2024 Planning Guide for Data Management

Published 4 October 2023 - ID G00796443 - 48 min read

By Analyst(s): Ramke Ramakrishnan, Joe Maguire, Jason Medd, Lyn Robison

Initiatives: Data Management Solutions for Technical Professionals; Evolve Technology and Process Capabilities to Support D&A

The rise of generative AI and LLMs offers exciting opportunities for delivering scalable data management architecture and governance. Data and analytics technical professionals should continuously enhance and modernize their data management practices to drive greater business outcomes.

Overview

Key Findings

- Advancing your data architecture for generative AI capabilities requires a holistic approach and a multidisciplinary effort involving the cross-functional coordination of multiple data and analytics teams.

- A unified architecture pattern such as lakehouse and LDW presents benefits in reducing data redundancy and integration complexity; enabling quicker data accessibility; simplifying design, maintenance and optimization; and enforcing data standards and governance.

- Data and analytics (D&A) governance delivers competitive advantage and business value while fulfilling regulatory requirements and compliance.

- Harnessing the power of GenAI, particularly large language models (LLMs), empowers data management professionals to automate and enhance various initiatives, thereby bringing efficiency and effectiveness to data management practices.

Recommendations

As a data and analytics technical professional charged with delivering effective data management solutions, you should:

- Cultivate data governance and advanced data processing techniques to support LLMs and generative AI capabilities. Data teams will be tasked with working with unstructured/semistructured data, information retrieval and knowledge graphs to enable trusted, up-to-date and accurate data for AI.

- Design your cloud data ecosystem and data management architecture, balancing optimization and flexibility by accounting for performance, latency, efficiency and operations to contribute to overall financial governance.

- Address data quality early on to prevent downstream data issues by augmenting data pipelines, defining capacity planning and production management with data observability, and deploying secure end-to-end D&A governance.

- Use generative AI cautiously and within an organizational AI strategy and roadmap. Scrutinize GenAI output continuously to mitigate risks associated with incorrectness, bias, legal, ethical, security and other concerns.

Data Management Trends

In 2024, data and analytics technical professionals deploying data and analytics solutions must strive to enhance and modernize their data management platforms to achieve greater business outcomes. However, modernization often leads to intricate and complex deployment scenarios. Technical professionals are being asked to deliver frictionless data management and to seamlessly integrate with architectures spanning diverse environments, including on-premises, hybrid, edge and cloud systems.

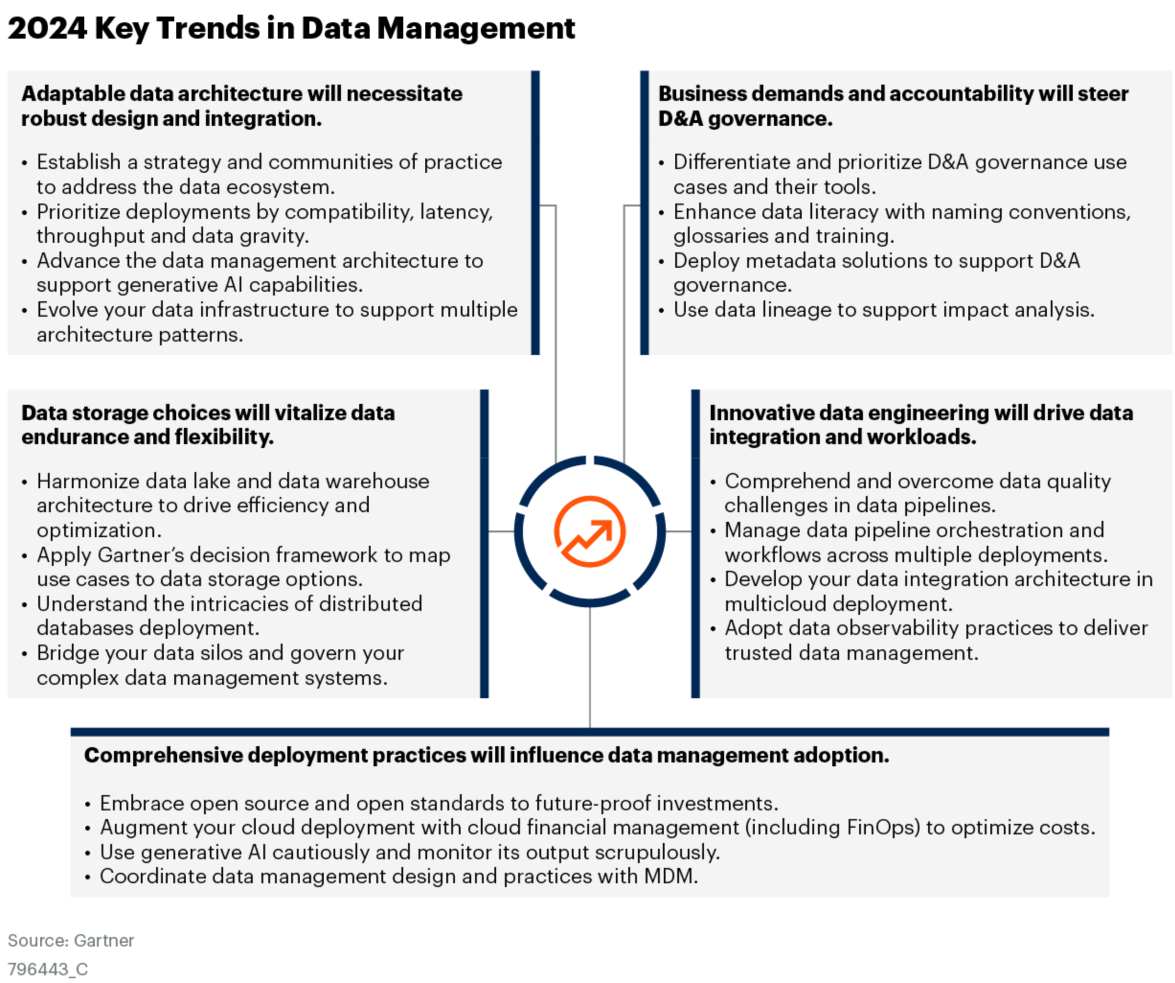
The status quo might not handle the velocity, variety and volume of data generated by modern application and analytical workloads. Technical professionals must adopt new and recent approaches to data storage and delivery. These approaches influence practices for data lakes, data warehouses, data hubs, logical data warehouses, data fabrics, data meshes, DataOps, data orchestration, data governance and the overall data ecosystem.

Additionally, harnessing the power of generative AI, particularly large language models (LLMs), empowers data management professionals to automate and enhance various endeavors such as data transformation, code development and documentation. This can make practices in data management and governance more effective and efficient. However, you must use them with careful consideration and continuously monitor the output.

Some common use cases for LLMs in data management include automating data interpretation and anomaly detection to improve data quality, utilizing natural language for constructing database queries, providing augmented assistance for database and code migrations, and identifying data privacy and compliance issues.

To help data and analytics technical professionals on their path to modernization, this Planning Guide identifies the key 2024 trends for data management, along with related planning considerations for developing and deploying a successful data management architecture and foundation (see Figure 1).

Figure 1: 2024 Key Trends in Data Management



The major trends for 2024 are:

- Adaptable data architecture will necessitate robust design and integration.
- Data storage choices will vitalize data endurance and flexibility.

- Business demands and accountability will steer D&A governance.
- Innovative data engineering will drive data integration and workloads.
- Comprehensive deployment practices will influence data management adoption.

