

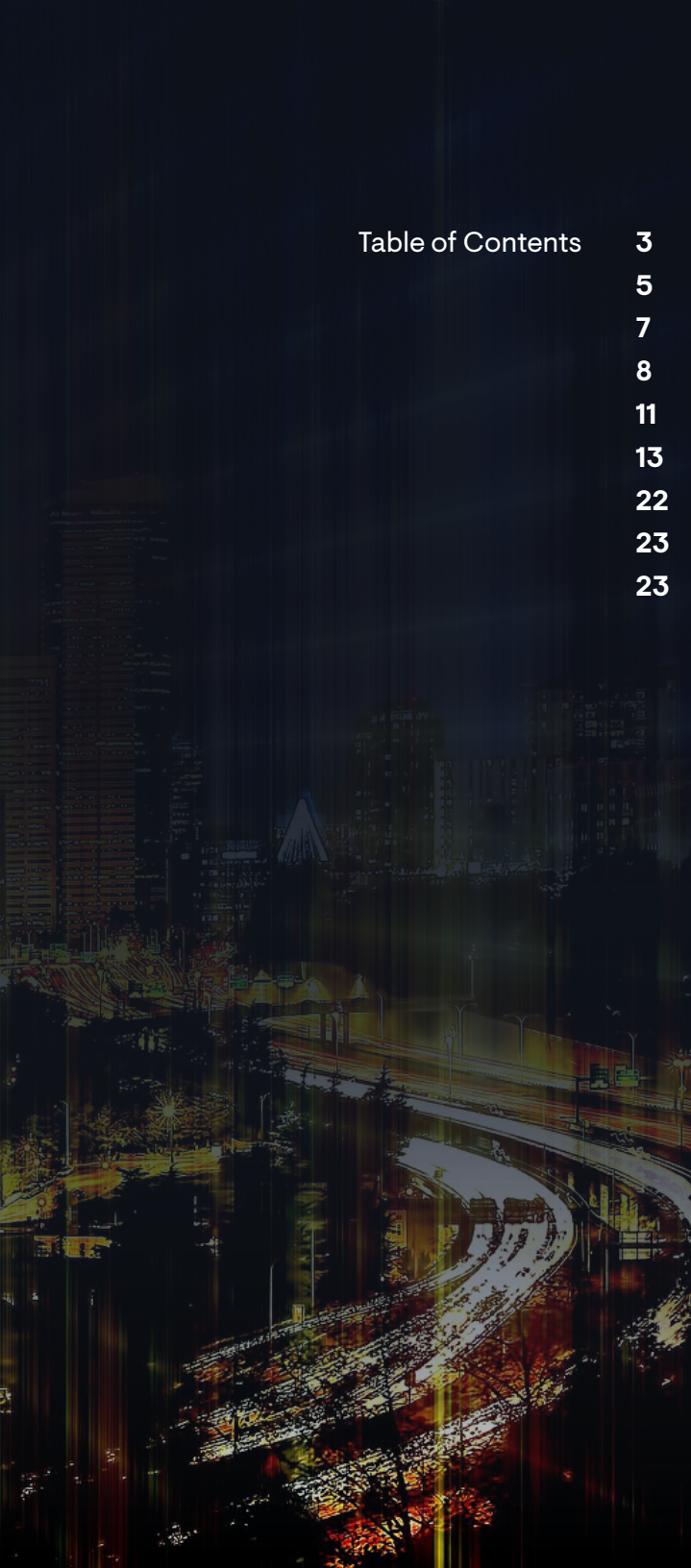
# Navigating the Complexity of Hybrid Data Ecosystems

## A Look at How to Modernize with Connected, Real-Time Intelligence

**May 2021 EMA Research Report Summary**

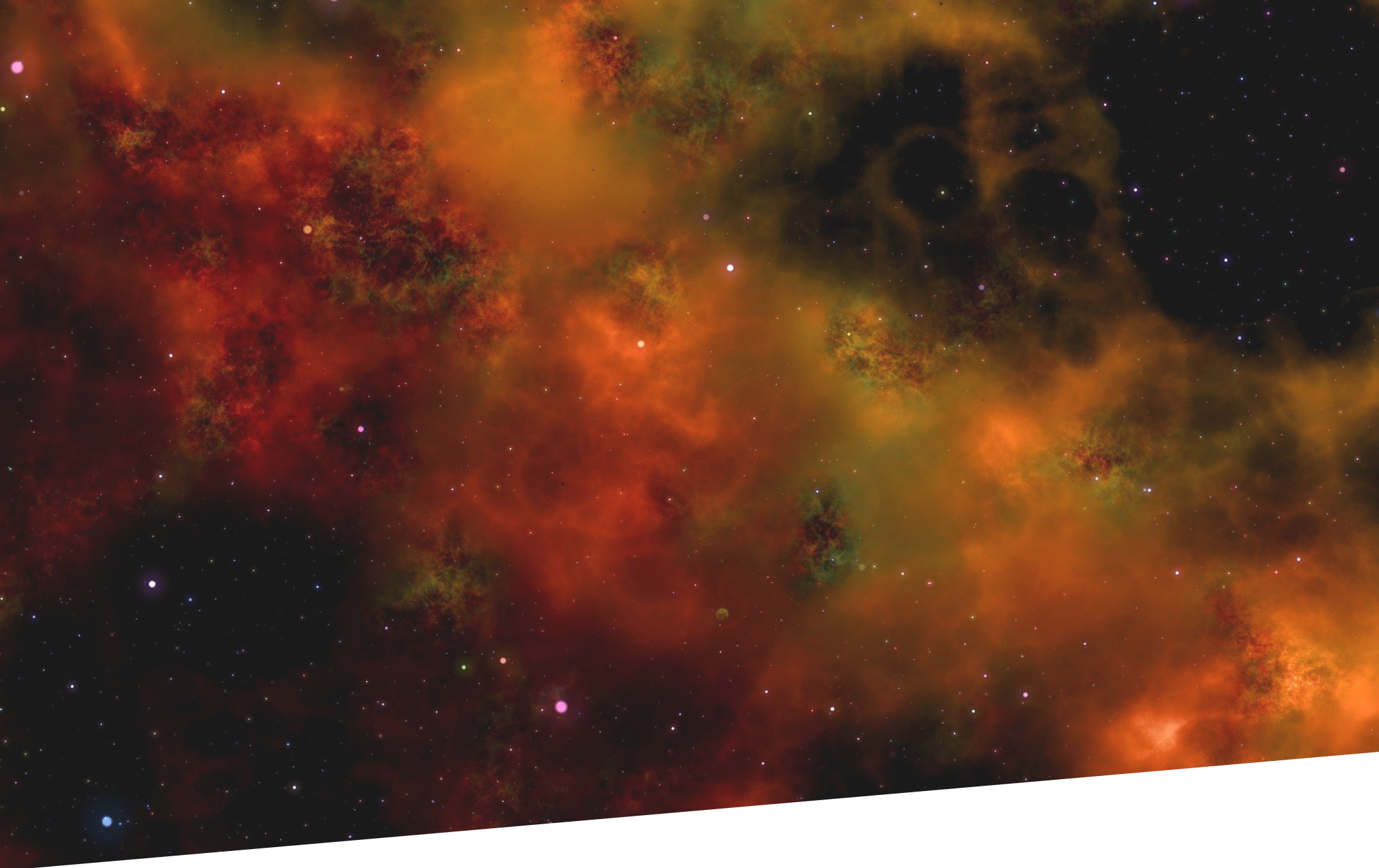
By John Santaferaro





<b>Table of Contents</b>	<b>3</b>	Introduction to Modernization and Hybrid Data Ecosystems
	<b>5</b>	What is a Hybrid Data Ecosystem?
	<b>7</b>	Modernization Priorities
	<b>8</b>	Modernization Priorities
	<b>11</b>	Hybrid Data Ecosystem Trends
	<b>13</b>	Hybrid Data Ecosystem Strategies
	<b>22</b>	Research Methodology and Demographics
	<b>23</b>	Methodology
	<b>23</b>	Demographics





