Trends in Data Management
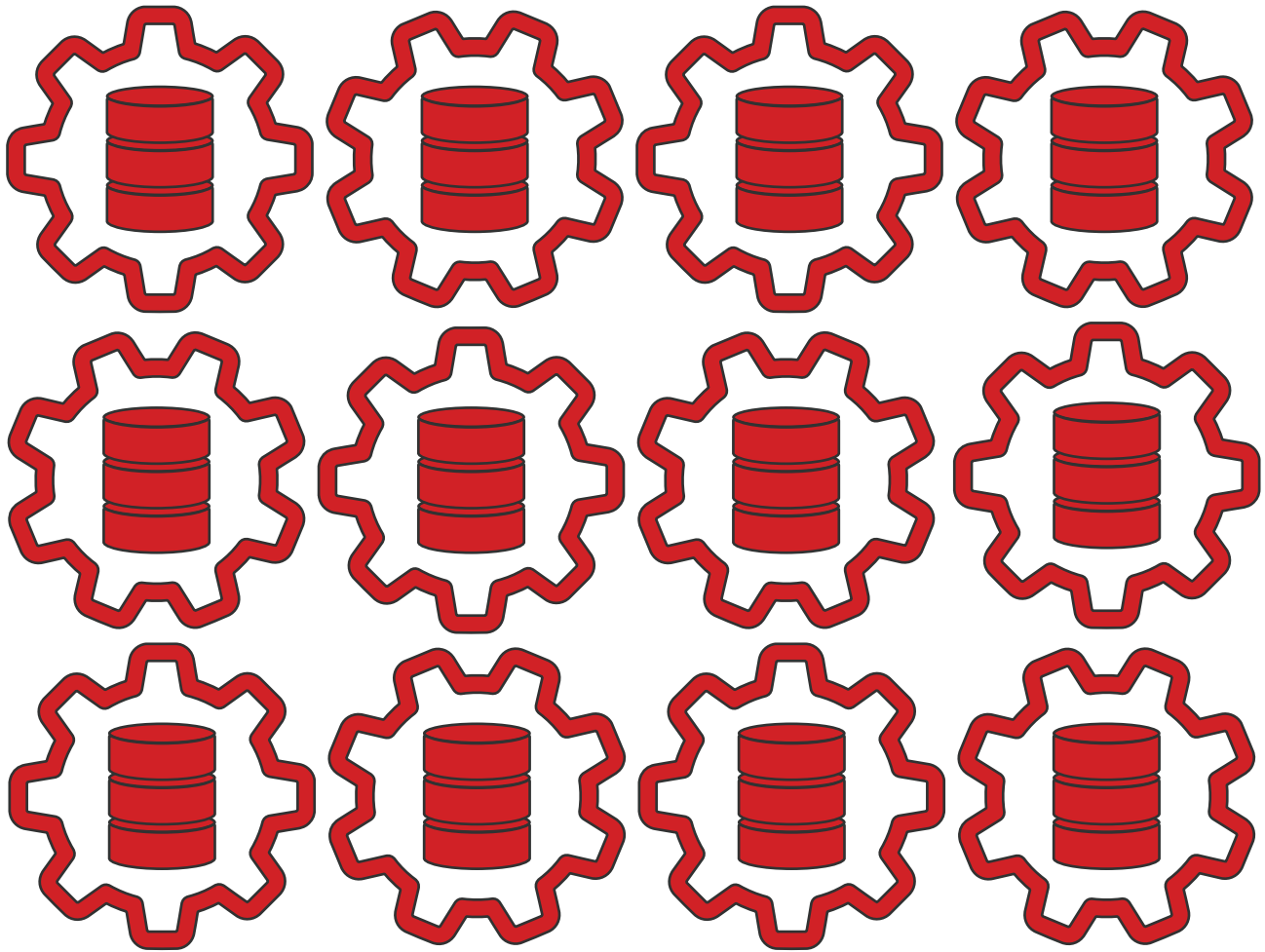## A 2019 DATAVERSITY® Report

by **Donna Burbank** and **Michelle Knight**

Sponsored by:

IDERA
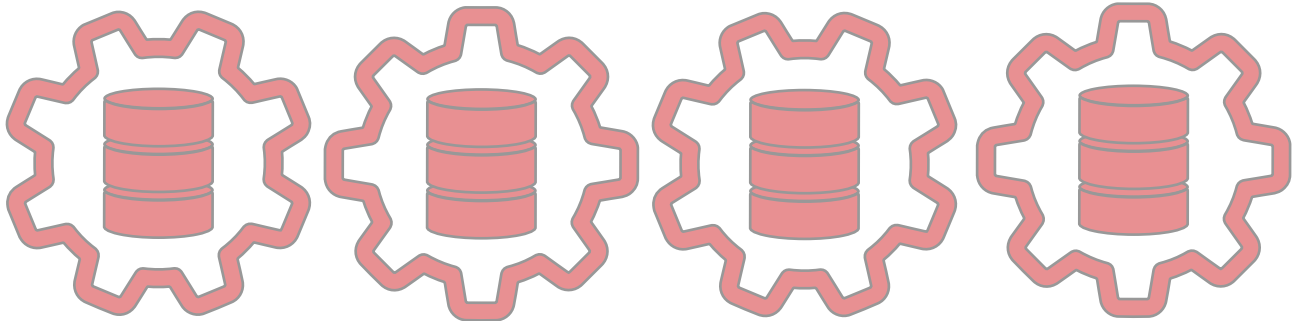
Produced by:

DATAVERSITY®

TABLE OF CONTENTS

1. EXECUTIVE SUMMARY

In today's fast-paced business environment, more and more organizations are looking to become data-driven, and with this comes a renewed interest in Data Management. In such dynamic times, Data Management becomes increasingly important both in "offense" mode – while driving business success and growth – and in "defense" mode – while protecting organizations against risk.

Due to the increased interest in data, more stakeholders across the organization are leveraging data, which requires an increased focus on Data Governance and collaboration. At the same time, these stakeholders need access to timely information, driving the need for newer, innovative data platforms to support faster, reliable access to an ever-growing amount of data.

These competing demands – the need for faster access to a greater amount of data in order to drive business success and the simultaneous need to protect and manage those same data assets to reduce risk – make for an interesting exploration into how real-world organizations respond to the challenge. How are companies practicing Data Management in this environment? What is the current state of Data Management? How are companies planning for the future? How do they balance new and old technologies? How do they manage data effectively for profit?

DATAVERSITY® asked these questions through a 2019 Trends in Data Management survey. This paper details and analyzes the latest thoughts, trends, and activities indicated by those who participated in the study. Some findings of the survey include:

- Organizations need effective Data Management to become data driven:

  · 70% of respondents feel that their organization considers data a strategic asset.

  · 42% have a Data Management effort in place across the enterprise.

  · More than half identified improved collaboration through using a defined Data Architecture.

- Businesses Intelligence and Analytics are key initiatives to gain critical business insights:

  · According to 80%, gaining insight through reporting and analytics motivates effective Data Management.

  · 87% of respondents are implementing Business Intelligence reporting and 87% have a data warehouse in place.

- Organizations realize the importance Data Governance and the methods that support it:

  · 76% have Data Governance in place or plan for the future.

  · Many responses point to Data Government as a critical piece that is gaining traction.

  · 86% of respondents consider data security a priority.

- Data Technology and Platforms continue to evolve:

  · Relational databases (on-prem or in the cloud), continue to be the leading platforms in use.

  · While relational database will continue to be popular in the future, a wider range of complementary technologies are being considered including graph databases, data streaming, IoT, etc.

  · 71% still use spreadsheets as a data platform.

We will analyze these results and investigate other key findings effecting Data Management today and into the future.

2. RESEARCH AND DEMOGRAPHICS

A. Scope of Research

The 2019 DATAVERSITY Trends in Data Management survey centered around the organizational pulse of Data Management. This study looked at the overall data landscape across the enterprise, including Big Data ecosystems, Data Modeling, Data Governance, Data Architecture, AI and machine learning, and newer trends such as blockchain, edge computing, Industry 4.0, and digital twins as each field pertains to Data Management.

The survey had thirty-six questions. Nine questions displayed only on the condition of a "yes" response to the previous question. Four questions posed open-ended queries. The survey broke down into nine areas:

- General demographics (four questions).

- Current state of Data Management (six questions, one open-ended).

- Goals, drivers, and training (five questions, two open-ended).

- Data Governance and Metadata Management (three questions, one conditional on a "yes" answer to having a Metadata Management effort in place).

- Data Architecture (three questions, two conditionals on having a defined Data Architecture foundation underlying Data Management goals and priorities).

- Data Modeling (three questions, two conditionals on actively using Data Modeling).

- Big Data ecosystem technologies (three questions, one conditional on using a Big Data platform, and one question on data lakes).

- Data platforms and storage (six questions, two conditionals on whether a cloud-based database was currently in use, and one conditional on whether blockchain was in use or in consideration).

- New technologies and trends (three questions, including one open-ended).

Most of the questions also contained extra space for written comments, and those comments will be discussed, when relevant, throughout the paper.

DATAVERSITY recruited survey participants from an email campaign, Data Education Month resources, and from the 2019 Enterprise Data World Conference. Respondents received links to the Trends in Data Management survey and did not have a time limit on answering the questions. No compensation was given for participation. The results and conclusions contained within this paper are representative of the survey respondents only, a group predisposed to interest in Data Management topics. 189 participants responded to the survey.

B. Principal Demographics

To begin, DATAVERSITY asked respondents three main demographics questions (outside of contact information): job function, respondent's industry, and number of employees in the respondent's company.

1. Job Function

The largest percentage of those surveyed, 43.11 %, identified themselves as working in Information/Data Governance [Figure 1]:
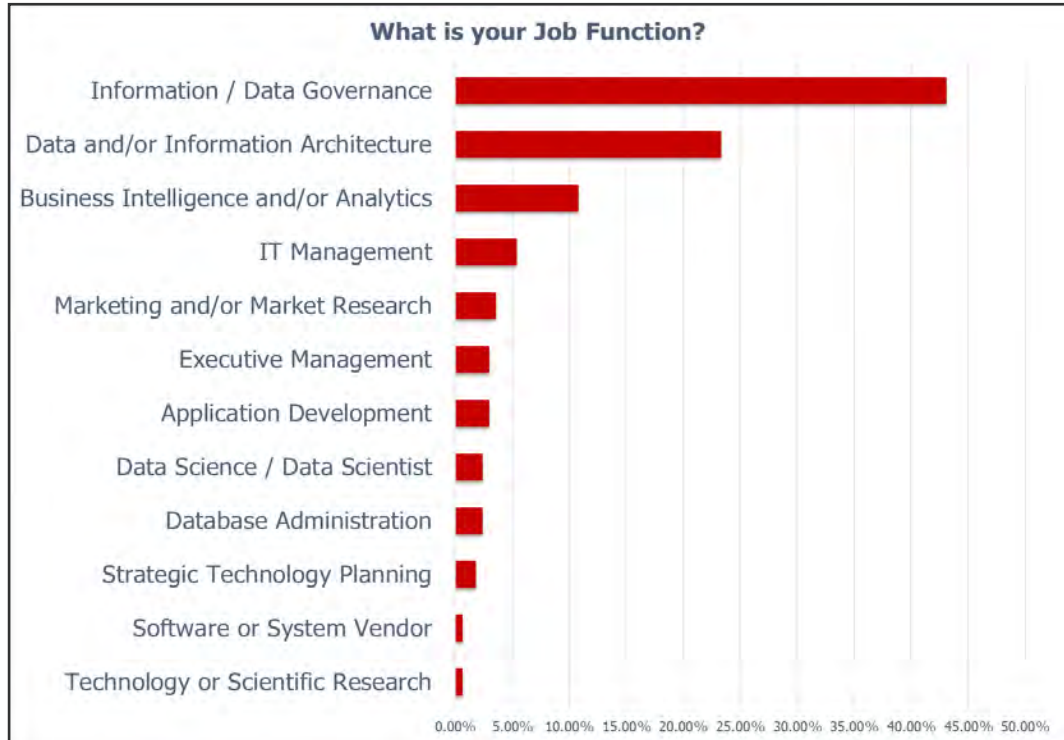

What is your Job Function?

Figure 1: Job Function

The top three professions broke down as follows:

- Information/Data Governance: 43.11%

- Data and/or Information Management: 23.35%

- Business Intelligence and/or Analytics: 10.78%

Those that identified as "Other" (10.78%) covered a wide range of areas including GIS data integration and faculty at higher education institutions.

2. Industry Representation

The wide range of industries included reflects the democratization and widespread adoption of Data Management. While in the early years of the industry, Data Management was largely skewed towards heavily regulated and historically data-driven industries such as banking, insurance, and government, today those industries make up less than 10-15% each of the survey respondents. Participants over 31 industries, from agriculture to utilities [Figure 2]:

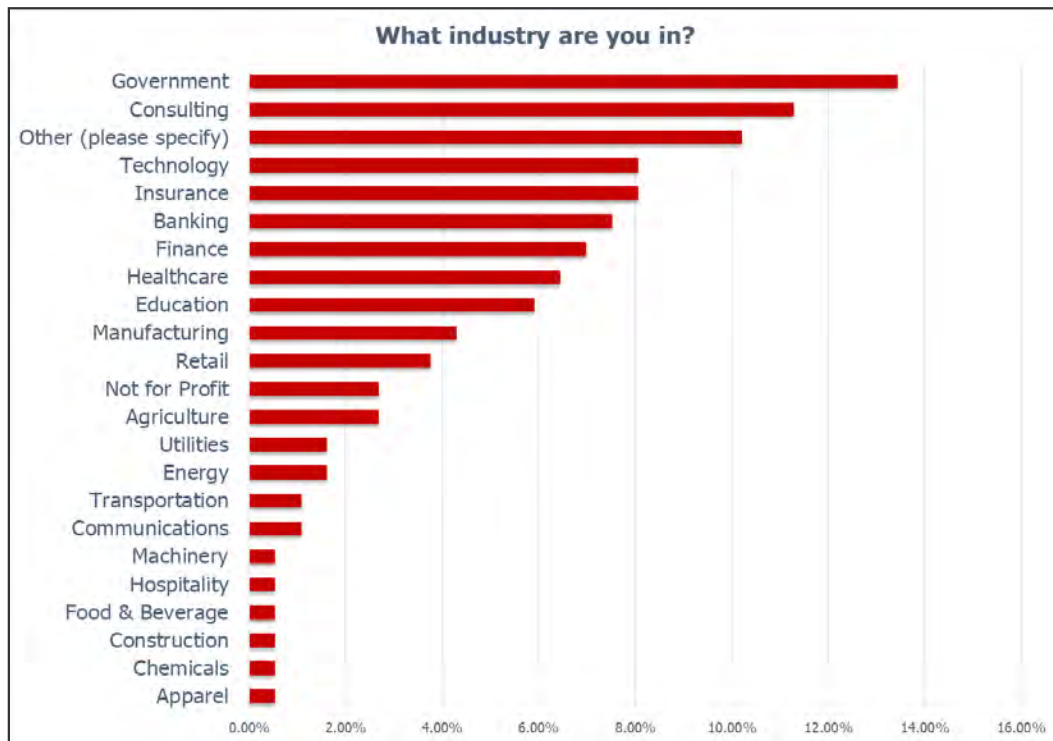Figure 2: Industry

Government made up the largest industry at 13.44%. The top four sectors were:

- Government: 13.44%

- Consulting: 11.29%

- Banking: 8.06%

- Insurance: 8.06%

Other sectors included health care, manufacturing, education, retail, agriculture, not for profit, energy and utilities, as well as write in industries such as lottery and gaming, law enforcement, life sciences, and travel.

3. Company Size

The democratization of Data Management is also shown in the distribution of organization sizes represented by the survey. Overall, participants came from a wide range of businesses – some with less than ten employees and others with more than 50,000. Again, whereas the early days of Data Management saw adoption primarily among the large players, current adoption is widespread across a range of organizational sizes. Data Management is not just for the big players anymore [Figure 3]:
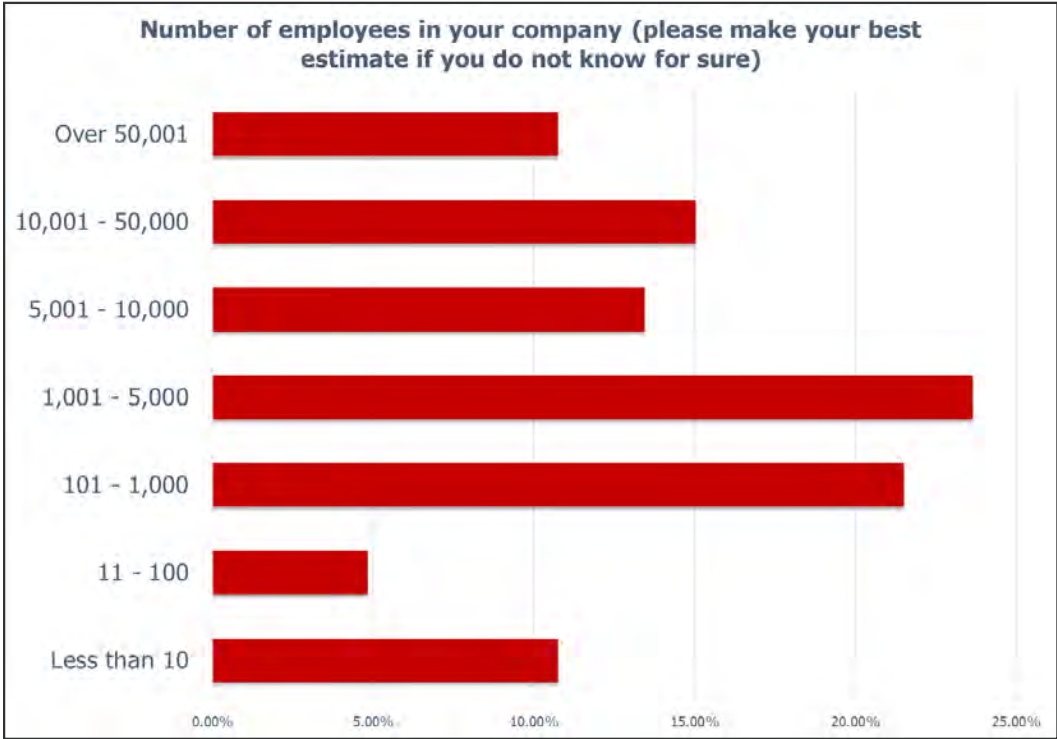

Figure 3: Number of Employees

The top three segments consisted of:

- 1,001 – 5,000: 23.66%

- 101 – 1,000: 21.51%

- 10,0001 – 50,000: 15.05%

3. INTRODUCTION: WHAT IS DATA MANAGEMENT?

For the purpose of alignment and clarity within this report, a working definition of Data Management is needed. DATAVERSITY specifies Data Management as:

> "A comprehensive collection of practices, concepts, procedures, processes, and a wide range of accompanying systems that allow for an organization to gain control of its data resources. Data Management as an overall practice is involved with the entire lifecycle of a given data asset from its original creation-point to its final retirement, how it progresses and changes throughout its lifetime through the internal (and external) data streams of an enterprise."

