

2024 Data Connectivity Report

A Look Back, a Look Ahead

WHITEPAPER



Introduction

The goal of the 2023 Data Connectivity Survey was to gather feedback about what will be important in the data space in 2024. Understanding these trends can help organizations achieve better analysis, reporting and operational results. We surveyed individuals across a variety of data management roles and revisited some of the topics surfaced in prior studies. Migration and modernization in the cloud is an ongoing theme. As organizations contend with increasingly distributed hybrid cloud platform environments, they must balance a dizzying array of data management directives imposed via laws, regulations and industry standards. While data connectivity methods are well-established in the traditional on-premises environments, the need to pivot from data extracts and reliance on ODBC and JDBC standards means increasing adoption of modern connectivity methods, particularly REST data services and APIs.

The intent with the survey was to understand the perceptions and opinions of corporate approaches in three key areas:

- Data connectivity
- Data protection, privacy and security
- Data governance and compliance

As in prior years, three distinct themes emerged from the responses:

- The need for improving enterprise data literacy
- The need to differentiate between traditional data security tactics and strategic data protection methods used to manage information risk
- The need for continued maturation of data governance operationalization through data policy management

The nexus of two of these three themes reinforces what we learned from our previous data connectivity report: businesses remain concerned about data security, data protection and other data governance issues, like regulatory compliance. One difference, however, was the emergence of data literacy as a prerequisite for maturing processes and procedures supporting data-driven operations and decision making.

Survey Response Roles

In terms of organizational roles, the participants could be mostly divided into two main categories: management or developer. Approximately 29% of the respondents were in management roles (e.g., IT Manager, IT Staff Manager, Product Manager, Vice President, CTO, CIO, CEO/owner/founder). Sixty percent of respondents worked in a technical or development role (Architect/Solutions architect, Developer, IT administrator, IT staff member, DevOps engineer, Data engineer, Data scientist, Security administrator, Database administrator, Database engineer). The remainder were either external consultants or business people (business analyst, marketing specialist, etc.).

The most frequently represented industries included software development, banking/finance, computer-related products or services, healthcare, education, manufacturing, retail, technology or business process outsourcing, online IT services and telecommunications. The full distribution by industry is captured in Figure 1.

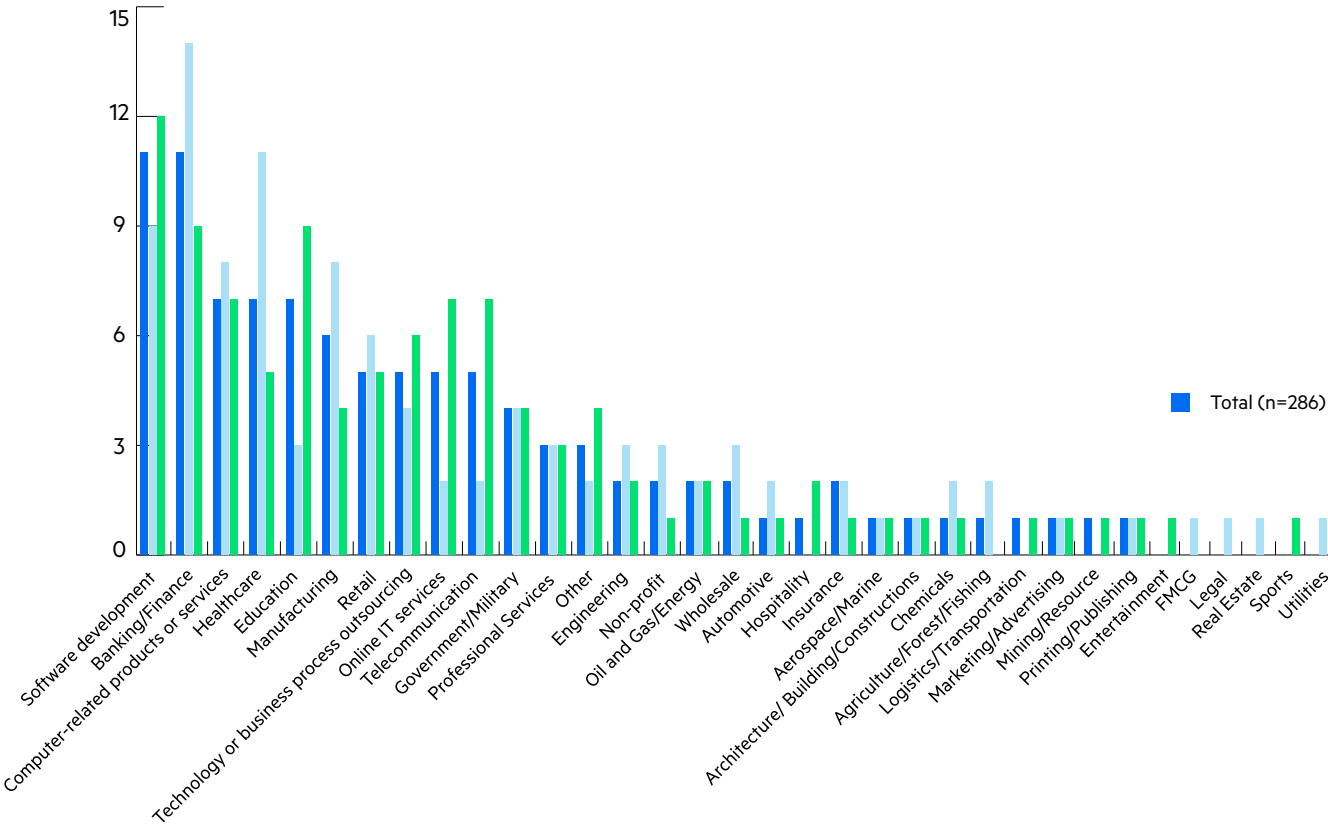


Figure 1: Distribution of respondents by industry

Of the total sample and those who managed to fully complete the survey, the largest cohort of respondents (39%) work in the US. Other represented countries include Canada, the United Kingdom and India—with respondents from each of these countries representing approximately 5–6% of the total participant populations.