# Introduction to Modernization and Hybrid Data Ecosystems

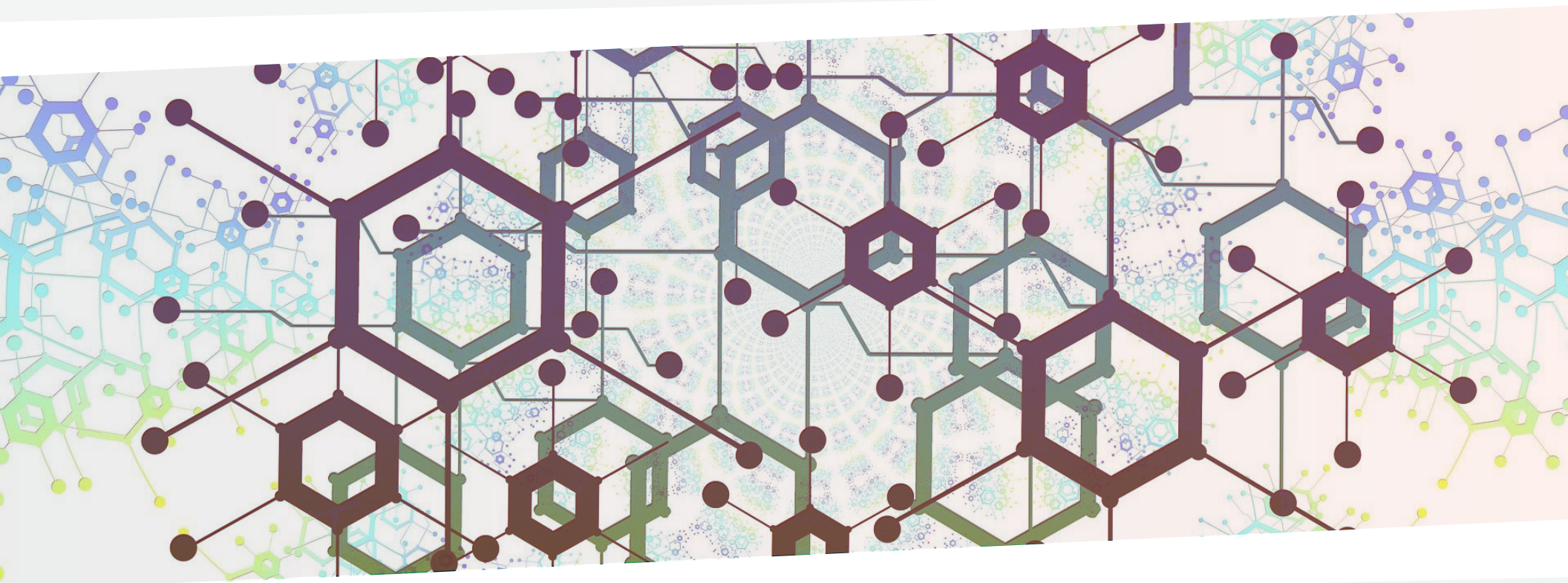


Digital strategies were supposed to transform, and the cloud was supposed to simplify, everything. Instead, digital investments bring new data types that are extremely valuable, but incompatible with structured data. The cloud and SaaS bring a complex network of data stored across multiple cloud and SaaS providers. As a result, most organizations manage a hybrid data ecosystem, a complex mix of different data management platforms positioned in different locations on-premises and in different clouds.

EMA hybrid data ecosystem research in early 2019 revealed that 32% of organizations operate more than five different platforms. For large organizations, analytical databases or data lake platforms can have as many as 700 instances. The addition of every new platform or instance carries additional cost and

complexity. Proven expansion and consolidation strategies are critical to the success of every analytics program.

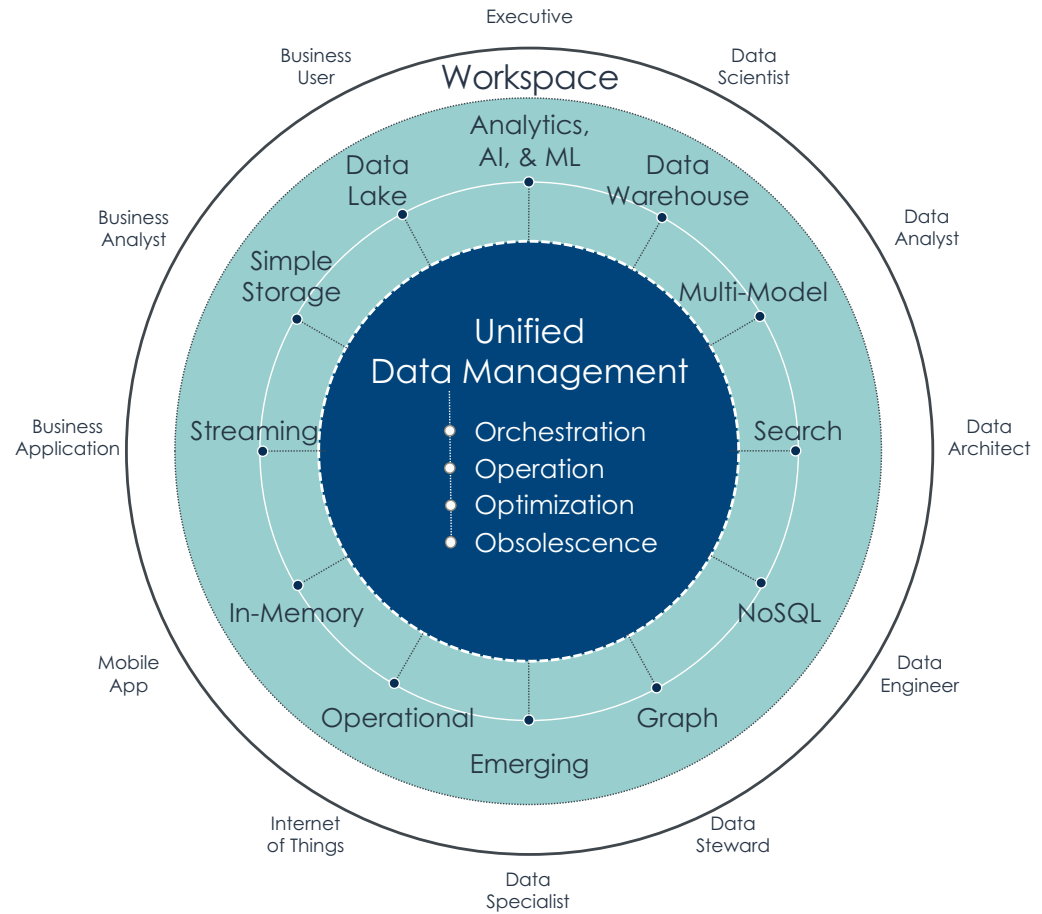
2020 created even more urgency for efficient management of hybrid data ecosystems. The global pandemic created a digital reset, causing leaders to immediately mandate a quest for digital superiority. Analytics fueled by the need for immediate, intelligent responses to business events is at the core of every digital strategy. The greatest challenge to immediacy comes from the fact that most organizations lack operational maturity, and they are struggling to navigate their hybrid data ecosystems.



# What is a Hybrid Data Ecosystem?

A single platform that can handle all the different digital workloads required by the business ecosystems does not exist. In the real world, most organizations operate more than five platforms. For large organizations, critical platforms like analytical databases can have more than 700 instances. A hybrid data ecosystem (HDE) is the full set of various platforms an entity manages.

For example, an insight-driven entity might run critical applications on operational databases and their mobile or internet apps on NoSQL distributed databases. Multifaceted analytics and business intelligence programs require a combination of analytic databases, data warehouse platforms, and data lake platforms. Because of cost constraints, IT leaders are moving data onto simple storage or deploying multi-model platforms for consolidation. In order to facilitate intelligent responses to real-time engagement, search platforms speed the discovery of insight and streaming platforms become the data nervous systems of HDEs for both integration and analytics. In-memory databases are changing the entire emerging data landscape with multi-use capabilities. Emerging platforms address specialized use cases, like graph databases for analyzing networks of networks and complex ecosystems like markets or the human body.



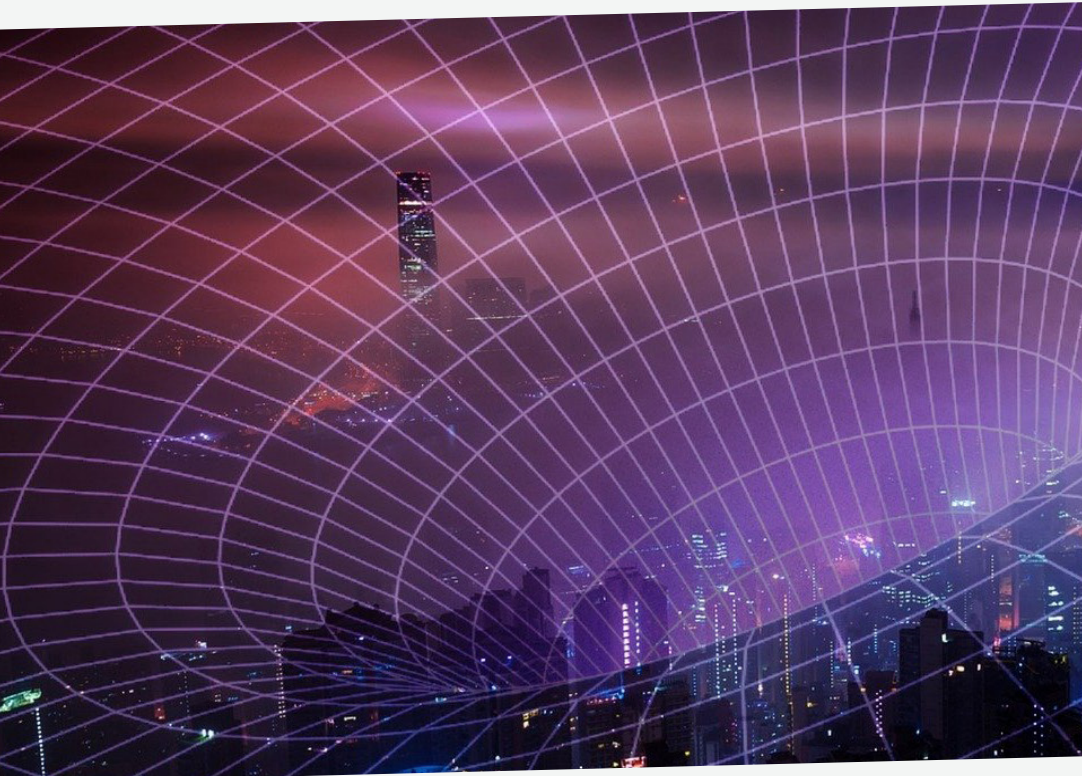


Developed, refined, and supported with end-user insights since 2012, the EMA hybrid data ecosystem architecture provides guidance for insight-driven organizations. In order to achieve deep interoperability, EMA recommends working toward a unified data management strategy, represented in the center of the diagram.

