Hype Cycle for Data Management, 2023

Published 26 July 2023 - ID G00792590 - 144 min read

By Analyst(s): Aaron Rosenbaum

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

Accelerated cloud adoption continues to enable enormous innovation across data management technologies, with some new disruptors like generative AI now entering the mix. Data and analytics leaders can use this research to strategize their system and architecture roadmaps and investments.

Analysis

What You Need to Know

Data and analytics (D&A) leaders must cope with the evolving requirements of digital business, the deepening impact of cloud and increasing data ecosystem complexity. Today, data is rapidly moving to the cloud. This migration is creating opportunities for innovation and accelerating the pace of change across many aspects of data management.

To address these challenges, new entrants on the Hype Cycle reflect two key themes: the push to distribute responsibility and accountability for data management across lines of business and domains, and the impact of generative AI on data management.

D&A leaders should use this Hype Cycle to identify promising technologies and practices, plus decide when it is appropriate to evaluate them for adoption. Some technologies are less mature but offer significant innovation or differentiation. Others have recently matured and are near ready for mainstream use.

Gartner provides several Hype Cycles that cover data management, analytics and related fields. Together, these Hype Cycles draw a holistic view of the emergent technologies and practices across the entirety of data and analytics:

- Hype Cycle for Data & Analytics Programs and Practices, 2023
- Hype Cycle for Data and Analytics Governance, 2023
- Hype Cycle for Analytics and Business Intelligence, 2023
- Hype Cycle for Data Science and Machine Learning, 2023
- Hype Cycle for Artificial Intelligence, 2023
- Hype Cycle for Data Security, 2023
- Hype Cycle for Privacy, 2023

The Hype Cycle

A Hype Cycle uses both current level of adoption and number of years to mainstream adoption to measure the maturity of technologies. With this in mind, we can see various levels of maturity (and various kinds of change) in this year’s Hype Cycle. Here are some highlights.

Changes in the 2023 Hype Cycle

New entrants: The 2023 Hype Cycle for Data Management includes four new entrants on its Innovation Trigger. (See the left side of Figure 1.) These are:

- Self-service data management
- Data product
- Generative AI for data management
- Vector database

Generative AI for data management and data product are both identified as transformational.

Maximum hype: At or just over the Peak of Inflated Expectations are some of the most hyped, albeit less mature, technologies in data management today. These include augmented data quality, lakehouse, distributed transactional databases and intercloud data management. Lakehouse, which just entered the Peak last year, has rapidly achieved significant market penetration despite its relative immaturity as a technology. Data mesh, despite its obsolescence and immaturity, continues to grow in hype.

Entering the Trough of Disillusionment: Several technologies and disciplines around future data architectures are currently heading into or passing through the trough: DataOps, data fabric, active metadata management, and augmented data cataloging/metadata management solutions. Several technologies, such as operational intelligence, graph DBMS, knowledge graphs, data engineering and augmented data management, are poised to enter the plateau within the next five years.

Plateau of Productivity:

- SQL interfaces to object stores moved into the plateau rather quickly. Originally a feature of emergent, primarily open-source, SQL engines, it's rare to find a SQL engine that cannot query data directly from object storage.
- Multimodel, also once a specialized niche within the database market, has become the standard for almost all databases. While different databases continue to have different levels of functionality and performance for different models, "relational model only" is now an exception rather than the majority of the market.

Obsolete before plateau:

This evolution to “obsolete before plateau” is frequent for database features and doesn’t reflect on the utility of these models, but the evolution of the products in that category:

- Data mesh: While it continues to grow in expectations, we expect the core capabilities of data mesh to be subsumed by data fabric.
- Private cloud database platform as a service (dbPaaS) as stand-alone technology: Both vendors and end users have primarily shifted to “public cloud first.”
- Ledger databases: The use of both immutable object storage and distributed ledgers, without database functionality, have found acceptance and maturity in the market.
- Wide-column databases: The technology has evolved into a feature of multimodel databases rather than a stand-alone technology.

The Current State of Data Management Maturity

This year’s Hype Cycle for Data Management has a wide distribution of technologies, reflecting an extremely high pace of innovation and change that has accompanied the migration to the cloud. The cloud makes it easier to innovate with very new technologies and also accelerates adoption by enterprises and maturity by vendors, moving technologies through the Hype Cycle more quickly. Lakehouse and DataOps, for example, moved significantly since last year.