Theme #1: Enterprise Data Literacy

There are three trends that influence organizational perspectives on treating data and information as resources to drive profitable business decisions:

- **Exploding data volumes:** The growth of data volumes continues to accelerate.
- **Being “data-driven”:** There are ongoing initiatives to focus corporate attention on leveraging data to improve the way the business operates.
- **Analytics democratization:** Simplified end-user analytics tools and increased access to data has empowered new communities of citizen data analysts.

Data analysts and data scientists alike are impeded by issues with data connectivity and access, along with a lack of knowledge of what available data sets best meet business analytics needs. Downstream data users are interested in using and analyzing enterprise data sets, yet as the number and sizes of the data sources increase, those users are hampered by insufficient data literacy and data awareness. This brings us to the first theme of this year’s survey: **enterprise data literacy**.

Gartner defines data literacy as “the ability to read, write and communicate data in context, including an understanding of data sources and constructs, analytical methods and techniques applied, and the ability to describe the use case, application and resulting value.”¹ Harvard Business School defines data literacy as “a term used to describe an individual’s ability to read, understand, and utilize data in different ways. It doesn’t require an individual to be an expert—as a data scientist or analyst might be considered—but rather, to show an understanding of basic concepts, such as: Different types of data, Common data sources, Types of analysis, Data hygiene, and Tools, techniques, and frameworks.”²

Training data users to become data literate involves:

- Giving them the basic knowledge of different types of data sources.
- Raising awareness of the role that data sources play in operationalizing business processes.
- Learning skills in data-driven problem-solving.
- Gaining the ability to identify and articulate information requirements to support problem-solving.
- Having a fundamental understanding of the data analytics lifecycle.
- Acquiring the ability to communicate a story using data.

¹ Panetta, Kasey, “A Data and Analytics Leader’s Guide to Data Literacy,” 08/21/2021, accessed 10/16/2023 via <https://www.gartner.com/smarterwithgartner/a-data-and-analytics-leaders-guide-to-data-literacy>

² Stobiersky, Tim “Data Literacy: An Introduction for Business,” 02/23/2021, Harvard Business School Online, accessed 10/16/2023 via <https://online.hbs.edu/blog/post/data-literacy>

The 2023 survey results demonstrated the need for improving data literacy and data awareness. In this section, we discuss the questions the survey asked about different aspects of data awareness: knowledge of the number, type and origin of data sources; how to find, connect to and access data sources; who is responsible for managing the data sources and how to bring new data sources into the organization. While some survey participants exhibited a growing level of data literacy, it remains an imperative for truly empowering corporate data users.

1.1 Knowledge of the Number of Data Sources

Only 17% of respondents had the impression that data users know the number of data sources very well.

The survey asked the respondents how well they believed data users in their organization knew the number of data sources available for use. The objective of the question was to assess the respondents' perception of the degree to which enterprise data users are aware of the data landscape. While 35% said the data users were moderately aware of the number, almost half indicated that data users had very limited knowledge, had no knowledge or were unsure.

There are two potential conclusions one can draw from these results. The first presumes that the respondents' perceptions about the enterprise data user audience are accurate. In this scenario, these numbers indicate an acknowledged gap in data awareness, which is a key component of data literacy. Only a little over half of the respondents indicated that their data users had what might be termed a reasonable working knowledge or awareness of the number of data sources available for use. This gap could be addressed through the introduction of a data awareness campaign.

The second conclusion drops the presumption of respondent accuracy and could indicate a different type of data literacy gap: a lack of visibility into the capabilities of the organizational data user communities. This would need to be addressed through improvements in operational data stewardship processes.

It is important to remember that, either way, these figures may be deceptive. Respondents self-report about the number of data sources available, when they may not be aware of all the data sources available. In those cases, these statistics might not reflect the actual number of available data sources.

1.2 Knowledge of Means for Accessing Data Sources

Sixteen percent believed the data users know how to access data sources very well. Thirty-seven percent said the data users know how to access data sources moderately well.

The survey asked participants how well they believed the data users in their respective organizations know how to access the data sources available for use. As opposed to asking about data awareness, this question focuses attention on perceptions about individual data users' hands-on familiarity with the methods by which the different data sources are accessed. The numbers for this question mostly mirrored the previous question about knowledge of the number of data sources. A surprisingly low percentage (48%) believed users' knowledge of accessing data sources was very limited, completely limited or were not sure.

The responses about number of data sources and familiarity with data accessibility raises two important questions about enterprise-wide data accessibility and familiarity with how source data sets are accessed:

1. How many staff members in the organization need to access data sources, and how many of the data users in the organization need to know how to access data sources? More to the point: how many data users are there, and how many of them are accessing the source data as opposed to looking at data products manufactured from source data?
2. To what extent are data users trained on data accessibility, and to what extent do data users need to be trained on data accessibility? In other words, are there opportunities for simplifying how individuals access multiple data sources?

These two questions highlight an emerging need to enable access with integrated governance. Increasingly, self-service mechanisms must provide authorized access to requested data in a way that makes enforcement of privileges opaque to the data user. At the same time, those self-service layers should simplify data access by hiding the complexity of data distribution and federating data accessed from the physically disparate parts of logically distributed data sources.

1.3 Data Stewards

The survey posed a question about corporate data governance: How well do you believe the data users in your organization know who the data stewards are? TechTarget defines the role of the data steward as the individual "responsible for carrying out data usage and security policies as determined through enterprise data governance initiatives, acting as a

50% of respondents believe data users know who the data stewards are moderately or very well.

liaison between the IT department and the business side of an organization.”³ Thirty-four percent believe that knowledge of who the data stewards are is very limited, while 15% believe that users do not know who the data stewards are at all or are unsure.

While this is really a governance and process question, it also ties into the questions about data awareness. More specifically, in a governed environment with mature processes in place, one might expect a level of opacity with respect to the individuals tasked with data stewardship. Do team members really need to know who the data stewards are as long as the process architecture ensures that data stewardship is effectively managed? If there are good practices and processes in place, workflow processes would enable communication between data users and the data stewards, even while maintaining barriers between them.

