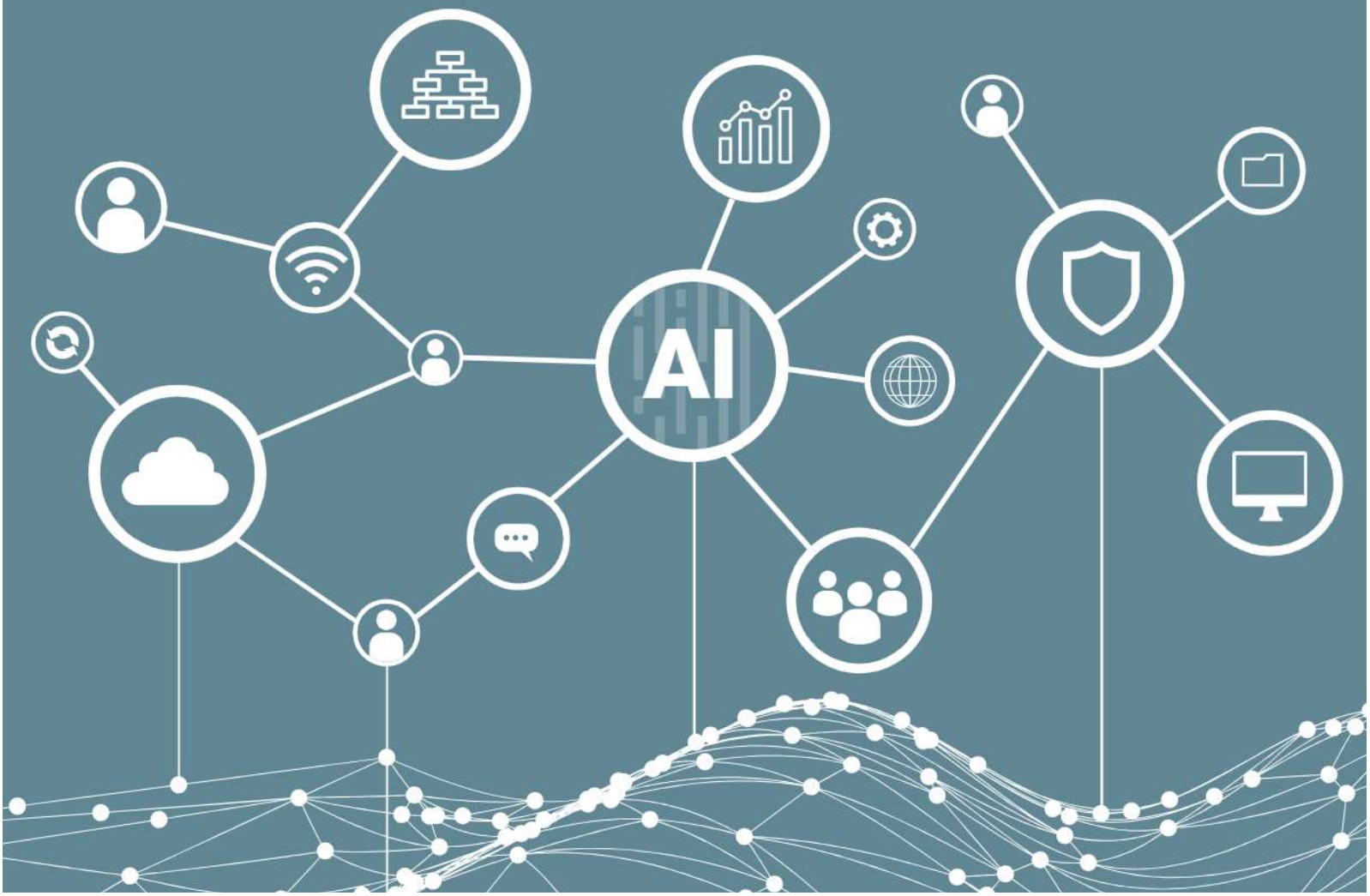
MIT Technology Review Insights

Produced in partnership with

databricks

How data and analytics leaders are delivering business results with cloud data and AI platforms

# Building a high-performance data and AI organization



# Preface

"Building a high-performance data and AI organization" is an MIT Technology Review Insights report sponsored by Databricks. To produce this report, MIT Technology Review Insights conducted a global survey of 351 chief data officers, chief analytics officers, chief information officers, and other senior technology executives. The respondents are evenly distributed among North America, Europe, and Asia-Pacific. There are 14 sectors represented in the sample and all respondents work in organizations earning $1 billion or more in annual revenue. The research also included a series of interviews with executives who have responsibility for their organizations' data management, analytics, and related infrastructure. Denis McCauley was the author of the report, Francesca Fanshawe was the editor, and Nicola Crepaldi was the producer. The research is editorially independent, and the views expressed are those of MIT Technology Review Insights.

We would like to thank the following individuals for providing their time and insights:

**Patrick Baginski,** Senior Director Data Science, McDonald's (United States)

**Bob Darin,** Chief Data Officer, CVS Health, and Chief Analytics Officer, CVS Pharmacy (United States)

**Naveen Jayaraman**, Vice President – Data, CRM & Analytics, L'Oréal (United States)

**Michel Lutz**, Group Chief Data Officer, Total (France)

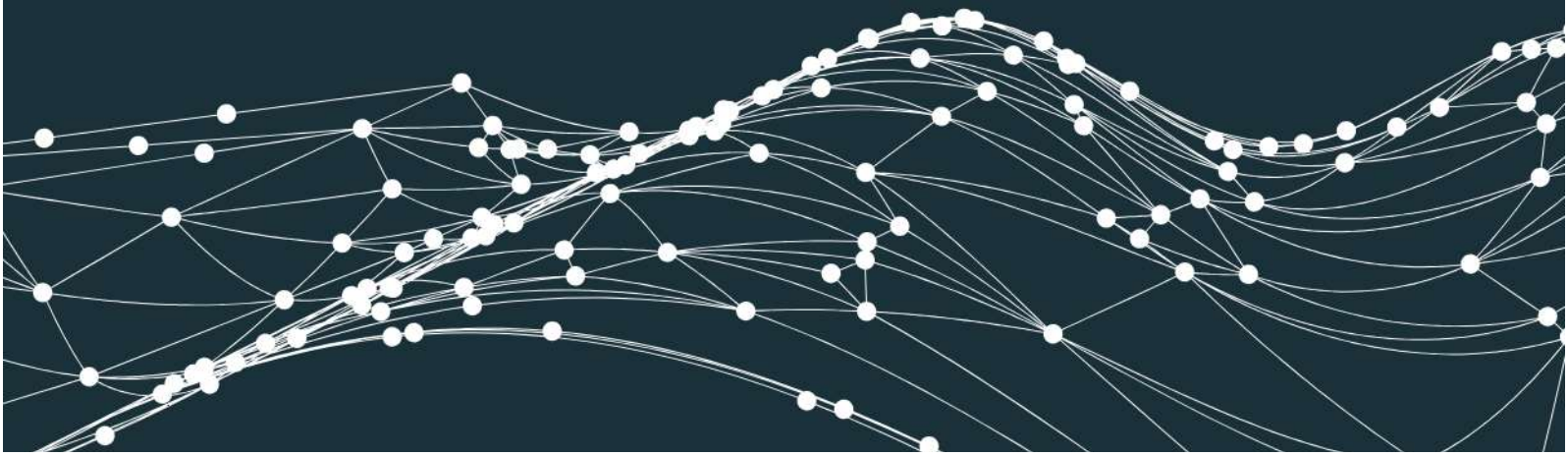
**Mainak Mazumdar,** Chief Data and Research Officer, Nielsen (United States)

