trend toward use of self-service visual analytics and data discovery tools. Now, in the same way, business leaders want to get away from on-premises data technology constraints limiting their potential to work with data in ways that could add business value. Cloud computing offers the promise of enabling organizations to further "get out of the IT business," at least to some extent, by relying on cloud-based services.

Use of the cloud does help organizations focus more on the business.

In our research, most participants do indeed view cloud computing as helpful in enabling their organizations to focus more attention on business-value-added activities and less on IT issues. They are able to move faster to deploy more advanced analytics, such as predictive analytics, through use of more packaged analytics services (and microservices) discussed previously.

Over half (57%; not shown) said cloud computing was either helpful or very helpful; 32% said it was at least somewhat helpful (4% said it was not helpful and 7% did not know). "Cloud computing" as it is used in this question includes public and private cloud, which in some cases could be simply the use of virtualization inside the firewall to enable greater scalability and ROI with existing on-premises hardware. We also found in interview research that IT, which typically manages virtualization, has been able to use virtualization to overcome some system constraints and support business needs with greater versatility and cost-effectiveness.

Opening Doors to Open Source Tools and Technologies

Open source technologies are of interest in the cloud.

The cloud can give organizations a space to experiment with new technologies and methods for developing BI, analytics, and data management systems. For many organizations, cloud enables them to experiment with open source technologies in particular, such as those that are part of the Apache Hadoop ecosystem. Nearly half (48%) of research participants said that their organizations are considering open source and that the cloud could help them experiment with the technologies. Almost a third (30%) indicated that they are believers in open source and the cloud will encourage further use of open source tools. Only 14% said they are not interested in open source in the cloud or otherwise (8% did not know).

What Drives Measurable Value in the Cloud?

We asked those respondents who are using the cloud now what value they are seeing from their cloud deployments. Fourteen percent (not shown) said that they had measured a positive top- or bottom-line impact using the cloud. Another 48% believe that they have become more effective or efficient but have been unable to measure the impact. The remainder had not measured value (12%) or did not know (25%).

Respondents do believe that the cloud has helped them become more effective.

To explore what drives measurable value in more detail, we broke the respondent pool into three groups: those who are using the cloud now and are measuring value, those who are using the cloud now and believe they are gaining value, and those who are seeing no value. Those measuring value and those not seeing value are both small groups (fewer than 30 respondents), so this analysis is qualitative at best. However, there were some interesting descriptive findings that emerged by examining the data this way.

Those who measured value and are using the cloud appear more likely to be dealing with larger amounts of data than the other two groups (i.e., terabytes and petabytes rather than gigabytes and terabytes). More of these respondents were also involved in projects that include large amounts of disparate data, such as IoT, which often use the cloud as part of the data architecture. These respondents may be more advanced analytically, in general, and are able to extract the most value possible from their analytics deployments that include the cloud. We have seen this kind of pattern in other research.

Additionally, those who measured value are typically very satisfied with the flexibility and scalability that the cloud provides. This is an important driver for them. They like the fact that there is less hardware to support because this gives them the opportunity to focus more on business issues than IT issues (as described previously). As one respondent said, "It's not the infrastructure that is your asset; it is the data that is the asset." This group is very business-value focused and is interested in dealing with business problems.

A Sample of Relevant Vendor Platforms and Tools

Because the firms that sponsored this report are all good examples of vendors that offer technologies to support analytics in the cloud, we present a brief look at the product portfolio of each. The sponsors form a representative sample of the vendor community, yet their offerings illustrate different approaches.[7]

Cloudera

Cloudera delivers a modern data management and analytics platform built on Apache Hadoop and the latest open source technologies. Cloudera Enterprise helps customers solve their most challenging business problems with data. The platform includes the open source Hadoop as well as advanced system management and data security tools plus dedicated support and community advocacy. Cloudera's goal is to provide the same platform and tooling, regardless of whether Hadoop is in the cloud or on premises in the data center. As part of this, Cloudera Director allows users to quickly deploy and manage the life cycle of Cloudera Enterprise clusters in any cloud environment, and it is deeply integrated with Cloudera Manager for a direct connection and consistent administration for cluster-level management and monitoring. Cloudera Enterprise components including Apache Impala (incubating), Apache Hive, and Apache Spark are also integrated with Amazon S3 cloud object storage, enabling companies to effectively use Hadoop as an analytics database for self-service BI, exploratory analytics, and ETL processing against both cloud-native and local storage.