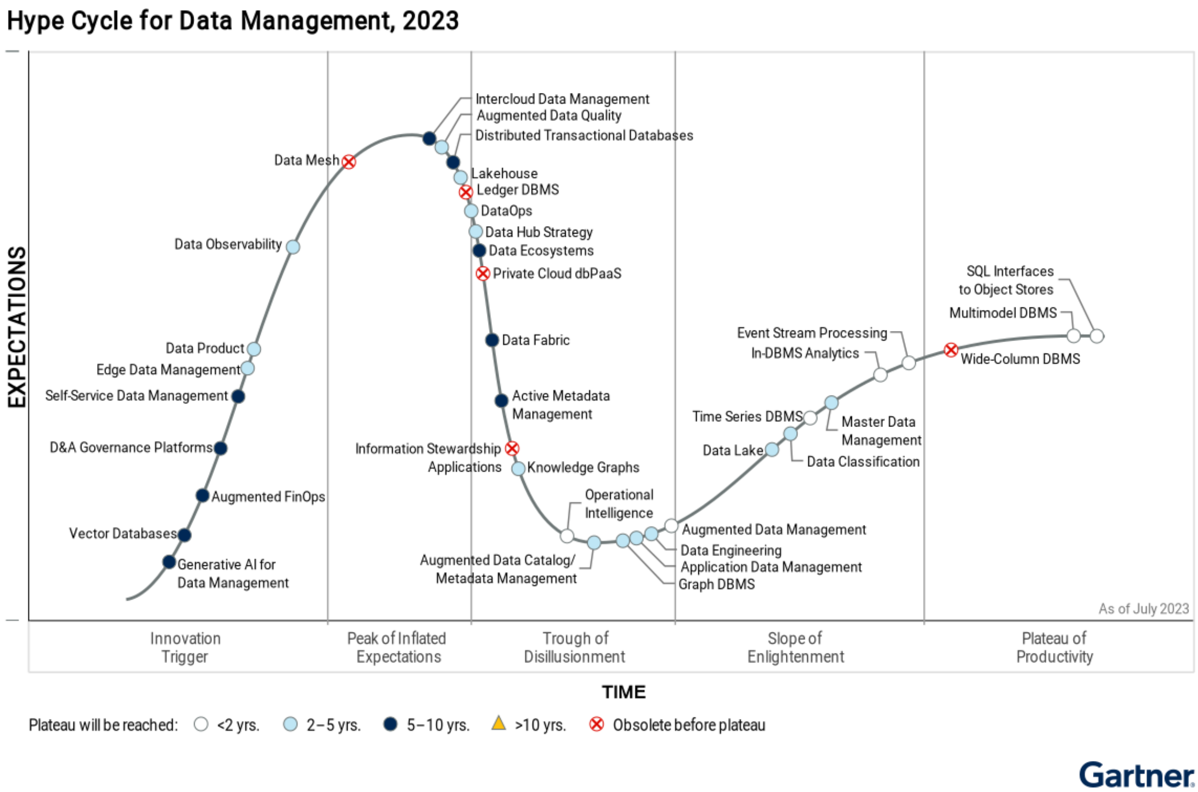
Some of the areas within data management that are driving innovation and pushing the need to mature rapidly include:

- **Data ecosystems:** The whole environment for data to be used in analytics, business intelligence, AI and machine learning is changing. The desire for less complex management of disparate data sources has led to the concepts of data ecosystems and data fabric. These concepts are being realized in technologies such as lakehouse, augmented catalogs and DataOps, all of which are advancing to useful products very quickly. Of course, all are tied together with metadata management.

- **Generative AI (GenAI):** While very early, the potential for transformational impact is high. The benefits of GenAI, if realized, are significantly disruptive to many current practices within data management. The GenAI market is, however, very immature and extremely fast-moving, so any near-term deployment should be tempered by appropriate risk management.

- **Data management operations:** Due in part to the growth in cloud along with intercloud and edge computing, operations management for data management is changing. The need to manage and govern across disparate platforms is increasing the need for DataOps, data engineering, augmented FinOps and more.

Figure 1: Hype Cycle for Data Management, 2023



The Priority Matrix

The Priority Matrix for Data Management, 2023 positions each technology tracked in this Hype Cycle according to two dimensions — business benefit, and years to mainstream adoption and maturity. In the matrix, benefit is quantified as transformational, high, moderate or low.

Transformational benefit: The adoption of some hyped technologies could dramatically change how data management is performed, as well as transform organizations that depend on data management. Two new entrants in 2023 with transformational benefits are generative AI for data management and data products.

Similarly, data fabric, enabled and closely linked to active metadata management, promises to elevate data management out of its silos of disconnected tools and datasets to an unprecedented level of integration, interoperability and innovative business applications. Although data fabric is still five to 10 years to the plateau, its pace has accelerated over the last year.

High benefit: Many of the data management technologies (the majority of which will be mainstream in five years or less) promise a high level of business benefit. In particular, the technologies described as “augmented” or “active” will automate many manual tasks and make it easy for data engineers to optimize those tasks as well. Data ecosystems fall in this category and will reduce the integration necessary across multiple tools, especially in the cloud. Many of those data ecosystems are built around the lakehouse, which we have moved up to high benefit this year. Of note, intercloud data management has advanced over the peak in the past year and offers significant benefits to many enterprises. It’s notable that intercloud data management offerings are advancing from both cloud providers and independent service providers.

Moderate benefit: With the exception of graph DBMS, other DBMS types fall in the moderate category and are on or approaching the plateau; this is due to the ubiquity and maturity of multimodel databases, including the functionality of many of these DBMS types.

Low benefit: Because of the overall criticality of data management, we consider all data management technologies to deliver at least moderate benefit.