1.4 Internally vs. Externally Sourced Data

The survey asked about the percentage of data sources that originate within the organization as well as the percentage made available from outside parties. The means of the responses were 69.4% shared from within the organization and 30.6% shared from outside parties. This indicates, that on average, respondents believe nearly a third of the data sources used within the enterprise originate from outside the organization.

In the earlier days of data warehousing and business intelligence (i.e., “pre-cloud”), most organizational analytical systems were populated with data mostly, if not totally, originating from sources behind the corporate firewall. And although organizations may have licensed third-party data from external vendors and aggregators, contractual stipulations may have prevented wholesale integration of externally acquired data sets.

The responses to this question were interesting, as they indicated the trend towards increased adoption and incorporation of external data into the enterprise data landscape. External sources might include the third-party aggregators, but also might include open data sets (such as those provided by governments or as a courtesy by some commercial organizations) or data provided through data marketplaces or facilitated via vendor “cleanroom” append environments.

The growing use of externally sourced data is not risk-free. An organization considering bringing in more external data must consider the governance and quality-assurance aspects of data created outside of the organization’s administrative oversight. Data stewardship for externally sourced open data requires specification of data quality

³ Definition of “data steward” accessed 2023-11-21 via <https://www.techtarget.com/searchdatamanagement/definition/data-stewardship>

expectations and methods for instituting the appropriate cleaning tactics. These methods help keep the data set in its raw form, which is vital since material changes might modify the semantics and utility of that data. When using data from third-party providers, organizations should review the licensing and data use agreements and identify any data policies that direct monitoring or mandatory reporting to the provider to better comply with contractual agreements.

1.5 Finding Data Sources

When asked about the ease of a data user to find data sources for their specific needs, we see a typical distribution. Overall, only 9% say that it is extremely easy (12% for Progress DataDirect users). While 30% say it is somewhat easy, 30% say it is somewhat difficult or extremely difficult, 31% said it is neither easy nor difficult.

This is not an unexpected distribution. Approximately 10% of the surveyed population could be characterized as “savvy” or “expert” data users. However, this again raises the meta-question of data literacy, which encompasses practical questions about the fundamental need for data users to know how to find data, and what their expectations should be about finding data sources to address their business needs. A different way to look at this question revolves around that 30% of the audience for whom it is somewhat or extremely difficult. These numbers suggest that it is good practice to seek to understand the root causes of their challenges and consider tools and techniques that can be used to alleviate that frustration.

1.6 Bringing New Data Sources Into the Organization

Our next question asked about the ease of bringing a new data source into the organization. The shape of the distribution of the responses reflects, for the most part, a normally distributed population, with 32% saying it is neither easy nor difficult, 28% saying it is somewhat difficult and 25% saying it is somewhat easy, with smaller percentages for either “extremely easy” or “extremely difficult.”

These results might hide a more complex set of questions about new data sources, especially because what it means to “bring in a new data source” can vary. It can range from simply importing or downloading a workbook or CSV file to someone’s desktop machine, to extracting data from an external repository, to continuously acquiring and storing streaming data via a set of APIs.

The ease of bringing in a new data source will differ depending on the type of data source, its size, the methods used to access it, where the data will be managed and how to document the data source and facilitate its general availability and accessibility. In other words, simplifying the integration of new data sources implies instituting governed processes for documenting data source metadata and data curation processes to make that data source a sharable and curated resource available to data users across the enterprise.

1.7 Types of Data Sources

Survey participants were asked to indicate the types of data sources their organizations are connecting to today. Seventy-six percent of respondents indicated they were accessing on premises, 62% are connecting to a cloud database, 31% reported connecting to data managed by a SaaS, 22% are connecting to a cloud data lake and 9% are connecting to a cloud data lakehouse.

The high percentage of users still accessing on-premises data sources suggests that there remains a strong reliance on them. Those data sources might be the primary sources (i.e., data pulled directly out of application systems) or on-premises data warehouses and data marts. At the same time, there is a relatively high number of individuals reporting that they are accessing databases in the cloud, indicating a growing adoption of cloud-based services and resources.

There are relatively lower numbers of individuals reporting that their organizations are accessing data in cloud data lakes or cloud lakehouses. It may be that there is still some confusion about what these architectural paradigms really mean in the context of an enterprise data landscape strategy. One reason might be that there are particular vendors that advocate for those data platforms as options, and they may be promoting those alternatives to their customers.

1.8 Future Data Sources

The survey also solicited information from the participants about their organizations' future plans, asking about other types of data sources the organization wanted to connect to in the future. Respondents were able to select multiple answers, with the results showing continued interest in connecting to data sources in the cloud. Two off-premises choices competed for the top future choice: 26% anticipated connecting to SaaS systems, 26% looked forward to accessing data in a cloud data lake, while cloud databases (23%) and cloud data lakehouses (22%) were close behind.

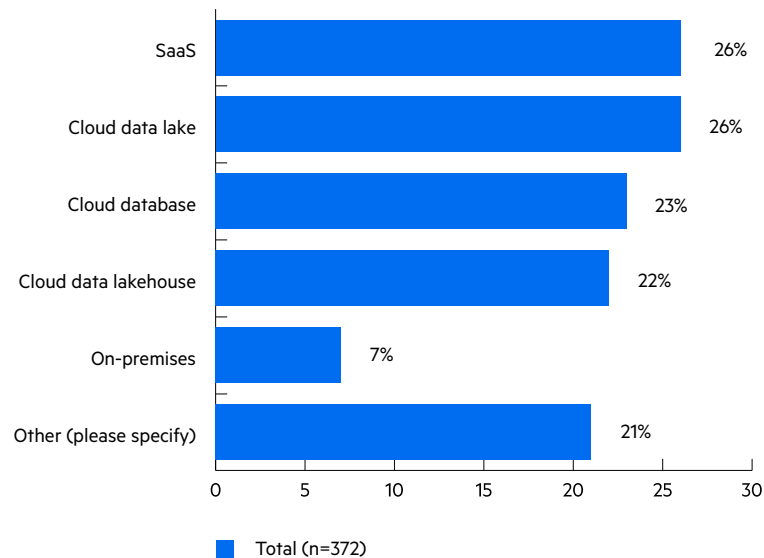


Figure 2: What other types of data sources do you want to connect to in the future?

