Trends in Data Management

A 2022 DATAVERSITY® Report

Donna Burbank and Keith D. Foote

TABLE OF CONTENTS

1. EXECUTIVE SUMMARY

In today's data-driven digital economy, organizations are increasingly looking for competitive advantages through reporting, analytics, and operational efficiencies. While this has been true for many years, there is an increasing maturity in the Data Management space as more organizations look to focus on Data Governance, Data Quality, and Data Security to ensure a solid data foundation for these efforts.

While Digital Transformation continues to be a priority for organizations, the importance has shifted toward using digital technologies and data to support customer satisfaction, rather than as an effort in and of itself.

With data becoming increasingly important as a business asset, more business stakeholders are looking to take an active role in Self-Service Reporting as well as Data Governance. To meet this demand, Data Literacy is becoming increasingly important, and was cited in the DATAVERSITY *2022 Trends in Data Management Survey*, as a current lack in the industry. Similarly on the technical side, data-centric skills are in high demand, with a cited gap in skills resources to meet this demand.

Some of the highlights of this paper include:

- 43.29% of the participants surveyed claimed they have an enterprise-wide effort around Data Management.

- When asked about the reasons for implementing a Data Management program, the top two choices were "gaining insights from Data Analytics" (74.29%) and "saving costs and increasing efficiency" (66.43%).

- A mature Data Governance program was expressed as a goal for many, and the use of metadata is an emerging interest by those not already yet using it.

- The lack of Data Literacy and a skilled workforce remains a source of frustration.

- On-premises relational databases were clearly the most popular form of platform currently being used at 75.00%. Spreadsheets are still at 70.16%. 21.77% are using on-premises, "non-relational" databases.

- 51.97% of the respondents claimed their organization was actively using Data Modeling.

- Data Mesh and Data Fabric were listed as top Data Management priorities for 2022–2023.

2. RESEARCH AND DEMOGRAPHICS

A. Scope of Research

DATAVERSITY's *2022 Trends in Data Management Report* offers insights about the directions and concerns businesses have as Data Management continues to evolve. The overall structure and methodology of the 2022 study is similar to those of previous years from 2019 to 2021. Previous *Trends in Data Management* surveys will be compared with this year's when relevant to determine long-term trends.

This year's survey had a total of 181 participants, from 35 countries, and over 32 industries.

The majority of respondents worked with Information/Data Governance (35.67%), followed by Data Information and/or Data Architecture (21.05%). People working with Business Intelligence, Analytics, and Data Science came in third (14.04%), and consultants came in fourth (13.45%). The remaining participants ranged from business executives to academic researchers. A broad selection of industries, from aerospace to recreation were represented. Businesses of all sizes were surveyed.

The survey had 35 questions. Three of the questions were open-ended, with the remaining questions offering a selection of answers, and instructions for some to check-off all answers that apply. The questions offering checked-off responses typically included an "Other (please specify)" box, with space available for a short description. Those comments will be included where relevant to the analysis.

The survey was broken down into eight sections, with each section followed by a space for additional comments. The eight sections of questions are:

- General Demographics (three questions)

- The Current State of Data Management (five questions)

- Goals, Drivers, and Training (six questions)

- Data Governance and Metadata Management (four questions)

- Data Architecture (four questions)

- Data Modeling (four questions)

- Data Platforms and Storage (five questions)

- New Technologies and Trends (four questions)

For the survey, participants were recruited using an email campaign and the resources of DATAVERSITY's Data Education Month. Participants received links for the 2022 *Trends in Data Management* Survey and there was no time limit for answering the questions. The participants did not receive any monetary compensation; however, they did receive compiled preliminary results from the study. The survey's responses and comments are representative of the respondents, only.

B. Principal Demographics

The survey's respondents were initially asked three basic demographics questions regarding their job function, their industry, and how many employees work at their company.

1. Job Function

The majority of respondents in the 2022 survey (and previous surveys) held data-centric positions, such as Data Architecture, Analytics and Business Intelligence, and Data Governance [Figure 1].
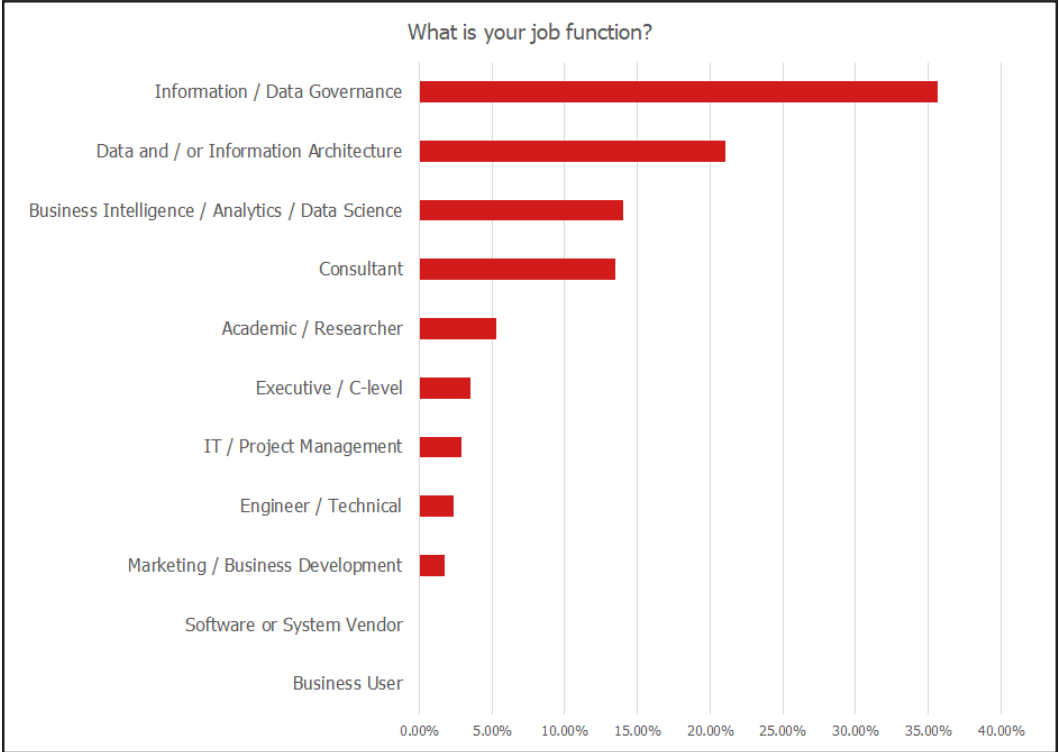

Figure 1: Job Function

Separating the participants by profession, the top three career groups responding in 2022 are:

- Information/Data Governance: 35.67%

- Data and/or Information Architecture: 21.05%

- Business Intelligence and/or Analytics: 14.04%

2. Industry Representation

The 2022 study represents a broad range of industries, ranging from retail to healthcare to government. The participants in this survey represent over 32 industries [Figure 2].
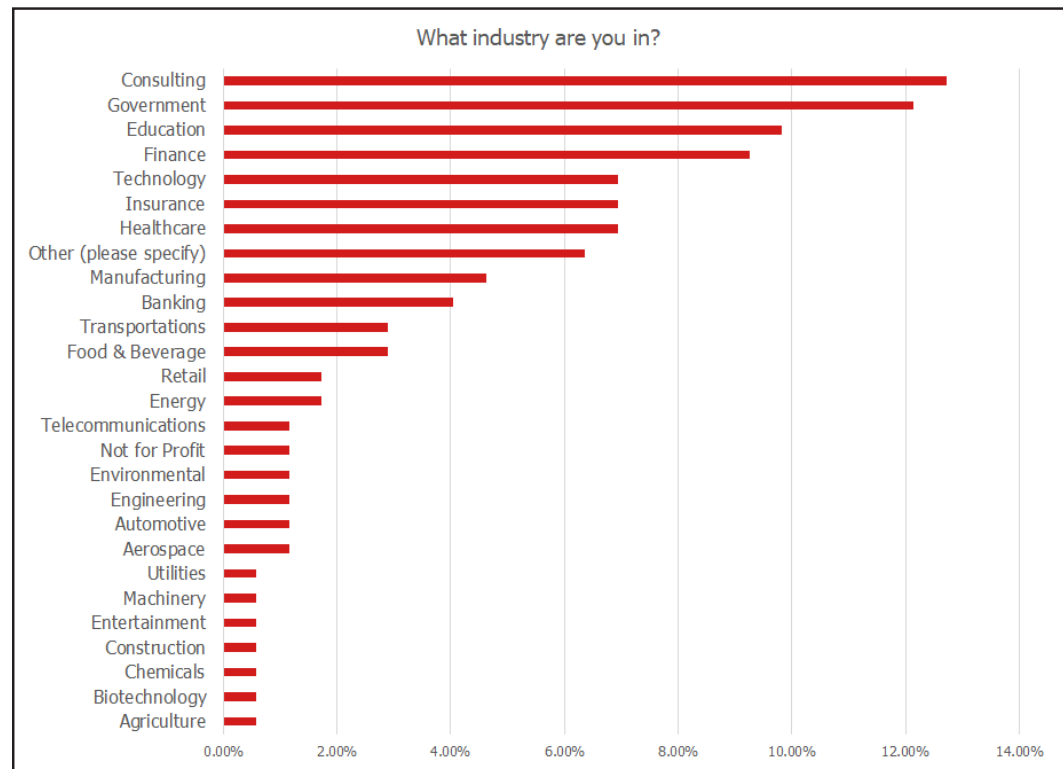

Figure 2: Industry

The top industries represented in the study are:

- Consulting: 12.72% (12.28% in 2021)

- Government: 12.14% (14.91% in 2021)

- Education: 9.83% (6.14% in 2021, 3.83 % in 2020)

- Finance: 9.25% (8.33% in 2021)

The education industry has seen an increase in participation over the last few years, becoming one of the top four industries responding to the survey. The consulting and finance industries have remained fairly consistent since last year's survey, and government, as an industry, has , slightly decreased participation (a drop of 2.77%).

Other industries that were represented on the list of offered options include not-for-profits, environmental groups, entertainment, utilities, and construction.

3. Company Size

Although businesses of many sizes participated in the survey, the majority of the respondents came from very large businesses, with 10,000 or more employees (25.43%). Slightly smaller businesses, with 1000–4999, came in second (24.86%). Mid-sized businesses, with 100-999 employees, came in third (24.27%) [Figure 3].
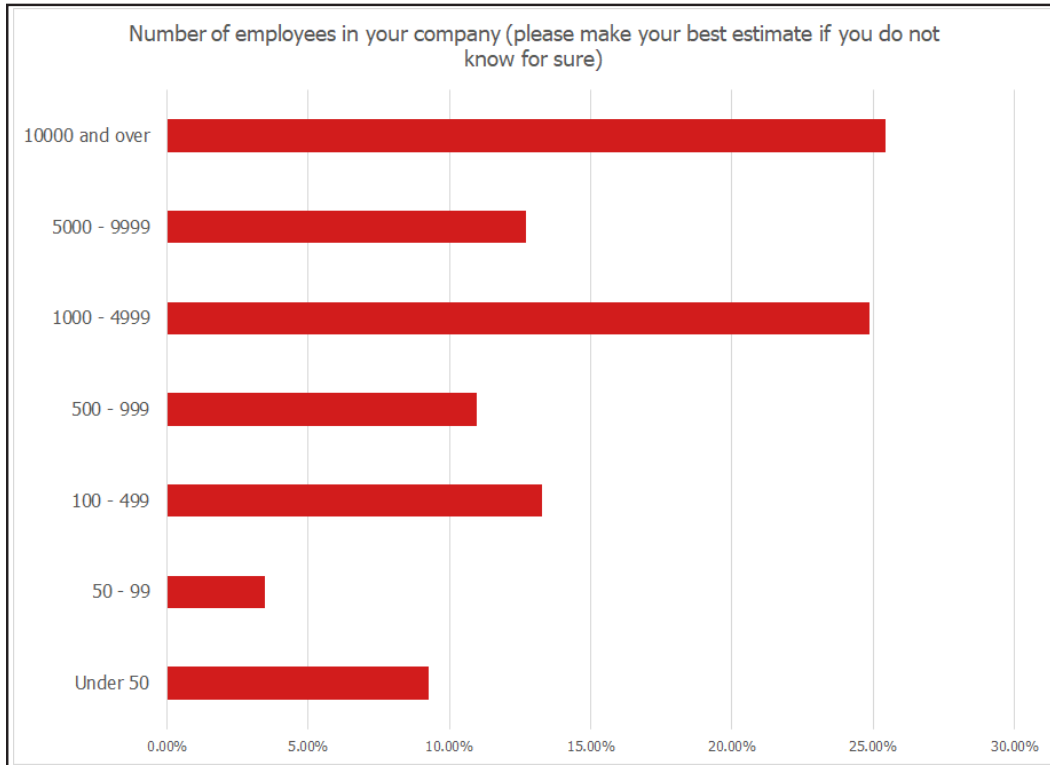

Figure 3: Number of Employees

Some industries listed in the "other (please specify)" section were: oil & gas exploration & production, medical instruments, and legal.

3. THE CURRENT STATE OF DATA MANAGEMENT

DAMA International's Data Management Body of Knowledge (DAMA DMBoK) has the recognized industry-standard Data Management definition. It says:

> "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 life cycles."

To gain a deeper and more holistic understanding of Data Management's scope, its uses, and its planned uses, DATAVERSITY asked four questions.

A. The Scope of Data Management

To understand how broadly Data Management was being used by organizations, the survey asked [Figure 4]:

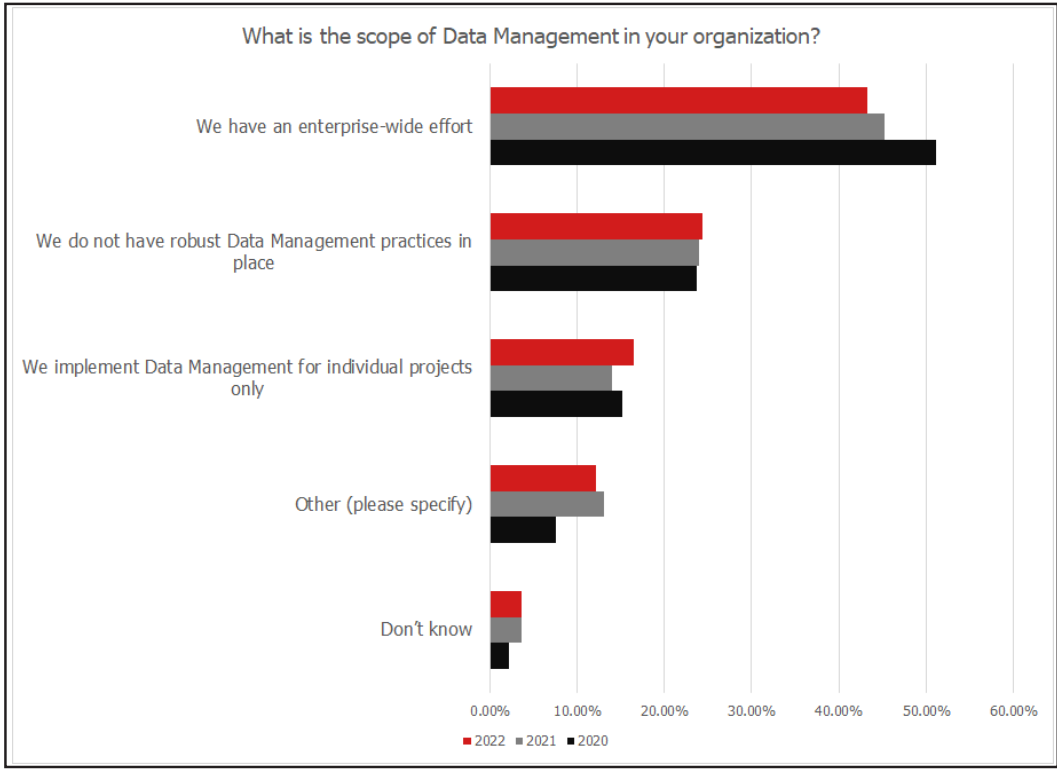▶ **"What is the scope of Data Management in your organization?"**

Figure 4: Scope of Data Management (2020 – 2022)

As in previous years, the most popular response is "we have an enterprise-wide effort," showing the overall maturity of the industry. At the same time, however, there is also a consistent percentage of respondents (24.39% in 2022, 23.98% in 2021, and 23.77% in 2020) that do not yet have robust Data Management practices in place, indicating a continuing need for Data Management solutions, particularly as industries and organizations that have not traditionally been data-driven become more involved with data and analytics.

Responses in the "other (please specify)" category came in at 12.20%. However, the respondents did not list alternatives to Data Management, but generally speaking, offered responses that might have fit in the "we do not have robust Data Management practices in place" category. Some examples of responses in the "other (please specify)" are:

- "We are currently doing it at the project level with the intent to establish an enterprise Data Management effort."

- "We are trying to have an enterprise-wide effort but it is still functioning as project specific activities and is NOT robust."

- "We are in the process of implementing a robust, enterprise-wide Data Management effort."

If responses to "other (please specify)" category are added to the "we do not have robust Data Management practices in place," the percentage of responses may increase to as high as 36.59%. This suggests a continuing need for Data Management solutions.

