



# Building a Winning Data Strategy

# CONTENTS

---

EXECUTIVE  
GUIDE

## Building a Winning Data Strategy

### 1 Why Culture Is the Greatest Barrier to Data Success

By Randy Bean

### 4 Getting Serious About Data and Data Science

By Thomas C. Redman and Thomas H. Davenport

### 7 Data Governance in the 21st-Century Organization

By Gregory Vial

### 11 Actioned Analytics Pave the Way to New Customer Value

By Barbara H. Wixom and Gabriele Piccoli

### 14 How Organizations Can Build Analytics Agility

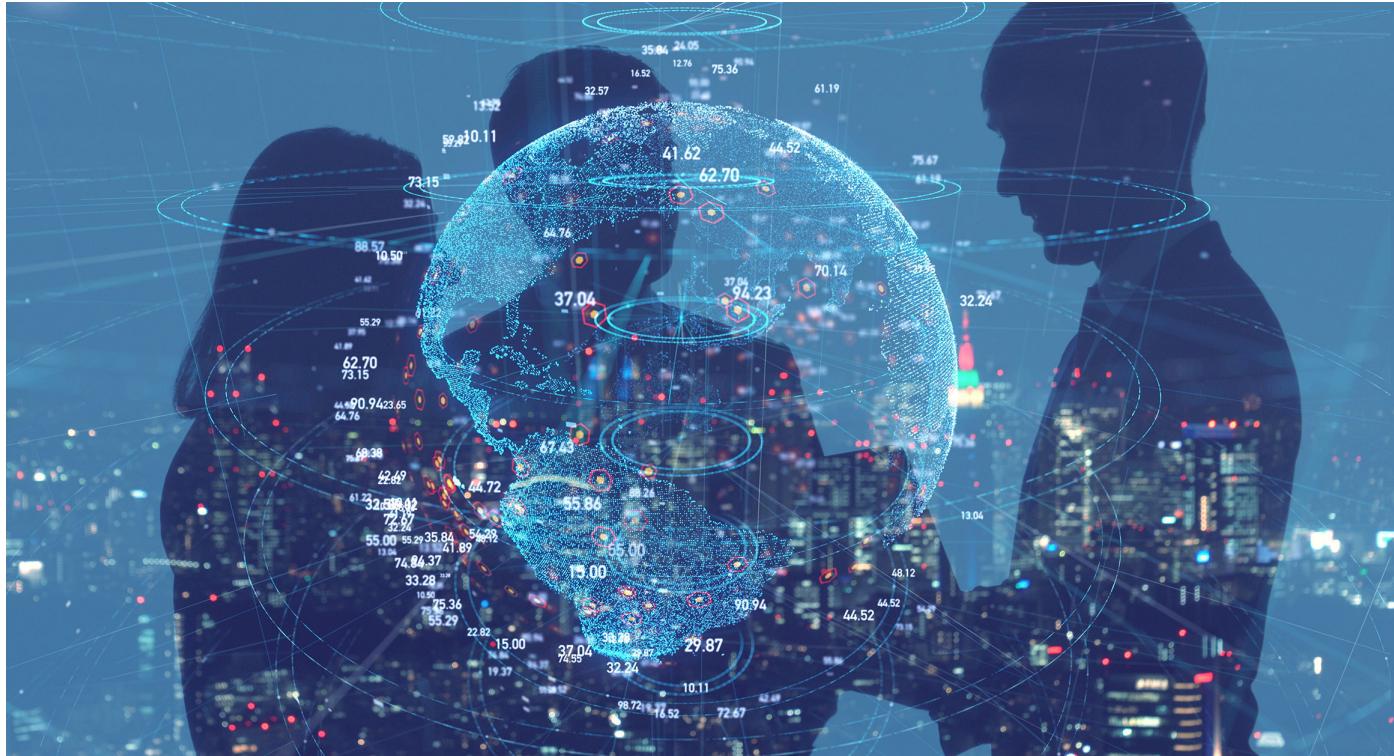
By Lori C. Bieda

### 17 The Enterprise Systems That Companies Need to Create

By David Waller and Paul Beswick

### 21 Becoming a Data-Driven Enterprise: Meeting the Challenges, Changing the Culture

Sponsor's Viewpoint from AWS



# Why Culture Is the Greatest Barrier to Data Success

To be successful with data and analytics, organizations must evolve and change the ways in which they structure current business processes.

BY RANDY BEAN

In order to compete in the new digital economy, businesses must become increasingly data-driven. Few executives would dispute this objective. Recent events, including the global outbreak of COVID-19, have underscored the critical importance of having reliable data to inform organizational decision-making. Yet companies continue to struggle to operate in a data-driven manner. Why?

Even though we are now decades into the age of competing with data, a 2020 NewVantage Partners survey of C-suite executives representing more than 70 Fortune 1000 companies found that only 37.8% of companies have created a data-driven organization. In the same survey, 45.1% of executives reported that they compete on data and analytics, but a majority — 54.9% — stated that they do not.

Meanwhile, investments in data are increasing, with 98.8% of organizations reporting that they are investing in data initiatives. Nearly a fifth of these companies — 18.3% — have invested more than \$500 million. Changes are also happening when it comes to enterprisewide governance. Large corporations have increasingly embraced the newly established role of chief data officer (CDO). While only 12%

of companies had appointed a CDO as of 2012, this percentage increased to a high of 67.9% by 2019 before falling back to 57.3% in 2020. Among major companies today, there is nearly universal acceptance that data-driven management is strongly preferable to the alternatives.

In spite of the growing consensus and investment levels, only half of organizations — exactly 50% — reported that they are managing data as a business asset. The advent of big-data solutions and a next generation of data management capabilities — Hadoop, data lakes, DataOps, and modern data architectures — have been helpful but have not assured successful business adoption or outcomes. Technology does not appear to be a barrier or the problem. Only 9.1% of executives pointed to technology as the principal challenge to becoming data-driven.

Albert Einstein is said to have remarked, “The world cannot be changed without changing our thinking.” What is clear is that the greatest barrier to data success today is business culture, not lagging technology. In fact, cultural factors — that is, people and process issues — were cited by 90% of executives as the principal obstacle that they face. It is not enough for companies to embrace modern data architectures, agile methodologies, and integrated business-data teams, or to establish centers of excellence to accelerate data initiatives, when only about 1 in 4 executives reported that their organization has successfully forged a data culture.

Cultural change and business transformation must be adopted at all levels of an organization for data-driven management to be truly embraced. Having professionally engaged with scores of large organizations over the course of the past two decades, each at a varying stage of maturity, I have found that certain actions distinguish successful data-driven companies from those that continue to struggle.

**1. Secure executive commitment, not just lip service.** Executive commitment is essential to building a culture where data is central. Pay attention to what companies do, not what they say. Most pay lip service to the criticality of data in their annual reports and company mission statements, but far fewer companies embody it in their DNA or in

their day-to-day business practices. Some companies, like Capital One and American Express, have a history of embedding data in all aspects of their business culture. For most legacy companies, however, building a data culture remains a challenge. Companies that have overcome it instantiate data processes throughout their supply chains, from data production to data consumption. I have met with organizations that have proclaimed their commitment to creating a culture where data is a priority without making the necessary investments and following through on their proclamations in order to drive real change and data-driven business outcomes. Don’t be one of them.

### **2. Expect to work hard, and forget about magic bullets.**

Companies that have succeeded in their data-driven efforts understand that forging a data culture is a relentless pursuit, and magic bullets and bromides do not deliver results. To borrow from Thomas Edison, becoming data-driven is 1% inspiration and 99% perspiration. I have seen too many companies undertake big-bang, overly ambitious initiatives that fail over time. Start simple. Focus

on key business questions. Tie data investments to business outcomes. Realizing quick wins enables organizations to build credibility and establish sustainable momentum. Companies such as Cigna, Nuveen, and Citizens Bank are embarking on long-term efforts to define use cases with strong business sponsorship. Companies that accept that there is no easy path to success fare best over time.

### **3. Establish realistic expectations, not unattainable goals.**

Only 28% of companies reported that the CDO role has been successfully established within their organization. Why is this? One reason is that many organizations have struggled to establish realistic and achievable objectives for the CDO. Data is an asset that flows across an organization, and managing data is therefore complex. CDOs must establish achievable expectations to ensure success. I was once asked by a business line president for a plan to make his organization data-driven within 90 days. Until organizations develop attainable goals, it will be impossible to achieve successful data outcomes.

The greatest barrier to data success today is business culture, not lagging technology.