**Andy McQuarrie,** Chief Technology Officer, Hivery (Australia)

**Sol Rashidi,** Chief Analytics Officer, The Estée Lauder Companies (United States)

**Ashwin Sinha,** Chief Data and Analytics Officer, Macquarie Bank (Australia)

**Don Vu,** Chief Data Officer, Northwestern Mutual (United States)



# CONTENTS

01 **Executive summary** ..... 4

02 **Growth and complexity** ..... 6

Databricks perspective: The rise of the lakehouse effect ..... 7

03 **Aligning and delivering on strategy** ..... 9

Data high-achievers ..... 11

Nielsen: data transformation for a data-reliant business ..... 13

04 **Scaling analytics and machine learning** ..... 14

A paradigm shift at CVS Health ..... 15

Barriers to scale ..... 16

Protecting return on investment ..... 17

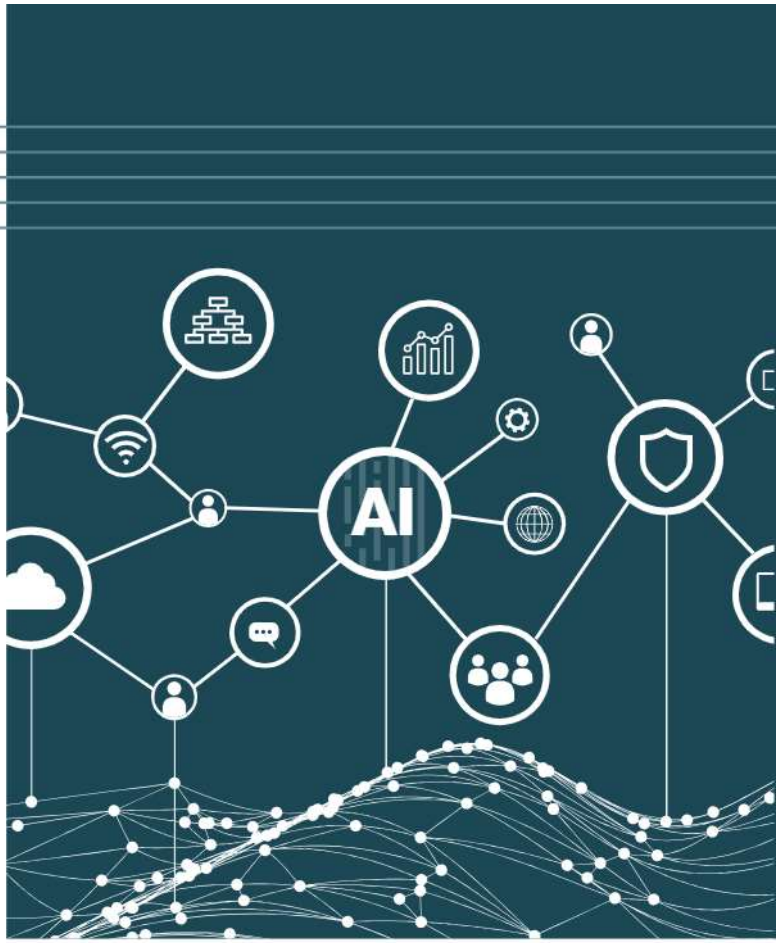
Technology, democracy, and culture ..... 18

05 **Visions of the future** ..... 19

A CDO wish-list for a new architecture ..... 19

06 **Conclusion** ..... 21

# 01 Executive summary

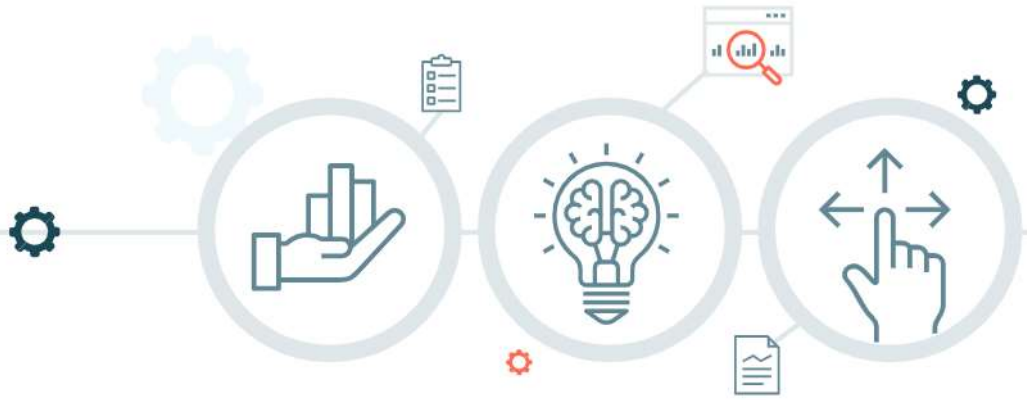



CxOs and boards recognize that their organization's ability to generate actionable insights from data, often in real-time, is of the highest strategic importance. If there were any doubts on this score, consumers' accelerated flight to digital in this past crisis year have dispelled them. To help them become data driven, companies are deploying increasingly advanced cloud-based technologies, including analytics tools with machine learning (ML) capabilities. What these tools deliver, however, will be of limited value without abundant, high-quality, and easily accessible data.

In this context, effective data management is one of the foundations of a data-driven organization. But managing data in an enterprise is highly complex. As new data technologies come on stream, the burden of legacy systems and data silos grows, unless they can be integrated or ring-fenced. Fragmentation of architecture is a headache for many a chief data officer (CDO), due not just to silos but also to the variety of on-premise and cloud-based tools many organizations use. Along with poor data quality, these issues combine to deprive organizations' data platforms—and the machine learning and analytics models they support—of the speed and scale needed to deliver the desired business results.

To understand how data management and the technologies it relies on are evolving amid such challenges, MIT Technology Review Insights surveyed 351 CDOs, chief analytics officers (CAOs; we refer to these and CDOs as "data leaders" at various points in the report) as well as chief information officers (CIOs), chief technology officers (CTOs), and other senior technology leaders. We also conducted in-depth interviews with several other senior technology leaders. Following are the key findings of this research:

- **Just 13% of organizations excel at delivering on their data strategy**. This select group of "high-achievers" deliver measurable business results across the enterprise. They are succeeding thanks to their attention to the foundations of sound data management and architecture, which enable them to "democratize" data and derive value from machine learning. The foundations ensure reduced data duplication, easy access to relevant data, the ability to process large amounts of data at high speeds, and improved data quality. The high-achievers are also advanced cloud adopters, with 74% running half or more of their data services or infrastructure in a cloud environment.



Organizations' top data priorities over the next two years fall into three areas, all supported by wider adoption of cloud platforms: **improving data management**, **enhancing data analytics and ML**, and **expanding the use of all types of enterprise data**, including streaming and unstructured data.

- **Technology-enabled collaboration is creating a working data culture**. The CDOs interviewed for the study ascribe great importance to democratizing analytics and ML capabilities. Pushing these to the edge with advanced data technologies will help end-users to make more informed business decisions—the hallmarks of a strong data culture. This is only possible with a modern data architecture. One CDO sums it up by saying that successful data management is achieved when the right users have access to the right data to quickly generate insights that drive business value.

- **ML's business impact is limited by difficulties managing its end-to-end lifecycle**. Scaling ML use cases is exceedingly complex for many organizations. The most significant challenge, according to 55% of respondents, is the lack of a central place to store and discover ML models. That absence, along with error-prone hand-offs between data science and production and a lack of skilled ML resources—both cited by 39% of respondents— suggest severe difficulties in making collaboration between ML, data, and business-user teams a reality.

- **Enterprises seek cloud-native platforms that support data management, analytics, and machine learning**. Organizations' top data priorities over the next two years fall into three areas, all supported by wider adoption of cloud platforms: improving data management, enhancing data analytics and ML, and expanding the use of all types of enterprise data, including streaming and unstructured data. For "low-achievers"–organizations having difficulty delivering on data strategy—improving data management overshadows all other priorities, cited by 59% of this group. Most high-achievers, by contrast (53%), are focused on advancing their ML use cases.

- **Open standards are the top requirement of future data architecture strategies**. If respondents could build a new data architecture for their business, the most critical advantage over the existing architecture would be a greater embrace of open-source standards and open data formats. Data leaders now realize the value of open-source standards to accelerate innovation and enable choice in leveraging best-of-breed third-party tools. Stronger security and governance, not surprisingly, are also near the top of respondents' list of requirements.