B. Roles Driving Data Management

The responsibilities of developing a Data Management program have been assigned to a wide range of job descriptions. To gain an understanding of who is guiding and developing Data Management, we asked [Figure 5]:

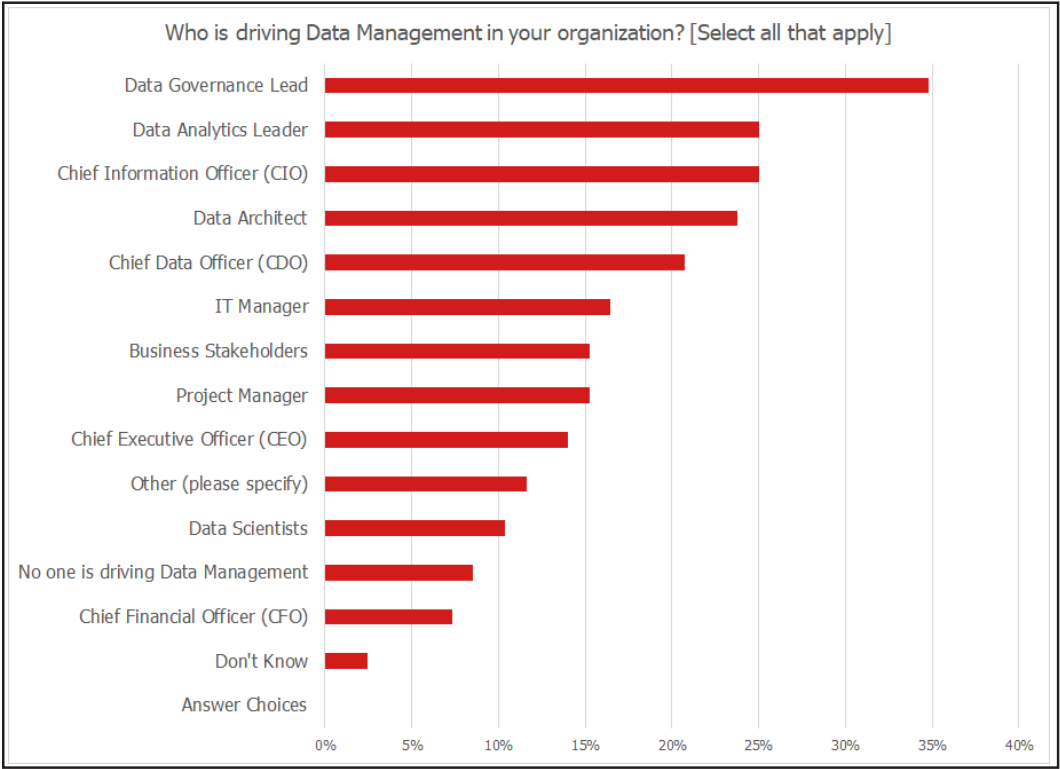▶ **"Who is driving Data Management in your organization?"**

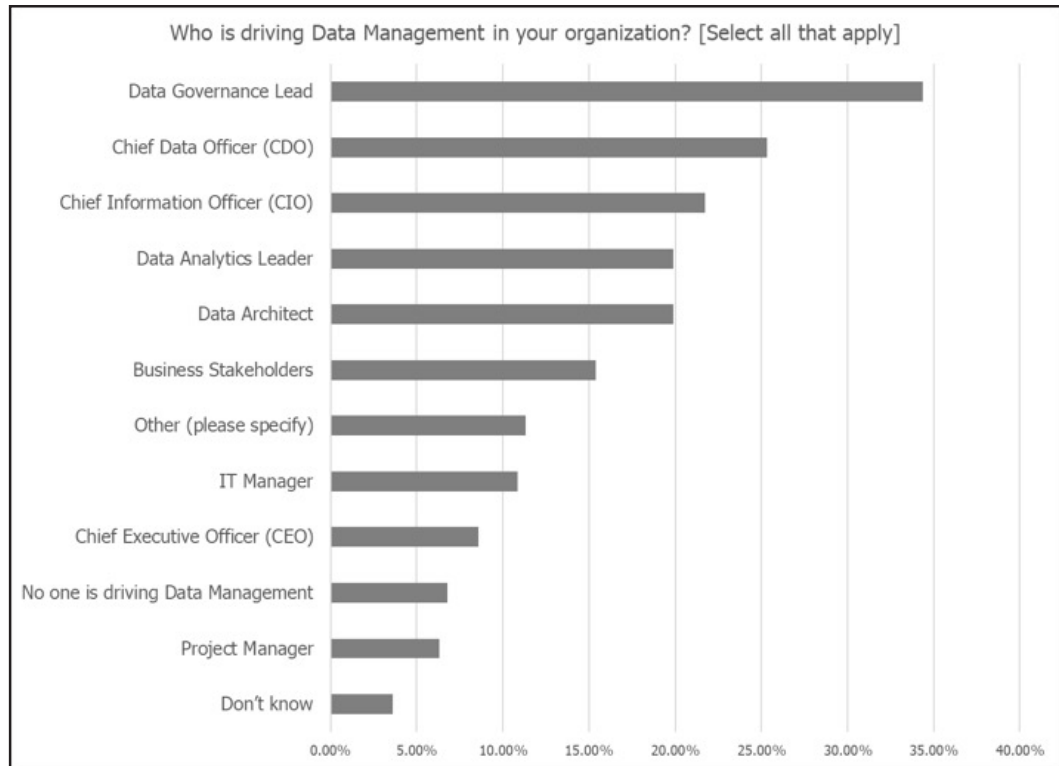Figure 5a: Who is Driving Data Management? (2022)

Figure 5b: Who is Driving Data Management? (2021)

The following list represents the Top 5 answers in 2022 for who is driving Data Management:

• Data Governance Lead: 34.76% (34.39% in 2021)

• Chief Information Officer (CIO): 25.00% (21.72% in 2021, 27.60% in 2020)

• Data Analytics Leader: 25.00% (19.91% in 2021, 18.55% in 2020)

• Data Architect: 23.78% (19.91% in 2021, 25.34% in 2020)

• Chief Data Officer (CDO): 20.73% (25.34% in 2021, 22.62% in 2020)

The Data Governance Lead continues to be seen as a driver for Data Management, with a similar percentage from 2021. A strong Data Governance Lead plays a critical role in aligning business and IT roles across the organization, which makes them well-placed as a driver for change in Data Management.

While the Top 5 roles remained unchanged from 2021, notable is that the Data Analytics Leader increased by 5% while the Chief Data Officer (CDO) decreased by a similar amount. The strong focus on Data Analytics in 2022 may explain this change.

Another promising trend is the number of C-Level roles involved in a Data Management effort from CIO to CFO, COO, and even CEO. As data continues to be seen as a business asset, we expect this trend to continue.

C. Components of Data Management

We wanted to know which Data Management components organizations are using today and their plans for the future. First, we asked [Figure 6]:

► **"Which of the following have you already implemented in your organization?"**
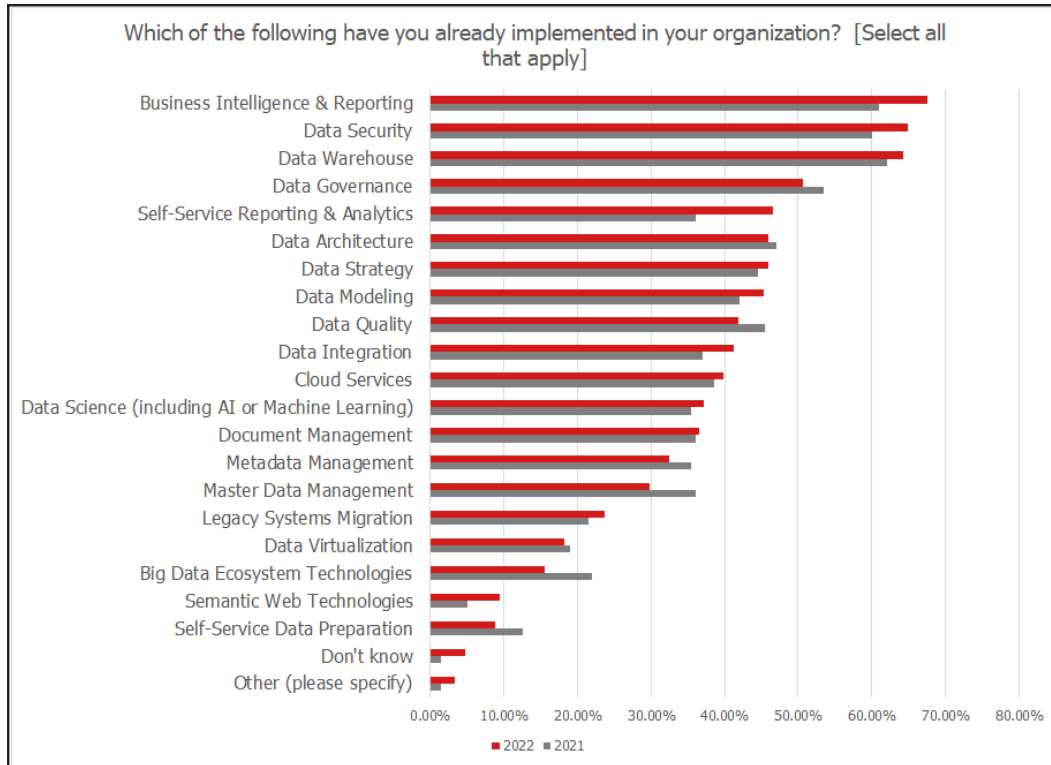

Figure 6: Current Implementation of Data Management (2021 & 2022)

The Top 5 practices implemented in 2022 were:

- Business Intelligence and Reporting: 67.57% (61.00% in 2021, 71.76 % in 2020)

- Data Security: 64.86% (60.00% in 2021, 60.00% in 2020)

- Data Warehouse: 64.19% (62.00% in 2021, 69.41% in 2020)

- Data Governance: 50.68% (53.50% in 2021, 52.94% in 2020)

- Self-Service Reporting: 46.00% (36.00% in 2021, 40.00% in 2020)

According to respondents, in 2022 these practices were implemented the least:

- Big Data Ecosystem Technologies: 15.54% (22.00% in 2021, 20.00% in 2020)

- Semantic Web Technologies: 9.46% (5.00% in 2021, 2.35% in 2020)

- Self-Service Data Preparation: 8.78% (12.50% in 2021, 8.24 in 2020)

Business Intelligence and Reporting, with Data Warehousing as a supporting foundation, continue to lead the initiatives underway, as they have in past years. Data Governance and Data Security have also continually been leaders as organizations look to protect, manage, and secure the data assets needed for this reporting and analysis.

A notable increase in 2022 is the rise in Self-Service Reporting and Analysis (increasing by 10.62%), as more business stakeholders look to be hands-on in driving their own Data Analytics. This is a positive sign for the rise of the data-driven organization as more stakeholders are looking to be involved in reporting and analysis.

The survey followed up with participants and their plans for improvements in the future [Figure 7]:

▶ **"Which of the following are you planning on implementing in the next 1–2 years in your organization?"**
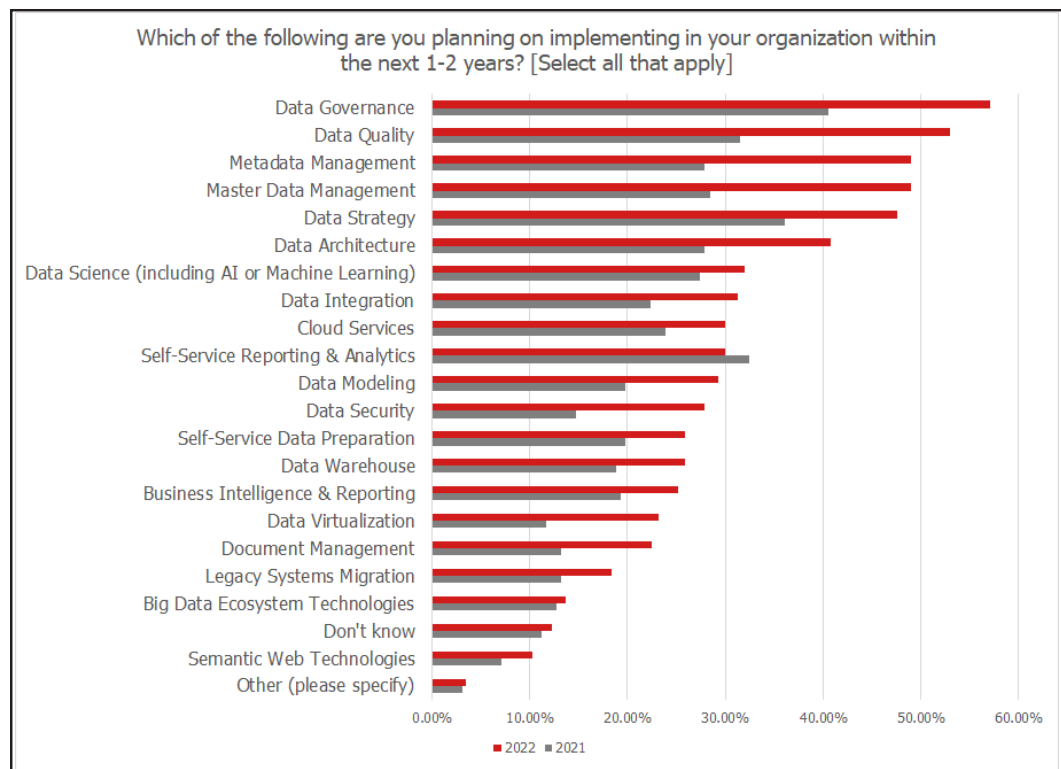
Figure 7: Future Implementation of Data Management (2021 & 2022)

The Top 5 Data Management improvements businesses are planning to implement between 2022 and 2024 are:

• Data Governance: 57.14% (40.61% in 2021, 49.40% in 2020)

• Data Quality: 53.06% (31.47% in 2021, 44.58% in 2020)

• Master Data Management: 48.98% (28.43% in 2021, 46.99% in 2020)

• Metadata Management: 48.98% (27.92% in 2021, 40.96% in 2020)

• Data Strategy: 47.62% (36.04% in 2021, 49.04% in 2020)

Plans for improvements that scored a lower response are:

- Legacy Systems Migration: 18.37%

- Big Data Ecosystem Technologies: 13.61%

- Semantic Web Technologies: 10.20%

Key initiatives that continue to drive future plans year over year include: Data Governance, Data Quality, Master Data Management, and Data Strategy as organizations look to better manage their data assets that drive analytics, operational efficiency, and more.

Notable is the fact that Self-Service Reporting & Analytics, which was a top "future initiative" in 2021, appeared as a top current initiative in 2022 – establishing the accuracy of that prediction. Its absence on this year's survey does not necessarily suggest a lack of importance. On the contrary, it suggests organizations have implemented self-service and see it as a "current priority," not a future plan.

The inclusion of Metadata Management in the top five responses of the 2022 survey makes sense, aligning with the rise of Self-Service Analytics & Reporting. As more business users actively become involved in using data for business decisions, they are likely to have more questions about the definition, provenance, and business rules regarding key data. Many organizations are looking to implement Data Catalogs to support their Metadata Management needs.

D. Concluding Comments

As business-driven Reporting & Analytics continue to be drivers for Data Management, the need for the foundational supporting initiatives of Data Governance, Metadata Data Management, Master Data Management, and Data Quality continue to be a high priority for both present and future plans. As more business users become hands-on with data via self-service, Metadata Management is of particular importance, with organizations looking to have a shared understanding of critical data assets.

4. GOALS, DRIVERS, AND TRAINING

This section focuses on needs and goals of organizations, and how these goals are used to improve their business and increase profits. Participants were presented with five questions about their organization's goals, drivers, and training.

A. Goals and Drivers

Respondents answered the following [Figure 8]:

▶ **"What are your main business goals and drivers for implementing Data Management in your organization?"**
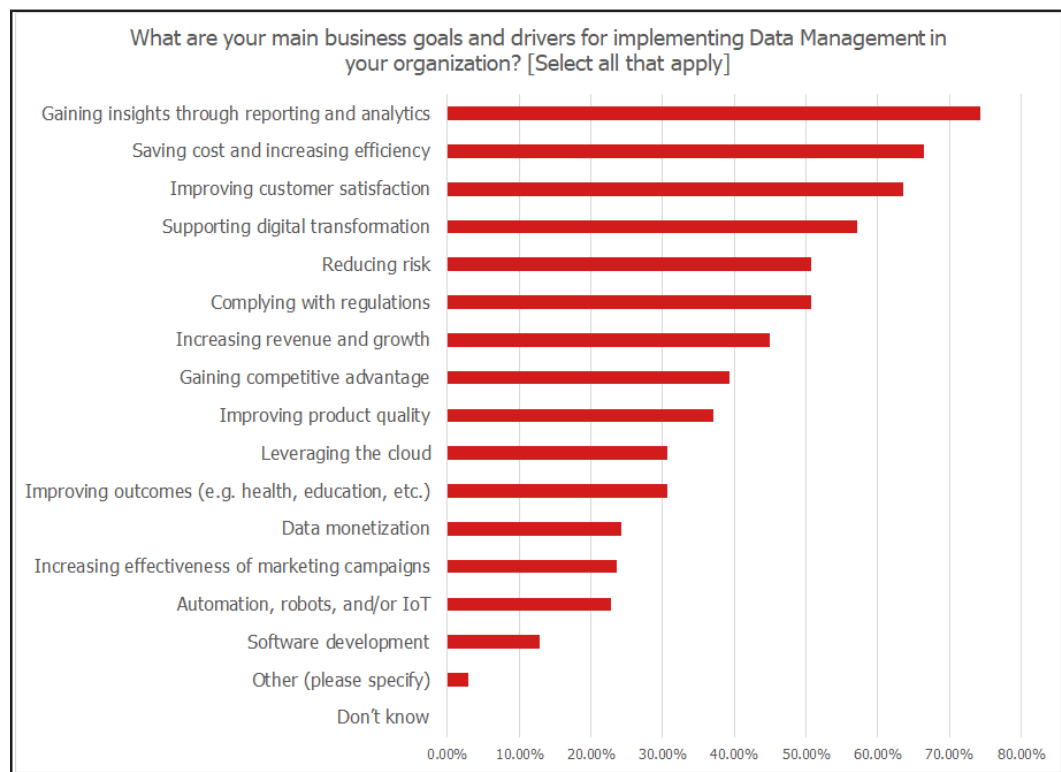

Figure 8a: Business Goals and Drivers (2022)

Figure 8b: Business Goals and Drivers (2021)

The survey shows the Top 5 goals and drivers are:

- Gaining insights through Reporting & Analytics: 72.29% (It was 69.40% in 2021, with a 2.89% increase)

- Supporting Digital Transformation: 57.14% (It was 64.48% in 2021, with a 7.34% decrease)

- Saving cost and increasing efficiency: 66.43% (It was 57.38% in 2021, with a 9.05% increase)

- Improving customer satisfaction: 63.57 (It was 51.37% in 2021, with a 12.20% increase.)