#### **4. Make steady progress, and overcome false starts.**

For many organizations, impatience to see results in the short term often leads to false starts, whether it is the investment in and then abandonment of technology initiatives or the launch of and lack of follow-through in data programs across the organization. The CDO's short tenure for many companies reflects both the nascence of the role and a lack of clear expectations. Some of the largest banks are on the fourth or fifth iteration of their CDO role within the past decade. For many organizations, the tenure of the CDO is averaging less than three years. Companies must reach a point of stability and consistency in leadership and approach to maximize the return on their data investments.

#### **5. Learn from the experience of others.**

Organizing around data is a new principle for most companies, many of which have continued to evolve from a product-centric view to a customer-centric view in recent decades. Many organizations tell me that they have their data and cultural challenges all figured out, but in practice, those that think they have it all worked out rarely do. They would benefit from understanding what other companies have done so that they can avoid the same pitfalls and replicate formulas for success. I have seen organizations seesaw in hiring external CDOs as change agents and revert back to longtime company insiders who know the culture of the company. Staying the course is important. Keep an open mind, learn from the experiences of others — their failures and successes — and look beyond your four walls for inspiration and success models.

#### **6. Maintain a long-term view, and expect a journey.**

It should be clear by now that achieving data success is a journey, not a sprint. Companies desire to accelerate their efforts to become data-driven, but consistency, patience, and steadfastness pay off in the long run. Companies that set a clear course, with reasonable expectations and phased results over a period of time, get to the destination faster. Develop a plan. Create a data strategy for your company if you do not already have one. If you do have a data strategy, make sure that it is updated annually to reflect changes in the business and the ongoing and rapid evolution of emerging data management capabilities. Define your future state, and build an execution road map that will take you from your current state to the target outcome. It is hard to reach any destina-

tion without a good road map. Companies need to maintain a long-term view and stick to it while making periodic adjustments. Patience, persistence, and commitment are the ingredients for ensuring a successful long-term outcome.

#### **Change Means Finding**

#### **New Ways of Doing Things**

Organizations must evolve and change the ways in which they structure current business processes if they expect to become more data-driven. In short, companies must be prepared to think differently. In 1997, Apple launched its legendary “Think Different” advertising campaign, noting that it’s often “the misfits, the rebels, the troublemakers, the round pegs in the square holes” who “see things differently … change things … change the world.” Those companies that recognize that competing with data and analytics requires them to do business a little differently and embrace change will likely be well positioned to realize the benefits of a data-driven culture.

**Randy Bean** (@randybeannvp) is an industry thought leader and author and the CEO of NewVantage Partners, a strategic advisory and management consulting firm he founded in 2001. He is a contributor to MIT Sloan Management Review, Forbes, Harvard Business Review, and The Wall Street Journal. You can contact him at rbean@newvantage.com.

---

**Copyright** © Massachusetts Institute of Technology, 2020.

All rights reserved.

# Getting Serious About Data and Data Science

To implement successful data programs, companies need to shift goals, muster resources, and align people.

BY THOMAS C. REDMAN AND THOMAS H. DAVENPORT

**D**ata science, including analytics, big data, and artificial intelligence, is no longer a novel concept. Nor is the important foundation of high-quality data. Both have contributed to impressive business successes — particularly among digital natives — yet overall progress among established companies has been painfully slow. Not only is the failure rate high, but companies have also proved unable to leverage successes in one part of the business to reap benefits in other areas. Too often, progress depends on a single leader, and it slows dramatically or reverses when that individual departs the company. In addition, companies are not seizing the strategic potential in their data. We'd estimate that less than 5% of companies use their data and data science to gain an effective competitive edge.

Over the years, we have worked with dozens of companies on their data journeys, advising them on the approaches, techniques, and organizational changes needed to succeed with data, including quality, data science, and AI. From our perspective, these are the two biggest mistakes organizations make:

1. They underinvest in the organization (people, structure, or culture), process, and the strategic transformations needed to get on offense — in other words, to take full advantage of their data and the data analytics technologies at their disposal.
2. They address data quality improperly, which leads them to waste critical resources (time and money) dealing with mundane issues. Bad data, in turn, breeds mistrust in the data, further slowing efforts to create advantage.

Although the details at each company differ, seeing data too narrowly — as the province of IT or the data science organization, not of the entire business — is a recurring theme. This causes companies to overlook the transformative potential in data and therefore underinvest in the organizational, process, and strategic



changes cited above. Similarly, they blame technology for their quality woes and failures to capitalize on data, when the real problem is poor management.

We've all observed how companies behave when they are truly serious about something — how the goal changes from incremental progress to rapid transformation; how they muster both breadth and depth of resources; how they align and train people; how they communicate new values and new ways of working; and how senior leaders drive the effort. Indeed, it almost seems as if companies go overboard when they are truly serious about something. Amazon's Project D initiative to develop the Echo/Alexa smart speaker is a great illustration of that seriousness, with hundreds of employees, several startup acquisitions, heavy CEO involvement, and no expense spared. DBS Bank's journey to being named World's Best Digital Bank by Euromoney is another good example. The company's CEO, Piyush Gupta, said the following upon receiving that award in 2018:

*At DBS, we believe that banks tomorrow will look fundamentally different from banks today. That's why we have spent the past three years deeply immersed in the digital agenda. This has been an all-encompassing journey, whether it is changing the culture and mindsets of our people, re-architecting our technology infrastructure, or leveraging big data, biometrics, and AI to make banking simple and seamless for customers.*

The contrast with most companies' data programs is stark — one can only conclude that many are not yet serious about data and data science. For those only beginning to explore data, this may be understandable. But, if you've been at it for three years or more, it is time to either get serious in addressing mistakes or invest your resources elsewhere — and expect to lose out to competitors.

### Stop Wasting Effort on Data Quality

The obvious approach to addressing these mistakes is to identify wasted resources and reallocate them to more productive uses of data. This is no small task. While there may be budget items and people assigned to support analytics, AI, architecture, monetization, and so on, there are no budgets and people assigned to waste time and money on bad data. Rather, this is hidden away in day-in, day-out work — the salesperson who corrects errors in data received from marketing, the data scientist who spends 80% of his or her time wrangling data, the finance team that spends three-quarters of its time reconciling reports, the decision maker who doesn't believe the numbers and instructs his or her staff to validate them, and so forth. Indeed, almost all work is plagued by bad data.

The secret to wasting less time and money involves changing one's approach from the current "buyer/user beware" mentality, where everyone is left on their own to deal with bad data, to creating data correctly — at the source. This works because finding and eliminating a single root cause can prevent thousands of future errors and eliminate

the need to correct them downstream. This saves time and money — lots of it! The cost of poor data is on the order of 20% of revenue, and much of that expense can be eliminated permanently. That's more than enough to fund the needed investments.

### Get On Offense

Now consider the budgets for AI (as an example of "offense-minded" data efforts). It appears to us that, in many cases, the data science work to develop a new algorithm is funded well enough. Algorithm development is getting cheaper anyway, given that automated machine learning programs are doing more of the work. But useful algorithms die on the

vine because the work to build processes, train people, address fear of change, and adapt the culture is substantially underfunded. Based on our experience, a good rule of thumb is that you should estimate that for every \$1 you spend devel-

oping an algorithm, you must spend \$100 to deploy and support it. A few of these dollars will go to building algorithms into work processes, and many more to training, building a culture that embraces data, and change management. Most companies aren't spending this money yet, and it explains their lack of production AI deployments.

### Make Bold Moves

What tangible steps should business leaders take to demonstrate that they are serious about data? First, they should more tightly couple their business and data strategies with an eye toward driving revenue growth. From the data perspective, opportunity abounds in fully exploiting proprietary data, driving analytics into every nook and cranny of the company, and augmenting virtually every decision using AI. You cannot — and should not — do them all, so you must select those most closely aligned with your business strategies. One sign that you're on the right track is that there will be fewer data efforts. But those you do have will be far larger, more comprehensive, and more closely managed.

Second, get everyone fully engaged. After all, everyone is technically involved in your data efforts

A good rule of thumb is that you should estimate that for every \$1 you spend developing an algorithm, you must spend \$100 to deploy and support it.

already. They interpret data correctly, or they do not; they create data correctly, or not; they use data to improve their work, or not; and they contribute to larger data initiatives, or not. Today, there are far too many “nots.” Similarly, managers push back against the nots, or they do not, and more senior leaders get in front of them, or not. So you must reach out to people, educate them, and enroll them in the effort, even as you grow increasingly intolerant of the inefficiencies stemming from bad data. This is going to take some time. One sign that you’re on the right track is that morale will improve. In our experience, once people get the hang of it, most of them find data work quite enjoyable. Importantly, in the data space, talent wins.