# 02 Growth and complexity

Cloud, once considered an optional technology environment, is today the foundation for modernizing data management: **63% of respondents** use cloud services or infrastructure widely in their data architecture.

The pace of change in how organizations manage their data has been both breathtaking and frustrating. Once viewed by senior management as a byproduct of operations, data is now regarded as a supreme driver of business value. The volumes of data generated continue to grow at a rapid pace across structured, semi-structured, and unstructured data types that businesses are now able to store and need to analyze.

Whereas not long ago organizations relied on a few technology giants to meet their needs for data infrastructure and tools, enterprise customers today are spoiled for choice from among hundreds of providers in a vast data ecosystem. These players continuously develop new analytics tools—now often powered by machine learning—that parse data at unprecedented speed, depth, and sophistication. Ever-expanding clouds provide organizations with vast space to store, and enormous power to crunch, their data, and in an increasingly cost-efficient manner. Last but not least, new roles and structures have emerged at different levels—witness the rise of chief data officers (CDOs) and chief analytics officers (CAOs), among others—to channel the organization's data capabilities toward creating new business value aligned with its strategic objectives.

"It used to be difficult and costly to me to get data about many elements of our customer experience," says Bob Darin, chief data officer of CVS Health (and chief analytics officer of CVS Pharmacy). "Now I can get insights about our customers, about our supply chain, about how people work that I just couldn't capture before. We have all the tools to analyze that data at scale, and the cost of those tools is coming down. This allows us to develop insights at a great scale and integrate them, so they are part of our patient and customer workflows, enabling us to provide a more personalized and relevant experience for our customers."

Cloud, once considered an optional technology environment, is today the foundation for modernizing data management, providing ever greater storage and computing power at declining cost. Among the companies in our survey, 63% use cloud services or infrastructure widely in their data architecture. Of these, just over one-third (34%) operate multiple clouds.

Nevertheless, frustrations abound with data management. As enterprises seek to upgrade their data platforms, many remain saddled by legacy, on-premise silos that resist easy integration, incur high costs, or cause problems

Partner perspective

## Databricks perspective: The rise of the lakehouse effect

Every company feels the pull to become a data company, and they are placing increasing importance on AI to deliver on the tremendous business potential it can offer. But, as indicated in this report, only 13% of organizations today are succeeding at delivering on their enterprise data strategy. Data and analytics leaders attribute much of their success to having a solid handle on data management basics. So why do so many others struggle?

The challenge starts with the data architecture. The research suggests organizations need to build four different stacks to handle all of their data workloads: business analytics, data engineering, streaming, and ML. All four of these stacks require very different technologies and, unfortunately, they sometimes don't work well together.

The technology ecosystem across data warehouses and data lakes further complicates the architecture. It ends up being expensive and resource-intensive to manage. That complexity impacts data teams. Data and organizational silos can accidently slow communication, hinder innovation and create different goals amongst the teams. The result is multiple copies of data, no consistent security/governance model, closed systems, and less productive data teams.

Meanwhile, ML remains an elusive goal. With the emergence of lakehouse architecture, organizations are no longer bound by the confines and complexity of legacy architectures. By combining the performance, reliability, and governance of data warehouses with the scalability, low cost, and workload flexibility of the data lake, lakehouse

Continued, next page

owing to data duplication and poor quality. This creates a good deal of complexity when it comes to data infrastructure. The cloud, for all its game-changing impact, can also increase complexity as organizations continue to store their data with multiple providers to hedge vendor lock-in, meet regional needs, or optimize for best-of-breed solutions. And data architectures have evolved in a relatively short space of time so that organizations may simultaneously be using on-premise databases, data warehouses, data lakes, or other emerging data architectures along with different cloud-based tools performing configuration, governance, or other functions.

"Architectures have gotten really complicated, but only because we tend to over-complicate them," says Sol Rashidi, chief analytics officer at The Estée Lauder Companies. "We do this because we lose sight of what matters most. We too often bring in the latest and greatest in technology and platforms, thinking they will solve the problem. But unless the business is ready to leverage the tools, has the maturity to extract the insights, and processes and logic are agreed upon, we're only adding to the spaghetti architecture."

If organizations are unable to manage the complexity, the consequences are usually a combination of missed opportunities (in the failure of ML use cases to deliver returns, for example), higher costs (such as from administering and supporting multiple overlapping systems), difficulty meeting the growing regulatory requirements on data, and, ultimately, considerable exposure to competition.

Nevertheless, as our research makes clear, enthusiasm and optimism outweigh any sense of frustration among data and technology leaders when it comes to their present and future ability to manage data effectively for their business.

## Partner perspective

architecture provides a flexible, high-performance design for diverse data applications– including real-time streaming, batch processing, SQL analytics, data science, and ML.

At Databricks, we bring the lakehouse architecture to life through the Databricks Lakehouse Platform. The key enabler behind this innovation is Delta Lake. Delta Lake is at the core of the platform, and it creates curated data lakes that add reliability, performance, and governance from data warehouses directly to the existing data lake. Organizations get a better grasp on enterprise-wide data management.

The Databricks Lakehouse Platform excels in 3 ways:

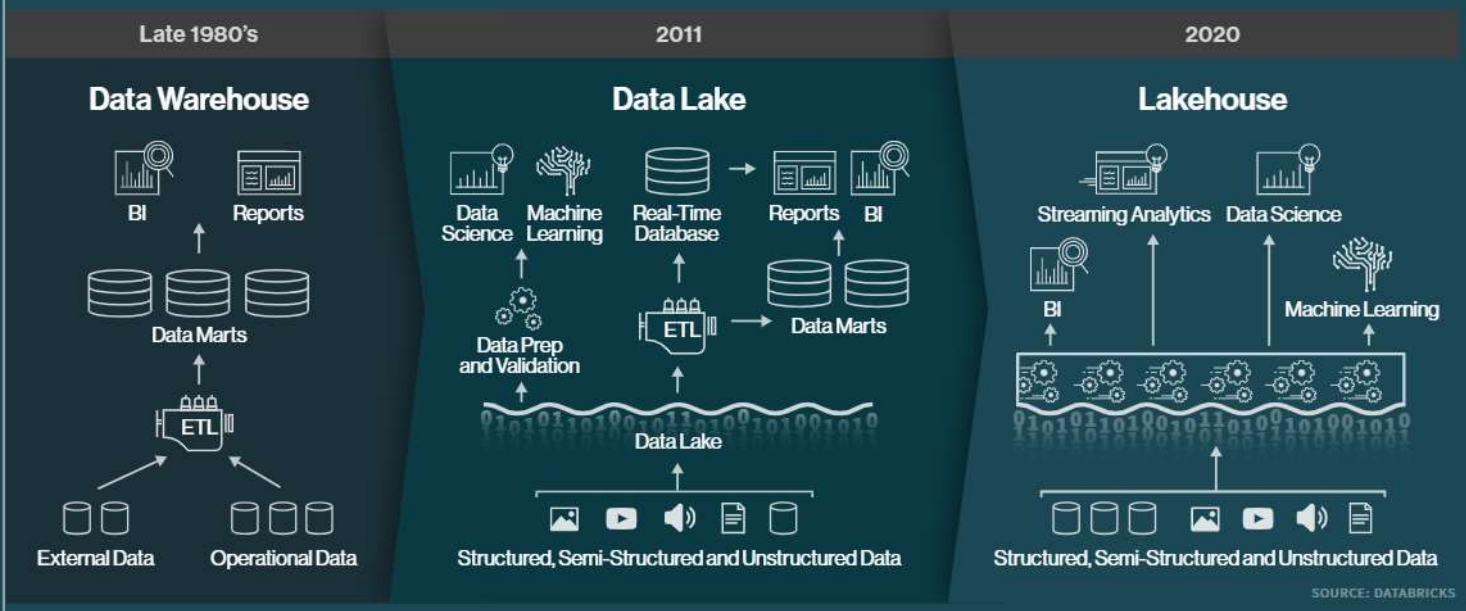
**It's simple:** Data only needs to exist once to support all workloads on one common platform.