- Two tied for fifth place: Complying with regulations and reducing risk: 50.71% (Complying with regulations was 52.46% in 2021, with a 1.75% decrease. Reducing risk was 57.38 in 2021, with a 6.67% decrease)

In looking at current priorities, common themes continue to emerge year after year, as gaining insights through "Reporting & Analytics", "saving cost and increasing efficiency", and "complying with regulations" continue to land in the Top 5. Notable is that while "supporting Digital Transformation" still continues to be a top driver, its importance has fallen 7%, while "Improving customer satisfaction" has increased by 12%, placing it into the Top 5 this year. As digital becomes the norm, particularly after the drive to digital as a result of COVID (a key finding in last year's survey), more organizations are realizing that it's not enough to be digital – consumers are looking for a positive and improved customer experience as a result of the digital engagement model.

1. Priorities and Goals

This question was open-ended, and participants were asked about their Data Management priorities and goals in the future:

▶ **"What are your top 2–3 Data Management priorities/goals for 2022–2023?"**

Top priorities for 2023 align with the key findings in other survey responses: Data Governance and an overall need for a Data Strategy are top responses. The related initiatives of Data Quality, Master Data Management, and Metadata Management, particularly using a Data Catalog were also noted themes.

Data Literacy was another common theme. As more organizations push to become data-driven, the need to increase the awareness and knowledge of core Data Management fundamentals is of greater importance. Comments included:

- "The majority of our senior managers seek solutions in technology rather than understanding and improving data. Data Literacy is critical and strong leadership from the top is needed."

- "Implementing a Data Strategy and supporting frameworks (i.e., Data Management, Data Governance, Data Quality, Data Architecture)."

- "Establish Data Lifecycle policies, set up Data Catalog for new development, enhanced security/compliance."

- "Implementation of Data Governance program and improving enterprise-wide Data Literacy."

- "Define an Enterprise Data Strategy, start a Data Governance program and begin looking at Master Data Management processes."
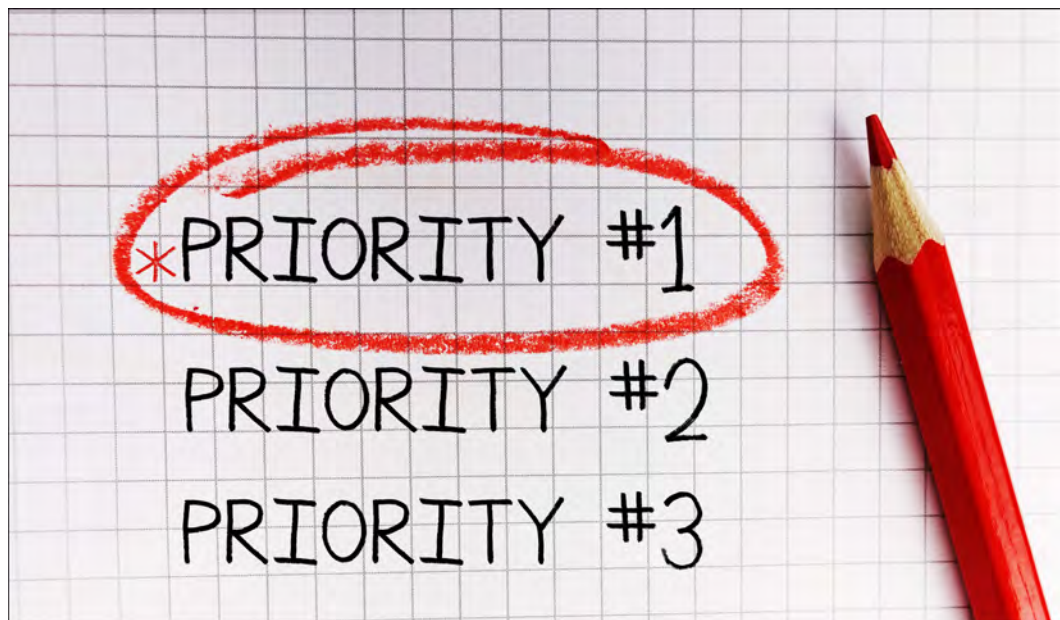
2. Challenges

To gain a better understanding of the challenges organizations are facing, we asked [Figure 9]:

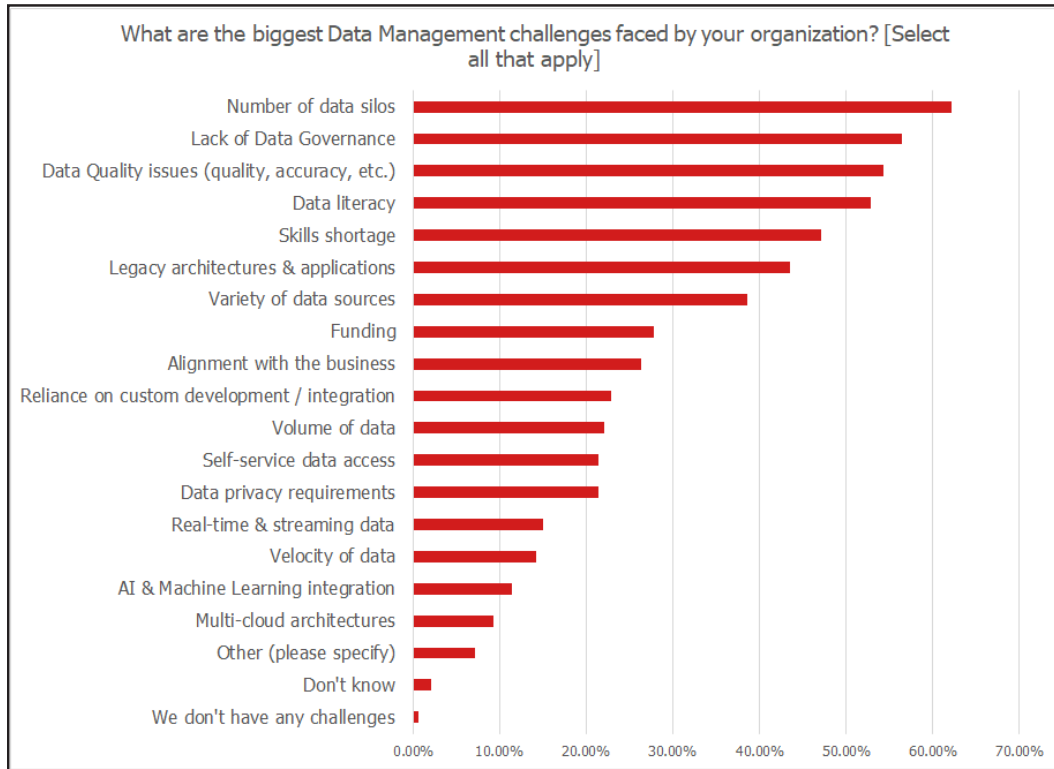► **"What are the biggest Data Management challenges faced by your organization?"**

Figure 9: Data Management Challenges

The Top 5 choices included:

• Number of data silos: 62.14 (65.75% in 2021)

• Lack of Data Governance 54.63% (new answer in 2022)

• Data Quality issues 54.29% (new answer in 2022)

• Data Literacy 52.86% (new answer in 2022)

• Skills shortage: 47.14 (46.96% in 2021)

The "number of data silos" continues to be a top Data Management challenge year over year, as does the "skills shortage" for key personnel to address the issue. New to the Top 5 are Data Governance and Data Literacy, both indicators of the awareness that reducing data silos requires involvement from the business, and the proper knowledge from business stakeholders to support effective resolutions. Data Quality was another newcomer to the Top 5, as organizations realize that it is not enough to remove silos, but that improving the quality of data is key to effective usage of that data for both operational and analytic use.

3. Tools and Technology

To find out how technology has influenced Data Management decisions, we asked [Figure 10]:

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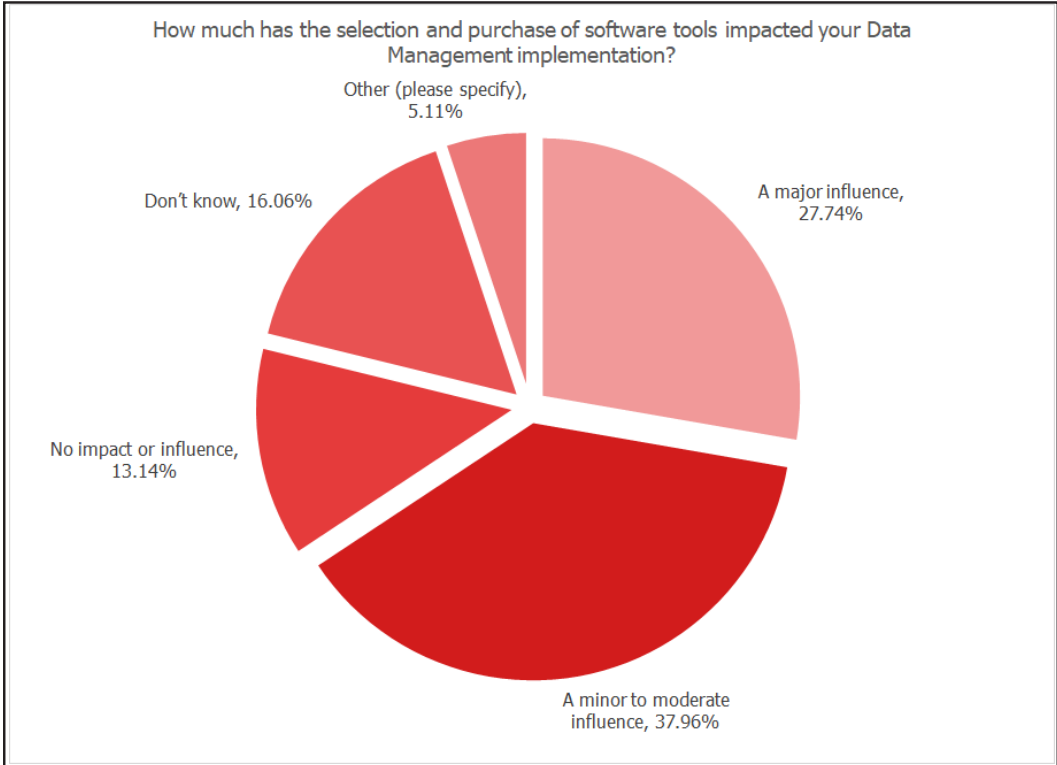

Figure 10: Software Tool Impact on Data Management Implementation

The responses strongly suggest the choices of software tools, and their purchase has an impact:

• A minor to moderate influence: 37.96% (It was 34.08%, with a 3.88% increase)

• A major influence: 27.74% (It was 29.61% in 2021, with a 1.87% decrease)

• No impact or influence: 13.14% (It was 13.97% in 2021, with a 0.83% decrease)

• Don't know: 16.06% (It was 15.08% in 2021, with a 0.98% decrease)

• Other (please specify): 5.11% (It was 7.26% in 2021, with a 2.15% increase)

37.96% of the participants indicated software choices had a minor to moderate influence, while 27.74% expressed software had a major influence. When combined, a total 65.70% claimed the selection and purchase of software had an impact on data implementation.

Software tools play an important role in how Data Management functions, but as shown below, some organizations are working without software that would improve their Data Management.

Many of the comments in the "other (please specify)" section supported a theme about a lack of purchased software:

- "We have not selected or purchased anything to date."
- "We have not selected or purchased software tools for our Data Management practices/ implementation yet."
- "We wrestle with what we can afford, as we seek to mature our program."
- "We do not have any Data Management tools yet."

B. Training

There are a variety of training tools, ranging from books to online courses. Direct, hands-on experience also qualifies as a training tool. We surveyed participants on their learning experiences about Data Management. [Figure 11]

▶ **"What type(s) of training have you received in Data Management?"**
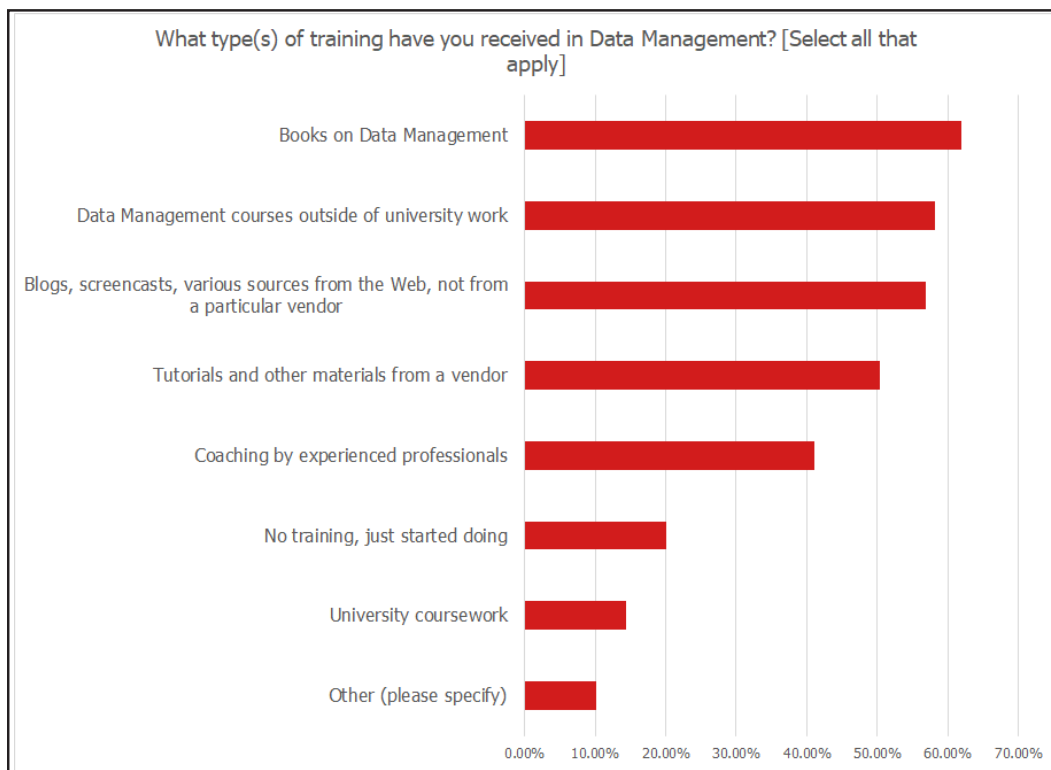

Figure 11: Training in Data Management

Finding the time for training is not an easy task. Many of the participants surveyed received their training from:

- Books on Data Management: 61.87% (66.12% in 2021, 58.76% in 2020)

- Blogs, screencasts, various sources from the web, not from a particular vendor: 58.63% (60.66% in 2021, 66.67% in 2020)

- Data Management courses outside of university work: 58.27% (48.63% in 2021, 49.15% in 2020)

Although the use of books for training declined slightly, many of the respondents chose it as their preferred method for adding to their education.

Only 14.39% of the responses claimed to have taken university classes for training in Data Management.

Training by way of "blogs, screencasts, various sources from the web, not from a particular vendor" has shown a steady decrease in popularity over the last few years.

Sources of training from the "other (please specify)" section included:

- DATAVERSITY Webinars

- Member of a local DAMA chapter

- Conference presentations

- Hands-on work

C. Concluding Comments

Initiatives that drive business success, such as Data Analytics, efficiency through automation, and supporting customer satisfaction through Digital Transformation are of high priority for organizations taking the survey. In order to support these business-centric initiatives, there is a strong need for initiatives that increase the involvement and education of business stakeholders. As a result, Data Governance and Data Literacy are top areas of importance to organizations in 2022 and beyond.

5. DATA GOVERNANCE AND METADATA MANAGEMENT

This section asks about the organizations' state of Data Governance and their use of Metadata Management.