- Orchestration defines how to acquire, prepare, provision, and adapt data for consistent use across the entire data landscape.
- Operation is the delivery, administration, security, availability, and automation of all platforms within the ecosystem.
- Optimization is the process of continually improving operations and orchestration.
- Obsolescence is the means by which legacy systems reach end of life and migrate to new systems.

Surrounding unified data management are the HDE data platforms, including space for emerging data platforms. Shared workspace includes data access via a broad range of end-user platforms. Additionally, between the data and users is technology for data security and privacy, collaboration, storytelling, and sharing insight.

The most important part of an HDE is outside the technology circle: the consumers. These are the people, processes, and machines that utilize the information resident within an HDE. Development of new services takes place from the outside in, with business requirements driving all portfolio management. Delivery takes place from the inside out, with technology innovation leading the way.



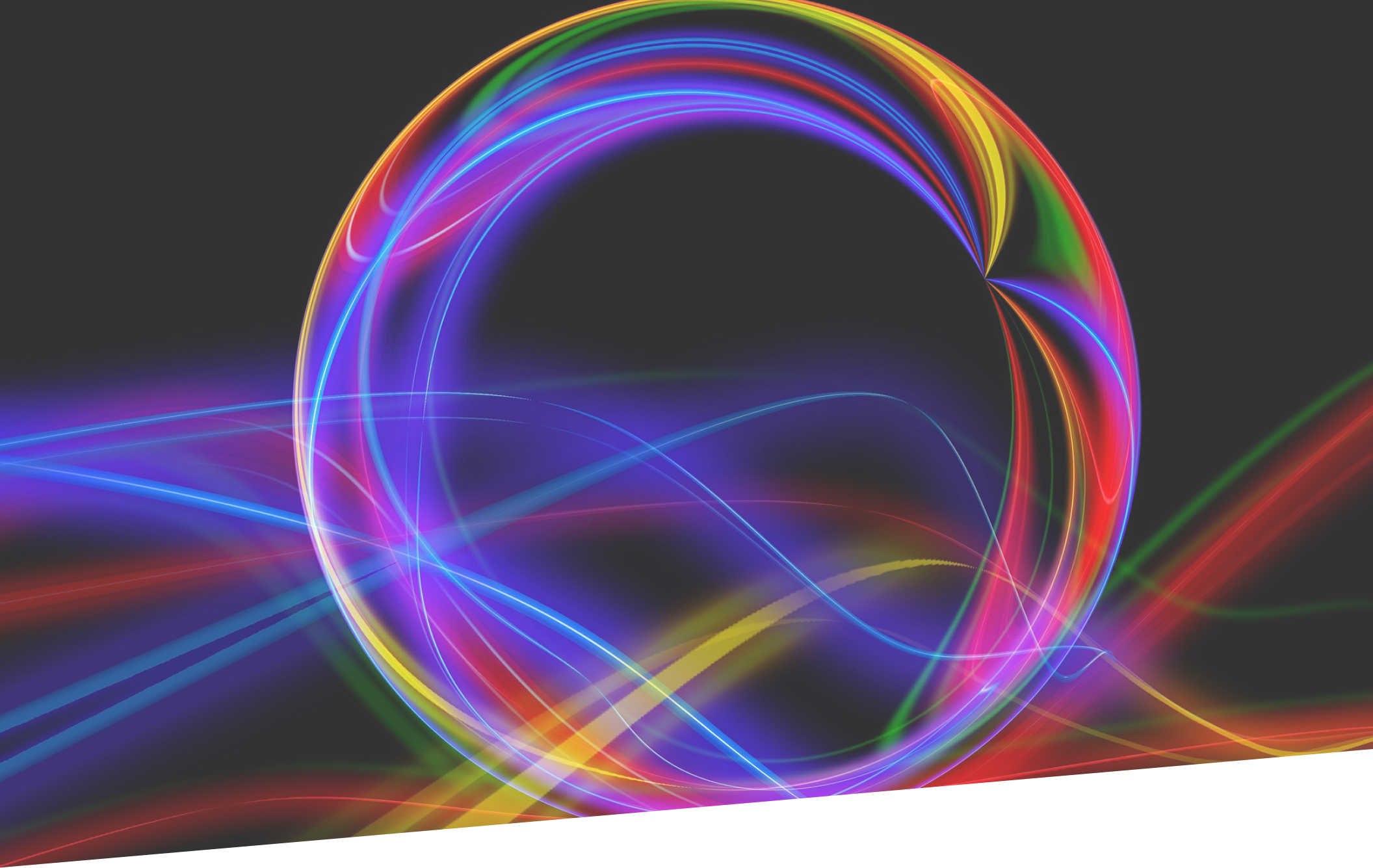
## Hybrid Data Ecosystem Highlights

**47%** of participants have more than five data platforms in their hybrid data ecosystem

**53%** of participants are either expanding or consolidating their number of platforms in the next 12 months

**Analytic platforms** and **data warehouse platforms** are the #1 and #2 platforms in HDEs

**36%** of participants plan to add a data lake platform in the next 12 months



# Modernization Priorities



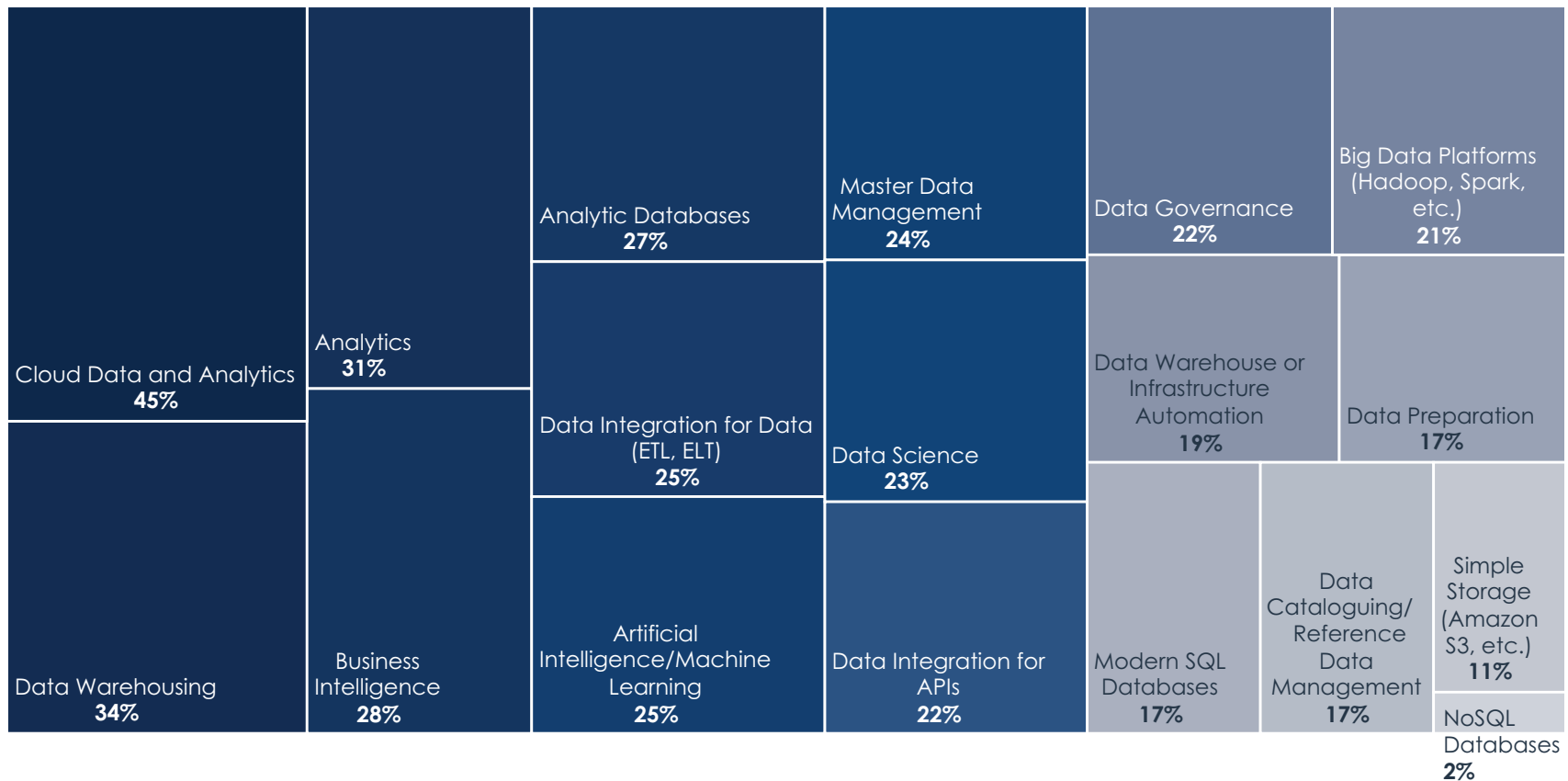
# Modernization Priorities

## Overall Analytics and Data Management Priorities

Cloud data and analytics came in as the number-one priority for modernization in 2021. Analytics and data warehousing continue to be top priorities for modernization, as they were in EMA's previous HDE study in 2019. This indicates