IBM

IBM has evolved to a cloud-first strategy; 98% of its solutions were built with that strategy. At a high level, IBM offers several flavors of the cloud to its clients. The first is analytics products in the cloud (e.g., a SaaS solution). IBM's analytics products such as Cognos and SPSS are offered this way. The other is via IBM Soft Layer, a dedicated, bare-metal, single-tenant environment. Organizations can choose what to use it for. Soft Layer is hosted in the IBM cloud or can be hosted on premises. IBM also offers Bluemix, a shared cloud PaaS environment where organizations can build analytics applications using a catalog of data and analytics services.

Informatica

Informatica provides data management solutions in the cloud, on premises, or in a hybrid environment. The company's goal is to provide end-to-end data integration and data management, on premises or in the cloud. Informatica Cloud Services provides data integration and data management capabilities. Tailored toward any user, it includes templates and wizards for data integration along with hundreds of pre-built connectors for transactional and historical data, both structured and unstructured. Informatica is now adding data quality and master data management as a cloud service. Informatica Cloud Services are fully HIPAA, SOX, and PCI compliant.

SAP

SAP BusinessObjects Cloud is a software-as-a-service (SaaS) solution that consolidates all analytics functionalities for all users in one product: planning and forecasting, discovery, visualization, and predictive functionalities. Built on SAP HANA Cloud Platform, SAP BusinessObjects Cloud is geared to help organizations overcome the challenge of point solutions and data silos spread throughout the organization with enterprisewide access to analytics delivered through a public cloud experience. Targeted at the line-of-business user, these analytics capabilities are offered together

[7] The vendors and products mentioned here are representative. The list is not intended to be comprehensive.

on one cloud platform to provide flexibility, speed, and openness for users. Users can analyze data from other SAP cloud-based products on this platform, in one place, where the data lives. SAP also offers predictive analytics algorithms via the HANA Cloud Platform, which are offered as analytics services that can be consumed through the platform, and a PaaS solution that enables developers to build their own services. All of these offerings are hosted by SAP in the 41 state-of-the-art, in-memory SAP HANA Cloud data centers around the world where you can choose a public, private, or hybrid cloud option. In addition to the PaaS and IaaS offerings with SAP HANA Cloud Platform, SAP has about 15 additional SaaS solutions.

SAS

SAS, founded in 1976, offers technology products and services that span analytics, data management, and business intelligence. The goal of the SAS cloud strategy is to meet customers where they are. For instance, SAS offers software-as-a-service as access to Visual Analytics and Visual Statistics. The company offers managed services, enterprise hosting, and remote managed services (in the cloud or on premises). SAS also offers SAS Viya, an open, cloud-ready, distributed, in-memory architecture. SAS Viya delivers a common, consolidated, and centralized environment for developing, deploying, and managing all aspects of the analytics life cycle—from data management to model management. For developers, SAS Viya provides API access to analytics through RESTful APIs to call analytics routines through native language support for Python, Java, and Lua and through public RESTful API access to the microservices-based middle-tier capabilities. Cloud Foundry is the predominant PaaS environment for SAS Viya deployment, orchestration, update, and management, but as a result of the microservices being delivered as containers, other PaaS and container services can be used as well.