Adaptable Data Architecture Will Necessitate Robust Design and Integration

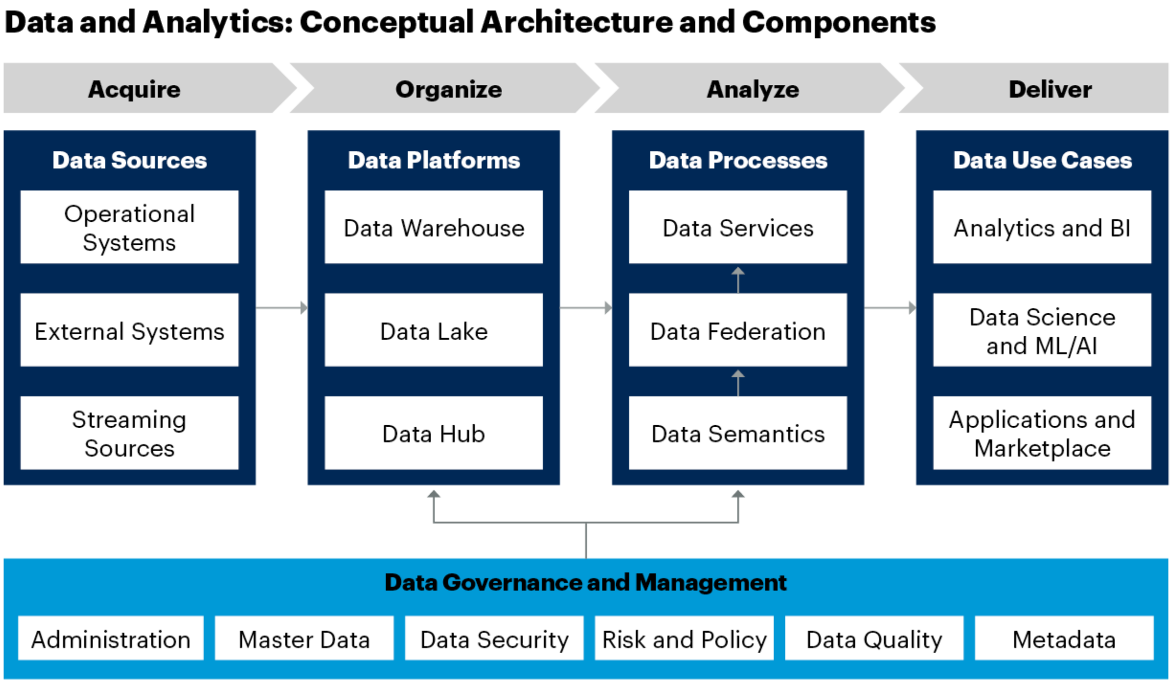
Data and analytics architectures are evolving rapidly, forcing technical professionals to adapt their current strategies to accommodate on-premises, cloud, multicloud, intercloud and hybrid deployment structures. To succeed, organizations must design adaptable data architecture that seamlessly integrates, enables collaboration, and delivers complementary analytics solutions.

Assembling a robust data and analytics (D&A) architecture necessitates detailed planning to ensure the unified design and integration of new technologies with existing infrastructure and services. The architecture should also consider the ability to cater to future demands for additional capabilities and align with the goals and objectives of the broader D&A governance program. Moreover, the effectiveness of such architecture increasingly relies on the seamless integration and compatibility of services spanning various domains, including cloud providers, software as a service (SaaS) solutions, on-premises resource deployments, and cloud data management and orchestration services.

As you evaluate your cloud-based services strategy to deploy your data architecture, it's essential to gauge its level of integration with the underlying cloud data ecosystem it operates within. A cohesive and well-connected network of services within a singular cloud platform might surpass the perceived advantages of adopting services across multiple services within and across different platforms. Aiming to create a more interconnected ecosystem ensures a clearly defined path to production and facilitates streamlined data and metadata management processes.

> **A cohesive data management environment supports all kinds of data workloads and is accessible via unified access with augmented and integrated capabilities. These include advancing generative AI capabilities with a robust framework for common governance and metadata management.**

Figure 2: Data and Analytics Architectural Components



Source: Gartner
796443_C

Planning Considerations

Based on this trend, you should focus your 2024 data architecture efforts on the following activities:

- Establish a strategy and communities of practice to address the data ecosystem.
- Prioritize deployments by compatibility, latency, throughput and data gravity.
- Advance the data management architecture to support generative AI capabilities.
- Evolve your data infrastructure to support multiple architecture patterns.

Establish a Strategy and Communities of Practice to Address the Data Ecosystem

Data ecosystems leverage distributed components that run on multiple clouds and/or on-premises but are treated as a cohesive whole:

- By 2025, 55% of IT will adopt data ecosystems, consolidating the vendor landscape by 40%, thereby reducing cost while reducing choice.

The 2022 Gartner State of Data and Analytics Cloud Adoption Survey found that 85% of those using the public cloud were using more than one. [1] Therefore, it is common for organizations to grow into a hybrid and multicloud environment using a mix of private and public cloud platforms for specific use cases.

Other factors that lead to multicloud include regulatory requirements, governance, mergers and acquisitions, and the nonavailability of primary cloud providers in certain regions. Multicloud architectures often arise organically for these reasons and may not align with the organization's primary cloud strategy. Thus, it is always best to establish an overarching cloud strategy with standard policies that prioritize providers and practices that govern additional cloud deployments. These policies and practices will help mitigate the risks that unsanctioned cloud deployments can pose to your overall data management solution.

Proper upfront planning of multicloud strategy brings numerous benefits, including localization, reliability, better price/performance, reduced vendor lock-in, and significantly reducing complexity in your cloud deployments. To manage and support your data management capabilities, you must:

- Devise a strategy that governs which primary and secondary providers to use for which use cases.
- Ensure that those practices are documented and communicated across your organization.

Data management strategies include establishing a data ecosystem that utilizes augmented data catalogs to create a unified metadata inventory spread across distributed data stores. The collected metadata should include technical, operational, business and social metadata. Once the metadata has been collected, data fabrics can analyze it continuously to monitor, maintain and continuously improve the data pipeline. As data ecosystems integrate (and analyze) metadata across a multicloud/hybrid ecosystem — and use this "active metadata" for decision insights and automation — they leverage the data fabric design patterns to support augmented data management and become increasingly self-correcting.

Prioritize Deployments by Compatibility, Latency, Throughput and Data Gravity

Traditionally, data management solutions have been designed and optimized for centralized deployments. However, distributed and complex use cases are now driving newer innovations that deliver business value. Organizations now aim to look beyond public cloud environments for their overall operational strategies.

- Creating or adapting to software as a service (SaaS) offerings
- Incorporating DevOps and DataOps methodologies
- Utilizing the cloud's flexible pricing and resource models
- Extending and facilitating more self-service capabilities in platform and delivering analytics

Achieving success in your deployment is dependent on the ability to cater to business users' specific needs and is not necessarily defined or limited by the cloud provider's capabilities. Implementations of those are likely to be distributed using multicloud and intercloud architecture with functions that are likely to run close to the data sources and operate at the edge and other devices. Some of the common best practices include:

- Adapt your design and architecture to suit the specific demands of each use case. Incorporate on-premises options to avoid cloud-based network latency and to achieve throughput requirements. Select your vendor offerings based on current and future centers of data gravity. As your data management framework evolves, maintain flexibility for cloud migration.
- Prioritize aligning hybrid, intercloud, or multicloud deployment tiers with the distinct optimization requirements of your use cases. This will guide you when addressing the trade-off between offering optimization and ensuring flexibility. Optimization includes the combination of costs, performance and integration.
- Harmonize business use cases and prioritize deployment strategies to empower your organization to be more robust, reliable, scalable and high-performing. This approach will also position you to embrace open-source innovations and foster seamless interoperability with other systems.
- Establish comprehensive policies and optimization tools for multicloud or intercloud strategies. These tools should encompass mechanisms for monitoring usage and adoption and preventing the adoption of nonstrategic solutions and unchecked expenses.

- Link the security context to the data movement. This strategy guarantees consistency across systems containing duplicate data. It’s also prudent to allow different security contexts to drive distinct technology stacks, particularly in cases involving different jurisdictions with varying regulations.

As data proliferation continues within the data ecosystem, tracking metadata and monitoring metadata organizationwide and across the cloud becomes essential. So establishing a companywide policy and a centralized data repository for metadata tracking empowers all organizations to swiftly locate pertinent information for their tasks.