To expand further, the DAMA International definition in the Data Management Body of Knowledge ([DAMA DMBOK](#)) states:

> "Data Management is the development, execution, and supervision of plans, policies, programs, and practices that deliver, control, protect, and enhance the value of data and information assets throughout their lifecycles."

Given such a broad concept, we wanted to know how survey respondents defined Data Management to better understand the multiple perspectives, commonalities, and discrepancies of how Data Management is understood in the real world. This survey asked the open-ended question:

- "What is your (or your organization's) definition of Data Management?"

Many of those surveyed agreed with the DAMA DBMOK for their definition, which points to a widespread adoption of this standard.

Those who thought of Data Management differently pointed to a broad alignment of people, processes, and technology to holistically protect data as an asset:

- "Data Management improves the quality and integrity of master data assets, so they are globally trusted, accessible, single sourced, enriched, and standardized."

- "Data Management describes people, process, and technology to optimize, protect, and leverage data as an asset."

- "Data Management ensures our data assets, throughout the data lifecycle, are known, protected, accessible, and valuable to our data consumers."

A large group also saw Data Governance as a critical part to Data Management, as these comments indicate:

- "Data Management is Data Governance and lifecycle management of all data sources."

- "Data Management is the processes that govern the acquisition, use, and disposal of data."

- "Data discipline of managing data and information within the organization through structured governance, policies, and standards."

> "
> *Data Management ensures our data assets, throughout the data lifecycle, are known, protected, accessible, and valuable to our data consumers.*
> "

Several respondents pointed to Data Management as a driver towards business insights and improvement:

- "Data Management is an organization capability supported by tools, processes, standards, and people."

- "Data Management cares and feeds our data stores in support of our business needs."

- "Data Management makes enterprise data effective and efficient by supporting business activities."

- "Data Management is the practice of organizing and maintaining data processes to meet our business requirements."

Only a few respondents considered Data Management as solely a technical or security mechanism without including other aspects like Data Governance and business requirements. This is constructive, as Data Governance and business insights are essential elements.

In a unique view, one respondent defined Data Management as "data ethics and integrity from initial collection to presentation, in the court system."

All of these Data Management facets – caring for data as an asset, driving business understanding, formally governing data along with technology, and security – are important in order to be successful.

"Data Management makes enterprise data effective and efficient by supporting business activities."

4. THE CURRENT STATE OF DATA MANAGEMENT

The survey asked four questions about Data Management's scope, roles, plans, and values within the organization to get a clearer picture of current Data Management practices as well as trends for the future.

A. Scope of Data Management

We asked, the following question [Figure 4]:

► **"What is the scope of Data Management in your organization?"**

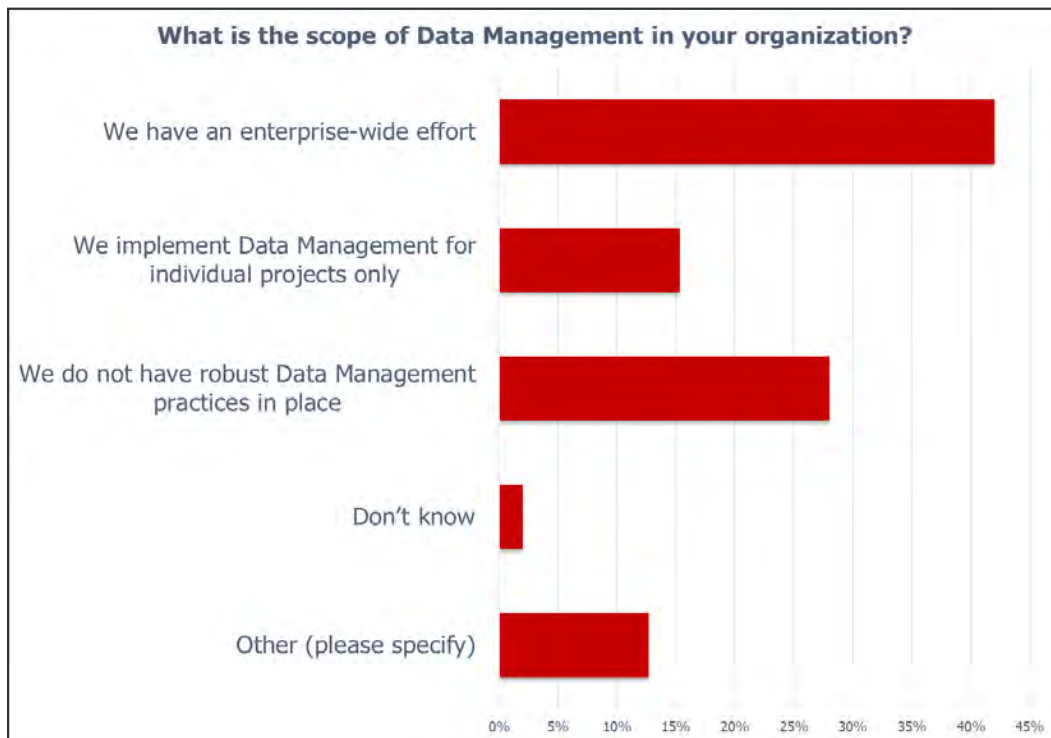**What is the scope of Data Management in your organization?**

Figure 4: Scope of Data Management

The results showed a large gap. The highest percentage reported having enterprise-wide scope, but the second-highest group reported not having any significant Data Management practices in place:

• "We have an enterprise-wide effort" was selected by 42%

• "We do not have robust Data Management practices in place" was chosen by 28%

One reason for this disparity may be the increase in the number of new entrants into the Data Management space, as shown by the wide range of industries shown in Figure 2. While those in the industry longer may have a more established practice, newcomers may not have had time to catch up, which is to be expected.

It is also important to remember that Data Management often builds in stages within an organization, with one team initially taking the lead and showing success before rolling out to the wider organization. This is likely the case for the 15% of respondents implementing individual projects and is reflected in the comments, as well:

- "One business segment has a Data Management practice. We are thinking about building this out for the rest of the bank."

- "Data Management is hybrid in scope. Some cover enterprise-level and some Data Management is specific to a business, subsidiary or data domain."

- "Overall, we have global enterprise scope, with phased implementation by data domain."

B. Roles Driving Data Management

With the rise in interest in Data Management comes a wide variety of roles involved in driving the effort, including both business and technical roles. The following question was asked [Figure 5]:

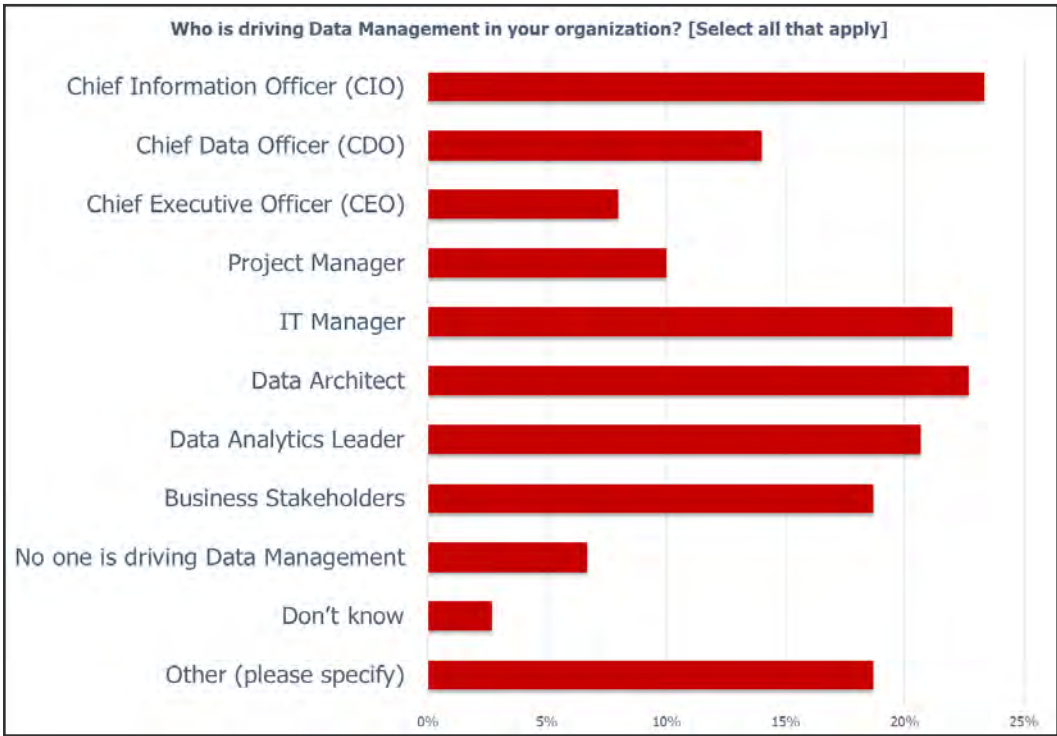▶ **"Who is driving Data Management in your organization? [Select all that apply]"**

Figure 5: Who is Driving Data Management

The top answers included:

- Chief Information Officer (CIO): 23.33%

- Data Architect: 22.67%

- IT Manager: 22.00%

- Data Analytics Leader: 20.67%

- Business Stakeholders: 18.67%

It is interesting to note the fairly equal distribution of a number of roles across the organization involving IT, business, architecture, and analytics – all of which play a critical part in a successful Data Management initiative. Notably, a number of respondents mentioned Data Governance in their comments as a way to align the various stakeholders around common goals.

C. Components of Data Management

Data Management is a varied discipline involving a number of components across people, process, and technology. The survey queried respondents to understand which aspects of Data Management they were implementing today, and what their plans for the future were. [Figure 6]:

▶ **"Which of the following have you already implemented or plan on implementing in the next 1-2 years in your organization? [Select all that apply]"**
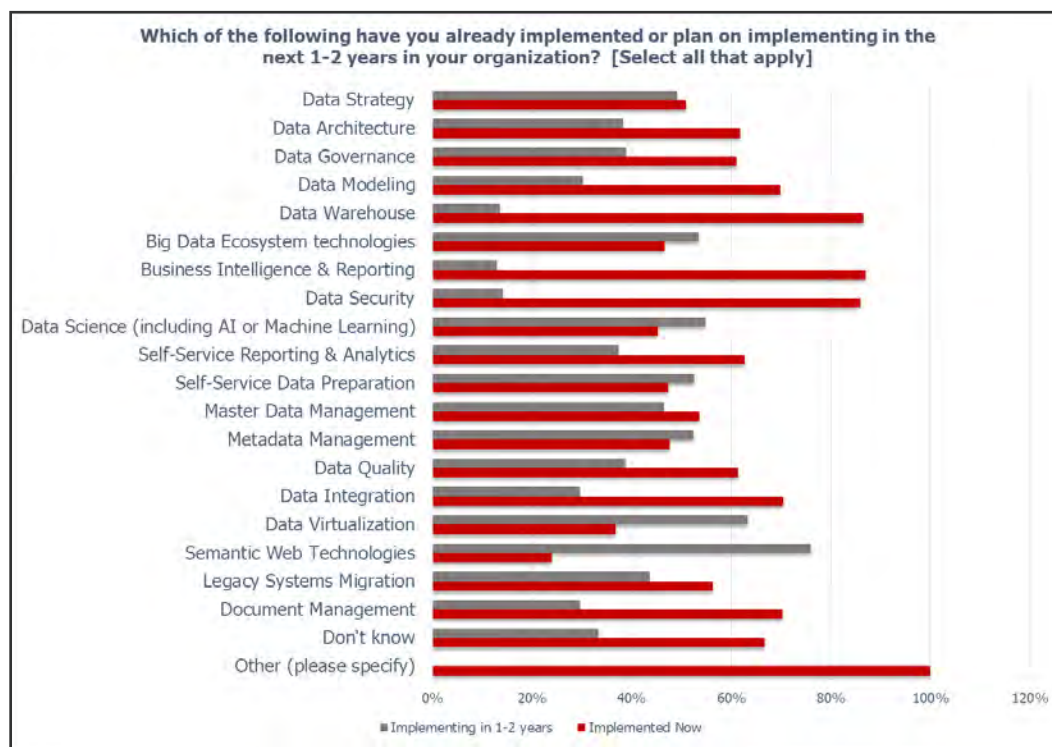

Figure 6: Implementation of Data Management

The top five Data Management components currently implemented are as follows:

- Business Intelligence and Reporting: 87.02%

- Data Warehouse: 86.55%

- Data Security: 85.95%

- Data Integration: 70.37%

- Document Management: 70.33%

A significant number responded the following has also already been implemented:

- Data Governance: 61.11%

- Data Quality: 61.29%

Respondents indicated the following technologies planned for implementation in the next 1-2 years:

- Semantic Web Technologies: 76.00%

- Data Virtualization: 63.24%

- Data Science (Including AI or Machine Learning): 54.74%

- Big Data Ecosystems: 53.42%

- Self-service Analytics: 52.63%

Organizations also plan on implementing core Data Management efforts, such as:

- Metadata Management: 52.43%

- Data Governance: 38.89%

Business Intelligence and data warehousing continue to be key drivers for Data Management, as organizations are looking to be more data-driven. Other top-ranking efforts center on security, integration, and document management, among others.

This is not surprising. With the number of high-profile breaches and the increasing number of privacy regulations, organizations feel pressure to secure data assets, including both database-centric information and documents. The high number of Data Governance and Data Quality initiatives supports this trend.

Survey participants also indicated future efforts will focus on modernizing tools that have already been implemented. For example, A data warehouse 2.0 would find more favor than a data lake.

These trends, through the next year, indicate that organizations are using data more strategically by leveraging existing rules and data definitions used in prior data endeavors, while also taking advantage of the faster response times and ease of access provided by new technologies. This thoughtful approach makes sense and leads to increased Data Governance and protection. Businesses realize that carrying out these fundamental tasks, in addition to Metadata Management, underpin successful data-driven business results, such as reporting, Data Science, self-service BI, etc.

Looking a year or two in the future, semantic web technologies and data virtualization appeal to a large number of organizations. The flexibility and power of semantic web technologies promises to handle an increasing data volume and variety, bridging these factors with common meanings and thereby providing a business edge. This may also explain data virtualization's current and anticipated rise – providing real-time access to all data systems without moving the data into a single data store. Investing in data virtualization is seen as more fruitful than moving everything into a more traditional data warehouse.

Many organizations are in the early stages of their exploration and understanding of AI and machine learning. More study needs to be done to clarify organizations intentions in using AI and machine learning as well of their technical understanding of the topic. At the end of the survey, we asked: "Which of the following is your organization currently leveraging or planning to implement within 1-2 years?"

Implementing deep learning was chosen by 82.35% of the respondents.

The difference in response between this question, represented by Figure 29, and the above question, represented by Figure 6, is surprising. One would expect a smaller gap between these values, as deep learning is a subset of Data Science (along with AI and machine learning). Yet, in that later question only 17.65% said they currently implement deep learning, so it seems that over the next couple of years deep learning is expected to grow substantially. (See Section 11, "New Technologies and Trends" for further discussion.)

D. Organizational Support for Best Practices

To figure out how organizations value their data assets, we asked the question [Figure 7]:

▶ **"Which of the statements below is accurate regarding your organization? [Select all that apply]"**
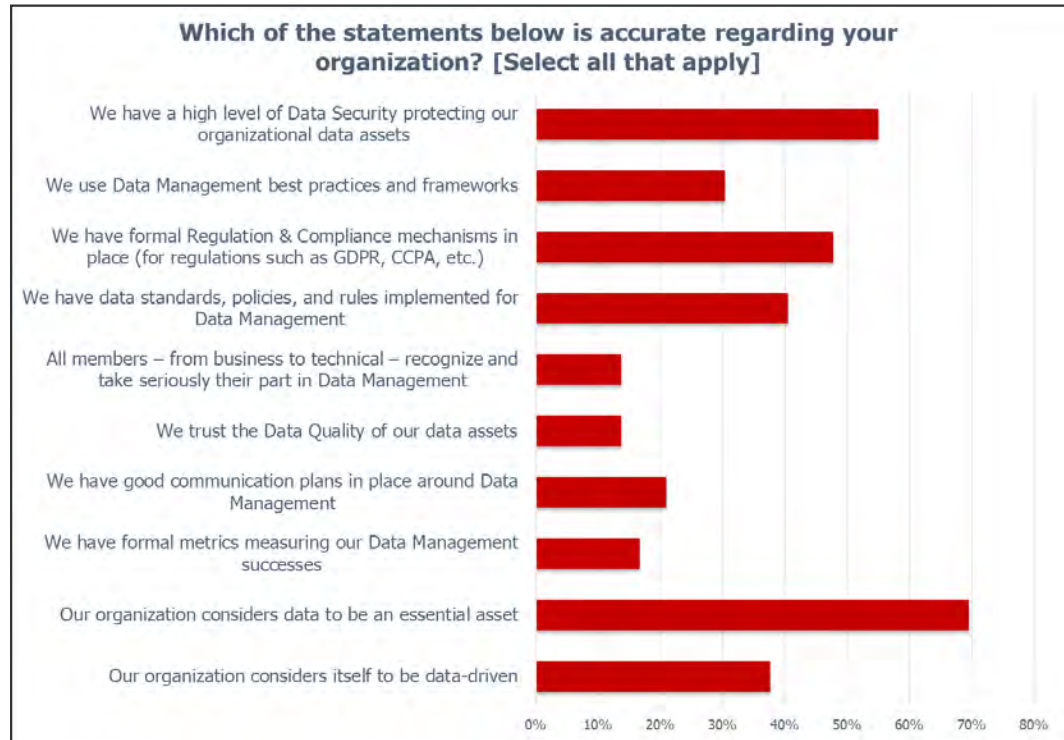

Figure 7: Organizational Data Value Statements

The highest response was for "Our organization considers data to be an essential asset" at 69.57%.