The most telling result was the low percentage associated with on-premises data sources (7%), indicating an expectation of decreased reliance on data coming from on-premises data sources.

Aside from the list of options provided, respondents indicated an interest in accessing other types of data sources as well as expanding the use of API access methods. Some of the data sources that respondents wanted the ability to connect to in the future include:

- IoT (Internet of Things)
- Office productivity tools
- Proprietary tools (e.g., Excel)
- Openly accessible tools (e.g., Google Sheets)
- Websites
- Shared file systems
- Vendor knowledge bases
- REST APIs
- Kafka or other data-streaming services

1.9 Methods to Connect to Data Sources

In addition to soliciting responses about the data sources to which the respondent organizations connected, the survey asked about the methods being used today to connect to those data sources. It was not surprising to see a relatively large percentage

accessing data through ODBC (59%) and to a lesser extent, JDBC (36%). The percentages of respondents indicating data extracts (48%) and application connectors (48%) reflect a natural balance between legacy approaches to data access and emerging frameworks that leverage software connectivity.

The most interesting result is the high percentage of organizations employing APIs (78%).

In our prior survey reports, we saw a continuous uptick in adoption of APIs for connectivity, and this high percentage demonstrates that APIs are rapidly becoming mainstream. The emergence of APIs as the dominant method of connecting to data is important, since it signifies a transition from a source-bound approach to data accessibility (i.e., using the data source's methods for connectivity) to a neutral approach. This neutral approach favors data users because it eliminates the need for those users to become proficient in the variety of data sources.

1.10 Future Connectivity

Participants were then asked about the methods they wanted their organizations to use in the future to connect to data sources. Considering the large percentage of respondents indicating their current use of APIs, it is not a shock to see that a relatively low percentage (10%) anticipate adopting APIs in the future.

The two methods that were most frequently reported were cloud-native methods (32%) and SaaS connectors (26%). Together, these figures suggest a continued migration of data and applications to cloud-based and externally hosted systems. Cloud-native revolves around connectivity services engineered to leverage cloud technology, which benefits those organizations considering implementing a hybrid cloud data fabric. Employing cloud-native connectivity services helps to seamlessly stream data from a variety of sources to their final destinations.

Another interesting artifact from this question is the percentage of individuals indicating an expectation of connecting via Hive (12%), which was a higher figure than those reporting current use of Hive at their organizations (8%). Although interest in using components of the Hadoop ecosystem has largely diminished, Hive provides a mechanism for introducing a data schema on top of structured data, particularly that managed in cloud object storage. This reinforces the impression of an ongoing march to the cloud.



Theme #2: Data Protection and Information Risk

Organizations are concerned about several factors: the increasing number and breadth of jurisdictional data privacy laws, susceptibility to data breaches, potential leaks of important corporate information and responsiveness to customer anxiety about the appropriate use of their data. This has prompted interest in corporate data security protocols and methods for protecting against improper use of sensitive information.

The conventional knowledge in the IT industry is that preventing unauthorized access to sensitive corporate data is a priority, whether your organization is engaged in a cloud-oriented digital transformation effort, a data architecture modernization initiative or has been cloud-native from the start. And for a long time, companies have instituted perimeter security measures to help protect against unauthorized entry through the corporate firewall.

However, there are some mitigating factors that render perimeter security insufficient for protecting data. Data volumes continue to explode, and enterprise data landscapes are becoming increasingly complex. Data integration and fusion, in which data sources are combined to create new information products, allow for sensitive information to be inferred through a combination of relatively benign data sources. Limited controls placed on data scientists allow for questionable choices to be made when accessing personal or private customer data. Insider threats are also not addressed in perimeter security tactics.

Complying with the data policies that govern data security and data protection has become increasingly complex. Without instituting the right tools, processes and practices, data protection becomes unscalable and ultimately unsustainable.

In other words, there are differences between data security tactics and methods, which are typically used to protect against data access and data protection tactics and methods intended to prevent unauthorized data use. This year's survey captures some of the differences the respondents perceived in distinguishing data security from data protection. The survey solicited opinions about corporate responsibility for data security and data protection, data protection methods and applying data protection at different points in the data lifecycle.

1.11 Responsibility for Data Protection

The survey asked respondents to indicate who they believed was responsible for data protection in their organization. For this multiple-selection question, nearly one-third (32%) indicated that their organization had a dedicated Data Protection/Compliance Office, suggesting that a corporate program for data protection had been established. Almost a quarter (24%) selected the CIO (Chief Information Officer) or the CISO (Chief Information Security Officer), while 20% chose the Chief Technology Officer. This might reflect a reasonable assumption about corporate responsibility.

More surprising, were some of the other choices and their respective percentages. Fifteen percent of respondents picked the CEO, which could indicate a justifiable assumption that the head of the organization is ultimately responsible for protecting sensitive information.

Seventeen percent selected the Legal department. Although the Legal department staff might be involved in reviewing the legal and compliance directives motivating the need for data protection, their expertise might be focused on interpreting the law and directing data policy, as opposed to implementing compliance with those policies.

Even more curious were the remaining percentages. Six percent indicated that an external consultant was responsible for data protection. It would be surprising (and risky!) for an organization to outsource such a critical function to a third party.

Fourteen percent of respondents indicated they were unsure as to where the responsibility for data protection lay. With growing concern about protecting sensitive data amid ongoing cyber threats and data breaches, everyone in an organization should be clear about who is responsible for data protection.

1.12 Responsibility for Data Security

The subtext of the next question in the survey was to assess the degree to which respondents differentiated between oversight of policies governing protection of sensitive data and management of the infrastructure and processes for data security. The survey asked, “Who is responsible for data security within the organization?” Participants were allowed to select multiple roles.

A relatively large percentage of respondents selected IT administrator (39%) and/or Security operator/administrator (SOC) (25%)—roles reasonably expected to be responsible for data security infrastructure. Those who selected CIO/CISO (27%), CTO (22%) or CEO (14%) may have interpreted the question as who is ultimately responsible. In most cases,

though, C-suite individuals are not likely to have an operational role in the day-to-day operations of data security.