Table 1: Priority Matrix for Data Management, 2023

(Enlarged table in Appendix)

Benefit	*Years to Mainstream Adoption*			
↓	*Less Than 2 Years* ↓	*2 - 5 Years* ↓	*5 - 10 Years* ↓	*More Than 10 Years* ↓
Transformational	Event Stream Processing	Augmented Data Quality Data Product	Active Metadata Management Augmented FinOps Data Fabric Generative AI for Data Management	
High	Augmented Data Management In-DBMS Analytics Operational Intelligence	Augmented Data Catalog/Metadata Management Data Classification Data Engineering Data Hub Strategy Data Observability DataOps Edge Data Management Graph DBMS Knowledge Graphs Lakehouse Master Data Management	D&A Governance Platforms Data Ecosystems Intercloud Data Management Vector Databases	
Moderate	Multimodel DBMS SQL Interfaces to Object Stores Time Series DBMS	Application Data Management Data Lake	Distributed Transactional Databases Self-Service Data Management	
Low				

Source: Gartner (July 2023)

Off the Hype Cycle

In 2023, a few innovations have changed or left the Hype Cycle for Data Management:

- **Fully mature innovations:** Data integration tools and data preparation tools attained maturity on the Plateau of Productivity, and they have now moved off this Hype Cycle into mainstream adoption.

- **Dropped from the Hype Cycle:** Data discovery and management, as it relates to data management, is now combined with augmented data catalog/metadata management.
- **Name changes:** Augmented data cataloging and MMS is now called “augmented data catalog/metadata management.” Cloud data ecosystems has been renamed “data ecosystems.” “Augmented transactions” has been renamed “operational intelligence.”

On the Rise

Generative AI for Data Management

Analysis By: Roxane Edjlali, Rita Sallam

Benefit Rating: Transformational

Market Penetration: Less than 1% of target audience

Maturity: Embryonic

Definition:

Generative AI technologies can generate new derived versions of content, strategies, designs and methods by learning from large repositories of original source content. Generative AI has profound business impacts, including on content discovery, creation, authenticity and regulations; automation of human work; and customer and employee experiences. GenAI transforms data management activities through natural language interfaces, making data management activities more widely accessible.

Why This Is Important

Although leveraging AI and ML as part of augmented data management is not new, leveraging generative AI products such as Azure OpenAI or Google Codey may transform the data management discipline, making self-service data management activities accessible to much larger audiences. Through integration with active metadata management tools and enrichment from semantic tools and knowledge graphs (by training LLMs on enterprise or industry-specific corpora), it will become possible to have conversational interfaces to perform data management tasks. This will not only reduce the skills barrier to data management but will also increase the productivity of data management specialists.

Business Impact

The Gartner 2023 CDAO survey has again identified availability of skills as a recurring issue for data and analytics. In parallel, business is demanding greater flexibility and agility for accessing data, which is driving a decentralization of data management activities. Using conversational interfaces will reduce the pressure on highly skilled roles such as data engineers and ease the job of data stewards and even DBAs. This can result in a transformation of the organizational models, roles and data management practice.

Drivers

- Allow organizations to innovate and implement data and analytics use cases faster.
- Open data management activities to new roles by reducing the skills barrier.
- Increase productivity of data management specialists.
- Allow to deploy data pipeline, monitor usage and perform CI/CD activities through natural language.
- Offer data stewards the option to express data quality rules in natural language and have the tool generate the corresponding scripts or pipelines automatically.
- Generate code for databases, scripts or APIs to interact with applications via natural language. This will simplify the work of developers, data integration specialists and DBAs, but also offer the possibility (through prompt engineering) to have increasingly system-to-system conversations.
- Simplify the work of data engineers by creating pipelines and data management workflows based on natural language.
- Enable augmented capabilities for testing, validation, error detection, schema drift detection and correction, PII detection and masking, and impact analysis reporting through interfaces that make it easier for SMEs to become a part of the data management and engineering journey.

Obstacles

- Generative AI techniques applied to data management are still in their infancy. Vendors are still in the very early stages of using this technology to power their data management tools. But with the vast investments in the area, we expect a rapid progression of product availability.
- Although generative AI can speed up the creation of new data pipelines or generate code, validation and execution, but ensuring security and access rights management will be left to the experts.
- Generative AI will be constrained by data and analytics governance, regulatory, and data security considerations, and will require human supervision.
- Generative AI is creating considerable fear and uncertainty in the market in general. This may delay adoption of these new capabilities by risk-averse organizations.
- Adoption will require dedicated training, education and development of new skills such as prompt engineering.

User Recommendations

- Monitor adoption of generative AI as part of your data management tool portfolio.
- Monitor product developments from your key data and analytics product suppliers.
- If you are selecting a cloud data ecosystem from a single cloud provider, consider generative AI capabilities across overall data management as part of that evaluation.
- Encourage your most curious team members now (data engineers, data quality engineers, DBAs or data analysts) to experiment with generative AI and explore the possibilities.
- Identify potential use cases that are being slowed down by engineering bottlenecks and the lack of business involvement. Those would be ideal use cases in data and analytics to investigate generative AI for data management.

Sample Vendors

Amazon Web Services; Databricks; Google; Informatica; Microsoft; PingCAP; Snaplogic

Gartner Recommended Reading

Innovation Insight for Generative AI

Innovation Insight: Data Ecosystems Will Reshape the Data Management Market

Innovation Insight: State of Data Management Support for Self-Service

Vector Databases

Analysis By: Arun Chandrasekaran, Radu Miclaus

Benefit Rating: High

Market Penetration: Less than 1% of target audience

Maturity: Embryonic

Definition:

Vector databases store numerical representations of data. In such databases, each point is represented by a vector with a fixed number of dimensions, which can be compared via mathematical operations, such as distance measures. Vector databases are commonly used in machine learning (ML) solutions, where vectors represent data features/attributes, such as text embeddings. Storing these vectors in a database enables users to search for similar data points with low latency.