Advance the Data Management Architecture to Support Generative AI Capabilities

The evolution of artificial intelligence and machine learning is changing how you define your data management architecture. Delivering AI capabilities that are intuitive, seamless and adaptive involves advancing and optimizing the way of storing and processing the data in the data management layer to support generative AI capabilities. Many of the advancements in the use of AI and machine learning in the data management space have been in the works for over a decade. However, the latest acceleration of generative AI initiatives makes it more meaningful and relevant for data management practitioners to advance and augment those capabilities to support a robust data architecture and foundation for AI initiatives. Some of the common use cases for data management include the use of NLP for query language processing, autonomous database management, vector storage, processing to support advanced use cases, augmented data management, integrated data quality checks and running embedded AI/ML models within the database.

Also, advancing your data architecture for generative AI requires a multidisciplinary effort involving several teams, such as data engineering, data science and machine learning, security, ethics, and governance, to deliver a holistic approach to deploying generative AI capabilities.

Although the traditional way of storing data in modern databases remains, we are now seeing newer concepts, including evolving capabilities such as vector searches and embedding through vectorized API capabilities added to the traditional databases or with newer dedicated vector databases. Vector representation allows embedding that captures semantic relationships between the data with the contextual aspects of the data supporting high dimensionality and efficient search by calculating distances and similarities between vectors. Storage of embeddings along with vector format in vector databases can optimize storage and improve data access compared to traditional databases. Also, by storing diverse embeddings and settings, vectorization can significantly enhance and support generative AI models with a broad range of scenarios and outputs. Not all generative use cases will require vector databases until you grow into something big. For the initial prototypes and proofs of concepts, any regular databases that can support vector formats will suffice.

From a data management perspective, NLP is getting strongly integrated with several databases that can help to prepare and manage datasets to find insights and relationships in structured, semistructured and unstructured contexts that are useful to train LLMs to learn, understand, manipulate and generate human language more effectively. Additionally, knowledge graphs can be used to support NLP with well-defined ontologies, taxonomies and vocabularies.

Evolve Your Data Infrastructure to Support Multiple Architecture Patterns

As you adopt a domain-centric approach for data management to drive common standards and agility across your deployment, you should investigate logical data warehouse, data mesh or data fabric design and other architectural options. In some circumstances, you should even consider using more than one as a complementary solution to simplify data management and access to your data products.

A data product is a curated and self-contained combination of data, metadata, semantics and templates. It includes access and implementation logic certified for tackling specific business scenarios and reuse. A data product must be consumption-ready (trusted by consumers), kept up to date (by engineering teams) and approved for use (governed). Data products enable various data and analytics (D&A) use cases, such as data sharing, data monetization, domain analytics and application integration.

Data mesh is domain-driven. It involves a “data-as-a-product” mindset and decentralized data ownership, with distributed teams managing the data. The data mesh approach shifts the responsibility of data integration and data cleansing from the central data engineering team to the business domains. Yet, it can still tap into centralized data infrastructure and common standards as part of the shared governance supported by the data fabric. The data fabric pattern is underpinned by metadata that can automate specific data management tasks, including cataloging data assets and publishing data products to the catalog.

In order to evolve your data infrastructure to support different architecture patterns, including data fabric:

- Inventory all types of your data assets including metadata and their associated relationships into a flexible data model to create an augmented data catalog.
- Ensure that the data fabric allows business teams to contribute and enrich the data models with semantics and taxonomies/ontologies.
- Provision a data fabric architecture that has the capability to combine various data integration styles.
- Assess your current data management tools and components to identify the existing, overlapping or missing capabilities needed to deliver the data fabric design.

Also read: Demystifying Taxonomies, Ontologies and Data Models.

Data Storage Choices Will Vitalize Data Endurance and Flexibility

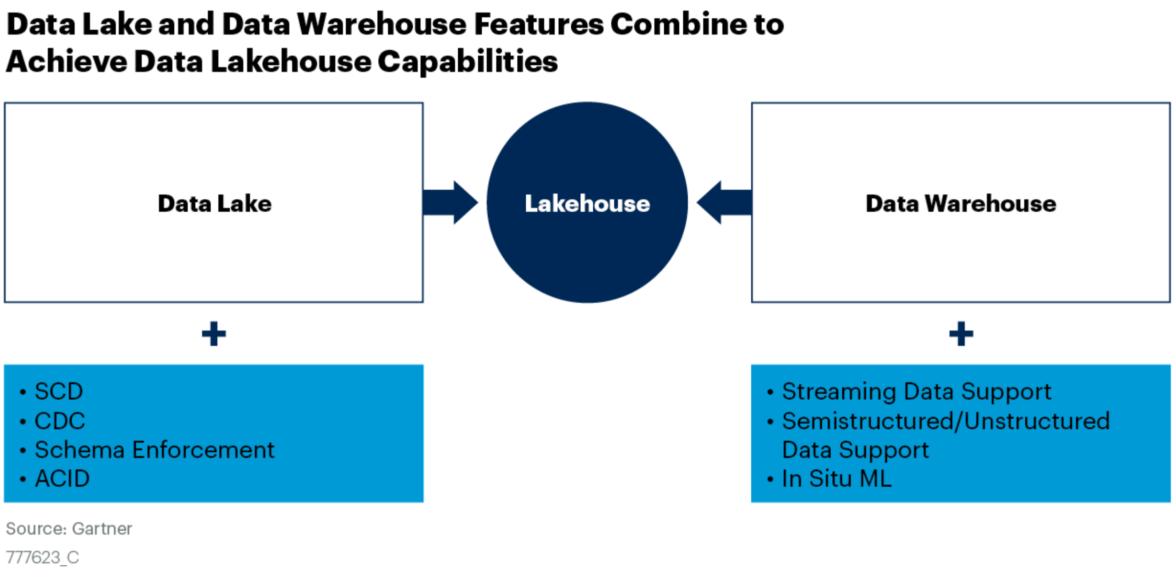
Data Storage choices are continuously evolving, increasing the importance of using multiple data storage and processing engines to maximize cost advantages in both on-premises and cloud environments. Not all data lives in one place. Data tends to be distributed across multiple systems, both inside and outside the enterprise, and it remains difficult to identify and integrate. Cloud data management also has nonrelational and special-purpose data stores such as graphs, time series and in-memory databases.

Data also needs to evolve as new digital scenarios emerge. The burden of administering infrastructure makes it harder for data and analytics professionals to deliver solutions that keep up with changing business needs. By deploying fully managed platforms, technical professionals can focus on delivering business values instead of mundane database administration tasks, as well as maintenance tasks like backup, patching and scaling.

A strategic approach to data storage can contribute significantly to an organization's overall data management strategy and its ability to derive value from its data assets.

With the evolution of multiple systems required to support analytical architecture, organizations are looking for ways to create a data ecosystem where they integrate multiple data management solutions to provide unified access. Multimodel approaches, where multiple data types are stored and managed in the same system, are also becoming increasingly common. An example is the data lakehouse architecture pattern, which combines the capabilities of the data lake and the data warehouse on a single platform.

Figure 3: Data Lakehouse



Planning Considerations

To further support this trend, focus your 2024 data storage deployment efforts on the following activities:

- Harmonize data lake and data warehouse architecture to drive efficiency and optimization.
- Apply Gartner's decision framework to map use cases to data storage options.

- Understand the intricacies of distributed databases deployment.
- Bridge your data silos and govern your complex data management systems.

Harmonize Data Lake and Data Warehouse Architecture to Drive Efficiency and Optimization

Cloud-based systems are moving toward a converged data architecture that combines the capabilities of the data lake and the data warehouse on a single data platform to drive efficiency, optimization and cost savings. A unified lakehouse architecture pattern presents benefits in reducing data redundancy and integration complexity; enabling quicker data accessibility; simplifying design, maintenance and optimization; and enforcing data standards and governance.

Lakehouse architecture can be either monolithic or distributed. The design principles can support either a single data storage layer and processing engine or multiple data storage layers and processing engines to maximize price/performance in the cloud. In both scenarios, different storage is utilized by the same compute engine, and multiple compute engines run against the same storage to help optimize resources.

The distributed architecture allows organizations to implement data lakes and data warehouses independently, and then unify them into a lakehouse architecture pattern. The distributed approach is an alternative to adopting an immature lakehouse solution, which may lack the essential capabilities, desired scale and performance to support advanced use cases.