Third, draw a clear distinction between the management of data and the management of technology. Just as a movie is a different sort of asset than streaming technology, data and tech are different sorts of assets. Each demands its own specialized management. Yet today, too many companies subordinate data to tech. The result is that topics such as data architecture do not get the attention they deserve, leading to such absurdities as a bank having 130,000 databases, not including spreadsheets. Meanwhile, technology programs spend too much time dealing with the consequences of having systems that don’t talk to one another and spending too little time introducing new technologies to employees. One sign that you’re on the right track is that technology departments will become more effective and, in time, strategic.

Finally, now is a good time to start thinking about the longer-term roles data will play in your company. It is easy enough to recite the mantra “Data is the new oil.” And according to *The Economist*, data is now worth up to \$2 trillion in the U.S. alone. But, of course, not all data is created equally. Some data — such as proprietary data, data needed to run the company, and data associated with other key assets — is so important that it should be treated as an asset in its own right. At the very least, you should make sure that end-to-end accountabilities for this data are clear.

We fully recognize how challenging these recommendations will prove to be. Yet they signal great opportunity, especially for the first companies in their sectors to embrace them. The needed approaches, methods, and technologies are widely available and have proved themselves over and over

among digital natives and at the department level for established companies. It is clear enough that the future depends on data, so sooner or later, you have no real choice. As in all things, *audentes Fortuna iuvat* — fortune favors the brave.

**Thomas C. Redman** (@thedataoc1) is president of Data Quality Solutions, a New Jersey-based consultancy, and coauthor of *The Real Work of Data Science: Turning Data Into Information, Better Decisions, and Stronger Organizations* (Wiley, 2019).

**Thomas H. Davenport** (@tdav) is the President’s Distinguished Professor of Information Technology and Management at Babson College, as well as a fellow at the MIT Initiative on the Digital Economy and a senior adviser to Deloitte’s Analytics and AI practice.

---

**Copyright** © Massachusetts Institute of Technology, 2020.

All rights reserved.



# Data Governance in the 21st-Century Organization

Good governance requires balance and adjustment. When done well, it can fuel digital innovation without compromising security.

BY GREGORY VIAL

In one of our recent research projects, a respondent told our team that “data is the food of AI. It’s what AI grows on.” This simple but powerful metaphor illustrates that generating value with data is not about having lots of data on hand; it is about using the *right* data (at the *right* time, one might argue), which explains why many organizations are still struggling to become data-driven.

At the same time, data breaches that we too frequently hear about underline the reality that having data is not without risk for organizations — as illustrated by a quick glance at GDPR rulings.<sup>1</sup> To paraphrase the Spider-Man comics, with great data comes great responsibility. So how do organizations generate value by leveraging data while avoiding the issues that stem from generating, collecting, and processing data?

To shed light on this pressing question, it’s important to discuss the relevance of data governance in data-driven organizations.

## Governance: Bridging Strategy and Operations

Broadly speaking, data governance builds on the concepts of governance found in other disciplines, such as management, accounting, and IT. Think of it as a set of

practices and guidelines that define the loci of accountability and responsibility related to data within the organization. These guidelines support the organization's business model through generating and consuming data.

A recurring question I hear from executives is whether data governance takes place at the strategic level. Although it's possible to approach it this way, it is not an ideal pattern, because it fails to translate strategy into concrete practices and guidelines. Where data governance really takes place is between strategy and the daily management of operations. Data governance should be a bridge that translates a strategic vision acknowledging the importance of data for the organization and codifying it into practices and guidelines that support operations, ensuring that products and services are delivered to customers.

By functioning as a bridge within the organizational design, data governance supports the execution of the strategy and enables innovation while providing the necessary safeguards to guarantee the security and confidentiality of information owned and/or processed by the organization. Unfortunately, in many organizations, data governance is implemented later and efforts to place it as a central piece of the organizational puzzle are not sufficient, leading to situations where it does not support the execution of the strategy and is perceived as an unnecessary nuisance by operations.

## Governing Data

Data governance initiatives often begin with a flurry of activity concerning a data governance "plan." Usually, this takes the form of a document outlining rules and procedures that put a coercive spin on data generation, collection, and use. Often, procedures focus heavily on security policies and can get quite technical. While necessary, these documents can be problematic for four reasons. First, as with many other corporate documents, employees may not read them even though they know they exist. Second, they are highly impersonal. The plurality of data means that it is very difficult to outline all possible scenarios in one document. Third, documents must be maintained and

evolve over time with the organization. Unfortunately, these changes are not always done in a timely manner. And finally, documentation can give a false impression that things are taken care of and create a sense of complacency. Things are done (or not) because "it's in the data governance plan" rather than because they continue to create value for the organization.

To address these issues, we can think about gov-

*erning* data, rather than data governance. The difference is subtle but ties back to placing governance between strategy and operations — because these activities bridge and evolve in step with both.

Literature highlights the reality that governance is not just about rules and procedures. Rather, there are three primary categories of mechanisms that complement one another and that leaders can leverage to govern data:

1. *Structural* mechanisms are the most formal and include elements such as the creation of special roles, official policies, and rules.
2. *Procedural* mechanisms are used by the organization to ensure compliance with structural mechanisms, such as data audits and reviews. (This is where IT plays an important role.)
3. *Relational* mechanisms are the least formal and include key activities such as communication and informal employee mentoring. In a large financial institution where I conducted research, for example, leaders relied heavily on the informal mentoring of junior AI developers to teach them about the ethics of using sensitive data for things such as credit ratings and loan applications.

While organizations must rely on a combination of all three mechanisms to successfully govern data, relational mechanisms are particularly important for creating a data-driven culture that serves the organization's strategic objectives.

## Governing Data Does Not Have To Be Monolithic

Organizations sometimes adopt a one-size-fits-all approach to data governance. Although this is easier to

create and maintain, it's often not ideal. For example, if your organization has different business units that use data with various levels of sensitivity, a monolithic approach based on the requirements of the most sensitive data may not meet the company's needs; other units may require additional flexibility to support digital innovation. The framework offered by Vijay Khatri and Carol Brown is a useful tool to diagnose or design an agile approach to data governance, recognizing the different needs of organizations.<sup>2</sup> It is built on five key dimensions that represent domains for management decision-making where a combination of structural, operational, and relational mechanisms can be implemented:

1. *Principles* are the foundation of the framework and ask questions related to the role of data as an asset for the organization.
2. *Quality* defines the requirements for data to be usable and the mechanisms in place to assess that those requirements are met.
3. *Metadata* defines the semantics that are crucial for interpreting and using data — for example, those found in a data catalog that data scientists use to work with large data sets hosted on a data lake.
4. *Accessibility* establishes the requirements related to gaining access to data, including security requirements and risk mitigation procedures.
5. *Life cycle* supports the production, retention, and disposal of data on the basis of organizational and/or legal requirements.

Although we can certainly apply this framework to the entire organization, the authors suggest creating a matrix where each dimension is evaluated on the basis of the location of the central point of authority and accountability. For example, an organization may need to comply with regulations related to data life cycle and keep decisions related to this domain in the hands of C-level executives relying more heavily on structural mechanisms (for example, organizational policies). Data quality decisions, on the other hand, may be deferred to business units on the basis of their own requirements.

## Good governance requires balance and adjustment, and when done well, it can fuel digital innovation without compromising security.

### Think About Services

When we think about data, we often think about bits and bytes at rest, stored in dedicated data structures such as databases or, more commonly nowadays, text files (in CSV or JSON format, for instance). Although data storage and management are important, data is often in flux, and interactions are increasingly taking place using services — programmatic interfaces that allow users to access and/or manipulate data over a network.

In data governance, services are useful because they describe the semantics of data and their methods of access, regardless of the structure and the location of the underlying data, and policies such as usage quotas can be enforced programmatically (and customized on a tier basis, for instance) directly within those services. An upside is that this makes it easier to design, scale, and customize services based

on the needs of a given business. On the flip side, it's crucial to harness services in a way that helps mitigate security risks to avoid undesired outcomes, such as a data leak to data consumers (for example, third parties).

Overall, services force us to think about governing data as a piece of software, which also means that we need to carefully consider how we evaluate the viability of those services — by ensuring that they provide only the data elements they were designed to provide. The good news is that services can easily be tested using continuous automation to ensure continued compliance with data governance practices and guidelines.

### Four Action Items to Govern Your Data

Good governance requires balance and adjustment, and when done well, it can fuel digital innovation without compromising security. Here are four simple action items to help govern data in an organization.

- **Start at the top.** To govern data, leaders need to acknowledge its strategic relevance. Taking advantage of the formulation of AI strategies or the revamping of existing business strategies in the current crisis context allows leaders to incorporate data as part of their strategy.