Why This Is Important

Vector databases serve such use cases as similarity search and product recommendation. Rapid innovation in generative AI and adoption of AI foundation models have spawned interest in vector databases. When customers adopt generative AI models, vector databases store the embeddings that result from the model training. Storing vector embeddings representing the model training, the database can do a similarity search, which matches a prompt (the question) with specific or similar vector embedding.

Business Impact

Businesses thrive by delivering differentiated customer experience (CX). Generative AI is increasingly embedded in applications to empower the human-machine symbiosis, and organizations need scalable and accelerated ways to build and support these applications long term. Vector databases are an important back-end service that allows businesses to future-proof and scale their generative-AI-enabled applications. These drive business value through customer engagement and adoption.

Drivers

- **Popularity of vector embeddings:** With the rise of AI foundational models, embeddings have become the cornerstone for semantic search. Hence, they are the working inputs for training large foundation models.

- **Performance and scalability needs:** The applications looking to embed generative methods that use embeddings-based models need back-end services that can respond with low latency to high concurrency requests (prompts) and responses (completions) for generative AI use cases.

- **Service architecture:** Because most applications are built on service-based architectures, vector databases are ideally presented to applications as services that communicate with the interface via APIs.

- **Hybrid implementation of retrieval and generative models:** Vector databases are optimal for both semantic search (retrieval based on vector similarity) and generative inference through foundation models. This hybrid combination of models drives the need for optimized vector databases, because both generative and retrieval are used together for grounding of facts.

- **Developer focus:** Developers of new applications are driving the demand for vector databases by presenting use cases that cannot scale without the ability for embeddings to be stored in an optimized structure for high-throughput production applications.

Obstacles

- Enterprises lack an understanding of what vector databases do and the unique use cases they enable.
- Vector databases are superspecialized databases that may cause challenges around data migration and integration and limited extensibility across use cases.
- Most vector databases are delivered as cloud-managed service — the complexity of deploying, configuring and operating them outside cloud environments requires deep technical skills and know-how.
- Vector databases can be expensive to implement, given the newness of the technology and lack of industry skills to deploy and manage it.
- The vector database market is nascent and populated mostly by startups, which may not have extensive experience working with enterprise clients, as well as unproven product market fit.

User Recommendations

- Determine whether your functional requirements can be satisfied by incumbent vendors that can support the storage and retrieval of vector embeddings — you may not always need a purpose-built vector database.
- Prioritize developer experience, ecosystem integration, use case fit, reliability and performance as important selection criteria, and validate them thoroughly via a POC process.
- Select managed, cloud-based vector databases as deployment modes, unless you have stringent requirements and deep technical skills for an on-premises, self-managed deployment mode.
- Conduct internal training and education on the appropriate use cases for vector databases, how to leverage their true potential, and effective ways to optimize their deployment and maximize their value.

Sample Vendors

Couchbase; Croma; Elastic; Google; Pinecone Systems; Qdrant; Redis; Weaviate; Zilliz

Gartner Recommended Reading

Innovation Insight for Artificial Intelligence Foundation Models

Quick Answer: What Is GPT-4?

Executive Pulse: AI Investment Gets a Boost From ChatGPT Hype

How Large Language Models and Knowledge Graphs Can Transform Enterprise Search

Augmented FinOps

Analysis By: Adam Ronthal, Dennis Smith

Benefit Rating: Transformational

Market Penetration: Less than 1% of target audience

Maturity: Embryonic

Definition:

FinOps applies the traditional DevOps concepts of agility, continuous integration and deployment, and end-user feedback to financial governance, budgeting and cost optimization efforts. Augmented FinOps automates this process through the application of artificial intelligence (AI) and machine learning (ML) practices — predominantly in the cloud — to enable environments that automatically optimize cost based on defined business objectives expressed in natural language.

Why This Is Important

In the cloud, it is now possible to assess the cost of a specific workload or collection of workloads assigned to a project. However, price/performance — the primary measure of cloud efficiency — is difficult to assess due to the complexity and diversity of choice in underlying cloud infrastructure and service offerings and a lack of consistency in pricing models. Augmented FinOps can automate this process by applying AI/ML techniques.

Business Impact

The automation of cloud budget planning and financial operations will allow businesses to express their objectives — ideally in natural language — and allow their cloud ecosystems to automatically optimize the underlying cloud resources to meet those objectives. This will result in more efficient use of resources and, therefore, optimal spend by reducing/eliminating misaligned or poor use of cloud infrastructure and service offerings.