Data that is secure and compliant with regulations was also important:

• "We have formal regulation & compliance mechanisms in place (for regulations such as GDPR, CCPA, etc.)" was chosen by 47.83%.

• "We have a high level of data security protecting our organizational data assets" was chosen by 55.07%.

Many people do not trust their organization's Data Quality and do not think that other staff take their part seriously:

• "We trust the Data Quality of our data assets." was only true for 13.77%.

• "All members – from business to technical – recognize and take seriously their part in Data Management" was also chosen by 13.77%.

We are glad to see such a high focus on being a data-driven organization, as well as the concomitant high ranking for data security, regulation, and policies to protect data assets. The combination of these values sends a strong signal that businesses truly consider data an asset – something to be protected and valued. Respondent comments also support this:

- "Acceptance by top level management that data is truly an asset is the most recent profound change."

- "Our company is focused on establishing ourselves as a data-driven organization. We will provide the right data to the right people at the right time."

Lower scores on Data Management metrics, Data Quality, and the wide-spread understanding of data across the organization suggest a lack of the skills and expertise needed to support the better Data Management that organizations would like to have. These responses highlight a need for more Data Management support:

- "This business considers itself 'data-driven' but most do not actually understand what that entails."

- "Data Management is something we do in response to a government/client requirement or reacting to an event. Data Management is implemented in pieces on an as-needed basis."

> *Acceptance by top level management that data is truly an asset is the most recent profound change.*

E. Concluding Comments

Organizations continue to grow and evolve, demonstrating a vibrant and motivated Data Management industry. Many organizations have only recently entered the data space, as pressure mounts to understand and use data as a strategic asset.

Two respondents indicated that their organizations used maturity assessments to further focus around Data Management. Additional comments indicated:

- Those identifying as organizations at a low level of maturity have specific plans for improvement.

- Others who had been in the industry longer mentioned the need to continually evolve with new technologies and business challenges.

- Organizations have some areas that are more evolved than others.

- A key challenge remains to evangelize good Data Management to all across the organization.

Tools such as the DAMA DMBOK, DATAVERSITY seminars, on-line training, articles, and ventures like the Data Literacy Project can provide valuable resources and best practices for these organizations to help them better manage and grow to their data assets.

5. GOALS, DRIVERS, AND TRAINING

In the previous section, the survey questions and answers provided a shape of current Data Management practices and planned future implementations. This section will build on this by showing the currents and undercurrents patterning Data Management. This survey asked five questions about Data Management's goals, drivers, and training.

A. Goals and Drivers

The survey asked [Figure 8]:

▶ **"What are your main business goals and drivers for implementing Data Management in your organization? [Select all that apply]"**

Figure 8: Main Business Goals and Drivers

The top three goals and drivers were:

- Gaining insights through reporting and analytics: 79.70%

- Saving cost and increasing efficiency: 68.42%

- Reducing risk: 66.92%

Other popular choices included:

- Improving customer satisfaction: 58.65%

- Driving revenue and growth: 57.14%

- Supporting digital transformations: 53.38%

Only 1.50% of those surveyed did not know the main business goals and drivers for implementing Data Management, reinforcing the fact that the majority of data initiatives are strongly aligned with business goals and drivers.

The respondents recognize that Data Management goals and drivers need to go hand-in-hand with business support. Business insights made up the top driver, which aligns with other responses throughout this report. To support increasing business value and perceptiveness, survey participants identified the need for cost savings and revenue growth.

Notably, the need to reduce risk was higher than that of driving revenue and growth (57.14%), which is again consistent with other findings from this survey. Organizations focus on compliance and regulation in the industry. An increasing media and customer focus on data security, privacy, and compliance – as well as tightened industry regulations such as GDPR – make this result not surprising.

Many organizations want to improve the customer experience through improved digital engagement, driven by data. This likely explains why customer satisfaction and digital transformations remain popular. Business patrons also react more positively towards better security and privacy, pushing these outcomes.

The goals and drivers mentioned in the section help maintain energy in and commitment to Data Management (e.g. in Data Governance and Metadata Management) across the enterprise.
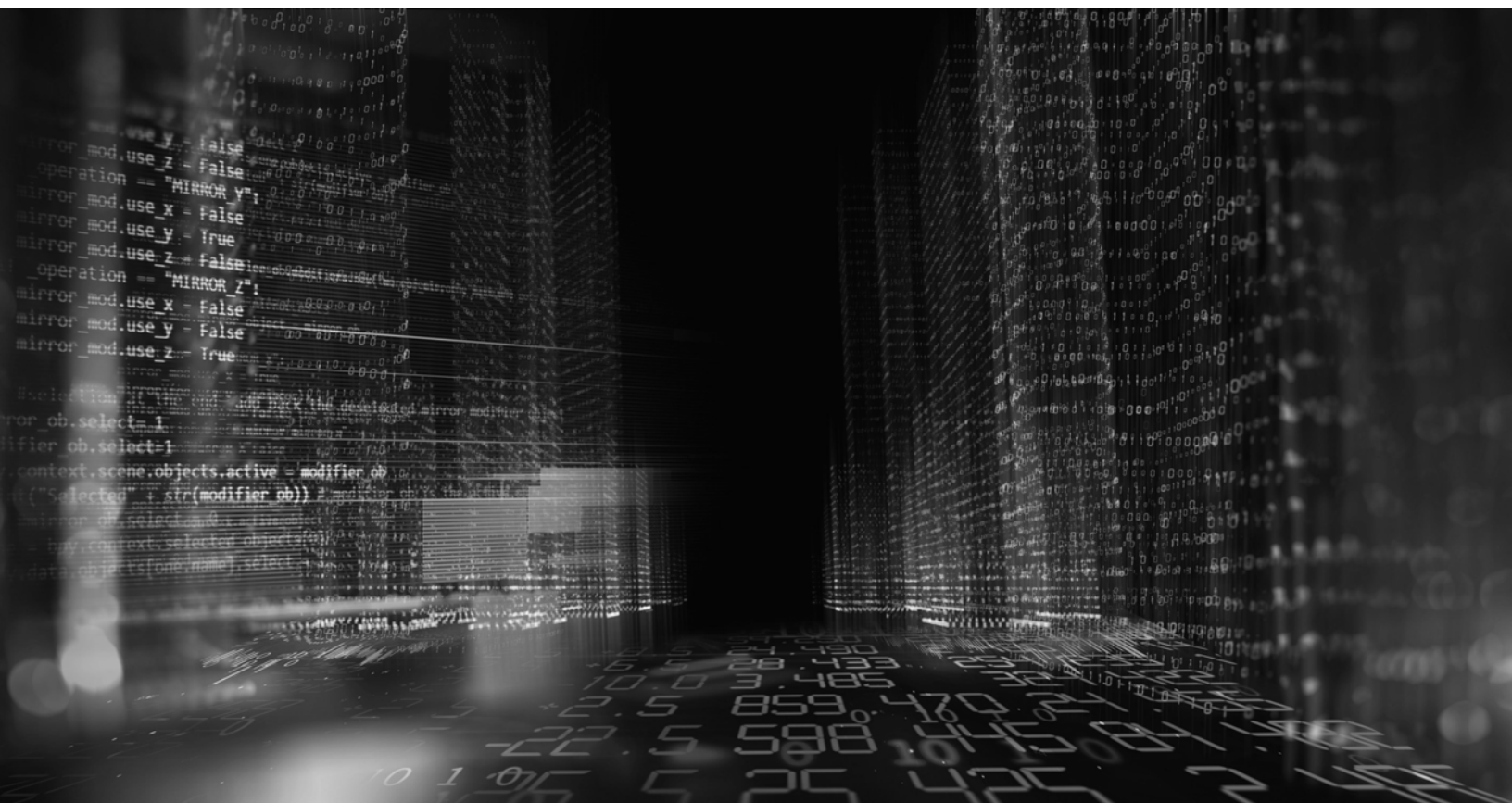
1. Priorities and Goals

How do survey participants see Data Management drivers in the future? To find out, we asked an open-ended question:

- "What are your top 2-3 Data Management priorities/goals for 2019-2020?"

Replies to this question indicated that the following were important:

- Data Governance: Respondents want an initial plan and roadmap towards managed governance programs
- Metadata Management: Those surveyed desire more mature Metadata Management.
- Data Quality: Comments emphasized trusted Data Quality and improvement of Data Quality to enable key priorities and to deliver improved outcomes.

The above list emphasizes the importance of ownership, accountability, and governance. This feedback reinforces a theme that we hear throughout this paper: Governance and risk reduction strongly drive organizations.

Various comments also reflect a need to engage more stakeholders across the organization with ownership and governance. This is particularly true for business stakeholders to increase their use and involvement with data assets. This is indicated through participant statements such as:

- "Data Management implementation establishes data ownership/accountability within the business."
- "Data Management goals and drivers establish a trusted Data Quality management process."

Opinions like these may seem to call for a "defensive" type of Data Management – to deflect any risks or to react to cost and time threats towards existing Data Management structures. A number of replies indicated a more proactive or "offensive" Data Management approach – using data more

strategically for gaining insights, data monetization, and competitive differentiation. For example:

- "Enabling decision science for cost reduction and improving customer experience in digital interactions are top priorities."
- "Support digital transformation and improve customer satisfaction drive Data Management."

Some respondents said their companies had difficulty prioritizing Data Management activities. This may indicate that an organization's maturity in Data Management plays a strong role in how future priorities and goals are chosen.

2. Challenges

Organizations certainly face difficulties in achieving Data Management goals. To understand this, we asked the open-ended question:

▶ **"What are the top 2-3 challenges you are facing with implementing Data Management?"**

Key difficulties centered around themes of ownership and accountability. As more and more roles in organizations need to get their hands on organizational data assets, people are clamoring for better Data Management resources and support. Comments included:

- "We need resources! We need time, people and money to get it done right."

- "We need more resources to make sure that due diligence is followed within the Data Management."

Along with sustaining Data Management practices, companies face challenges in managing competing priorities – especially in allocating more energy to data ideas and initiatives. Responses indicating this include:

- "The lack of leadership is a top challenge."

- "Managers are too busy on business tasks to see the importance of Data Management."

- "Top Data Management challenges include a lack of clear roles and responsibilities that are coordinated with top priorities."

Those that answered the survey recognized the importance of collaborating across the organization around data assets and effectively using information stored in data siloes. As a result, many responded that business culture posed problems in Data Management that need to be solved. For example:

- "Upper management has mandated a move to the cloud, but its motivations have not trickled down to the trenches, yet."

- "Culture and mindset change are the biggest obstacles."

Participants also identified training as an ongoing dilemma in implementing Data Management. This will be discussed further in the "Training" section below.

In conclusion, one response wisely observed:

- "Data is a constant challenge. It changes so quickly and is so critical."

> " *Top Data Management challenges include a lack of clear roles and responsibilities that are coordinated with top priorities.* "

> " *Culture and mindset change are the biggest obstacles.* "

3. Tools and Technology

What about technology? How does it impact Data Management decisions? To find out, the survey asked [Figure 9]:

▶ **"How much has the selection and purchase of software tools impacted your Data Management implementation?"**
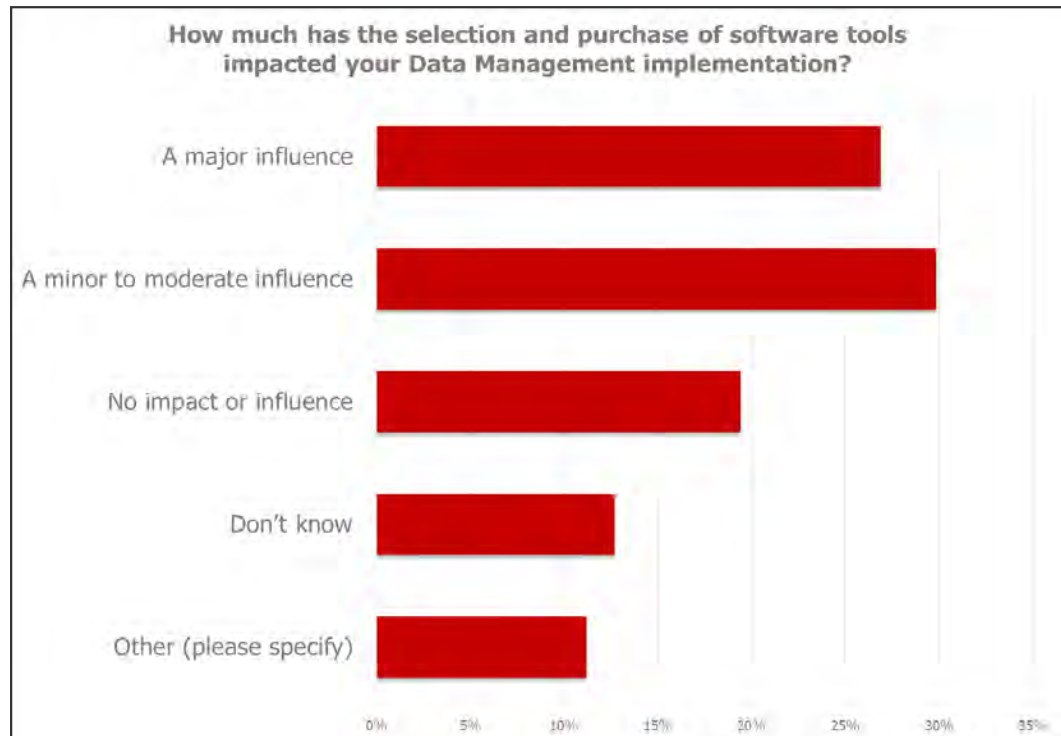

Figure 9: Software Tool Impact on Data Management Implementation

No strong trend was noted in the responses. The answer broke down as follows:

- A minor to moderate influence: 29.85%

- A major influence: 26.87%

- No impact or influence: 19.40%

Comments indicated that while some organizations were driven by tool selection, many were driven by business needs. Software selection is shifting away from monolithic enterprise applications, such as ERP systems, to technologies enabling Data Management in the business. Various respondents mentioned specific tools their organization used to aid in accelerating their organization's progress toward the goal of managing all data assets. Software tool selection is becoming more customized to companies and how they do business.

B. Training

How do participants educate themselves about Data Management and feel about their learning experience? This survey posed the question:

▶ **"What type(s) of training have you received in Data Management? [Select all that apply]"**

Survey participants cited a wide variety of sources, a necessity in today's ever-changing technology ecosystem [Figure 10]:
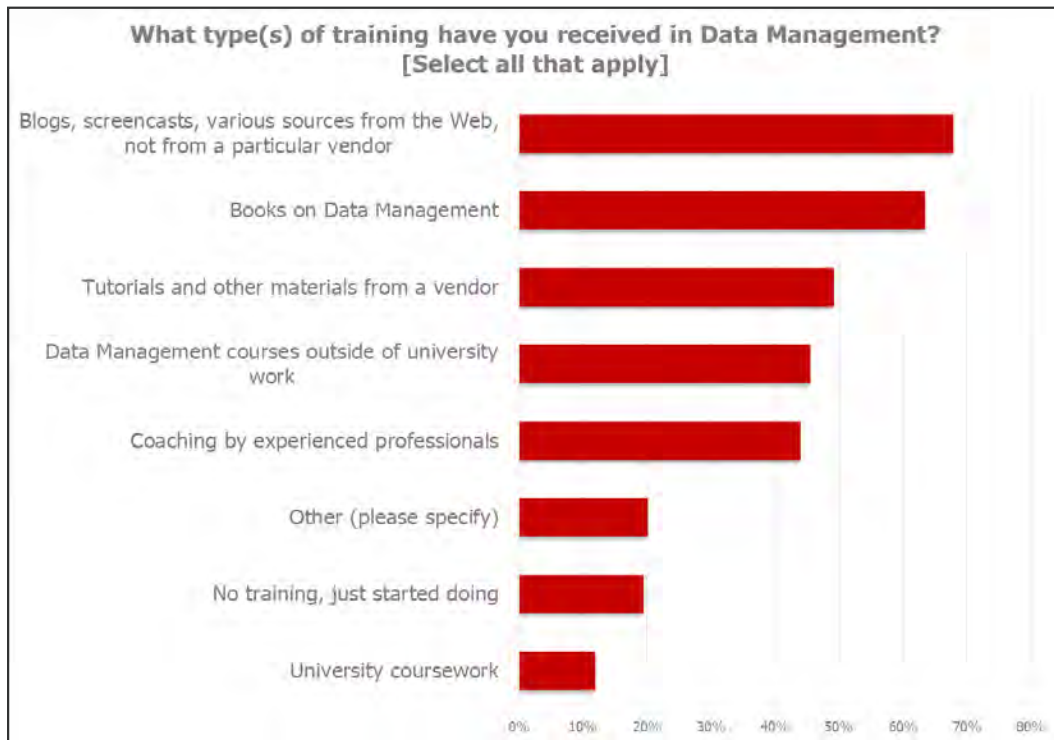

Figure 10: Training in Data Management

Top responses include learning from:

- Blogs, screencasts, various sources from the Web, not from a particular vendor: 67.91%.

- Books on Data Management: 63.43%.

According to the survey, universities were not a big source of Data Management training at this time. Only 11.94% received training through university coursework.

For those who take the initiative, there are a number of sources to get training: articles, books, and videos. However, the survey responses indicate that few people in an organization – beyond data professionals themselves – will seek out training. In the comments, many participants emphasized the need for more available and frequent data education for all:

- "Data fluency training for ALL employees, not just data employees are needed."

- "Frequent user training and data literacy are very important."

- "We need to train more people in Data Governance, and how to best automate these processes so that they are not perceived as roadblocks to product development and deployment."

In general, survey participants talked about the need for training across a number or roles, not just the traditional technical ones.

6. DATA GOVERNANCE AND METADATA MANAGEMENT

In the earlier sections, "Current State of Data Management" (Section 4) and "Goals, Drivers, and Training" (Section 5) those surveyed advocated for Data Governance and its supporting practices, such as metadata. In this section, three questions were posed to go deeper into the specific practices of Data Governance and Metadata Management across organizations.