To learn more about data lakehouse architecture and how to benefit from it, see Does My Organization Need a Data Lakehouse?

Apply Gartner's Decision Framework to Map Use Cases to Data Storage Options

Organizations are continuously assessing and comparing storage possibilities to determine which options can best meet their use cases and business requirements. As part of their cloud migration strategies, organizations running a single vendor database stack on-premises or in a private cloud have started considering data storage options available on the public cloud to modernize their data architectures.

Although overlapping capabilities exist across multiple databases, selecting specific data stores or DBMSs without considering specific use cases can have practical implications on scale, performance, cost and maintenance, leading to suboptimal usability. However, mapping use-case requirements to a specific data storage solution can be complex. Moreover, organizations need a solid understanding of the different types of workloads to determine the optimal DBMS.

To address these challenges, Gartner has created a decision framework and tool. See Decision Point for Selecting the Right DBMS to determine the right data store for your data management and analytics requirements.

Understand the Intricacies of Distributed Databases Deployment

Distributed database systems offer numerous advantages that address the challenges posed by the management of large-scale data in modern computing environments. Most of the database architectures that are deployed are either a centralized transactional architecture or an eventually consistent distributed architecture. These two architectures are complementary, and the lists of their capabilities are almost reversed images of each other. Distributed databases are generally used for transactional systems and not analytical systems, and for most common use cases, a typical centralized or somewhat distributed database system will suffice. Also, most of the transactional systems that are not dependent on high performance have moved toward NoSQL databases.

On the other hand, distributed SQL database architectures replicate data across different regions while maintaining strong consistency guarantees. The distributed SQL database was really made possible because of the cloud due to its ability to easily scale with built-in resiliency and disaster recovery. It combines the resilience and scalability of a NoSQL database with the full functionality of a relational database to support high-end transactional systems.

NewSQL databases are a new generation of databases that aim to provide the scalability and performance benefits of NoSQL databases while retaining the consistency and transactional guarantees of traditional relational databases. NewSQL databases are designed to work well in environments where there is a high volume of concurrent transactions, and where low latency is a key requirement. These databases can scale horizontally to handle increasing workloads and rapid increases in transactions, which makes it ideal for cloud-based deployments. In addition, NewSQL databases often offer strong consistency guarantees, meaning that changes to the data are immediately visible to all nodes in the cluster.

Distributed SQL may be the better choice for applications that require the consistency and transactional guarantees of traditional relational databases. For those that require the scalability and performance advantages of NoSQL databases, NewSQL may be a better fit.

To learn more about the distributed database capabilities and options, read Assessing Distributed SQL Databases.

Bridge Your Data Silos to Govern Your Complex Data Management Systems

Organizations are developing flexible, versatile data management architectures using logical data warehouse (LDW) principles that can provide unified access, augmented capabilities and holistic governance across diversified data stores to create a data ecosystem. The LDW can be extended to offer unified access to distributed data management using data virtualization/federation.

Data virtualization provides cohesive and unified access to business data across distributed data management, both real-time and near real-time. Data virtualization can also be used independently by connecting existing data sources and other databases via an integrated layer.

LDW principles and data virtualization can further evolve into data fabric design, enabling flexible, reusable and augmented data integration pipelines, services and semantics that support various operational and analytics use cases across multiple deployments and orchestration platforms. A data fabric supports a combination of different data integration styles and utilizes active metadata, knowledge graphs, semantics and ML to augment data integration design and delivery.

To learn how to best integrate and govern diverse and distributed data, read Assessing the Relevance of Data Virtualization in Modern Data Architectures.

Business Demands and Accountability Will Steer D&A Governance

Business demands and outcomes will dictate a complete, distributed and coordinated strategy for governance, including financial governance.

Organizations pursue D&A governance to meet regulatory and compliance requirements, such as privacy and security, because the inability to satisfy such directives can lead to costly penalties and other legal repercussions. In addition to responding to regulatory and compliance needs through governance, organizations are now looking at governance as a means to achieve strategic initiatives and business goals.

D&A governance is difficult to operationalize and coordinate with the organization's overall strategy because business users sometimes cannot perceive how governance improves business outcomes. Thus, the original motivations for governance were externally imposed regulatory compliance. But when organizations established outcome-driven D&A governance, they began to reap benefits unrelated to compliance.

In many organizations, technology professionals participate in data and analytics governance because business leaders ask them to. But in some organizations, IT teams value data and analytics governance even if business leaders do not. These IT teams try to establish and lead D&A governance programs themselves. Such efforts can fall short in several ways. First, D&A governance programs thrive best when focused on business-driven goals, but IT personnel might perceive those goals incompletely. Second, IT personnel tend to monitor D&A governance programs exclusively with technical metrics such as objective quality and completeness of metadata. A fuller assessment of D&A governance will measure business outcomes like rates of employee turnover or customer satisfaction.

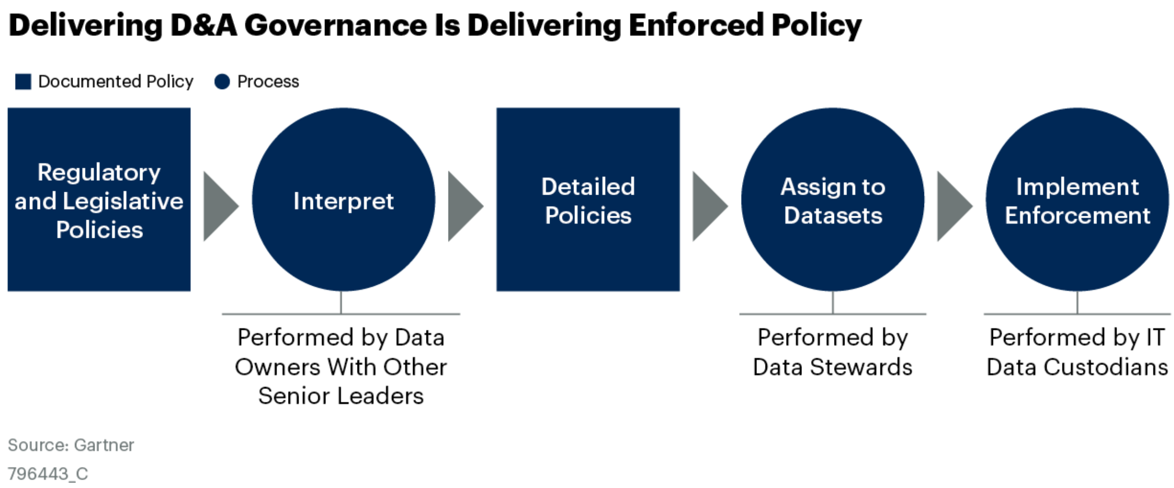
In the 2021 Gartner Data and Analytics Governance Survey, only 18% of all surveyed organizations said that their D&A governance was mature and scaling across the enterprise. [2] Most still needed IT support to enforce policies. Thus, it is unrealistic to rush the creation of data domains and believe that subject matter experts (SMEs) at the domain level will be both adept at and willing to enforce the policy.

Developing a CDAO Mindset: A Quick Guide for Technical Professionals provides guidance for technical professionals who may be attempting to lead data projects in the absence of appropriate leadership guidance, such as a CDAO.

Organizations now realize that D&A governance can drive competitive advantage and business value while satisfying regulatory requirements.

All manifestations of D&A governance share an overarching workflow. That workflow converts high-level policy requirements (imposed on the organization via external laws and regulations) into detailed policy that is enforced on specific data assets. Figure 4 outlines this workflow.

Figure 4: Delivering D&A Governance Is Delivering Enforced Policy



Note that data and analytics governance is closely related to AI governance. A workflow like the one above can be used to articulate policies for responsible and ethical use of AI. For more information, see the 2024 Planning Guide for Analytics and Artificial Intelligence.

Planning Considerations

To contribute effectively to D&A governance programs, focus your 2024 planning efforts on the following activities:

- Differentiate and prioritize D&A governance use cases and their tools.
- Enhance data literacy with naming conventions, glossaries and training.
- Deploy metadata solutions to support D&A governance.
- Use data lineage to support impact analysis.

