

EDM COUNCIL

CLOUD DATA MANAGEMENT

BENCHMARK REPORT

July 2023



Executive Summary

Executive Summary

Welcome! It is my pleasure to present to you the 2023 EDM Council's Cloud Data Management Benchmark Survey Report.

There is no debate. Cloud is the platform of today—and tomorrow. If you're not in the cloud, or not planning to be in the cloud, chances are, you will not survive. Most experts predict that, by 2025, 50% of all globally stored data will be in the cloud.

We are in the throes of a digital transformation. Nearly everything we do, how we do business, socialize, exercise, shop, even order food, is happening online. With this transformation comes a huge increase in the amount of data being introduced into our information ecosystem. Where is all this data being stored? In the cloud. Cloud offers one of the greatest opportunities to enable our data, to democratize our data, and to realize value from our data, but it also presents us with the challenge of properly managing this data. How do we ensure that the information floating in the cloud is being curated properly, protected, and used ethically?

To address these questions and ascertain where the industry stands in its cloud planning and migration, the EDM Council conducted the 2023 EDMC Cloud Data Management Benchmark Survey. The survey is based on EDM Council's newly published Cloud Data Management Capabilities

Framework (CDMC™). The CDMC is the product of a collective 18-month effort, involving more than 100 companies and 300 industry professionals and subject matter experts, who took on the task of codifying the best practices for migrating and managing data in the cloud.

Cloud is in mass adoption

The transition from on-premises data management to cloud computing is well under way, with 92% of respondents indicating that their organizations are engaged in cloud implementation. Many of these initiatives are enterprise-wide, indicating an understanding of the significant impact of cloud technology throughout the organization. There's a diversity of cloud strategies—cloud, multi-cloud, and hybrid cloud—with a slight majority favoring a hybrid-cloud approach.

Cloud data management capabilities, however, are still in the early stages of implementation. Most capabilities are in the 'Developmental' or 'Defined' stages, with few, if any, reaching the 'Achieved' or 'Enhanced' stages of maturity. Progress is being made, but there is still significant room for growth.

Areas in need of concerted attention include the management of sensitive data in the cloud, the management of data sovereignty and cross-border movement, and the stewardship of data to ensure ethical access, use, and outcomes of data in the cloud. Eighty-

four percent of respondents have not implemented "ethical access, use, and outcome" policies and procedures when executing their AI/analytics agendas in the cloud.

Need for automated cloud controls

Automated controls for managing data across jurisdictions in the cloud are not prevalent, with 55% of respondents indicating their organizations do not have such controls in place. Lack of such controls presents significant risk and could lead to breaches of data privacy laws in various jurisdictions and hefty penalties.

A significant proportion of data management activities related to operational controls also lack automation. Eighty-eight percent of respondents say that they lack automated catalog and classification processes, which are fundamental to data risk lifecycle management. Manual data management processes increase the risk of human error and reduce the effectiveness of data management. Areas such as data control compliance, ownership field population, and data protection impact assessments show significant room for improvement through automation.

Strong demand for cloud data management frameworks

A large percentage of respondents (98%) agreed that it would be helpful to have a framework to manage data

throughout its lifecycle in the cloud. It's recommended that organizations adopt or adapt existing frameworks, such as the EDM Council's Cloud Data Management Capabilities (CDMC) framework, ISO standards, and the NIST guidelines, according to their needs.

Now is the time to act

Invest in data management training and skills development

Organizations need to develop and refine their cloud data management capabilities. As firms aspire to be data driven and to manage data effectively in the cloud, organizations need a workforce competent in both technical and governance aspects of data management. Investing in training and skills development can help build this competency and ensure that all data professionals, regardless of their role, understand the importance of effective data management and their responsibilities in this area.

Conduct cloud data management assessments

An important part of embracing a new journey is understanding the baseline from which you're starting. Seventy-one percent of survey respondents stated that they have not conducted a formal cloud assessment. Assessments provide an awareness of where capability gaps exist and what areas need attention.

It is important to establish a baseline, then assess on a regular basis, to ensure forward-looking and sustainable progress. Assessments should focus not only on technical capabilities, but also on governance structures, compliance, and overall strategic alignment to help identify gaps, prioritize areas for improvement, and track progress over time. Particular attention should be paid to areas like data sovereignty and cross-border data movement, as well as ethical data access, use, and outcomes.

Develop and implement data management policies and procedures

The sensitivity of data and the potential legal implications of its misuse underscore the need for organizations to develop comprehensive data management policies and procedures. These should encompass all aspects of data management (from collection and storage to use and disposition), be communicated widely, and enforced across the organization.

Establish cross-functional data management teams and get them trained

Given the cross-cutting impact of data management, it's beneficial to establish cross-functional teams to oversee and coordinate data management activities. These teams should include representatives from various

departments and roles, with support from the top, to ensure an holistic approach to data management in the cloud, build a strong data management culture, and support the growing inventory and end-user demand for data.

The journey is just beginning

Overall, while the survey shows that cloud initiatives are widespread and the importance of effective data management in the cloud is recognized, it also highlights significant gaps in maturity and implementation. By focusing on these areas, organizations may ensure that they are maximizing the benefits of cloud technology while also managing the associated risks and challenges effectively.

The CDMC Framework is a great place to start. To learn more about the CDMC Framework, and download your free copy, visit <https://edmcouncil.org/frameworks/cdmc/> and begin your own journey...

Sincerely,



John A Bottega

President, EDM Council

July 2023

Table of Contents

Executive Summary	2
Table of Contents	5
Survey Questions	6
CDMC™ – The Cloud Data Management Capabilities Framework	7
How Is CDMC Scored?	9
Crossing the Capability Chasm	10
Demographics & Profile of Participants	11
Scope of Operations	12
Cloud Data Management Drivers & Priorities	13
Methodology	16
Survey Results & Observations	17
1.0 Governance & Accountability	17
2.0 Cataloging & Classification	23
3.0 Accessibility & Use	27
4.0 Protection & Privacy	31
5.0 Data Lifecycle	35
6.0 Data & Technical Architecture	40
CDMC™ Composite Scores	44
2023 Benchmark Advisory Team	48

Note: The 2023 EDM Council Cloud Data Management Benchmark Report represents the opinion of the EDM Council, not individual members nor organizations. This analysis is intended for informational purposes only.

© 2023 EDM Council. All rights reserved. EDM Council and CDMC are registered trademarks of EDM Council. All other marks are the property of their respective owners.

Survey Questions

1.0

Governance & Accountability

- 1.1 Cloud data management business cases are defined and governed.
- 1.2 Data ownership is established for both migrated and cloud-generated data.
- 1.3 Data sourcing and consumption are governed and supported by automation.
- 1.4 Data sovereignty and cross-border data movement are managed.

2.0

Cataloging & Classification

- 2.1 Data catalogs are implemented, used, and interoperable.
- 2.2 Data classifications are defined and used.

3.0

Accessibility & Use

- 3.1 Data entitlements are managed, enforced, and tracked.
- 3.2 Ethical access, use, and outcomes of data are managed.

4.0

Protection & Privacy

- 4.1 Data are secured, and controls are evidenced.
- 4.2 A data privacy framework is defined and operational.

5.0

Data Lifecycle

- 5.1 The data lifecycle is planned and managed.
- 5.2 Data quality is managed.

6.0

Data & Technical Architecture

- 6.1 Technical design principles are established and applied.
- 6.2 Data provenance and lineage are understood.



CDMC™
**The Cloud Data
Management
Capabilities
Framework**

CDMC™ – The Cloud Data Management Capabilities Framework



The CDMC framework outlines best practices and capabilities that help organizations ensure seamless cloud migration, effective data protection, and robust data management in the cloud.



The EDM Council's Cloud Data Management Capabilities (CDMC™) framework is a comprehensive set of best practices and standards designed to guide organizations in the migration, management, and protection of sensitive data within cloud environments.

Developed over an 18-month period, the CDMC framework is the product

of a collaborative effort involving more than 300 professionals from 100 firms, including leading cross-industry firms, consultancies, technology companies, and major cloud service providers Amazon Web Services (AWS), Google, IBM, and Microsoft.

The CDMC framework outlines best practices and capabilities that help organizations ensure seamless cloud migration, effective data protection, and robust data management in the cloud. This includes requirements for governance and accountability, cataloging and classification, and accessibility and use of sensitive data.

For example, compliance must be monitored for all data assets containing sensitive data through metrics and automated notifications. Also, cataloging and classification must be automated for all data at the point of creation or ingestion, with consistency across all environments. Additionally, entitlements and access for sensitive data must default to the creator and owner, and access must be tracked for all sensitive data.

Organizations may use the CDMC framework to assess their maturity in implementing cloud data management programs, and it forms the basis of the EDM Council's Cloud Data Management Benchmark Survey. The insights gained from this survey and subsequent benchmark analysis assist organizations in their journey toward becoming data-driven and optimizing the value of their data assets. ■

How Is CDMC Scored?

CDMC is structured to define and measure data management and analytics capability. The singular goal of CDMC is the achievement of the capabilities necessary to develop, implement, and sustain an effective cloud data management program.

CDMC is scored on a 1 through 6 scale, describing the current state of each CDMC capability—from 'Not Initiated' to 'Enhanced'.

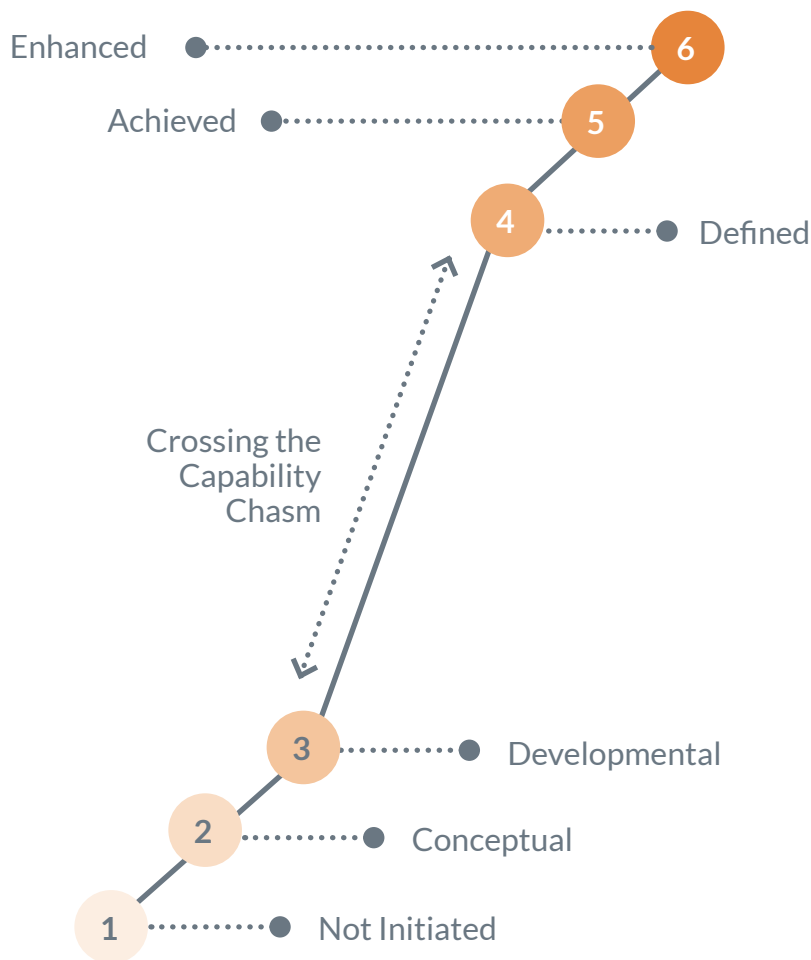
Each requirement clearly defines the rationale for the specified capability

and the dependencies among the components of cloud data management into a cohesive program. CDMC is linked to a scoring matrix developed by the EDM Council to evaluate achievement of capability from three critical dimensions:

 Engagement Ensure that the right people with the appropriate levels of authority are participating in the cloud data management program.	 Process Measure the degree to which cloud data management processes are established, structured, and repeatable.	 Evidence Produce business artifacts necessary to audit against each capability statement.
---	---	--

Score	Category	Description
1	Not Initiated	Ad hoc data management (performed by heroes)
2	Conceptual	Initial planning activities (white board sessions)
3	Developmental	Engagement under way (stakeholders being recruited and initial discussions about roles, responsibilities, standards, and processes)
4	Defined	Data management capabilities established and verified by stakeholders (roles and responsibilities structured, policy and standards implemented, glossaries and identifiers established, sustainable funding sourced)
5	Achieved	Data management capabilities adopted and compliance enforced (sanctioned by executive management, activity coordinated, adherence audited, strategic funding)
6	Enhanced	Data management capabilities fully integrated into operations (continuous improvement)

Crossing the Capability Chasm




Since its inception, the EDM Council has observed and continues to observe how organizations develop and mature their data management programs.