The interesting results are those that indicate a disconnect between understanding data protection and data security. Thirteen percent selected the Legal department. While the Legal department might be instrumental in interpreting how policy dictates the need for protection, lawyers are not likely managing the security perimeter. As with the previous question, those indicating an external consultant (5%) and those who are unsure (7%) raise the risk of introducing more vulnerabilities into the enterprise.

1.13 Data Security Processes

Data protection is critical, and there are many different data security tools and techniques put in place to support protection of sensitive information. Survey respondents were asked about which data security processes and systems were currently applied in their company.

The most popular selections included: authentication (65%), data backups & recovery (60%), anti-virus (60%), role-based access control (RBAC) (57%), data encryption (47%), perimeter security (firewall, IDS etc.) (46%) and end-point security (41%).

Based on the responses we've seen, the processes, methods and tools used for promoting data security can often be confusing. The survey asked the participants to select the data security processes and systems they perceived as being the most challenging to integrate.

The most frequently selected method was role-based access control (RBAC), selected by 33% of the respondents, followed by intrusion detection and prevention systems (30%), authentication (29%), data encryption (28%) and data leak prevention (28%).

If data protection involved only one of these methods, these percentages might cause a bit of consternation. However, because data protection implementations must employ many methods simultaneously, challenges in integration of any single method will inherently impact the others. This is especially true when different teams are tasked with managing various methods and techniques without enterprise-wide coordination.

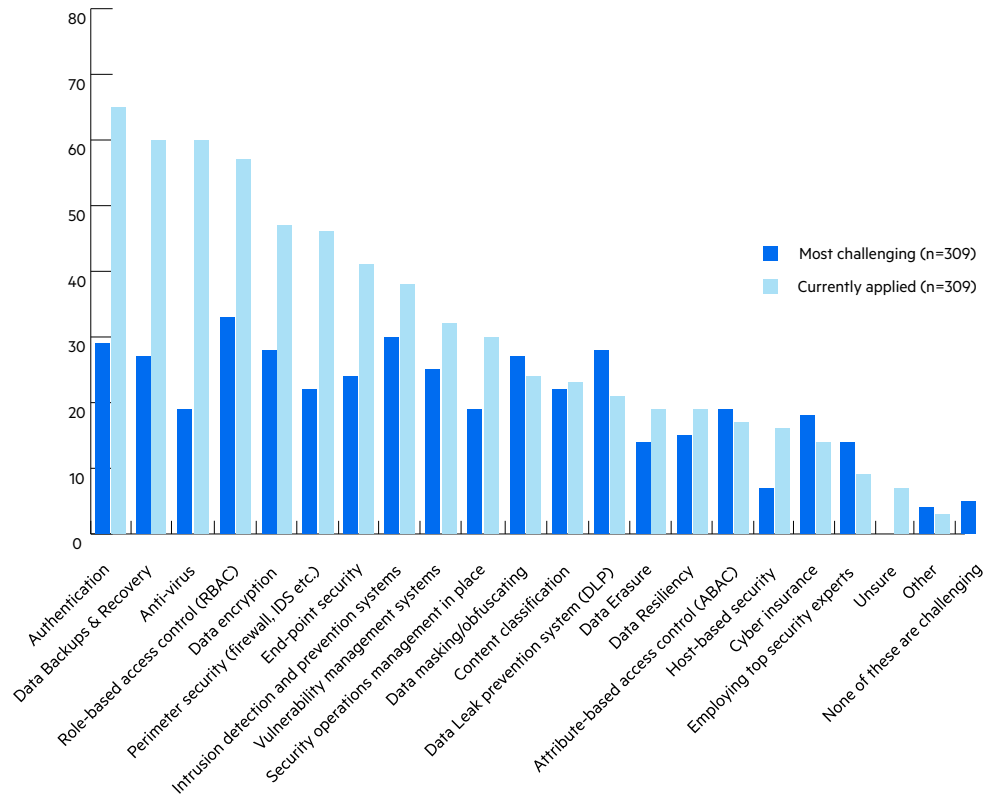


Figure 3: Currently applied data security processes compared to data security processes that are most challenging.

1.14 Masking and Encryption

Data masking is a process that rearranges data values or replaces parts of the data values so that the masked data retains the format and type of the original data, without the sensitive elements. Encryption is a method of protecting data using cryptographic (mathematical) algorithms that transform the data into an unreadable form which is only reversible using a decryption key. Encryption is a powerful data protection technique that can be used on stored data (data at rest) as well as data in transit. Processes that employ sensitive data can use data masking and encryption tools to support directives associated with sensitive-data protection, such as privacy laws (e.g., GDPR or CPRA).

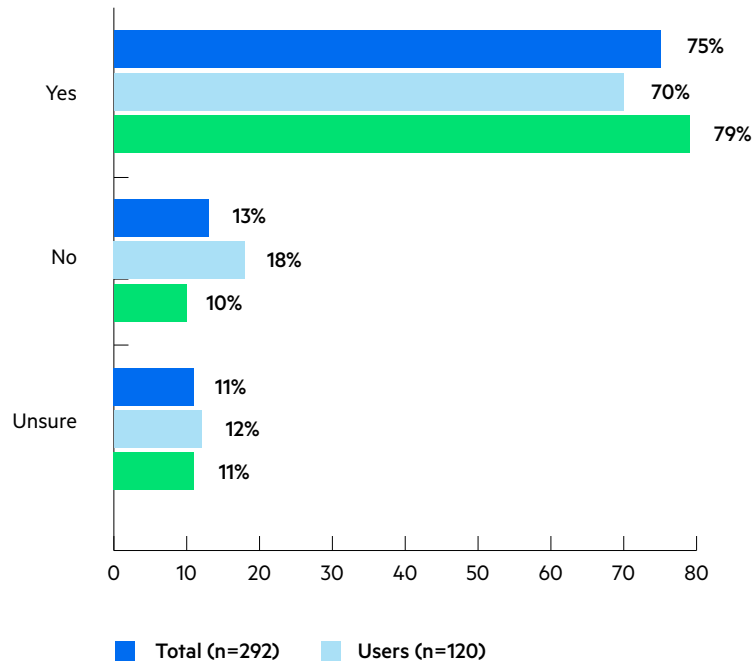


Figure 4: Does your organization use data masking to hide sensitive data?