Drivers

- Practitioners are increasingly realizing that cloud is fundamentally a complex cost optimization exercise.
- Cloud adopters have a strong desire for transparency into cloud spending.
- Buyer inexperience is leading to either under-provisioning and associated resource contention or overprovisioning and spending more than is needed.
- Vendors are positioning cost-effectiveness as a competitive differentiator in their go-to-market strategies.
- Practitioners need to reduce the unpredictability of cloud spending when using cloud infrastructure and services for analytics, operational database management systems (DBMSs), data lakes and other applications, including custom IT infrastructure.
- Consumption-based usage remains common in earlier stages of cloud adoption, driving the need for augmented FinOps, although commit-based usage mitigates some unpredictability.
- Cost overruns are often obscured, downplayed, or dismissed by line of business implementers, requiring augmentation to achieve holistic and comprehensive cost optimization.
- Automation of financial governance controls in cloud environments provides increased predictability and cost optimization with less operational effort.
- Solid financial governance frameworks are positioning organizations to take advantage of FinOps.
- Emergence of specific roles — like FinOps practitioner or cloud economist — focused on FinOps practices and cost optimization means organizations have the expertise to address augmented FinOps.
- Owing to their complexity, cloud environments are ideally suited for the application of ML and AI methods to automate processes and track price and performance.
- Core FinOps capabilities are being delivered in three ways: Homegrown solutions, cloud service provider (CSP) instrumentation and third-party vendors. Increasingly practitioners are seeking out third-party or CSP tools to address their needs. All of these have a broad objective of adopting augmented capabilities as a means of competitive differentiation.

Obstacles

- Cloud service provider pricing models remain needlessly complex and diverse.
- Cloud ecosystems are (and will remain) open to third-party participants, which implies multiple commercial arrangements with multiple providers.
- Standards for cloud cost, usage and billing data like the FinOps Foundation's FOCUS proposal have yet to be broadly adopted. APIs for communicating performance data within the context of a broader ecosystem have yet to emerge. Both of these are required to assess the primary measure of success: price/performance.

User Recommendations

- Seek out service offerings to automate (via AI/ML) performance, consumption and pricing options. Increasingly, incorporate these capabilities into cloud data ecosystems that will learn from consumption patterns as they seek to optimize the underlying resources, and by extension, cloud spending through orchestration and optimization.
- Apply Gartner's FinOps Maturity Model to assess FinOps offerings in terms of their ability to address the following core capabilities: Observe, report, recommend, predict and optimize. The last three introduce augmented FinOps capabilities.
- Plan to use multiple tools to address the full scope of requirements. Many tools are broad in reach, but do not go deep into prescriptive recommendations. Others are tightly scoped and provide very targeted optimizations. Expect to spend time combining multiple tools to achieve broad and deep capabilities.

Sample Vendors

Acceldata; Anodot; Apptio; Capital One Software; Densify; Enteros; Finout; OtterTune; Sync Computing; Unravel Data

Gartner Recommended Reading

How to Identify Solutions for Managing Costs in Public Cloud IaaS

A Guidance Framework for Selecting Cloud Management Tools

Emerging Tech: Data Management Product Leaders Must Implement Augmented FinOps in Their Cloud Solutions

CDAOs and CFOs Must Drive Business Value in the Cloud Through Collaboration

Financial Governance Is Essential to Successful Cloud Data and Analytics

D&A Governance Platforms

Analysis By: Guido De Simoni

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

A data and analytics (D&A) governance platform represents a set of integrated technology capabilities that help govern and steward a range of policies spanning security, quality, access, retention, privacy and ethics. It exposes a user experience for policy setting and enforcement to all relevant participants (e.g., data stewards in business roles, business analysts, line of business [LOB] users, data scientists and governance board members).

Why This Is Important

The most complex governance challenges can no longer be met with siloed approaches. Convergence of capabilities is mandatory. Today, the execution of D&A governance is inconsistent, with different organizations using different types of technology. These disparities impede the success of digital business initiatives. Governance needs have grown more diverse and complex; all aspects of governance for all types of policies can benefit from cohesive technology support.

Business Impact

D&A leaders adopting D&A governance platforms will benefit from:

- An ability to mitigate risk from most complex, cross-organizational governance challenges
- Enhanced productivity and efficiency in governance processes, more rigor in enforcement of policies, and therefore more control and trust in data and analytics
- Emerging augmented data management capabilities that discover data and its relationships to seed and power various governance work efforts

- Converging long-term, discrete markets that will collide into one

Drivers

- Increasing complexity from data sovereignty requirements and digital strategies is forcing organizations to simplify and coordinate governance efforts globally across privacy, security, storage, access, use and sharing.
- Organizations want to have automated, synchronized, integrated, cost-effective and efficient D&A governance solutions with a central design, yet a distributed deployment. This requirement is driven by the growing recognition that the work of data and analytics governance is different from the work of data management, but that augmented data management supports the growth of these platforms of convergence.
- All of these aspects are operationalized, and more efficiency is gained when identification of data sources, curation of data, application of workflow, harmonization, reporting and visualization are provided in a coherent platform with automation. For example, you can address autogeneration of data quality rules using a number of methods. These include rule definitions and automated execution of data quality checks, AI-assisted data curation and association of business terms to technical artifacts, automated classification of sensitive data, and build subject registry.

Obstacles

- D&A governance today is served by discrete markets, each with its own solution. Inertia and sunk costs will slow down the emergence of this newer market.
- The current convergence within data management may not satisfy the needs of organizations across D&A governance.
- Incompatibility between what vendors can support and what different customer environments require will likely necessitate multiple metadata management solutions.
- Data management executes the policy that D&A covernance sets. The work — policy setting, enforcement and execution — is different, so the technology capacities, roles and value propositions of the platforms are different.
- Other obstacles reside in the cultural shift that many organizations must address in leveraging the inherent value of D&A governance. When organizations are committing to data and analytics initiatives aligned to mission-critical priorities, such obstacles can jeopardize the adoption of these platforms as enablers for continuous improvement. We estimate that this innovation will reach the Plateau of Productivity in more than 10 years.