As the diagram at left depicts, once a program is initiated, firms will move quickly to the conceptual phase. In this phase, the program is described, objectives are identified, and management, business, and technology agree on a vision and approach.

As program development begins, firms must 'jump the chasm' from developmental to defined, as stakeholders are recruited, and the real work of implementation takes shape. In this phase, business and data architecture are defined, technology is aligned, and data governance is deployed.

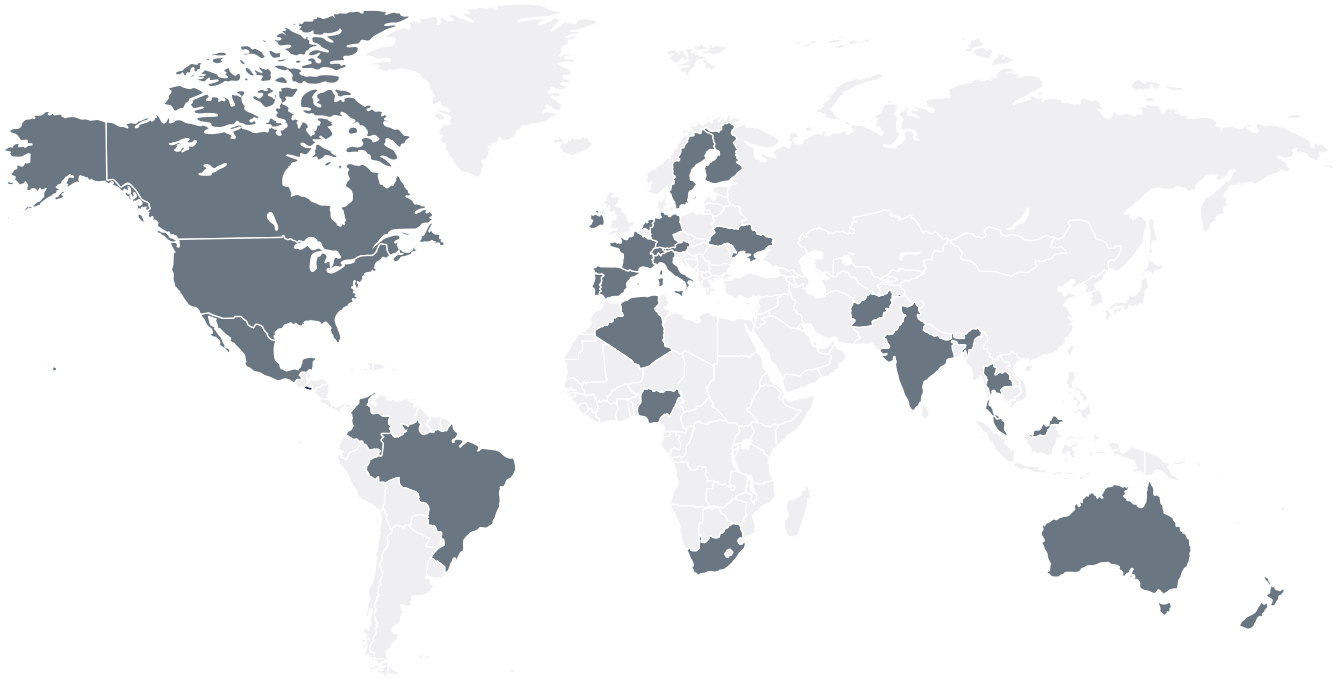
It is not uncommon to see organizations hover at the developmental stage for a while. To progress, organizations must commit to advancing their programs, continue to garner support from management and the business, and focus their efforts on achieving the objectives of the business and industry.

Once all of these capabilities are defined and supported, organizations move toward achievement and enhancement of capabilities.

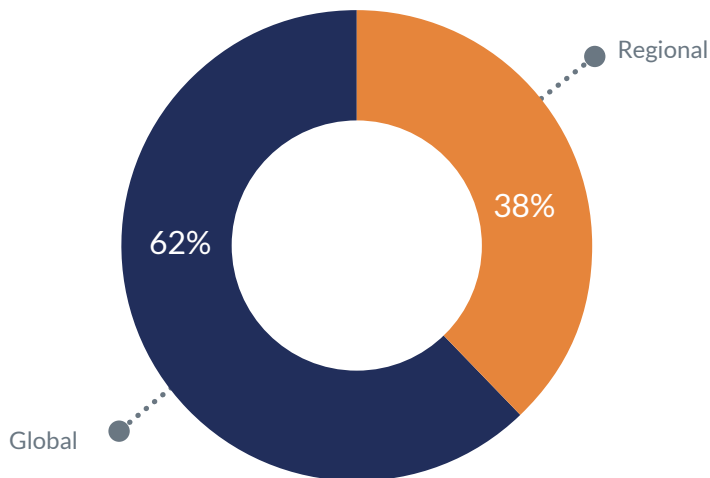
An aerial photograph of a river delta, likely the Amazon, showing a large central island and a complex network of channels and distributaries. The water is a vibrant turquoise color, and the land is a mix of green and brown. The text "Demographics & Profile of Participants" is overlaid in white, bold font on the lower left side of the image.

Demographics & Profile of Participants

The 2023 Cloud Data Management Benchmark Survey comprises responses from more than 250 data professionals in more than 30 countries across the globe.



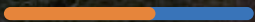
Scope of Operations



62%
of respondents report their organization's primary area of operations is global.

An aerial photograph of a road intersection. A road curves from the top left towards the bottom left, meeting a straight road that runs vertically through the center. A large white triangular road marking is painted on the asphalt at the intersection. The surrounding area includes green grass and some dry vegetation. The text 'Cloud Data Management Drivers & Priorities' is overlaid in white on the lower-left portion of the image.

Cloud Data Management Drivers & Priorities



Cloud Data Management Drivers & Priorities



Across all categories, automation levels are relatively low.

Since 2015, insights from the EDM Council's Global Benchmark Surveys have helped shape the programming of the EDM Council. We learn about the issues facing data management professionals and facilitate Special Interest Groups on those topics to gather best practices and produce guidance and tools to aid in the development, implementation, and evolution of a successful data management program.

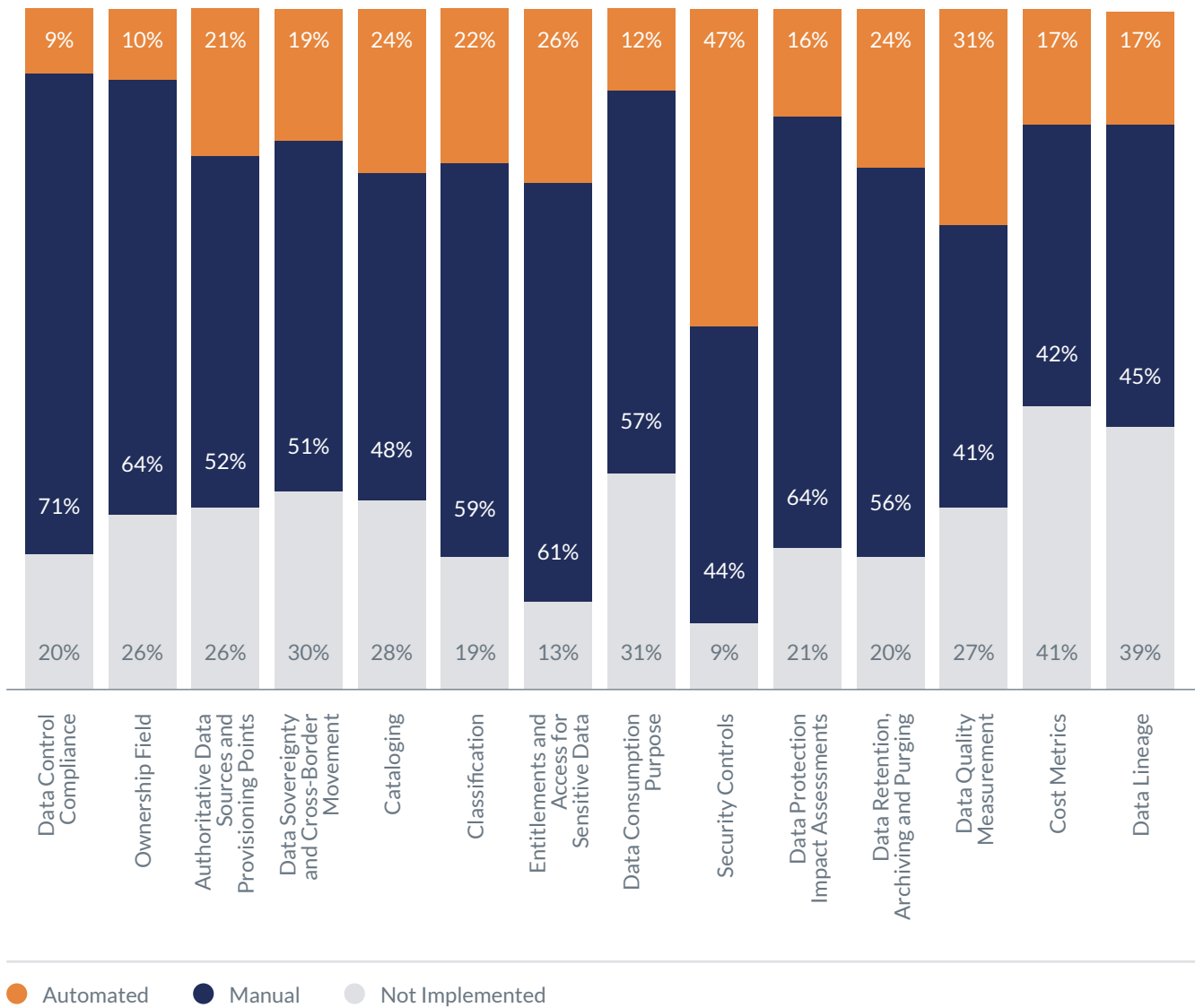
We asked participants to indicate the status of the cloud data management controls that are operational in their organizations, using the 14 key controls of the CDMC™ framework.

