How IT Taps Big Data

to Optimize Digital Operations and Drive Business Advantage
Table of Contents

**Executive Summary: Advancing Big Data Analytics for Digital Business** ................................................................. 1

**How IT Taps Big Data to Optimize Digital Operations and Drive Business Advantage** ........................................ 3

- IT’s Priority for Action: Data Quality .................................................................................................................. 3
- Analytics Goals Set the Data Management Agenda ............................................................................................. 6
- Investing Wisely in Real-Time Data .................................................................................................................. 7
- Chasing the Single View of Data, Plunging Into the Data Lake ........................................................................ 8
- The Challenge of Cloud .......................................................................................................................................... 9
- How CIOs Can Drive Analytics Forward ........................................................................................................... 10

**About the Research** ............................................................................................................................................. 11
**Acknowledgments** .............................................................................................................................................. 11
**Sponsor’s Viewpoint** .......................................................................................................................................... 12
Executive Summary: Advancing Big Data Analytics for Digital Business

Big data is the fuel that powers business innovation and optimization. Leveraged through a robust analytics practice, it has the potential to help companies improve their operations, products, and customer experience. It can build lasting competitive advantage and even lead to new business models — but attaining that level of payoff requires multifaceted efforts. What strategies and tactics are the most advanced analytics users pursuing?

To gain insight into analytics-led transformation, we looked to CIOs and IT executives, who are typically at the heart of these initiatives. IT maintains the systems and networks through which all data flows and typically enables the collection, cleaning, integration, storage, and access to data for analytics. And IT itself uses data analytics to keep the digital business infrastructure running securely and at peak performance.

In October 2018, we surveyed more than 300 executives about their data analytics goals and practices. How are they managing the data infrastructure to support analytics within both IT and the business? What actions do they see as most necessary to improve analytics? What are their biggest challenges? How advanced are they on the analytics journey? What are organizations with the most advanced data analytics practices — characterized throughout this report as “leaders” — doing differently?

Data quality is the top priority, with more than 1 in 5 overall and more than 1 in 4 leaders ranking it as the top priority action for improving analytics. However, data integration/transformation, storage efficiency, and data collection clustered close behind as the remaining top four priorities for all respondents. Interviews with data and analytics experts underscored the
Leaders expressed the greatest confidence in data quality and also placed the greatest emphasis on ongoing efforts to maintain trustworthy data.

Interplay between these activities and how, at different stages and for different use cases, they are all critical to serving the analytics mission.

Companies trust their data to be accurate, complete, and up-to-date — faith that some experts believe may be misplaced. Nevertheless, it appears that many companies acknowledge that data quality is not a “one and done” activity: Leaders both expressed the greatest confidence in data quality and also placed the greatest emphasis on ongoing efforts to maintain trustworthy data.

Companies are moving toward single, multiuse data lakes. Big data leaders are committed to data lakes: 96% either already have or are planning to implement one. But data silos persist, perhaps reflecting the reality that sometimes the most useful source of data is the point of origin.

Cloud/data center integrations are a big pain point. Organizations identify the three most significant challenges to achieving a single view of data as cloud/data center integration, incompatible technologies, and siloed infrastructure.

Goals for analytics become more business oriented as practice advances. While respondents overall listed their top goal for analytics as managing and improving IT infrastructure, more-advanced practitioners are likelier to have broader goals that include delivering results to the business through predictive analytics.

Budgets are up, having increased across the board for analytics, especially for business analytics, where 82% of leaders report an increase.

Real-time data is considered key to a range of use cases, including IT operations and security, customer experience, network operations, and business unit operations.

Respondents ultimately want to ensure data security and performance across organizational infrastructure, and support analytics that optimize the business of the company. But our research also reveals that many still face significant challenges as they work to advance the creation and use of data-driven insights within and beyond IT.