that analytics are core to the business and data warehousing maintains its position as a trusted source for mission-critical analytics.

Which of the following technologies are most important to your modernization efforts?



Sample Size = 206, Valid Cases = 206, Total Mentions = 842



## Data Warehouse and Data Lake Priorities

In 2021, EMA expects a convergence of the data warehouse and the data lake into a unified analytics warehouse (UAW). This new multi-use data platform is unified because it adequately handles multi-structured, multi-latency data in a single platform. It is an analytics platform because the primary use case for both the data lake and the data warehouse has always been analytics. It is a warehouse because it stores multi-structured data in an organized and accessible manner.

Cloud data warehousing clearly beat out cloud data lakes as the number-one modernization priority, with 36% more participants selecting the data warehouse over the data lake. Data sharing was the number-two choice, indicating

that this new concept has been popularized in the last two years. Just under 34% selected data sharing as a top priority.

In addition, with an increase in multi-cloud and hybrid deployments, interoperability and portability across different deployment options both scored high. EMA sees a definite advantage with platform vendors that have a single software solution, no matter where it is deployed.

**36%** more participants chose cloud data warehousing as a priority compared to those who chose cloud data lakes.

Which of the following modern capabilities is most important to your data warehouse and data lake programs?

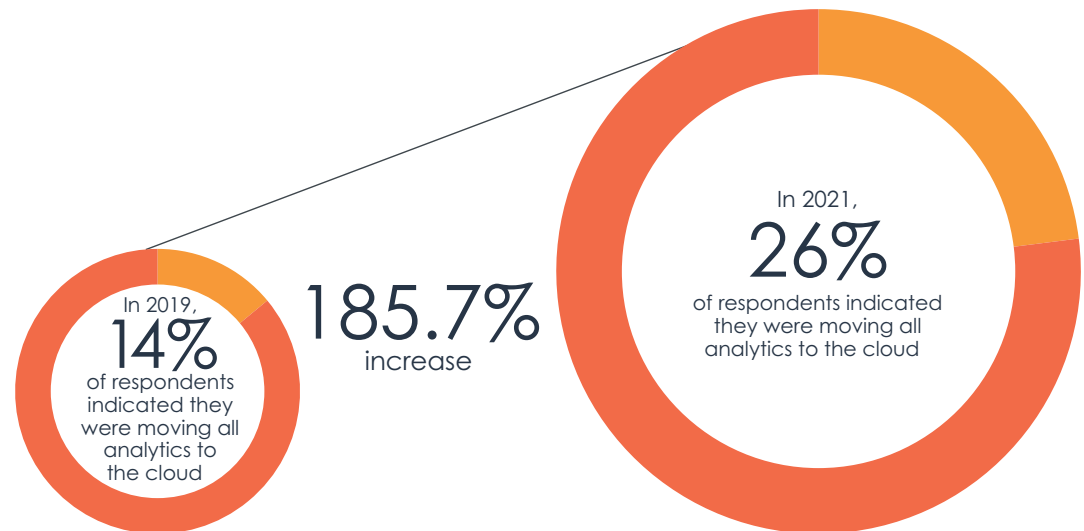
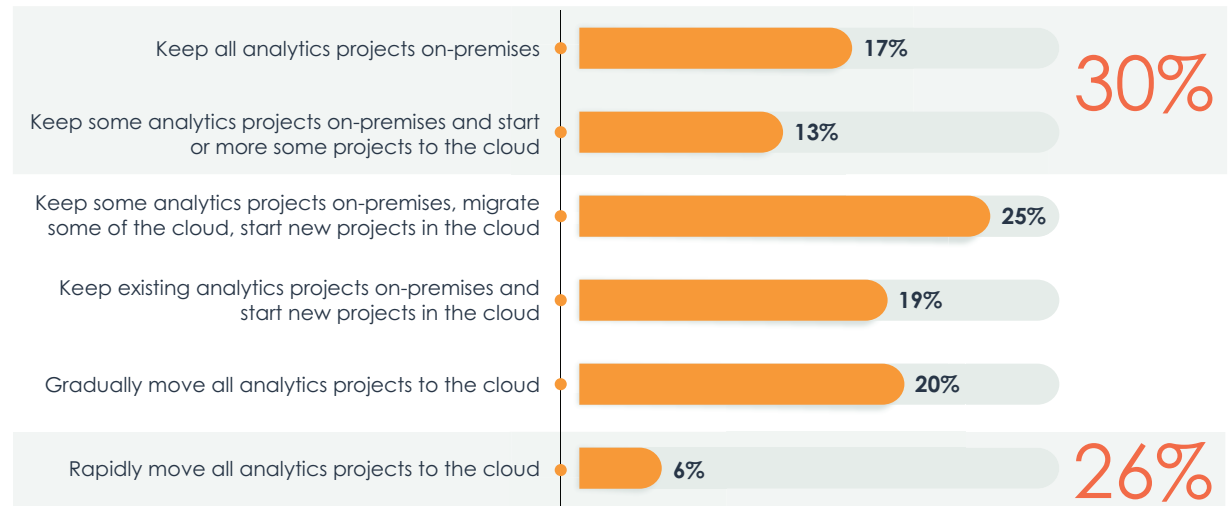
Cloud data warehousing <b>45%</b>	Data sharing <b>33%</b>	Unified security and privacy <b>26%</b>	Automation <b>24%</b>	Support for multi-structured data <b>23%</b>	Unified analytics - combined data warehouse and data lake <b>20%</b>		
		Multi-cloud and hybrid deployment <b>25%</b>			Multi-workload support (operational, analytical, streaming, etc.) <b>17%</b>	Interoperability with other platforms and applications <b>14%</b>	Multi-model adaptability <b>14%</b>
Cloud data lakes <b>33%</b>	Technology portability across deployment options <b>26%</b>	Analytics embedded in the data platform <b>24%</b>	Data gravity <b>18%</b>	Multidimensional agility (new data, new queries, new users, etc.) <b>18%</b>	Delivery acceleration <b>16%</b>	Automated performance tuning <b>13%</b>	Preinstalled business logic <b>4%</b>

Sample Size = 206, Valid Cases = 206, Total Mentions = 807

## Cloud Analytics and Data Management Priorities

Since EMA first conducted this research in 2019, there has been a major shift toward the cloud as the strategic platform for analytics and data management. In 2019, 47% of all participants wanted to keep a significant portion of their analytics on-premises. In 2021, that number dropped to 30%. In 2019, a very small percentage of participants (14%) were looking to move all analytics and data management to the cloud. In 2020, that number increased to 26%. That is an increase of almost double.

Which of the following best describes your cloud strategy as it relates to analytics and data management?