A. Data Governance

To better understand the maturity level of Data Governance within the organizations surveyed, they were asked [Figure 12]:

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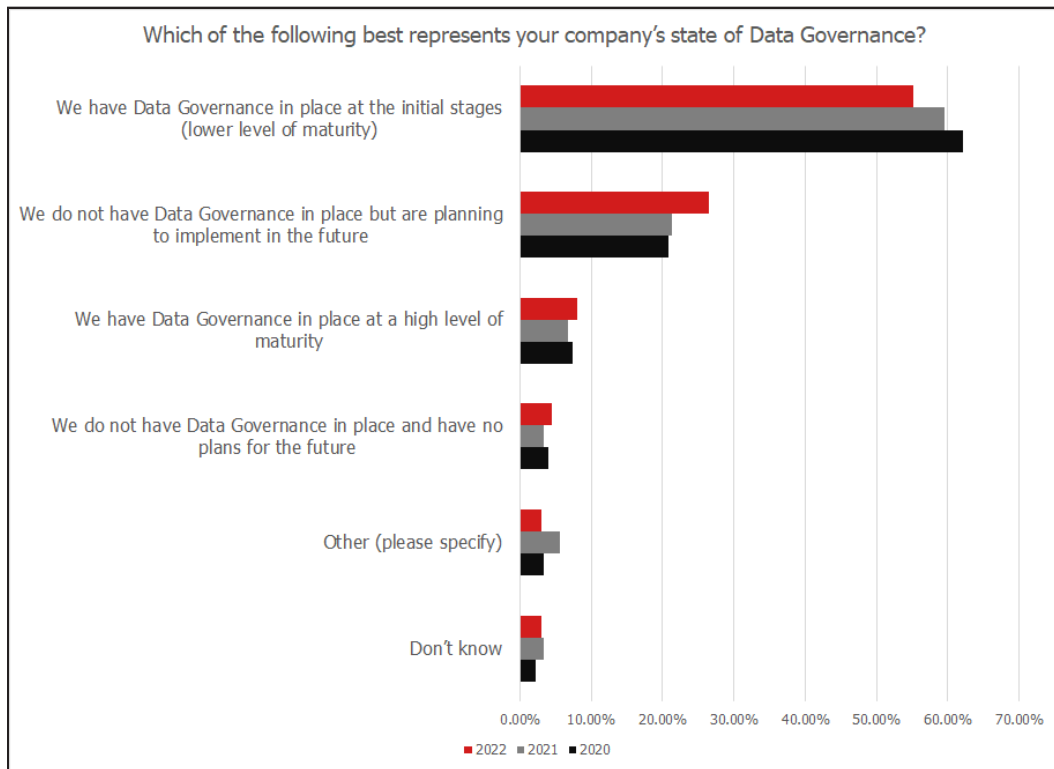

Figure 12: State of Data Governance (2020 – 2022)

Organizations continue to support Data Governance programs, but at low maturity levels.

The survey's Top 5 responses were:

- Data Governance is in place at the initial stages (lower level of maturity): 55.15% (59.55% in 2021, 62.15% in 2020, and 53.38% in 2019)

- No Data Governance in place but a plan to implement it in the future: 26.47% (21.35% in 2021, 20.90% in 2020, and 22.66% in 2019)

- Data Governance at a high level of maturity: 8.09% (6.74% in 2021, 7.34% in 2020, and 9.02% in 2019)

- No Data Governance program in place nor any plans for the future: 4.41% (3.37% in 2021, 3.95% in 2020, and 3.76% in 2019)

- Two responses tied for fifth place: Don't know and Other (please specify). Don't know -2.94% (3.37% in 2021, 2.26%, and 3.76% in 2019). Other (please specify) -4.41% (5.62% in 2021, 3.39% in 2020, and 7.52% in 2019).

Data Governance continues to be a priority for organizations, but many feel that there is still more work to be done. While the majority of organizations have a Data Governance program in place (55.15 %) or have plans to initiate one (26.74%), for a total of 81.89%, 55.15% feel that they are still at a low level of maturity. Since successful Data Governance requires a mix of people, process, technology, as well as organizational and cultural change, it often takes organizations several years before they feel that all areas of governance are working at an advanced level.

Only 4.41% of the respondents believed there was no need for a Data Governance program.

Comments from the Other (please specify) category were generally more specific descriptions of their Data Governance program's maturity level. Examples are:

- "We have initial stages of Data Governance around material master data only."

- "In place at a moderate level of maturity."

- "We have Data Governance in place in varying stages of maturity across the company. Some parts are very mature, others have a much lower level of maturity."

B. Metadata Management

Metadata, or data in context, supports Data Governance, and other Data Management processes. The survey asked respondents about their current Metadata Management use cases [Figure 13]:

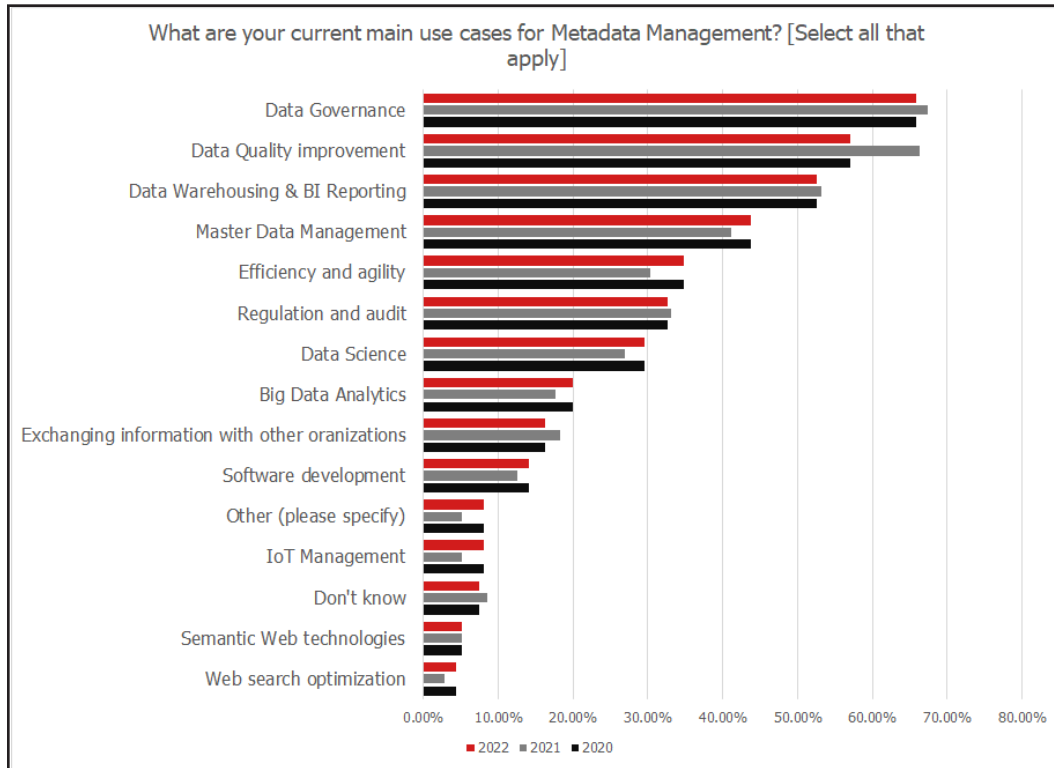► **"What are your current main use cases for Metadata Management?"**

Figure 13: Metadata Management Use Cases (2020 – 2022)

The survey shows Data Governance is used quite often with Metadata Management, followed by Data Quality improvement, Data Warehousing and BI reporting, and Master Data Management.

- Data Governance: 65.93% (67.43% in 2021, 77.91% in 2020)

- Data Quality improvement: 57.04% (66.29% in 2021, 60.47% in 2020)

- Data Warehousing & BI Reporting: 52.59% (53.14% in 2021, 58.14% in 2020)

- Master Data Management: 43.70% (41.14% in 2021, 53.49% in 2020)

- Efficiency and Agility: 34.81% (30.29% in 2021, 30.23% in 2020)

Top drivers for Metadata Management continue to be Data Governance, Data Quality improvement, Data Warehousing and BI Reporting, and Master Data Management. New to the Top 5 list in 2022 is "Efficiency and Agility", replacing last year's fifth place "Regulation and Audit". This is a positive trend, as more organizations are realizing that properly managed metadata can help improve agility and efficiency. Far from slowing teams down by having better documentation, well-managed metadata makes data consumers' jobs easier in the long run.

C. Additional Comments

Several of the respondents described how they are dealing with Data Governance and Metadata Management, and its maturity level. Some of the more interesting additional comments describing the uses of Data Governance and Metadata Management are:

- "We have a vocabulary team that has collected agency glossaries, developed a metadata schema, and is trying to establish a governance practice for terminology."

- "This is the primary focus for Data Management development as a keystone for development of Data Governance and other aspects of Data Management development."

- "Data Governance became the central discipline for business regulations compliance."

- "The current Data Governance and Metadata Management is primarily driven by requirements on research data search/discovery and access."

6. DATA ARCHITECTURE

Data Architecture, as described by the DAMA DMBoK:

> "Defines the blueprint for managing data assets by aligning with organizational strategy to establish strategic data requirements and designs to meet these requirements."

The purpose of Data Architecture is to apply the business' needs to its data and system requirements. It is used to manage the data, and its flow, throughout the enterprise.

We wanted to know how popular the use of Data Lakes had become and the challenges of using Data Lakes and Data Warehouses. We wanted to know if organizations had a Data Architecture supporting their Data Management initiative. If they had an architecture, we wanted to know its value and problems. We also provided an area for additional comments.

A. State of Data Architecture Within Data Management

We asked the following about Data Architecture in Data Management [Figure 14]:

► **"How has a defined Data Architecture helped your organization?"**
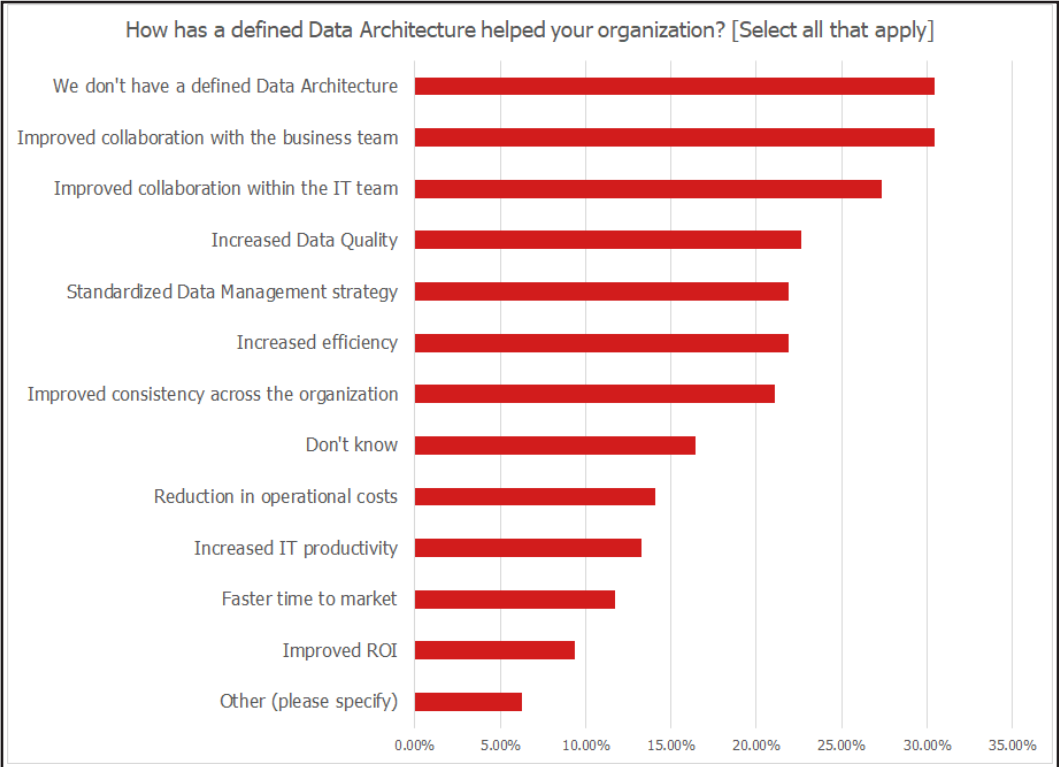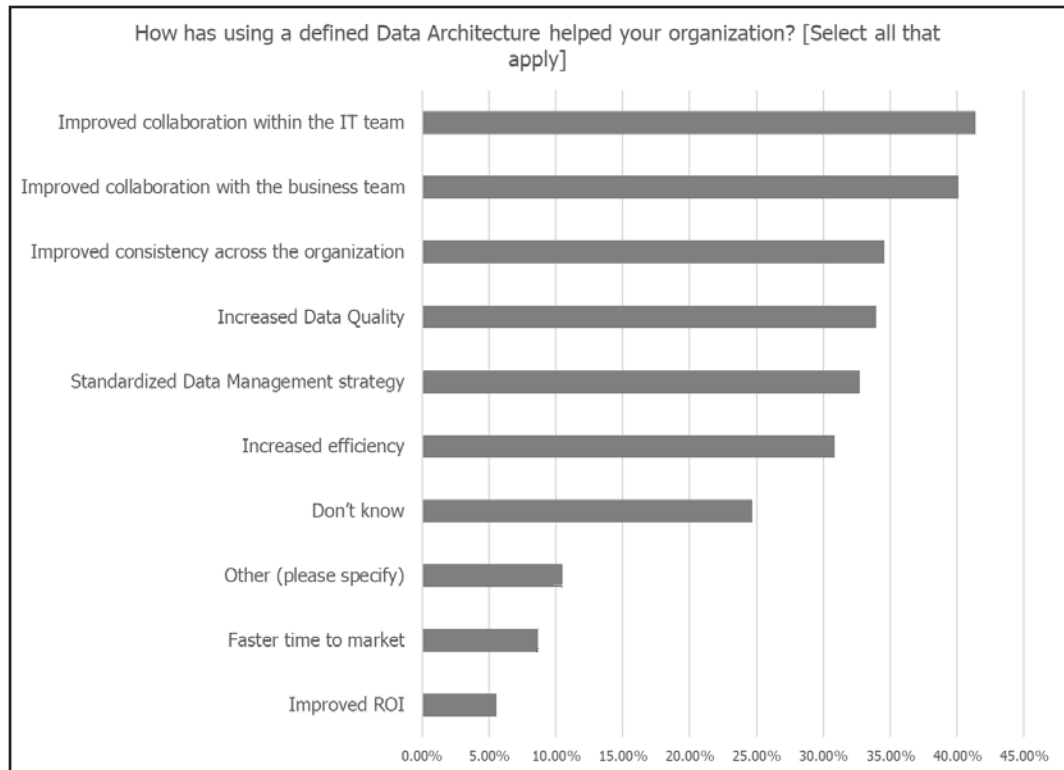

Figure 14a: Data Architecture Advantages (2022)

Figure 14b: Data Architecture Advantages (2021)

The Top 5 responses for describing how a defined Data Architecture helped an organization are:

- Improved collaboration with the business team: 30.47% (40.12% in 2021, 58.33% in 2020, 57.14% in 2019)

- We don't have a defined Data Architecture: 30.47% in 2022 (new answer in 2022)

- Improved collaboration within the IT team: 27.34% (40.12% in 2021, 58.33% in 2020, and 57.14% in 2019)

- Increased Data Quality: 22.66% (33.95% in 2021, 44.44% in 2020, and 28.57% in 2019)

- Increased Efficiency: 21.88% (30.86% in 2021, 33.33% in 2020, and 38.10% in 2019)

For organizations implementing a Data Architecture, collaboration is a clear benefit, with 30.47% citing improved business collaboration and 27.34% showing improved collaboration with IT. Increased Efficiency and Data Quality were also benefits aligned with this collaboration.

Unfortunately, many organizations (30.47%) still do not have a defined Data Architecture, partly due to the difficulty in finding skilled Data Architect resources. Several of the "other (please specify)" comments show their organizations were in the beginnings of developing a Data Architecture solution.

- "We are in the beginning stages of having a defined Data Architecture."

- "There is a developing body of experience and recognition of Data Architecture. It has been done previously on a piecemeal basis by engineers focused on engineering needs. It is still early in development."

- "We are working on this, but have a hard time with collaboration with our IT department…. sigh."

- "We are hiring a Data Architect and working on the Data Architecture piece now."
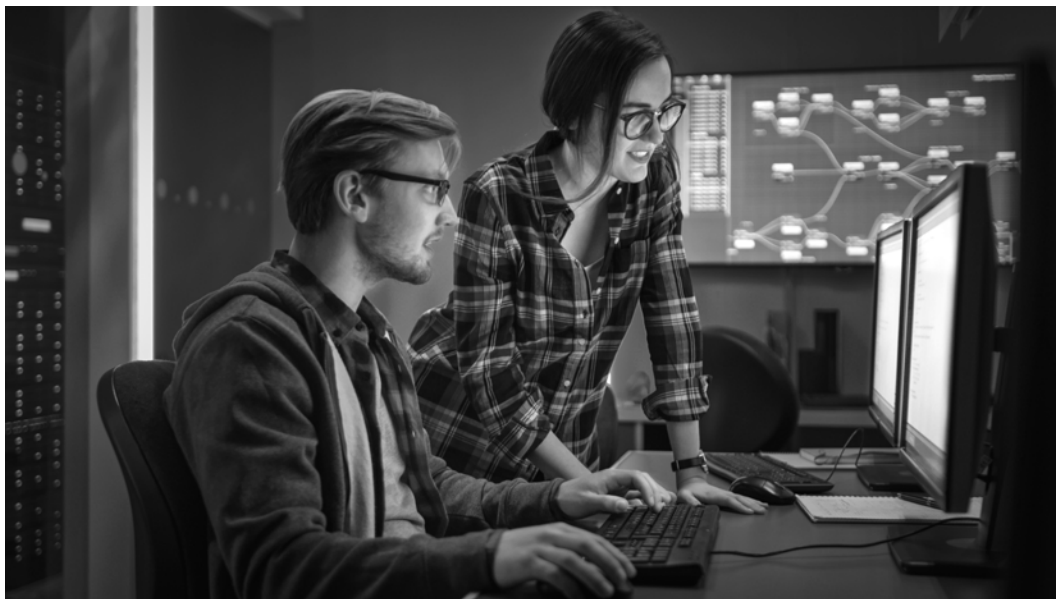
B. Use of Data Lakes

A Data Lake is a storage platform used for structured, unstructured, and semi-structured data at any scale. Typically, this type of storage is used specifically for Data Analytics. Data is gathered and stored in its original format in Data Lakes, with no format transformations, indexing, or prep work required.