**It's open:** Based on open source and open standards, it's easy to work with existing tools and avoid proprietary formats.

**It's collaborative:** Data engineers, analysts, and data scientists can work together and more efficiently.

The cost savings, efficiencies, and productivity-gains offered by the Databricks Lakehouse Platform are already making a bottom-line impact on enterprises in every industry and geography. Freed from overly complex architecture, Databricks provides one common cloud-based data foundation for all data and workloads across all major cloud providers. Data and analytics leaders can foster a data-driven culture that focuses on adding value by relieving the daily grind of planning and all its complexities, with predictive maintenance.

From video streaming analytics to customer lifetime value, and from disease prevention to finding life on Mars, data is part of the solution. Understanding data is the key that opens the doors.




SOURCE: DATABRICKS

# 03 Aligning and delivering on strategy

Amid a global economic downturn of a scale not seen for nearly a century, businesses might be expected to be reining in their ambitions and focusing on the bottom line. Many of those represented in our survey, however, appear growth oriented. When asked about the most important business objectives they have set for their enterprise data strategy over the next two years, more respondents stress top-line growth, in the form of expanded sales and service channels (cited by 45%) than those who point to improved efficiency (43%). Following closely (at 42%) is improving innovation and reducing time to market of new or improved products.

A look at the surveyed firms' principal data initiatives over the next two years suggests a substantial degree of alignment with a growth-oriented business strategy. It also reflects their recognition of the urgency of improving data management in order to support those business objectives. The wider adoption of cloud-native platforms will underpin this and other initiatives.

The most frequently cited priority is achieving better data management by improving data quality and processing, mentioned by 48% of respondents. (That figure is 74% among those working in oil and gas companies and 67% in consumer products firms.) Such efforts are critical to

**Figure 1: Companies' most important business objectives for enterprise data strategy over the next two years** (top responses; % of respondents)

MIT Technology Review Insights survey, 2021

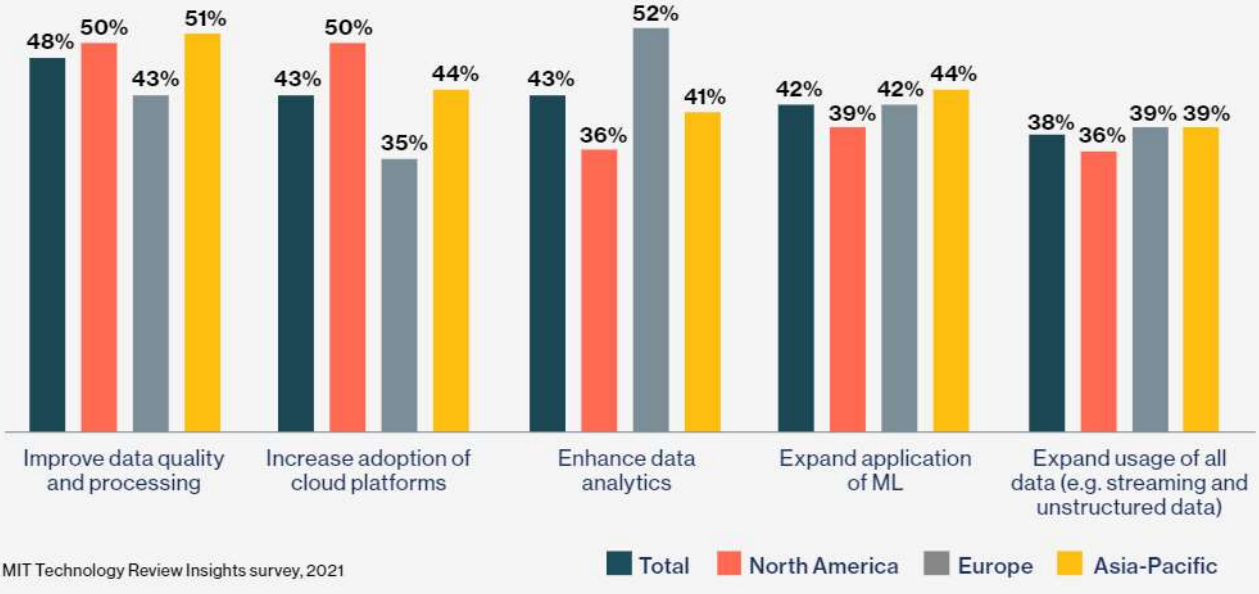
## Achieving better data management by improving data quality and processing is critical to enabling growth-oriented efforts, like those driven by ML, to move ahead at speed.

enabling growth-oriented efforts, such as those driven by ML, to move ahead at speed. For Hivery, an Australia-based retail technology firm whose products are powered by artificial intelligence (AI), the quality of its customers' data matters even more than its ability to ingest large volumes. In fact, says Andy McQuarrie, Hivery's chief technology officer, the cleaner its customers' data, the fewer ingestion problems it encounters.

The other top data priorities of the surveyed firms—increasing the adoption of cloud platforms (cited by 43%), enhancing data analytics (43%), and expanding the application of ML (42%)—if met, will provide data teams with additional capacity, power, and scale to, among other things, quickly tap new sales and service opportunities and support new data product development. They also, of course, fully support the goal of improving operational efficiency. Another priority (cited by 38%) is expanding the use of streaming, unstructured, and other varieties of data.

That data strategy should be closely aligned with the overall business objectives seems self-evident today, but the importance of alignment has not always been clear. According to Don Vu, chief data officer of US financial services firm Northwestern Mutual, alignment of data and business strategy has become much tighter at many companies as CDOs have exerted their influence, and data responsibilities have been brought together in streamlined organizational structures. At his firm, says Vu, "people knew that alignment was important, but that became crystallized as our teams dug deeper into how we're actually going to deliver on the various business strategy initiatives. The link to business strategy from notions such as trust in essential sources of truth, or democratizing the use of data, became much clearer."

**Figure 2: Companies' most important enterprise-wide data strategy initiatives over the next two years** (top responses; % of respondents)




MIT Technology Review Insights survey, 2021

## Data high-achievers

Not many large enterprises excel at data management. This is reflected in the survey, where only 13% of respondents rate their organization's performance highly when it comes to delivering on data strategy, scoring it at the top end (9-10) of a 1-10 scale. These data "high-achievers" deliver with measurable business impact across multiple business units, say their executives. They are contrasted with a similarly sized group of "low-achievers" (12% of the sample), whose data performance is rated at 6 or lower on the scale.

Large gaps separate these two groups in certain attributes as well as intentions. For example, cloud features more prominently in the data architecture of

**Figure 3a: The extent to which organizations are successfully delivering on the enterprise data strategy**
(self-assessed rating on a 1-10 scale where 10=succeeding)

