Prepared for

The 2022 State of Data and What's Next

February 2022 EMA White Paper
By Will Schoeppner, Research Director

Table of Contents

Executive Summary

Today's business relies on access to critical data more than ever to remain ahead of the competition. Due to the challenges over the last two years, organizations must respond to greater demand for data access and a shift to supporting AI/ML in a complex, hybrid multi-cloud environment. Organizations need to become nimble in providing fast, reliable access to data anytime, anywhere, throughout their increasingly complex, integrated enterprise.

Research proves that a growing trend in business operation functions is to move from a traditional centralized model to a decentralized model. This movement is seen across a company's data platforms and architecture. This trend is in line with a data mesh approach based on modern architecture for analytical data management, enabling organizations to have fast, reliable access to data to meet business demands.

There are many factors that add significant levels of complexity in today's data environment. Organizations have an average of 4-6 data platforms and as many as 12 separate data platforms, creating an intricate data landscape that includes many applications and systems, and the integration layer that connects them. This complexity can create additional challenges for these enterprise-class organizations on their journey to be data-driven. Inversely, emerging startups can create a robust data strategy without the complexity and, in many cases, may be further along in maturity in the data-driven journey. Key to overcoming this complexity is rooted in an organization's strategy and focus on automation of key process in infrastructure.

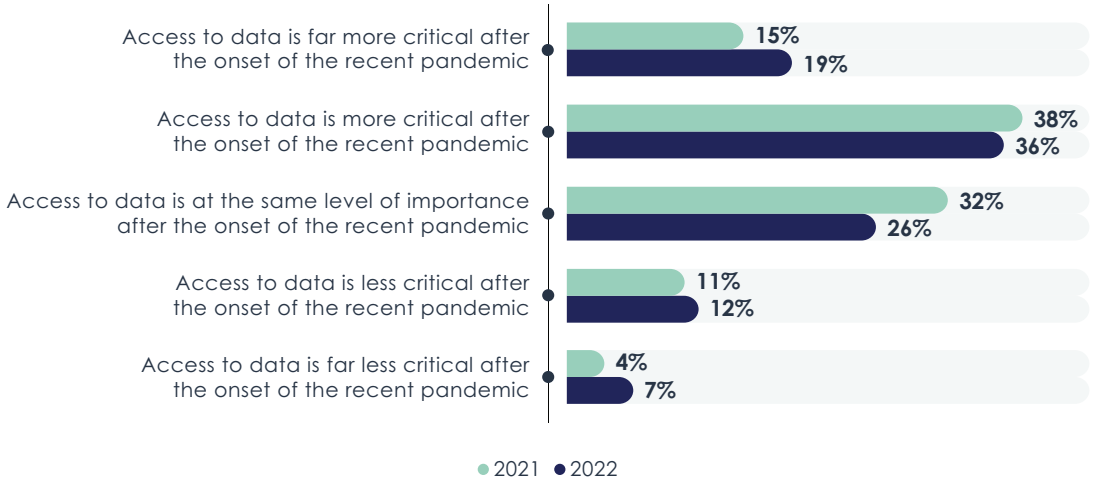
The Pandemic's Impact on Data Access

The COVID-19 pandemic continues to transform how organizations operate and rely on access to critical data to drive business functions and insights into customer experience. Organizations with a highly developed strategy and culture that embrace a data architecture centered on a hybrid, multi-cloud solution have adapted quickly in today's fast-paced environment.

EMA research indicates that the world will continue to see increasing reliance on data, with pressures for data access growing to meet customer demands in an advancing digital, highly mobile landscape.

To meet the stringent challenges businesses are facing, EMA predicts that technology functions will shift to be quick and flexible. This will require a new data platform architecture to meet the demands for faster, more secure, and more reliable access to data.

What impact has the recent pandemic had on the need for data access at your company?



When asked about their organization's data-driven business strategy, only 41% of respondents said their organization has a fully integrated data-driven strategy.

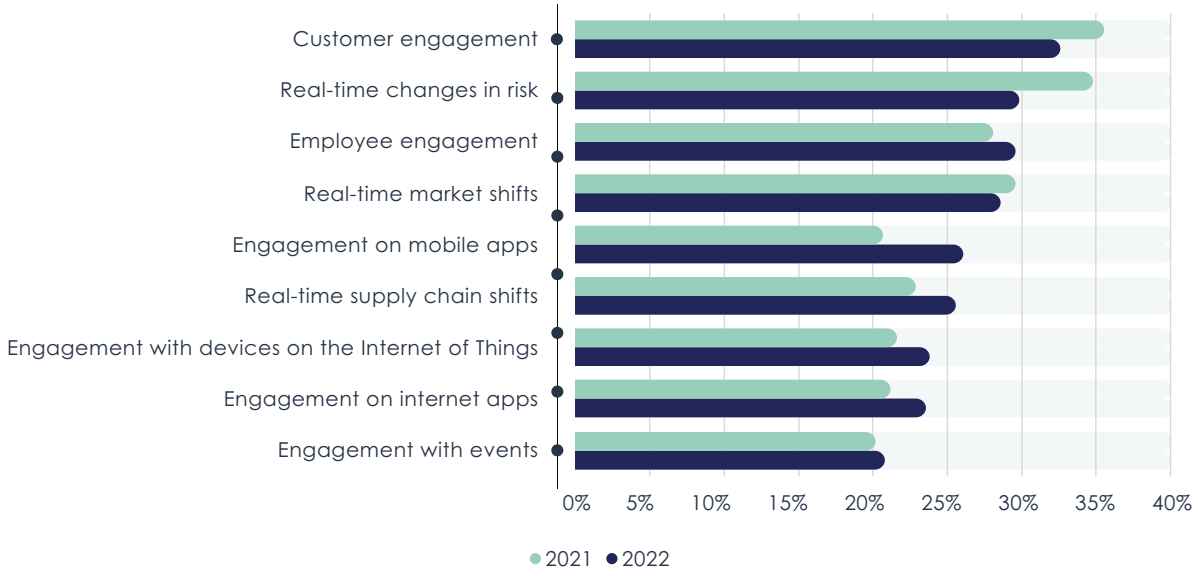
Move to Faster Data Access

As a result of the challenges over the last two years, organizations must respond to greater demand for data access and a shift to supporting AI/ML in a complex, hybrid multi-cloud environment. Organizations need to become nimble in providing fast, reliable access to data anytime, anywhere. Research proves that demands on customer experience, the ever-growing challenge of staying ahead of risk and market swings, and employee engagement are the driving factors for these shifts.