We hope this insight into the state of practice, combined with the perspectives of seasoned analytics veterans, will help CIOs and other technology leaders gain advantage from their data assets.
For the past several years, organizations have struggled to get their arms around the exponential growth in data. They face an array of choices: What should they put in the cloud, and which cloud should they put it in? What should they keep on premises? Which data should be available to which departments and job functions?

But informing all of those tactical decisions about big data is an even more important question: How might data analytics build competitive advantage or add value for the organization? The most successful data initiatives must drive to the top business problems, says Donna Roy, executive director of the Information Sharing and Services Office at the U.S. Department of Homeland Security (DHS). Without that alignment between analytics and C-suite priorities, “you’re not getting the visibility to fix the problems, getting access to the data, getting quality data, and getting decisions on the data,” she says.

Our research sought to investigate technology management’s perspectives on big data and analytics because IT has a uniquely panoramic view. Not only does it build and maintain the data infrastructure that serves business analytics users, but IT also uses data to optimize the entire digital infrastructure underpinning business. About 80% of our respondents come from IT, with the remainder in data analytics or C-suite roles with significant analytics oversight. We looked for insight into organizations’ analytics goals, what data management strategies they are pursuing, and what challenges they are encountering. We also sought to identify what the most advanced organizations are doing that may be enabling them to derive more value from their analytics practice.

We identified the advanced cohort from the 56% of survey respondents who characterized their organizations as having strong data and analytics practices. This group was evenly split between leaders, who are gaining significant advantages from data and analytics relative to their competitors, and mature practitioners, who are widely using data and analytics to advance decision-making and are advancing their data science practice.

**IT’s Priority for Action: Data Quality**

For all, but especially so for leaders, the classic tagline “Quality Is Job One” could be the motto of today’s enterprise data analytics efforts.

Asked to choose the single most important area for action that would improve analytics, leaders were more likely to choose data quality than respondents overall were; for the surveyed group as a whole, data quality and data integration/transformation were equally ranked priorities (see Figure 1). Storage efficiency and data collection tied in the next two spots overall — but among leaders, data collection came in a strong...
second to data quality. That may indicate that more-advanced practitioners have developed more potential use cases and are seeking data to feed new analyses. It may also reflect a realization, gained through experience, about how data collection processes may influence quality.

The only others to rank collection highly were organizations just getting started — in fact, they rated collection, rather than quality, as the top priority for action. That makes sense, since an organization just dipping its toes into analytics must first establish what data it needs and how it will collect it.

For purposes of the survey, we defined quality in simple terms: Is it accurate, is it complete, and is it up-to-date? Using those measures, we found respondents place a high level of trust in the quality of their data. Not surprisingly, the more mature the organization, the higher the confidence level (see Figure 2). More than 90% of leaders usually trust data to be accurate, complete, and up-to-date. Among those just starting, confidence levels drop below 70%.

Despite this apparent confidence, when asked how data quality issues affect analytics efforts, 54% of the overall group said it slows their ability to use data for certain use cases, while 50% said it prevents them from implementing certain use cases.

![Figure 2: Confidence in Data Quality Grows With Analytic Maturity](image)

In addition, respondents indicated that a fair proportion of data requires some amount of work before it can be analyzed. When we asked respondents to gauge the extent to which their data was ready to analyze, 58% of leaders reported the data they retrieve is ready; 30% said it requires minimal transformation/cleansing (see Figure 3). Notably, the second-most-advanced group (mature) are less confident and reversed those two ratings: Only 26% reported the data ready to use; the majority (62%) said it requires a minimal amount of transformation/cleansing.

Clearly, companies realize it takes ongoing work and resources to maintain trustworthy data. At the same time, the survey results also may reflect some wishful thinking.