- **Think beyond coercion to support data-based innovation.** Governing data is often perceived as a way to rein in data within the organization. While that is important, making sure that data governance also supports innovation is equally important.
- **Design and assert data governance frequently using frameworks.** Delving into the details of daily governance mechanisms is a daunting task. Designing and asserting how data is governed within an organization using simple frameworks such as the one presented here is less tedious, more flexible, and more amenable to the way executives think about their organizations.
- **Think beyond data at rest.** Data in flux is an important area for data governance in 21st-century organizations. Although services are software, they must be designed and tested to comply with the data governance practices and guidelines of the organization.

Governing data is not easy, but it is well worth the effort. Not only does it help an organization keep up with the changing legal and ethical landscape of data production and use; it also helps safeguard a precious strategic asset while supporting digital innovation.

## REFERENCES

1. "GDPR Enforcement Tracker," CMS, accessed July 10, 2020, [www.enforcementtracker.com](http://www.enforcementtracker.com).
2. V. Khatri and C.V. Brown, "Designing Data Governance," *Communications of the ACM* 53, no. 1 (January 2010): 148-152.

**Gregory Vial** is an assistant professor of IT at HEC Montréal.

---

**Copyright** © Massachusetts Institute of Technology, 2020.  
All rights reserved.

# Actioned Analytics Pave the Way to New Customer Value

The more companies deliver actioned analytics, the more they can measure, monitor, influence, and drive real value creation.

BY BARBARA H. WIXOM AND GABRIELE PICCOLI

**T**he typical modern company is swimming in data. Whether in the form of inventory flows, manufacturing decisions, customer inquiries, or app and website clicks, there are vast quantities of data created and consumed every day. Yet managers continue to find it challenging to use this data to create new value for customers.

We blame this data paradox on the classic data-insight-action framework. The familiar process lulls managers into thinking that producing “valuable” data or surfacing insights for their customers is enough. This encourages passivity, and companies become much too trusting that customers will actually use and act upon the information provided. As a result, companies deploy analytics-based features and experiences that customers don’t use. And, because there is no action, there also is no value creation, and certainly no value capture, by the company.

Consider consumer banking. When banks initially established mobile banking apps, they focused on providing customers with data (for example, transaction histories) and insights (such as spend patterns, often in the form of a pie chart broken out into categories like rent, utilities, and food). Banks proficient at delighting customers, however, went a step further. Global bank group BBVA, for example, bundled spend insights with customer stories, tutorial videos, and recommended next steps intended to motivate customers to benefit from their information. Over time, BBVA introduced features into its mobile app that integrated predicted outflows into a customer’s digital calendar, automatically saved money toward a desired savings goal, and connected the customer with a BBVA financial adviser to execute a suggested investment opportunity.

Our research (see “The Research”) indicates



that analytics-based experiences and product features for customers pay off only when combined with aggressive tactics that all but guarantee that customer value materializes. Companies that do this well rely less on a sequential data-insight-action process and focus instead on interweaving insights and action activities to deliver *actioned analytics* to their customers.

### Three Principles for Actioned Analytics

Fortunately, data-intensive businesses — such as data aggregators, digital natives, and progressive financial institutions — have developed an array of tactics to ensure that data offerings are activated by, or for, customers. These tactics can be organized into three key principles that can help managers produce actioned analytics and create real customer value for their company to capture.

#### **1. Prioritize customers who are action-ready.**

Data-intensive businesses start by developing information offerings with or for the customers who are easiest to inspire. This means prioritizing both self-sufficient customers who are willing and able to act on insights and open-minded customers poised to respond and change.

At PepsiCo, leaders were eager to serve the company's retail partners using PepWorx, a suite of capabilities built to deliver information solutions such as tailored product assortment strategies, innovative marketing programs, and space-optimization suggestions for stores. But the leaders recognized that the success of PepWorx required retailer follow-through, so they purposely worked with retail partners who had "will and skill." According to PepsiCo's vice president of shopper analytics and insights, "We began to model ourselves like a consulting firm, helping retailers understand how to solve their operating problems and demonstrating how we could help them do that. ... It was imperative to understand the retailers who were eager and able to take action so that we could focus our efforts on them."

At times, customers are action-ready because they have been hungering for information that helps them resolve compelling problems. Cochlear, the hearing-implant provider, discovered this after it released an external sound processor that worked with an iPhone. In 2017, the company deployed insights to a mobile app to help customers optimize their hearing and better use the physical device, which is worn behind the ear. A "find my sound processor" feature

was particularly well received (and used) by parents of young patients who had long shared concerns about losing the costly physical unit.

#### **2. Bundle insights with action-inspiring activities.**

Data-intensive businesses deliver insights in consumable ways that make it easy for customers to access and use them. They also might establish other customer-centered efforts, such as advisory units, automated services, and onsite support.

Healthcare IQ is a health care spend analytics company that for decades helped hospital customers lower the cost of care by comparing a hospital's supply costs with those of peers; identifying process inefficiencies, such as off-contract orders; and finding savings opportunities, such as bulk-order discounts. Healthcare IQ leaders learned, however, that activating these insights required their customers to negotiate new pricing rates, restructure procurement contracts, and reengineer how materials were stocked and replenished — all of which are significant commitments for a hospital.

Healthcare IQ created reporting and analytics offerings that indicated to customers exactly what they should be learning from the data and how to apply new insights and lessons to generate business value. To do this, the company used data visualization, alerting and exception reporting, and hands-on support. The company established its IQ Center, a customer support unit staffed with experienced analysts and clinicians with years of hospital experience, to assist hospitals with strategic projects, special initiatives, and specialized reports. In some cases, support took the form of onsite consultants who helped the hospital's value analysis team or negotiated with vendors.

When possible, data-intensive businesses inspire action through automation — in fact, this is the ideal state of actioned analytics. Tripbam is the leading hotel-rates shopping and compliance engine for travel managers, serving half of the Fortune 100 and a sizable proportion of Fortune 500 companies. Its core offering initially suggested rebooking opportunities as they became available as a result of rate fluctuations at the booked hotel and at nearby competitors. Ultimately, Tripbam created an automatic rebooking feature, which guaranteed that a busy traveler cashed in on the lower-cost opportunities.

As a byproduct of its continuous rate shopping, Tripbam produced insights regarding the hotels' compliance with negotiated contract terms such as

discounted-rate availability. However, only Tripbam's largest corporate clients had travel units capable of follow-up. The company's CEO, Steve Reynolds, explained, "In the early days, we would show reports to our customers, and they would say, 'This is great insight, but I don't have any staff to action it.'" In response, Tripbam created alerts that automatically generate emails to hotels with a warning to comply with specific client contract terms — or face getting kicked out of the client's travel program.

### **3. Monitor action, and remediate inaction.**

Customer nonresponse can be astonishing. For Cochlear, product managers initially were surprised to learn that some hearing patients did not optimize their hearing settings even after receiving recommendation alerts on their mobile apps. Product managers discovered this during routine meetings with analytics colleagues to review and interpret processor usage data and while talking with customers about the feature during clinical trial processes. In response, developers designed a new technology, Scan, to analyze the sound environment and adjust the settings automatically for customers in order to deliver the best hearing for the conditions.

Duetto Research, a data-intensive hotel revenue management company, formed a customer success team that was accountable for ensuring that customers could extract the maximum value from using Duetto offerings. The customer success team lead explained, "Customer success is not customer support. The latter is reactive — customers open tickets and report bugs, and customer service representatives troubleshoot them. Customer success is proactive. It's our job to ensure that customers see the reason to stay with the company: They make better decisions and more money."

As these examples illustrate, data-intensive companies learn to expect customer inaction and proactively intervene with action-inspiring activities. As a result, these companies first uncover what's going on with customers, typically by tracking usage or soliciting input from customer-facing employees. Then they jump in as needed with value-sharing incentives, gamification, education, and other methods designed to motivate customers to take action.

### **It's Time to Embrace Actioned Analytics**

For Tripbam, customer retention and competitive advantage are closely related. The company monitors savings in relation to monthly subscription

costs and computes the ROI that each client receives from its service. Leaders aim to deliver customers value that equates to fourfold their costs. If a customer's ROI drops below that target — an early warning of decreasing customer loyalty and possible churn — the team proactively reaches out to the customer to remediate the problem. When a customer's ROI increases significantly above the target, leadership may evaluate a possible contract renegotiation to maintain the desired value-capture balance. As a result, the company enjoys a near-perfect customer retention rate. As Tripbam's CEO noted, "We make sure each customer has at least a fourfold ROI. Why would they churn?"

Delivering actioned analytics is now within reach of most companies, not just seasoned, data-intensive ones. And those who commit to bridging the gap between data and action for their customers stand to gain tremendously. Such a commitment positions a company to measure, influence, and adjust customer value creation — and, in effect, establishes vertical integration with the customer's data-insight-action process. This is an enviable competitive position to achieve, and one that ensures both customer loyalty and company value capture.