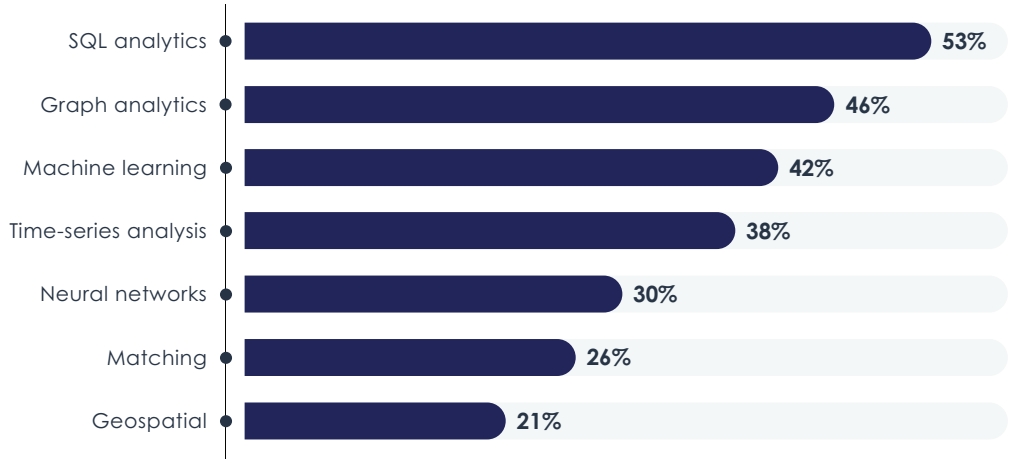
A constant trend remains to drive business decisions with SQL analytics at the forefront, with 53% of respondents rating it the highest importance to their analytical program.

EMA's research shows that organizations are moving data pipelines into production faster than ever before, with 51% of organizations moving pipelines into operation within one day, compared to 48% last year. Healthcare and government entities are the quickest, moving 55% of data pipelines into production within one day.

What is driving your company's need for more real-time access to data or analytics?



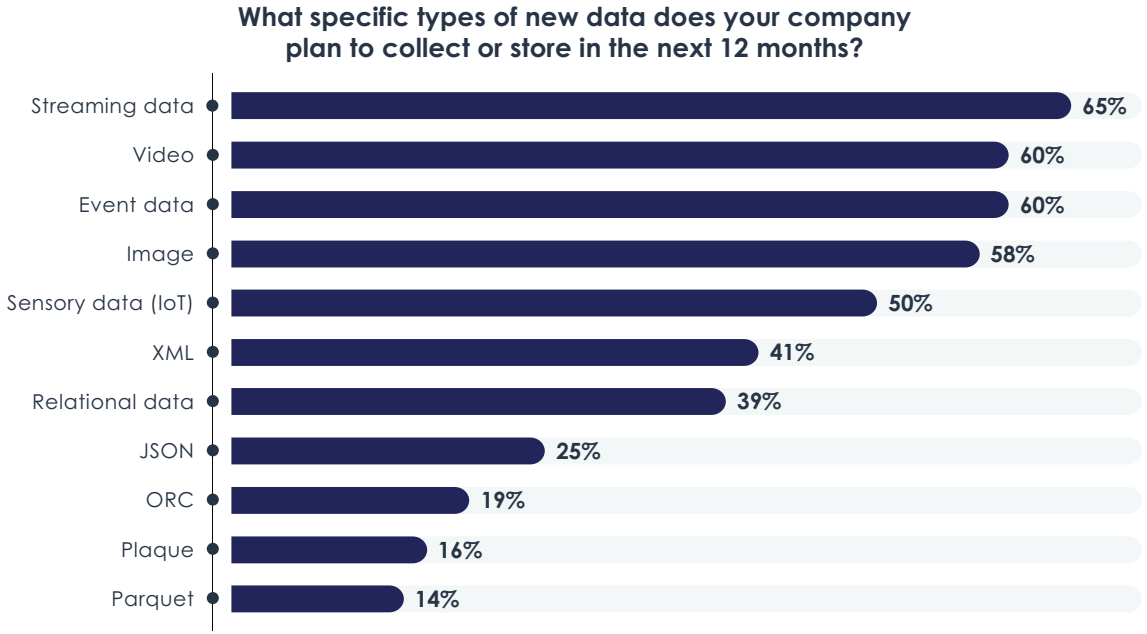
Which types of analytical workloads are important to your overall analytics program?



Data Sprawl Has Continued

Data sprawl is a phenomenon that continues to plague businesses. Data storage complexity persists as data is held locally on-premises or in multi-cloud storage platforms, and depending on the organization, may be geographically dispersed. This complex, distributed, and dynamic IT environment creates several barriers for businesses to access data that is reliable and fast. A growing concern is data security in a complex storage environment, which creates a cybersecurity challenge with organizations' critical data dispersed across multiple platforms.