Differentiate and Prioritize D&A Governance Use Cases and Their Tools

D&A governance is about articulating and enforcing policies on data. Those policy objectives fall into several categories including:

- Data quality
- Master data management
- Identity, security and privacy
- Data life cycle management

Different organizations will prioritize these policy objectives differently. For example, if you run your own data center and want to control storage costs, then data life cycle management might be prioritized. Likewise, if you maintain a lot of personal data about your customers, then identity, security and privacy could be prioritized.

Note that different policy objectives will require different governance processes and, in many cases, different tools. (Of course, some tooling, like data catalogs, will help many of the use cases.)

Your plan for delivering governance should include an incremental roadmap. Such a roadmap would allow for gradual acquisition of some technology. For example, identity, security and privacy use cases require additional tools and techniques, including any or all of the following:

- Identity governance and administration (e.g., One Identity Safeguard by Quest Software and SailPoint Identity Security Platform)
- Access control frameworks (e.g., Apache Ranger and Apache Sentry)
- Data-masking tools (e.g., Redgate Software Data Masker and Privitar Data Privacy Platform)
- Data loss prevention (DLP) tools (e.g., Broadcom [Symantec] Data Loss Prevention and Digital Guardian Endpoint DLP)
- Native access control features of DBMSs and other tools

Technical professionals responsible for technical implementation of D&A governance policies should assist the program with:

- Auditing and classifying data assets to find sensitive data. Prioritize the portions of the data architecture that are most likely to contain sensitive data.

- Differentiating imposed policy from delivered policy. Imposed policy is regulatory or legislative; it constitutes requirements. Delivered policy is more detailed — it is enforceable through technology.

- Prioritizing D&A governance deliverables based on desired business outcomes through collaboration with business stakeholders. Adaptive governance provides the flexibility to deliver innovative business outcomes.

Enhance Data Literacy With Naming Conventions, Glossaries and Training

D&A governance programs depend on data literacy — that is, the individual and institutional ability to read, write and communicate data in context. To support your D&A governance program, you should attend to data literacy in the following ways:

- Establish semantically helpful naming conventions for data artifacts, and teach those conventions widely.

- Build a business data glossary.

- Deploy a formal training program in data literacy.

Establish semantically helpful naming conventions for data artifacts. Typical naming conventions focus on superficial standards, such as preferring underscores over hyphens or preferring camel case over initial caps. Such standards are useful, but data literacy demands more. Specifically, data literacy demands that each data artifact has a name that accurately describes its data payload. For example, a column named “weight” is a poor name because it fails to indicate units and it fails to describe semantics. A better name is weight in kilograms because it includes units. Better still is birth weight in kilograms, because the name now provides interpretive context.

Standards for semantically meaningful names can apply to individual columns, to entire datasets (such as tables and query results) and to metrics. For more information, see Improve Data Literacy and Governance by Using Accurate, Meaningful Names for Data Artifacts.

Note that superficial naming standards are typically the purview of IT professionals only because they want consistent formats for the names of database tables, database columns and program variables. By contrast, semantically meaningful names should be taught to both IT professionals and business personnel. That's because business personnel will often encounter a dataset through its name. Furthermore, some business personnel such as self-service analysts will create datasets, and those datasets should also be given semantically revealing names.

Build a business data glossary. Data literacy is the ability to read, write and communicate about data in context. That context can vary across industries, across companies and across business units within a company. (For example, the legal department and the finance department might use subtly different interpretations for the term customer.) Such variation can cause confusion and misunderstanding, both of which threaten data and analytics programs and their attendant governance. You should build a business data glossary, and it should include names for datasets (both persisted and calculated), scalar values (such as database columns or data-entry fields), and common metrics and business KPIs. Where applicable, you should apply the semantically meaningful naming conventions to the terms in the business data glossary.

Deploy a formal training program in data literacy. Start by conducting a data literacy assessment to determine the current state of data literacy in the organization. Leverage the results of this assessment to determine core competencies for analytics and existing skill gaps. Then develop an incremental roadmap to deliver data literacy gradually. The roadmap can prioritize some data-literacy topics over others, some data role players over others and even some data domains over others.

Note that one part of data literacy training should include the semantically meaningful naming conventions described earlier in this section.

For more information on developing a data literacy program, see Tackle Data Literacy Head-On to Avoid Data and Analytics Program Failure.

Deploy Metadata Solutions to Support D&A Governance

Metadata is essential for D&A governance. But metadata is a complex phenomenon with multiple manifestations, including:

- Technical metadata (e.g., database schemas, table names and columns names)
- Operational metadata (e.g., extraction, lineage, transformation and loading [ETL] operations and their orchestration)

- Business metadata (e.g., terms in a business data glossary)
- Social metadata (e.g., annotations attached to the other forms of metadata)

Furthermore, metadata can exist in many locations, including:

- Persistent stores (e.g., the catalog of relational DBMSs)
- Virtualization engines (e.g., data virtualization tools and query accelerators)
- Data movement and orchestration tools (e.g., data warehouse automation tools)
- Multiple places throughout the data architecture (e.g., business intelligence [BI], data warehouses, data marts and semantic layers)
- Security brokers (e.g., data access control tools)
- Message brokers such as Apache Kafka and real-time engines
- ML life cycle management tools (e.g., MLflow, feature stores, vector databases)

Because metadata is everywhere, maintaining a uniform view of it is difficult. Data catalog tools can help consolidate metadata in a way that supports D&A governance throughout the software development life cycle (SDLC).

As a technical professional supporting D&A governance via metadata management, you should:

- Integrate metadata management roadmaps with overall data strategy. If the data strategy roadmap calls for incremental delivery of data management solutions, then the roadmap for metadata management should use the same priorities.
- Calibrate your metadata management aspirations with the traits of your data architecture. Those traits will impose certain demands on metadata management. For example, data mesh requires a mesh-wide registry of data products, where each entry describes the syntax and semantics of a data product available on the mesh. Likewise, data fabric requires active metadata management to detect opportunities for fine-tuning your pipeline and its orchestration.

- Differentiate metadata management along phases of the SDLC. For example, your governance program can have one process that protects production databases, but another process that protects development and test databases. Your roadmap for delivering governance should implement processes for these different phases incrementally.

- Use various techniques to populate the metadata repository automatically. Techniques include:

    - Direct harvesting of persistent schemas.

    - Data profiling, such as detecting credit card numbers by running Card Verification Value 2 (CVV2) algorithms.

    - Automatic metadata discovery. For example, tools may infer likely primary key/foreign key relationships by scrutinizing the contents of data lakes and other data assets.

Use Data Lineage to Support Impact Analysis

Data lineage is usually visualized via a map that displays the journey of the data from its source database to its final analytical use or disposal. By its very nature, lineage involves connecting metadata from multiple parts of an architecture.

Within that architecture, some components can provide limited lineage capabilities. For example, a data preparation tool may be able to document the lineage of how the data was prepared in detail, but that lineage is limited to the product itself.

A complete end-to-end lineage is more likely to be deployed by combining or independently using multiple stand-alone solutions, such as Informatica's data catalog, Manta's data lineage solution or Tableau's analytics platform.

End-to-end lineage can reveal the impact of upstream changes to downstream reports and models.

Technical professionals responsible for supporting D&A governance with data lineage should:

- Determine which products are involved in the data pipeline.

- Understand exactly which metadata can be harvested by data catalogs. For example, most products will be able to gather metadata from Tableau reports, but only some can handle custom SQL statements inside Tableau.
- Devise and support a plan to manually document lineage where automated scans are not supported or to utilize and integrate multiple metadata management tools where required to create a comprehensive view of lineage across the data landscape.
- Formalize data lineage processes for different phases of the SDLC. For example:
  - Impact analysis is typically a design-time activity, such as assessing how changing the data type of a column in an operational system will influence downstream artifacts.
  - Root cause analysis can be a runtime activity, such as when data stewards or self-service analysts inspect the data integration pipeline to diagnose why a data visualization that worked yesterday is producing incorrect data today.