Low Levels of Automation Across all categories, automation levels are relatively low, with the highest being

'Security Controls' at 47%. Most organizations rely heavily on manual processes, which are more labor-intensive and prone to human error. This suggests there's significant room for improvement through automation.

Lack of Implementation 'Cost Metrics' and 'Data Lineage' have the highest percentage of non-implementation, at 41% and 39%, respectively. These areas are essential for budgeting and data tracking. Their non-implementation may lead to financial inefficiencies and difficulty tracing data origins and transformations.

Heavy Reliance on Manual Controls 'Data Control Compliance,' 'Ownership Field,' and 'Data Protection Impact Assessments' show high manual process dependencies, each exceeding 60%. These controls are critical for regulatory compliance and data ownership clarity, and manual processes may increase risks related to non-compliance and data mismanagement.



Based on these insights, EDM Council makes the following recommendations:

Invest in Automation Organizations should prioritize automating their cloud data management controls. This could involve investing in modern data management solutions or upskilling existing staff to implement more automated processes.


Implement Missing Controls

Organizations should focus on implementing controls that are currently absent, especially for ‘Cost Metrics’ and ‘Data Lineage.’ Without these controls, it’s challenging to manage costs effectively or understand data’s journey throughout its lifecycle.

Improve Manual Processes For controls still dependent on manual processes, organizations should aim to

improve these workflows. This might involve further training, streamlining of processes, or the introduction of semi-automated controls to reduce error rates.

Prioritize Security Controls Given the relatively high automation in ‘Security Controls,’ organizations appear to be prioritizing security, which is crucial. However, with 9% of respondents reporting non-implementation, there’s still room for improvement. ■

A high-angle, wide shot of a modern library. The space is filled with white bookshelves on multiple levels, connected by a light-colored metal railing. The shelves are filled with books. In the foreground, there are several blue modular sofas and a small table. The lighting is bright and even, creating a clean and organized atmosphere.

An original web-based survey informed by evidence from past literature and validated scales was developed. An email campaign was used to disseminate an online Qualtrics survey link to EDM Council membership, partners, webinar attendees, and a list of key data-related industry participants between April 18 and May 18, 2023.

The next section of this report contains tables and charts describing the 'Average CDMC' score and 'Achieved / Enhanced'

percentages for each of the survey questions pertaining to maturity level. The individual CDMC scores provided by the respondents were combined into an Average CDMC Score. The percentage of responses with scores at the highest maturity levels for each question determined the percentage 'Achieved / Enhanced.'

These aggregate scores and individual score distributions are used as a baseline for comparison to future iterations of the survey. ■

Methodology



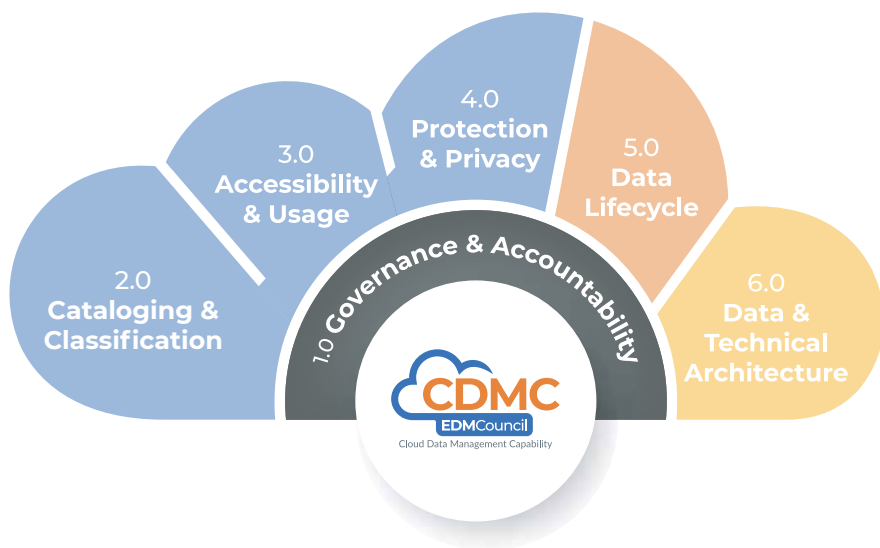
Survey Results & Observations

1.0 Governance & Accountability

1.0 Governance & Accountability



The Governance & Accountability component is a set of capabilities that ensures an organization has clear accountability, controls, and governance for data migrated to or created in cloud environments.



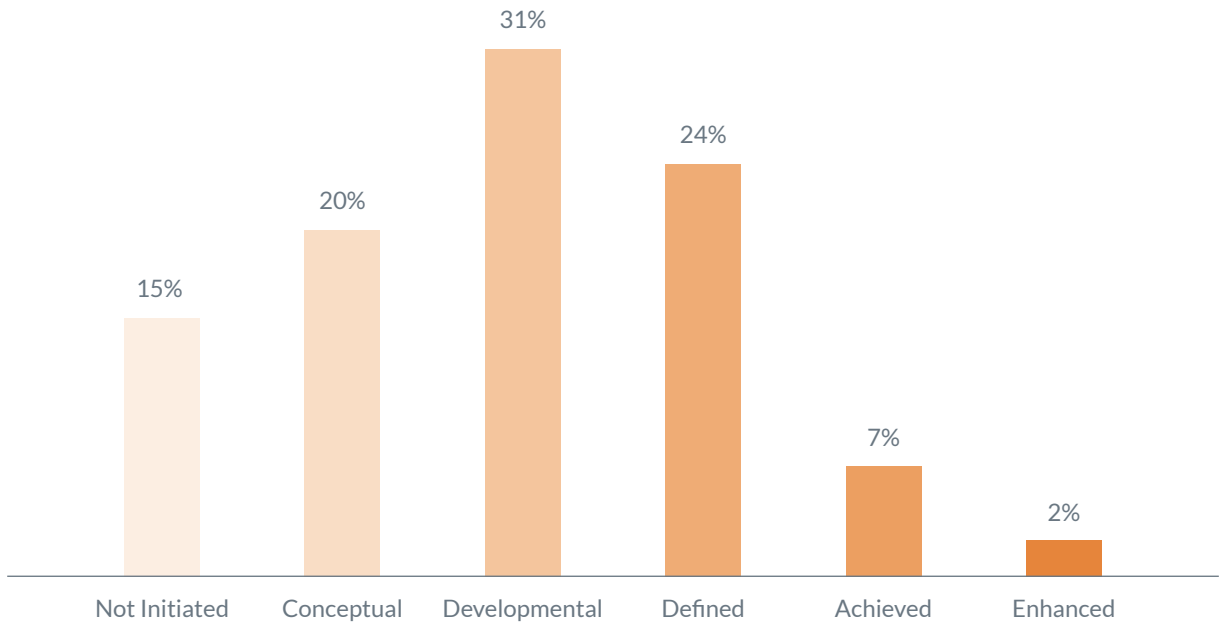
Governance and accountability are the backbones of the successful management of data in cloud environments. The cloud environment introduces challenges and opportunities for scale, standardization, automation, and the shared responsibility model.

Consequently, it is important to apply an effective data governance program to data that resides in a cloud environment. All stakeholders should have a clear understanding of data controls and accountability for each role. The approach is similar to how data governance, controls, and accountability are applied to conventional data management in an organization.

The Governance & Accountability component is a set of capabilities that ensures an organization has clear accountability, controls, and governance for data migrated to or created in cloud environments. These capabilities provide the foundation of well-governed business cases, effective data ownership, governance of data sourcing and consumption, and management of data sovereignty and cross-border data movement risks.

Organizations that establish strong governance and data management controls over data residing in cloud applications have an opportunity to realize all of the benefits of a cloud implementation while managing the associated risks. Data governance and accountability in cloud environments help define effective business case processes, identify accountable stakeholders and data owners, ensure the proper management of data sourcing, and provide proper tracking and control of data movement concerning data sovereignty guidelines.

1.1 Cloud data management business cases are defined and governed. Data control compliance must be monitored for all data assets containing sensitive data via metrics and automated notifications.



2.94
Average Score

9%
Achieved / Enhanced

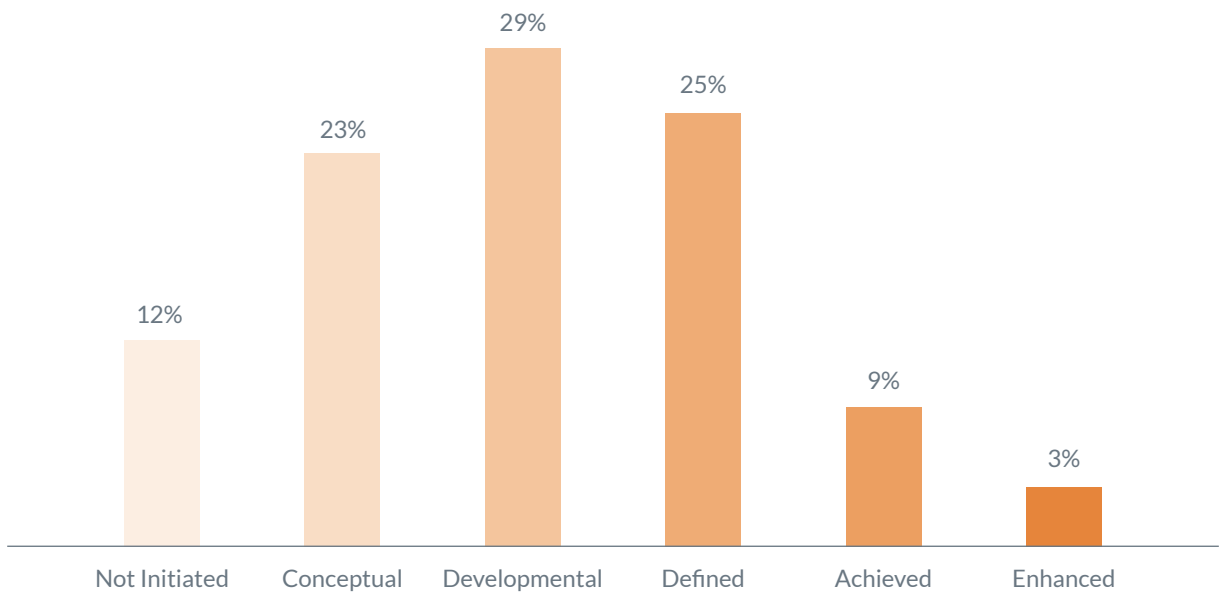
Analysis

Most organizations (31%) are still in the developmental stage, indicating that organizations are working toward developing clear business cases for cloud data management. A substantial portion (15%) have not initiated this process, highlighting a potential blind spot in data management strategy.

Organizations that establish strong governance and data management

controls over data residing in cloud applications have an opportunity to realize all of the benefits of a cloud implementation while managing the associated risks. Defining and governing cloud data management business cases are instrumental to this aim, yet respondents of the survey indicate that their organizations have significant work to do to achieve maturity on this subcapability.

1.2 Data ownership is established for both migrated and cloud-generated data. The ownership field in a data catalog must be populated for all sensitive data or otherwise reported to a defined workflow.