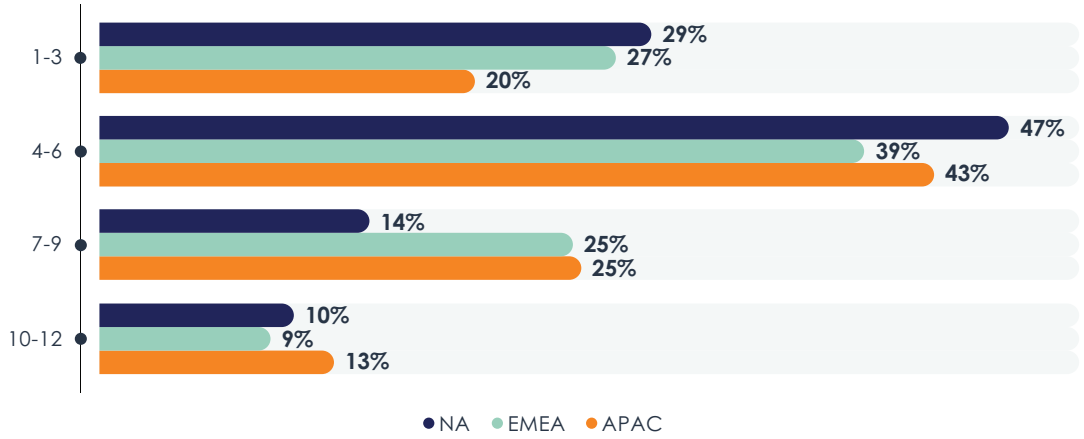
Organizations are adding new data to their environments at an increasing rate. In today's digital and remote environment, it's not surprising that 65% of respondents show streaming data as the primary data that they will collect in the next year, followed by 60% of respondents indicating video and event data as secondarily important. Image data and sensory data also are significant, with half or more of businesses indicating there will be an increase in data collection in these areas over the next year.



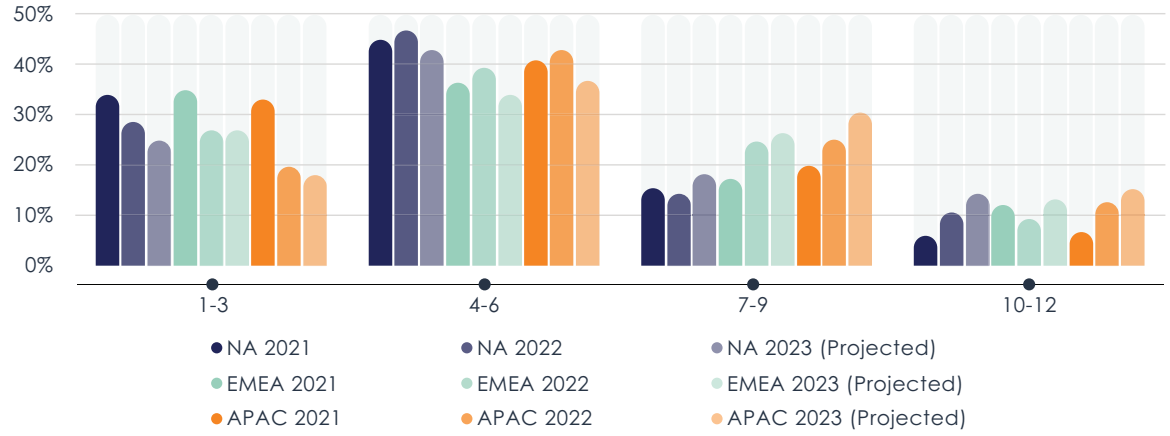
Timely availability of data is important. When asked, "Thinking about your experience in having all the data you need, please indicate a percentage for each data availability description," 43% responded, "Data is available where I need it."

Data sprawl complexity issues are universal across industry and region. When asked how many different data platforms, such as analytic platforms, data lakes, data warehouses, object storage, and other types of data storage platforms respondents currently have in their data ecosystem, the average number in most organizations is 4-6 platforms, with at least 11% of organizations having 10-12 platforms. As data sprawl concerns spread, so does the number of platforms in an organization's data ecosystem. Interestingly, we are seeing the highest increase in APAC countries, with projections showing platforms increasing last year and projected to continue to rise into 2023.

How many different data platforms do you currently have in your data ecosystem?



How many different data platforms do you currently have in your data ecosystem?

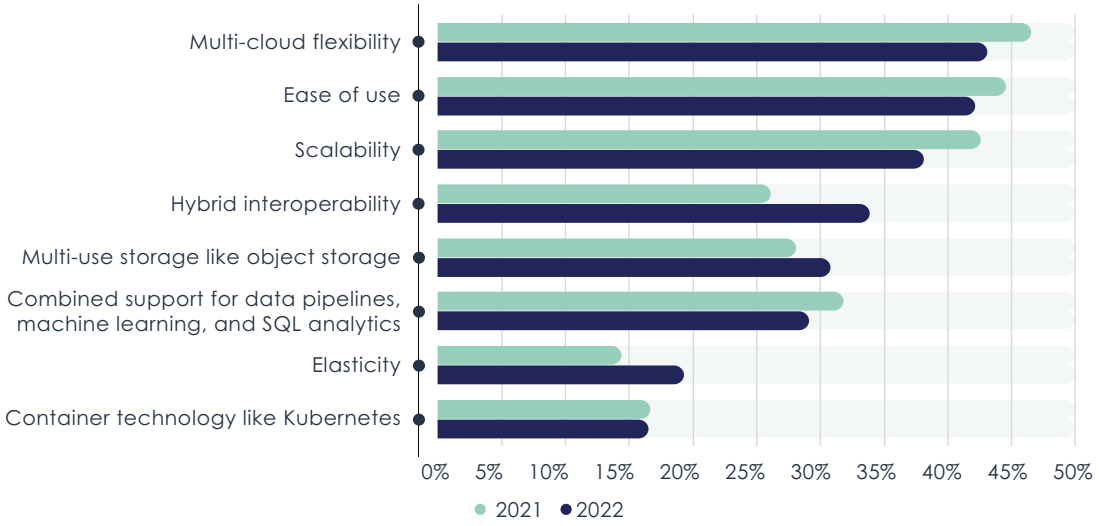


Trending Shifts

EMA continues to see a growing trend in highly scalable hybrid multi-cloud environments. Companies were already on a digital transformation journey and with the COVID-19 pandemic, this accelerated. A hybrid cloud strategy gives organizations the flexibility of security for sensitive customer data on-premises and the scalability, speed, and flexibility that a public cloud platform offers. However, the digital transformation journey requires a well-developed strategy with thorough planning to be successful. This strategy needs to be based on measurable business outcomes, a defined innovation roadmap, and an architecture that incorporates all aspects of the business.