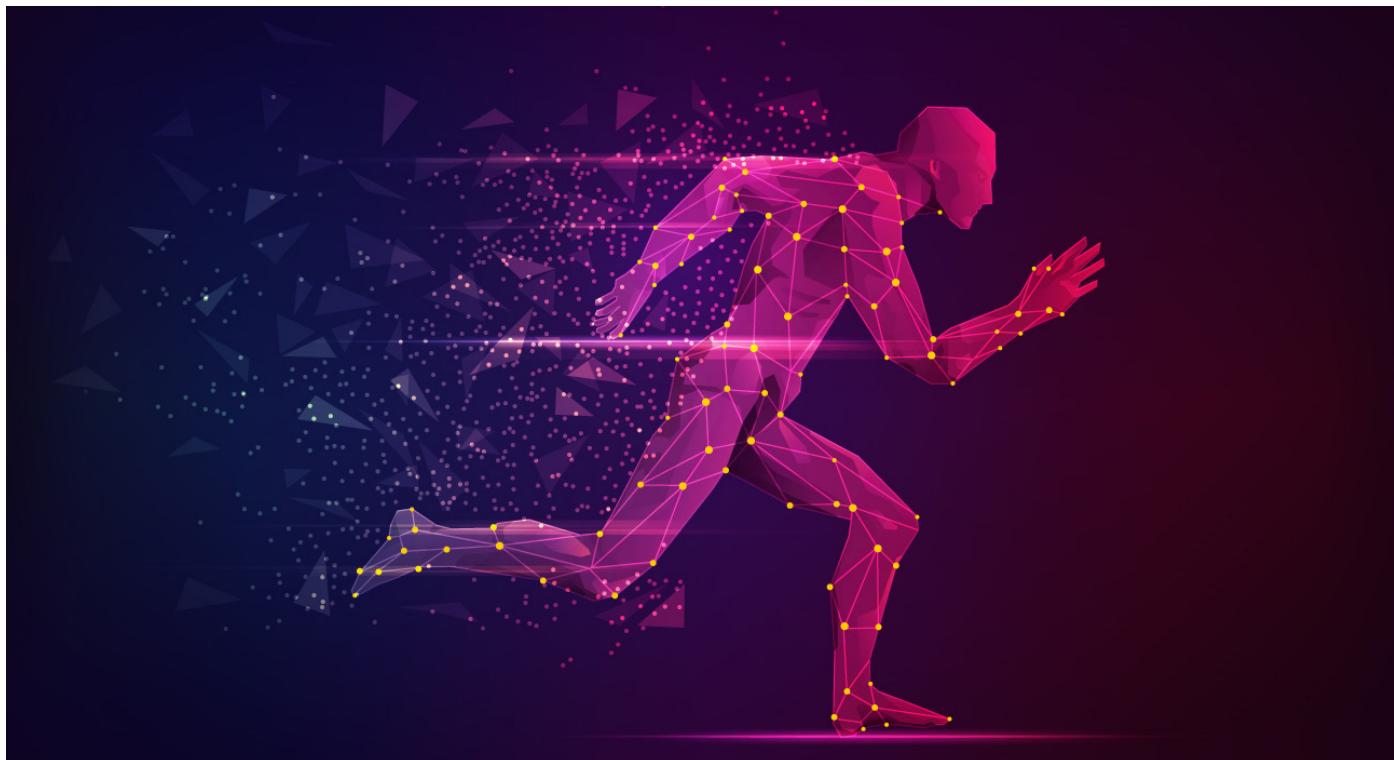
### **ABOUT THE RESEARCH**

Since 2013, the authors have been investigating data-intensive businesses to understand what it takes to generate competitive advantage and new revenue streams from data monetization.

They conducted dozens of executive and project team interviews; developed in-depth data monetization cases about companies, including Tripbam, Duetto, BBVA, Healthcare IQ, and PepsiCo; and surveyed hundreds of product managers and executives regarding data monetization activities and outcomes.

**Barbara H. Wixom** (@barbwixom) is a principal research scientist at the MIT Center for Information Systems Research in Cambridge, Massachusetts. **Gabriele Piccoli** is the Edward G. Schleider Chair of Information Sciences at Louisiana State University and an associate professor of information sciences at the University of Pavia.

**Copyright** © Massachusetts Institute of Technology, 2020.  
All rights reserved.



# How Organizations Can Build Analytics Agility

Companies must increase their analytical fitness and develop strong muscle memory when they are tested by disruptive events.

BY LORI C. BIEDA

In an era of constant change, companies' data and analytics capabilities must rapidly adapt to ensure that the business survives, never mind competes. Organizations seek insights from their data to inform strategic priorities in real time, yet much of the historical data and modeling formerly applied to predict future behavior and guide actions are proving to be far less predictive, or even irrelevant, in our current normal with COVID-19.

In order to survive through crises, proactively detect trends, and respond to new challenges, companies need to develop greater analytical agility. This agility comes from three areas: improving the quality and connections of the data itself, augmenting analytical "horsepower" at the organization level, and leveraging talent that is capable of bridging business needs with analytics to find opportunity in the data.

## The Answers Are in the Data

The quest for better data is not new, but the cost of *not* having it is easier to substantiate and understand in a time of crisis. Gaps in data quality — whether it's time-lagged, disconnected, insufficient in granularity, or poorly curated (rendering analysis slow or impossible) — become intolerable amid chaos when companies must act quickly. Crises can be opportunities to augment data quality and further enrich the data to better serve customers and the company.

Making the business case for data investments suddenly makes sense as business leaders live through data gap implications in real time. Monetizing data typically comes from four sources:

- Connecting data with other data differently than before.
- Getting new data sources, or more specific levels of the same data you already had.
- Putting data to better or faster use than the competition.
- Getting data faster.

Data and analytics leaders must frame investments in the current context and prioritize data investments wisely by taking a complete view of what is happening to the business across a number of functions. For example, customers bank very differently in a time of crisis, and this requires banks to change how they operate in order to accommodate them. The COVID-19 pandemic forced banks to take another look at the multiple channels their customers traverse — branches, mobile, online banking, ATMs — and how their comfort levels with each shifted. How customers bank, and what journeys they engage in at what times and in what sequence, are all highly relevant to helping them achieve their financial goals. The rapid collection and analysis of data from across channels, paired with key economic factors, provided context that allowed banks to better serve customers in the moment. New and different sources of information — be it transaction-level data, payment behaviors, or real-time credit bureau information — can help

ensure that customer credit is protected and that fraudulent activity is kept at bay.

Every data and analytics team has a roster of data demands that outpace budgets. Foundational data investments often languish and are perpetually underfunded because their value is difficult to isolate and describe to others. Events such as COVID-19 pinpoint investments that are likely already on the list but lack sufficient organizational buy-in to propel them forward. As the adage goes, never let a good crisis go to waste — use it to enrich your data and customer understanding.

## Augment Analytical Horsepower With Business Parameters

Experienced analysts recognize that all analysis requires a blend of art and science. However, the nature of crises and unusual events means that analysts simply do not have the standard observation windows on which to build their typical projections and formulate baselines. History is no longer as useful. Moreover, rapidly changing market conditions require teams to constantly readjust models and analytical approaches to stay current. This requires adequate *datasense* (a blend of analytical facts and business intuition) in order to set the right parameters and optimize the business.

We often think of business rules as the antithesis of data-driven decisions or models, because they introduce a layer of subjectivity into a world that thrives on objectivity. In an unrecognizable crisis, however, establishing these parameters is important for a business to function practically and make basic analytical decisions. In our banking example, consider the customer conversations happening within a branch network throughout COVID-19. Historically, there has been a set capacity available to handle and fulfill requests, and the nature of conversations — whether financial check-ins, mortgage renewals, or efforts to ensure that customers have the right financial solutions to meet their immediate needs based on emerging life events — has been known and model-driven.

In the current reality, that constraint-based optimization problem is turned on its head because

**Making the business case for data investments suddenly makes sense as business leaders live through data gap implications in real time.**

customer needs are vastly different, staff availability is in flux, branch hours are changing, and orchestrating effective customer conversations requires new analytics. Augmenting the process with business rules helps define the parameters of what can be done. Next, the whole approach to customer conversation optimization needs to be recalibrated (across branches and every channel) based on circumstances that are changing daily because what you are optimizing *for* has changed. And although machine learning and automation can help, the vast aberration in the data from the pandemic means that it will take time for such solutions to adapt and become relevant.

This reveals a need for a different and augmented approach to analytics that suits the time. It necessitates a greater blend of “art and science,” of “(wo)man and machine,” and of business rules and models in order to navigate gray areas.

### **Business-Analytics Hybrids**

Navigating uncertainty and responding to change requires an exceptional translation layer — a team of individuals whose skill sets blend a superior understanding of the business with the technical acumen to transform data into insight. The world is *always* hungry for these skills, and in the future it will make the difference between brands that are nimble and thrive and those that languish.