Sample Size = 206





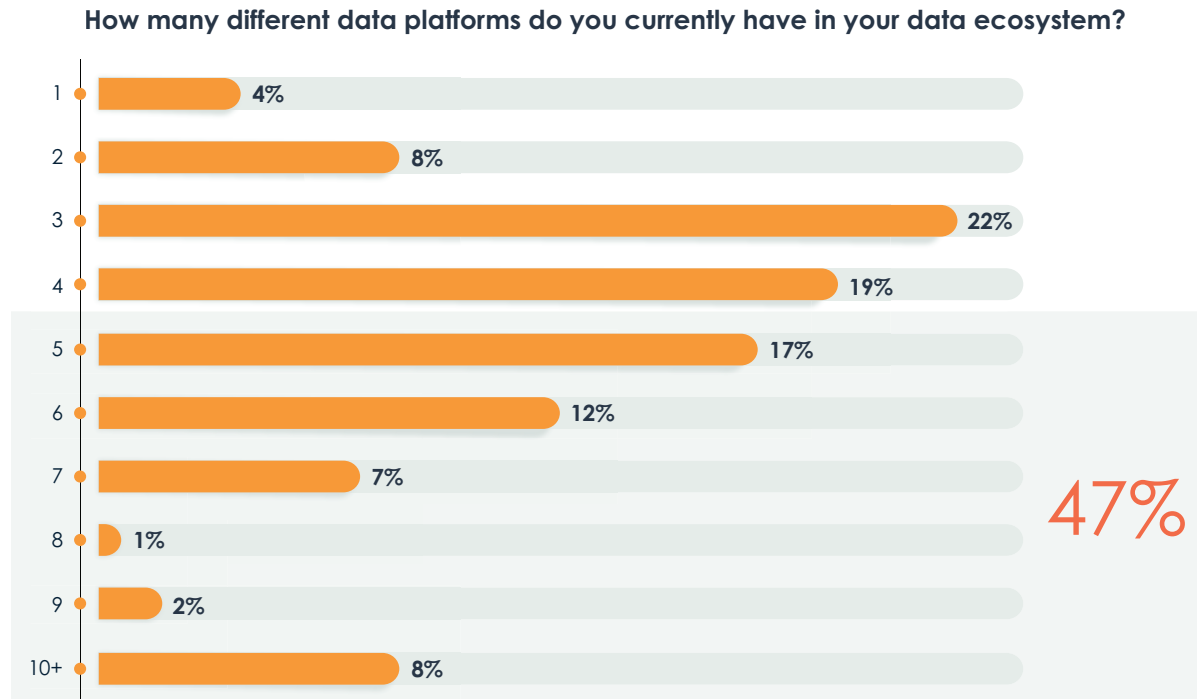
# Hybrid Data Ecosystem Trends



The need to drive insight from new types of data that were not well-supported in relational databases drove much of the initial phase of modernization, adding data lake platforms to the data warehouse and analytical platforms. Since that time, analytical workloads have also grown more complex. As a result, many organizations continue to add special use platforms like graph platforms, or time-series and NoSQL databases.

Today, 47% of participants indicated that they currently have more than five platforms in their hybrid data ecosystem. A full 8% have 10 or more platforms under management, and the number continues to grow. EMA asked participants how many platforms they planned to have in the next 12 months. The results show that there is a continual move toward more platforms. The percentage of participants with more than five platforms is expected to grow from 47% to 52%.

**47%-52%**  
 The number of participants with more than five platforms in their hybrid data ecosystem will grow from 47% to 52% in the next 12 months.



Sample Size = 206



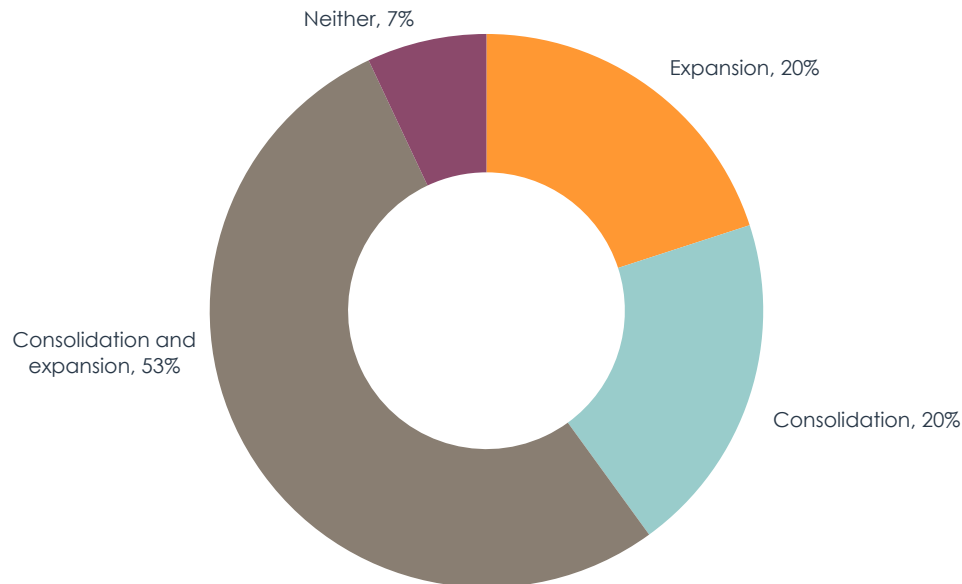
## Hybrid Data Ecosystem Strategies

In 2021, data ecosystems will continue to be in flux. Only 7% of participants indicated that they are neither adding nor consolidating platforms. More than half of all participants (53%) indicated that they are both expanding and consolidating platforms. An additional 20% are expanding the number of platforms, and equally, 20% are consolidating.

EMA believes that the next five years will continue to produce great opportunities for services and consulting firms, especially for those that have very targeted services for both expansion and consolidation. In addition, the opportunity is ripe for platform vendors to expand their unified analytics offerings to help consolidate even more specialized platforms.

**93%** of participants have ecosystems that are in flux, either expanding or consolidating.

Does your data ecosystem growth strategy move toward consolidation or expansion in the number of platforms?



Sample Size = 206

## Best Practices for Expansion

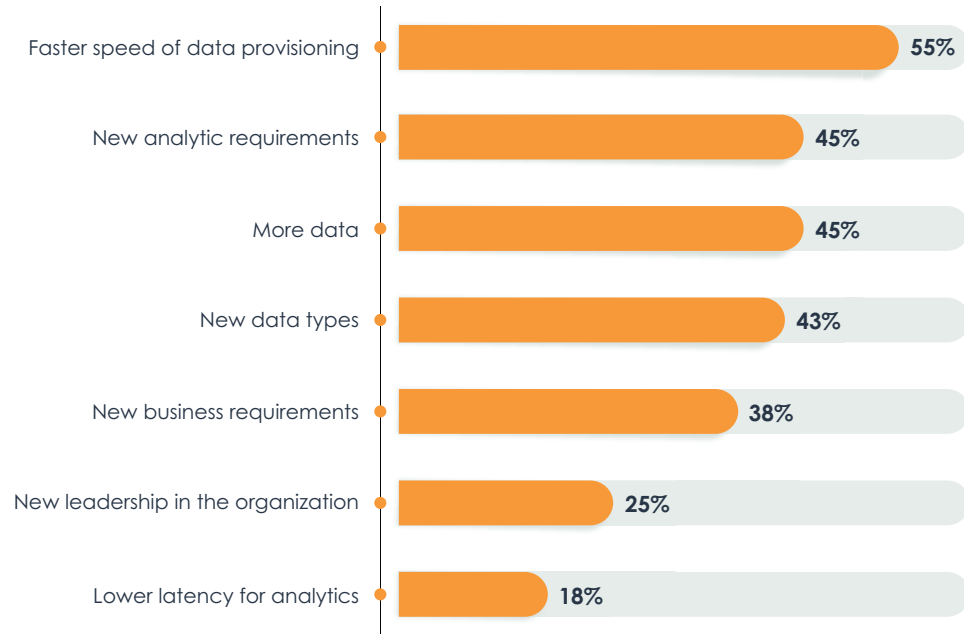
### 01. Expand platforms for faster time to decisions and more competitive analytical use cases.

In 2021, speed matters, especially considering the uncertainty brought about by the pandemic. Fifty-five percent of participants indicated they are adding new data platforms to their environment to gain a faster speed of data provisioning. The second-highest

choice among participants was new analytic requirements, with 45%. Since the highest-value use of data is for analytics, both choices support the need for insight-driven decisions.

**SPEED and ANALYTICS**  
drive hybrid data ecosystem expansion.

### If your data ecosystem strategy is to expand and add more data platforms, what is driving the need to expand?



Sample Size = 152, Valid Cases = 152, Total Mentions = 410



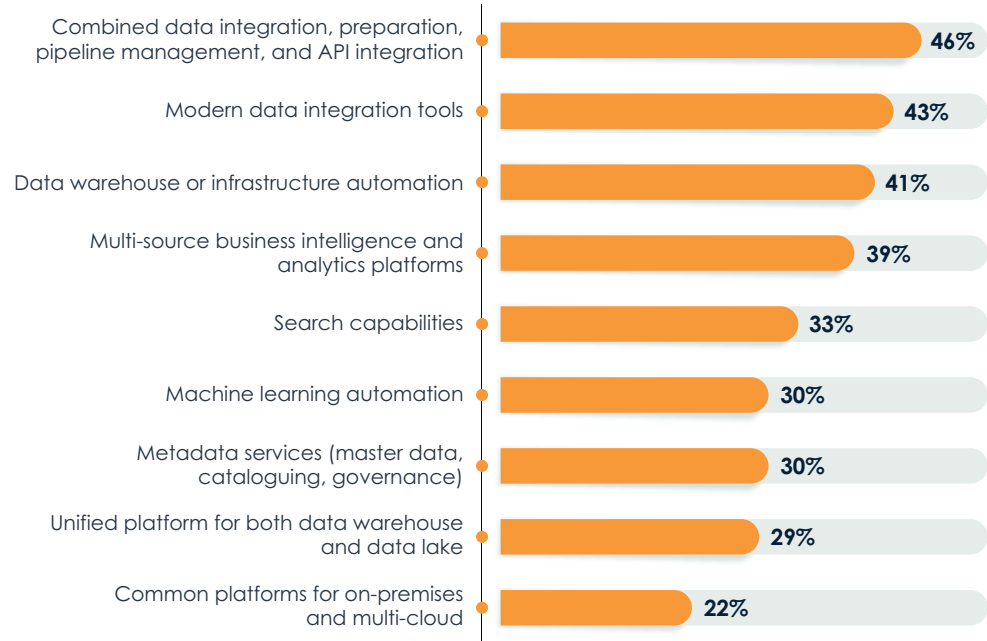
**02.** HDE expansion requires the implementation of modern, multi-modal data integration.