When asked which aspects of cloud data storage and access most impact a business's buying decisions, multi-cloud flexibility, ease of use, and scalability remain the top reasons. However, EMA's research is seeing a shift with hybrid interoperability, multi-use storage, and elasticity growing in importance from 2021.

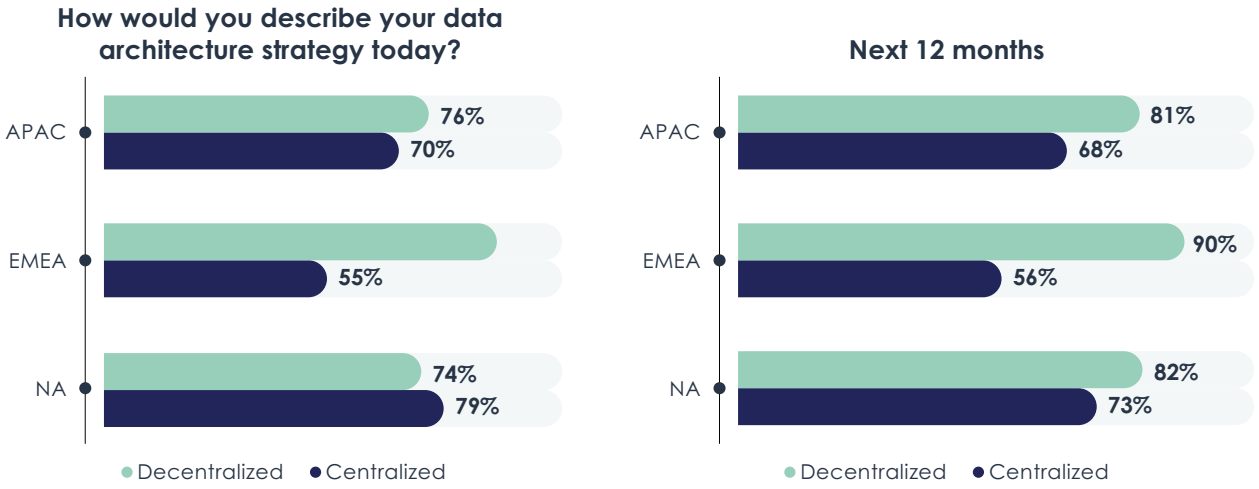
Which aspects of cloud data storage and access most impact your buying decisions?



Shift to a More Decentralized Model

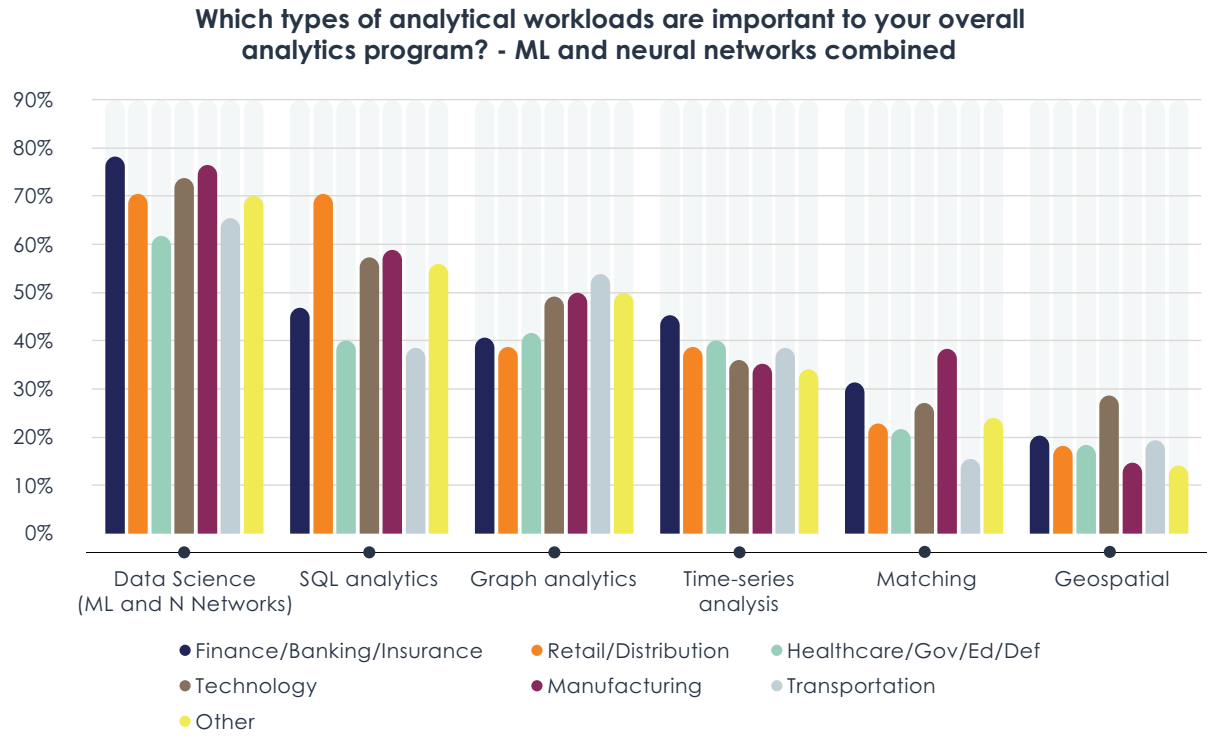
Another growing trend EMA is seeing across technology functions and even business operation functions is the notion to move away from a centralized model to one that is more decentralized. Undoubtedly, the benefits of centralization remain. These benefits include consolidated cost, high level of control, and ease of management. However, centralization also comes with increased risk, with a single point of failure and a centralized model lacking business flexibility. This lack of flexibility can leave a business slow to adapt in a rapidly changing environment.