The COVID-19 pandemic has been characterized by rapid unpredictability and never-before-seen trends. For businesses to adapt and make key strategic changes in short order, they need teams with hybrid skills, capable of both finding opportunity in data and executing quickly and accurately when the business knows what it wants to do.

In a crisis, data and analytics can overwhelm leaders and prevent them from acting quickly on account of the sheer volume, pace, and continuous shifting of data. In order to break down emergent trends and properly contextualize them, organizational functions must come together in new ways. While few businesses ever anticipated needing to transform into operating as fully remote workforces, many are seeing that the intense collaboration and connectedness of their people have formed strong virtual networks that are at the heart of their survival.

### **A Culture of Analytics and Business Collaboration**

When teams come together to interrogate new, shifting data from multiple perspectives, they begin to gain comfort in establishing more “knowns” in an unknown world. The result is that leaders can make the best possible decisions in the most rapid time frame. Done well, this will help companies thrive in disruption and gain competitive advantage, but it requires a high level of analytical literacy throughout the business and, most important, a culture of collaboration.

Facilitating these ongoing exchanges can happen in a variety of forums, ranging from the formal (such as customer optimization forums and risk exchanges) to the informal (such as real-time dashboards and views of computations on the fly). It’s important that this collaboration is continuous, interactive, and inclusive, with both business and analytical teams present so that the data is properly interpreted and all stakeholders understand any actions that are required.

In order to detect and respond to disruptive events with agility, companies must increase their analytical fitness and develop strong muscle memory when they are put to the test. Navigating the pandemic often feels like running a marathon at sprint speed into the dark, and an agile approach to data and analytics will be the headlamp that companies cannot do without.

**Lori C. Bieda** is an analytics, technology, and marketing thought leader with more than 20 years of experience driving profitable business growth through the strategic use of data and analytics. She is chief data and analytics officer, North American Personal and Business Banking, at BMO Financial Group.

---

**Copyright** © Massachusetts Institute of Technology, 2020.  
All rights reserved

# The Enterprise Systems That Companies Need to Create

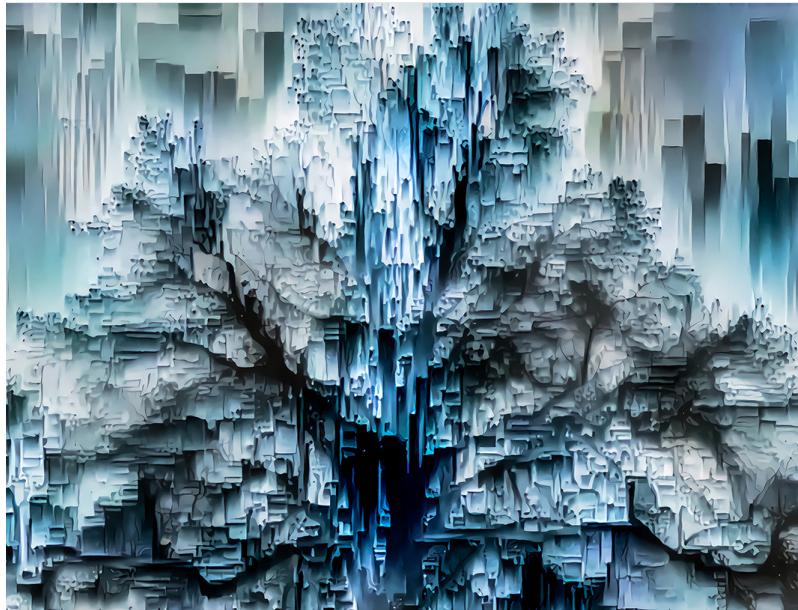
Companies should focus on creating enterprise systems that allow them to make fundamentally better business decisions.

BY DAVID WALLER AND PAUL BESWICK

**T**he swiftness of technology's progress in the past decade has convinced legions of companies that their survival depends on jettisoning their legacy systems as soon as budgets permit such an overhaul. Computing power has surged, storage costs have plummeted, and networking speeds have approached theoretical limits. All the while, companies and consumers are generating ever-growing floods of data packed with clues on how individuals behave and how products perform. Many companies thus risk upgrading technology purely for its own sake. In doing so, they overlook what may be the greatest opportunity presented by the modern technology stack: the chance to mobilize new tools in a way that empowers managers and technologists alike to make fundamentally better business decisions.

To illustrate, consider the curiously old-fashioned approach companies typically take to upgrading their legacy systems. It starts when something old stops working. Perhaps an aging mainframe fails often, resulting in seemingly never-ending maintenance costs, or an outage destroys transaction data. A decade ago, the natural response was to check for an updated version of yesterday's software that could stamp out the bugs — particularly if paired with newer hardware. Now, companies look to the cloud for the latest collection of computing services, storage technologies, and performance guarantees.

Both scenarios share the same mentality and prevailing aim: to replicate yesterday's functionality at today's prices. "I want what I have today, only faster, or cheaper, or simpler." The habit of colossal, periodic technology projects persists, justified by often-strained



business cases that hinge on cost improvements or risk reductions spread across a wide swath of deeply entrenched systems.

There is a better way. Tech upgrades can be revenue generators, not just cost sinks, and they need not saddle you with soon-to-be legacy burdens. Our experience suggests that three strategies can position companies to carry out technology transformations that can create value and enable continuous innovation.

**Redefine success.** Companies that reap the greatest rewards from technical improvements recognize that it's not only technology that changes: It's also their leaders' minds, priorities, and circum-

stances. Legacy systems aren't bad because they're outmoded — they're bad because they're almost invariably hard to deprecate.

To skillfully keep pace with technology, companies therefore need to develop what we call *second derivative thinking*: They must work to increase the rate of change of *change*. To build systems that improve the velocity of change in practice, companies need to identify the structural impediments that act as brakes on their ability to deliver technically, and then insist that each change project aims to remove at least one of those obstacles. In addition to achieving the project's immediate goals, the effort also clears roadblocks that would otherwise bog down future efforts. With fewer impediments, subsequent projects automatically accelerate. And because those individual solutions are delivered in concert with existing work, teams don't need to contrive laborious business cases to address them in isolation.

Take the example of a bank that wanted to grow by expanding its geographic footprint. Many aspects of banking vary from one country to the next — regulations and requirements, consumer habits and preferences. Idiosyncrasies aside, however, much remains the same about the core proposition and the span of products and services; in every nation, people save, spend, and borrow. Rather than inventing new infrastructure to conform to the nuances of a given location or region, the bank instead sought to leverage a common systems core where possible and then engineered custom services that could be "swapped in" to meet the particular needs of any given area.

Because the bank set the goal of being able to rapidly deploy operations in a new geography, it had no choice but to engineer the ability to rapidly adapt its systems. If a new regulation in, say, Singapore alters local banks' identity verification requirements, the bank can update that single, isolated service rather than try to retool core banking applications in their entirety.

Of course, not every company has the time or wherewithal to build cloud-native applications from scratch. Many sit in a technical limbo, unable to reach for state-of-the-art tech but unwilling to hobble along with turn-of-the-millennium tools. To resolve that dilemma, companies often try to build data lakes as a stepping stone to more comprehensive system upgrades. They reason that if they can systematically

hoover up the information stored in those source systems and grant more widespread access to it, they can create more modern applications while letting older systems lumber along in the background. This tactic can work, but it comes with a cost.

To pipe data out of legacy systems reliably, you have to build and maintain reliable pipelines. With a handful of systems and sources, this isn't hard. But multiply that across the vast array of systems inside most large organizations and you get what has been called *pipeline jungles* — thickets of expensive data-integration jobs that no one really owns. Anyone who's ever written or read a service-level agreement knows that a collection of individually strong components can produce a brittle system.

A novel solution to this problem is emerging in the form of data virtualization: a logical data layer that unifies data siloed in disparate systems without actually physically integrating it. At first, the idea sounds a bit silly. Don't *actually* pull all of your data together into a communal, capacious tank. Instead, use technology that lets you *pretend* that you did. Rather than yanking data out of systems, you reach into them and fetch information when you need to use it.

Accessing data *in situ*, rather than creating infrastructure to shuffle it around, offers a few benefits. You'll reduce unneeded copying and the wasteful expense of storing duplicate data. You'll use tools that give you a single route to reach upstream data, avoiding pipeline jungles to wade through. And by combining these tools with gateways and intelligently constructed interfaces, you can implement detailed permissions and security protocols far more easily.