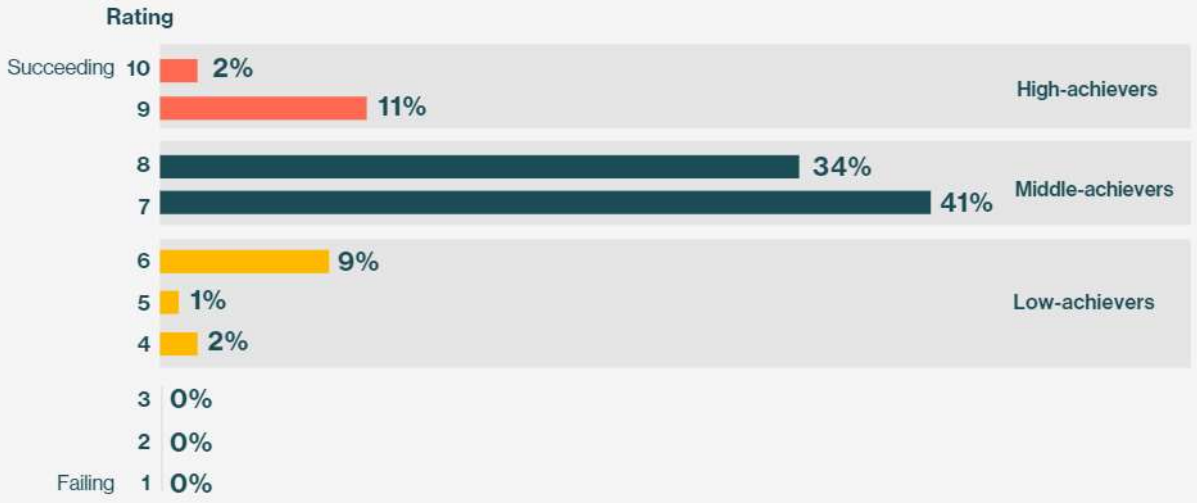
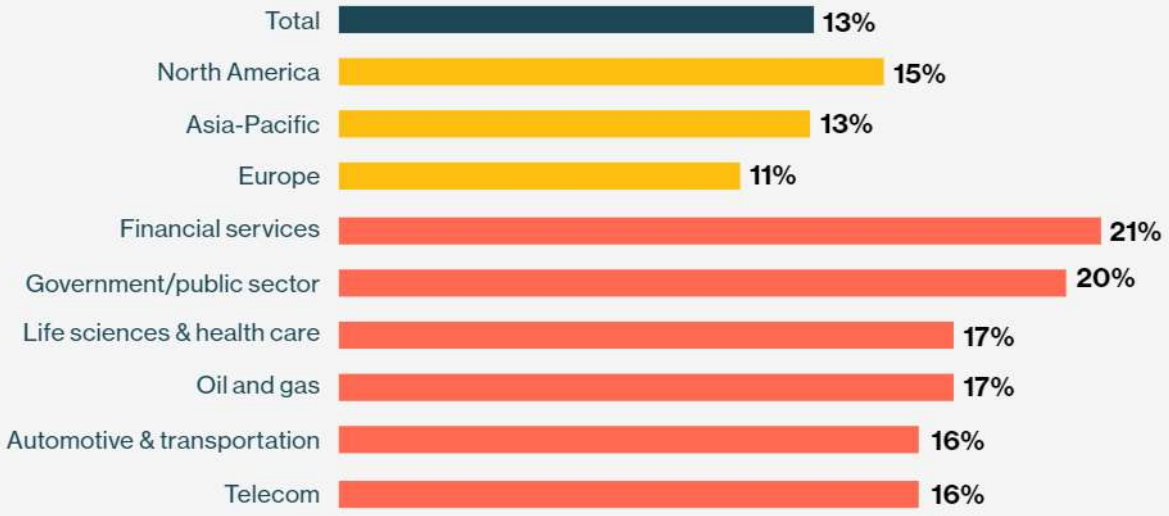



**Figure 3b: High-achievers: respondents rating their organizations 9 or 10 on their delivery of enterprise data strategy, with measurable business impact across multiple business units**
(total, regions and selected industries)




MIT Technology Review Insights survey, 2021

**Figure 4: The main success factors enabling "high-achiever" organizations to deliver on their data strategy initiatives** (top responses; % of respondents)

MIT Technology Review Insights survey, 2021

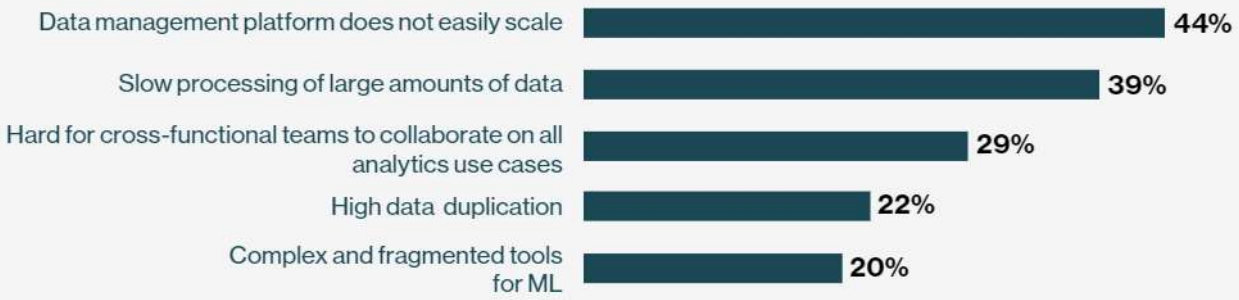
high-achievers: 74% of this group run at least half of their data services or infrastructure in a cloud environment, compared with 60% of low-achievers who do the same. When it comes to data priorities, most low-achievers (59%) are focused on improving data management (data quality and processing) over the next two years, while high-achievers' most frequently cited initiative (by 53%) is expanding the application of ML.

Far from taking the basics for granted, the high-achievers attribute their success to their close attention to the foundations of sound data management. These include the reduction of data duplication, ease of data access for enterprise end-users—a hallmark of data "democratization"–and the processing of large data volumes at high speeds.

Duplication of data in large organizations happens at multiple levels such as data warehouses, operational systems, reports, dashboards, and desktop tools. This has significant cost, risk management, and reliability implications, says Ashwin Sinha, chief data and analytics officer at Macquarie Bank. Data duplication also impacts the ability to scale and make effective use of machine learning across the organization.

Asked what's holding back their progress, the largest percentage of low-achievers point to limited scalability of their data management platform. Other often-cited impediments are slow processing of large data volumes and difficulties in facilitating collaboration. Achieving scale, speed, and collaboration, as we will see, are challenges for organizations right across the span of data operations.

**Figure 5: The main challenges keeping "low-achiever" organizations from delivering on their data strategy initiatives** (top responses; % of respondents)




MIT Technology Review Insights survey, 2021

## Nielsen: data transformation for a data-reliant business

It is hard to overestimate the importance of sound data management to Nielsen, one of a few century-old organizations in which data has been central to the business model from day one. Nielsen's panels tell consumer goods companies what products customers are buying and how behaviors are changing. The panels also advise such companies on where and when they should place their television advertising. Now in his second year as the firm's chief data officer and fifth as chief research officer, Mainak Mazumdar has presided over a transformation of its data management and infrastructure. "Just a few years ago," he recalls, "we struggled with fragmentation–lots of data in silos and tribal knowledge needed to access it–a lack of metadata and very little governance, all while data volumes were growing by petabytes each day."

Mazumdar paints a different picture today: "Now we're able to scale quickly from having 20-30 specialists on a platform to 300-plus. We're on a cloud platform with a data lake where data is curated, labelled, defined, has metastores, and is consolidated. We've built our own analytics engine. In fact, much of what was done by software engineers in the past is now done by my team, deploying directly into production." The changes, says Mazumdar, have reduced his team's cycle time by 50%. "The speed of our models is now about 50x. What used to take 20 minutes we can now do in a minute or less. At the same time, we are ingesting and processing massive amounts of data, which are easily accessible to data science. It's a huge change."

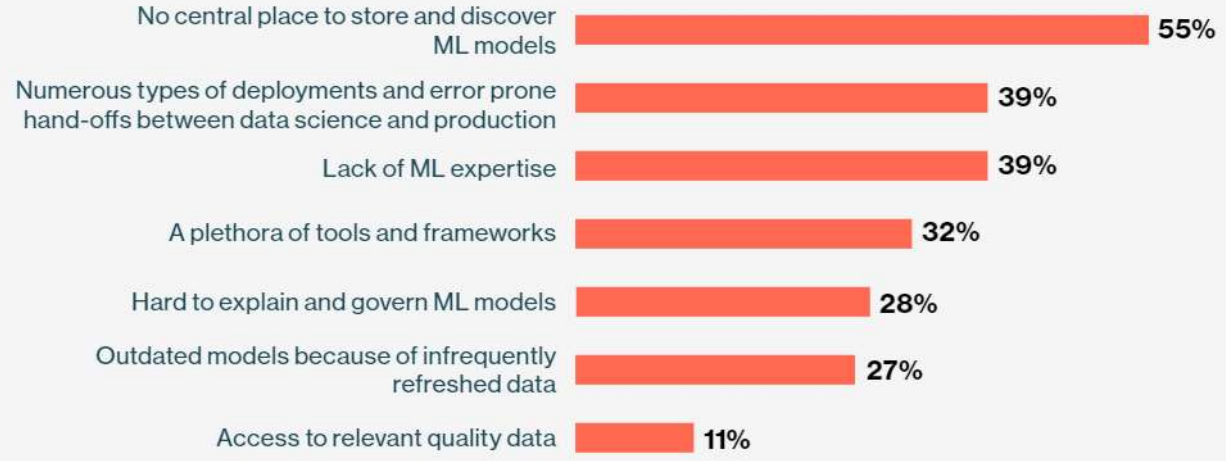
An example of how Nielsen has used these capabilities to enable growth is the roll-out of a new ratings product in the company's roughly 200 local markets in the US. Crunching large amounts of data from TV set-top boxes, the "high-recognition deep-learning model" enables Nielsen to predict for customers not only what a viewer is likely to watch at any given time, but also who in each household is doing the watching, something not previously possible. "We could not have rolled out this product without the changes made to how we manage data, and we could not have ingested such volumes of data," says Mazumdar. "It's with that ingestion of ever greater volumes of data that the models–and the product–get better and better."