For those that answered "yes" to having a Metadata Management initiative, survey questions asked further details on how their metadata assets are administered.

A. Data Governance

We wished to better understand how organizations formally practice and process data assets. Respondents were asked:

▶ **"Which of the following best represents your company's state of Data Governance?"**

Most organizations are beginning or about to begin Data Governance initiatives [Figure 11]:
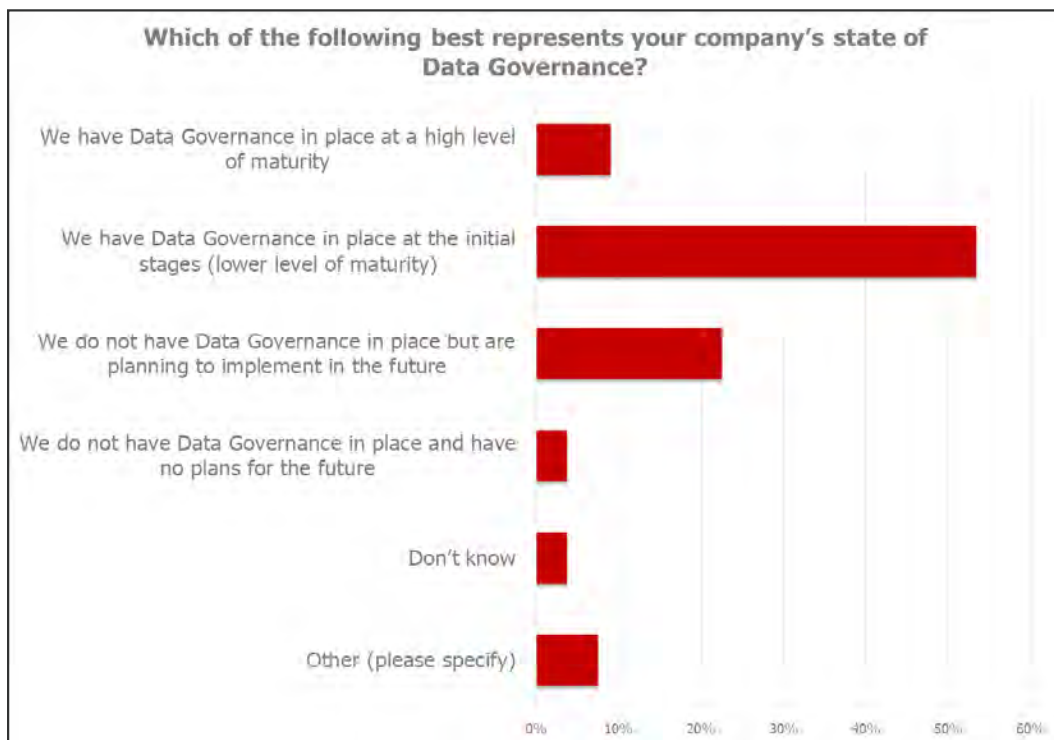

Figure 11: State of Data Governance

The responses broke down as follows:

- Data Governance is in place at the initial stages (lower level of maturity): 53.38%.

- No Data Governance in place but planning to implement in the future: 22.56%.

- Data Governance at a high level of maturity: 9.02%

- No Data Governance program nor any plans in the future: 3.76%.

Survey participants commented that Data Governance existed in different states across the organization, with older technologies having more maturity than newer ones:

- "We have a very mature mainframe data environment. The newer data and applications developed using AGILE are less governed."

- "The level of maturity is diverse and depends on the business area."

From the responses provided, Data Governance is clearly a growing trend. This is consistent with the theme of the report.

B. Metadata Management

Metadata, or data in context, fuels Data Governance success. We wanted to know whether organizations were managing metadata and asked the following [Figure 12]:

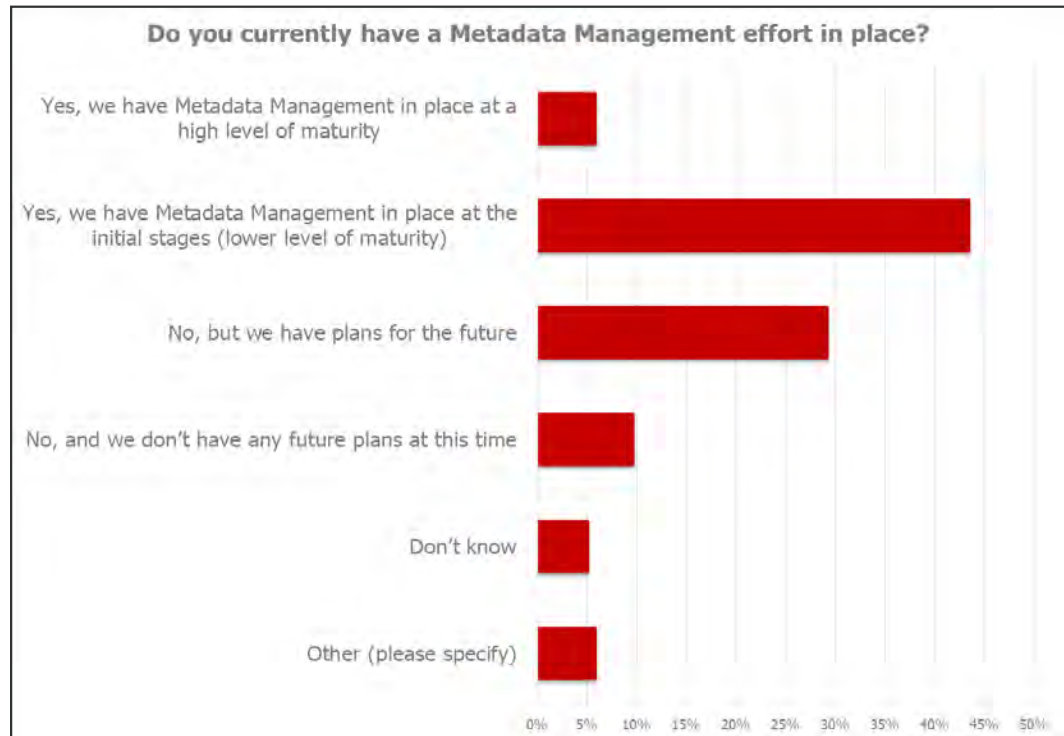▶ **"Do you currently have a Metadata Management effort in place?"**

Figure 12: Metadata Management Efforts in Place

The numbers show metadata's growth and importance:

- 43.61% indicated that Metadata Management is in place at the initial stages (lower level of maturity).

- 29.32% of participants indicated, "No, but we have plans for the future."

Clearly, there is plenty of room for Metadata Management to grow. Only 6% indicated they had a Metadata Management effort in place at a high level of maturity. A high interest in Metadata Management makes sense, as it provides a key foundation for Data Governance.

For those organizations that said "yes" to having Metadata Management, we posed an additional query [Figure 13]:

► **"What are your current main use cases for Metadata Management? [Select all that apply]"**
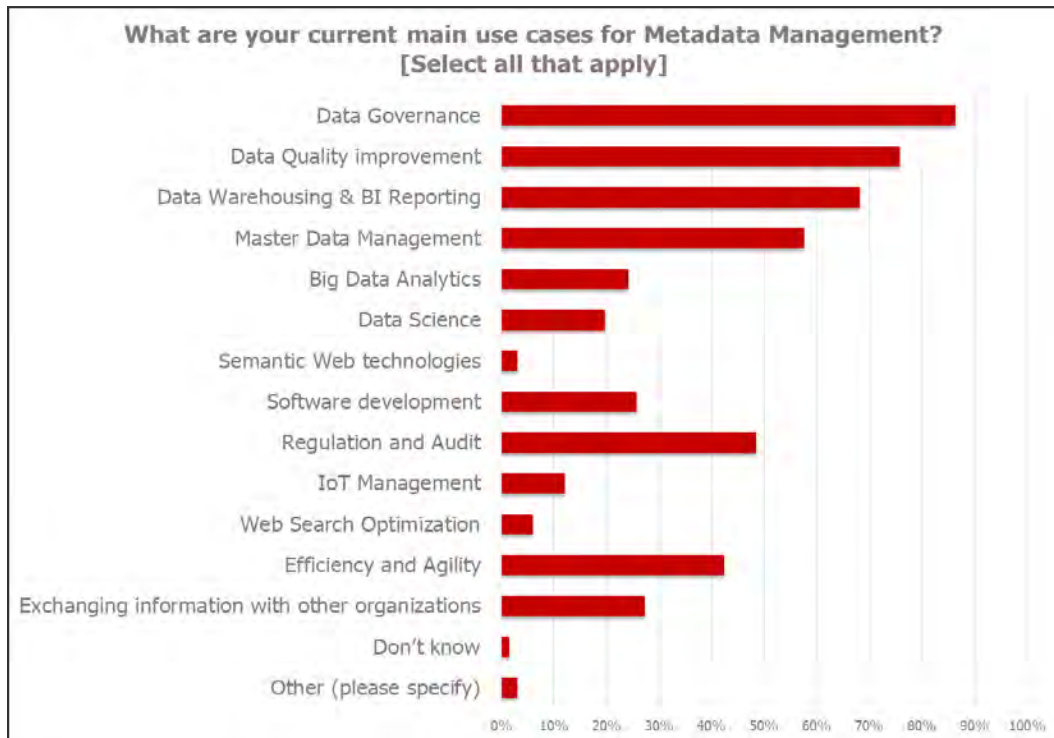

Figure 13: Metadata Management Use Cases

The top three Metadata Management use cases were:

• Data Governance: 83.36%

• Data Quality Improvement: 75.76%

• Data Warehousing & BI reporting: 68.18%

Legacy systems migration and data lineage for effective impact assessment were listed under written comments for the "Other" choice.

Metadata Management is seen as a trust/communication tool and not solely a technical component. It continues to grow in importance to the business. In the 2016 DATAVERSITY *Emerging Trends in Metadata Management Report*, we found that two-thirds of respondents said metadata is more important than it was ten years ago. This is still the case and likely will not change moving in the coming years.

Key drivers for good metadata include strong Data Governance and related areas of Data Quality and regulation/audit. Gaining business insight from data and becoming more a data-driven organization are also important drivers of Metadata Management efforts. The high ranking for data warehousing and business reporting (68.18%) shows a strong desire to get a consistent and true handle of an organization's data assets.

C. Concluding Comments

Many businesses understand or are starting to see that Data Governance and Metadata Management are necessary and provide value. Survey participants see business metadata as of the utmost importance as well as the most challenging to implement because of its complexities and need for business involvement. Respondents commented:

- "Technical Metadata Management is easier to implement than Business Metadata Management."

- "The business will not trust and use the data without full transparency of how it is being modified."

A few comments stressed that Metadata Management and Data Governance need to align with technological implementations like Data Modeling (See Section 8), Artificial Intelligence, or Machine Learning (See Section 11):

- "The possibilities of discovery for metadata are very exciting and open up a whole new approach for both Data Governance and metadata. AI and ML will help to achieve the discipline missing from both Data Governance and Metadata Management."

- "Data Modeling is a key contributor to enterprise Metadata Management which in turn provides a broad support to effective Data Management."

Data Governance and Metadata Management assist in bridging the business and the technical so that users can bank on the accuracy of an organization's data assets.

METADATA

7. DATA ARCHITECTURE

Data Architecture connects business strategy and technical execution. It produces outcomes, like models; activities, like software deployment; and behaviors, like collaborations between IT and business. As such, Data Architecture is a very important aspect of Data Management.

Three questions were asked in the survey. We wanted to know if organizations had a Data Architecture supporting their Data Management objectives. For those who did, we wanted to know its value and the problems companies face with their Data Architecture.

A. State of Data Architecture within Data Management

To comprehend Data Architecture's role in Data Management, survey participants were asked the following:

► **"Does your organization have a defined Data Architecture foundation underlying your Data Management goals and priorities?"**

Currently, from the results, it looks like effective Data Architecture is sporadic, with slightly more organizations not having a strong architecture than those that do. Survey results showed the following [Figure 14]:
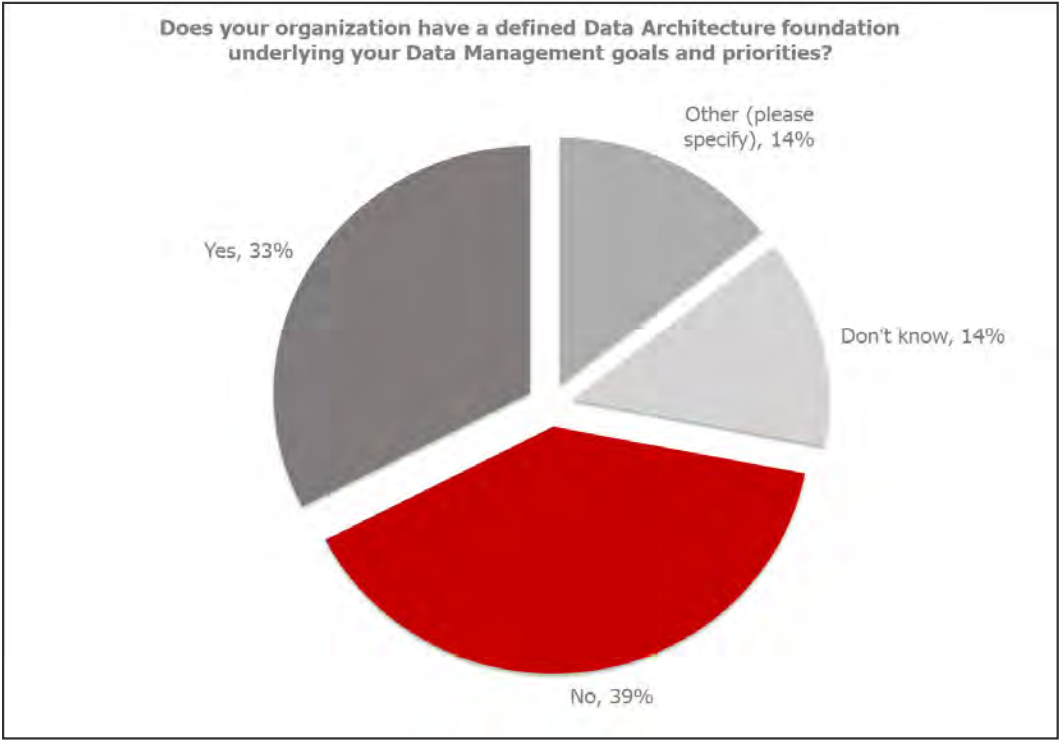

Figure 14: Data Architecture Foundation

Top responses broke down into:

- Do not have a defined Data Architecture: 39.39 %

- Do have a defined Data Architecture: 32.58%

The high percentage of "no" responses (39.39%) causes concern about Data Management risks, such as processes around security or data integration. However, many who choose "Other" (14.39%) specified that new initiatives are underway to begin a Data Architecture program, which is somewhat of a relief.

B. Value of Data Architecture

Of those organizations that have a Data Architecture supporting their Data Management, we wanted to know what value they were receiving from it. Participants were asked [Figure 15]:

▶ **"How has using a defined Data Architecture helped your organization? [Select all that apply]"**
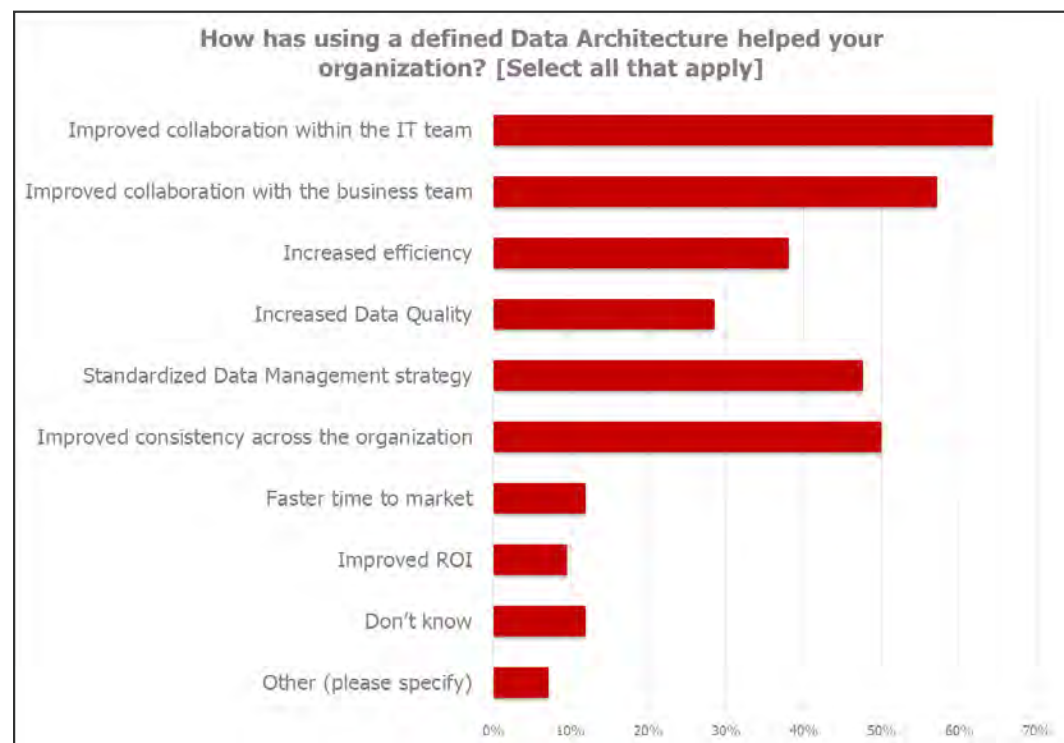

Figure 15: Data Architecture Advantages

Notable statistics included:

- Improved collaboration within the IT team: 64.29%

- Improved collaboration with the business team: 57.14%

- Improved consistency across the organization: 50%

Participant replies stressed that collaboration both within and beyond the IT team was helped by Data Architecture, which improved consistency. This ties in nicely with Data Governance and the benefits of using architectural tools (e.g. data models, glossaries, etc.) to support a Data Governance program.

C. Challenges

Data Architecture often faces roadblocks in its conception and implementation. To understand how this effects the organization, we asked the following [Figure 16]:

► **"What are your top Data Architecture challenges? [Select all that apply]"**

Figure 16: Data Architecture Challenges

The top four challenges were:

- Determining and implementing an enterprise strategy: 54.69%

- Managing change and impact analysis: 50.78%

- Gathering metadata from existing data sources: 43.75%

- Collaborating with business stakeholders: 42.19%

It is not surprising that the biggest difficulty appears to be creating and implementing this strategy amidst an ever-changing environment. In the "Goals, Drivers and Training" section above, the question, "What are the top 2-3 challenges you are facing with implementing Data

Management?" highlighted competing priorities and problems with ownership and accountability.