Innovative Data Engineering Will Drive Data Integration Pipelines and Workloads

Enrich your data engineering effort with DataOps, an agile and collaborative data management practice focused on improving the communication, integration, automation, observability and operations of data flows between data managers and data consumers.

Data management platforms are evolving rapidly. Consequently, data engineers are involved in various efforts to replatform their data, including architecting, designing and migrating data integration pipelines. As they architect, design and develop these data integration pipelines, they need to address a number of challenges, such as agility, a data-centric approach for development and a framework-agnostic solution to define scalability and latency requirements. Some of the other data engineering practices to drive data integration pipelines and workloads include:

- Devise a configuration-driven approach for building end-to-end data pipelines, allowing customization across the different components of the pipeline. Create reusable abstract functions/steps that can accept parameters. Decouple development and deployment using metadata and configuration-driven architecture.

- Build an abstraction layer, and slot tools into the data engineering pipeline based on proper functionality. This abstraction approach makes it easier to add, replace and test components in a data pipeline.

- Enable refactoring of common elements into libraries to support reusability and time to market, and remove incidental complexities by providing standardized, reusable solutions.

- Automate manual and repetitive tasks by using declarative semantics and workflows. Automation is necessary for data engineering organizations and should be done at each layer of the tasks to allow for seamless repeatability, reproducibility and rollback.

- Ensure workloads can run on any environment with a containerized and Kubernetes-based approach.

- Apply a serverless approach to developing data systems where startup latency and operational costs permit. Organizations should apply DataOps and should not expect data engineers to spend time managing and deploying infrastructure.

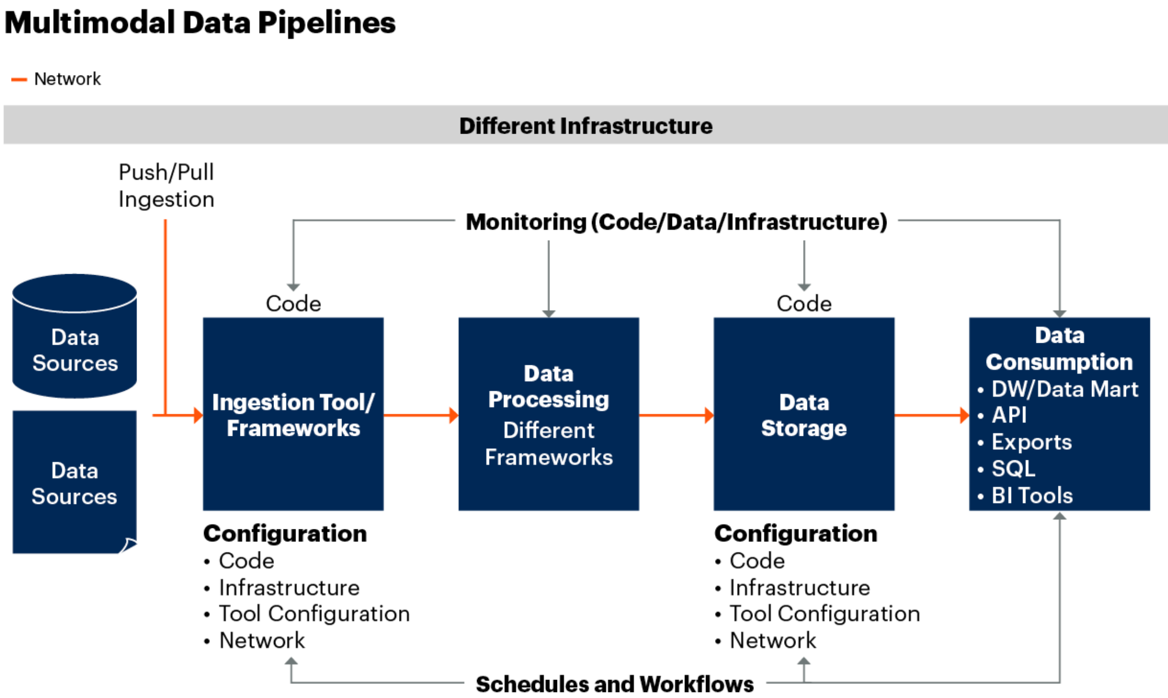
- If necessary, isolate individual pipelines on completely disjointed technology stacks. Several use cases would require such isolation. One is if one pipeline generates real-time data or analyses that you sell to your customers; that pipeline requires high availability, disaster recovery and security far beyond what's needed by your pipeline that supports your internal analytics program. Another example is if different jurisdictions impose vastly different governance or data-residency regulations. In such cases, each jurisdiction might require its own pipeline, separate from all others.

- Leverage data-as-code tools like Project Nessie for versioning the data, and reduce overhead of replicating data across development, testing and production, such as with Dremio Arctic.

By 2025, data engineering teams guided by DataOps practices and tools will be 10 times more productive than teams that do not use DataOps.

Figure 5 summarizes the data engineering life cycle. To learn more about applying data engineering principles and frameworks to modernize your data platforms, see Data Engineering Essentials, Patterns and Best Practices.

Figure 5: Data Engineering — Multimodal Data Pipelines



Source: Gartner
BI = Business Intelligence; DW = Data Warehouse
784750_C

Planning Considerations

Given this trend, focus your 2024 data integration efforts on the following activities:

- Comprehend and overcome data quality challenges in data pipelines.
- Manage data pipeline orchestration and workflows across multiple deployments.
- Develop your data integration architecture in multicloud deployment.
- Adopt data observability practices to deliver trusted data management.

Comprehend and Overcome Data Quality Challenges in Data Pipelines

Data quality is an integral part of the data management pipeline. Typically, data quality tools are deployed alongside other tools, such as data governance, MDM and data integration tools. Data quality is usually an enterprise initiative and is sometimes linked to the overall data strategy and data governance process. Achieving data quality as an organizational objective involves leveraging the entire data architecture to deliver on key data quality dimensions. Also, it is crucial to understand how evolving technologies, such as LLM, impact data quality initiatives. For more details, see Quick Answer: How Will LLMs Impact Data Quality Initiatives?

Common data quality challenges include

- Aligning your data that is fit for purpose and ensuring that you have appropriate access for acquisition
- Identifying the level of preprocessing required to cleanse and normalize the data
- Validating the data to ensure that it can be appropriately tested to meet the requirements
- Monitoring the data for changes, drift and other issues that require further analysis and remediation

Modern data quality tools utilize AI/ML capabilities to establish the relevance and accuracy of data while reducing user oversight and interaction. This augmentation greatly improves user productivity while allowing data consumers to focus on getting value from their data.

To learn more about how to overcome some of the common data quality challenges, read Overcoming Data Quality Risks When Using Semistructured and Unstructured Data for AI/ML Models.

Manage Data Pipeline Orchestration and Workflows Across Multiple Deployments

Managing the data pipeline will require a logical grouping of activities to support both batch and real-time interactive data processing. As organizations start executing workloads across hybrid environments and multiple cloud service providers (CSPs), orchestrating the data pipeline via a centralized, unified interface will become a critical element. Several vendors are working on creating an enhanced user experience for managing and running both intercloud and multicloud workloads.

Orchestration of data pipelines that ingest, prepare, transform and analyze data before it is consumed can be manual or automated. These pipelines should be developed such that they can be maintained, versioned, tested and integrated with other tools. However, orchestration requires scalability, including the ability to programmatically create, schedule and monitor data flows using a data pipeline.

Apache Airflow, a data orchestration tool, is becoming more mainstream across various cloud providers and runs on-premises. In addition, two open-source options, Kubeflow and MLflow, are extending the concept of automating pipelines to machine learning workflows.

As a data and analytics professional responsible for data orchestration tasks, you should investigate opportunities to incorporate data pipeline orchestration options. For example:

- Explore orchestration products for use cases such as migrating to the cloud, managing Apache Spark jobs, and retrying automatically in the event of failures.
- Embrace the use of orchestration in your existing ecosystem to automate tasks and address dependencies on data integration tasks.
- Develop your data orchestration with a clear governance framework and standards to meet data security and compliance requirements.