The survey asked participants about their organizations' use of masking and encryption. When it came to data masking, 53% of respondents said their organization used it, 24% said they did not and 23% were unsure. A larger percentage of the respondents indicated that their organization used encryption (75%), while 13% said their organization did not and 11% were unsure.

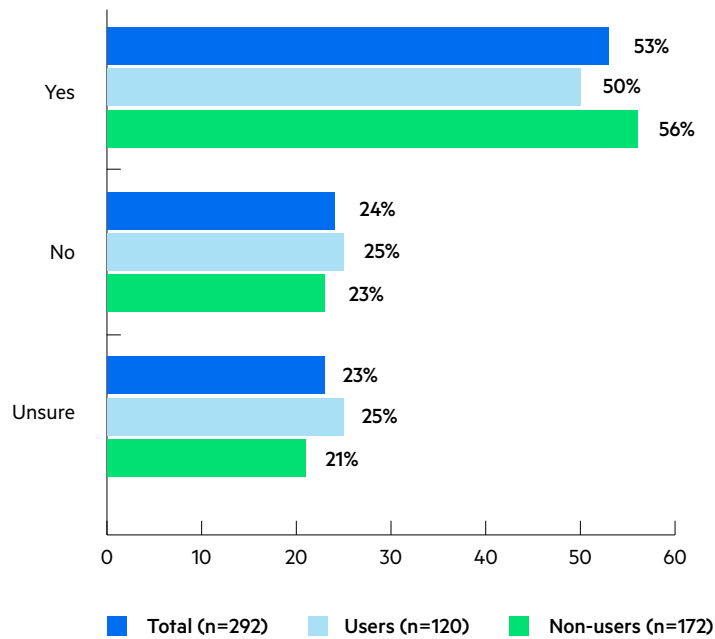


Figure 5: Does your organization use encryption to hide sensitive data?

1.15 Protection of Data at Rest and Data in Motion

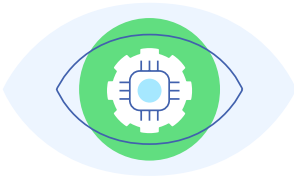
Even in an organization that has a variety of controls in place for data protection, it is still imperative to recognize the challenges of moving sensitive data used in application development or analytics. Any time data is provided for these purposes, the data team must be aware of the data pipelines, information flows and data touch points.

Data protection depends on the ability to assess the full data lifecycle to determine any points at which data exposure can occur. That includes identifying those data sets that contain sensitive information requiring protection, and determining where the data set has been stored, which processes might request the data, the pipelines through which the data set will move and how the data is to be presented to the requestor.

The survey asked respondents about whether their organization applies data protections to data at rest. About two-thirds of respondents replied that the organization applied data protections to data at rest, 12% said their organization did not and 24% said they were unsure.

The survey asked respondents about whether their organization applies data protections to data in motion. In both cases, about two-thirds of respondents replied that the organization applied data protections to data in motion, 13% said their organization did not and 24% said they were unsure.

We can interpret the percentages of unsure respondents in two different ways. The first is to take it at face value and assume the survey participant was truly unaware of what data protection techniques were (or were not) applied to either data at rest or data in motion. The other possibility is that the organizations do apply data protection (to both data at rest and data in motion), but provide techniques to present the data in its unprotected form to the data users, thereby shielding them from the details of how and where data protection techniques are applied.



Theme #3: Operationalizing Data Governance and Policy Compliance

Laws, regulations and industry standards are examples of externally defined policies intended to direct organizational behavior associated with data collection, connectivity, access and use. While data privacy laws like GDPR remain top of mind for many data professionals, there are numerous other laws and regulations that impose constraints or controls on how data is used. Constraints may originate from several sources, including:

- Government requirements for mandatory reporting
- Quality controls for data sets made publicly available by government entities
- Obligations for data retention and disposition schedules
- Requirements for ensuring integrity of results of studies and analyses

No matter the originating source, government and industry mandates impose operational policies on the management and use of information needing ongoing auditing and monitoring.

Data policies must provide assurance that data consumers are able to access the data they need under the appropriate circumstances and usage scenarios. At the same time, organizations must define data policies that guard against inappropriate use. This year's survey solicited opinions from respondents on:

- Laws and regulations
- Compliance support methods
- Issues with the evolving data landscape
- Monitoring and auditing processes
- Limitations resulting from compliance directives

The results suggest that opportunities still exist for increasing the level of maturity for data governance. Organizations need to increasingly refocus their data governance efforts on connecting business objectives with data usability

The first step in operationalizing data policies that deliver quantified business value is interpreting externally imposed policy directives and converting them into data policy rules. The data governance team's role is to articulate the data restrictions and constraints that are implied by policy mandates.

1.16 Subjected to Regulatory Compliance

The survey asked respondents if their organizations were subject to any regulatory or compliance requirements that explicitly or implicitly specified policies for data protection. Seventy-four percent of the surveyed population answered in the affirmative, 14% said “no” and approximately 11% were unsure.

The respondents who indicated that their organizations were subject to compliance were asked to select the laws and regulations their organizations followed. Not surprisingly, the European Union’s General Data Protection Regulation (GDPR) was most frequently selected (52%).