Survey participants also acknowledged that removing siloes and adequate documentation of both business and technical processes was a common problem.

Participants mentioned that a good, supportive Data Architecture needs to contain a variety of tools in the right amounts. Some indicated education/training about the Data Architect role is needed across the organization. One respondent observed:

- "Not a lot of people in an organization understand the Data Architecture/Data Architect concept."
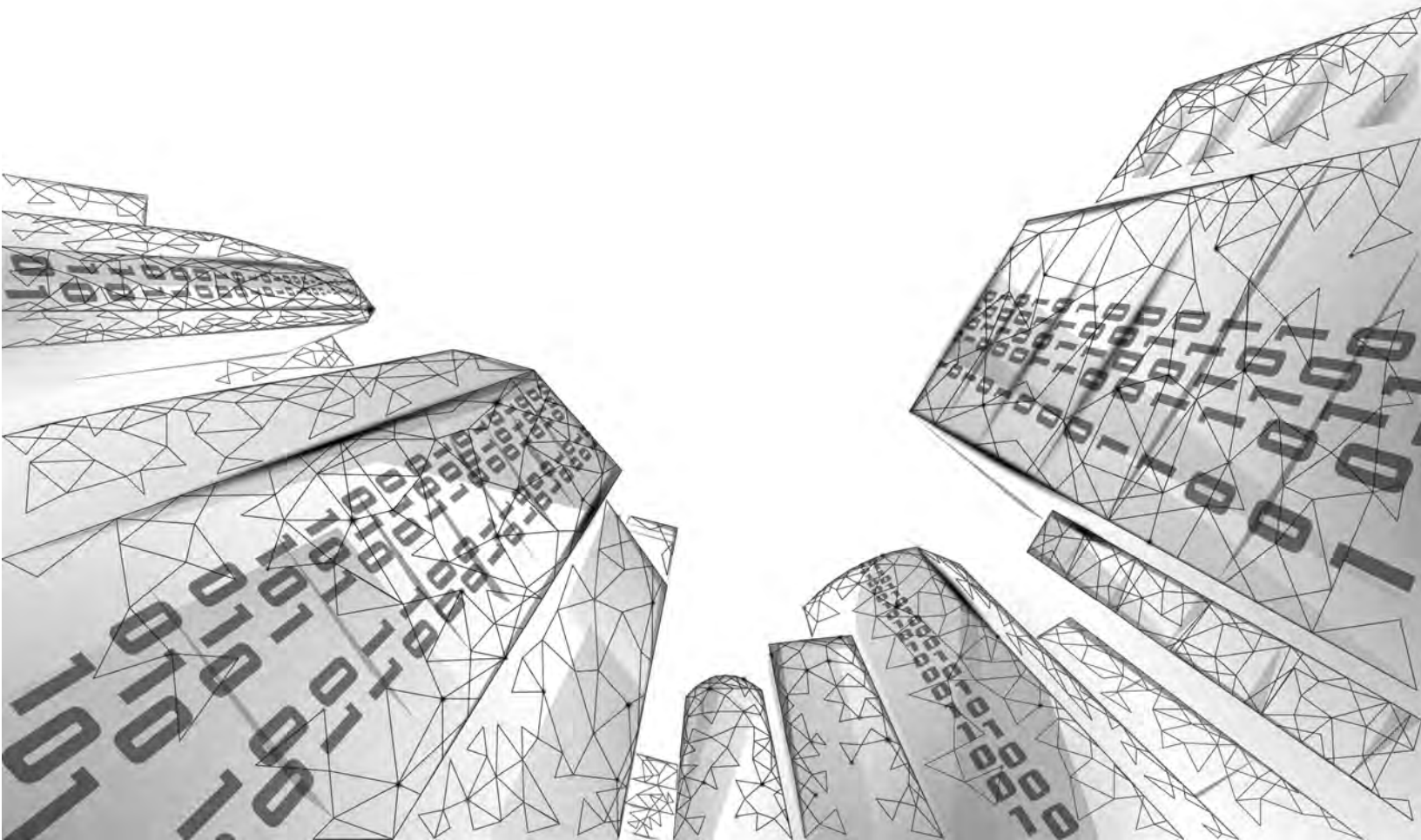
8. DATA MODELING

Data Modeling – the documentation describing the core business rules and definitions around data – plays a critical role in Data Management. We asked three questions about Data Modeling activities in the organization. Additional comments were provided to gain insight into Data Management and business alignment.

A. State of Data Modeling within Data Management

We posed the question [Figure 17]:

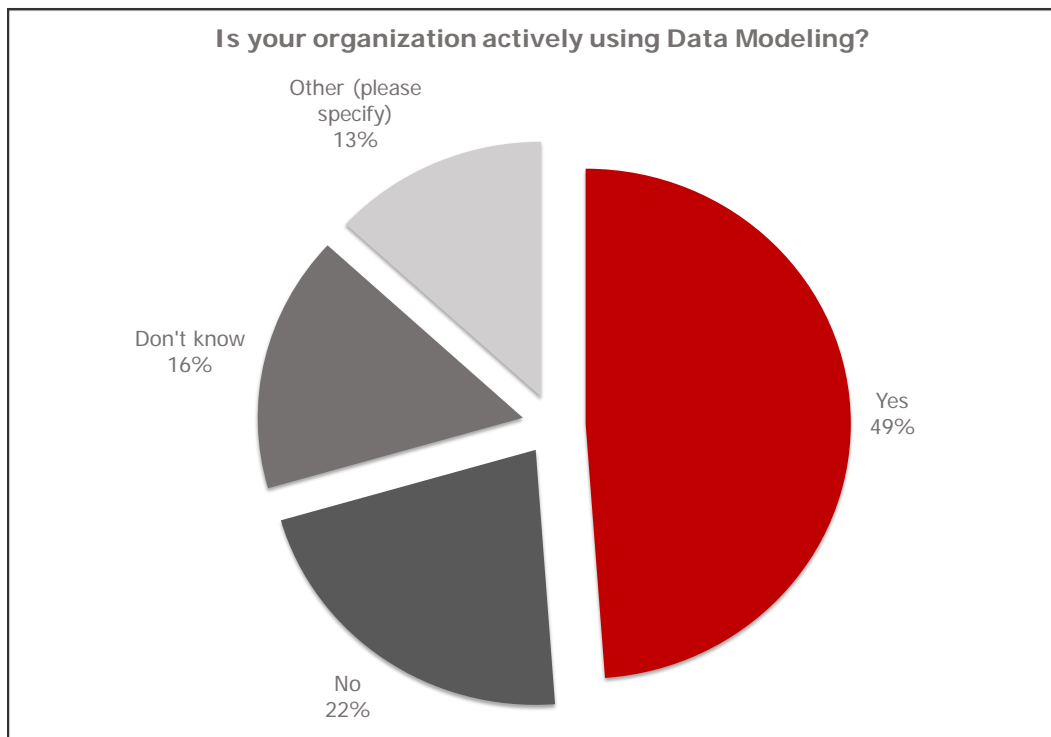▶ **"Is your organization actively using Data Modeling?"**

Figure 17: Actively Using Data Modeling

Top responses were as follows:

* Yes, actively using Data Modeling: 48.84%

* No, not actively using Data Modeling: 21.71%

Compared to previous research, as seen in the DATAVERSITY *2017 Trends in Data Architecture Report* where over 95% of respondents reported actively using Data Modeling, a lower number (48.84%) affirmed they do so now. Perhaps respondents felt Data Modeling usage was not comprehensive enough to say "yes," or perhaps they have had negative experiences with Data Modeling. There could also be discrepancies in who was answering the survey between both reports.

Those answering this question mentioned they were only beginning efforts, either as a new project or bringing back previous efforts, as organizations want to minimize the pain of not having Data Modeling. Data Modeling was noted to be inconsistent across an organization, which is seen as causing negative impacts where not implemented:

- "Yes, but Data Architecture is used often retrospectively, as an afterthought instead of a specification and is documented after a project finishes."

- "Data Modeling usage has decreased over the past couple years. Now the pain of this decision is hurting those that made the decision."

B. Data Modeling Methods

For organizations that said they were using data models, this study wanted to know their methods and model types. We asked the question [Figure 18]:

► **"What methods of Data Modeling do you use in your organization? [Select all that apply]"**
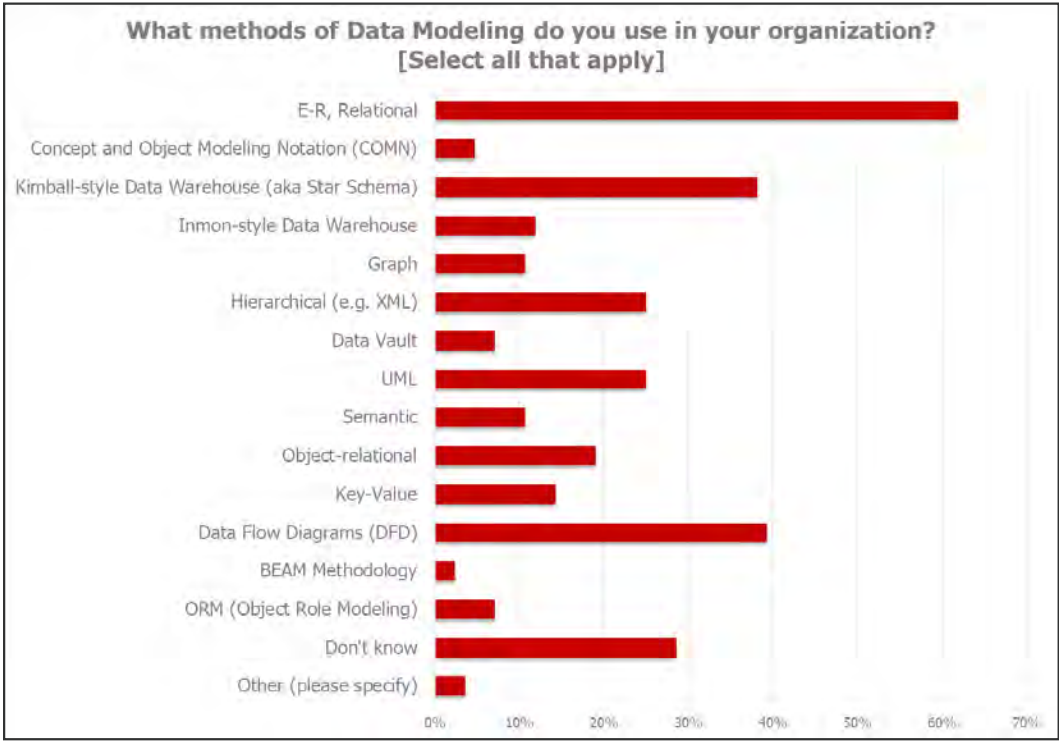

Figure 18: Data Modeling Methods

The highest-ranking Data Modeling methods included:

- E-R, Relational: 61.90%

- Data Flow Diagrams (DFD): 39.29%

- Kimball-style Data Warehouse (aka Star Schema): 38.10%

Relational and data warehousing models appeared to be the most popular, along with accompanying data flow diagrams, which makes sense given the strong focus from many organizations on Business Intelligence and reporting. Figure 6 above shows the majority of respondents have implemented Business Intelligence and reporting (87.02%) as well as data warehouses (86.55%). The preferred Data Modeling methods align with this information.

C. Types of Models and Diagrams

The next question asked centered on those who are actively practicing Data Modeling [Figure 19]:

► **"What types of models and diagrams do you use in your Data/Enterprise Architecture? [Select all that apply]"**
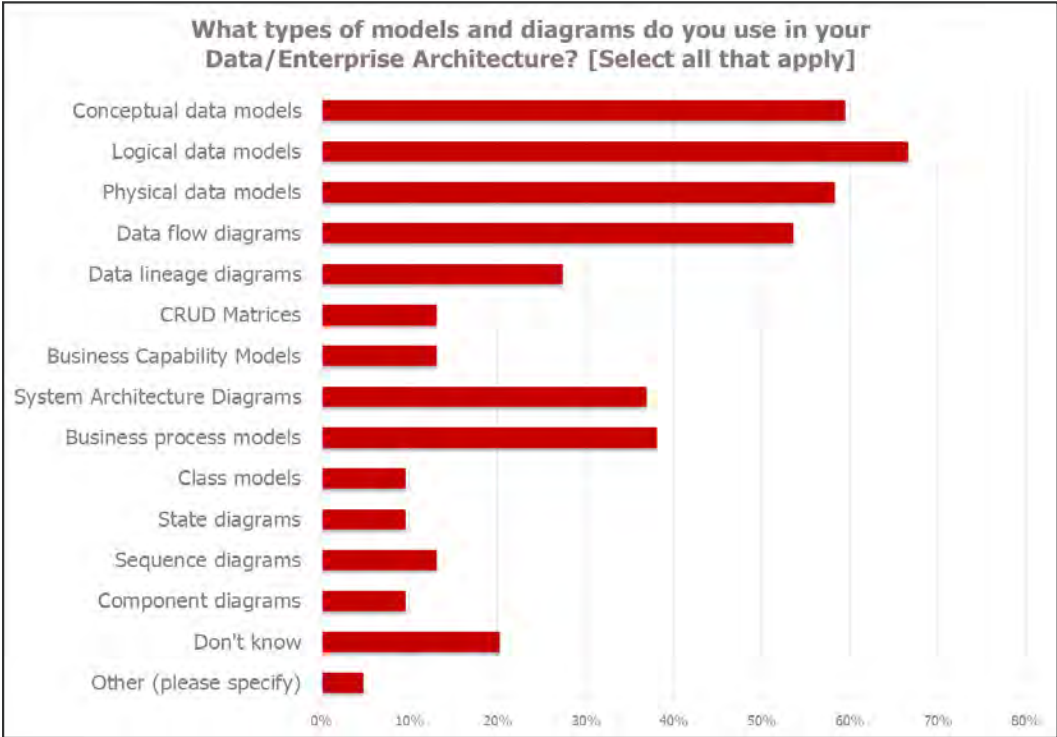

Figure 19: Data Models and Diagrams

Respondents prefer (with a small spread in model choice):

- Logical data models: 66.67%

- Conceptual data models: 59.52%

- Physical data models: 58.33%

- Data Flow Diagrams: 53.57%

- Business process models: 38.10%

Fewer use class models, state diagrams, and components diagrams, which all came in under 10%.

Top data models and diagrams show workflow that corresponds with previous discussions, especially around processes and actions, including business metadata and Data Governance. Data flow (53.57%), business process models (38.10%), and system architecture diagrams (36.90%) show an increased need to align with business needs, remove siloes, and look holistically at data as a strategic business asset.
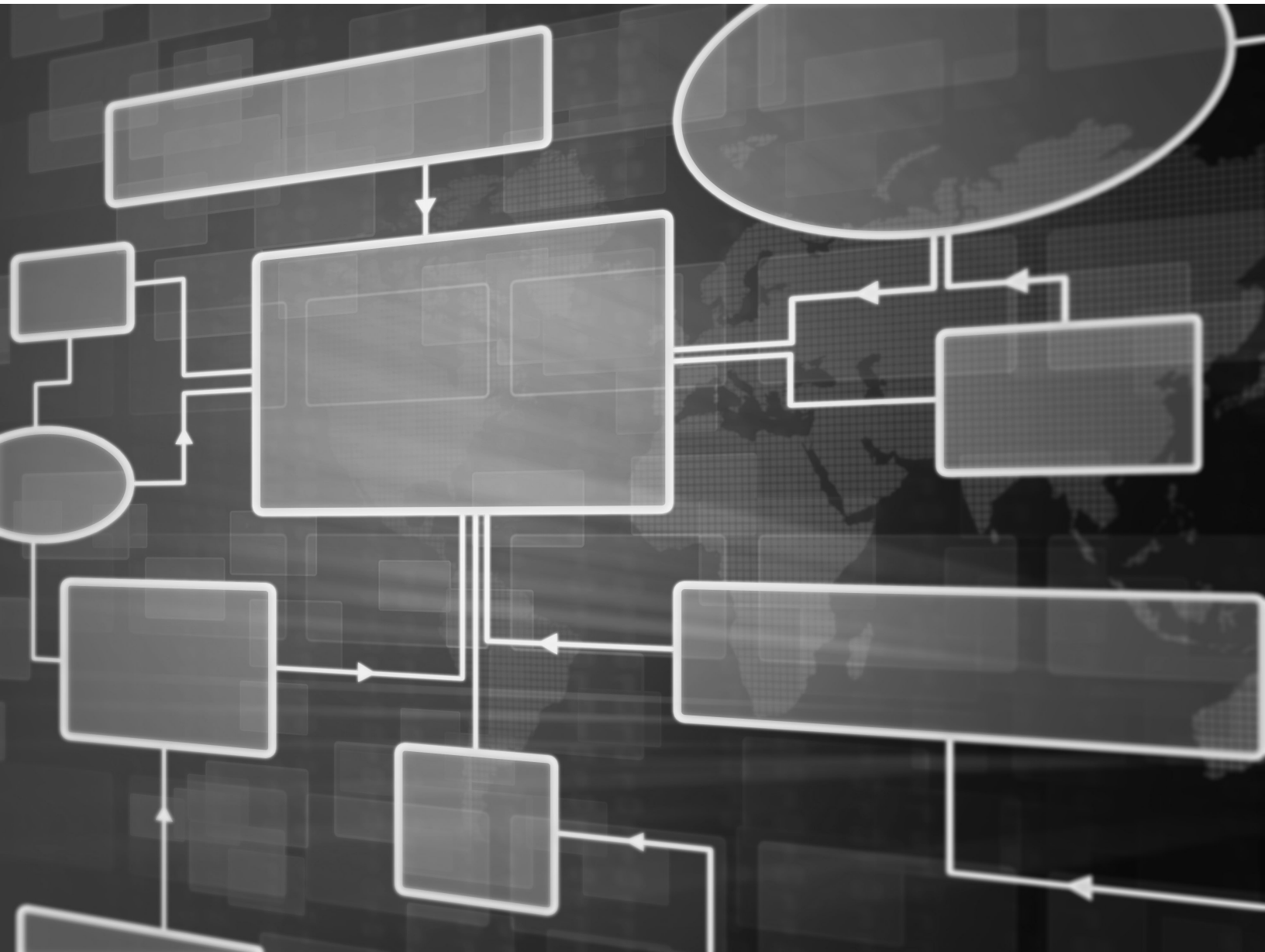
9. BIG DATA ECOSYSTEM TECHNOLOGIES

Big Data refers to extremely large data sets of varying types of data – structured, unstructured, and semi-structured – that can be collected, stored, and analyzed to provide insights for organizations. We asked three questions to see if and how people are using Big Data systems.