Snowflake Computing

Snowflake Computing provides a cloud native data warehouse solution delivered as a data warehouse as a service and utilizing an architecture designed specifically for the cloud. Snowflake's multi-cluster, shared data architecture decouples data storage, query processing, and metadata management, making it possible to bring together data in a single location while independently scaling computing horsepower on the fly. The data warehouse can be scaled up or scaled down at any time without redistribution of data, read-only downtime, or delay. Snowflake was designed for semistructured and structured data that can come from both discrete and streaming sources. The service allows users to load semistructured data without having to define a fixed schema, and then query that data, in combination with structured data using SQL. A recent addition is a multicluster warehouse capability, where a customer can predefine minimum and maximum number of clusters and allow Snowflake to scale up and down that number automatically to support varying levels of concurrency with consistent performance.

Tableau Software

Tableau Software provides highly visual, easy-to-use tools to help democratize analytics across the organization. It offers a range of capabilities from mobile BI to analyzing big data. Tableau offers Tableau Online, which is the hosted version of Tableau Server for organizations to share dashboards and visual insights. Tableau also offers the ability for customers who are already using Amazon Web Services, Microsoft Azure, or Google to host Tableau there. Tableau's goal is to work with data where it is. For instance, a user can access on-premises data from Tableau online via an agent.

Top 12 Best Practices for Cloud Analytics

In closing, we summarize the report by listing the top 12 best practices for cloud BI and analytics with a few comments about why each is important. Think of the best practices as recommendations that can guide your organization into successful implementations of cloud BI and analytics.

1. **Embrace change.** Many organizations state that they will not move BI and analytics to the cloud (and especially the public cloud) because of (perceived) concerns about data security and privacy. They are concerned about performing analysis in the cloud with company data. In many cases, the real issue is about a fear of losing control over their data. However, organizations that have moved to the cloud are reaping the benefits of its flexibility and agility. It may be time to revisit past assumptions about security.

2. **Address cultural issues.** Tied to number one, change can be hard. Education is critical here as is changing the mindset. Some people don't get it. Some have legitimate concerns. Some are concerned about their jobs. It will be important for those driving change to get executive support (someone who is the champion and can deliver the message) and help to communicate the message.

3. **Use the cloud to help users move faster to realize value from BI and analytics.** Enabling cloud services to support users' BI and analytics requirements can accelerate their progress in applying more advanced and sophisticated tools than common spreadsheets. Many organizations have invested in dashboards, but due to technology, platform, management, and other constraints associated with on-premises deployment, they have been unable to advance beyond simple visual portals and interfaces. Cloud BI and analytics can enable greater depth and flexibility for users' data interaction, which can increase the relevance of dashboards and other elements to their decision making.

4. **Pick the right analytics projects for success.** As is the case in most successful analytics projects, start with a plan. For instance, it may make sense for a public cloud deployment to start with sales or marketing applications because their data may not need to be on premises. This might involve understanding customer behavior to reduce churn using cloud CRM data. Start with the business use case and figure out what make sense. A cloud-to-cloud analysis may make sense when data is generated in the cloud. Always understand the business problem the user is trying to solve, *then* prototype it in the cloud. Plan in phases. This requires collaboration between different parts of the business to be most effective.

5. **Use cloud services to modernize integration between BI, analytics, and business applications.** Organizations can take advantage of modern APIs and microservices approaches to better integrate often highly separate BI and analytics applications and an organization's CRM, financial forecasting, planning, and other business applications. Some solution providers make it straightforward to embed BI and analytics routines inside these applications; cloud-based options give organizations new potential to upgrade analytics, BI, and data visualization functionality more rapidly for users of these applications. Organizations should evaluate users' needs for these capabilities.

6. **Cloud architecture is important.** Often, the cloud is part of a larger, overall enterprise architecture that may include the data warehouse and other data management platforms. As our survey data illustrates, the majority of respondents believe that their cloud deployments will ultimately be hybrid clouds. Users of BI and analytics tools and applications will routinely need to access data from different types of cloud platforms, including SaaS, as well as on-premises systems. The

architecture can become complex, but it will be manageable if a plan is in place. This means reworking architecture plans to determine how platforms will integrate and operate together. If this isn't done, then data and analytics silos will result.