3.05
Average Score

12%
Achieved / Enhanced

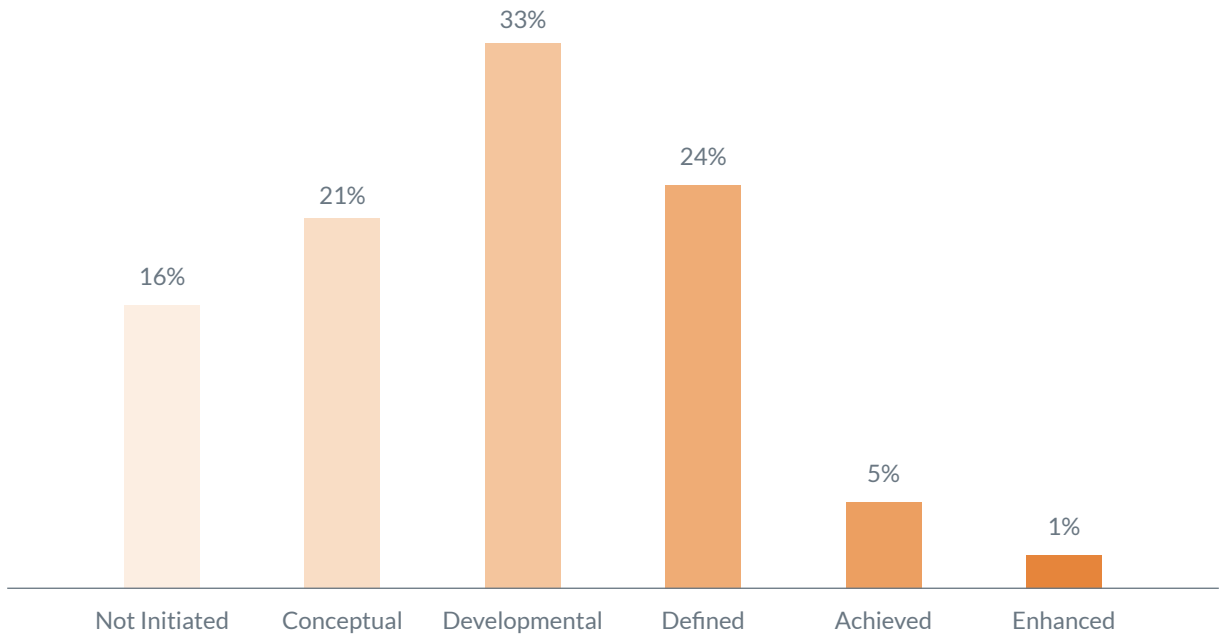
Analysis

Similar to 1.1, most organizations (29%) are in the developmental stage in establishing data ownership. However, 12% have not initiated this process, suggesting that these organizations need to prioritize establishing clear data ownership.

Identifying and assigning ownership for data that resides in a cloud

environment should follow the same guidelines for on-premises data ownership. Ownership of all data elements in any data domain within a cloud environment is mandatory and specified by data management policy and standards. Most organizations are in the early stages of this subcapability.

1.3 Data sourcing and consumption are governed and supported by automation. A register of authoritative data sources and provisioning points must be populated for all data assets containing sensitive data or otherwise must be reported to a defined workflow.



2.84
Average Score

6%
Achieved / Enhanced

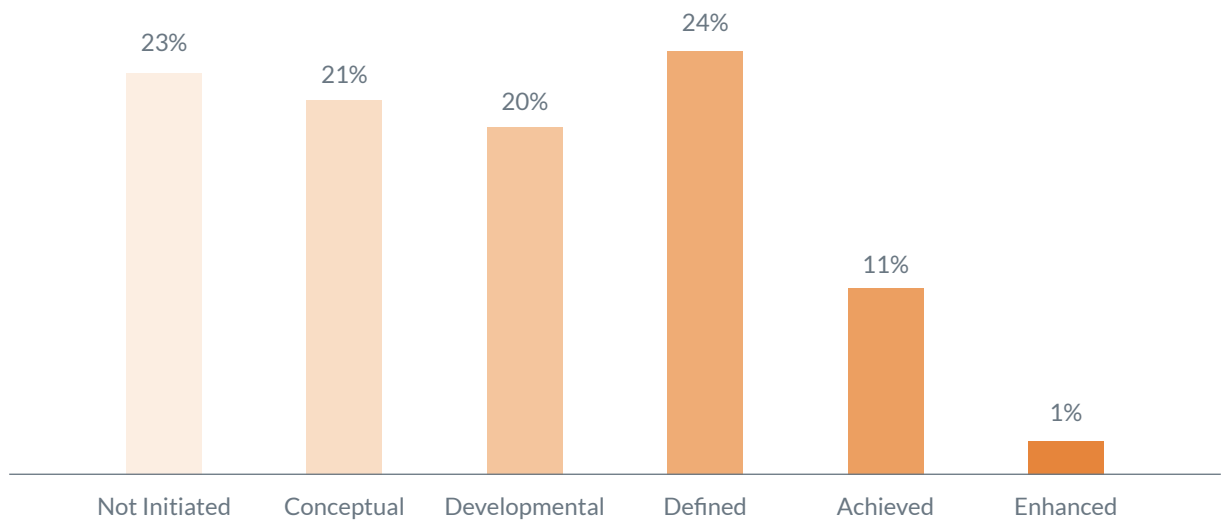
Analysis

Again, most organizations are in the developmental stage (33%), but a significant portion have not initiated this process (16%). It signals a potential risk of non-compliance with regulations and standards.

The organization must ensure that data is consumed from authoritative sources or authorized distributors,

with data governance that manages the designation of this authority. Cloud platforms must provide automation to enforce consumption from authoritative sources and authorized distributors or highlight consumption from non-authoritative sources. Most organizations are in the early stages of this subcapability.

1.4 Data sovereignty and cross-border data movement are managed. The data sovereignty and cross-border movement of sensitive data must be recorded, auditable, and controlled according to defined policy.



2.82
Average Score

12%
Achieved / Enhanced

Analysis

This seems to be a challenging area for many organizations, with a significant percentage (23%) not having initiated it at all.

The sovereignty of data in cloud environments must be tracked. This information must be used to ensure

that the storage and cross-border movement and use of data conform to the relevant jurisdictional requirements. Most organizations are in the early stages of this subcapability.



Survey Results & Observations

2.0 Cataloging & Classification

2.0 Cataloging & Classification

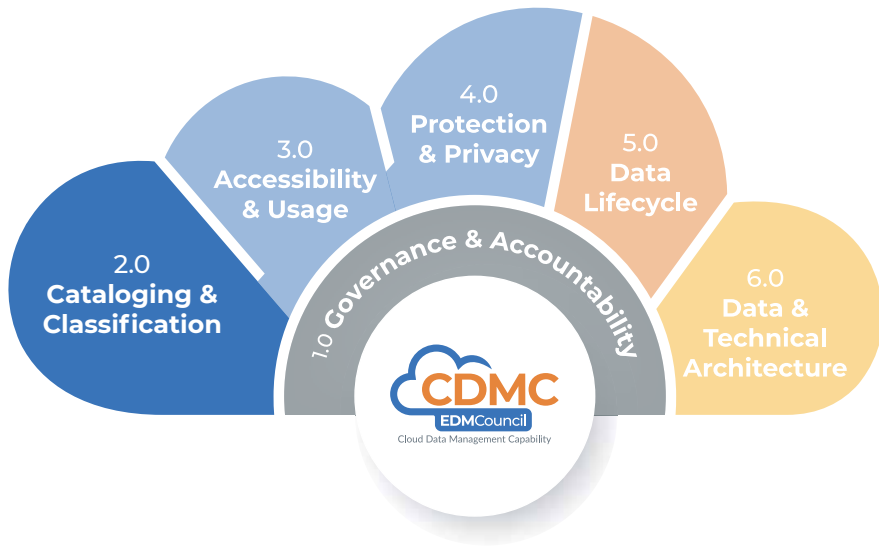


Organizations that create, maintain, and share comprehensive data catalogs gain the ability to maximize controlled reuse of data assets.

business definitions, classifications, sourcing, retention, physical location, and ownership details. Together, these characteristics comprise the data catalog.

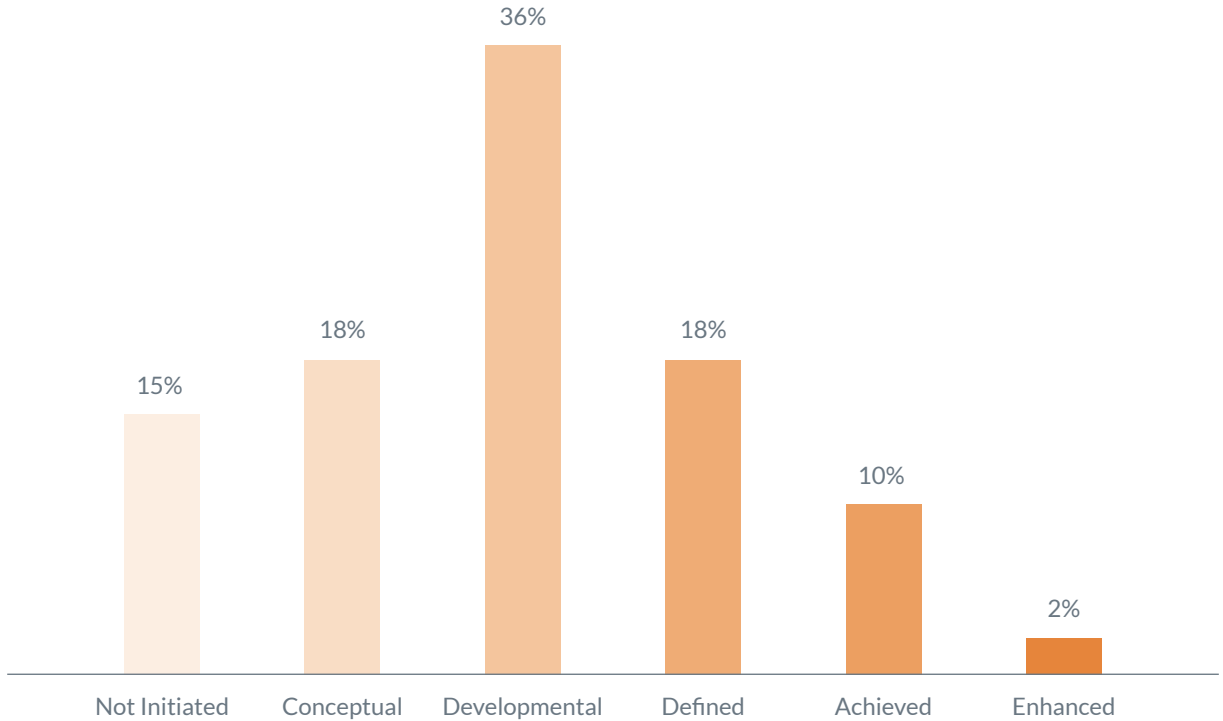
The Data Cataloging & Classification component is a set of capabilities for creating, maintaining, and using data catalogs that are both comprehensive and consistent. This component includes classifications for information sensitivity. These capabilities ensure that data managed in cloud environments are easily discoverable, readily understandable, and support well-controlled, efficient data use and reuse.

Organizations that create, maintain, and share comprehensive data catalogs gain the ability to maximize controlled reuse of data assets. Organizations that effectively support the information sensitivity classifications may benefit from enhancements in transparency and consistent treatment of data classifications. Maximum transparency on the precise locations of data storage and data transfer routes will enable automatic processes to manage, monitor, and enforce the consistent data treatment according to a specific information sensitivity classification. Standardizing an information sensitivity classification functionality also enables an authorized automatic process to manage, monitor, and enforce security and regulatory compliance across multiple jurisdictions. In some instances, information sensitivity classification applies to additional classification metadata.