The top two responses to this question on navigating complex data ecosystems demonstrate a major shift in IT strategy from integration to interoperability. Modern data pipeline automation platforms replace what was formerly 4-5 integration platforms, including data integration, pipelines, preparation, and API integration. Forty-six

percent of participants plan to move this direction, consolidating integration platforms to manage even more data platforms. Forty-three percent are likely identifying these multi-modal integration tools as modern data integration tools.

**46%** of participants plan to use multi-modal, combined data integration to manage the expanding complexity of HDEs.

**If your data ecosystem strategy is to expand and add more data platforms, how do you plan on managing the expanding complexity of your environment?**



Sample Size = 152, Valid Cases = 152, Total Mentions = 477

## Best Practices for Consolidation

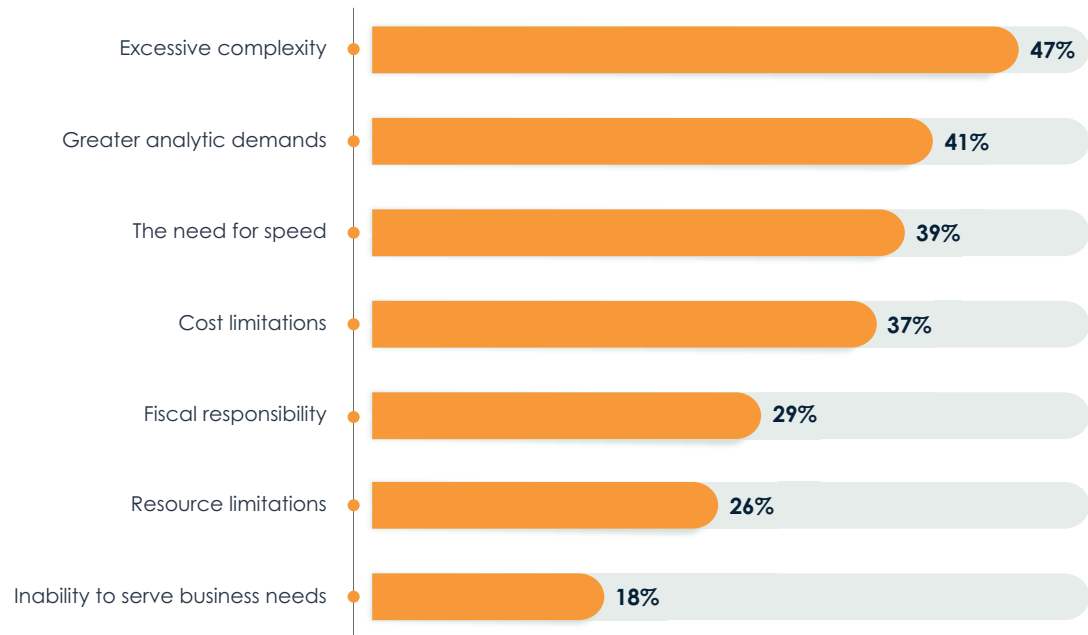
### 01. Track cost and time as a baseline to measure real savings from consolidation.

IT organizations are tired of the complexity necessary to manage hybrid data ecosystems, especially during a time when the business continues to demand more data and more analytics than ever before. The complexity of multiple systems slows down the speed at which data and analytic professionals can operate and increases the overhead necessary to provide analytics for the business. It is not surprising that the top four drivers for consolidation are tightly knit.

Forty-seven percent are driven to consolidate because of excessive complexity in their current systems. Forty-one percent indicated they are consolidating because of greater analytic demands from the business. The combination of these two foremost drivers leads directly to a slowdown of response

times and an increase in costs, therefore, the need for speed and cost limitations as the third and fourth drivers for consolidation. EMA recommends tracking a baseline of platform costs to better understand real savings from consolidation.

### If your data ecosystem strategy is to consolidate and reduce the number of data platforms, what is driving the need for consolidation?



**47%** of respondents are consolidating because of excessive complexity in their current systems.

Sample Size = 150, Valid Cases = 150, Total Mentions = 357

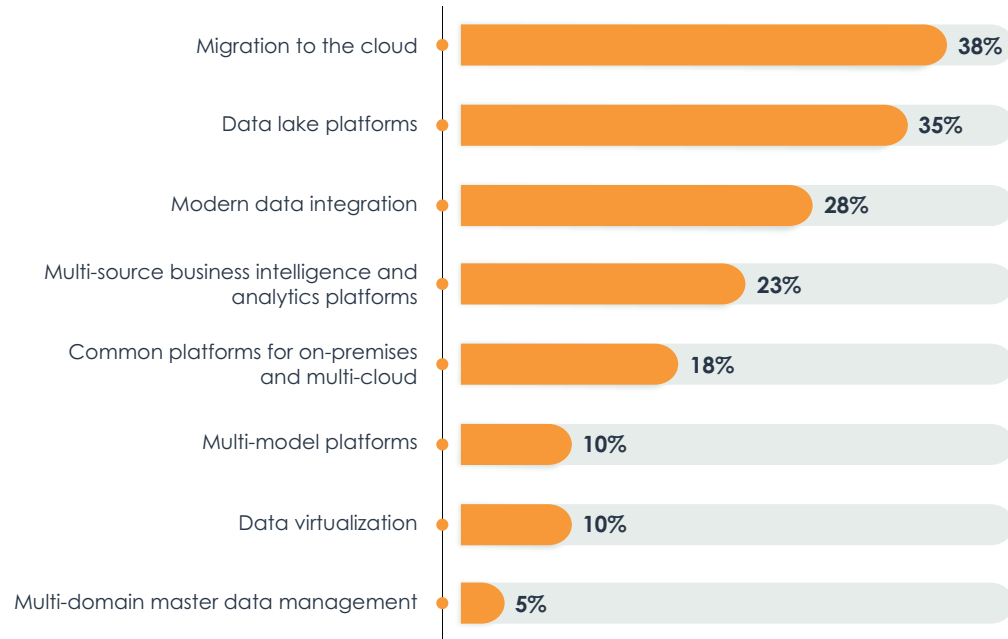
**02.** Consider a cloud platform capable of handling multi-structured data to consolidate legacy systems.

Migration to the cloud and data lake platforms are the primary choices for consolidation, with 38% and 35%, respectively. Many leaders perceive the cloud as a place to reduce cost and the data lake as a place to analyze multi-structured data.

Not far behind the frontrunners, with 28% and 23%, are modern data integration and multi-source business intelligence and analytic platforms. Both technologies work to connect diverse data platforms, indicating that while consolidation use cases may continue to increase in 2021, most organizations will still need to facilitate interoperability among their platforms.

In addition, while migration to the cloud is the number-one answer for consolidation, the cloud is not always the answer. EMA has seen a wave of people who went to the cloud thinking it would reduce costs, but they are now moving some analytical projects back on-premises, especially as container technology matures and makes the private cloud more tenable.

**If your data ecosystem strategy is to consolidate and reduce the number of data platforms, how do you plan to consolidate?**



Sample Size = 40, Valid Cases = 40, Total Mentions = 67



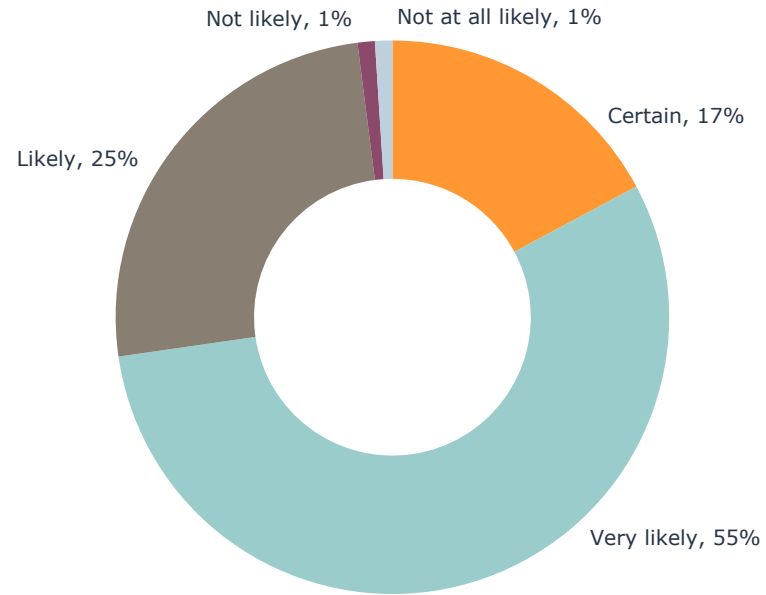
### 03. Begin planning for the adoption of multi-model platforms.