# 04 Scaling analytics and machine learning

Business leaders know that their company's ability to keep pace with and anticipate demand, to manage competitive pressures, to innovate effectively, and to operate efficiently is coming to rest on their mastery of analytics and ML. Organizations in virtually every industry are busy developing analytics and ML use cases that will deliver greater business impact. For most large enterprises, a wide portfolio of use cases in production and at scale is no longer a nice-to-have but a must-have. CDOs and their teams are increasingly judged on their contributions to delivering such cases.

Many organizations struggle with this, and particularly with achieving the scale needed to generate a sizeable impact. According to Sol Rashidi of Estée Lauder, one reason is over-ambition: "Too often companies want to skip crawling and walking with ML and go straight to running, without having mastered the basics." For other CDOs, such as Don Vu of Northwestern Mutual, the key challenges lie in selecting the right use cases to deploy into production. Without business user input, he says, the probability rises of selecting cases that do not clearly map to a business objective.

**Figure 6: The main difficulties companies encounter in scaling ML use cases**
(top responses; % of respondents)




MIT Technology Review Insights survey, 2021

## A paradigm shift at CVS Health

Pharmacies have always played an essential role in societies, but arguably never more so than during the past year. Pharmacy chains such as CVS, America's largest by revenue, are using ever more advanced data capabilities to ensure that their customers are up to date on their medications and use of other health services, lest underlying conditions lead to more serious health consequences. Bob Darin is leading many of those efforts as chief data officer of CVS Health and as chief analytics officer of its retail pharmacy business.

The company has long used data systems to prod customers to stay current with their medications. This has involved, for example, patient outreach through phone calls and texts, prompts at the pharmacy counter, or recommendations for the patient to talk with their health-care provider about specific follow-up or medication reviews. In recent years, those initiatives have become embedded with data science in order to make them more personalized, says Darin. "We now know which reminders, which programs, what modes of communication, and what messaging are going to be most effective in helping people remember to take their medications, schedule their pick-ups, and understand cost-saving opportunities." Those are all now driven by analytics models, he says, whereas it used to be a one-size-fits-all process.

"That's been a paradigm shift for us," says Darin. "Not just in how we use analytics but also in how we manage our data platform and architecture." It has involved investments in data warehousing, operational reporting, and analytics, he says. His teams are on a journey, however, "to move more of our business functions into a hybrid, multi-dimensional data environment—one that that can support not just data warehousing and descriptive analytics use cases, but can use more advanced machine learning, algorithm development, and optimization techniques at scale."

This type of environment, says Darin, "needs to support more advanced data science applications as well as day-to-day business analytics, such as descriptives and ad hoc queries – the types of information people need to make their business decisions every day. We need architectures and data platforms that support both."

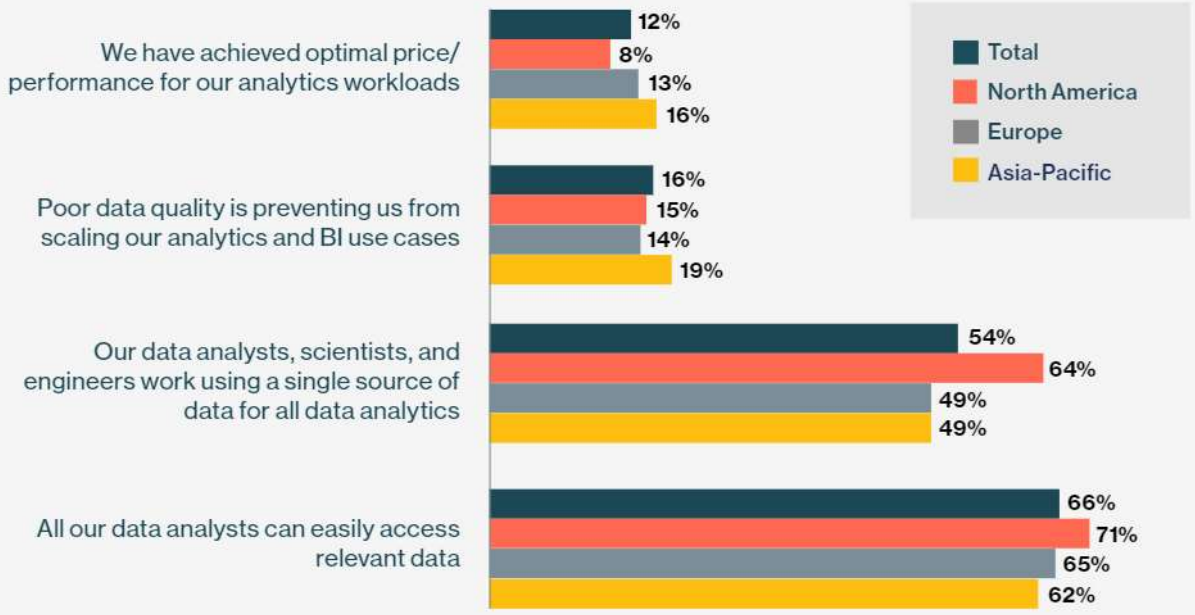
## Three-quarters of those working in the power sector say the lack of a central place to store and discover ML models is a major hindrance to scaling, as do over two-thirds of those in the government and consumer goods sectors.

Ensuring a robust dialog between ML model builders and business users insures against the "shiny algorithm" effect, where the data science team prioritizes the most exciting ML use cases that business users find little value in. "If your machine learning use case isn't mapping very closely to a business use case," says Andy McQuarrie of Hivery, "then I'm wondering what you're doing with it."

### Barriers to scale

Companies clearly struggle with the complexity of managing the end-to-end ML lifecycle. This is illustrated in the list of difficulties the respondents' organizations encounter in scaling use cases (Figure 6). The most common problem, cited by 55%, is the lack of a central place to store and discover ML models. Three-quarters of those working in power and other utilities say this is a major hindrance to scaling, as do over two-thirds of those in government agencies and in consumer product firms. Inadequate collaboration between data science and production, reflected in multiple deployments and error-prone hand-offs, is another key impediment to scaling, according to 39%. This figure is 50% among respondents from life sciences and health-care organizations. And many organizations grapple with a multiplicity of tools and frameworks, as well as infrequent refreshing of model data (cited by 32% and 27% respectively).

**Figure 7: Respondents who strongly or somewhat agree with statements about analytics and business intelligence (BI) at their organization** (% of respondents)




MIT Technology Review Insights survey, 2021

"The issue of multiple deployments and error-prone hand-offs between data science and production is a huge issue. There's often a gap between the data science output and the results we get after operationalizing it."

– Naveen Jayaraman, Vice President – Data, CRM & Analytics, L'Oréal

The issue of multiple deployments and error-prone hand-offs between data science and production is one that resonates with the CDOs we interviewed. "This is a huge challenge," says Naveen Jayaraman, vice president – data, CRM & analytics at L'Oréal, a provider of personal care products. "There's often a gap between the data science output and the results we get after operationalizing it."