Effective cloud data management depends on having full control of all data assets. This understanding must include technical characteristics such as formats and data types and contextual information supporting the full CDMC capabilities. These capabilities include

2.1 Data catalogs are implemented, used, and interoperable. Cataloging must be automated for all data at the point of creation or ingestion, with consistency across all environments.



2.96
Average Score

12%
Achieved / Enhanced

Analysis

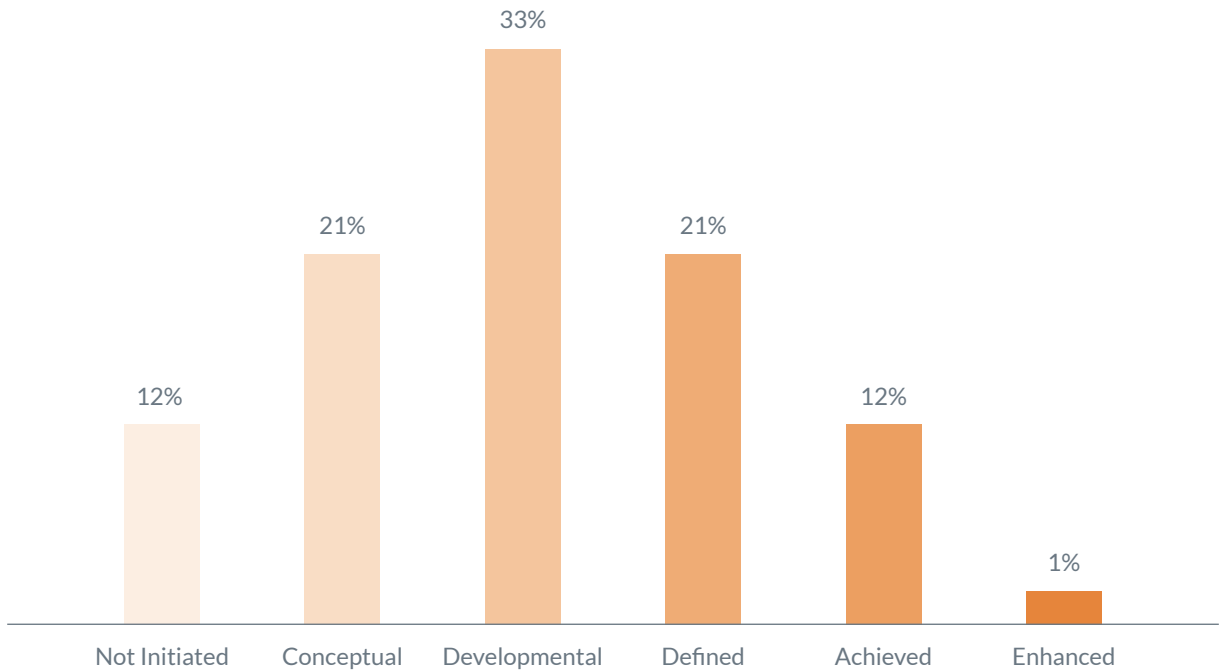
Most organizations are in the developmental stage (36%), showing an increased focus on creating and maintaining data catalogs, but a lot more effort is required for full implementation and enhancement.

Data catalogs describe an organization's data as metadata for documentation, discovery, and understanding. The data cataloging

scope and approach must be defined. Catalogs must be implemented and populated with the metadata that describes the data. This metadata must be exposed to both users and applications, and standards should be defined and adopted to ensure that metadata may be exchanged between catalogs on different platforms. Most organizations are in the early stages of this subcapability.

2.2 Data classifications are defined and used. Classification must be automated for all data at the point of creation or ingestion and must be always on.

- Personally identifiable information auto-discovery
- Information sensitivity classification auto-discovery
- Material non-public information (MNPI) auto-discovery
- Client identifiable information auto-discovery
- Organization-defined classification auto-discovery



Analysis

A significant proportion of organizations are in the developmental stage (33%), showing efforts toward automating data classifications, yet there's still a lot to be done.

From the moment it is created, data can be both liabilities and assets. Poorly managed data are likely to pose a risk if used inappropriately or unauthorized users access it. Such risks increase in a cloud environment, and many organizations increase their exposure as they move massive amounts of critical data into the cloud.

An information sensitivity classification is a schema for labeling data elements according to business risk level or value. Data present a business risk if disclosure, unauthorized use, modification, or destruction could affect compliance, financial, operational, reputational, or strategic risk. Information sensitivity specifies how to access, treat, and manage a data element through each stage of its lifecycle. This labeling is essential to security and regulatory compliance in all applications and a growing portion of cloud environments. Most organizations are in the early stages of this subcapability.

3.03
Average Score

13%
Achieved / Enhanced



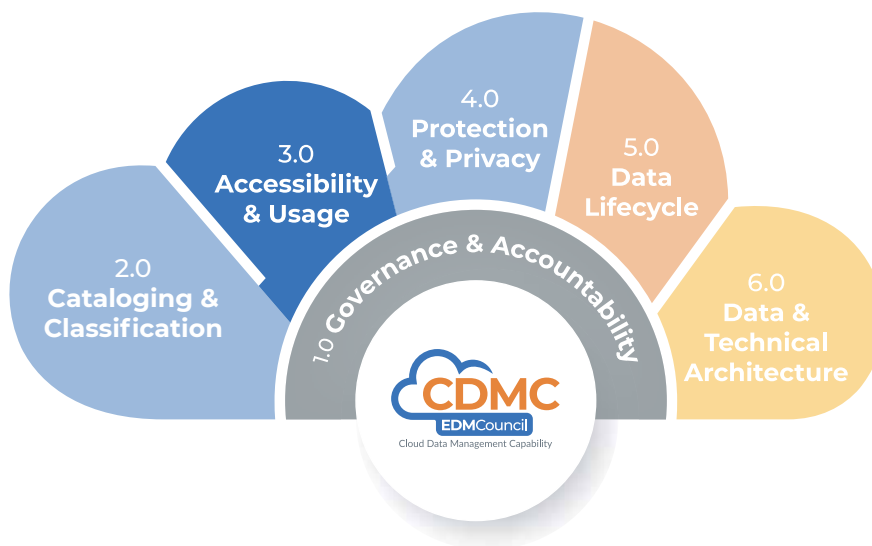
Survey Results & Observations

3.0 Accessibility & Use

3.0 Accessibility & Use



The Accessibility & Use component is a set of capabilities to manage, enforce, and track entitlements and to ensure that data access, use, and outcomes of data operations are done in an appropriate and ethical manner.



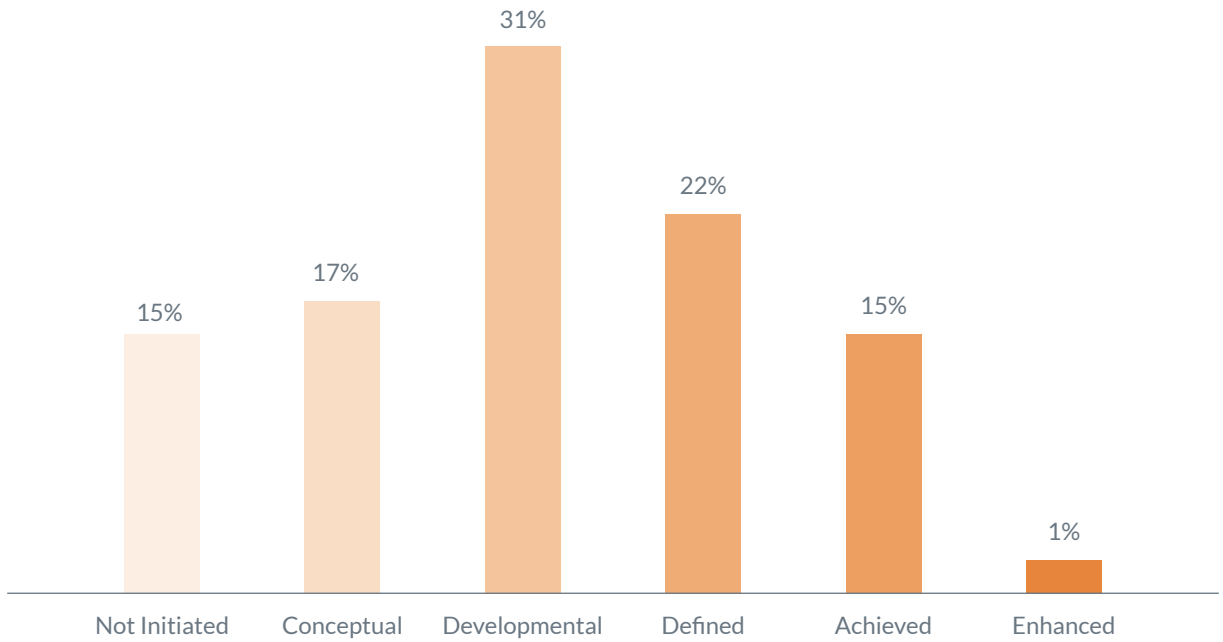
Cloud technology offers significant opportunities for organizations to leverage their data assets in new ways. Many cloud service providers enable machine learning and advanced analytics by many more people than on-premises environments. Cloud computing technologies enable the combination of

data sets in ways that were previously impractical. When an organization seeks to maximize business value by making data and tools widely available, it must ensure that employees may only access the data to which they have proper entitlements in service of the organization.

The Accessibility & Use component is a set of capabilities to manage, enforce, and track entitlements and to ensure that data access, use, and outcomes of data operations are done in an appropriate and ethical manner.

Organizations that implement metadata-driven data access control drive business value by making data readily available for innovative use, while minimizing the legal and reputational risk of unauthorized access. Organizations that achieve a culture of ethical data use and outcomes protect and enhance their businesses by gaining and maintaining customers' trust.

- 3.1 Data entitlements are managed, enforced, and tracked.
 - 1. Entitlements and access for sensitive data must default to creator and owner until explicitly and authoritatively granted.
 - 2. Access must be tracked for all sensitive data.



3.08
Average Score

16%
Achieved / Enhanced

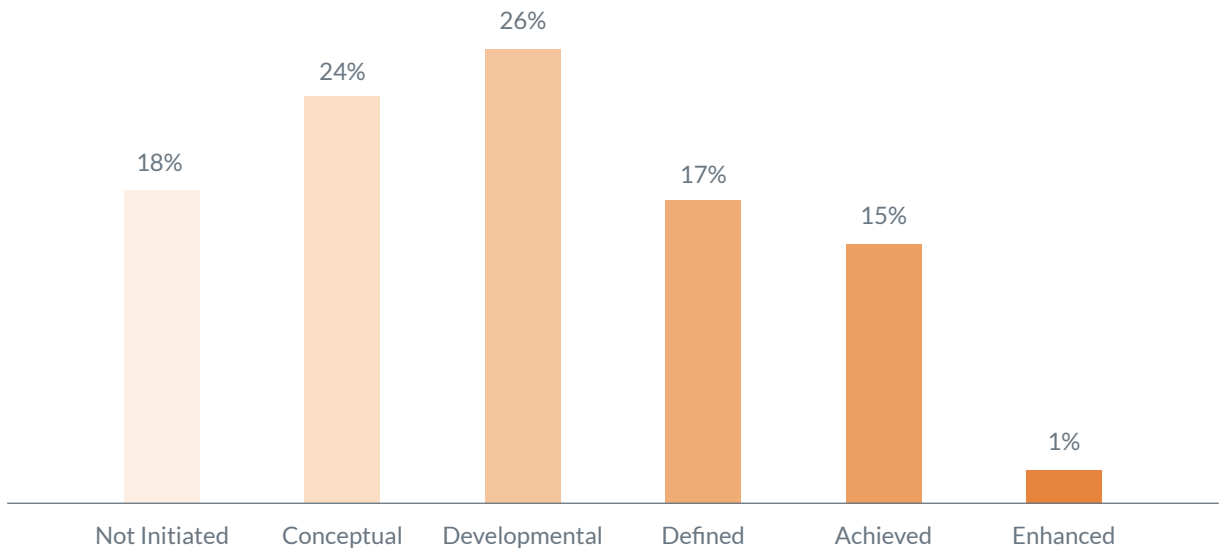
Analysis

Organizations seem to be better at managing data entitlements, with a lower percentage in the not initiated stage (15%) compared to others. However, only 15% have achieved this, suggesting the need for more efforts to enhance data entitlement management.