EMA sees a significant rise in the adoption of multi-use and multi-model platforms, like in-memory data stores or analytical platforms

with intelligent query routing capabilities. EMA asked participants about the likelihood of using these innovative platforms. A surprising 98% said they would likely use these platforms, with 55% of respondents indicating they are very likely to consider their use.

**98%** of consolidation respondents indicated they are likely to use multi-use or multi-model platforms to consolidate their hybrid data ecosystems.

As multi-use and multi-model platforms grow in market acceptance, how likely are you to consider using these platforms to consolidate your data ecosystem?



Sample Size = 412

## Important Considerations for Multi-Model Platforms

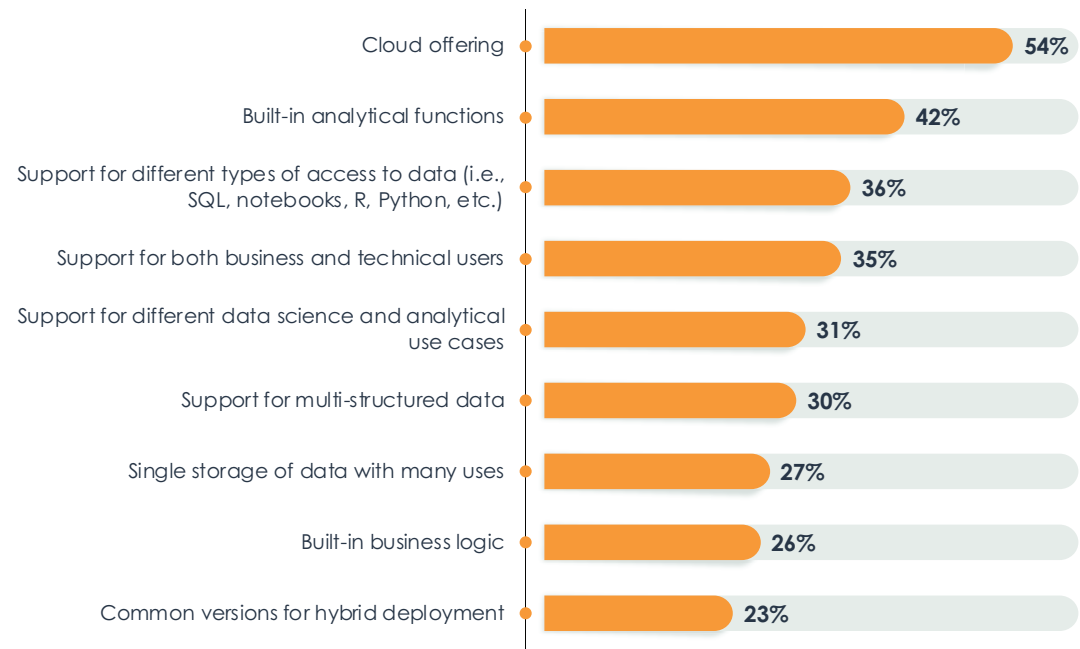
Because of the rising importance of multi-model platforms, EMA wanted to better understand which aspects of these new platforms were most important to users and buyers. New unified analytics platforms are enabling the analysis of both multi-structured and multi-latency data. In addition, they span use cases for business and technical users using a broad set of tools to access the data and understand it from different perspectives. Consider the following three recommendations based on data collected.

### 01. Consider a move to the cloud with multi-use analytics for technical and business users.

While 54% of respondents indicated that a cloud offering is a high priority for the selection of a multi-model platform, it is equally impressive, in this age of cloud migration, that 46% of respondents do not see a cloud offering as essential. EMA suspects that discerning buyers are still somewhat wary of cloud-only vendors and are looking to consolidate on multi-use platforms that will provide solutions for both cloud and on-premises needs.

In addition, because most organizations are already growing their internal analytics and data science practices, 42% want a solution that already has built-in analytical functions to support their existing projects. Support for different types of access for both business and technical users is also highly important, especially as business users become increasingly proficient in analytics.

**As you consider the purchase of multi-use and multi-model platforms, what capabilities are most important to you?**

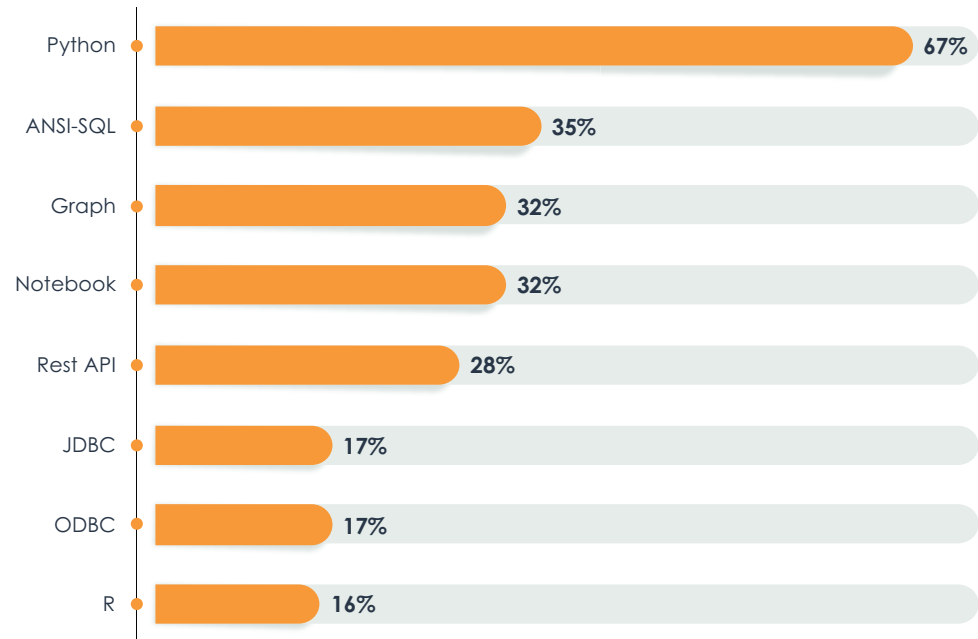


Sample Size = 206, Valid Cases = 206, Total Mentions = 629

**02.** Plan for the massive Python wave, but maintain your commitment to SQL.

When it comes to access to multi-use data platforms, Python has become the dominant player and the once-dominant, R, has fallen out of favor. Python came in at 67% almost two times—more than the nearest response, ANSI-SQL, which received a 35% response. While the temptation may be to pour all your efforts into Python, there are still far more SQL users and SQL-based access tools than any other mode of access.

Which of the following access types would you want supported on your multi-use, multi-model platforms?



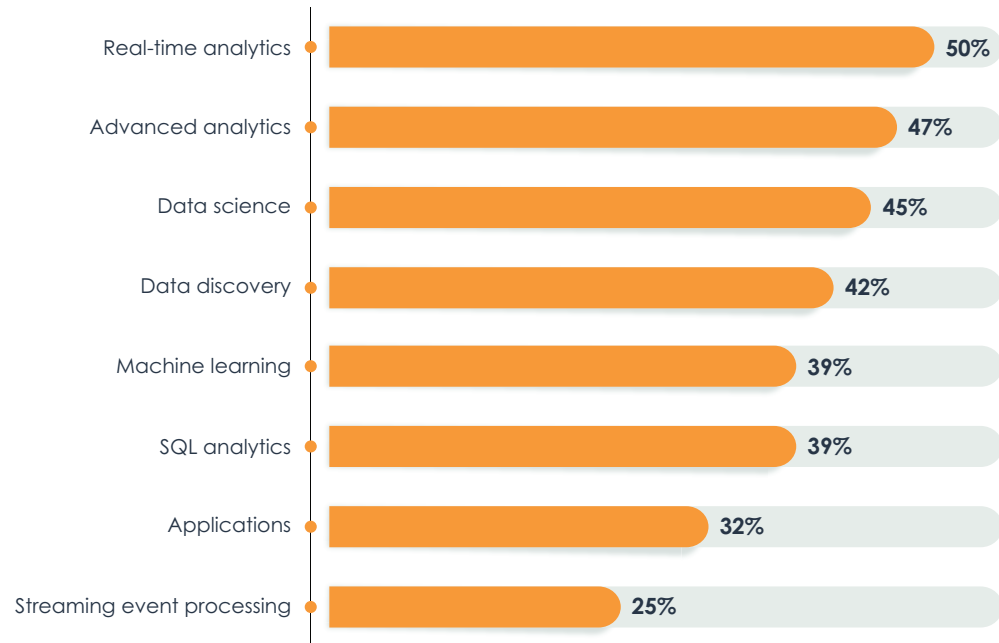
Sample Size = 206, Valid Cases = 206, Total Mentions = 502



**03.** Push for the realm of real time and build for the continued growth of advanced analytics and data science.

Global shutdowns forced almost every business to be digital, making the quest for digital superiority universal. As a result, immediate intelligent responses to digital events is now more important than ever. In the number-one response, 50% of participants now find it necessary for multi-use platforms to support real-time analytics. Along with real time, these modern platforms must also support advanced analytics and data science.