#### **Orient technology around decision-making.**

To create value from investing in technology, companies need to be clear on where the value lies in the first place. Companies can't function without *any* technology, but surely the goal of better systems should be to function more effectively. In business, this imperative boils down to the goal of making better decisions. Generally, businesses don't make money by chance but by choice. Behind every value-creating action, there is a long string of decisions about how, when, and by whom related tasks need to get done. For instance, long before a bank collects payments on a loan, it first needs to identify target customers, devise ways to solicit them, estimate

their creditworthiness, create the product, set a rate, originate the loan, and ensure that the systems are ready to service it.

In general, companies should strive to make high-stakes decisions *more effective* and low-stakes ones *more efficient*. One national retailer noticed that when its senior-most executives gathered for quarterly planning meetings, they spent nearly all of their time scrutinizing historical sales reports and almost none of it making strategic choices. Why? Culture played a part — dwelling on anomalies and nitpicking was a well-known habit. (“Retail is detail,” as many say.) But there was a deeper, technological failing too.

Over time, the company had steadily improved its reporting capabilities to let users see fresher, more granular information. In contrast, it hadn’t invested in technologies to make predictions or to model scenarios. Users could pinpoint the store that sold the most blueberry yogurt yesterday. But ask them which flavors customers would buy if the blueberry were removed, and they’d have no idea. Ask them whether they should invest to build more stores or to reduce prices, and you’d hear a similar silence.

The executives hatched a bold plan to reorient themselves. They resolved to create what they called “living, 18-month plans” that codified their strategic choices — for instance, levels of discounting or footprint expansion — as well as robust forecasts of expected performance. During quarterly meetings, leaders updated the plan. What’s more, they insisted on the ability to quantify the business impacts of varying those choices as interactively as they could. The company needed to redirect investments away from reporting and toward modeling — fewer dashboards and more decision-support tools. But tailoring their technology to better suit the company’s key decisions meant that the leadership team could focus its collective attention on the choices that mattered most.

Enhancing low-stakes decisions can generate sizable profits too, if done frequently or on a large scale. A grocery store chain spent years agonizing over an elaborate, item-level pricing system. To

work, it needed lots of data. Some inputs should have been simple to obtain, like statistics on past prices and sales, but analysts were stymied by having to collate data from different legacy systems. Other inputs, like competitive price data, were costly to buy and error-prone. To function, the system needed to build, store, and update tens of thousands of narrow statistical models, each one of which could fail from an input glitch or return nonsensical results because of an outlier. It was, in other words, a mess.

Dismayed by what he saw, the CEO asked a provocative question: Was “the juice worth the squeeze”? In other words, what if, instead of updating the complicated system, the retailer simply stopped

setting item prices itself?

Instead of deploying large teams to chase down tiny details, the retailer could ask small teams of buyers and category managers to apply the same margin to all items in a category, effectively shifting the burden of setting product pricing back to the vendors.

If a brand wished to undercut its rivals, it could do so by lowering its cost to the retailer — but that dynamic would require zero effort from the retailer. The retailer, in turn, could adjust its overall level of price competitiveness by altering margin targets across categories.

Though the idea struck some as heresy, it gained currency as its assumptions and implications became clear. It *might* be possible to measure differences in price sensitivities between two similar cans of beans, as pricing folklore suggests, but the evidence is weak. Hence, this kind of decision is likely to be a low-stakes one where the best way to win is to be more efficient. Automating such decisions often makes them more efficient; offloading them entirely always does so.

**Reduce your delivery “chunk size.”** One final strategy that companies can use to good effect when revamping enterprise systems is to deliver smaller, self-contained, complete units of work. As a thought experiment, think of how much you’re willing to spend on your next tech transformation and how long you expect it to take. Divide both numbers by 50. You should aim to organize your work so that if you spend that diminished sum over that brief time interval, you get an entirely functional systems com-

In general, companies should strive to make high-stakes decisions *more effective* and low-stakes ones *more efficient*.

ponent, like a Lego brick that's yours to build with. Maybe it's a service encapsulated in a container, or a set of interfaces for accessing data programmatically — but whatever you build, you want it to be feature-complete and immediately reusable.

Operating in this way embeds the kind of second derivative thinking described above. Your thinking may evolve substantially as you upgrade your systems, but working in small chunks makes pivoting far easier. You have room to change your mind.

Furthermore, teams can use and derive value from individual components as they come online, rather than waiting for the entire edifice to be built. And giving teams new technical possibilities is a great way to unleash their creativity and empower them to solve problems you may not have considered.

The experience of a large commercial insurer provides an example of how to put these ideas into practice. Like many of its peers, this company felt both the drag of legacy systems and the fear of taking them off life support. Rather than rebuilding them in one go, the company chose to isolate its internal systems by putting legacy applications in a distinct service tier. Then it built a separate services layer on top of those applications. Those services provided users with access to the data and functions of the underlying legacy systems while abstracting away the need to interface with them directly. Legacy systems were in effect shrink-wrapped for a longer shelf life, and more legacy systems could be moved into the first, internal tier one by one.

To move a system behind this services layer, teams needed to think carefully about which of its data elements and functions were critical to the company's operations. Not only did the exercise allow them to incrementally disentangle their tangled legacy setup; it also gave them a road map for progressively building more modern applications.

### **Leave an Agile Legacy**

You don't have to be a technical expert to be astonished by the increasing potential for technologies to transform organizations. Fifty-one years ago, the United States piloted a spaceship to the surface of the moon using a 70-pound computer that could perform 14,245 calculations per second. In September of this year, Nvidia introduced a graphics processor that is more than 2.5 billion times as fast, weighing little more than a book.

Currently, high-capacity hard drives can store data at a cost of just over \$15 per terabyte. When we first walked on the moon, storing that terabyte would have cost \$1.7 billion in today's dollars. This August, researchers pushed data through a single fiber-optic cable at a rate of 178 terabits per second — enough to transfer about 1,500 4K movies in the time it takes to say "one, Mississippi." When the Apollo 11 astronauts splashed back down to Earth, the first computer-to-computer link hadn't yet been invented. (It would come three months later, with Arpanet.)

Given the giant leaps forward that technology continually makes, it's now possible for companies to replace their legacy systems and aging computing platforms with systems that will enable them to realize breakthroughs in terms of greater efficiencies, but also with greater product and service offerings. The key is to recognize how your legacy systems must adapt in a world where accelerating change is the only constant.

**David Waller** is head of data science and engineering at Oliver Wyman Digital. **Paul Beswick** is deputy CIO at Marsh & McLennan.

---

**Copyright** © Massachusetts Institute of Technology, 2020.  
All rights reserved.



**Rahul Pathak**, general manager for analytics at AWS, is responsible for Amazon Athena, Amazon Elasticsearch, Amazon EMR, AWS Glue, AWS Lake Formation, and Amazon Redshift. Previously, he was AWS general manager for emerging databases and blockchain, AWS general manager for several managed database and analytics services, and the product manager who oversaw the launch of the Redshift cloud data warehouse service. He has cofounded companies focused on digital media analytics and IP geolocation. He received a bachelor's degree in electrical engineering and computer science from MIT and an executive MBA from the University of Washington.

**aws** data and analytics

# Becoming a Data-Driven Enterprise: Meeting the Challenges, Changing the Culture

In this Q&A, Rahul Pathak, general manager for analytics at Amazon Web Services (AWS), describes why data-driven enterprises matter, explains how companies can build strong data foundations, discusses technical and organizational challenges, and offers forward-looking advice.

*This conversation has been edited for clarity, length, and editorial style.*

**Q: How do you define a data-driven enterprise, and why is it important?**

**Pathak:** Data-driven enterprises base their strategic decisions on data. This requires thinking about data as an asset, figuring out how to gather it from across the enterprise, bringing it together, and analyzing and using it to inform decision-making. These decisions can be about driving efficiencies or reducing costs, but also about finding new opportunities and new areas to expand. So it's not just a technical challenge; it requires companies to change how they operate and think.

Companies are looking for ways to get early signals about how to respond and react. Being data-driven helps them make decisions based on what they're actually able to measure and see, rather than acting on gut feelings.

For instance: How can you maximize your ability to interact with your customers? How can you retain them? How do you reduce churn? How do you think about what else to offer them? How can you find more customers like them? That all comes back to how efficiently you use data about customers and about how the organization is operating.

**Q: Flexibility and agility seem to be more important than ever now. Why is that the case?**

**Pathak:** It's impossible to plan for every decision that companies face. In just the last few months, we've seen so much volatility in the environment, and companies need to be able to adapt quickly.

## SPONSOR'S VIEWPOINT

Again, being data-driven enables you to start getting all the signals about what's happening, and then you can put yourself in a position to respond. Once you've got the ability to understand what's happening in your business and the ability to act on it, you can start to make good decisions. The faster you can adapt to the changing environment, the more successful you'll be at navigating it.