A. State of Big Data Management

First, we wanted to know if people were using Big Data ecosystem technologies. We asked [Figure 20]

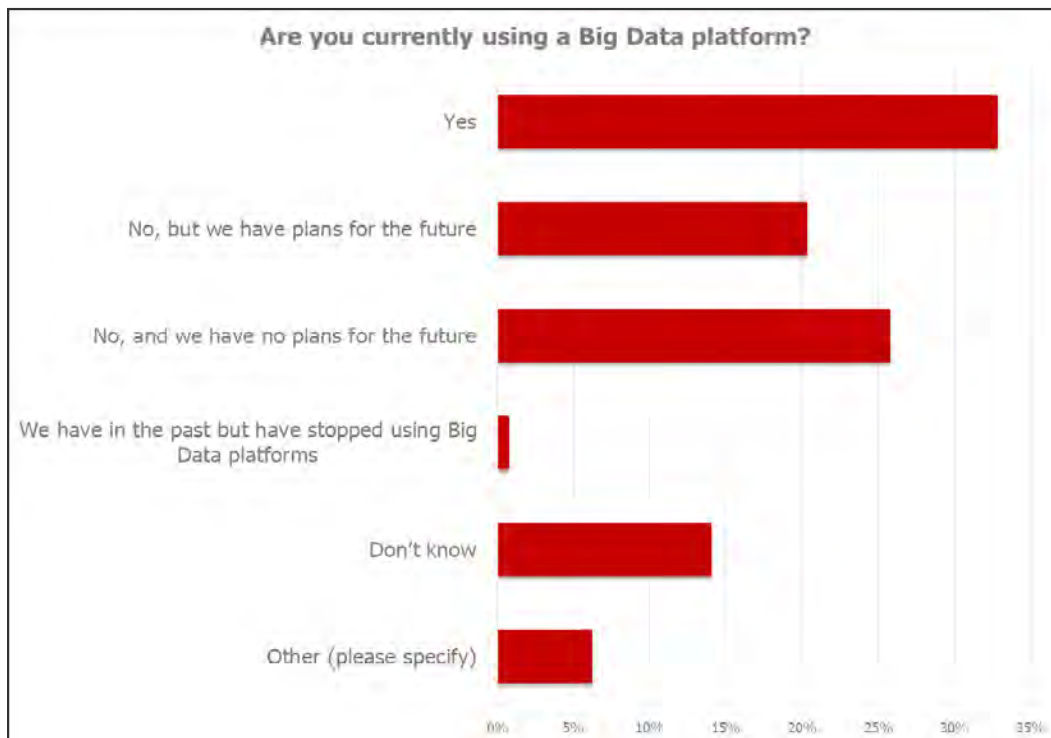▶ **"Are you currently using a Big Data platform?"**

Figure 20: Big Data Platforms

Percentage breakdowns were as follows:

- "Yes, we have a Big Data platform": 32.81%

- "No, and we have no plans for the future": 25.78%

- "No, but we have plans for the future": 20.31%

- "We have in the past and have stopped using it": 0.78%

Many of the 32.81% who said "yes" use Big Data platforms to gain insights from a wide range of data sources. This purpose differs between organizations, as the use cases in the following section will show.

B. Big Data Use Cases

To gain more clarity into why organizations were using Big Data technologies, we asked those who answered "yes" to having a Big Data platform the following [Figure 21]:

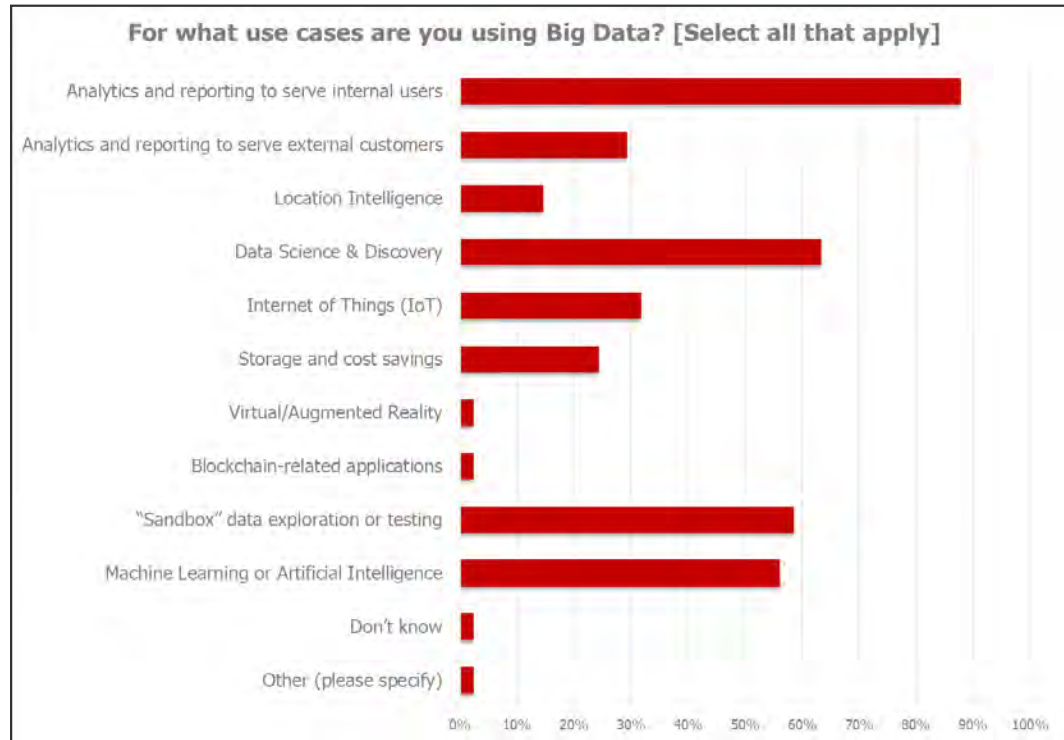▶ **"For what use cases are you using Big Data? [Select all that apply]**

Figure 21: Big Data Use Cases

The top Big Data use cases include:

- Analytics and reporting to serve internal users: 87.80%

- Data Science & Discovery: 63.41%

- "Sandbox" data exploration or testing: 58.54%

This survey also demonstrated that 56.1% use Big Data for machine learning or AI. Big Data is also applied (31.71%) to the Internet of Things (IoT). However, in general, Big Data platforms are less used with emerging technologies like blockchain and virtual/augmented reality (both under 3%).

Survey respondents strongly preferred to use Big Data for analytics and other reporting functions. This response is not surprising given the feedback received across multiple survey questions. For example, Figure 6 shows the majority have implemented Business Intelligence and reporting (87.02%) as well as data warehouses (86.55%).

C. Data Lakes

A Big Data platform typically utilizes a data lake– an environment where vast amounts and types of data can be ingested – to function properly. We wanted to know from those using a Big Data platform [Figure 22]:

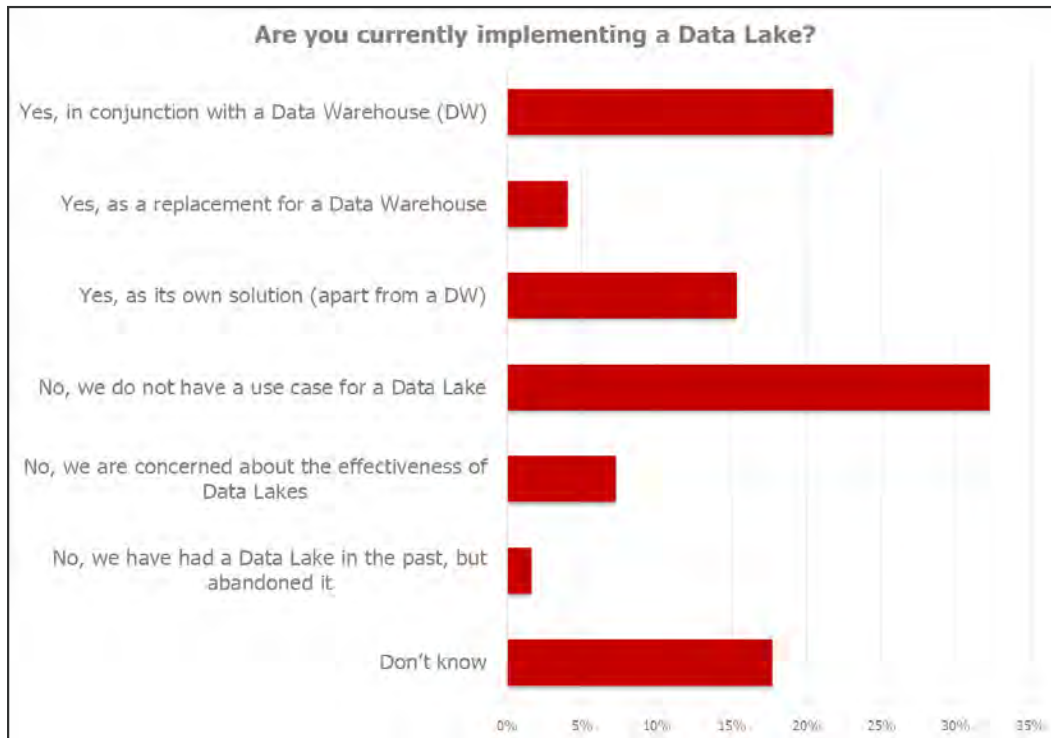► **"Are you currently implementing a Data Lake?"**

Figure 22: Data Lake Implementation

Top responses were as follows:

- Most of those surveyed did not see a use for a data lake: 32.26%

- About one fifth are implementing a data lake along with a data warehouse: 21.77%.

These numbers reinforce that data warehouse technologies will remain and evolve by embracing new technologies and methods. Quite a few organizations just do not require a Data Lake, and those that do may be using data warehousing alongside it.

D. Additional Comments

Participants also indicated confusion as to what Big Data means, noting that different people have different definitions. Perhaps more Big Data education and training across the organization would be useful.

Also, organizations continued to feel overwhelmed by Big Data technologies and unclear about what to do with them. One respondent commented:

- "Big Data is too large, complex, with too many technical decisions and layers."

10. DATA PLATFORMS AND STORAGE

This study wanted to know what data platforms and storage those surveyed were using in their organizations. The survey asked six questions. The first set centered around data sources or platforms used now as well as plans for the next one to two years. Two other queries sought out information on blockchain usage and were conditional on a "yes" answer from earlier questions. Where companies applied cloud technologies, two other questions were posed about the reasons and concerns moving to the cloud. Additional comments were provided to gain new understanding of data platforms and storage and its impact within Data Management.

A. Data Platforms

To gain a clearer understanding around platforms being used we asked [Figure 23]:

► **"Which of the following data sources or platforms are you currently using? [Select all that apply]"**
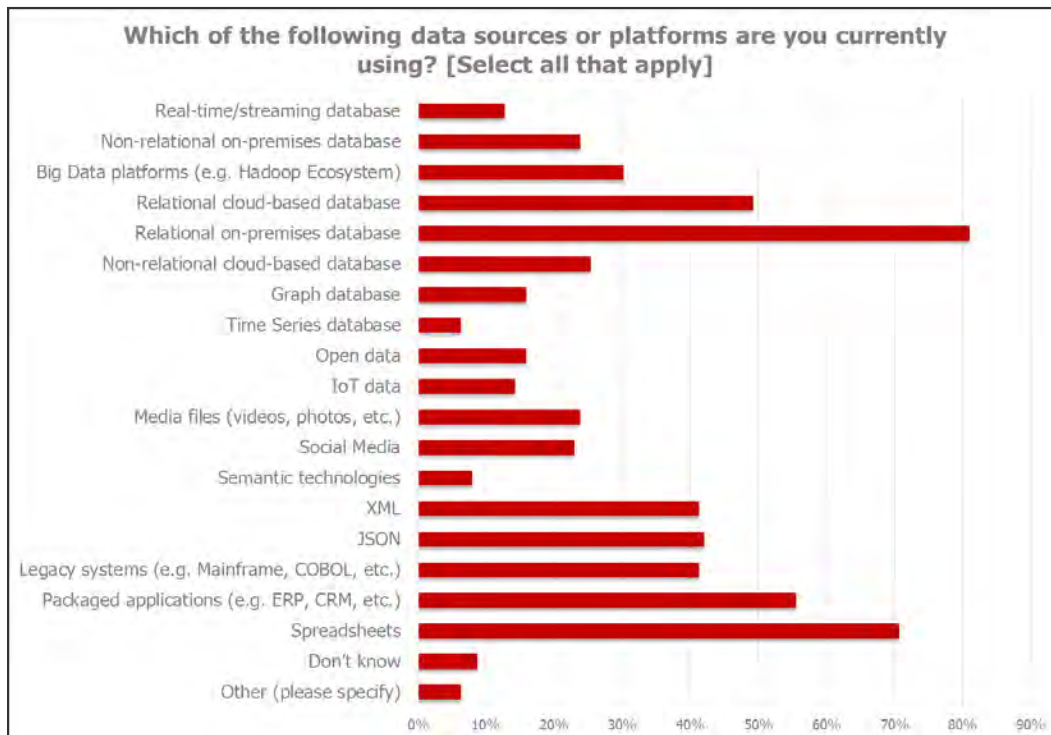

Figure 23: Data Sources and Platforms

Top data sources or platforms were as follows:

- Relational on-premise database (e.g. Microsoft Access): 80.95%

- Spreadsheets: 70.63%

- Packaged applications (e.g. ERP, CRM, etc.): 55.56%

- Relational cloud-based database: 49.21%

The least used technologies included:

- Time series database: 6.35%

- Semantic technologies: 7.94%

Quite a few respondents use a cloud-based database (74.61%), which includes relational (49.21%) and non-relational data systems (25.40%). The survey asked those who did use a cloud-based database about their reasons and concerns.

Spreadsheets, unfortunately, still see high usage due to their ease of use, particularly as more business roles create and get their hands-on corporate data. Reliance on spreadsheets stresses the need for better Data Governance and collaboration across teams and data sources.

B. Future Plans for Data Platforms and Storage

We asked the following to find out future platform and storage plans [Figure 24]:

▶ **"Which of the following data platform/data storage technologies do you plan to use in the next 1-2 years? [Select all that apply]"**
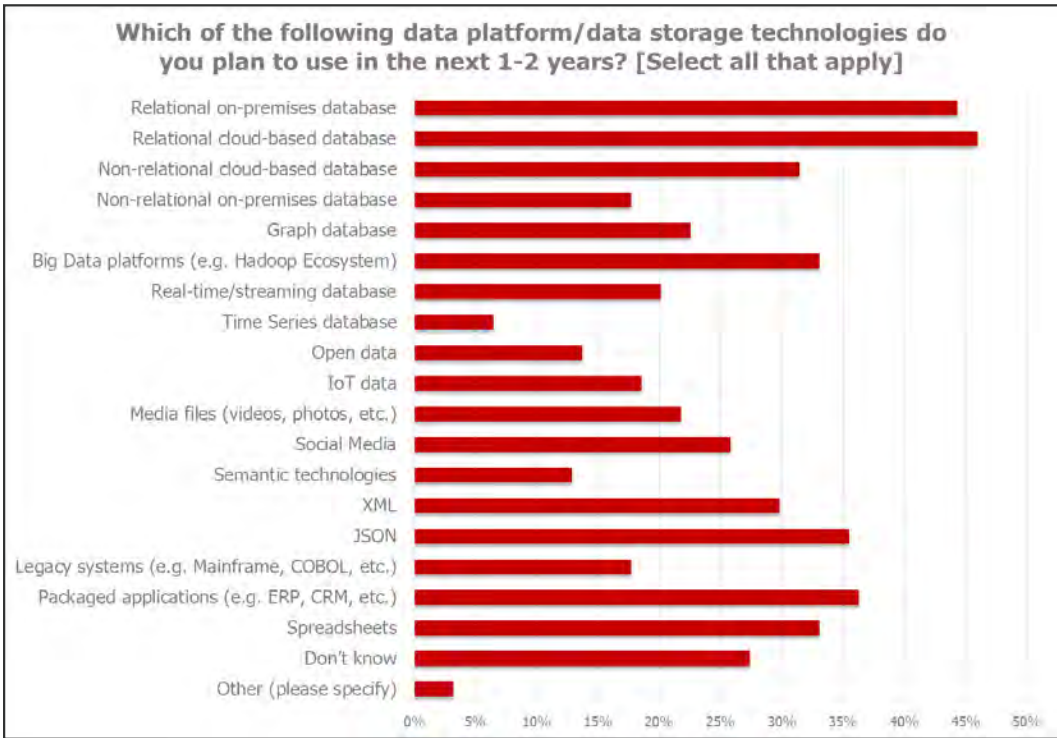

Figure 24: Data Sources and Platforms Plans in Next 1-2 Years

Top responses included:

- Relational cloud-based databases: 45.97%

- Relational on-premise databases: 44.35%

- Packaged applications (e.g. ERP, CRM, etc.): 36.29%

The least popular platforms or storage options included:

- Time series databases: 6.45%

- Semantic technologies: 12.90%

Relational databases will continue to be strong, with a higher percentage of these in the cloud. Anticipated spreadsheet usage (33.06%), though significant, is expected to decrease by 37% from the current use at 70.63%.

This drop, however, may be due to the fact that survey respondents aren't likely to admit to planning to use spreadsheets in the future, although the reality may be different. We can look to the future *2021 Data Management Trends* survey to see whether decreasing trends in spreadsheet usage pan out.

Although less preferred than relational databases, graph databases (22.58%), streaming (20.16%), video/media (21.77%), and IoT data (18.55%) will remain viable options among organizations. Note that usage across a number of platforms distributes evenly. This indicates organizations plan to leverage fit-for-purpose platforms available in a myriad of new technologies.