The organization must capture data assets' entitlement rights and

obligations as metadata and use this information to enforce its policies for accessing and using the data. Enforcement of data entitlements must be evidenced via automated tracking and reporting. Most organizations are at the developmental and defined levels of this subcapability.

3.2 Ethical access, use, and outcomes of data are managed. Data consumption purpose must be provided for all data sharing agreements involving sensitive data. The purpose must specify the type of data required and include country or legal entity scope for complex international organizations.



2.90
Average Score

16%
Achieved / Enhanced

Analysis

This is a relatively challenging area, with 18% not initiating and only 16% achieving this. Organizations should prioritize ethical data management, given its importance in today's business environment.

Organizations must have structures in place to manage the ethical access and use of data, as well as to ensure ethical

outcomes of data use. The organization must establish operational processes to report, review, and address ethical issues arising from data access, use, and outcomes. Most organizations are in the early stages of this subcapability.



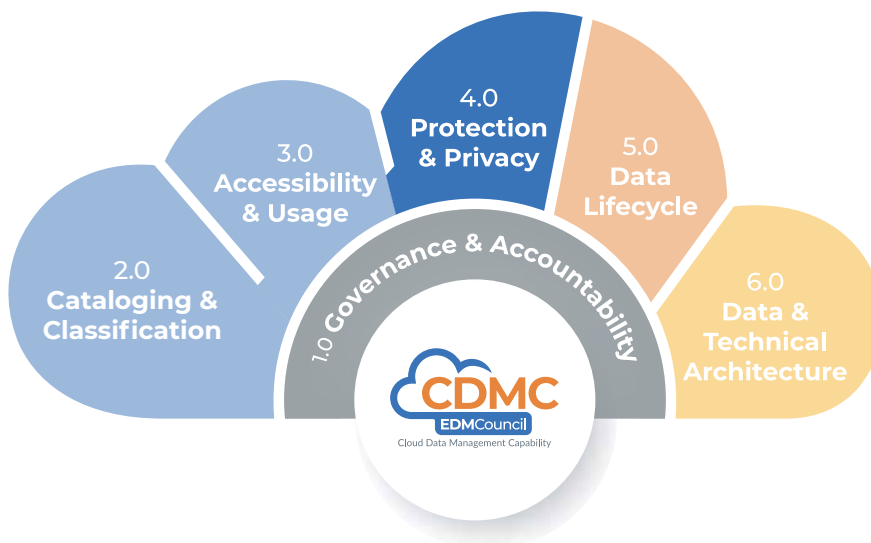
Survey Results & Observations

4.0 Protection & Privacy

4.0 Protection & Privacy



Teams planning integration with a cloud service provider (CSP) must exhibit data protection and privacy capabilities that are sufficient to meet both internal policy mandates and external regulatory requirements.

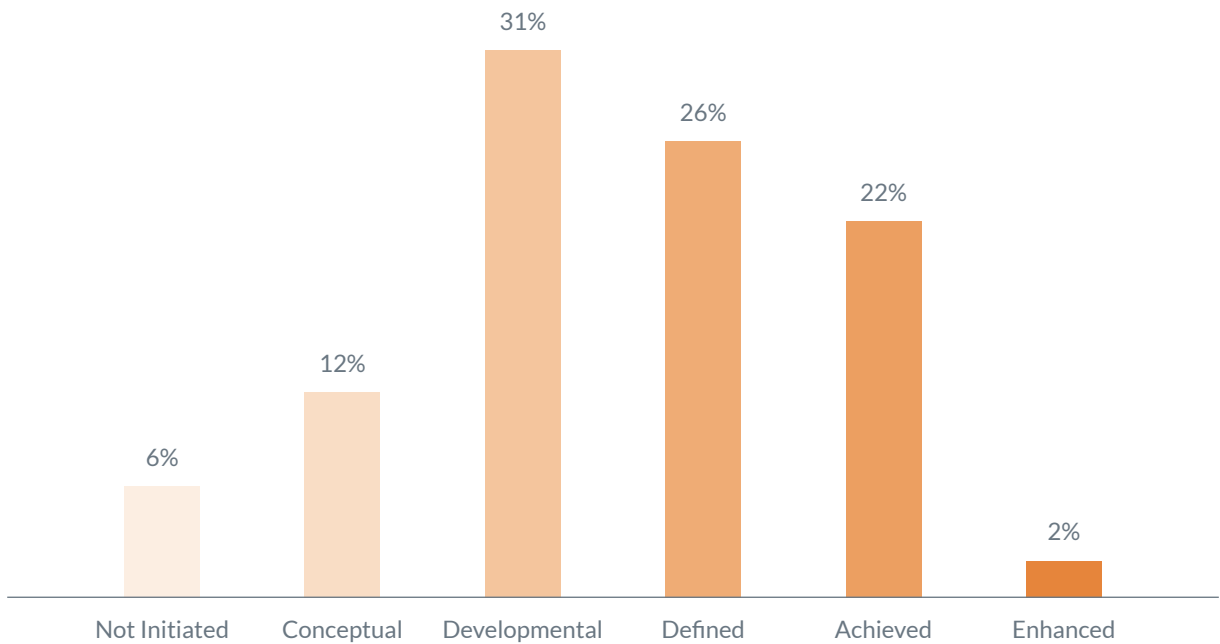


Protecting the content and the privacy of data in the cloud is a critical requirement in today's cloud environments. Organizations that employ cloud computing technology may be required to comply with multiple jurisdictions' data protection and privacy legislation. The compliance burden can be quite heavy

for an organization that is a member of a regulated industry. Teams planning integration with a cloud service provider (CSP) must exhibit data protection and privacy capabilities that are sufficient to meet both internal policy mandates and external regulatory requirements.

The Protection & Privacy component is a set of capabilities for collecting evidence that demonstrates compliance with the organizational policy for data sensitivity and protection. The purpose of these capabilities is to ensure that all sensitive data has adequate protection from compromise or loss as required by regulatory, industry, and ethical obligations.

4.1 Data are secured, and controls are evidenced. 1. Appropriate security controls must be enabled for sensitive data. 2. Security control evidence must be recorded in the data catalog for all sensitive data.



3.53
Average Score

24%
Achieved / Enhanced

Analysis

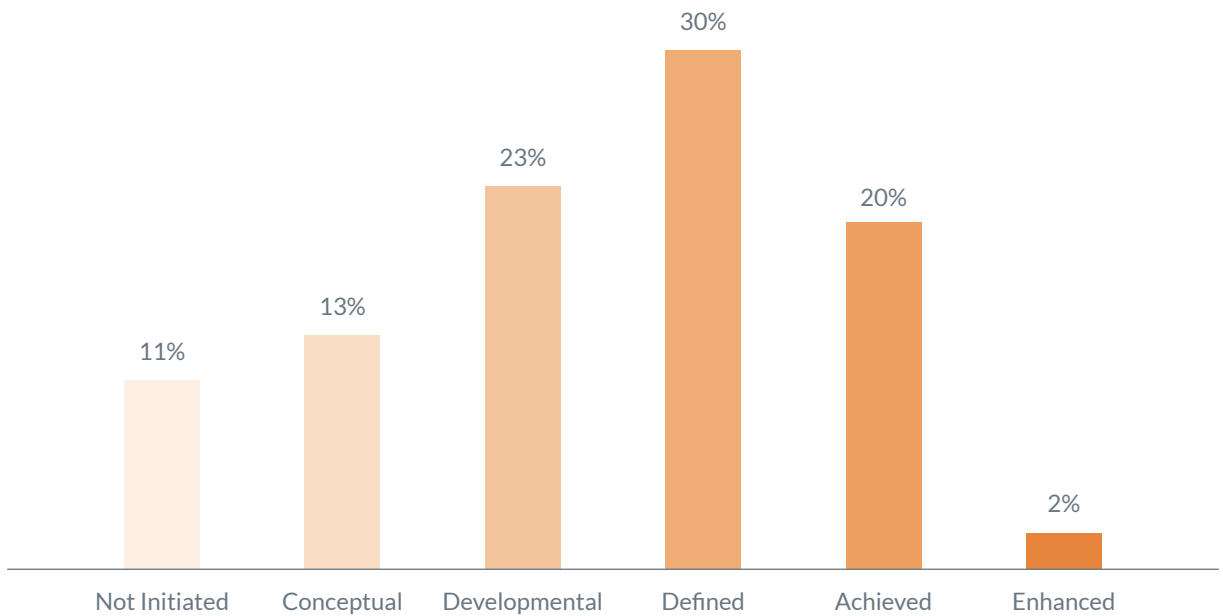
This is one area where organizations seem to have made significant progress, with only 6% not initiated and a substantial 24% having achieved it.

The organization's policy for the encryption of data must be extended to cloud environments. This policy must be enforced for data at rest, in motion, and in use. Evidence of the

implementation of these controls must be captured.

Securing data goes beyond encryption. Techniques for the obfuscation of sensitive data must be supported and adopted in all environments. A data loss prevention regime must be in place and must cover both on-premises and cloud environments.

4.2 A data privacy framework is defined and operational.
Data protection impact assessments (DPIAs) must be triggered automatically for all personal data according to its jurisdiction.



3.41
Average Score

22%
Achieved / Enhanced

Analysis

Similarly, data privacy has seen significant attention, with 22% achieving this and only 11% not initiated.

The organization's data privacy framework must be updated to address

cloud-specific requirements and considerations. Once defined, processes and controls must be established to operationalize the framework.