EMA is seeing a similar tendency across data platforms and architecture. Most organizations have a data architecture strategy centered around a decentralized model. EMEA organizations have progressed the most in the decentralized movement; however, countries in North America and APAC are closing the gap and moving to a more decentralized architecture strategy for their data platforms. This movement fits a data mesh approach based on a modern architecture for analytical data management, enabling companies to easily access and query data where it resides without transporting it to a data lake or warehouse.



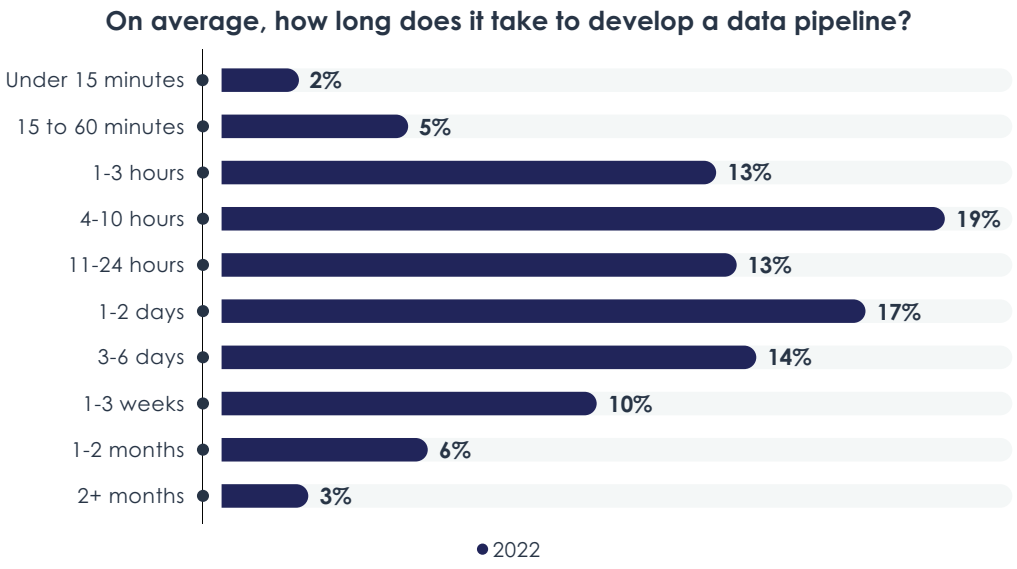
AI/ML Increasing and Placing Greater Pressure on a Variety of Systems

The need for faster, more secure access to data across multiple platforms has increased with the need for organizations to consume and process vast amounts of AI/ML data. Data science workloads are increasing, applying pressure to already complex data platforms. Along with the struggle to find data science resources, companies are looking for solutions that can automate data science workloads, saving time and resources. Historically, when asked which types of analytical workloads are important to an overall analytics program, SQL analytics topped the list. Currently, EMA is seeing a shift for nearly every industry wherein data science (ML and neural networks) is now rated as the most important analytical workload.



The Data Pipeline/ELT Problem Persists... Including Creating Bottlenecks

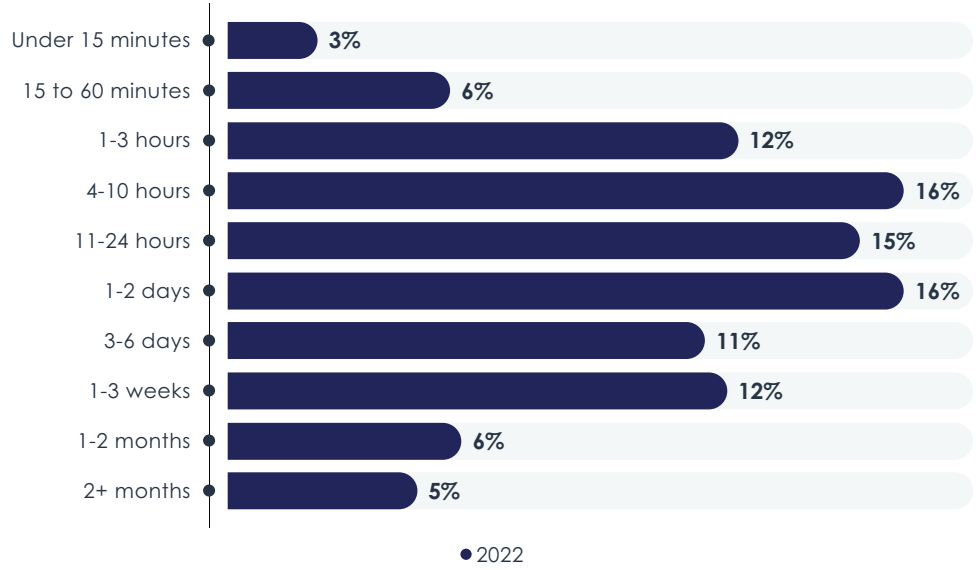
Digital transformation and today's unprecedented rapid change require even faster responses to business events, and data pipelines are needed to process and deliver valuable insights. As shown in this research study, the complexity of data pipelines can take too much time to develop and place into production, creating a backlog of data and preventing real-time decisions. In most cases, developers manually create data pipelines without the ability or resources to automate development, significantly impacting real-time business operations. Over 48% of respondents said that they take more than a business day to develop a data pipeline, with 32% taking three days to two months.