With the data ecosystem constantly evolving, determining how best to operationalize and orchestrate your data pipelines will become increasingly challenging. For guidance, see How to Operationalize Data Workloads and Building Data Orchestration and Workflows.

Develop Your Data Integration Architecture in Multicloud Deployment

Modern data management platforms require deploying various data integration use cases and styles across complex environments, including data sources, storage and compute, to enable data processing and access in multiple clouds and on-premises locations. Multicloud and hybrid environments have increased the complexity of implementing data integration pipelines. Multicloud architecture often becomes independent of data integration concerns, even though such architecture decisions and designs significantly impact integration patterns.

Some of the most common data integration use cases involve combining data pipeline patterns and mapping them to the different integration styles.

Some of the basic patterns include:

- **Data ingestion:** Data is ingested into the pipeline. Data may be from batch, streaming, Internet of Things (IoT) or other application sources.
- **Data consistency:** Different databases or applications agree to share data collectively, rather than maintain it independently. An example use case is to address overlapping or inconsistent customer data.
- **Multistep process:** Independent applications collaborate to automate a business process. An example is the straight-through processing of shipments for goods ordered.
- **Composite services:** New applications consume APIs or data from other applications. An example use case is to create a single interface for purchase order approval.

These patterns can be matched with the integration styles below to support broader data integration use cases:

- **Data-centric:** Primarily batch-driven with a "data model" worldview. The focus is on datasets and larger data unit sizes.
- **Event-centric:** Streaming data with an "event" worldview. The focus is on streams of data and smaller data unit sizes.
- **Application-centric:** Typically API-delivered data with a "process" worldview. The focus is on moving data between applications as processes with smaller data unit sizes.

It is also important to assess these patterns against additional dimensions, such as data persistence, data attributes, data movement and data complexity. For more details, see Implementing Multicloud Data Integration.

Adopt Data Observability Practices to Deliver Trusted Data Management

"Data observability" applies the observability concept from the DevOps world to the context of data, data pipelines and data platforms (i.e., DataOps), which seeks to reduce the friction around the consumption of data across the organization. It considers data issues an engineering problem and strives to empower data engineers to provide accurate and reliable data to consumers and applications within expected time frames. Data observability uses automation to identify data quality issues, prevent downstream data issues, and augment performance management, capacity planning and production management. Additionally, it allows IT and business leaders to have a degree of control over data usage and capacity planning. The most important underlying features supporting data observability are "signal collections" to drive correlation and relationship analysis using data profiling, data monitoring, anomaly detection, active metadata and data lineage.

To learn more about the data observability landscape and how to prepare your organization to adopt those practices to deliver trust, see Deliver Trust by Adopting Data Observability Practices.

Comprehensive Deployment Practices Will Influence Data Management Adoption

Some practices and technologies are so far-reaching that they influence many aspects of contemporary business and technology. These influences can be felt during all phases of the SDLC, at all phases of the data life cycle, and at all layers of a typical application stack. The influences can alter best practices, desirable design patterns, the scope and coordination of various aspects of governance, and the management of fundamental forms of data that are used everywhere.

Planning Considerations

Some initiatives and practices can influence many aspects of your data architecture. You should appreciate the far-reaching nature of these initiatives and practices:

- Embrace open source and open standards to future-proof investments.
- Augment your cloud deployment with cloud financial management (including FinOps) to optimize costs.
- Use generative AI cautiously and monitor its output scrupulously.
- Coordinate data management design and practices with MDM.

Embrace Open Source and Open Standards to Future-Proof Investments

Open source is the prevalent software model for open innovation efforts in the digital economy. Many enterprises seek open-source standards as an alternative to developing homegrown software or licensing software from third parties. They also see open source as a way to avert vendor lock-in.

Open source allows enterprises to tap into a wider pool of innovative talent and access software features faster from public repositories. It is widely adopted in several mission-critical solutions, including database management systems (DBMS), DevOps and other analytics platforms. With open source, enterprises also gain the flexibility to choose between self-support, commercial support or managed service options in the cloud.

To be successful with open source, you must identify its strategic importance to your business, enforce policies for effective governance and communicate its value to various stakeholders. Any open-source effort needs to be tackled on an organizationwide basis, with participation from leadership across enterprise architecture, engineering, security and risk, infrastructure and operations, and sourcing.

Another way to future-proof your data management infrastructure is to think of your public cloud providers as strategic partners. Because of the proprietary nature of these providers' data-related PaaS services, it is not realistic to attempt a vendor-neutral approach with public cloud platforms. Consequently, you can future-proof your data management investment by selecting a leading cloud platform — a platform that is highly reliable and innovative — as your strategic platform.

Enterprises that use this approach typically standardize on one primary cloud vendor for the majority of their operational solutions, analytics solutions and data-centric pipelines. In addition, they engage a second (and maybe a third) cloud platform for a small number of other, unrelated solutions and workloads, to avoid too great a dependency on a single cloud platform.

To future-proof data management investments further, explore the following recommendations:

- Familiarize yourself with open-source pricing models in the cloud, including charges for compute and storage resources. Many cloud vendors use open-source software as the foundation for paid managed services, with multiple service tiers and options.

- Use standards that are open or provider-neutral. As much as possible, cloud solutions should rely on widely accepted standards, such as SQL, RESTful APIs, JavaScript Object Notation (JSON) or Apache Parquet. Do not overlook specialized standards that apply to specific portions of the data and analytics architecture, such as Open Neural Network Exchange (ONNX) or Predictive Model Markup Language (PMML).

- Understand the options for open-source data stores. From a data management standpoint, the earliest open-source options were relational databases, such as MySQL and PostgreSQL. The possibilities have since expanded. Apache Hudi, Delta Lake and Apache Iceberg for data lakehouse. MongoDB and Couchbase offer document storage. Wide-column databases, such as Apache Cassandra and Apache HBase, are also available to support large datasets. (For more details on different types of data stores, see Comparison of Data Stores to Support Modern Use Cases.)

- Adopt open-source metadata standards like Egeria and OpenLineage. These standards aim to make metadata shareable across platforms in an enterprise environment. They allow organizations to maintain a wider variety of tools and applications that might otherwise be unable to connect and share their metadata easily.

- Accept that, by itself, open-source community support is insufficient. Community support may help identify the problem, but not necessarily a fix to address it. Therefore, you should approach community support with caution. Always have a contingency plan with commercial support and/or managed services options.

- Conduct separate risk assessment of each open-source software as it might have inherent limitations/issues in term security (e.g., recent Apache Log4J issue), HA/DR or enterprise level support.

Augment Your Cloud Deployment With Cloud Financial Management (Including FinOps) to Optimize Costs

D&A governance exists alongside other forms of governance, including governance for IT, HR, operations and finance. Certain situations will require a coordinated response from several of these programs. For example, promoting a self-service report from an individual analyst's workspace to a shared location will require coordination of the following:

- D&A governance (e.g., updating the business data glossary)

- IT governance (e.g., performing a code review of the calculations in the report)

Likewise, some data governance processes must coordinate with financial governance. Consider the execution of runaway queries. In on-premises ecosystems, such queries would exhaust existing resources, perhaps leading to a crash. But in a pay-as-you-go cloud ecosystem with elastically provisioned resources, the query might run to completion, ultimately yielding a pauperizing end-of-month bill from the service providers. This possibility should remain front of mind for professionals participating in data management or D&A governance.

As a technical professional participating in D&A governance, you should:

- Coordinate IT governance with D&A governance where appropriate. Expect governance processes for each phenomenon to interact.
- Ensure your D&A governance program explicitly addresses financial considerations. For example, such governance can include monitoring the cost of operations throughout the data pipeline. It can also include collecting cost data during proof-of-concept investigations and pilot implementations.
- Implement data observability tools to monitor not only the data but also the usage costs. Monitor frequently; do not wait for end-of-month spending reports.
- Use the features provided by CSPs to limit runaway resource consumption caused by poorly designed queries and applications.
- Establish collaboration between the IT and business teams to find the right balance between performance and cost.