CVS Health has managed to reduce hand-off problems, according to Bob Darin: "In the past, we had data scientists who would build models and come up with insights. They would hand them to another team to implement. That 'waterfall' approach didn't work, and the models didn't scale well." For CVS Health, he says, integrating data science and production into single teams working in partnership with IT has helped smooth these difficulties. "It's much better than having one team develop the insights and another do the scaling."

The skills gap remains a systemic issue for organizations looking to build successful ML practices. The lack of ML expertise is cited by 39% of respondents as a major barrier to scaling use cases, and is a particular problem for the manufacturers in the survey (64%). The high-achievers, however, appear able to surmount any such gaps, with just 27% citing this as an ML challenge. Low-achievers, by contrast, struggle mightily with it, with 59% saying it was a major challenge for them.

## Protecting return on investment

Difficulties building ML and analytics use cases to the intended scale have a return on investment (ROI) impact. For example, just 12% of survey respondents say they have achieved optimal price/performance for their analytics workloads. The data duplication issues cited earlier are likely to detract from performance, as is the cost of allocating scarce resources to focus on the low-level task of cleaning poor-quality data. And reaching enterprise-wide scale often entails investment in cloud infrastructure and data management, says Jayaraman. "This is tricky territory," he says. "Unless a use case shows big wins in sales or efficiency gains, you will always have a big ROI challenge."

Meeting or exceeding performance expectations, on the other hand, can have more than just an ROI benefit, says Patrick Baginski, senior director data science at fast food giant McDonald's. "The time to value for ML is critical. The quicker you can demonstrate value from ML and data science, the quicker you get users' buy-in and build management confidence in the value of both to the organization." These in turn are important contributors to the development of data culture.

## Technology, democracy, and culture

For all the transformation that data management has undergone in systems, leadership, and perceptions of business value, there remains a sizeable divide at most organizations between data teams and end-users, the front or back-office employees who need data insights to make decisions on a daily basis. Many CDOs are seeking to bridge the divide by embedding data scientists directly into business units where they regularly interact with users. Another way of bridging the gap is to put analytics at the direct disposal of users so that they can draw insights themselves as needed. That, says Patrick Baginski of McDonald's, also requires pushing data closer "to the edge", to where the user is.

This is an objective that all the CDOs and CAOs we interviewed are pursuing. "As tools advance and people become more familiar with advanced analytics and data science, we've got to give users the ability to run the analytics themselves, rather than just consume analytics that someone else produces," says Bob Darin.

This is data democratization in practice. It is also, for some CDOs, a means of not only building data-literacy but also encouraging the growth of data culture in an organization. Data infrastructure and tools contribute directly to this. "Our aim, as a central data team, is to create tooling that federates a lot of the responsibility and power to others," says Don Vu. "So we're trying to push the power of our data platform to the edge. We want to create gravity among application teams and users from a features perspective. We want people to want to be on this platform because it takes care of data quality, it takes care of privacy, and it takes care of access."

According to Ashwin Sinha of Macquarie Bank, improving end-user ease of access to data and analytics has made a direct contribution to enhancing the organization's data culture. "This is not just at analyst levels," he says. "Senior leaders in the organization who have an analytical mindset use data visualization tools with simple interfaces to analyze data and get insights." Visualization tools have been important to enhancing access, Sinha says, but so has the team's use of cloud infrastructure in improving reliability and performance. "Our data workloads—regulatory, machine learning, and analytical—have all been shifted to a cloud data platform over the last three years. The data platform uses a variety of cloud services, open-source packages, and data management tools to ensure that data across the organization is standardized, integrated, and provisioned in a fit-for-purpose governed manner, with full lineage and traceability of data."

Strong data governance, ease of access, and simplification all combine to enhance user trust in data and analytics, says Sinha. A data culture cannot develop without trust and reliability in a data platform.

"Know what data matters most, prioritize it, build the discipline to protect and govern it, then democratize the data to enable your data specialists and end-users to extract the insights they need to innovate."

– Sol Rashidi, Chief Analytics Officer, The Estée Lauder Companies

# 05 Visions of the future

Half of the surveyed executives (and around two-thirds of those in technology and manufacturing firms) say they are currently evaluating or implementing a new data platform to address their current data challenges. Another 9% would do so but face blockers in changing their architecture. Here again, high- and low-achievers diverge markedly. Most high-achievers (56%) are satisfied with their current architecture, although nearly one-third of these are, nevertheless, researching new platforms or evaluating vendors. Just 24% of low-achievers, by contrast, are happy with the existing architecture, and 59% are actively seeking to change it. Technology firms and manufacturers are far more likely than those in other sectors to be on the path of upgrading their platform.

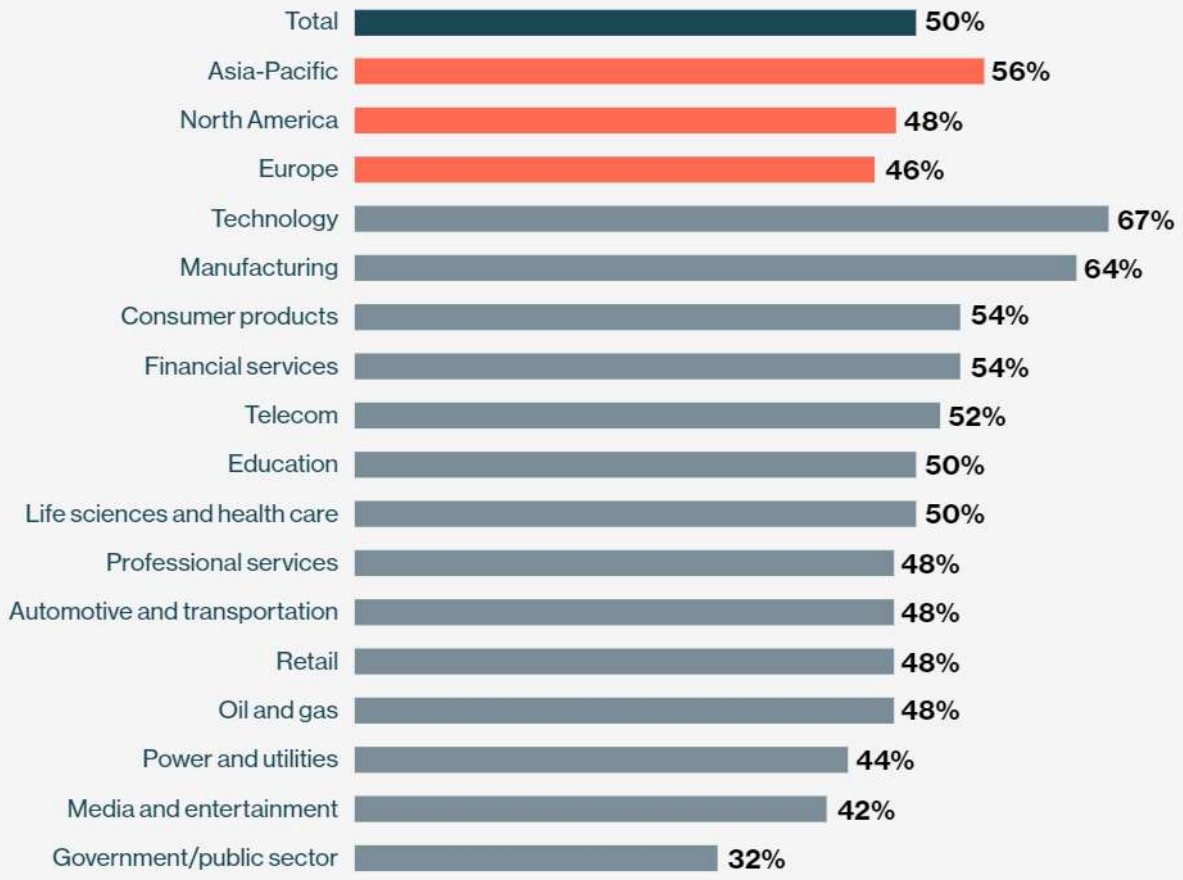
## A CDO wish-list for a new architecture

What if data and technology leaders could build a new data architecture for their business? We asked them what its key advantages would be over their existing one. At the top of their wish-list, cited by 50% (and by 60% or higher among technology, utilities, and government respondents), are open-source standards and open-data formats. This is no surprise to Andy McQuarrie of Hivery: "Open standards are what allows you to move through the maturity curve quite easily. You can consume services with a managed service in the short term. When your business matures and you find you need to add a component, with open source you've got that option without it affecting the entire architecture. You don't need to be changing technologies the way that you might have had to in the past."