User Recommendations

- Design proofs of concept that will capitalize on the required critical technology capabilities. Identify the relevance of these technologies and their connection to business outcomes as a first step. Then look into their ability to support specific use cases (such as risk management and compliance).
- Minimize the number of tools and solutions deployed by analyzing your strategic approach to D&A governance and by using available market technology capabilities in end-to-end scenarios supported by emerging D&A governance platforms.

Sample Vendors

Alex Solutions; Collibra; data.world; Global Data Excellence; IBM; Informatica; OvalEdge

Gartner Recommended Reading

The Role of Technology in Data and Analytics Governance Policy Management

Market Guide for Data and Analytics Governance Platforms

Tool: Vendor Identification for Data and Analytics Governance Platforms

Data Product

Analysis By: Ehtisham Zaidi, Roxane Edjlali, Aaron Rosenbaum, Robert Thanaraj

Benefit Rating: Transformational

Market Penetration: Less than 1% of target audience

Maturity: Embryonic

Definition:

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.

Why This Is Important

In today's business environment, data and analytics leaders (and their teams) seek faster cycle times for "integrated" and "governed" data that is ready for consumption. However, data engineering teams still struggle with provisioning optimal and trusted datasets for mission-critical use cases. Data products are becoming important, because they provide integrated datasets that are trusted, self-contained, well-governed and certified for reuse. They focus on tackling targeted business use cases in D&A.

Business Impact

Major impacts to the business include:

- Data products enable self-service, quicker response times and business accountability over data.
- Data products are cataloged, discoverable, preintegrated and targeted at specific use cases, enabling domain teams to focus on outcomes, rather than wait for IT to integrate, prepare and share data.

- Successful data products enable reuse. They not only serve as common inputs to internal analytics and data science projects, but also open up new revenue streams through data sharing and data monetization.

Drivers

- **Time to analytics (speed):** Organizations that operate with a product management mindset for data (with associated operating models, roles and skills) are more likely to achieve faster turnaround times for their analytics use cases. Data products include preintegrated data (from data silos) that is approved for use (and reuse) by governance teams. Hence, this data is trusted by consumers, improving the cycle time to analytics delivery.

- **Reuse (scale):** Product management principles support iterative and agile best practices for data product creation, maintenance, operationalization and life cycle management. Thus, data products can be reused and scaled across use cases quickly when needed.

- **Trust:** Data products enable domain teams to get closer to data. Data products provide prepackaged and preintegrated datasets with well-documented schema, semantics, ontology and usage guidance. This allows domain experts, analytics and SMEs to confidently use the data products to support their business outcomes.

- **New revenue streams (monetization):** Monetizing data, by sharing it across data exchanges and marketplaces, requires the creation of data products.

- **DataOps support (agility):** One of the biggest challenges with data pipelines is that they break, and it takes a long time to find issues and fix them. Another challenge is that pipelines are developed through waterfall approaches, leading to painfully long wait times for the business. A critical premise behind data products is that they must support DataOps tools and best practices for agile data integration (CI/CD), change management and issue resolution (such as schema drift).

Obstacles

- **"Data product washing":** Not every integrated dataset is a product, and not every use case needs a data product. Data products must be created to support focused and repeatable business challenges, not one-offs. If created without scalability and reusability in mind, data products may end up being expensive to maintain.

- **Operating model shift:** Data products require a shift in D&A organizational structure and the introduction of product management principles. They also need a formal data product manager role to oversee their life cycle.

- **Governance and value justification:** Organizations struggle to train SMEs on proper governance and use of data products. Moreover, teams struggle to reach consensus on KPIs, which are necessary to measure the success of data products.

- **Poor planning**: Organizations that create one-off data products and don't plan for management of the entire data product life cycle usually struggle with data products not delivering value.

User Recommendations

- Plan for scale but start small. Collaborate with domain teams to identify use cases that are well-established and have the broadest-understood context. These are ideal use cases where data products can be beneficial.

- Begin with business-critical use cases that need to be scaled. Then, correlate them with use cases that are impeded by central IT's delivery bottlenecks.

- Focus on delivering a minimum viable data product that teams will improve upon over time, instead of on perfecting a data product for specialized delivery.

- Ensure that producing and consuming teams reach a consensus on KPIs for measuring data product success.

- Invest in the data product manager role. Besides being skilled in product management, data product managers must be adept at translating business requirements for data engineering teams.

- Train SMEs and analysts on using data products. Ensure that data engineers are trained in DataOps tools and practices to enable agile maintenance of data products.

- Measure and communicate the success of data products to garner continued investment and support from business domains.

Gartner Recommended Reading

Quick Answer: What is Data Mesh?