In an effort to determine the general usage of Data Lakes, we asked [Figure 15]:

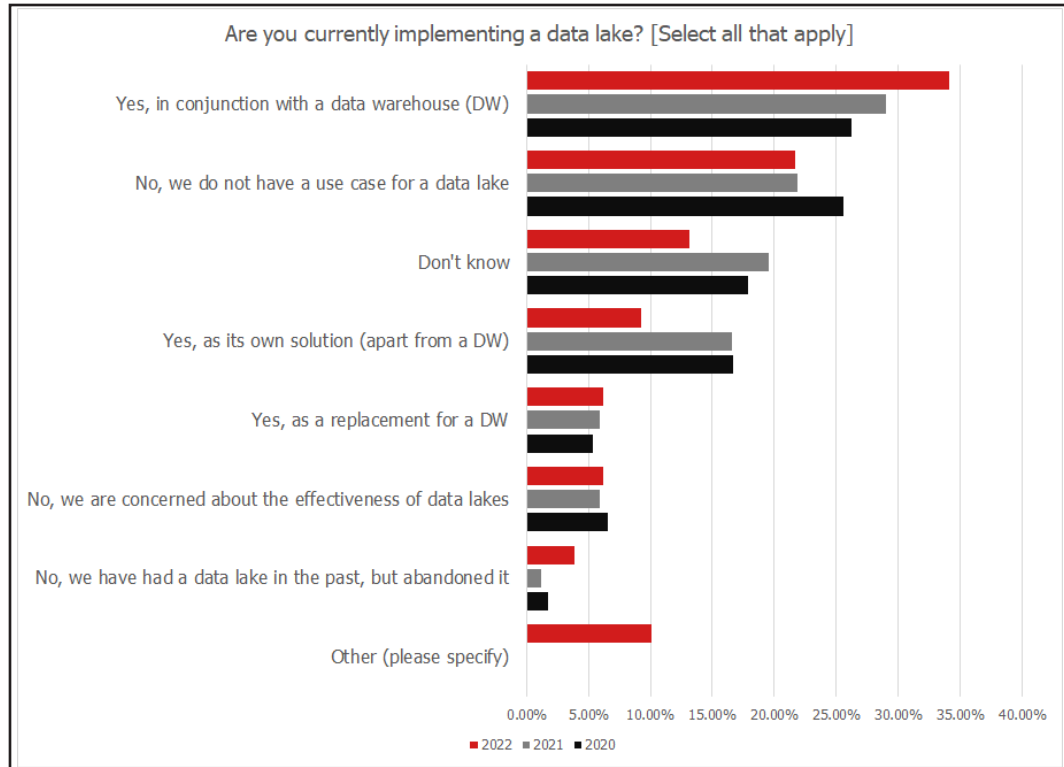► **"Are you currently implementing a Data Lake?"**

Figure 15: Data Lake Implementation (2020 – 2022)

As the usage of Data Lakes matures, an increasing number of organizations see them as a way to augment a Data Warehouse, rather than a replacement for Data Warehouse and other structured storage mechanisms. The highest response at 34.11% showed a 5% increase for organizations using a Data Lake in conjunction with a Data Warehouse. At the same time, the number of organizations using a pure-play Data Lake has decreased by an even greater amount (7.27%).

A large number of organizations are not using a Data Lake (31.79%) or are not aware of a Data Lake in use (13.18%). This, combined with the strong support for structured Data Management such as Data Warehousing and Master Data Management in other survey responses, show that Data Lakes are not a priority for many organizations.

C. The Challenges of Data Warehouses and Data Lakes

A Data Warehouse is storage for relational data taken from operational databases, business applications, and transactional systems with the goal of using it for reporting and Data Analytics. To discover the major Data Warehouse/Data Lake challenges businesses are facing, we asked [Figure 16]:

▶ **"What are the major challenges that you are facing with respect to Data Warehousing/ Data Lake?"**
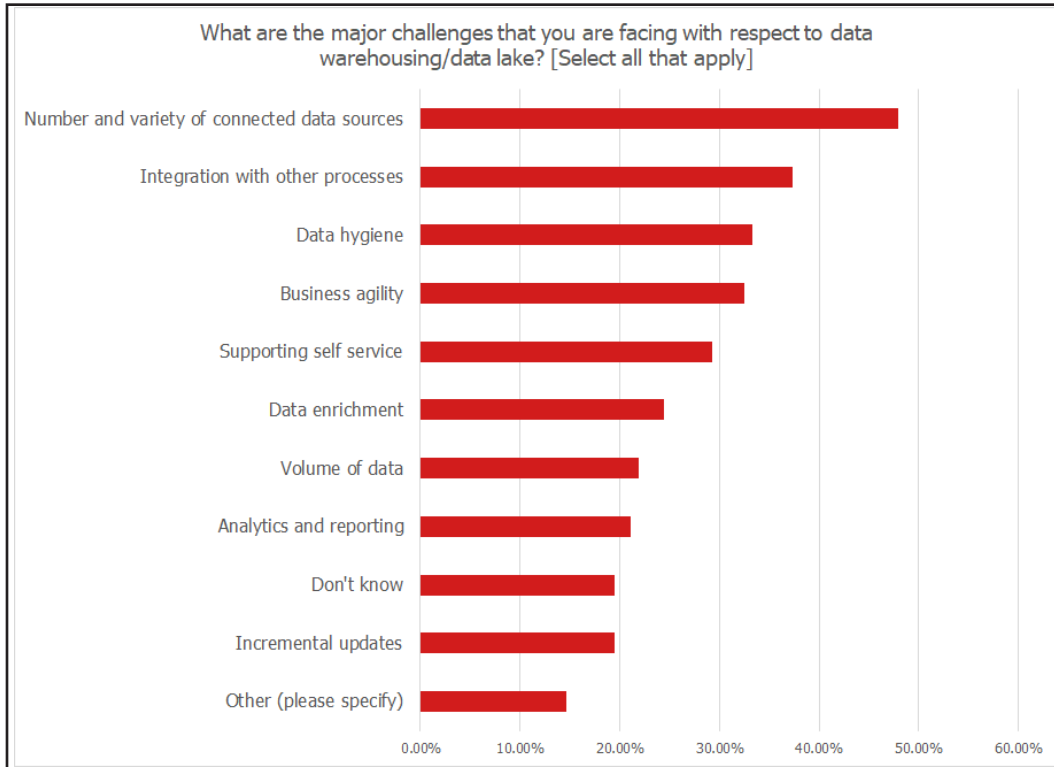

Figure 16: Data Warehouse and Data Lake Challenges

The survey's results showed the Top 5 challenges:

• Number and variety of connected data sources: 47.97% (46.54% in 2021)

• Integration with other processes: 37.40% (33.96% in 2021)

• Data hygiene: 33.33% (new answer in 2022)

• Business agility: 32.52% (new answer in 2022)

• Supporting self-service: 29.27% (27.04 in 2021)

These challenges for the use of Data Warehouses and Data Lakes express the complexity of producing quality, fit-for-purpose data at the speed of the business. The need to integrate a variety of data sources (47.97%) and "integrate with other processes" (37.40%) continues to be a key challenge.

Combining the need for quality data (Data hygiene: 33.33%) with "business agility" (32.52%) emphasizes the need to increase "supporting self-service" (29.27%). This is important as business users often want to create their own reporting and data sets to meet increasing demands.

"Other (please specify)" comments on challenges pointed to the lack of skilled resources and need for Data Literacy:

- Lack of skilled resources to lead these efforts

- Knowledgeable staff

- Lack of dedicated resources, lack of organizational vision

- Lack of education and understanding of modern thinking

7. DATA MODELING

Data Modeling, as described by the DAMA DMBoK:

> "Is the process of discovering, analyzing, and scoping data requirements, and then representing and communicating these data requirements in a precise form called the data model."

Data Modeling plays a critical role in Data Management. We posed three questions about these activities. From there, our survey offered insights into Data Management, Data Modeling, and business alignment.

A. State of Data Modeling Within Data Management

We posed the question [Figure 17]:

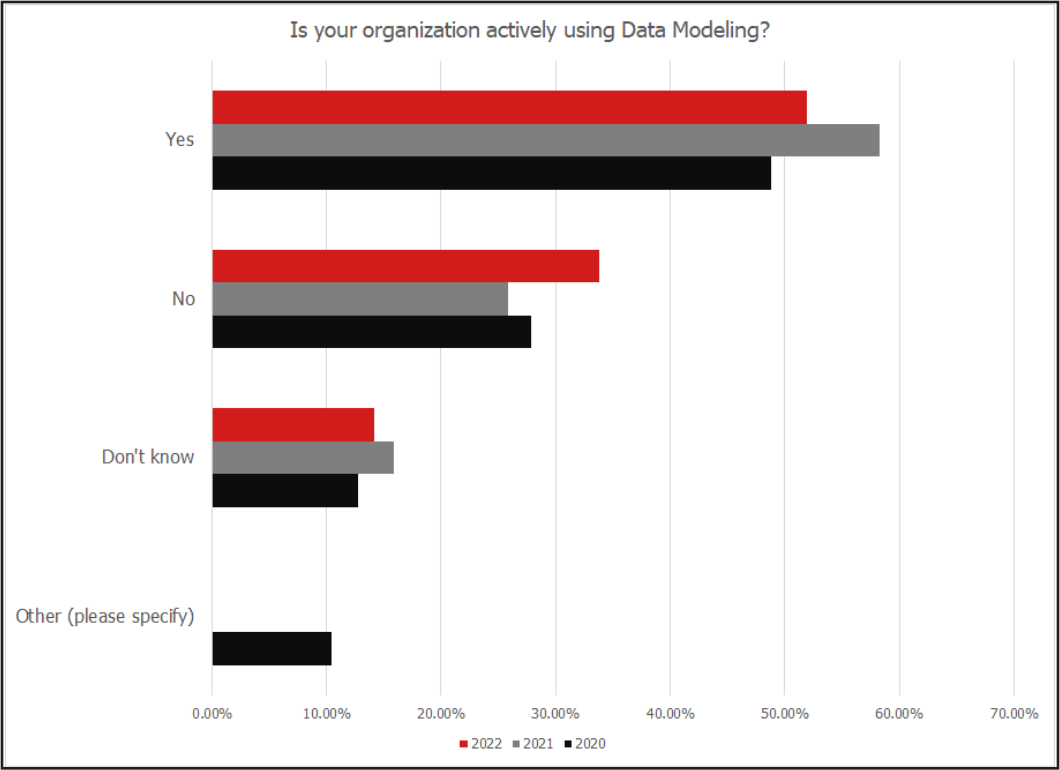► **"Is your organization actively using Data Modeling?"**

Figure 17: State of Data Modeling (2020 – 2022)

The percentages in 2022 were:

- Yes: 51.97% (58.24% in 2021, 48.84% in 2020)

- No: 33.86% (25.88% in 2021, 27.91% in 2020)

- Don't know: 14.17% (15.88% in 2021, 12.79% in 2020)

The use of Data Modeling rose by 9.40% in 2021 (58.24%) from 2020 (48.84%), but decreased in 2022 (51.97%) by -6.27%. Overall, Data Modeling usage continues to be strong year over year.

B. Data Modeling Methods

There are a variety of data models available for use, depending on the needs of the organization. This survey wanted information about the models being used. Survey participants were asked the question [Figure 18]:

► **"What methods of Data Modeling do you use in your organization?"**
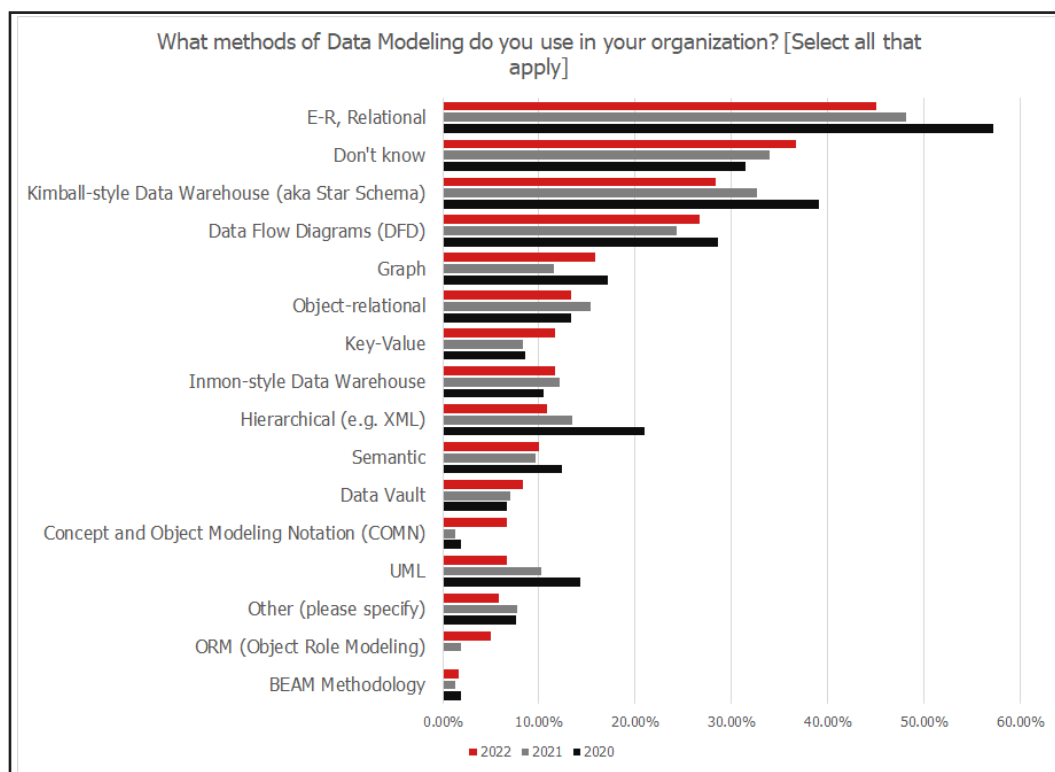

Figure 18: Data Modeling Methods (2020 – 2022)

The most preferred Data Modeling methods selected were:

- E-R, Relational: 45.00% (48.08% in 2021, 57.14% in 2020)

- Don't know: 36.67% (33.97% in 2021, 31.43% in 2020)

- Kimball-style Data Warehouse (aka Star Schema): 28.33% (32.69% in 2021, 39.05% in 2020)

- Data Flow Diagrams (DFD): 26.67% (24.36% in 2021, 28.57% in 2020)

- Graph: 15.83% (11.54% in 2021, 17.14% in 2020)

ER (Entity-Relationship) modeling continues to be the preferred Data Modeling method, at 45.00%, which aligns with the strong focus on Data Quality and Data Governance.

It is worth noting that 36.67% did not know what model was being used, coming in second. Kimball-style Data Warehouses (aka Star Schema) came in third, which makes sense given the strong need for reporting and analytics.

Data flow diagrams came in fourth (26.67%) as a Data Modeling method. Data flow diagrams are often used to help integrate data silos (listed as the largest Data Management challenge by participants).

C. Types of Models and Diagrams

The following question was asked to gain a better understanding of how models and diagrams are used [Figure 19]:

► **"What types of models and diagrams do you use in your data/enterprise architecture?"**
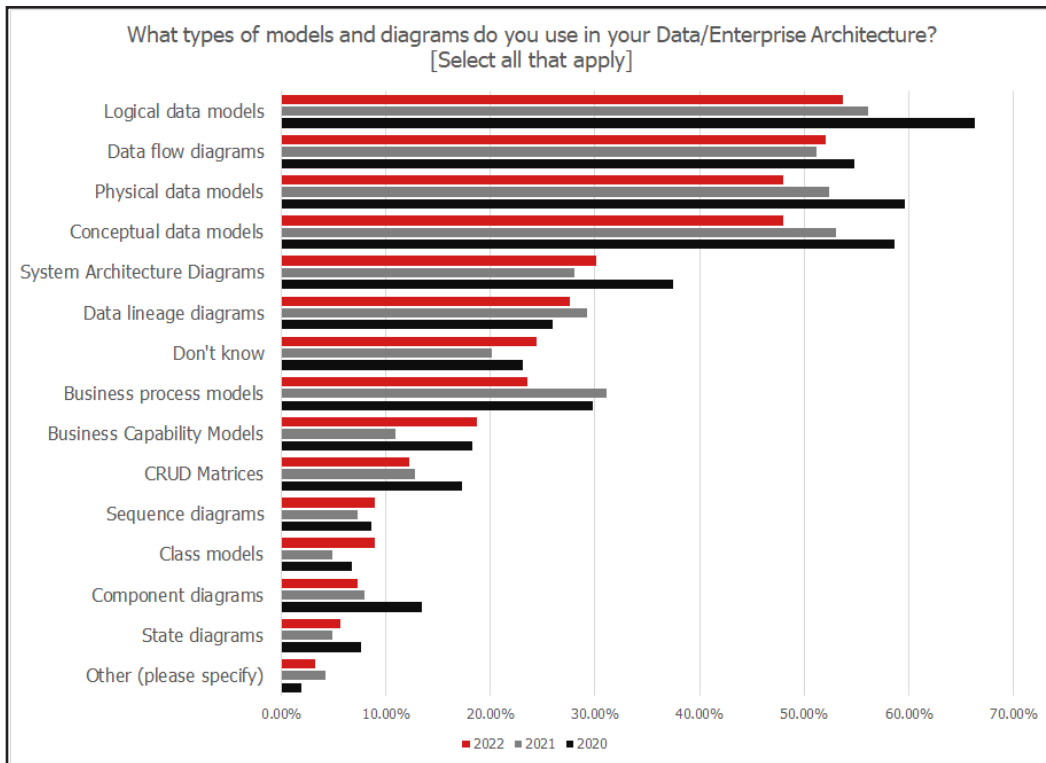

Figure 19: Data Models and Diagrams (2020 – 2022)

The Top 5 types of models and diagrams being used roughly continued to support the results of previous years:

- Logical data models: 53.66% (56.10% in 2021, 66.35% in 2020)

- Data flow diagrams: 52.03% (51.22% in 2021, 55.00% approximately in 2020)

- Physical data models: 47.97% (52.44% in 2021, 59.62% in 2020)

- Conceptual data models: 47.97% (53.05% in 2021, 58.65% in 2020)

- System architecture diagrams: 30.08% (28.05% in 2021, 37.50% in 2020)

The logical model continues to be the most popular data model, which makes sense given that it can act as a bridge between business and IT with both detailed business rules and definitions as well as a foundation for physical implementations. The physical and conceptual models have dropped and data flow diagrams have shown gains.