Survey Results & Observations

5.0 Data Lifecycle



5.0 Data Lifecycle



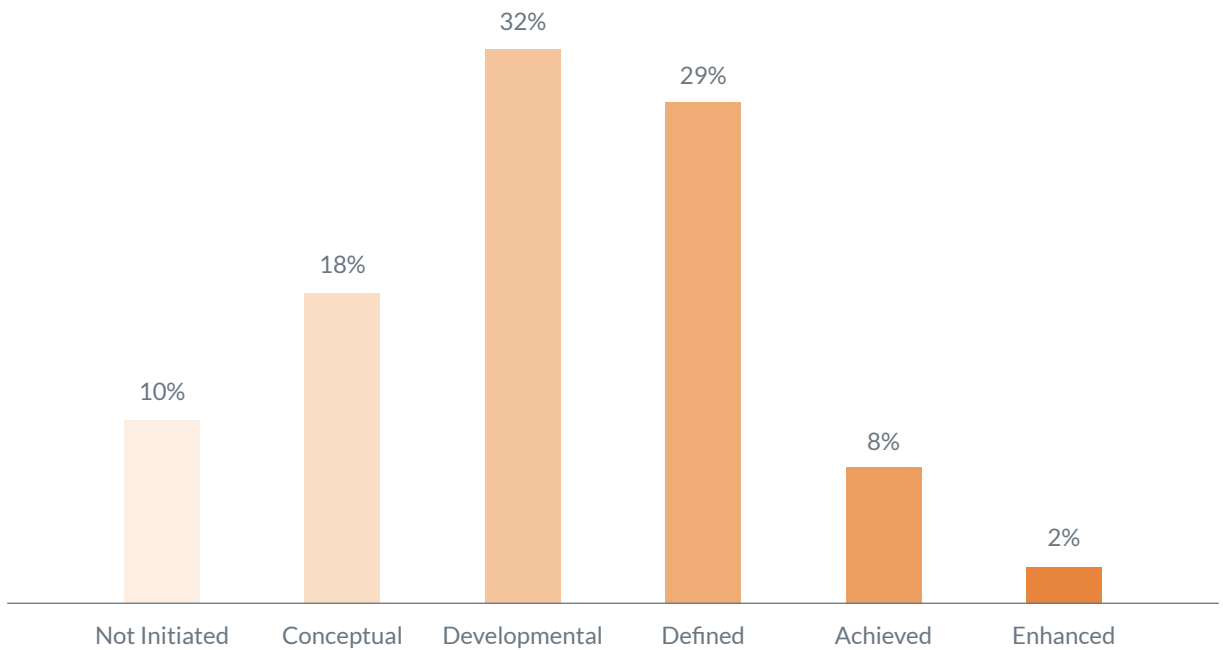
Practitioners must apply proper data management practices and controls across all lifecycle stages to maintain data quality and consistency.



A data lifecycle describes the sequence of stages data traverse including creation, use, consumption, archiving, and destruction. Data may reside in or move through cloud environments at any of these stages. It may be consumed and used at different stages in different environments. Practitioners must apply proper data management practices and controls across all lifecycle stages to maintain data quality and consistency.

The Data Lifecycle component is a set of capabilities for defining and applying a data lifecycle management framework and ensuring that data quality in cloud environments is managed across the data lifecycle.

5.1 The data lifecycle is planned and managed. Data retention, archiving, and purging must be managed according to a defined retention schedule.



3.09
Average Score

10%
Achieved / Enhanced

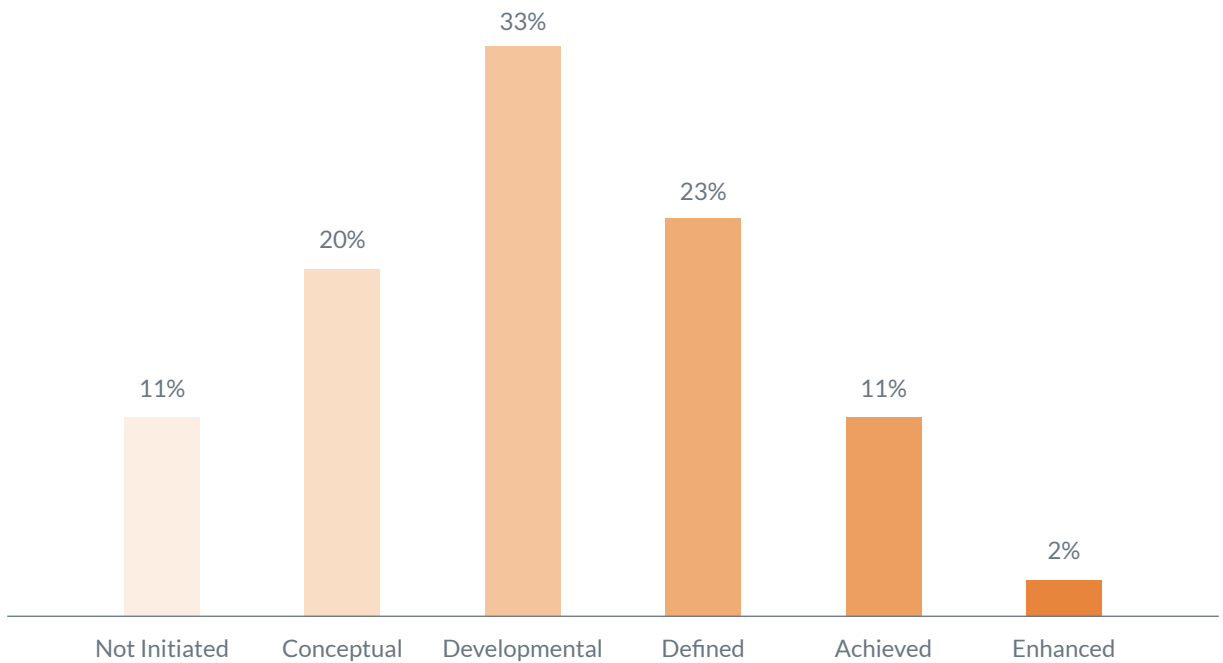
Analysis

Most organizations are either in the developmental or defined stages of managing the data lifecycle, showing awareness but indicating more work is needed.

Effective management of data throughout its lifecycle requires a data

lifecycle management framework to be defined and enshrined in policies, standards, and procedures. The lifecycle then must be managed for all data assets, whether on-premises or in cloud environments. Most organizations are in the middle stages of this subcapability.

5.2 Data quality is managed. Data quality measurement must be enabled for sensitive data with metrics distributed when available.



3.09
Average Score

13%
Achieved / Enhanced

Analysis

Data quality management shows a similar trend to data lifecycle management, highlighting the need for continuous improvement in managing data quality.

Management of data quality starts with the management of data quality rules. These rules then must be deployed and executed to operationalize data quality

measurement. Data quality metrics that result from this measurement must be reported and made available to owners, data producers, and data consumers. Processes must be established to manage the reporting, tracking, and resolution of data quality issues that are identified. Most organizations are in the middle stages of this subcapability.

Survey Results & Observations

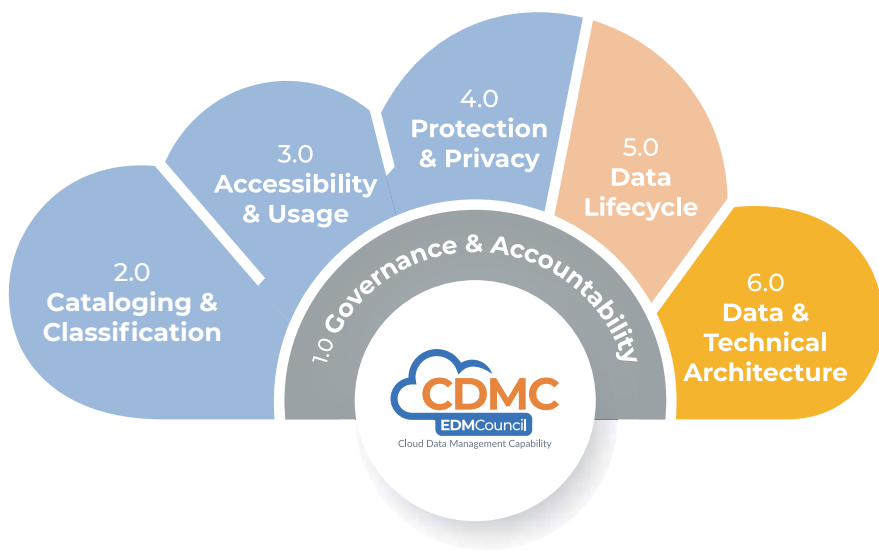
6.0 Data & Technical Architecture



6.0 Data & Technical Architecture



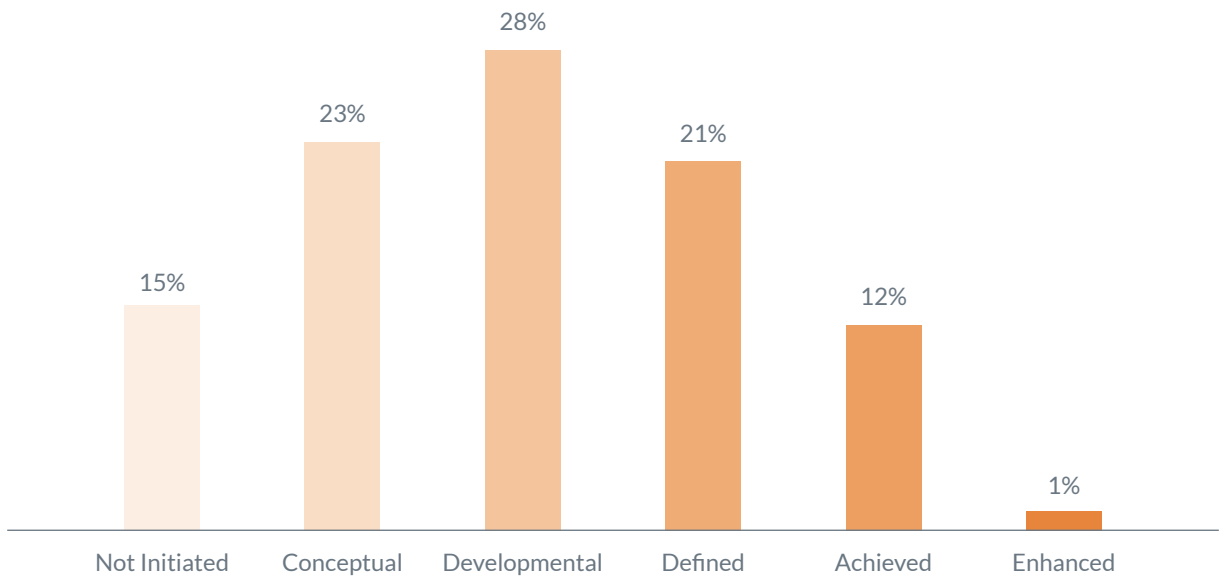
Developing and adopting specific architecture patterns and guidelines may provide a foundation for best practice data management in cloud environments.



Data and technical architecture address architectural issues unique to cloud computing and affect how data are managed in a cloud environment. Most cloud service providers offer many options for the design and implementation of business solutions in cloud environments, with a variety of services and configurations. Developing and adopting specific architecture patterns and guidelines may provide a foundation for best practice data management in cloud environments.

The Data & Technical Architecture component is a set of capabilities for ensuring that data movement into, out of, and within cloud environments is understood and that architectural guidance is provided on key aspects of the design of cloud computing solutions.

6.1 Technical design principles are established and applied. Cost metrics directly associated with data use, storage, and movement must be available in the catalog.