Next on the wish-list are not new capabilities, but rather strengthened ones in areas that technology leaders never stop seeking improvement in: stronger security, stronger governance, and better price/performance for infrastructure, operations, maintenance, and other architecture elements. The respondents are also insistent that any new architecture supports all analytics use cases, whether based on ML, data science, or business intelligence.

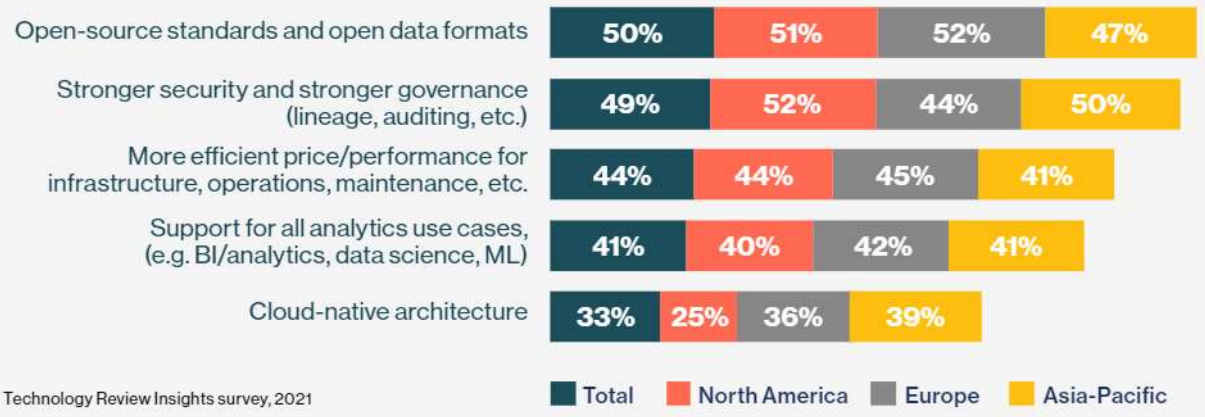
Open-source standards and open-data formats are top of the wish-list for technology leaders, followed by stronger security and governance, and better price/performance for architecture elements.

**Figure 8: Respondents currently evaluating or implementing a new data platform to enable their enterprise-wide data strategy or solutions to address their current challenges**
(% of respondents)




MIT Technology Review Insights survey, 2021

**Figure 9: The most critical advantages of respondents' ideal new architecture over the existing one**
(top responses; % of respondents)




MIT Technology Review Insights survey, 2021

> "Native integration with customer systems is something we are really focused on. How can we make it easier for our customers to transfer their data to us?"
>
> – Andy McQuarrie, Chief Technology Officer, Hivery

The CDOs we interviewed have additional hopes and plans as their data management and infrastructure evolve. Michel Lutz, group chief data officer at energy major Total, aims to expand ML-based automation of its metadata management. This has already been done for geoscience data, but he plans to extend it to all the group's data domains in the next year. The potential of data mesh architecture is also on Lutz's radar. "It's certainly a next step for our data architecture because it allows a higher organizational scalability and enables more domain specialization," he says.

L'Oréal's Naveen Jayaraman similarly sees the advantages of domain-oriented data products for enterprises like his that look to distribute data management functions widely. Architectural approaches using data mesh have potential for his purposes, Jayaraman says, provided that some elements, such as polyglot storage, can be centralized.

Mainak Mazumdar of Nielsen looks forward to implementing "AI for AI"—essentially the automated selection of AI models. "We hope to build in another layer of intelligence that will make the best decisions about models to use."

On a more distant horizon, Ashwin Sinha thinks about the possibilities of "agile and real-time architecture", one that is agile enough to adapt to future innovation in data engineering. It would be a way, he says, of future-proofing a recently designed architecture, at least for the medium term. "We'd like to be able to adopt an architectural approach flexible enough to ensure it won't become outdated in two or three years."

# 06 Conclusion

A common thread in our research, and particularly our discussions with CDOs, has been recognition of a direct line running from data infrastructure through data science to the attitudes towards data held by employees up and down the organization, and tying it all back to business impact–i.e. the data culture. The research offers some lessons for CDOs looking to use their assets to nurture a data culture in their business.

- **Keep it simple and flexible.** Many CDOs want to simplify overly complex architectures, but it is just as important to ensure that interfaces for analytics end-users are easy to use and, to the extent possible, fun. The architecture must also remain flexible enough to ensure future business needs are met without requiring large migrations or redesigns to add new technologies. As more users are able to pull and play with constantly refreshed data using their preferred tools, and derive insights from it on-demand, engagement will grow.

- **Keep it well-governed.** Nothing will mar employees' experience of analytics more than the recurring discovery of errors in the data. CDOs do not need to be reminded of the importance of sound data governance, but not all may realize how far the implications of governance failures can reach in the organization. Building and maintaining a single source of truth matters to everyone.

- **Explain and evangelize.** Some CDOs task parts of their teams to educate business units and other staff about data science and train them in the use of analytics and other tools. Many also find that advertising widely the positive outcomes achieved, and value gained, from the use of data and analytics can expand their employees' desire to use them. Advertising data accuracy achieved through strong governance will also help build trust.

- **Bond with C-suite peers.** The virtue of deepening ties with fellow C-levels may seem self-evident to CDOs and CAOs, but any disconnects on data management with CIOs and CTOs could, if not addressed, become sources of mistrust. The data strategy needs to factor in the existing technology infrastructure and ensure it is always focused on delivering the corporate priorities.

## About MIT Technology Review Insights

MIT Technology Review Insights is the custom publishing division of *MIT Technology Review*, the world's longest-running technology magazine, backed by the world's foremost technology institution—producing live events and research on the leading technology and business challenges of the day. Insights conducts qualitative and quantitative research and analysis in the US and abroad and publishes a wide variety of content, including articles, reports, infographics, videos, and podcasts. And through its growing MIT Technology Review Global Panel, Insights has unparalleled access to senior-level executives, innovators, and thought leaders worldwide for surveys and in-depth interviews.

## From the sponsor

Databricks is the data + AI company. Built on a modern lakehouse architecture, Databricks combines the best of data warehouses and data lakes to offer a simple, open, and collaborative platform for all data workloads. More than 5,000 organizations worldwide—including Shell, Comcast, CVS Health, HSBC, T-Mobile, and Regeneron—rely on Databricks to enable massive-scale data engineering, exploratory data science, full-lifecycle machine learning, and business analytics. With a global presence and hundreds of partners, including Microsoft, Amazon, Google, and Tableau, Databricks is on a mission to help data teams solve the world's toughest problems.

**Illustrations**
Davooda: 5, 9, 13, 14; fad82:13; Friday Studio: 2, 3, 6,15 ,17, 22; Nadiinko: 18; ninamalin:18, papipo:15; Prostock Studio: Covers, front and back,2, 3, 4, 5, 13,19; Rashas Ashor: 19; Starline: Covers, front and back, 2. All Illustrations provided by Shutterstock, assembled by Scott Shultz Design.

*While every effort has been taken to verify the accuracy of this information, MIT Technology Review Insights cannot accept any responsibility or liability for reliance by any person in this report or any of the information, opinions, or conclusions set out in this report.*

© Copyright MIT Technology Review Insights, 2021. All rights reserved.

**MIT Technology Review Insights**

www.technologyreview.com

@techreview @mit_insights

insights@technologyreview.com