“The simple reality is that most data is pretty darn bad,” says Tom Redman, president of Data Quality Solutions, a consulting company. “People spend an awful lot of time correcting the data, making sure they understand what it means, lining up units of measure,” he says. “Data scientists spend about 80% of their time just cleaning up the data, and they are not very good at it, so a lot of errors still get through.” That undermines analytics consumers’ trust in data, and that in turn undermines trust in analytics.
Redman says that looking at accuracy, completeness, and timeliness gives just a partial view of data quality. The ways in which data is collected, cleansed, transformed, and integrated also contribute to whether data is ready for analytics. In broadly assessing data quality for an analytics application, it’s important to consider whether you have assembled a complete data set, and whether you’ve pulled things together in the right way, he says. For example, one set of data might use yards as a measure while another uses meters, he points out. “I think there is this belief that your [data is] in pretty good shape, but a nagging little voice in the back of your head is whispering, ‘Maybe we should worry about this,’” he says.

Roy of DHS agrees that data quality is a continuous process. “Data quality is something that happens at different intervals, depending on the subject of the data,” she says. “If it’s about humans, data quality is a daily, weekly, real-time issue, much more than it is about something that doesn’t change very often.”

“Right now, we deal with data on a very transactional level as opposed to an event-driven model,” Roy says. “We treat data as a point-in-time fact, and we store it in our data lake and forget that it represents something that can change.” Instead, she says, “Data should be treated more like a living organism that evolves as the facts it represents evolve.”

In addition, quality is not binary. Roy says it’s important to consider the level of quality needed for any given application, as well as the potential impact of getting things wrong. In some situations, such as determining whether to allow someone to board an airplane based on indicators of terrorist-related activity, the most-recent information is critical and the consequences of a mistake potentially lethal. In less critical applications, it’s a matter of deciding what level of quality is “good enough.”

Another aspect of quality revolves around how and when data is collected. Companies have routinely collected certain data for years but may have only recently started using it for analytics, Redman explains. However, most do not reevaluate their collection process to make sure it matches the new use case. For example, companies are often enthusiastic about using sensor data from internet of things (IoT) devices, but they fail to check the device’s data quality. “When I ask them when was the last time that particular measurement device was calibrated, I usually get this blank look,” he says. “Nobody is responsible for making sure that data is correct.”

Given the many ways that companies assess data quality, data governance policies are critical, says Claire Walsh, who leads the data and analytics practice at consulting firm Excella. “If you have a data inventory and you have some data quality standards in place, then users will have an idea of how important various quality characteristics are for a particular application.”

As head of data governance at a large consumer electronics company, Stephen Gatchell is a firm believer in the value of data inventories or catalogs. Without them, companies have no way to know where the information originated, who owns it, and how it should be used, he says. “If you have an issue with the data’s quality but don’t know where it came from, how do you get that data quality problem resolved?”

Good data governance requires organizations to hash out standard definitions, a vital foundational element for accurate analytics. Gatchell is now engaged in that effort at his company, inventorying the data on all of its systems, identifying data owners, and establishing a business glossary. Even a simple word like customer can have different meanings across a company depending on whether the context is IT, sales, or procurement.
Analytics Goals Set the Data Management Agenda

While many companies default to data management as a starting point for streamlining data optimization, many expert practitioners recommend taking a step back. For them, understanding business goals for analytics projects — what questions someone hopes to answer with data or processes they wish to drive — must ultimately shape those data management activities. An organization’s specific analytics goals will also affect considerations such as how broadly data access must be enabled and which non-IT stakeholders or data owners must be included in planning and execution.

Most respondents to our survey focused on analytics goals centered around IT operations (see Figure 4), but differences in emphasis emerge between leaders and other survey respondents. The greatest percentage of leaders (64%) cite improving application delivery and performance as a goal, while the goal of the highest percentage of others (46%) is managing and improving performance of IT infrastructure. Better identification and remediation of security threats and improving customer experience round out the top goals.

Respondents were able to choose all goals that applied to their use of analytics, and notably, leaders were far more likely to choose multiple goals. In doing so, leaders demonstrate the broadest sense of mission, with many more listing analytics goals that deliver business results, such as predictive analytics for optimizing business processes and business decisions as well as optimizing digital business operations.