You need to trust that your *data foundation* – your combined cloud-based technology infrastructure, your culture, and your processes – will support any manner of future change. Data can inspire confidence during unexpected times, although it requires meaningful data and the right type of data foundation.

### Q: What does a data foundation require in terms of both technology and policy?

**Pathak:** A strong data foundation starts with low-cost, reliable, highly durable data storage that's scalable. It's not about hoarding data in silos; the exponential value comes from combing and using data over and over at scale, quickly. To gain the agility needed to respond to the unexpected, you need a data foundation that allows you to capture, clean, and curate data from across the business to be shared and reused for making everyday decisions.

Then you need to think about security and governance. We're finding that our customers want to give data access broadly to analysts and users in their organizations – but in a controlled way. So you have to define your governance policies, your security, who's allowed to see what. Once you've got those guardrails in place, then you're able to let people experiment and innovate using that data, secure in the knowledge that you've set the foundation for ensuring that data is used in ways consistent with what you're trying to do from a data policy perspective.

### Q: How should organizations interested in becoming data-driven enterprises be thinking about their infrastructure?

**Pathak:** The first step is to move away from antiquated, monolithic apps that run on one-size-fits-all relational databases toward highly distributed microservice-based systems running on multiple purpose-built databases. We've evolved to a world where it's no longer one-size-fits-all. You want to be able to use a service that's mapped to the workload in question.

The other part is thinking about moving to managed services, where you allow someone like AWS to take care of the undifferentiated heavy lifting – the provisioning, patching, scaling, and securing of the systems that you're using – and you focus instead on working with [application programming interfaces] and what differentiates your business. [For a real-life example, see "From Struggling to Seamless: Ensuring a 'Flawless' Global Education Event."]

You need to think about similar activities for analytics and how you get your data into one place using managed systems and a mix of open-source or specific purpose-built systems.

We also see a lot of modernization of data warehouses. Customers typically have been running expensive, constrained data-warehouse appliances on premises. Now they're moving to more decoupled models in the cloud, where you've got a data lake integrated with a data warehouse, which we call a *lakehouse architecture*. Customers are modernizing and moving to services like Amazon Redshift, which gives you performance at scale at a radically different cost point than customers had before.

This also involves building modern applications using those purpose-built databases. Once you've got all your data operating in the right set of services, you want to derive insight from it. So it's also about breaking down silos within the organization so that you can build a complete picture of your data, then providing democratized access to users who can do something with it. By providing controlled access to a broader set of people in the company, you're able to innovate and experiment faster, figure out what's working and what isn't, and double down on the areas that are successful.

### Q: What role does organizational culture play here?

**Pathak:** When companies want to become more data-driven, they've also got to teach people in the company to become more data-driven. That requires reskilling and retraining. But it also requires the ability to put data in everybody's hands so that the whole organization can experiment and interact with data.

So this is about helping to accelerate learning and building data literacy into the organization. That starts with finding and elim-

inating silos, but also by establishing a culture where folks think about data as a strategic asset and think of what they want to measure when they put new initiatives in place, and where that data needs to end up, and how it needs to be acted upon. Accomplishing that requires thinking about questions like “What’s the data to back

**“Leaders need to set the tone that data is a tool, an asset they can use to make better decisions, and then reinforce that message at all levels of the organization. That’s how culture can help transform companies in terms of how they use data.”**

this up?” and “How can we use this data to make a better decision?” and making that part of everyday conversations within the company. Essentially, leaders need to set the tone that data is a tool, an asset they can use to make better decisions, and then reinforce that message at all levels of the organization. That’s how culture can help transform companies in terms of how they use data.

#### **Q: What challenges do companies face in transitioning to data-driven enterprises?**

**Pathak:** Typically, the biggest challenge is inertia and processes that have been ingrained in the company over time. It starts with moving away from the mindset of “This is how we’ve always done it.”

A simple example is in the legacy world, when we had really expensive systems; we would make decisions about what data to keep and what to throw away. We found that the vast majority of data that companies had was discarded because it was too expensive to put it into a system and think about how to analyze it when it wasn’t precisely clear what it would do for you.

But when you transition to a modern cloud-based architecture, we’ve really tried to remove that cost constraint. And it turns out that much of that discarded data might actually be valuable. So now the default is “Let’s just store everything because we might not know what we want to do with the data or what insights we might get from it down the road.” But going that route involves a mindset shift. It requires a company to set up those guardrails

## **From Struggling to Seamless: Ensuring a ‘Flawless’ Global Education Event**

**Code.org**’s signature initiative is known as the Hour of Code, but that’s a deceptively simple name. After all, the annual event involves millions of students and teachers worldwide engaging with hundreds of computer science tutorials offered in more than 45 languages.

For many students, the Hour of Code provides first-time access to computer science – a potentially life-changing experience. But for Code.org, a Seattle-based nonprofit dedicated to expanding access to computer science education, the week-long event also creates an enormous technology challenge.

“During the school year, our online platform is already a very high-usage learning management system, executing about 1,000 database write queries per second,” says Will Jordan, Code.org’s lead infrastructure engineer. Usage surges by two to three times that amount during the Hour of Code. As Jordan notes: “That’s an extraordinary workload.”

And as more people participated in the Hour of Code each year, that workload grew as well. By 2018, the database was clearly reaching its performance limits, and event organizers struggled to provide consistently reliable service for that year’s event, disabling some features to reduce database workload. Ultimately, Code.org migrated its MySQL databases to Amazon Aurora, a fully managed database service.

The result? “In December 2019, we had an absolutely flawless Hour of Code,” Jordan recalls. “We offered the full range of functionality on our learning platform to all students around the world for the entire week.” The 2019 event saw 100% uptime, 50% fewer support tickets than were created in 2018, and 10 times better database write latency than Code.org’s previous solution.

Today, about 30% of all U.S. students have Code.org accounts, and the organization is working to expand its reach internationally. “Thanks to AWS, the performance of our systems for students around the world is very high,” Jordan notes. “They get a seamless experience, no matter where they are.”

**"We recommend that companies think about modernizing, that they think about data strategically, and that they set up to future-proof themselves."**

we talked about and then let people loose with the data, train them to do something useful with it, give them the tools that they need to iterate at their own pace, and have the ability to adapt.

As companies move from legacy systems to more modern ones, they're able to take advantage of the best tooling that's in place at any time without having to do expensive migrations or re-platforming. That allows them to do new things with data that they couldn't do before.

A simple example: You could use machine learning services to transcribe audio and then start to run sentiment analysis from that audio. So now what used to be a recording is actually data that you can start to run queries on. For instance, "How many of the calls that happened in the last week were positive versus negative, and what does the trend look like?" That's enabled by new technology that any customer on AWS can leverage, even without prior machine learning expertise.

**Q: What else should companies know about becoming more data-driven?**

**Pathak:** We recommend that companies think about modernizing, that they think about data strategically, and that they set up to future-proof themselves. That means thinking about embracing open data formats and data lakes, thinking about lakehouse architectures so they get the best of data warehousing and data lakes, and thinking about how to focus on what differentiates them rather than focusing on the classic undifferentiated IT. These things together put you in a position to iterate rapidly, to learn rapidly, as an organization. The faster you're able to learn, the more successful you're going to be, especially as the environment continues to be dynamic.

---

## ABOUT AWS

For 14 years, Amazon Web Services has been the world's most comprehensive and broadly adopted cloud platform. AWS offers more than 175 fully featured services for compute, storage, databases, networking, analytics, robotics, machine learning and artificial intelligence, internet of things, mobile, security, hybrid, virtual and augmented reality, media, and application development, deployment, and management. Services are offered from 77 Availability Zones (AZs) within 24 geographic regions, with announced plans for nine more AZs and three more AWS Regions in Indonesia, Japan, and Spain. Millions of customers — including the fastest-growing startups, largest enterprises, and leading government agencies — trust AWS to power their infrastructure, become more agile, and lower costs.

Learn more at [aws.com/big-data](http://aws.com/big-data) and [aws.com/databases](http://aws.com/databases).



### **PDFs • Reprints • Permission to Copy • Back Issues**

Articles published in *MIT Sloan Management Review* are copyrighted by the Massachusetts Institute of Technology unless otherwise specified.

*MIT Sloan Management Review* articles, permissions, and back issues can be purchased on our website, [shop.sloanreview.mit.edu](http://shop.sloanreview.mit.edu), or you may order through our Business Service Center (9 a.m. - 5 p.m. ET) at the phone number listed below.

Reproducing or distributing one or more *MIT Sloan Management Review* articles **requires written permission.**

To request permission, use our website [shop.sloanreview.mit.edu/store/faq](http://shop.sloanreview.mit.edu/store/faq), email [smr-help@mit.edu](mailto:smr-help@mit.edu), or call 617-253-7170.