Use Generative AI Cautiously and Monitor Its Output Scrupulously

Gartner observes a headlong rush to capitalize on large language models (LLMs) and generative AI techniques. The enthusiasm — and the fear of being left behind by competitors — entices many organizations to pursue these techniques heedlessly. Sensible organizations will proceed with caution. You should proceed with GenAI only with the official approval of strategic leadership, and only with formalized processes for governing GenAI usage. Remember that some vendor products include features powered by GenAI; use of such features should also be governed accordingly.

In particular, organizations using LLM will be vigilant for the following problems:

- Hallucinations
- Over-reliance on conventional wisdom

- Failure to account for changing technology landscape
- Failure to account for rapidly changing business conditions
- Automation bias
- Loss of human justification for and understanding of design decisions

Hallucinations: Hallucination is a high-profile problem in part because it is so serious and in part because the concept is charmingly (or disturbingly) anthropomorphic. LLM hallucinations fall along a spectrum, from the blatantly false assertions that are not supported by facts to assertions that are false because they rely on some facts and ignore others.

Over-reliance on conventional wisdom: These facts come from three sources: public data, private data controlled by the makers of the LLM and local data controlled by your organization (the organization using the LLM). The largest of these is public data — the corpus of information that is widely and publicly available. This data is subject to mass delusion. Over-reliance on this data can cause an LLM to reinforce this mass delusion even if — or especially if — that delusion is considered conventional wisdom.

Failure to account for changing technology landscape: Even if the conventional wisdom is not delusional, it still might have a finite shelf life. That is, recent technological breakthroughs might fundamentally change long-standing realities. A historical example: For decades, long-standing conventional wisdom held — accurately — that isolating databases for write-heavy operational/transactional workloads from databases for read-heavy analytical workloads was a necessary best practice. But then, along came SAP HANA (High-performance ANalytic Appliance) and other products classified as hybrid transactional analytical processors (HTAP), and the conventional wisdom changed. Admittedly, LLMs have access to facts about HTAP. But that's only because HTAP is no longer new. For any technology or practice that is very recent, LLMs might not have access to facts about it.

Failure to account for changing business conditions: The technology landscape can change quickly, but business conditions can change even faster. If you ask an LLM to draft a competitive strategy for you, the output is unlikely to account for a merger/acquisition that your two biggest competitors announced ten minutes ago.

Automation bias: Automation bias is the unexamined faith in technology solutions — because those solutions are technological. [3] You should use LLMs to catalyze or supplement your own thinking, not to replace it.

Loss of human justification for and understanding of design decisions: If you accept the recommendations of LLMs uncritically, the long-term consequences can be unpleasant. Admittedly, the long-term consequences are unknown because LLMs are such a fresh technology. Nevertheless, some negative outcomes are easy to imagine. Consider what happens today when you migrate a data management solution from one platform to another (say from on-premises to the cloud). Inevitably during such migrations, a moment arises when technologists scrutinizing the existing system ask “why did we design the database that way?” Ideally, some long-serving member of the technical staff can provide the answer, because that person remembers the chosen approach, the rejected alternatives and the attendant discussions. In the future, if such design discussions are replaced by an uncritical acceptance of LLM recommendations, the loss of human understanding will be significant and problematic. You won’t understand your own systems.

As a technical professional using LLMs to improve data management and governance, you should:

- Avoid bland, undifferentiated results by feeding the LLM prompts that specify distinguishing details of your problem. For example, don’t ask the LLM to design a curriculum for data literacy. Rather, ask it to design a data-literacy curriculum for a company that uses self-service analytics and whose overall mission is to supply operational support for human clinical trials to pharmaceutical companies specializing in cardiac disorders.
- Minimize the influence of conventional wisdom by supplementing the out-of-the-box LLM with your own well-governed proprietary data.
- Use LLM output to stimulate or catalyze your thinking. Design your own solutions.
- If you operate in an extremely dynamic environment, be vigilant for LLM output that relies on an out-of-date fact base that cannot keep up with constantly changing conditions.

Coordinate Data Management Design and Practices With MDM

Gartner defines MDM as a technology-enabled business discipline in which business and IT work together to ensure the uniformity, accuracy, stewardship, governance, semantic consistency and accountability of an enterprise's official shared master data assets.

MDM represents a concerted effort to guide business users in consolidating and mastering their data and making informed decisions about data usage. Moreover, technology experts need to recognize that master data extends beyond technology itself — it's about dismantling silos and enhancing data quality to achieve meaningful business outcomes for their organization.

MDM's core objective is to optimize human processes and workflows, all underpinned by the underlying technology, and its "business discipline" encompasses data governance and data stewardship. So business leaders exercise concerted discipline by using data governance to establish appropriate policies for creating, storing and utilizing enterprise data in their implementations. Together, they also exercise business discipline through data stewardship that involves applying governance policies within a business context thus ensuring that enterprise data meets the requisite quality standards for business applications.

Generally, technical professionals are responsible for deploying MDM solutions to enable businesses to conduct data governance and stewardship to master their data. Notably, robust metadata plays an essential role in ensuring the proper functioning of MDM and data governance. MDM also identifies core data domains to be governed and used to support the business.

As generative AI becomes mainstream, MDM can leverage AI services for data cleansing, standardization and enrichment capabilities within the MDM stack, which can increase efficiency and accelerate the overall MDM deployment. AI can also help with the automatic mapping of data sources and the enforcement of new data governance policies. We are already seeing some mainstream MDM vendors use generative AI capabilities.

Lastly, MDM initiatives can be complemented with data fabric for accelerated data discovery, data preparation activities for mastering the data and enabling shared data assets across the enterprise. Conversely, data fabric can leverage MDM hubs as the trusted source for business metadata and master data.

To understand more about how to implement the technical architecture and practices for MDM, please refer to Implementing the Technical Architecture for Master Data Management.

Evidence

[1] **2022 Gartner State of Data and Analytics Cloud Adoption Survey:** This survey was conducted to validate and understand how end-user organizations are practicing and planning their journeys to perform data and analytics in the cloud, and what the key drivers are for such a transition. The research was conducted online from October through November 2022 among 461 respondents from North America, EMEA and Asia/Pacific. The respondents were from the manufacturing, natural resources, healthcare provider, banking and finance, IT, retail and wholesale, government, education, media and communications, transportation, and utility industries, in organizations with more than 250 full-time employees. The respondents were screened for being in director roles and higher, having responsibility for adopting cloud or planning it, and having some visibility or involvement in financial decision making for cloud adoption. Disclaimer: Results of this survey do not represent global findings or the market as a whole, but reflect the sentiments of the respondents and companies surveyed.

[2] **2021 Gartner Data and Analytics Governance Survey** was conducted online from 12 July through 22 July 2021 to test our assumption that organizations with distributed, business-outcomes-based governance achieve better business results than centralized, IT/D&A-led initiatives. In total, 105 IT and Business Leaders Research Circle members participated. Fifty-seven were from Gartner's ITL Research Circle — a Gartner-managed panel — and 48 were from an external sample. Members from North America (51%), EMEA region (35%), Asia/Pacific (3%) and Latin America (11%) responded to the survey. Respondents were qualified based on their involvement and participation in decision making for data and analytics governance at their organizations.

[3] In his delightfully cranky book *Save Everything, Click Here*, Evgeny Morozov labels this phenomenon "technological solutionism."

Document Revision History

2023 Planning Guide for Data Management - 10 October 2022

2022 Planning Guide for Data Management - 11 October 2021

2021 Planning Guide for Data Management - 9 October 2020

2020 Planning Guide for Data Management - 7 October 2019

Recommended by the Authors

Some documents may not be available as part of your current Gartner subscription.

Migrating Enterprise Databases and Data to the Cloud

Assessing the Relevance of Data Virtualization in Modern Data Architectures

Best Practices for Migrating Data and Analytics Governance to the Cloud

Assessing Distributed SQL Databases

Data Engineering Essentials, Patterns and Best Practices

Graph Technology Applications and Use Cases

A Guidance Framework for Deploying Data and Analytics in the Cloud

How to Operationalize Data Workloads