Leaders’ broader range of goals and greater orientation toward the business sets them apart from less analytically mature organizations, where database administrators (DBAs) and data scientists may work together but isolated from business stakeholders.

That functional siloing at many would-be data-driven organizations has been observed by Dean Abbott, chief data scientist at SmarterHQ, a behavioral marketing firm that does advanced analytics for large retailers.

In that scenario, “You have DBAs and data scientists in the room, trying to make their best guess of what to do with the data, and they get it wrong because they don’t understand the business. And you’ve got business stakeholders desperately trying to use the data, but they don’t know the right analytics question to ask,” says Abbott.

“If you want to be a data-driven company, the data has to be organized in a way that enables you to make decisions easily,” Abbott adds. “The core and the key to all that is the data model and infrastructure.” The infrastructure makes the right information available at the right time to the right people, while the data model is the plan for collecting, organizing, and making data available. The streams of data collected from a website, for
instance, need to be sorted depending on what you’re trying to do. Do business users want the data organized by customer? By product?

Projects can also fail if organizations don’t have an accurate understanding of where data resides and how it can be accessed. Abbott cites a health insurance provider that wanted to use analytics to preemptively identify Medicaid fraud rather than respond to it after the fact. Late in the project, however, the company realized that certain data elements about higher-risk physicians resided in external systems from which they could not be imported as quickly as needed. Mapping out the systems and network infrastructure at the beginning would have revealed the flaw.

Abbott says developing a solid data model requires strong communication among three groups:

- Business-side personnel, who know what questions they want answered.
- IT, which understands the infrastructure and how and where data is stored.
- The data science group, which can build the right models.

Without proper orchestration among the three, projects can go off the rails. He’s seen cases where data scientists “build fantastic, insightful models that solve irrelevant problems because they haven’t tapped into what the business really needs. They can spend six months and $10 million on all this infrastructure only to discover that they’re scratching the wrong itch.”

**Investing Wisely in Real-Time Data**

Companies are ready to spend more money on analytics, the survey confirms, with a majority of respondents reporting budget increases (see Figure 5). The area leaders most frequently cite as receiving an increase is business analytics (82% among leaders), followed by security and network/application/cloud analytics.

Survey respondents consider real-time data to be key across a wide range of analytics use cases (see Figure 6). More than 70% say it’s important for IT operations and security; more than 60% consider it very important for customer experience, network operations, and business unit operations; and more than 50% deem it very important for sales and marketing.

It’s easy to understand why real-time data is key for IT operations, particularly in guarding against security threats; it can also be important to determining and quickly deliver-
ing on customer needs in some business applications such as e-commerce. But to make sure money is wisely spent, companies should determine exactly what data they need in real time, which isn’t as easy as it sounds. For example, DHS’s Roy says that while real-time data is often requested, make sure to verify the business case first, since using real-time data can increase the effort and expense of a project. “People always believe they need real-time or near real-time data, but if you push them with specific questions, you usually find that’s not the case,” she says.

Even worse, “People often can’t tell us what questions they need answered,” says Walsh. In a project for an assisted living facility, for example, the business groups were stymied by this question. “They said they wanted to do more with data, but they couldn’t point to the business problem they wanted to solve,” she says. In such cases, it’s better to do a small, limited pilot project first. After running the pilot, “You have something to show people as an example, and they can react to that,” Walsh says.

### Chasing the Single View of Data, Plunging Into the Data Lake

Organizations surveyed have clearly bought into the ideal of having a single view of data. Among leaders, 65% report they are implementing or have already accomplished a single view. For the overall group, 48% are implementing or have accomplished that single view, with another 28% in the planning stage (see Figure 7).

Data lakes, which generally store data before any cleansing or transformation takes place, are one of the key elements used to build that view. The survey found that the most advanced practitioners are the most committed, as some 96% of leaders say they have or plan to implement a single, multiuse data lake, versus 79% overall.