**Which of the following workload types would you want supported on your multi-use, multi-model platforms?**



Sample Size = 206, Valid Cases = 206, Total Mentions = 711



# Research Methodology and Demographics

## Methodology

All research results in this report are based on EMA’s survey of 206 randomly selected North American enterprise and midmarket data and analytics professionals.

## Demographics

The following is a demographic overview of 206 randomly selected data, analytics, and business professionals. The first question was used as a primary qualification. If the survey candidate did not indicate awareness or involvement in any of the data and analytics activities in that question, they were not included in the survey results.

### Company Size by Number of Employees

250-499	<b>30</b>
500-999	<b>31</b>
1,000-2,499	<b>44</b>
2,500-4,999	<b>34</b>
5,000-9,999	<b>33</b>
10,000-19,999	<b>10</b>
20,000 or more	<b>24</b>

### Role in the Organization

Database Administrator	<b>36</b>
Data Engineer	<b>29</b>
Data Analyst	<b>54</b>
Data Architect	<b>30</b>
Data Scientist	<b>28</b>
Data Steward	<b>10</b>
Systems Administrator	<b>21</b>
Business Intelligence Specialist	<b>11</b>
Data Integration/Preparation Specialist	<b>34</b>
Data Warehousing Leader	<b>20</b>
Analytics Leader/Executive	<b>17</b>
IT Specialist	<b>57</b>
IT Leader/Executive	<b>70</b>
IT Administrator	<b>55</b>
CIO/CTO/CAO/CDO	<b>29</b>
Business Analyst	<b>17</b>
Business User	<b>10</b>
Business Line Leader	<b>8</b>
Business Executive	<b>18</b>



## Industry

Aerospace/Defense	<b>2</b>
Consulting: Computer- or Networking-Related	<b>2</b>
Consulting: All Other (Not Computer- or Networking-Related)	<b>1</b>
Education	<b>8</b>
Finance/Banking/Insurance	<b>21</b>
Government	<b>6</b>
Healthcare/Medical	<b>13</b>
Life Sciences (Pharmaceutical, MedTech, BioTech)	<b>2</b>
High Technology: Software	<b>41</b>
High Technology: Reseller/VAR/Systems Integrator	<b>2</b>
High Technology: Application/Internet/Managed/Network Service Provider	<b>14</b>
Manufacturing: Computer Hardware- or Networking-Related	<b>5</b>
Manufacturing: All Other (Not Computer- or Networking-Related)	<b>17</b>
Marketing/Advertising/PR Agency/Market Research	<b>2</b>
Media/Entertainment: Publishing/Broadcasting	<b>2</b>
Nonprofit/Not for Profit	<b>1</b>
Oil/Gas/Chemicals	<b>2</b>
Professional Services: Computer- or Networking-Related	<b>16</b>
Professional Services: All Other (Not Computer- or Networking-Related)	<b>4</b>
Retail/Wholesale/Distribution	<b>18</b>
Telecommunications	<b>7</b>
Transportation/Airlines/Trucking/Rail	<b>8</b>
Utilities/Energy	<b>3</b>

## Annual IT budget

Less than \$1 Million	<b>7</b>
\$1 Million to less than \$5 Million	<b>20</b>
\$5 Million to less than \$10 Million	<b>30</b>
\$10 Million to less than \$25 Million	<b>37</b>
\$25 Million to less than \$50 Million	<b>41</b>
\$50 Million to less than \$100 Million	<b>27</b>
\$100 Million or more	<b>35</b>

# About Teradata

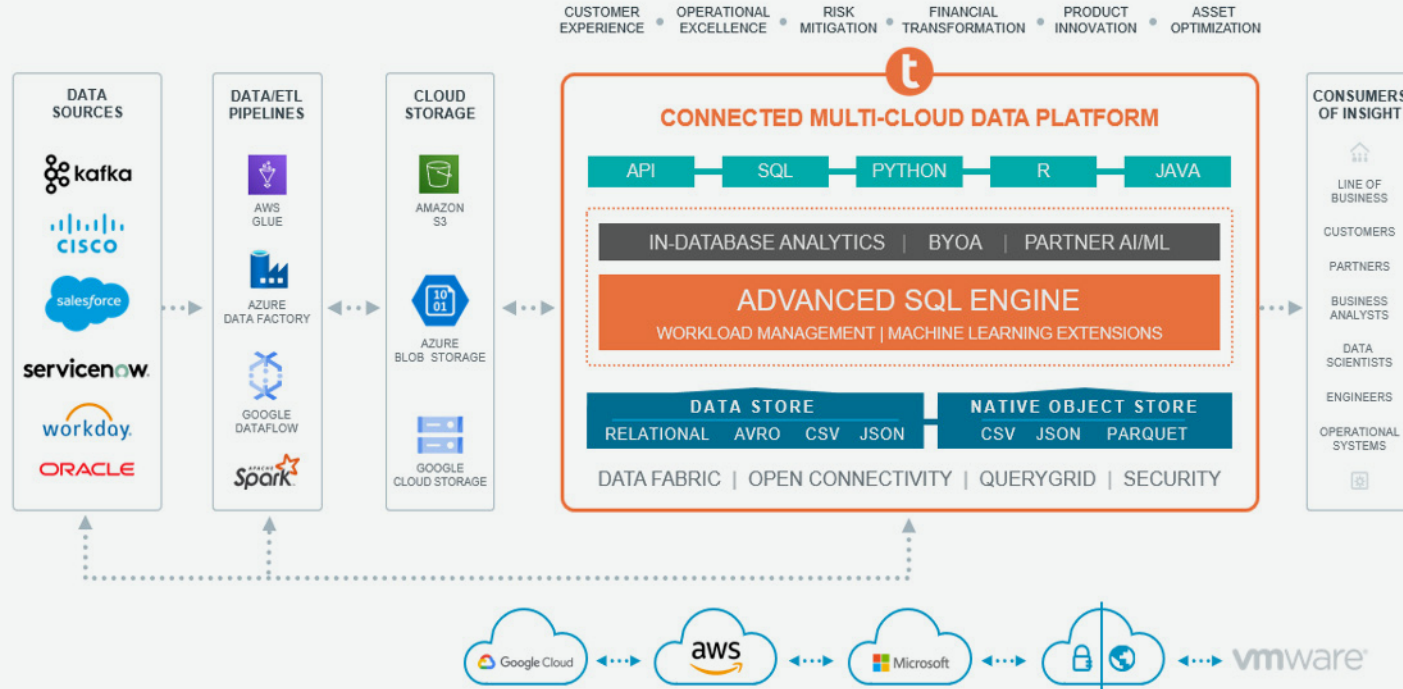
Teradata is the connected multi-cloud data platform company for enterprise analytics, solving data challenges from start to scale. Only Teradata gives you the flexibility to handle the massive and mixed data workloads of the future, today.

The Teradata Vantage architecture is cloud-native, delivered as a service, and built on an open ecosystem. Vantage takes advantage of the core underlying capabilities in the cloud, including native object store and separation of compute and storage. These design features make Vantage the ideal platform to optimize price/performance in a multi-cloud environment.

Our open approach embraces the modern ecosystem to create a seamless experience for ingestion, exploration, development, and operationalization. We provide the fastest path to analytics in the cloud at scale with the lowest risk. Our experience working with thousands of customers and partners around the world, across a wide range of verticals and industries, makes us the most effective platform for delivering business outcomes and unlocking unlimited value by turning data into your greatest asset. Learn more at [Teradata.com](https://www.teradata.com).

## Teradata Vantage

The Connected Multi-Cloud Data Platform for Enterprise Analytics











**25**  
YEARS

#### 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](http://www.enterprisemanagement.com). You can also follow EMA on [Twitter](#) or [LinkedIn](#).

---

This report, in whole or in part, may not be duplicated, reproduced, stored in a retrieval system or retransmitted without prior written permission of Enterprise Management Associates, Inc. All opinions and estimates herein constitute our judgement as of this date and are subject to change without notice. Product names mentioned herein may be trademarks and/or registered trademarks of their respective companies. "EMA" and "Enterprise Management Associates" are trademarks of Enterprise Management Associates, Inc. in the United States and other countries.

©2021 Enterprise Management Associates, Inc. All Rights Reserved. EMA™, ENTERPRISE MANAGEMENT ASSOCIATES®, and the mobius symbol are registered trademarks or common law trademarks of Enterprise Management Associates, Inc.