Some participants suggested future factors in data platform employment over the next few years. Written replies included:

- "Data storage will become decentralized, held in multiple forms and systems."

- "IT will have less control."

- "The cloud will continue to dominate and trend into the future."

> *"*
> *The cloud will continue to dominate and trend into the future.*
> *"*
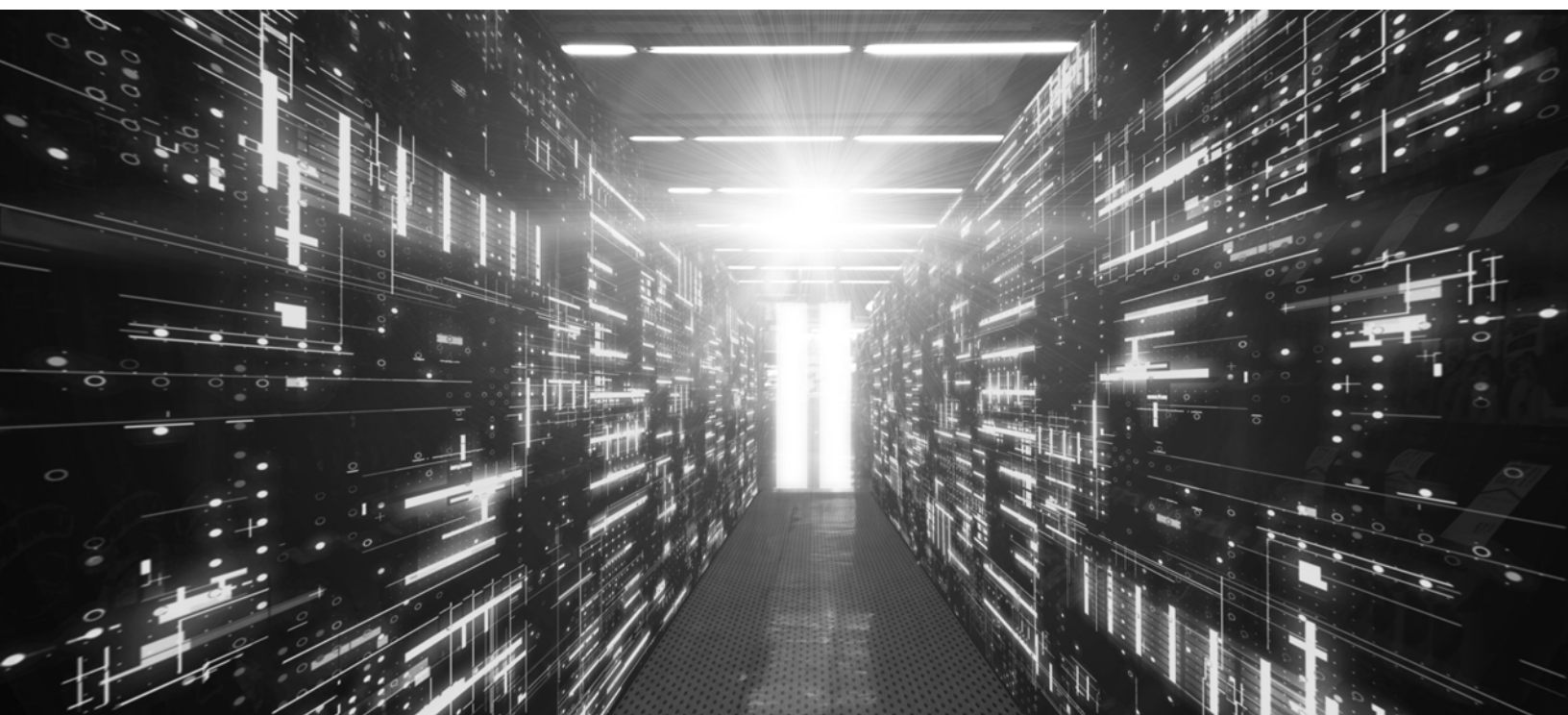
C. Cloud

As mentioned in the above section under "Data Platforms," the cloud dominates many survey participants' plans for tool usage. Cloud computing provides computation, software, data access, and storage systems without requiring knowledge of the physical location and configuration. For those who use cloud-based databases, we posed two questions. The first was [Figure 25]:

▶ **"What were your reasons for moving to the cloud? [Select all that apply]"**
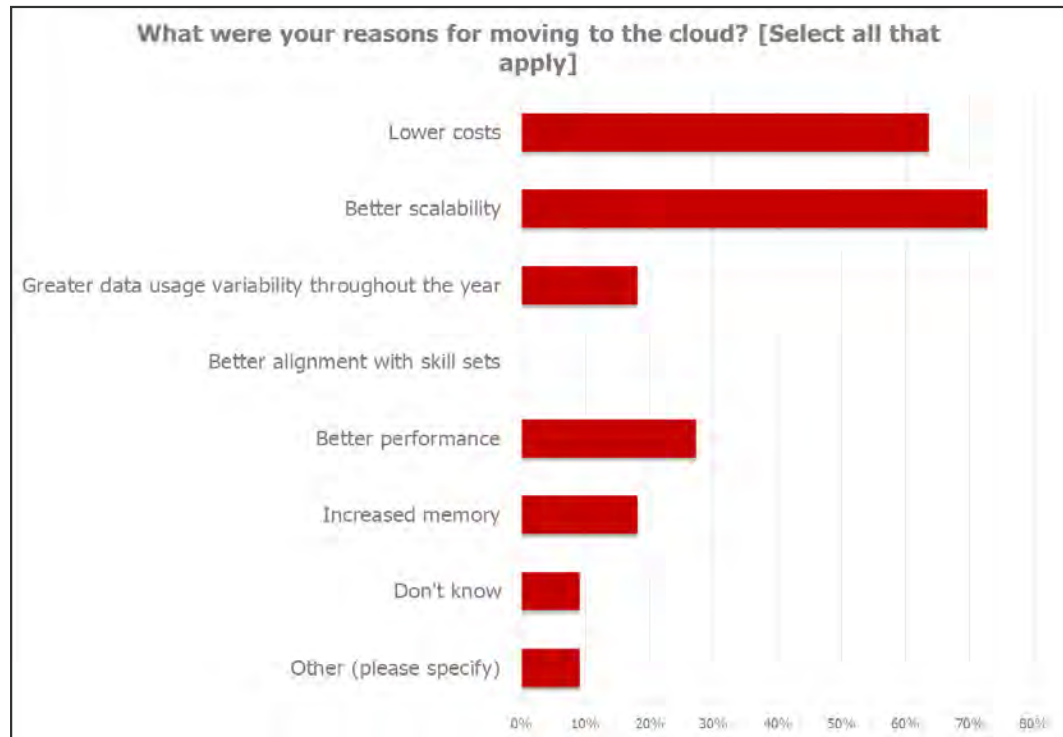

Figure 25: Reasons to Move to the Cloud

Those surveyed preferred the following reasons:

- Better Scalability: 72.73%

- Lower Costs: 63.64%

"Better alignment with skill sets" was not a consideration at 0%.

Other common responses included: better performance (27.27%), greater data usage variability throughout the year (18.18%), and increased memory (18.18%).

Scalability and lower costs clearly drive cloud use by many organizations, as does the flexibility offered by the cloud. Some possible case examples include:

- Retail companies who have high data volumes during peak sales seasons.

- Testing new applications, without investing time and costs for creating and/or maintaining an internal infrastructure**.**

Second, we asked respondents [Figure 26]:

► **"What are your concerns regarding moving data to the cloud? [Select all that apply]"**
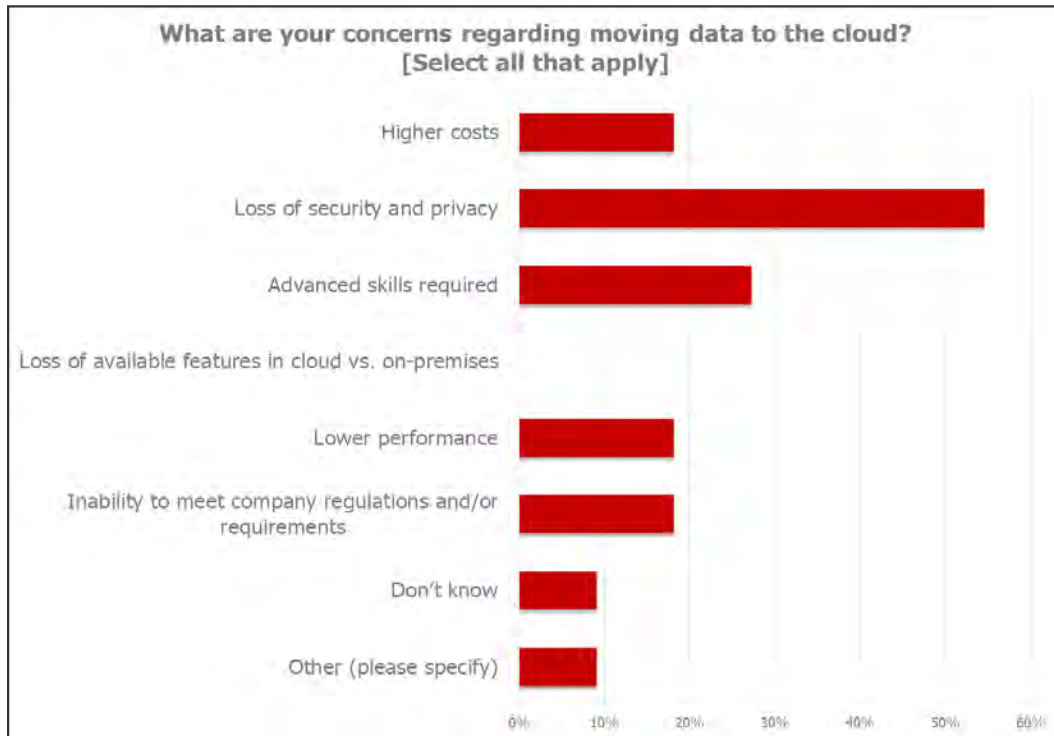

Figure 26: Concerns About Moving to the Cloud

Top concerns center around:

- Loss of security and privacy: 54.55%

- Needing advanced skills: 27.27%

None thought that loss of available features in cloud vs. on-premise would be an issue.

Securing data and respecting data privacy proved to be the largest concern. This is evident throughout this report. For example, Figure 6 demonstrated that data security has been implemented by 85.95% of the respondents' enterprises. Organizations must have some sort of Data Governance to be able to access and respect the security and privacy of strategic data assets.

There seems to be an unwritten expectation among people taking the survey that cloud applications will have the same or better capabilities as those on-premise. Cloud vendors, take note.

D. Blockchain

Blockchain – a combination of distributed databases and storage devices for data – has promise for adding trust in an untrusted environment. As a newer technology, we were curious as to its use and impact in Data Management. The survey asked [Figure 27]:

▶ **"Is your organization currently using blockchain technologies?"**
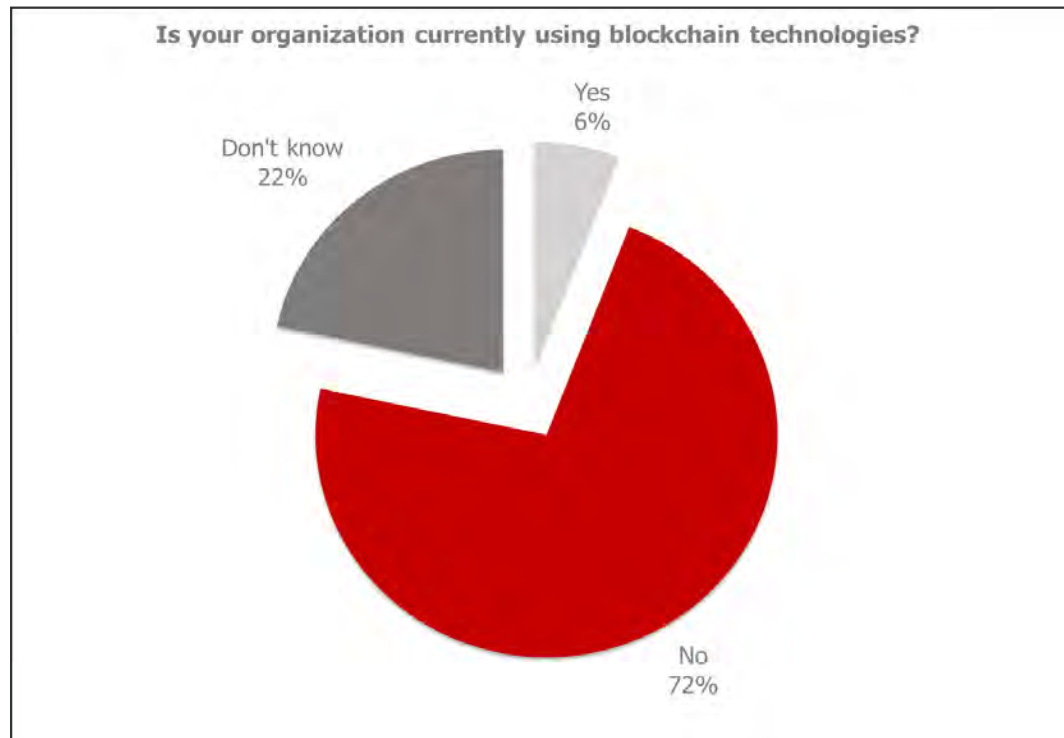

Figure 27: Blockchain Use

Responses were:

• No: 72.50%

• Don't Know: 21.67%

• Yes: 5.83%

Blockchain technology does not appear to be gaining wide adoption among survey participants as 72.50% said they are not using the technology. This continues the observations found in the DATAVERSITY *2017 Trends in Data Architecture Report* where only 7.7% said they would use or were actively using blockchain. This could be due to a lack of understanding, lack of use case necessity, or even security concerns, among many possibilities.

For those who said they used blockchain, DATAVERSITY wanted to know how they applied the technology. The following question was posed [Figure 28]:

► **"For which use cases are you using or considering use of blockchain technology? [Select all that apply]"**
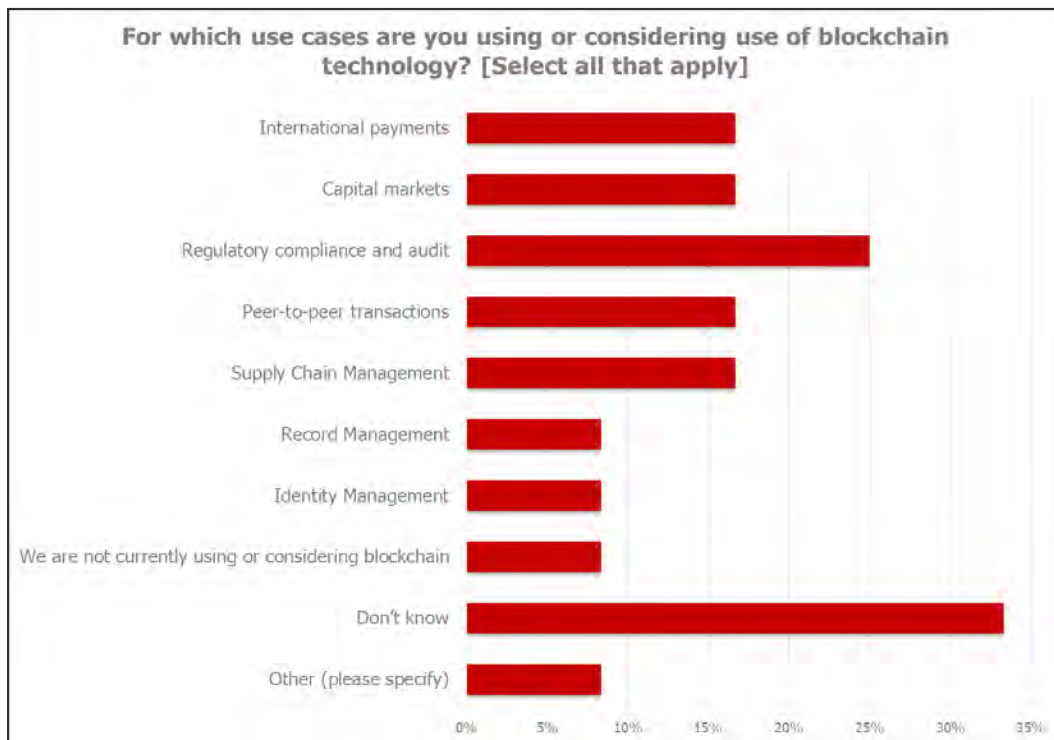

Figure 28: Specific Blockchain Use Cases

The top responses were:

- Don't Know: 33.33%

- Regulatory compliance and audit: 25%

An equal percentage (16.67%) use blockchain for international payments, capital markets, peer-to-peer transactions, and supply chain management. One respondent chose "Other" and commented:

- "We support clients to implement hyperledger as blockchain solutions."

E. Concluding Comments

With the wide range of options in data technologies, organizations have a wide range of options to choose from in supporting the data-driven organization. While relational databases still dominate, usage is being augmented by a wide range of fit-for-purpose technologies to fit specific use cases.