On the surface, a data lake is appealing. “You don’t have all the overhead associated with traditional file systems and databases,” Abbott says, while Walsh points out that centralized data lakes offer great exploratory avenues for data in its raw form. “It doesn’t all have to be cleansed anymore as people try to answer particular business questions or build out a machine learning model, for example,” she says.

However, data lakes are not a silver bullet when it comes to storage. While most respondents report they have centralized data storage (74% for all, 81% among leaders), many are also storing at least some of it at point of collection or in silos (see Figure 8). In fact, leaders are more likely to use a combination of data stores, or even multiple data lakes.

---

**Figure 7: Leaders Are Closest to a Single View of Data**

![Leaders Are Closest to a Single View of Data](image)

**Figure 8: Siloed Data Storage Persists**

![Siloed Data Storage Persists](image)

*Note: Percentages may not total 100 due to rounding.*
Why? While having a central, consistent, enterprise-wide view of certain data is important, much data analysis uses smaller slices of data. The DHS’s Roy, for one, favors having different lakes for different applications, because different types of decision-making require different data with different levels of quality. “You might need to build in different tools or re-form your data for different analyses,” she says.

DHS learned this lesson several years ago when it tried to use data from a single lake for vetting individuals at border crossings and TSA checkpoints. “We ultimately realized that Customs and Border Protection was closer to the data, closer to understanding the data, and had an operational tempo that was [faster] than headquarters,” Roy explains. Now, DHS is moving toward an ecosystem of data lakes, each built for the operational needs of a particular agency or purpose but still interconnected.

Companies also need to institute rules to keep their data lakes tranquil. That same lack of overhead can also lead to organizational chaos. “At some point, you have to organize your data,” Abbott says. “It can easily become unwieldy and so unmanageable as to be unusable.”

Moreover, data lakes are not for beginning swimmers. Walsh, for example, recommends access limits.

“Users need to have the appropriate skills to go swim around in that data and explore,” she says. “You have to have some type of subject matter expertise about the data, and you have to have some data analysis and querying skills to be effectively using lake data. If you don’t have those, then you should be using data that has already been cleansed and aggregated and put in a nice, easy-to-use format.” Otherwise, she says, “if you torture data long enough, it’ll confess to anything.”

On the other hand, Gatchell says that having all data in one environment doesn’t necessarily make it the sole source of data, even for experts. “Certain analysts and scientists might use the data lake, but other analysts are going directly to source systems like SAP or Oracle,” he says. “If people are going to SAP and using that information, then why am I putting that into my data lake? Now I have two copies of it, which increases my risk.” He says he would build ways to connect multiple data sources rather than physically move the data.

### The Challenge of Cloud

Even as they aim for a single view of data, respondents acknowledge significant challenges, with the most prominent being cloud/data center integration, incompatible technologies, and siloed IT infrastructure (see Figure 9). Leaders are somewhat more likely to see cloud/data center integration as their top challenge compared with less advanced practitioners (46% versus 39%), which may reflect both their greater adoption of multiple clouds or hybrid architectures, as well as greater awareness of the visibility challenges created by multicloud implementations.
The challenges will remain because hybrid architectures are adopted for good reasons: Health care companies, for example, have sensitive data they want to keep on premises, but they want to move aggregated, de-identified data to the cloud for analysis, notes Walsh. “It’s always a struggle to pull different information from different sources,” she says. And combining such data in real time is especially challenging, particularly for web-based applications. “People expect information to pop up quickly,” she notes.

That can also be the case among different departments of the same company. Siloed infrastructure persists even among those also building central stores — more so among leader companies, which may have to accommodate more legacy data stores.