D. Concluding Comments

Data Modeling is an important part of Data Management and continues to show strong usage in this survey. Data models offer flexibility in supporting business needs via logical and conceptual models, and focus on accurately representing the data at both the business and technical levels.
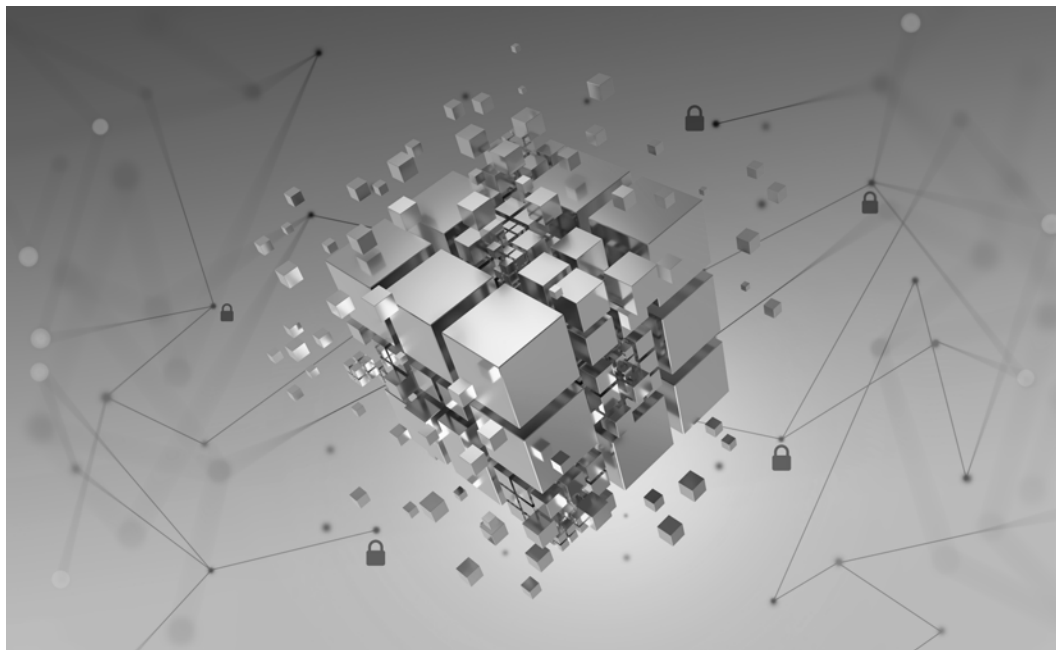
8. DATA PLATFORMS AND STORAGE

We wanted to know about the storage and data platforms used by the survey respondents, and their plans. We asked four questions with the opportunity to make additional comments.

The first set of questions are aimed at which data platforms and storage are being used, and future plans for them. The second set of questions focused on reasons for moving to the cloud and concerns prior to moving to the cloud.

A. Data Platforms

With the goal of understanding which platforms and storage applications are currently being used the most, respondents were asked [Figure 20]:

► **"Which of the following data sources or platforms are you currently using?"**
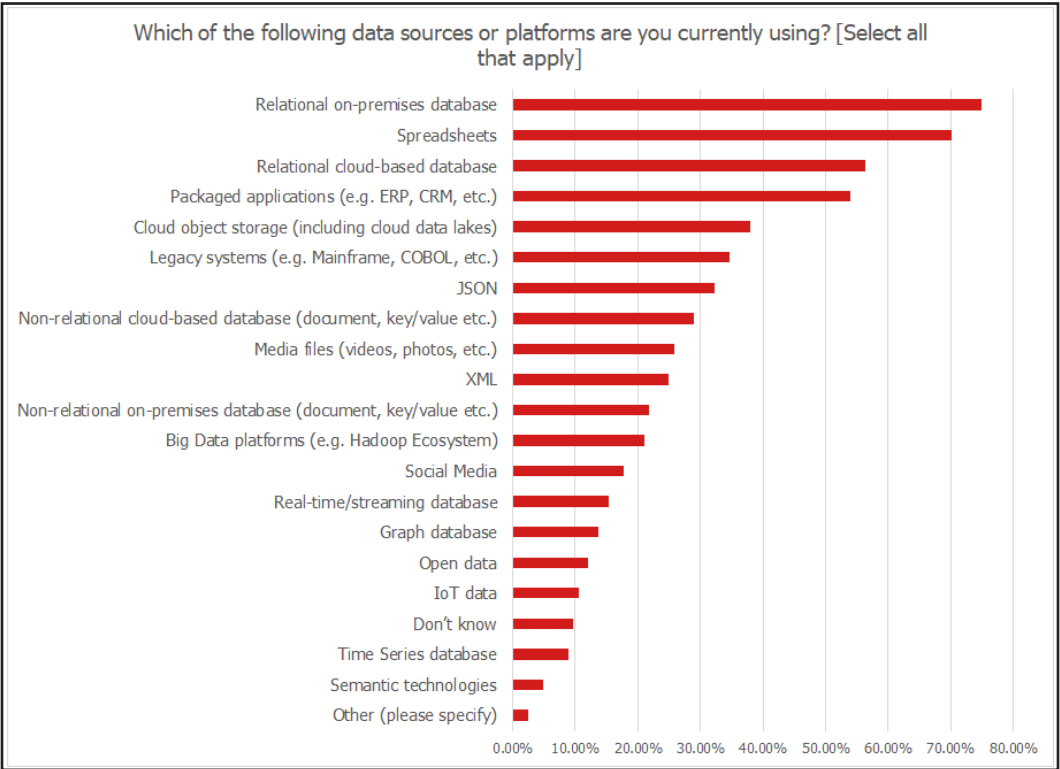

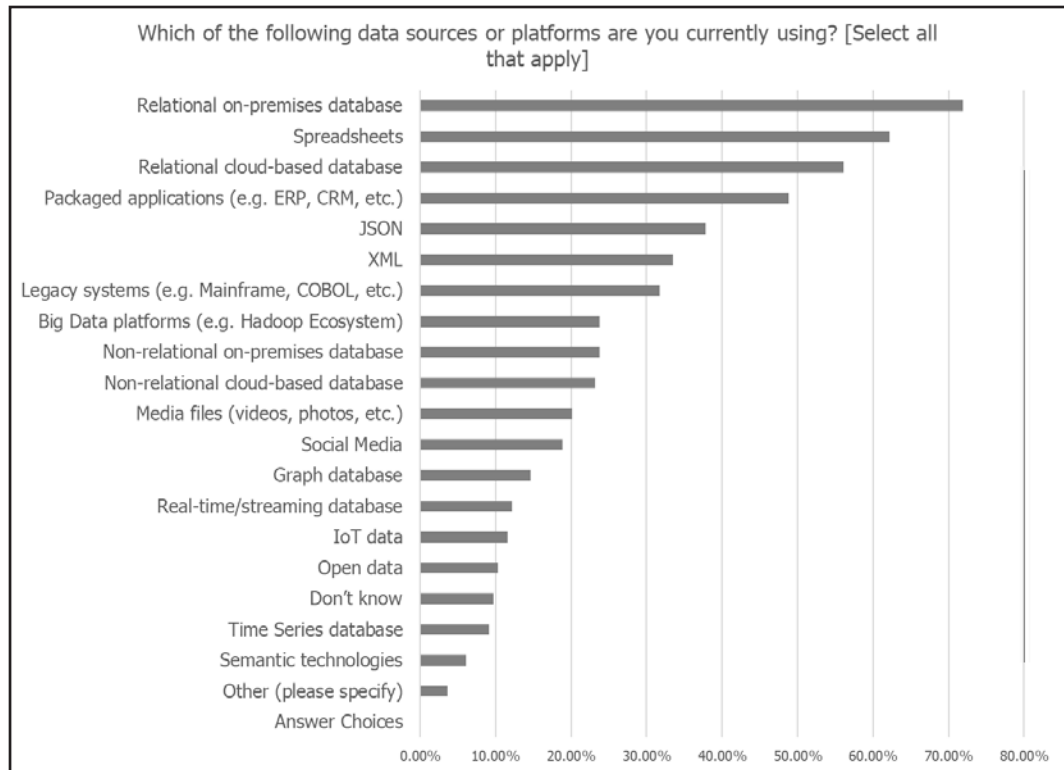Figure 20a: Data Sources or Platforms (2022)

Figure 20b: Data Sources or Platforms (2021)

In 2022, the Top 5 sources were:

- Relational on-premises database: 75.00 % (71.95% in 2021)

- Spreadsheets: 70.16% (62.20% in 2021)

- Relational cloud-based database: 56.45% (56.10% in 2021)

- Packaged applications (e.g., ERP, CRM, etc.): 54.03% (48.78% in 2021)

- Cloud object storage (including cloud Data Lakes): 37.90 (new answer in 2022)

The least-used technologies in 2022 were:

- Semantic technologies: 4.84% (6.10% in 2021)

- Time Series database: 8.87% (9.15% in 2021)

The commonly used relational databases are still the most popular form of database, which includes both on-premises (75.00%) and cloud-based (56.45%). The familiarity of the relational database model, with its referential integrity and structured storage mechanism, makes it a very popular technology, particularly with the importance of Business Intelligence, Data Warehousing, and Data Quality initiatives.

Results also show that organizations often struggle with general Data Literacy and Data Management fundamentals. Many continue to use spreadsheets to manage enterprise data due to their ubiquity and ease of use.

Cloud object storage is new to the Top 5 (at 37.90%) and shows a trend of organizations making use of the lower-cost and ease of use of this type of storage.

B. Future Plans for Data Platforms and Storage

Participants were asked about their future plans regarding platforms and storage. [Figure 21]

► **"Which of the following data platforms/data storage technologies do you plan to use in the next 1–2 years?"**
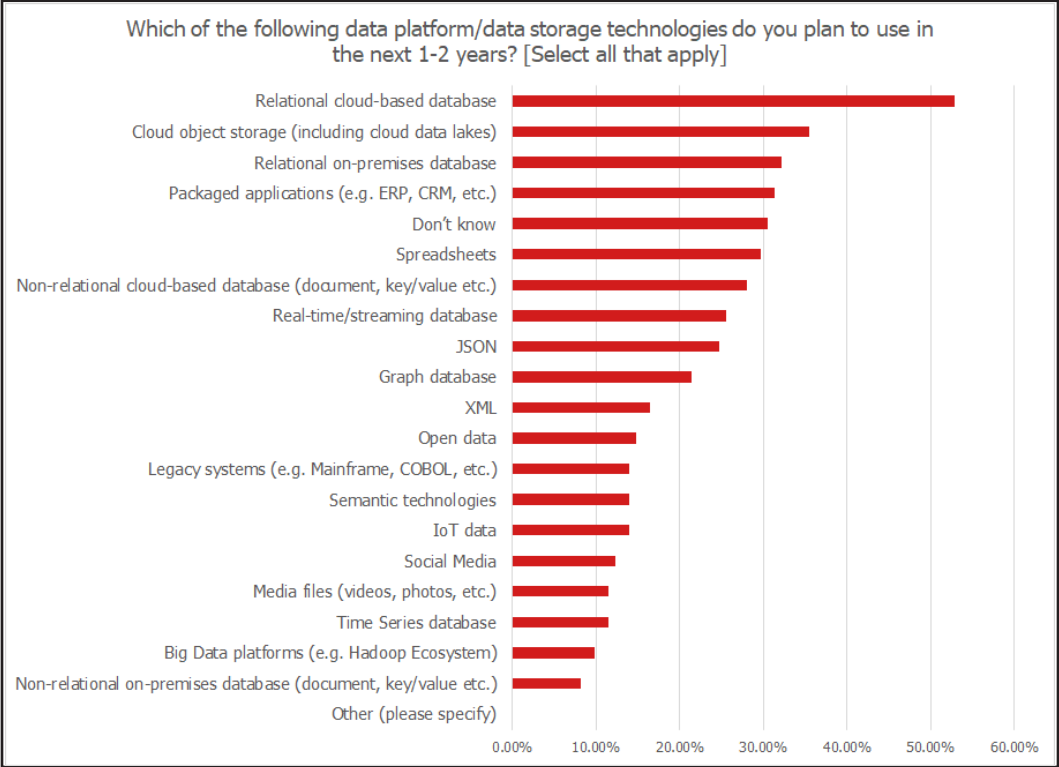

Figure 21a: Data Sources or Platforms Plans in Next 1–2 Years (2022)

Which of the following data platform/data storage technologies do you plan to use in the next 1-2 years? [Select all that apply]
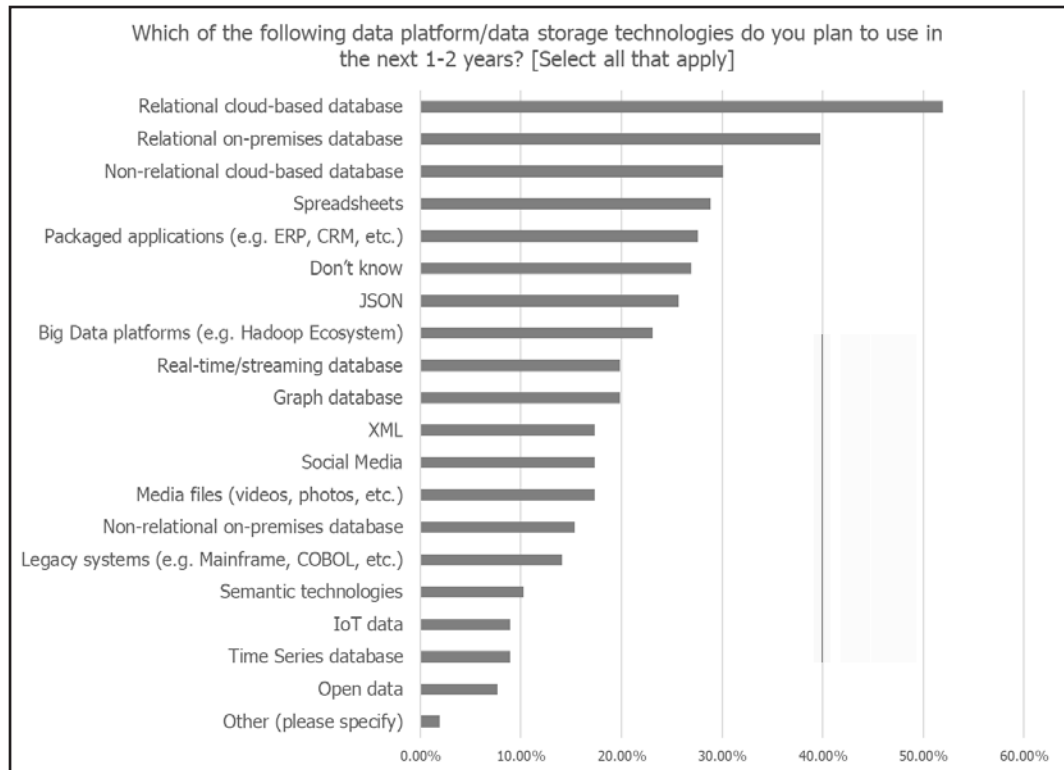
Figure 21b: Data Sources or Platforms Plans in Next 1–2 Years (2021)

The leading Top 5 responses were:

- Relational cloud-based database: 52.89% (51.92% in 2021, 54.17% in 2020, and 45.97% in 2019)

- Cloud object storage (including cloud Data Lakes): 35.54% (new answer in 2022)

- Relational on-premises database: 32.23% (39.74% in 2021, 47.02% in 2020, and 44.35% in 2019)

- Packaged applications (e.g., ERP, CRM, etc.): 31.40% (27.56% in 2021, 33.93% in 2020, and 36.29% in 2019)

- Don't know: 30.58% (26.92% in 2021, 13.59% in 2020, and 27.42% in 2019)

The least popular platforms were:

- Non-relational on-premises database (document, key/value etc.): 8.26%

- Big Data platforms (e.g., Hadoop Ecosystem): 9.92%

- Time Series database: 11.57%

- Media files (videos, photos, etc.): 11.57%

Implementing relational databases in the future is a strong preference for many organizations. The majority of businesses will be shifting to relational databases in the cloud (52.89%; 51.92% in 2021) as compared to on-premises relational databases (39.74%; 32.23% in 2021).

With the myriad choices of technologies in the current market, it is notable that 30.58% are still unsure of what they will choose moving forward.

C. The Cloud

Over the last decade, the cloud has increasingly offered businesses a growing number of benefits. It has become an important part of their business process and planning.

The respondents were asked five questions about how they use cloud technologies, and their reasons and concerns.