7. **Make cloud data governance a priority.** In line with number three, plan for an expanded governance model when moving to the cloud. This is the case whether you're using SaaS or something more complex. Many organizations expand their current governance strategy to include the cloud. This includes monitoring cloud services and maintaining visibility into cloud operations.

8. **Think about performance and latency up front.** A number of users we spoke with talked about performance issues and data loading. For instance, some were expecting to load large amounts of data (e.g., billions of rows) into a cloud data warehouse and have it respond quickly. However, they still had to do tuning, tweaking, and aggregation to get performance to an acceptable level. Others mentioned that they had not taken their own networks or cost into consideration when thinking about data loading and movement. We recommend that organizations learn about performance tuning prior to launching cloud applications. Think about your network. Consider the initial data load and ongoing data transport (if appropriate) and cost.

9. **Use cloud as an opportunity to experiment with open source technologies.** New projects launched in the cloud give organizations an opportunity to try innovative new technologies based on Apache open source projects. These include technologies in the Hadoop ecosystem such as Apache Spark as well as other open source analytics tools. Give your developers the room to learn about open source technologies and experiment by using them with cloud-based data.

10. **Do not assume that cloud providers have security and disaster recovery covered.** Security is the topmost concern regarding cloud computing. Yet some organizations fail to complete their due diligence and fully evaluate how particular public cloud and SaaS providers keep data secure. The same holds true for disaster recovery in the event of system or database failures. Be sure to vet fully how cloud and SaaS providers handle security and disaster recovery and make sure that their procedures fit with the kind of data your organization is putting in their hands.

11. **Ensure data management can provide the elasticity your BI and analytics need.** "Elasticity" is one of the biggest promises of cloud computing. Yet it can be elusive if BI, analytics, data management, and application development cannot take advantage of parallel processing and other technologies that are important to handling analytics workloads. Also, make sure you understand how elasticity relates to cost. Examine how your cloud provider charges for scaling systems out and up to meet immediate needs and scaling them back when the need passes. This is important for analytics where experimentation is often required.

12. **Focus on providing good integration between on-premises and cloud-based systems.** You do not want cloud-based systems to become new data silos that are difficult to bridge with existing on-premises systems or vice versa. BI and analytics users will need to access and interact with data across both types of systems. Data integration, data management, data quality, and other related technologies will need to be modernized to span both types of systems if users and automated applications are to gain comprehensive views of all relevant data.

Cloudera
www.cloudera.com

Cloudera is revolutionizing enterprise data management by offering the first unified platform for big data, an enterprise data hub built on Apache Hadoop. Cloudera offers enterprises one place to store, access, process, secure, and analyze all their data, empowering them to extend the value of existing investments while enabling fundamental new ways to derive value from their data. Cloudera's open source big data platform is the most widely adopted in the world, and Cloudera is the most prolific contributor to the open source Hadoop ecosystem. As the leading educator of Hadoop professionals, Cloudera has trained over 40,000 individuals worldwide. Over 1,700 partners and a seasoned professional services team help deliver greater time to value. Leading organizations in every industry plus top public sector organizations globally run Cloudera in production.

IBM
ibm.com

IBM has the most complete set of capabilities to enable today's logical data warehouse. IBM's best-in-class data warehousing solutions span from on-premises appliances (PureData System for Analytics) to Hadoop solutions (BigInsights) to cloud-based data warehouse (dashDB) deployments that address virtually every information need: structured, semistructured, or unstructured. IBM's commitment to innovation focuses on designing the right mix of data platforms and integration capabilities for our clients' changing business requirements to handle any type of data, whether streaming or at rest. To learn more visit: **ibm.com/data-warehousing** and **ibm.com/Hadoop**.