Data and Analytics Essentials: DataOps

Quick Answer: Comparing Data Fabric and Data Mesh

5 Ways to Enhance Your Data Engineering Practices

Self-Service Data Management

Analysis By: Roxane Edjlali, Ehtisham Zaidi, Michele Launi

Benefit Rating: Moderate

Market Penetration: 1% to 5% of target audience

Maturity: Embryonic

Definition:

Self-service data management is the ability for line-of-business (LOB) professionals to participate in data management disciplines with minimal IT support. These disciplines include data preparation, augmented data cataloging, data quality, data integration and DBMS administration and management.

Why This Is Important

Organizations experience bottlenecks in their data management processes, especially due to an overreliance on their IT department's data engineering or a centralized data and analytics team, which struggle to keep pace. Because of the increased demand for data throughout lines of businesses, self-service is no longer optional, but a necessity. Increased data literacy and augmented capabilities have reduced the skills barrier, growing the demand and support for self-service.

Business Impact

The 2023 Gartner CDAO Survey ranks talent shortages as a major challenge. In parallel, two out of three teams are boosting investments in data management. Self-service data management is powered by augmented data management. It has reduced the skills barrier for data management disciplines, which allows data management teams to free up scarce resources and employ them in other strategic data management tasks, and allows less-skilled business teams to benefit from the more efficient access to data.

Drivers

- The growing need for decentralization of data management activities within the LOBs and distributed domains is driving a growing number of citizen roles to participate in data management.
- Cloud data management services have made possible for lines of businesses to become more autonomous.
- Augmented data management and also upcoming generative AI capabilities will make these disciplines more accessible to new audiences, even for more technical disciplines like DBMS administration.
- Adoption of data literacy programs across enterprises over the last few years is creating both greater demand for data and analytics but also greater base level of competence across the business.
- Emerging data management approaches, such as data mesh, encourage domain-led data management activities.

Obstacles

- Self-service capabilities can introduce risks related to data governance, data quality, data loss, security and privacy. Incorrect data usage or problems with data quality can result in a false sense of confidence, allowing mistakes to compound.
- Data governance maturity is the major obstacle to self-service data management adoption. Organizations have to support and train in adaptive governance principles that balance central data governance with decentralized data governance models. This requires a shift in operating models, something which historically takes a significant amount of time and investment.
- Decentralization of data management activities can increase the proliferation of domain data silos and data product silos, making shareability and reusability of data an issue.
- Even with increased ease of use advancement that comes from self-service data management, availability, distribution of skills and data literacy across disciplines continues to be a challenge.

User Recommendations

- Set up clear policies about who has access to which data, how often and the context in which data can be reused.
- Make ease of use a central selection criterion across all data management tools, as it will benefit both technical and less technical users.
- Rethink organizational models and data governance by balancing a centralized and decentralized approach through adaptive governance practice.
- Ensure that your selection of tools can actually support central data governance and operationalization when needed (ensuring alignment with central IT teams).
- Do not overlook data literacy. Deep knowledge of data management is less needed but a greater data culture among business teams is required to avoid pitfalls.

Sample Vendors

Alteryx; Atlan; Google; Informatica; K2View; Microsoft; Oracle

Gartner Recommended Reading

Innovation Insight: State of Data Management Support for Self-Service

The State of Data Quality Solutions: Augment, Automate and Simplify

Market Guide for Active Metadata Management

3 Ways for CDAOs to Enable an "Integration Always" Approach to Data Management

Edge Data Management

Analysis By: Aaron Rosenbaum

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

Edge data management comprises the capabilities and practices required to capture, organize, store, integrate and govern data outside of traditional data center and public cloud environments. An increasing number of digital business use cases, including those based on IoT solutions, will leverage data in edge environments. This expansion creates tremendous opportunities to optimize resources and drive real-time decisions and actions, but also brings complexity and governance challenges.

Why This Is Important

Valuable data is increasingly generated and used outside of traditional data centers and cloud environments. This data often has a shorter useful life span, requiring value to be captured near the place and time of its origin. This is the role of edge-computing environments deployed closer to assets in the physical world. Edge data management will both impact and enable IT leaders and their teams, requiring new capabilities and skills while also opening up new opportunities to deliver value.

Business Impact

Edge data management creates value in various ways:

- By distributing data management to the edge, data-centric solutions better support demand for local and real-time data.
- More solutions, such as for IoT use cases, must operate in disconnected (or intermittently connected and low-bandwidth) scenarios.

- It enables smarter physical assets and collections of assets, including remote management or autonomous behavior, via edge data.
- It addresses inconsistencies, protection, sovereignty and other governance issues arising from siloed edge environments.

Drivers

- **Extreme speed:** By placing data, data management capabilities and analytics workloads at optimal points ranging all the way out to endpoint devices, enterprises can enable more real-time use cases. In addition, the flexibility to move data management workloads up and down the continuum from centralized data centers or the cloud to edge devices will enable greater optimization of resources.
- **Data gravity:** Bandwidth costs and scenarios with limited or intermittent connectivity demand the ability to organize and process data closer to the edge.
- **Expanded scale and reach:** By using distributed computing resources, and spreading the load across the ecosystem, enterprises can broadly scale their capabilities and extend their impact into more areas of the business. These areas include use cases and outcomes traditionally managed only via operational technology (OT) teams, such as those managing equipment in industrial settings. Dedicated hardware for edge processing of data will continue to amplify these benefits.
- **Resiliency:** Pushing data management capabilities toward edge environments can also bring benefits in the form of greater fault tolerance and autonomous behavior. If edge environments do not require centralized resources, then issues with connectivity to, or unplanned downtime of, those centralized resources don't disrupt processes that rely on local edge capabilities.