Further Delays in Data Pipeline Operationalization

Once the data pipeline is developed it needs to be moved into operation, which takes additional time and has an additional impact on business operations. Half of the respondents said it takes at least another full business day to get new data pipelines into production, with 23% indicating that production takes a week or even more.

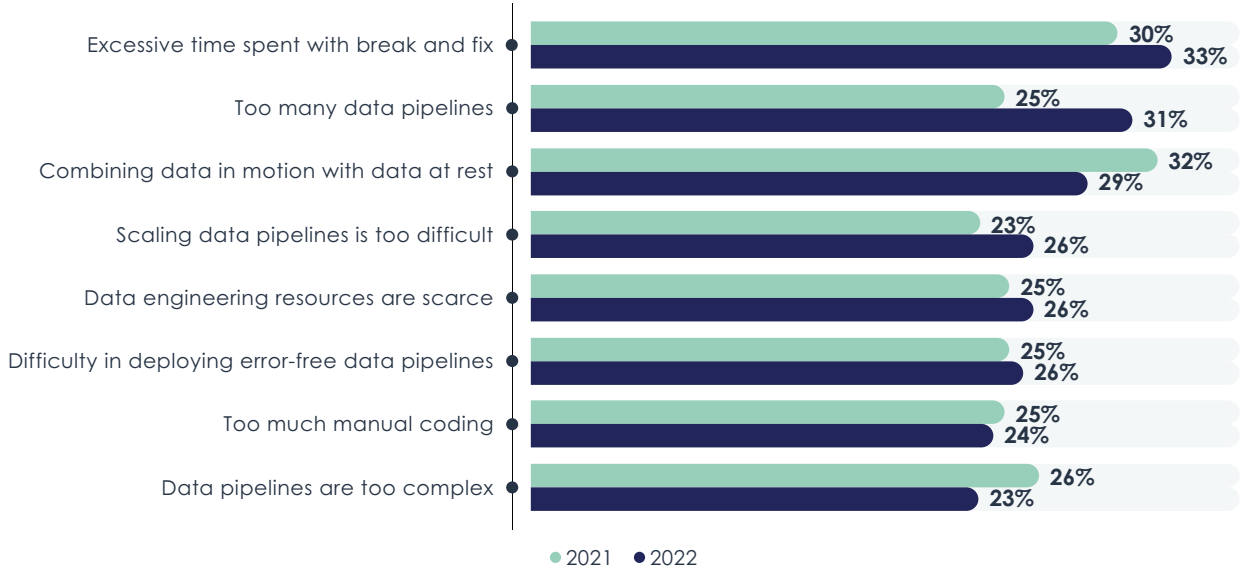
On average, how long does it take to make a data pipeline operational in production?



The Difficulties of Data Pipelines

As seen in this research, developing pipelines and moving them into operation is complex and takes a significant amount of time. When asked about the biggest challenges a respondent faces in building and deploying data pipelines, key themes are seen over the past two years. Due to the inherent complexity and the reliance on a manual process for coding and development, organizations spent valuable resources on break/fix across multiple pipelines and struggled to combine data in motion with data at rest.

What are the biggest challenges you face in building and deploying data pipelines?



Thirty-one percent of respondents said that data constantly being moved and changed makes finding data difficult.

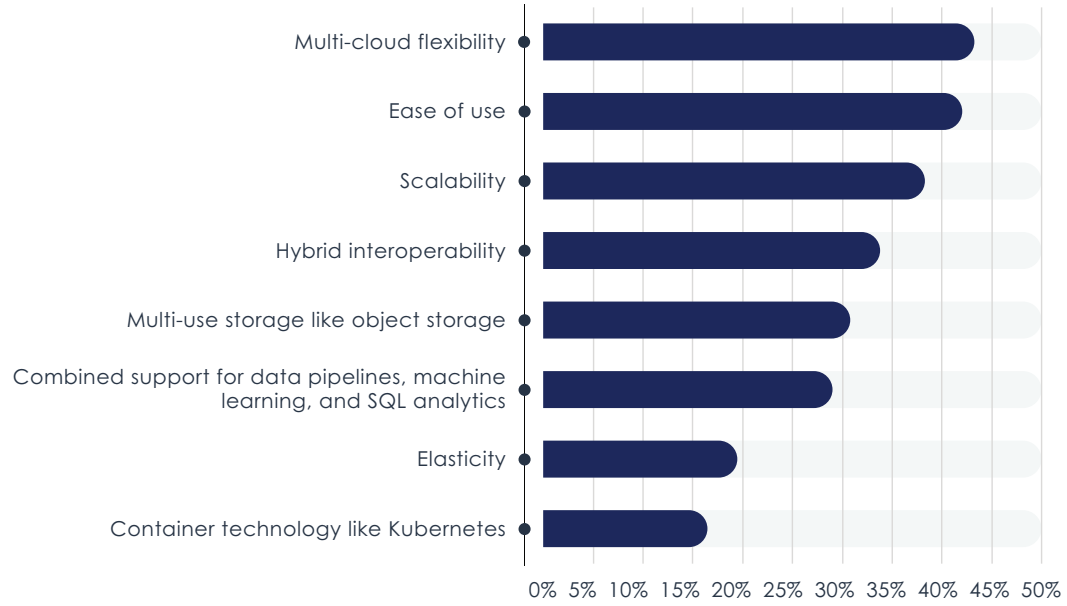
Trends in the Next 12 Months

The first and most obvious solution for data dispersion and multi-platform complexity is a move to the cloud. Moving to the cloud has been accelerated by the testing of digital business models in the global shutdowns. Respondents cited that in 2021, 56% of their data was in the cloud versus 44% on-premises. When asked the same question this year, respondents stated that 59% of their organization's data resides in the cloud compared to 41% on-premises. EMA predicts this trend will persist as companies continue their digital transformation journey.