The first question asked was [Figure 22]:

► **"What are/were your reasons for moving to the cloud?"**
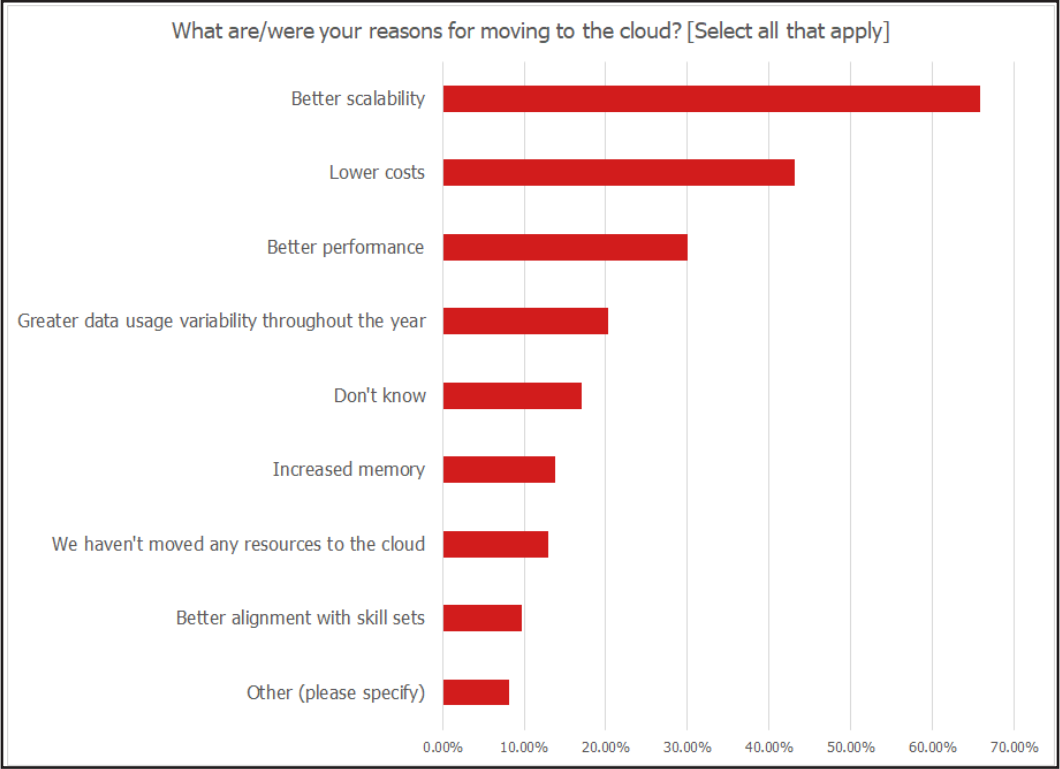

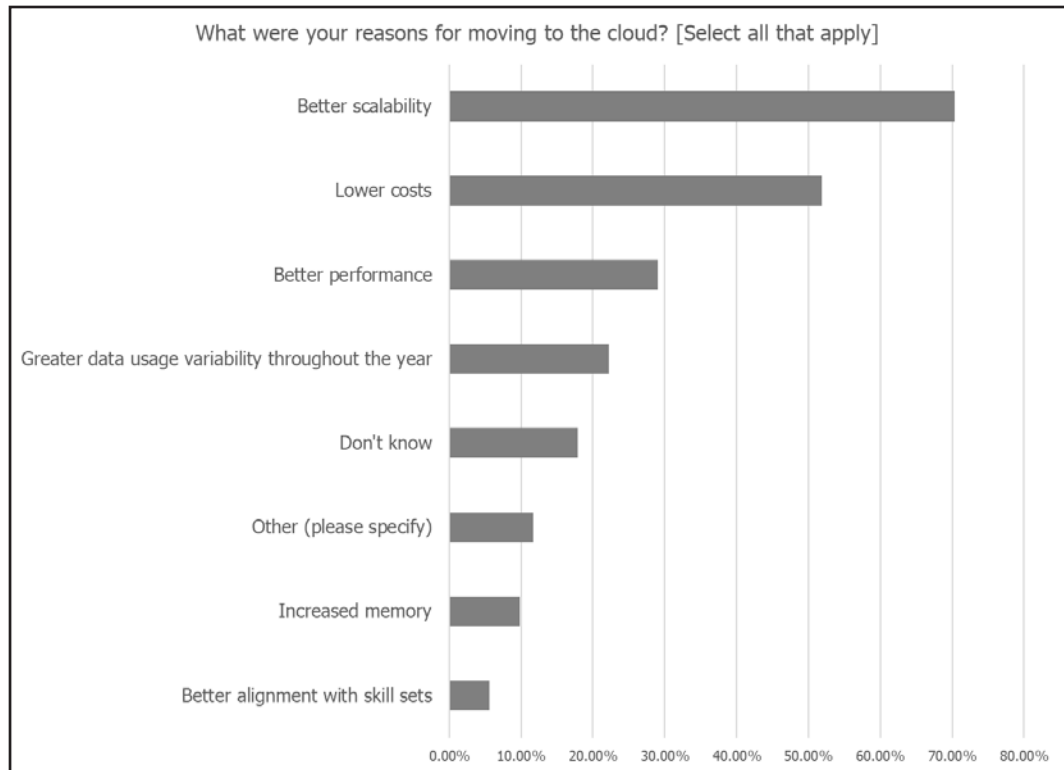Figure 22a: Reasons to Move to the Cloud (2022)

Figure 22b: Reasons to Move to the Cloud (2021)

The Top 5 reasons given for moving to the cloud are:

- Better scalability: 65.85% (70.37% in 2021, 52.94% in 2020, and 72.73% in 2019)

- Lower costs: 43.09% (51.85% in 2021, 47.06% in 2020, and 63.64% in 2019)

- Better performance: 30.08% (29.01% in 2021, 29.41% in 2020, and 27.27% in 2019)

- Greater data usage variability throughout the year: 20.33% (22.22% in 2021, 17.65% in 2020, and 18.18% in 2019)

- Don't know: 17.07% (17.90% in 2021, 17.65% in 2020, and 9.09% in 2019)

These top reasons have been consistent year over year in this survey. Then we asked about their concerns of moving data to the cloud [Figure 23]:

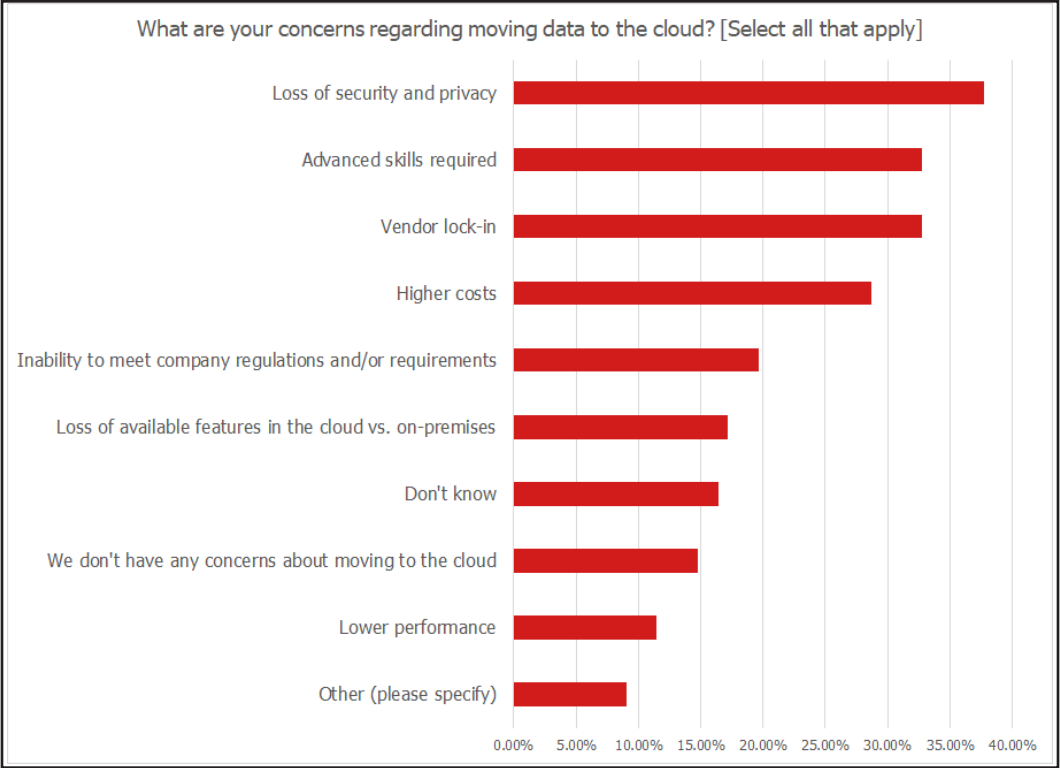► **"What are your concerns regarding moving data to the cloud?"**

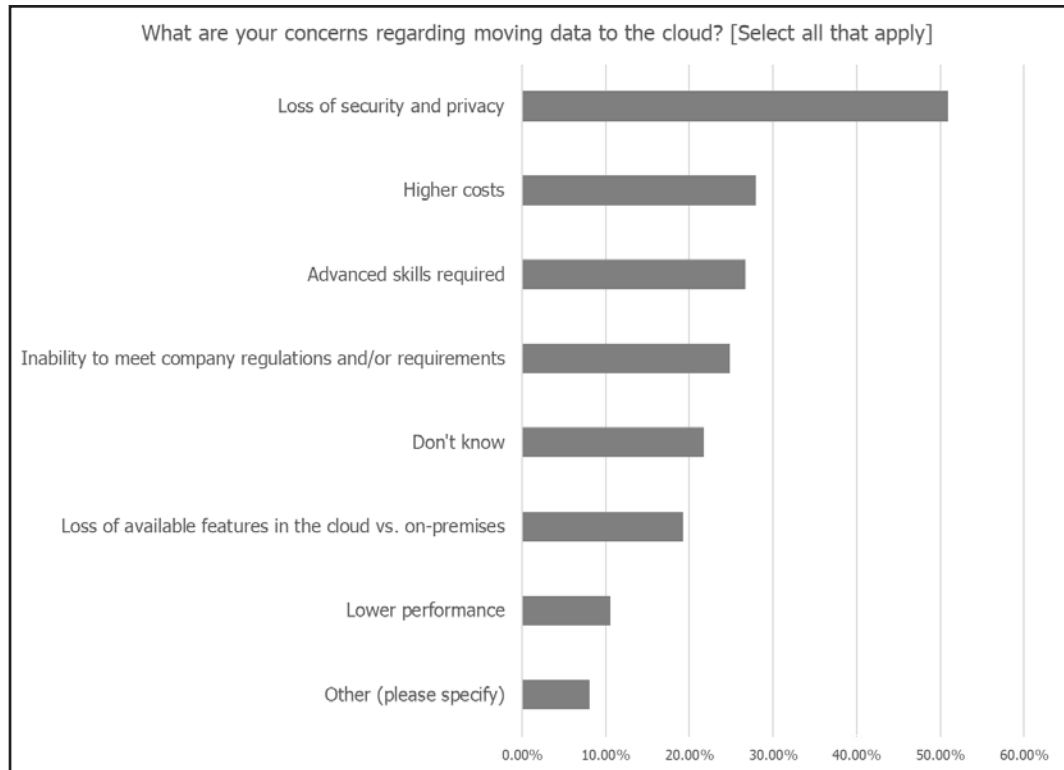Figure 23a: Concerns About Moving to the Cloud (2022)

Figure 23b: Concerns About Moving to the Cloud (2021)

The Top 5 leading concerns expressed in the survey are:

- Loss of security and privacy: 37.70% (50.93% in 2021, 16.67% in 2020, and 54.55% in 2019)

- Vendor lock-in: 32.79% (new answer in 2022)

- Advanced skills required: 32.79% (26.71% in 2021, 16.67% in 2020, 27.27% in 2019)

- Higher costs: 28.69% (27.95% in 2021, 11.11% in 2020, 9.09% in 2019)

- Inability to meet company regulations and/or requirements:19.67% (24.84% in 2021, 11.11% in 2020, 18.18% in 2019)

The top concern was a "loss of security and privacy", although this concern has decreased by over 13%, since 2021. Vendor lock-in is another key concern as is the need for advanced skills.

The 8.13% who answered "other (please specify)" listed the following reasons for moving to the cloud:

- Transferring risk to a vendor strategy

- Increase mobile access

- Beginning stages of cloud acquisition and work

- Higher reliability/availability, focus on business value rather than running infrastructure

- Move data closer to customer

- Simplify self-service capability

- Scaled back IT staffing

- Reliability

9. NEW TECHNOLOGIES AND TRENDS

This part of the survey explores the trends taking place with newer technologies. Innovations in managing data have leapt forward in the last decade.

A. New Technologies

The first question asked was: [Figure 24]

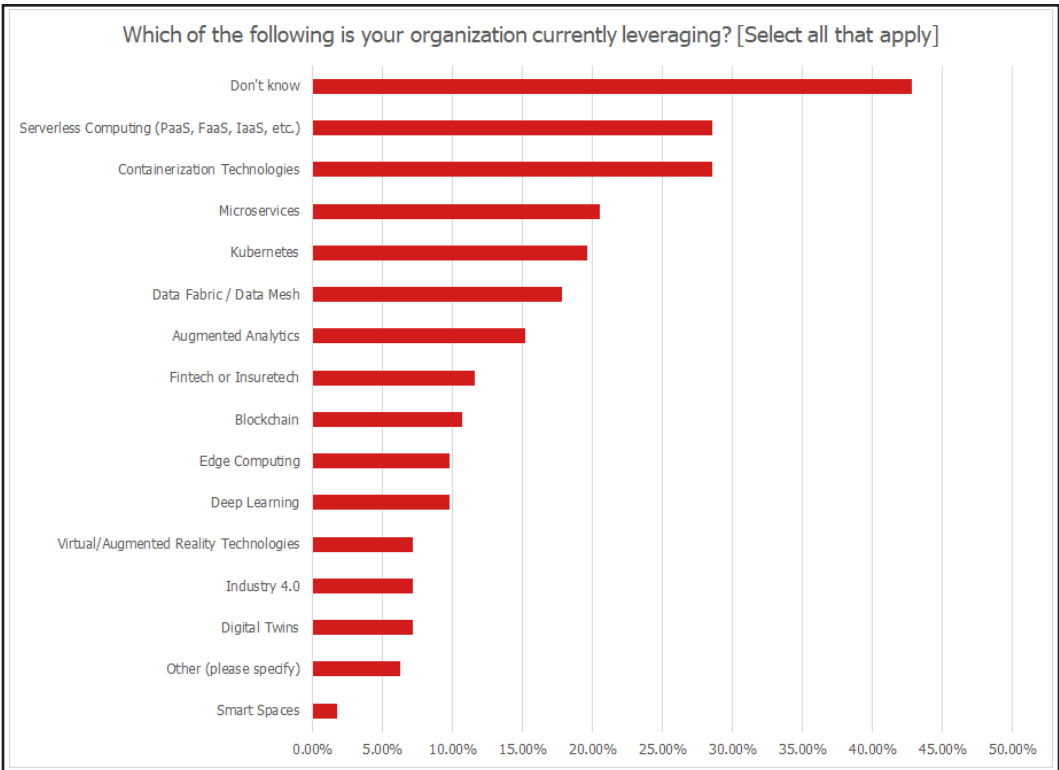► **"Which of the following is your organization currently leveraging?"**

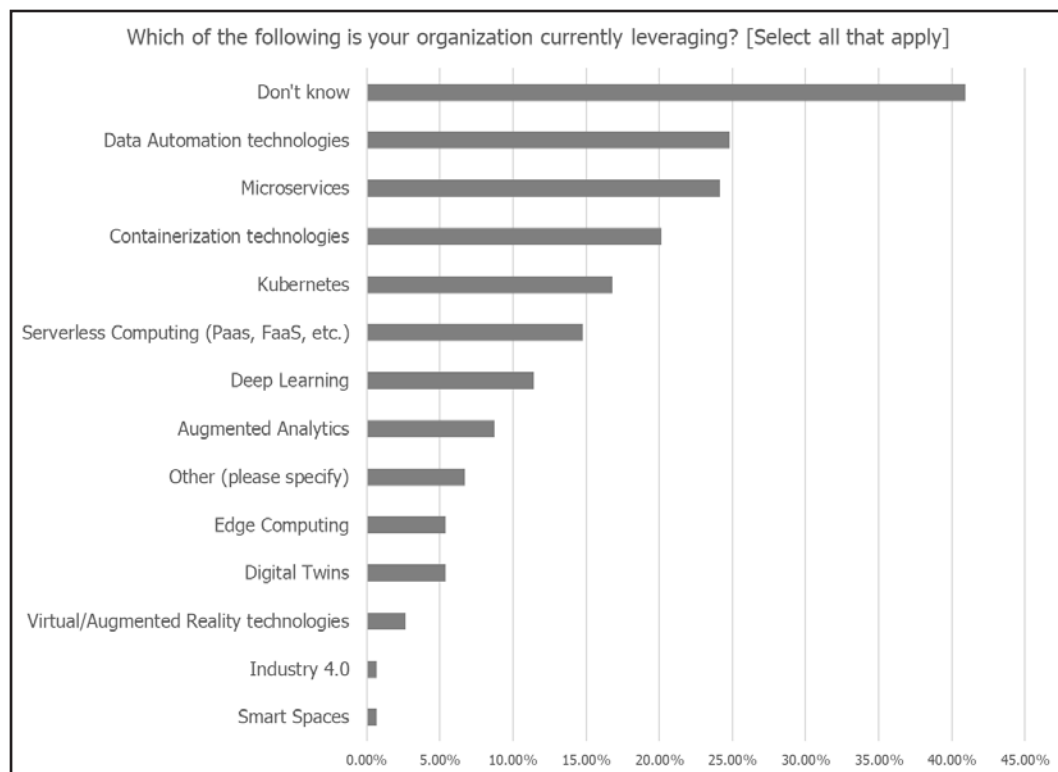Figure 24a: New Technologies Leveraged (2022)

Figure 24b: New Technologies Leveraged (2021)

The response "don't know" received the highest percentage of replies, coming in at 42.86%. Many of these technologies qualify as unknown territory by many of the survey's respondents. While many have probably heard, or read, a little about the new technologies listed, a lack of understanding and familiarity can make new technologies seem confusing. The Top 5 responses to the question, "which of the following is your organization currently leveraging?" are:

- Don't know: 42.86% (40.94% in 2021, 40.98% in 2020, and 54.72% in 2019)

- Containerization Technologies: 28.57% (20.13% in 2121, 26.23% in 2020, and 55.17% in 2019)

- Serverless Computing (PaaS, FaaS, IaaS, etc.): 28.57% (14.77% in 2021, 22.95% in 2020, and 45.45% in 2019)

- Microservices: 20.54% (24.16% in 2021, 27.87% in 2020, and 45.95% in 2019)

- Kubernetes: 19.64% (16.78% in 2021, 18.03% in 2020, 53.57% in 2019)

Container Technologies, Kubernetes, Microservices, and Serverless Computing provide tightly focused software services. A container is a "package of software" containing a software process or microservice that is executable in all computing systems/environments. Kubernetes provides a highly effective system for operating a container's applications at scale. Microservices are based on a philosophy that builds applications as a collection of small independent services that can be combined to create a larger, comprehensive software package, and is generally

delivered by way of containers. Unlike Container Technology, Serverless Computing uses the cloud to provide "backend services" to develop software and apps, which are used by customers working within the cloud.