Informatica
www.informatica.com

Informatica is 100% focused on data because the world runs on data. Organizations need business solutions around data for the cloud, big data, real time, and streaming. Informatica is the world's No. 1 provider of data management solutions in the cloud, on premises, or in a hybrid environment. The Informatica Platform is an intelligent data platform that includes data integration, data quality, master data management, data archiving, and data security capabilities to empower customers to transform raw, possibly unsafe data into great data for more accurate insights and more effective decisions. More than 7,000 organizations around the world turn to Informatica for data solutions that power their businesses.

SAP
www.sap.com/cloud

SAP is a cloud company driven by SAP HANA in-memory technology. We have 110 million cloud subscribers and 41 state-of-the-art data centers around the world. We offer cloud apps (software-as-a-service, or SaaS) for all lines of business, a market-leading cloud platform (PaaS), and flexible on-demand infrastructure (IaaS). You can rely on proven enterprise cloud security and hosting services—and choose from a public, private, or hybrid cloud environment.

In addition to the PaaS and IaaS offerings with SAP HANA Cloud Platform, SAP has around 15 SaaS solutions.

SAP BusinessObjects Cloud is a next-generation SaaS solution that combines all analytics capabilities—including planning, predictive analytics, and business intelligence—in a single solution. Take advantage of a modern, intuitive user experience—and save time by planning, analyzing, predicting, and collaborating in context. SAP BusinessObjects Cloud has native integration with on-premises applications and data sources to allow your business users access to insights no matter where the data is located.

The SAP Digital Boardroom equips C-level executives with real-time contextual information and ad hoc analysis. Built on SAP BusinessObjects Cloud, this next-generation board portal leverages SAP S/4HANA lines-of-business data to provide a single source of truth for the company.

SAS
sas.com/visualanalytics

SAS Visual Analytics is the flagship product for interactive, self-service business intelligence and analytics. It allows business users and analysts to visually integrate and explore data, design and share interactive reports, and employ an array of easy-to-use analytics capabilities for informed decision making.

For citizen data scientists and statisticians interested in predictive analytics, SAS Visual Statistics (an add-on to SAS Visual Analytics) can interactively explore data and build descriptive and predictive models in tandem.

What does this mean for users?

- No matter the size of your organization or data, explore all relevant data quickly, identify relationships, and make more precise decisions faster than before. Self-service visual data discovery puts insights within reach in a governed, reusable manner.
- Quickly design reports and dashboards that are attractive and meaningful, and then easily distribute them via Web, Microsoft applications, or mobile.
- Build queries to perform joins, add calculated columns, subset and sort data, load data in memory, and more with Web-based data mashup and modeling capabilities.
- Interactively customize models—adding or changing variables, removing outliers, etc.—and instantly see how those changes affect outcomes.
- Rightsize your environment with options that fit your needs. Deploy on site or in the cloud.

Learn more at sas.com/tryva.

Snowflake Computing
snowflake.net

Snowflake Computing, the cloud data warehousing company, has reinvented the data warehouse for the cloud and today's data. The Snowflake Elastic Data Warehouse is built from the cloud up with a patent-pending new architecture that delivers 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. The company is backed by leading investors including Altimeter Capital, Redpoint Ventures, Sutter Hill Ventures, and Wing Ventures. Snowflake is headquartered in Silicon Valley and can be found online at snowflake.net.

Tableau Software
tableau.com

Tableau helps people quickly analyze, visualize, and share secure data and interactive dashboards. From small businesses to the world's largest companies, governments, and universities, people everywhere use Tableau to transform raw data into actionable insight. Tableau Desktop lets you connect to nearly any data source with just a few clicks and build live visualizations via an intuitive drag-and-drop interface. Tableau Server allows you to share governed access to analytics across your organization—enabling the right people to collaborate, analyze, and find answers with trusted data, all from a browser or mobile device. Tableau Server can easily be deployed on all major public cloud providers. Tableau Online is the software-as-a-service implementation of Tableau Server. It enables you to share, manage, and secure your analytics without the need to configure or manage software. With Tableau Online, your analytics is securely hosted in the cloud, accessible across your organization and to customers and clients anywhere, at any time.

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.

tdwi
Advancing all things data.
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