Percentage of respondents' data that was in the cloud

Digital transformation and cloud migration also include a search for solutions for multi-cloud interoperability. As part of the journey to the cloud, the number-one impact on buying decisions is multi-cloud flexibility (43%). The second key factor impacting business buying decisions is ease of use (42%).

Which aspects of cloud data storage and access most impact your buying decisions?

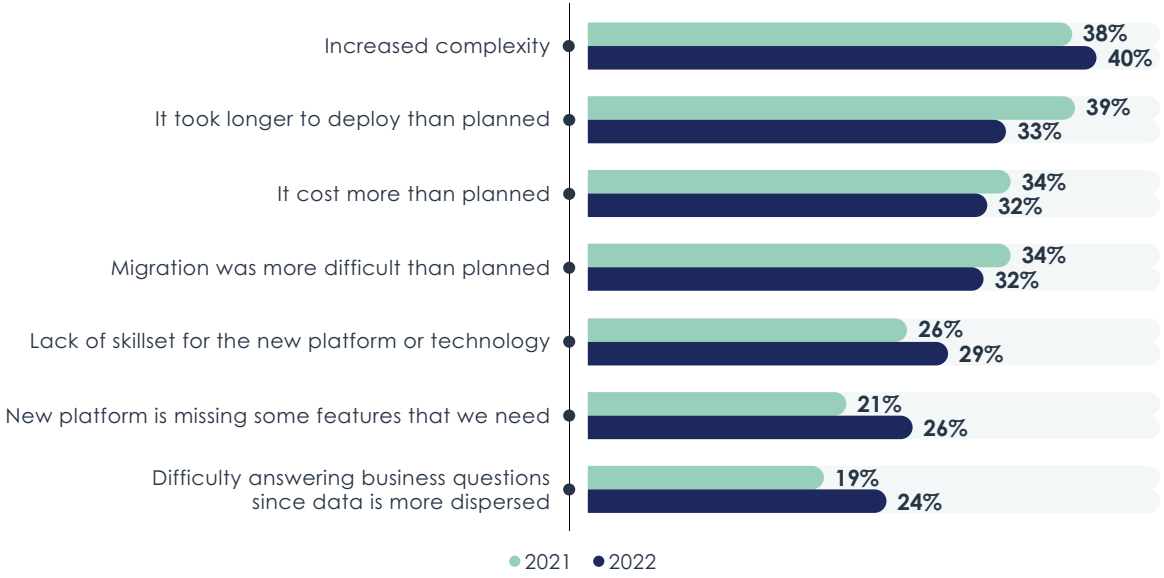


The current rapid pace of digital transformation and the need for faster, secure, and reliable access to data to meet growing business demands has exponentially increased the need for advanced analytics in decision making. As a result, current manual processes are inadequate for keeping up with demand. High-value data workers of all types are being asked to do more with less. It is no surprise that resource constraint is the number-one driver for automation across the entire information supply chain.

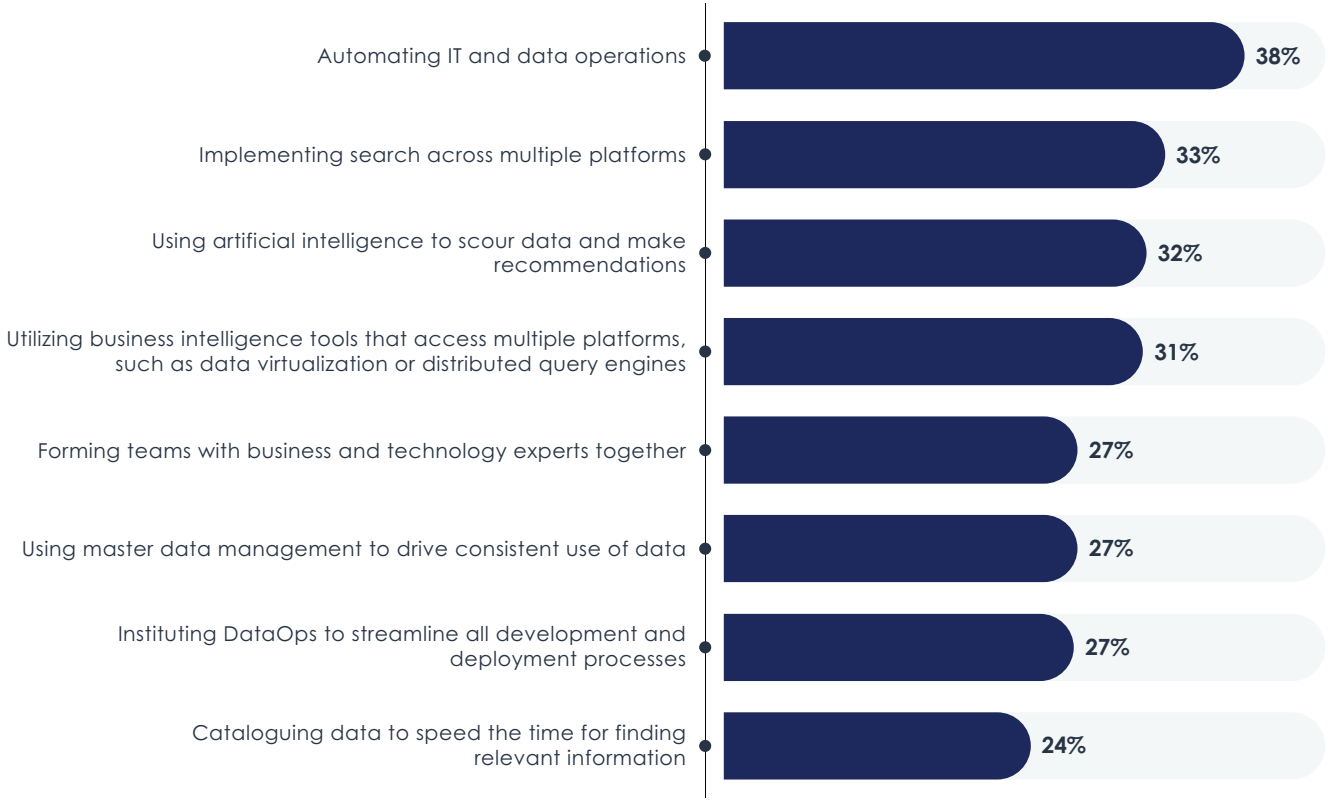
EMA has seen a trend in the use of machine learning in data management and analytics platforms for the past few years. A measure of maturity for an organization is the increased ability to move several formerly manual functions into completely automated processes. AI-enabled analytics promise to scale up data and analytics programs to impact more business decisions with fewer resources.