2.95
Average Score

13%
Achieved / Enhanced

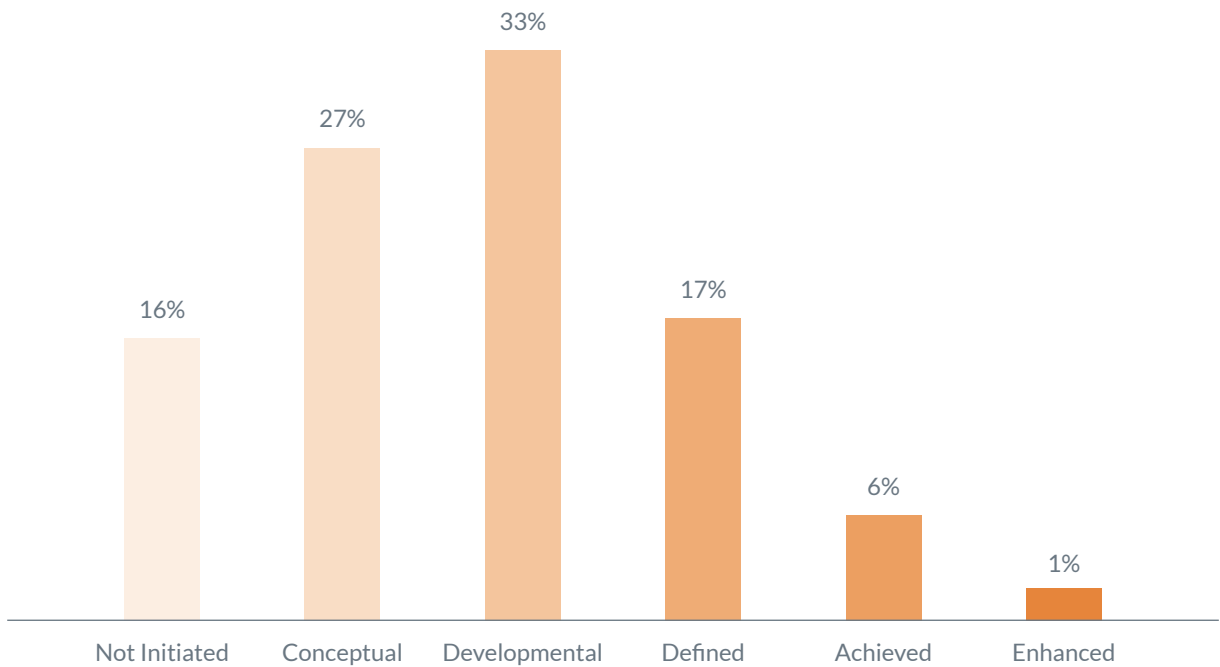
Analysis

Most organizations are in the developmental stage (28%), but a large portion have not initiated this (15%), showing a need for greater focus on technical design principles.

Technical design principles must be established to facilitate the optimization of cloud use and cost efficiency. They must guide the

implementation of solutions that meet availability, resilience, backup, and point-in-time recovery requirements. The ability to exit cloud services must be planned and tested, facilitated by data portability between cloud and on-premises environments. Most organizations are in the early to middle stages of this subcapability.

6.2 Data provenance and lineage are understood. Data lineage information must be available for all sensitive data. This must, at a minimum, include the source from which the data was ingested or in which it was created in a cloud environment.



2.73
Average Score

7%
Achieved / Enhanced

Analysis

This seems to be a challenging area, with 16% not initiating and only 7% having achieved this.

For the operational data management controls in their organizations, the majority are manual, followed by not implemented, and the least proportion is automated. This suggests a need to focus on automating these controls, given the efficiency and effectiveness of automation.

The data lineage in cloud environments must be captured automatically, and changes to lineage must be tracked and managed. Visualization and reporting of lineage must be implemented to meet the needs of both business and technical users. Most organizations are in the early stages of this subcapability.



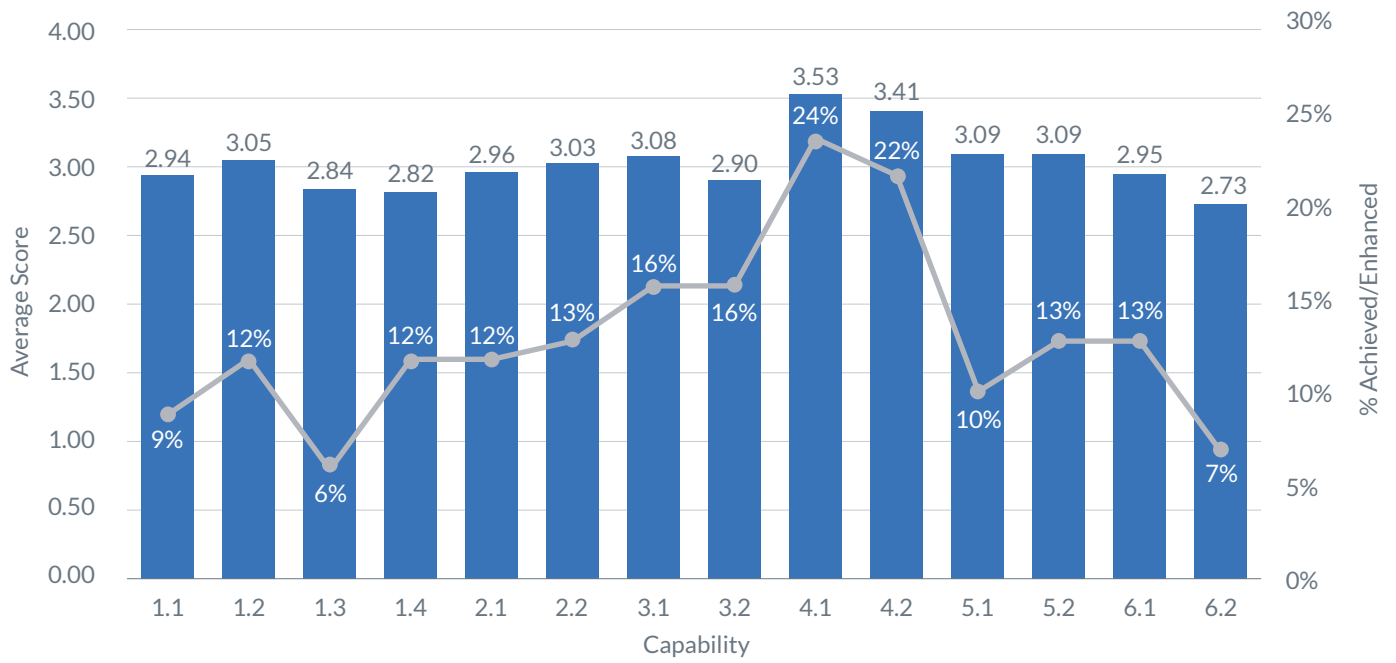
CDMC™ Composite Scores

CDMC™ Composite Scores

CDMC Key Control	CDMC Component	Capability	Average Score	% Achieved / Enhanced
Data Control Compliance	1.0 Governance & Accountability	1.1 Cloud data management business cases are defined and governed. Data control compliance must be monitored for all data assets containing sensitive data via metrics and automated notifications.	2.94	9%
Ownership Field	1.0 Governance & Accountability	1.2 Data ownership is established for both migrated and cloud-generated data. The ownership field in a data catalog must be populated for all sensitive data or otherwise reported to a defined workflow.	3.05	12%
Authoritative Data Sources and Provisioning Points	1.0 Governance & Accountability	1.3 Data sourcing and consumption are governed and supported by automation. A register of authoritative data sources and provisioning points must be populated for all data assets containing sensitive data or otherwise must be reported to a defined workflow.	2.84	6%
Data Sovereignty and Cross-Border Movement	1.0 Governance & Accountability	1.4 Data sovereignty and cross-border data movement are managed. The data sovereignty and cross-border movement of sensitive data must be recorded, auditable and controlled according to defined policy.	2.82	12%
Cataloging	2.0 Cataloging & Classification	2.1 Data catalogs are implemented, used, and interoperable. Cataloging must be automated for all data at the point of creation or ingestion, with consistency across all environments.	2.96	12%
Classification	2.0 Cataloging & Classification	2.2 Data classifications are defined and used. Classification must be automated for all data at the point of creation or ingestion and must be always on.	3.03	13%

Entitlements and Access for Sensitive Data	3.0 Accessibility & Use	<p>3.1 Data entitlements are managed, enforced, and tracked.</p> <p>1. Entitlements and access for sensitive data must default to creator and owner until explicitly and authoritatively granted.</p> <p>2. Access must be tracked for all sensitive data.</p>	3.08	16%
Data Consumption Purpose	3.0 Accessibility & Use	<p>3.2 Ethical access, use, and outcomes of data are managed. Data consumption purpose must be provided for all data sharing agreements involving sensitive data. The purpose must specify the type of data required and include country or legal entity scope for complex international organizations.</p>	2.90	16%
Security Controls	4.0 Protection & Privacy	<p>4.1 Data are secured, and controls are evidenced.</p> <p>1. Appropriate security controls must be enabled for sensitive data. 2. Security control evidence must be recorded in the data catalog for all sensitive data.</p>	3.53	24%
Data Protection Impact Assessments	4.0 Protection & Privacy	<p>4.2 A data privacy framework is defined and operational. Data protection impact assessments (DPIAs) must be triggered automatically for all personal data according to its jurisdiction.</p>	3.41	22%
Data Retention, Archiving, and Purging	5.0 Data Lifecycle	<p>5.1 The data lifecycle is planned and managed. Data retention, archiving, and purging must be managed according to a defined retention schedule.</p>	3.09	10%
Data Quality Measurement	5.0 Data Lifecycle	<p>5.2 Data quality is managed. Data quality measurement must be enabled for sensitive data with metrics distributed when available.</p>	3.09	13%
Cost Metrics	6.0 Data & Technical Architecture	<p>6.1 Technical design principles are established and applied. Cost metrics directly associated with data use, storage, and movement must be available in the catalog.</p>	2.95	13%
Data Lineage	6.0 Data & Technical Architecture	<p>6.2 Data provenance and lineage are understood. Data lineage information must be available for all sensitive data. This must at a minimum include the source from which the data was ingested or in which it was created in a cloud environment.</p>	2.73	7%

CDMC™ Composite Scores



● Average Score — % Achieved/Enhanced



Global Advocates for **Data & Analytics** Management

The EDM Council is the leading global trade association providing best practices, standards and education to data and business professionals in our data-driven world. We collaborate with our members to advance cross-industry data management and analytics programs and increase the business value of data assets.



Best Practices



Data Standards



Training &
Certification



Research &
Benchmarking



Regulatory
Engagement



Networking

Your data is extremely valuable. Make sure your organization is getting the most out of it. Join our community of 350+ organizations and 25,000+ professionals.

edmcouncil.org



2023 Benchmark Advisory Team



JOHN A. BOTTEGA
President, EDM Council

John Bottega began working with the EDM Council as an industry contributor in 2005 and served as Chairman from 2007 to 2014. He joined the Council's executive team as a Senior Advisor in 2014, took over as Executive Director of the EDM Council in 2017, and was promoted to President in 2020. John is a senior strategy and data management executive with more than 30 years of experience in the industry. Over his career, John has served as Chief Data Officer in both the private and public sectors, serving as CDO for Citi and Bank of America, as well as for the Federal Reserve Bank of New York.



DIANA ASCHER
Head of Research, EDM Council

As head of research and senior advisor on data ethics and responsible artificial intelligence to the EDM Council, Diana Ascher helps ensure the world's leading organizations develop ethical data management policies and practices. She is the founder of the Information Ethics & Equity Institute. Previously, Diana gained expertise in information science research and instruction as director of the IS Lab at UCLA and manager of the UCLA Anderson Laboratory in Behavioral Decision Making; and in data and knowledge management, custom-publishing technologies, and financial communication and literacy at American Funds and Bloomberg.



MICHAEL MERITON
Co-Founder, EDM Council

Mike Meriton is a Co-Founder of the EDM Council and served as the first Chairman and active Board member since inception in 2005. Mike joined in 2015 as a Senior Advisor, promoted to COO in 2020, to lead industry engagement strategy, new member services and council operations. Previously, Mike was the CEO of GoldenSource and held key executive roles at CheckFree (Fiserv), D&B, and Oracle.



JIM HALCOMB
Global Head of Product Management,
EDM Council

Jim Halcomb is Head of Product Management with over 20 years' experience as a strategy, data, and analytics leader across industries. Jim has served as chief data executive for companies in Financial Services, Healthcare, and Cybersecurity; primary author of the CMMI's Data Management Maturity Model; and author of data management principles for ISACA's privacy and security training and certifications programs.