Next, at 38%, was ISO/IEC 27001, which is the international standard for managing information security. This standard encompasses three directives for ensuring confidentiality, integrity and availability:

- **Confidentiality:** allowing only authorized persons access to information
- **Integrity:** limiting the ability to change information to authorized persons
- **Availability:** enabling authorized persons to access the information when needed

The remaining percentages can be seen in Figure 4. on page 16

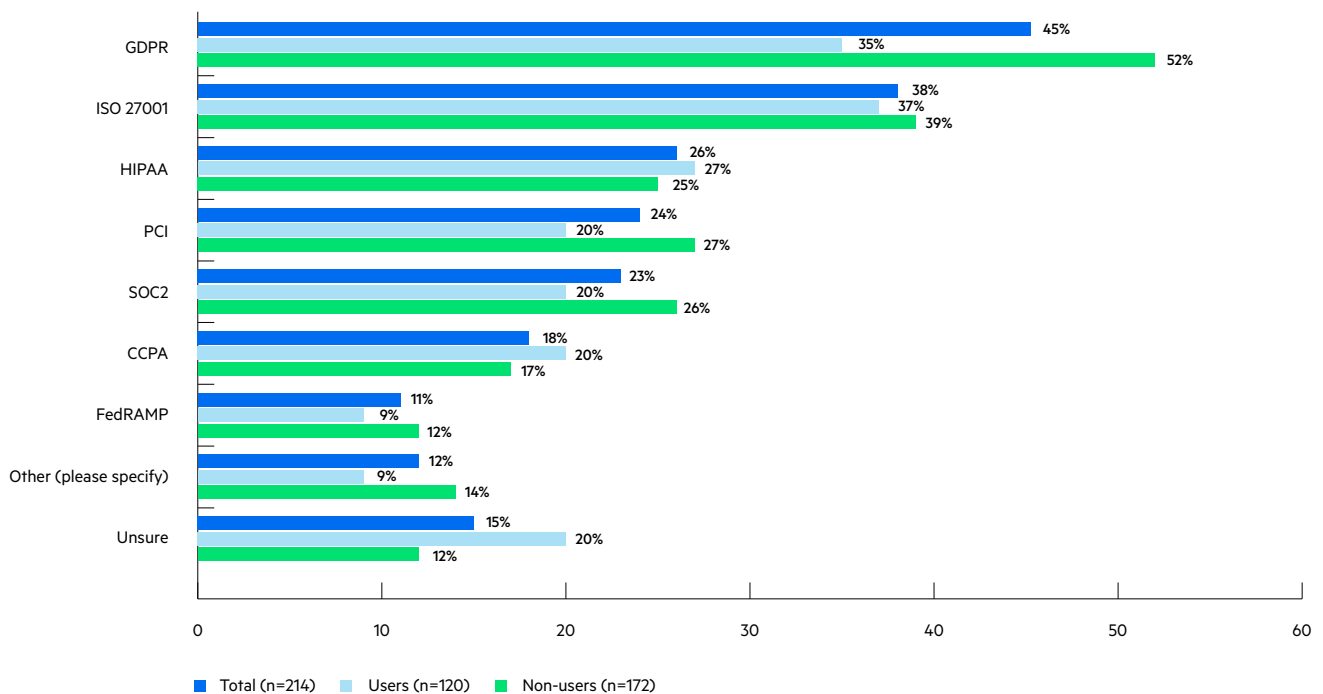


Figure 6: Which regulatory or compliance requirements does your organization follow? (Multiselect)

These findings are illuminating in two ways. First, it is reassuring to see that nearly three-fourths of respondents expressed awareness of their organizations' need to observe regulatory or compliance directives. Second, it is a bit alarming that the remaining quarter of respondents reported that their organizations were not subject to any requirements related to data protection—or were uncertain if they were or not. It would be unusual for an organization not to be governed by some set of laws or regulations, unless the company never handled information deemed sensitive in relation to privacy law. This is reflected somewhat by the 15% of respondents indicating that their organization was subject to compliance but were unsure of which regulations their organization followed. All of this suggests that discussions of legal compliance might be a worthwhile addition to a corporate data awareness campaign.



1.17 Managing Compliance

The survey asked respondents to grade how easy or difficult it was for their organization to manage compliance with data privacy laws. Ten percent of the audience indicated that it was extremely difficult, 45% said it was somewhat difficult, 28% noted it was neither easy nor difficult and 17% said it was somewhat or extremely easy.

The survey also asked participants to indicate the methods used to reach compliance from a limited selection of methods associated with data security processes. The resulting percentages, shown in Figure 5 on page 16, indicate that the perception of data privacy compliance remains rooted in some very basic tactics.

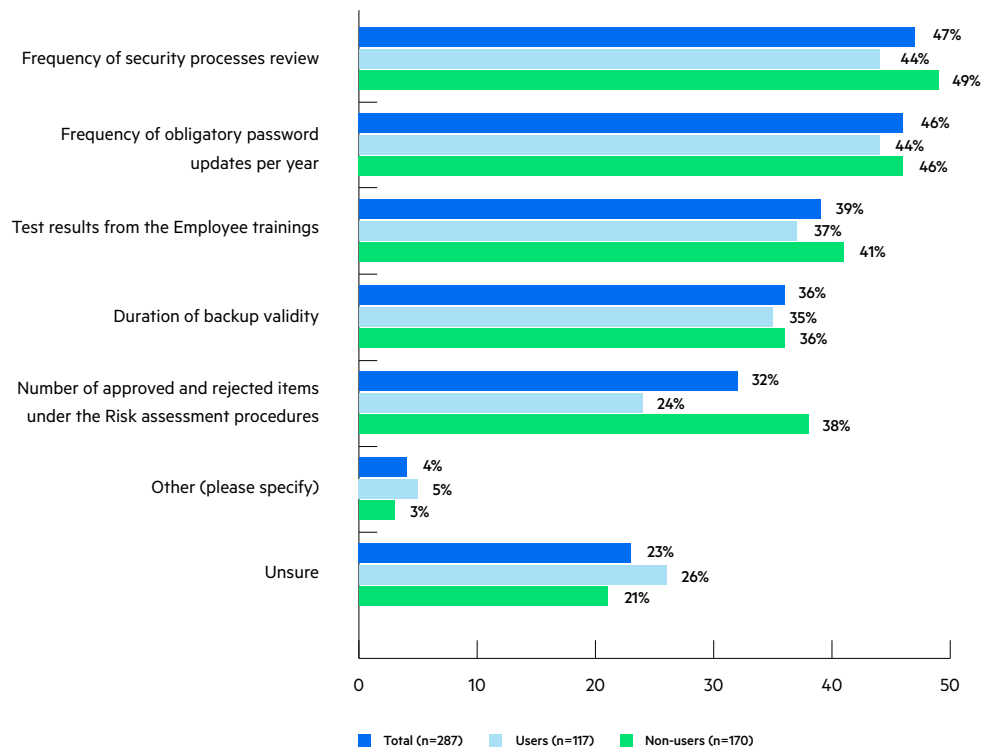


Figure 7: Methods used for reaching compliance

For example, the most frequently selected option, “Frequency of security processes review,” is related to operational techniques, and the second most frequently selected option, “Frequency of obligatory password updates per year,” is centered on system access controls for known entities. Neither of these address vulnerabilities associated with inadvertent data exposure, either to unauthorized insiders or via data breaches.