While the number of choices can be daunting, innovation in the industry provides an unprecedented array of technology choices. It is certainly worth investigating some of these new technologies in your organization as usage and use cases for data continue to expand.
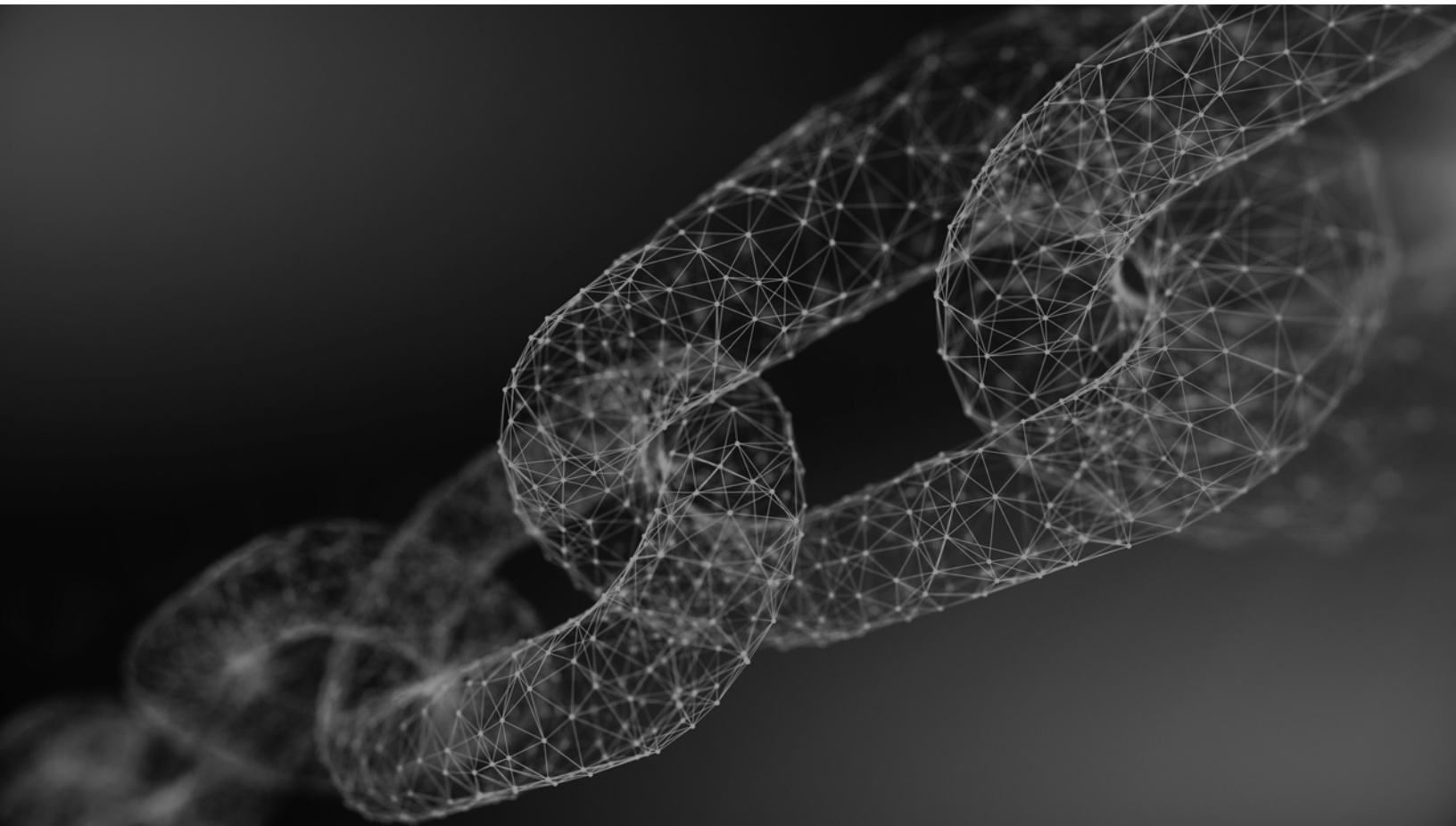
11. NEW TECHNOLOGIES AND TRENDS

Data-centric technologies are innovating at a rapid pace, and it can be difficult to keep track of the wide selection in the market. This section looks at newer technology trends and their adoption, including:

- Internet of Things (IoT)

- Containerization and Kubernetes

- Serverless computing practices such as PaaS and FaaS

- Data automation advances

- AI, machine learning, and deep learning

- Digital twins, which provide the virtualization of physical products or assets

Included in that discussion, and as an overarching theme, is Industry 4.0.

According to MIT's research community and several industry leaders, Industry 4.0 describes "smart factories." This term, first coined by the German government in 2003, describes a new revolution where sensors generate enormous volumes of data that can track every manufacturing process stage and whole production systems.

Forbes writer Bernard Marr describes Industry 4.0 implementation as interconnected computer networks that communicate and make decisions without the need for human intervention. Logistics and supply chains will improve from Industry 4.0 digital transformation.

DATAVERSITY wanted to know how much participants and their organizations were aware of Industry 4.0, its business impact, and related technologies. Each of the above trends promises to progress Data Management to the next level, but how widespread is their use today and within the next couple of years?
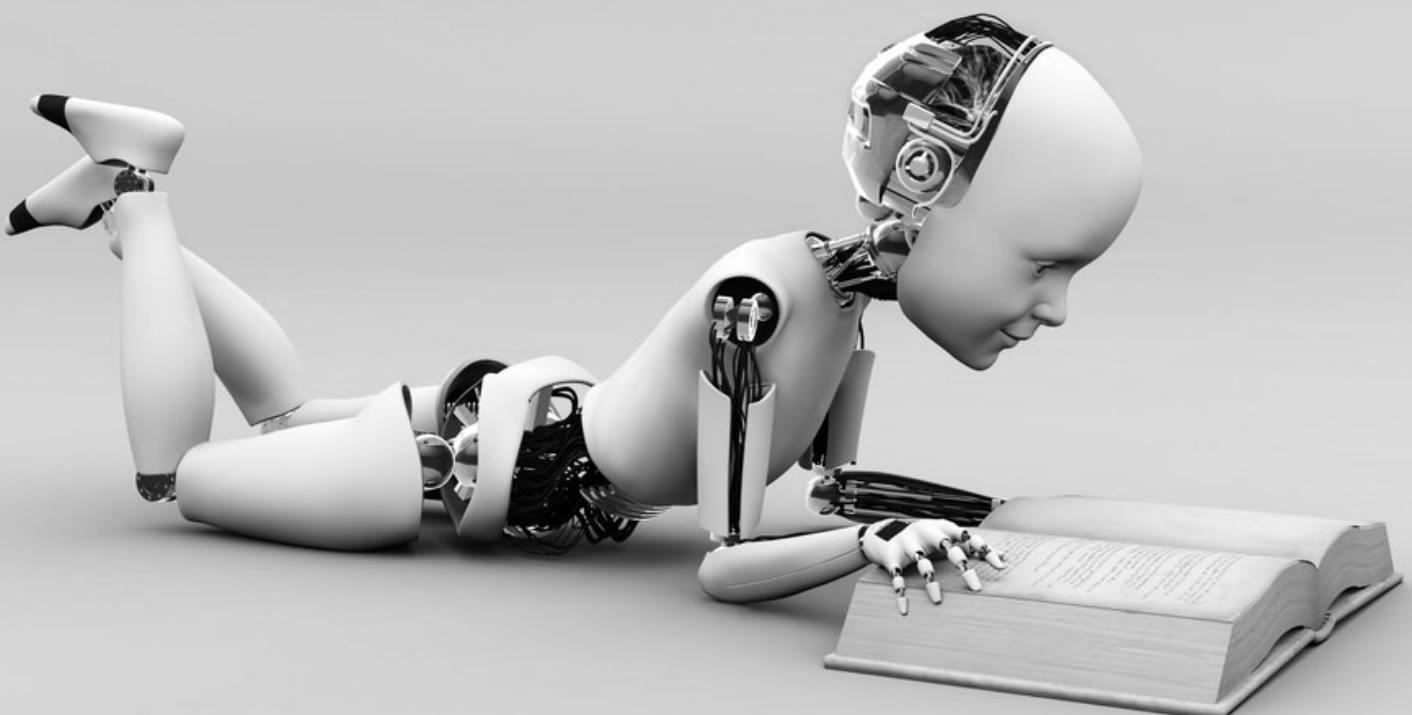
A. New Technologies

We asked participants [Figure 29]:

▶ **"Which of the following is your organization currently leveraging or planning to implement within 1-2 years? [Select all that apply]"**
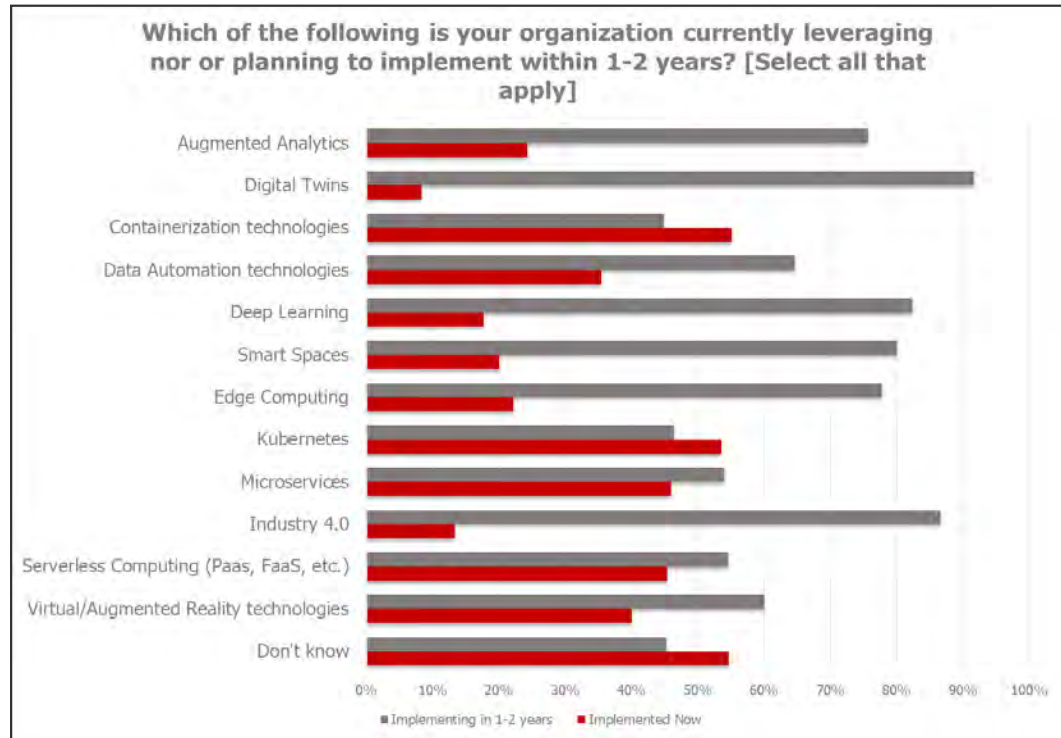
Figure 29: New Technologies Leveraged or Planning to Implement

Of those that did know which technologies had already been leveraged, respondents mentioned the following emerging technologies most often:

- Containerized technologies: 55.17%

- Kubernetes: 53.57%

- Serverless Computing (PaaS, FaaS, etc.): 45.45%

Businesses have yet to implement to the same degree:

- Deep Learning: 17.65%

- Industry 4.0: 13.33%

- Digital Twins: 8.33%

But, when looked at in terms of expectations about what would be implemented in the next few years, those three choices show great gains, along with smart spaces – visual and audio sensing systems that perceive and react to people:

- Digital Twins: 91.67%

- Industry 4.0: 86.67%

- Deep Learning: 82.35%

- Smart Spaces: 80.00%

Organizations clearly prefer new technologies that drive the development of Industry 4.0, the transformation of current business with autonomous systems, and machine learning. Networked machines are expected to do most of the work. The existing implementation of containerized technologies and Kubernetes allow applications to work flexibly in compatible operating systems and run reliably between computer systems. This is the beginning of getting machines to collaborate more efficiently, helping organizations to manage, orchestrate, and govern.

The high drive to implement digital twins (91.67%), Internet of Things 31.71% (also refer to Figure 21), and smart spaces (80%) in the next one to two years suggests that industrial use cases clearly lead the way.

Consideration that any work can be replaced by intelligent networks of machines will spill over into business analytics. Modern analytic approaches such as Deep Learning (82.35%) have ranked highly from survey participants. Participants would like to gain better insights faster using Artificial Intelligence.

Many people (54.72%) don't know what their organization is implementing. This may indicate room for education opportunities within organizations about Industry 4.0 and the technologies used to support it.
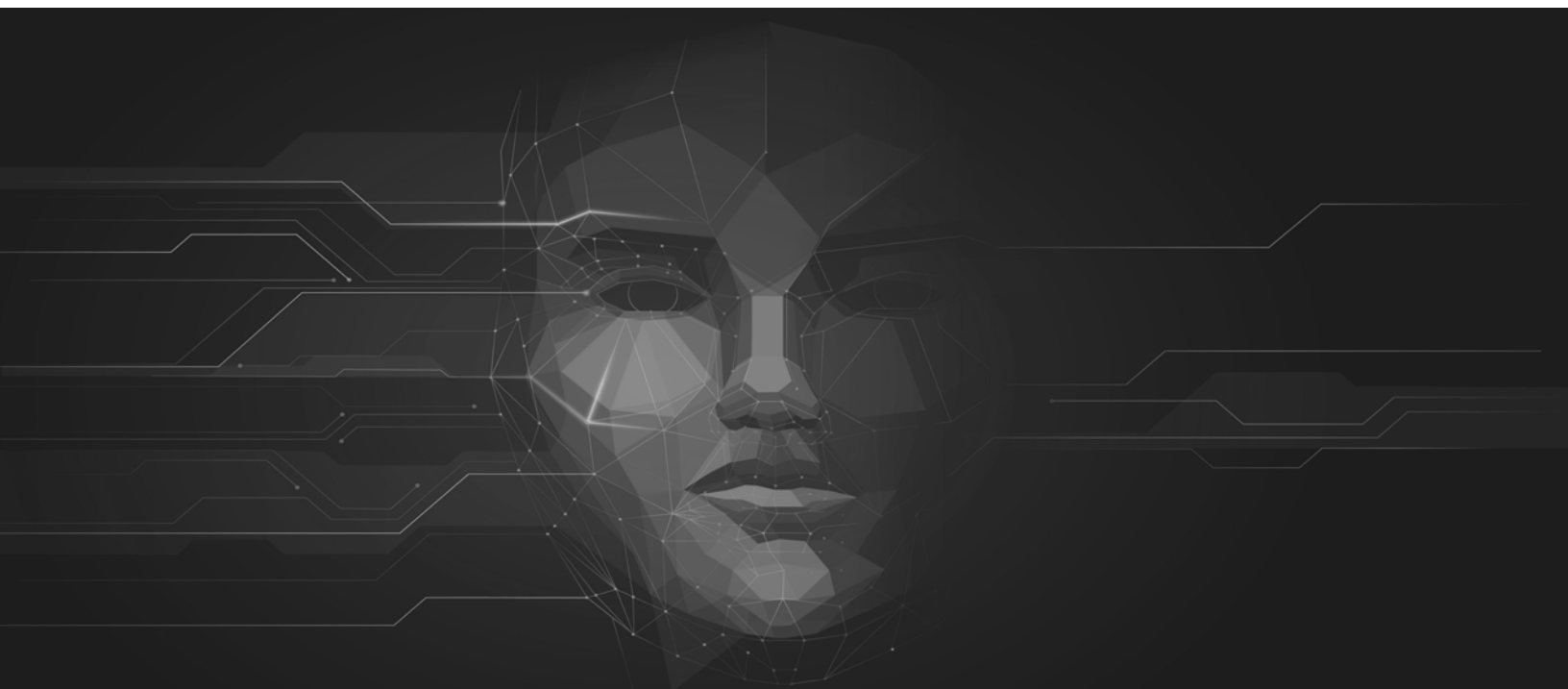
> *"*
> *We are drowning in too much data and this will grow by several orders of magnitude. We are going to have to figure out better systems with emerging technologies like AI/ML and other approaches to even have a hope to manage all of it.*
> *"*

B. Future Trends

The survey asked respondents the following open-ended question to finish out the survey:

- "What do you see as the next top three emerging trends in Data Management that will cause the most profound changes in the industry? [open-ended question]"

Comments strongly agreed that technologies backing Industry 4.0 and IoT will be disruptive. System foundations for Industry 4.0 and IoT generated the most discussion. This is true especially for the cloud, machine learning, and data virtualization. As one respondent stated:

- "We are drowning in too much data and this will grow by several orders of magnitude. We are going to have to figure out better systems with emerging technologies like AI/ML and other approaches to even have a hope to manage all of it."

Security and regulation will not go away and will cause some deep changes. Data Management will need to adapt to an ever-changing political environment. Respondents anticipate an increased trend toward ethical and trustworthy data.

12. CONCLUSION

In today's digital economy, more and more organizations are understanding the importance of data as a strategic asset, driving an increased interest in Data Management. As organizations look to gain increased insights around current and future trends through data, they also see the need for increased collaboration across roles and stronger Data Governance to protect data assets. With increased value and opportunity from data comes increased risk, which drives the need for expanded Data Governance and related efforts such as Metadata Management and Data Quality.

While innovation in the Data Management industry is growing at a rapid pace, there continues to be a focus on fundamental technologies and methods such as relational databases, metadata, and Data Management tools. Technological advances, such as a move to the cloud, have made these technologies more accessible and easier to use with expanded functionality, but core use cases continue to focus on increased business insight and greater risk protection through Data Governance.

Advances in technology – such as Industry 4.0, AI, and machine learning – will continue to offer new opportunities and challenges, but one of the largest factors in determining the effectiveness and scope of Data Management efforts continues to be the "people factor." The success of a data-driven organization lies heavily on building a data-driven culture which is supported by a Data Governance framework outlining clear roles and responsibilities for data.

Finding the right balance of technologies amid a growing array of choices and establishing the right roles and responsibilities across an even wider range of stakeholder roles will prove to be key in building a successful data-driven organization.

SPONSORS

IDERA

IDERA provides comprehensive database tools that help IT professionals improve data quality, boost performance, and ensure data integrity across the complex environments. A community of more than 100,000 users in healthcare, financial services, retail and technology rely on IDERA products to optimize business data and take advantage of new business opportunities. IDERA is a division of IDERA, Inc.

ABOUT THE AUTHORS

DONNA BURBANK is a recognized industry expert in information management with over 20 years of experience helping organizations enrich their business opportunities through data and information. She currently is the Managing Director of Global Data Strategy Ltd, where she assists organizations around the globe in driving value from their data. She has worked with dozens of Fortune 500 companies worldwide in the Americas, Europe, Asia, and Africa and speaks regularly at industry conferences. She has co-authored several books on Data Management and is a regular contributor to industry publications. She can be reached at donna.burbank@globaldatastrategy.com, and you can follow her on Twitter @donnaburbank.

MICHELLE P. KNIGHT is a freelance writer with a Master's in Library and Information from Simmons College and a software testing background. She has written about Data Management, Information Technology, and Information Management, including a wide range of articles for DATAVERSITY. She has also written about W3C Validation for Site Pro News, SEO Competitive Analysis for a Special Library Association, Future Ready 365 blog, and a world of blogs and search engines for a Librarian Journal. Michelle is a Smith College alumna who enjoys martial arts and writing creatively in her spare time. She can be reached at mknight113071@gmail.com.

Picture taken by Photo Yoshino