With data spread throughout organizations, the number-one challenge organizations face is the increased complexity of a hybrid, multi-cloud environment. Organizations strive to maintain a competitive edge in the industry and look to automation of critical technology systems to make this happen. At the forefront, a focus on the underlying data infrastructure with automation of data pipelines, the adoption of intelligent search, and implementing AI and ML in data processing can drive business decisions.

What challenges have you experienced when adopting new technologies or implementing new platforms in the past year?



With data spread across your company in different systems, what practices are most important to making the systems work together?



About Starburst and Red Hat

Starburst and Red Hat provide modern solutions that address data silo and speed of access problems. With Starburst's fast, distributed query Trino engine, Starburst Enterprise, and the leading enterprise Kubernetes platform, Red Hat OpenShift®, organizations can now run analytics anywhere to make better business decisions with minimal additional load on their operations teams.

Traditional data warehouse products approach data silo problems with outdated, monolithic solutions that breed inefficiency and ultimately cannot help business analysts run fast analytics on all their data. This prevents the business from making better and more timely decisions to improve their company's performance.

Companies everywhere are building distributed cloud and hybrid cloud applications. Many of these organizations rely on both traditional applications and modern applications to run their business and make critical business decisions. Likewise, their data sources include both traditional data sources and new data sources located everywhere—in data centers, cloud, and even vendor environments.

Instead of rebuilding data infrastructure from scratch, Starburst lets you add a straightforward, easy-to-manage and operate, real-time, distributed query engine that accesses your data no matter what form it appears in or where it resides.

With Starburst Enterprise, you can perform federated data queries across your different data sources—whether structured, semi-structured, or unstructured—even using different protocols. You can also perform onsite data analysis across various file systems, databases, and object stores delivered in storage platforms, such as Red Hat OpenShift Container Storage, Red Hat Ceph Storage, and more.

Starburst Enterprise provides distributed query support for varied data sources, such as Apache Cassandra, Hive (HDFS), S3 (HDFS), Microsoft SQL Server, MySQL, and PostgreSQL data sources. Starburst Trino Operators delivered with Red Hat OpenShift Container Platform automate installation, upgrades, and lifecycle management throughout the container stack.

Together, Red Hat and Starburst provide a simple, cost-effective, straightforward way to manage architecture that gives organizations fast access to all of their data to make better business decisions on more complete data.

For more information, visit **https://www.starburst.io/platform/deployment-options/red-hat/**.

Appendix: Demographics – 400 Participants

Company size by employees

500-999	**69**
1,000-2,499	**106**
2,500-4,999	**99**
5,000-9,999	**54**
10,000-19,999	**22**
20,000 or more	**50**

Company size by revenue

Less than $1 Million	**8**
$1 Million to less than $50 Million	**60**
$50 Million to less than $100 Million	**47**
$100 Million to less than $500 Million	**75**
$500 Million to less than $1 Billion	**78**
$1 Billion to less than $10 Billion	**101**
$10 Billion or more	**31**

Location by country

Australia	**72**
Canada	**33**
France	**59**
Germany	**60**
Singapore	**40**
United Kingdom	**64**
United States	**72**

Industry

Aerospace/Defense	**8**
Consulting: Computer-Related	**15**
Consulting: All Other	**2**
Consumer Goods	**11**
Education	**10**
Finance/Banking/Insurance	**64**
Government	**18**
Healthcare/Medical	**24**
Life Sciences	**11**
High Technology: Software	**46**
High Technology: Reseller	**12**
High Technology: Application, Services	**35**
Hospitality/Entertainment/Rec/Travel	**3**
Manufacturing: Computer-Related	**8**
Manufacturing: All Other	**26**
Media Entertainment	**2**
Oil/Gas/Chemicals	**9**
Professional Services: Computer- Related	**14**
Professional Services: All Other	**5**
Retail/Wholesale/Distribution	**33**
Telecommunications	**12**
Transportation/Airlines/Trucking/Rail	**18**
Utilities/Energy	**8**
Other	**6**

About Enterprise Management Associates, Inc.

Founded in 1996, Enterprise Management Associates (EMA) is a leading industry analyst firm that provides deep insight across the full spectrum of IT and data management technologies. EMA analysts leverage a unique combination of practical experience, insight into industry best practices, and in-depth knowledge of current and planned vendor solutions to help EMA's clients achieve their goals. Learn more about EMA research, analysis, and consulting services for enterprise line of business users, IT professionals, and IT vendors at **www.enterprisemanagement.com**. You can also follow EMA on **Twitter** or **LinkedIn**.

1995 North 57th Court, Suite 120, Boulder, CO 80301 +1 303.543.9500 **www.enterprisemanagement.com**

4155.030922