The historical comparison shows a significant drop in the use of the Top 5 new technologies after 2019. This may be the result of a decreasing need, combined with increasing security concerns. Much of the container technology in 2019, and before, was used to upgrade legacy systems. After 2019, the majority of legacy systems may have been upgraded or replaced, reducing the need for it.

The survey followed up with a question about their plans for the future: [Figure 25]

▶ **"Which of the following is your organization planning to implement within 1–2 years?"**

Which of the following is your organization planning to implement within 1-2 years?
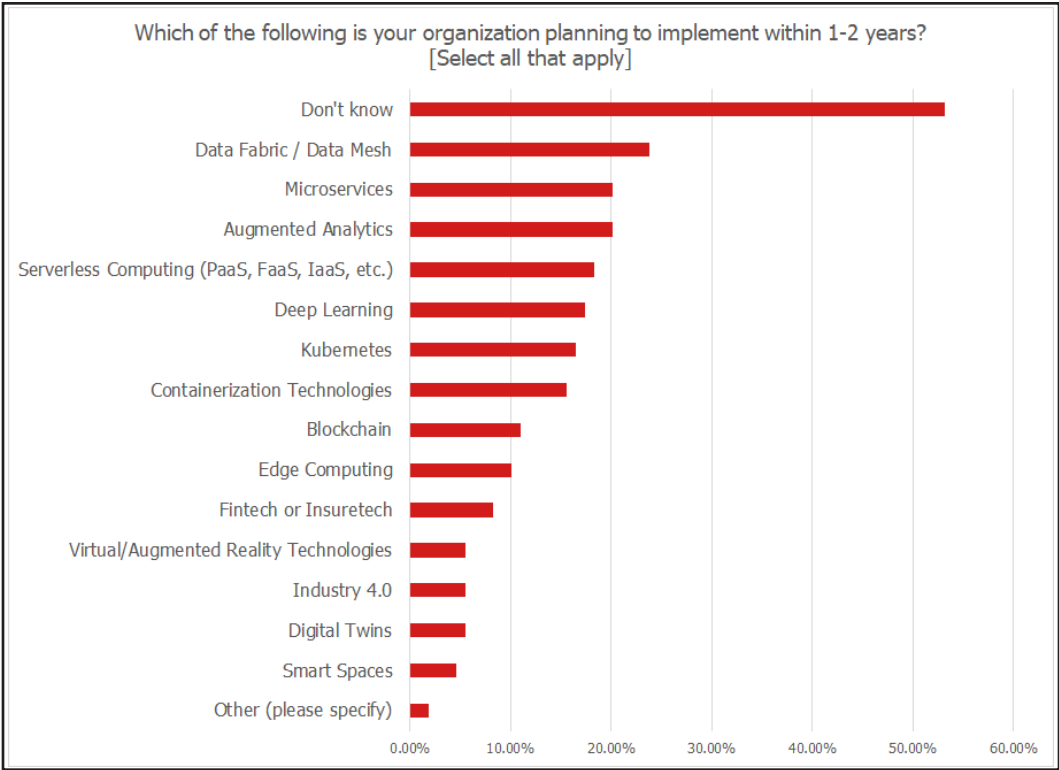[Select all that apply]

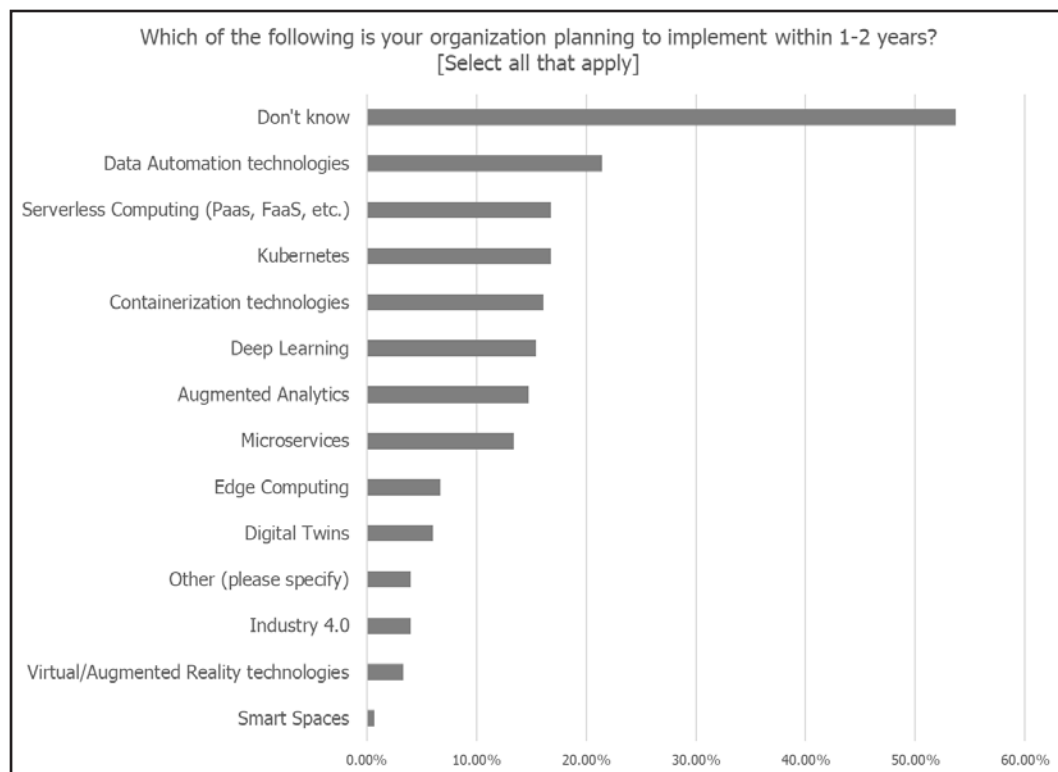Figure 25a: New Technologies Planned to Implement (2022)

Figure 25a: New Technologies Planned to Implement (2021)

Once again, the "don't know" response received the highest score, with 53.21%. This response is completely appropriate when the participants are dealing with unknown variables. The Top 5 New Technologies being planned for implementation are:

- Don't know: 53.21%

- Data Fabric/Data Mesh: 23.85%

- Augmented Analytics: 20.18%

- Microservices: 20.18%

- Serverless Computing (PaaS, FaaS, IaaS, etc.): 18.35%

Data Fabric/Data Mesh came in second, showing there is interest in these two new technologies. They offer organizations a holistic way to manage their data assets across differing systems. While there are similarities between Data Fabric and Data Mesh, there are also significant differences.

Data Fabric looks to build a single, virtual management layer atop distributed data. This allows the data to be accessed from a single location, easily and efficiently.

Data Mesh, on the other hand, can be described as a decentralized form of architecture that standardizes Data Management practices, providing uniform data for a system of Data Centers, Multi-clouds, and Edge Systems, and offering seamless access to these "partners," and their data.

Augmented Analytics and Microservices, tied with both having 20.18%. Serverless Computing (PaaS, FaaS, IaaS, etc.) came in fifth. The least popular new technologies listed were:

- Digital Twins: 5.50%

- Industry 4.0: 5.50%

- Virtual/Augmented Reality Technologies: 5.50%

- Smart Spaces: 4.59%

Despite the uncertainty over new technologies, a few trends have remained strong, such as Automation, Containerization, and Microservices. Likewise, there remains little interest in Smart Spaces, Industry 4.0, or Virtual/Augmented Reality technologies.

B. Future Trends

To finish the survey, participants were asked the following open-ended question:

▶ **What do you see as the next top three emerging trends in Data Management that will cause the most profound changes in the industry?**

Data Mesh and Data Fabric were mentioned the most, supporting the interest displayed in the previous question. With the number of data silos continuing to be a top challenge for organizations (see Section 4), and with the number of disparate platform options increasing (see Section 8), Data Mesh and Data Fabric offer a way to gain visibility across these disparate silos in a flexible and efficient way.

Artificial Intelligence and Machine Learning were also top choices. AI and ML continue to evolve and are giving computers the ability to process massive amounts of data. They allow computers to provide insights and make optimal decisions in a fraction of the time it would take humans. AI and ML are being used for research ranging from medical breakthroughs to climate change.

When asked about emerging trends in Data Management, survey participants wrote:

- "1) Self-service AI/ML Analytics in no/low code environments to be able to bring Advanced Analytics to organizations much easier. 2) Increased access and integration to public and organizational shared datasets to be able to access, swap and integrate data in a public, shared-B2B or team-based environment. 3) Increased security capabilities to reduce security threats such as ransomware, data-theft, etc."

- "1) Intelligent Data Catalogs. 2) Automated data product measurement/valuation. 3) Data Mesh."

- "1) Data Virtualization. 2) Metadata driven information provision. 3) AI for managing data platforms."

- "1) AI. 2) Data Literacy. 3) Metaverse."

10. CONCLUSION

Data Management continues to be a top priority for organizations in 2022. Data Analytics & Reporting drive more effective decision-making, and operational efficiencies are driven by data-driven automation. Digital Transformation is becoming the accepted norm, and the standard is now set higher for organizations to use data more effectively in managing the customer experience, while at the same time reducing costs and supporting regulations and risk reduction.

This increased importance of data has driven the need for Data Management fundamentals to establish trusted data that can drive organizational success. As a result, Data Governance, Data Quality, Data Security, and a solid Data Architecture are in greater demand.

To support this demand, Data Literacy and technical data skills are needed, and many organizations struggle to find qualified staff with the skills needed to support a data-driven digital organization. The need for education and training around data is a growing priority.

As organizations continue to expand their data practices, fundamentals such as Data Governance and Data Architecture will become business-as-usual activities with both business and technical resources becoming more proficient in the language of data to drive organizational success.

11. SPONSORS

Incorta provides a unified data and analytics platform that makes it quick and easy to unlock the full potential of data from multiple complex source systems by making it instantly ready for analysis. Backed by GV, Kleiner Perkins, M12, Prysm Capital, Telstra Ventures and Sorenson Capital, Incorta empowers the most forward-thinking companies to tackle their toughest data challenges, from innovators in the midmarket to Fortune 1000 category leaders such as Broadcom, Comcast and Shutterfly. For more information, visit www.incorta.com.

14. PRODUCERS

DATAVERSITY Education, LLC is a producer of educational resources for business and Information Technology (IT) professionals on the uses and management of data. Our team strives to provide high-quality content to our worldwide community of practitioners, experts, and developers who participate in and benefit from face-to-face hosted conferences, free online events, live webinars, white papers, online training, daily news, articles and blogs, and much more.

Global Data Strategy is an international information management consulting company specializing in the alignment of business drivers with data-centric technology. Our passion is data, and helping organizations enrich their business opportunities through data and information.

Our services and training offerings include:

- Data Strategy

- Data Governance

- Metadata Management

- Data Architecture & Data Modeling

- Data Quality

- Analytics & Business Intelligence

- Master Data Management (MDM)

Our expertise has been gained in helping organizations across the globe drive value from their data, ranging from large multinational corporations to small nonprofit organizations. We are able to provide our years of experience to help you achieve your data-driven goals, and are passionate about helping customers make the most of their data assets. Find out more at www.globaldatastrategy.com or email info@globaldatastrategy.com

15. ABOUT THE AUTHORS

DONNA BURBANK is a recognized industry expert in information management with over 25 years of experience helping organizations enrich their business opportunities through data and information. She is currently 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 regularly speaks at industry conferences. She has co-authored several books on Data Management and is a regular contributor to industry publications.
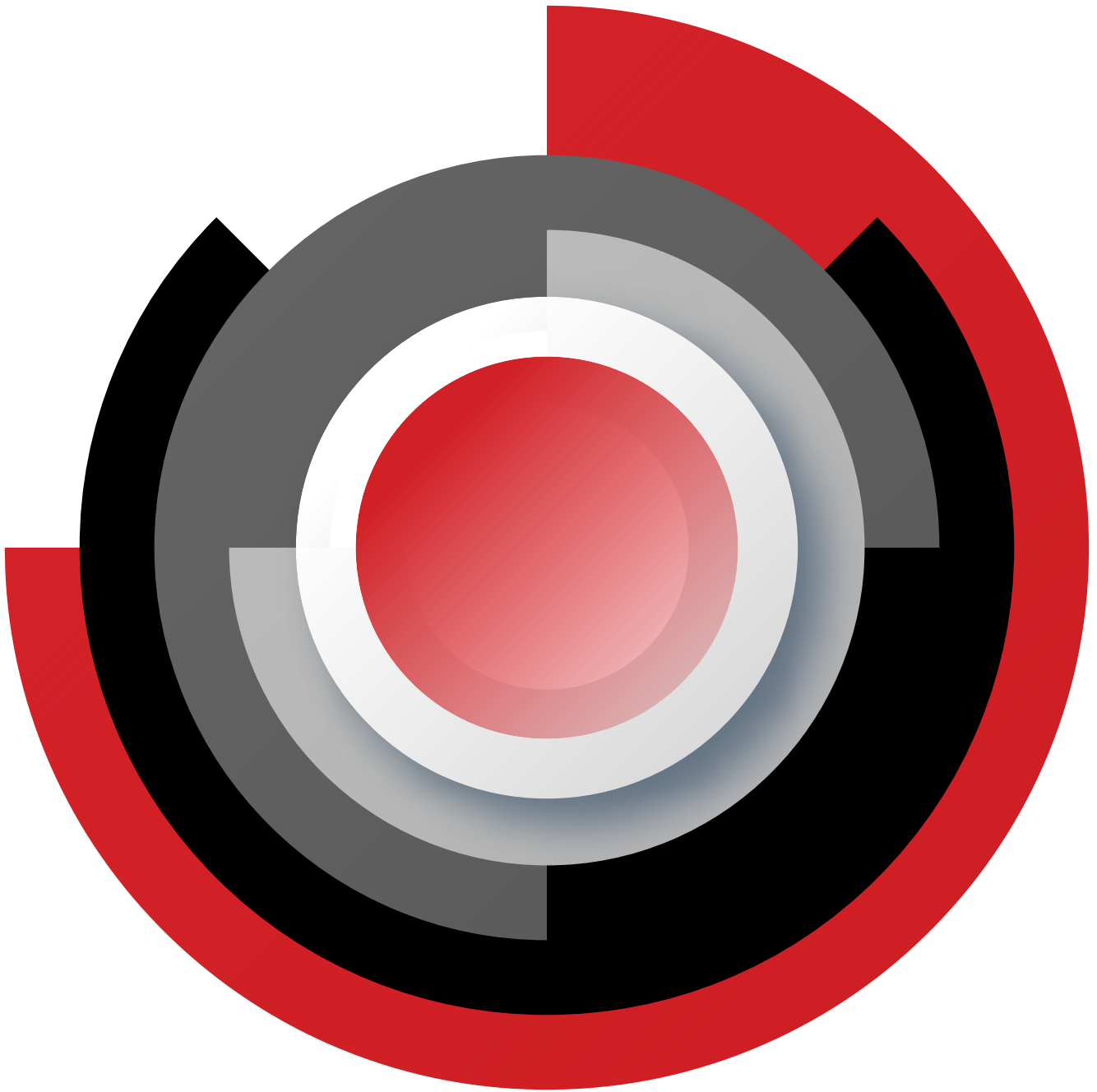
She can be reached at donna.burbank@globaldatastrategy.com, and you can follow her on Twitter @donnaburbank.

KEITH D. FOOTE has written numerous Data Management articles, with a recent focus on Data Science, and is the author of *Moving Your Business to the Cloud*, two sci-fi novels, and an alternative physics model (*The Ultra-Space Field Theory*). Keith has a strong technological background, but also enjoys the arts. He gave up his car for environmental reasons, preferring to bicycle and use public transport. Keith enjoys sci fi, yoga, mini adventures, spirituality, and chocolate ice cream.

Mr. Foote can be reached by way of LinkedIn:
https://www.linkedin.com/in/keith-d-foote-0a363868/

Sponsored by:

INCORTA

Produced by:

Global Data Strategy
The Business of Data

DATAVERSITY