How CIOs Can Drive Analytics Forward

In the race to gain operational and competitive advantage from big data analytics, there are no shortcut solutions that IT can deliver to solve data management challenges. But strategically minded CIOs know that ultimately, advancing analytics — and equally data-driven technologies like machine learning — is a top priority to effect digital transformation. Here’s what those IT leaders should keep in mind to help drive competitive advantage for their organizations:

1. Look ahead to automation for data quality help.

DHS’s Roy, whose staff spends about 80% of their time collecting, ingesting, and preparing data, envisions a day when technologies like software bots will help with that workload. Such bots could make data smart — able to describe itself and what it can and cannot be used for. Data would be tagged in ways that enable it to clean and update itself, not just in one silo but everywhere that particular piece of data lives.

Maintaining consistent, high-quality data is going to become even more challenging as 5G broadband wireless technology comes online and data pours in from the internet of things, she warns.

“If we don’t start thinking about data quality differently, we’re going to be much, much further behind than we already are,” says Roy.

2. Align data storage strategies with analytics uses.

Despite the investment in a single view of data and to data lakes among our survey respondents, experts interviewed earlier in this report have pointed out that those approaches don’t create data accessibility and quality nirvana. Specific analytics use cases will determine the extent to which data should be transformed and integrated, and then where it is stored. For example, “The quality required for data in a data lake is quite different than quality in a data mart,” says Excella’s Walsh. A data lake typically holds the most up-to-date data, but it’s in raw form. In a data mart, data is not only up-to-date, but it is cleansed and consistent. Also consider whether sometimes data is best left siloed rather than risk proliferating copies around the organization.

3. Look for analytics opportunities across the company.

CIOs who want to lead digital transformations in their companies are looking at how analytics can optimize both IT and business operations. Just as the leaders among our survey respondents had a more expansive set of goals for analytics than others and were more likely to have business applications in mind, CIOs should look across the business to understand the organization’s top strategic objectives. As corporate technology investments increasingly flow to business analytics, CIOs who proactively take an expansive view of analytics’ potential will maintain leadership and increase their influence in the data-driven revolution.
About the Research

We surveyed 309 respondents in North America in October 2018. Of those responding, 80% identified as being in IT management, with the remainder being data analytics leaders or C-suite executives with significant responsibility for big data and analytics strategies and activities.

Acknowledgments

**Dean Abbott,**
chief data scientist, SmarterHQ

**Stephen Gatchell,**
data governance head, consumer electronics company

**Tam Harbert,**
independent writer and researcher

**Tom Redman,**
president of Data Quality Solutions

**Donna Roy,**
executive director, Information Sharing and Services Office at the U.S. Department of Homeland Security

**Claire Walsh,**
practice lead, Data & Analytics, Excella
SPONSOR’S VIEWPOINT

Maximizing the Big Data Advantage

When it comes to digital business, we live in a world of instant gratification. Customers expect accurate answers at the swipe of a screen, while companies churn out goods and services at an ever-faster pace. And in this increasingly connected digital ecosystem, the coin of the realm is data, as companies increasingly turn to big data analytics to drive innovation and run their companies more efficiently. But what does it mean to optimize data across an organization?

After all, it’s pretty clear by now that building a data-driven business model is key to survival in this day and age. But like any major disruptor, this new digital world didn’t come with a road map, and indeed, the landscape seems to change on a daily basis.

Today, companies collect truly astounding amounts of data — and then must rapidly transform and integrate it in order to react to customer demands. Similarly, our astoundingly connected world sprawls from data center to cloud services, across partner systems and global networks — but how do we make sure that infrastructure performs at the speed customers expect, and does so securely?

This puts both IT and business-side workers at an interesting crossroads as they balance the need to both create high-quality big data and then deliver it reliably. Companies want real-time data, and many aspire to a single source of truth. But that’s easier said than done, as most organizations must first integrate data centers and cloud services, flatten technical and operational silos, and somehow unite incompatible technologies.

As it turns out, taking full advantage of data assets involves far more than analytics, and ensuring data security and performance is a critical piece of the big data equation.

— NETSCOUT