Organizations will require more comprehensive information risk assessment processes and procedures that will augment the types of tactical security methods listed as the options for this question (such as “security process reviews,” “obligatory password updates,” “test result from employee training,”). These tactics are necessary but not sufficient to satisfactorily address the expanding array of vulnerabilities that pose critical information risks.

1.18 Concerns About Cloud

The survey posed a straightforward question: Is the use of cloud providers concerning when trying to meet your compliance goals? Generally, the responses were evenly split: 40% said yes, 40% said no and 20% were unsure. Increasing confidence in cloud service providers will continue to be critical when supporting any type of policy compliance practices.

1.19 Monitoring for Compliance and Authorized Use

Monitoring and auditing are key practices for data governance and policy compliance. The survey asked participants two questions about monitoring and auditing: To what extent does your organization monitor and audit compliance, and to what extent does your organization monitor and audit for unauthorized data access and use? The results are encouraging. In both cases (compliance and unauthorized data access), the combined percentages of respondents indicating their organizations extensively or moderately monitored and audited compliance and unauthorized data access and use were both 79%. This suggests a growing level of maturity for operationalized data policy governance practices.

1.20 Limitations on Amounts of Data and the Need for a US Data Privacy Law

The final questions the survey posed to the participants focused on perceptions about the intersections between data privacy laws and limitations on data use. The survey asked if respondent organizations had limited the amount of data that could be accessed and used for analytics and business intelligence due to regulations. Almost half (47%) of the respondents replied “yes,” approximately one-third (32%) replied “no” and 21% said they were unsure.

These statistics corroborate a growing concern about the impacts of data privacy. For instance, in the absence of a mature set of processes for instituting appropriate controls, organizations are more likely to cut off access to data rather than risk exposing sensitive data in ways that might lead to noncompliance penalties.

Data Literacy and Data Awareness Is an Imperative for 2024

Decision-makers adopting a “data-driven” approach will typically ask several key questions about their processes for business intelligence, reporting and analytics:

- What data sources are available to inform data-driven decision-making?
- How can I find, become authorized to use and access those data sources?
- How is data from these data sources combined to produce the information necessary for business decisioning?
- What other decision makers rely on the same information?
- What are the organizational practices for ensuring protection and appropriate use of those data sources?

Increasingly distributed data architectures complicate business user initiatives to maximize data use. Therefore, instituting data literacy techniques will help raise data awareness by informing individuals about:

- Data inventory: What data sources are available
- Data source metadata: What information is contained within each data source
- Data type: The type/format of each data asset (e.g., structured vs. unstructured)
- Data quality: Quality characteristics of each data source
- Data lineage: How the data flowed through the data pipelines
- Data insights: Collaborative feedback about data usability of each data source

Looking forward to the next year, the results of this survey imply a need for augmenting the data connectivity strata and the underlying data fabric with the tools and techniques that raise data literacy and data awareness. Some key examples of this include:

- A data catalog that captures “data intelligence” about the data sources
- Data lineage tools that can map the data pipelines and identify data product “manufacturing” dependencies
- A robust framework for data connectors that seamlessly support access to all enterprise data resources
- Integrated data protection methods such as masking, encryption and access controls to authenticate access for authorized data users

Real-world examples of data users seeking guidance on fulfilling their business application’s information requirements can inspire data practitioners to take the practical steps for improving data awareness and increasing data literacy. Data stewards and other data governance professionals who understand the complexity of the data landscape and its plethora of data pipelines can re-engineer the enterprise data fabric and create a semantic partition that simplifies user data access. Visibility of the data landscape through a data catalog gives users the ability to understand the contents of the data sources, while a semantic connectivity layer that integrates data policy monitoring, auditing, protection and compliance will reassure users that data safeguards are being observed.

Survey Methodology

Survey responses were collected between March 30 and May 26, 2023. The respondents consisted of Progress® DataDirect® contacts who opted to receive surveys, users of other Progress products and individuals contacted via an external agency. All were invited to complete a questionnaire via email, and 285 successfully completed the survey.



About the Author

David Loshin, president of data strategy consulting company Knowledge Integrity, Inc. (www.knowledge-integrity.com), is a recognized thought leader and expert consultant in the areas of data governance, quality, management, and analytics. David is a prolific author regarding information management best practices as the author of numerous books and papers, including Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools, Techniques, NoSQL, and Graph, The Practitioner's Guide to Data Quality Improvement, and Master Data Management. David is a frequent invited speaker at conferences, web seminars, and sponsored websites and channels.

David is also a Senior Lecturer and Faculty Lead for Careers and External Relations, as well as the former director of the Master of Information Management program at University of Maryland's College of Information Studies.

David can be reached through LinkedIn, or email via loshin@knowledge-integrity.com.

The 2024 Data Connectivity Report was prepared by David Loshin in their personal capacity. The opinions or representations expressed herein are the author's own and do not necessarily reflect the views of Progress Software Corporation, or any of its affiliates or subsidiaries. All liability with respect to actions taken or not taken based on the contents The 2024 Data Connectivity Report are hereby expressly disclaimed. The content on this posting is provided "as is" with no representations made that the content is error-free.








Get the most from your data. Learn how Progress DataDirect can help your business.

About Progress

Progress (Nasdaq: PRGS) provides software that enables organizations to develop and deploy their mission-critical applications and experiences, as well as effectively manage their data platforms, cloud and IT infrastructure. As an experienced, trusted provider, we make the lives of technology professionals easier. Over 4 million developers and technologists at hundreds of thousands of enterprises depend on Progress. Learn more at www.progress.com

© 2024 Progress Software Corporation and/or its subsidiaries or affiliates.
All rights reserved. Rev 2024/01 | 1206214-071863559

 /progresssw
 /progresssw
 /progresssw
 /progress-software
 /progress_sw_