Obstacles

- **Management of distributed data architectures:** Data management has been largely based on principles of centralization — bringing data to central data stores (e.g., data warehouses), and then processing that data to create value. Edge environments break that model via distributed data architectures, raising complex choices about where to locate and aggregate data on the continuum of cloud/data center to edge. Determining the right balance of latency and consistency is one such choice.

- **Governance and security:** With the distribution and complexity of edge environments, data governance and security become challenging. Organizations should extend their governance practices and policies to address edge-resident data storage and processing capabilities, including disposal of ephemeral or nonvalue event data.

- **Organizational and skills considerations:** Many modern applications are being developed and deployed by OT teams lacking data management skills and oversight, or by IT teams lacking edge computing skills and experience.

User Recommendations

- Identify use cases where data management capabilities in edge environments can enable differentiated products and services by collaborating with OT and IT personnel working in edge locations.

- Expand the skill sets of IT and OT teams to include edge platforms and the technologies required to manage data and data-intensive workloads on them.

- Augment existing data management infrastructure to support edge deployment by partnering with product teams that are implementing IoT platforms and similar distributed computing architectures.

- Place a greater emphasis on end-to-end system design. Understanding the dependencies between all components of distributed data pipelines, analytics workloads and AI models will be crucial to success.

- Ensure safety and control by extending existing governance capabilities to edge data environments.

Sample Vendors

Couchbase; FairCom; IBM; Macrometa; Microsoft; MongoDB; ObjectBox; Xencia

Gartner Recommended Reading

Get Ready for Data Management at the Edge: Key Considerations and Actions

Building an Edge Computing Strategy

Forecast: Internet of Things, Endpoints and Communications, Worldwide, 2022-2032, 1Q23 Update

Forecast Analysis: Edge Hardware Infrastructure, Worldwide

Data Observability

Analysis By: Melody Chien, Ankush Jain

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

Data observability is a technology that supports an organizations' ability to understand the health of an organization's data, data pipelines, data landscape, and data infrastructure by continuously monitoring, tracking, alerting and troubleshooting issues to reduce and prevent data errors or system downtime. It tells us what went wrong based on agreed upon SLAs for data quality and usage; reasons; assesses the impacts; and recommends solutions. Data observability improves reliability of data by increasing our ability to observe changes, discover unknowns and take appropriate actions.

Why This Is Important

Data observability uses data profiling, AI/ML, lineage and active metadata to provide the following benefits:

- **Monitor & Detect:** Provide a holistic view to determine how components of data pipelines are operating, evaluate whether data quality meets expectations, and detect data related issues.
- **Alert & Troubleshoot:** Send right alerts to the right people at the right time and perform root cause analysis.

- **Resolve & Prevent:** Provide recommendations to fix the issues or optimize data pipelines to meet business requirements with the goal to prevent downtime or critical data issues before affecting business.

Business Impact

- Data observability allows technical teams to gain visibility of the health of data pipelines and infrastructure. They can identify possible drifts in various areas, and minimize the time to investigate and solve issues, preventing unplanned outages or critical data errors.
- Business users will also gain visibility of data quality and associated financial impacts. This will ensure appropriate use and management of data to meet governance requirements.
- Data observability allows facilitation and improvement of the data fabric with continuous observations and evaluations of the data and analytics ecosystem.

Drivers

- Data and analytics leaders face a growing number of mixed data stacks, diversity of datasets, unexpected data drifts such as change in schema or business context, high demand for data quality and near zero tolerance of downtime. All these add to the challenges in data management. They need a holistic view of the state of data quality and data pipelines within interconnected systems.
- Data pipelines move data from point to point and deliver data to consumers. This journey can be disrupted by unexpected events such as data quality issues or a lack of infrastructure resources. The data that flows through these pipelines needs to be monitored for loss of quality, performance or efficiency. Organizations need to be able to identify points of failure before they have a chance to propagate. Data observability automatically detects important events and analyzes various signals to troubleshoot the issues, and provides actionable insights of what to do next.
- Data observability goes beyond traditional monitoring. It provides a multidimensional view of data including performance, quality, usage and financial impacts to the downstream applications. Leveraging active metadata, lineage of data and AI/ML, data observability generates real-time insight by monitoring the business context and analyzing data pattern, comparing history, and developing a semantic understanding of the data. It provides an end-to-end observability to help organizations be better equipped to handle critical events and prevent business disruptions.
- This capability is essential to the data fabric design concept and becomes an important building block to further automation in data management practices.

Obstacles

- There is no standard definition of what constitutes a data observability solution. Vendors offer a range of different capabilities often branded as data observability which is causing confusion in the market and leading to issues adopting the tools.

- The current vendor landscapes are very fragmented based on coverage areas and data environments supported. Most vendors focus on observability of the data quality and data pipelines, and are less concerned about data usages and financial impacts. The full end-to-end observations are not quite there yet from individual vendors.

- Most data observability tools only support the modern data stack. This limits their application in large enterprise environments with more complex data environments in many cases using legacy data management tools.

- Most data observability tools target the data engineer persona and are positioned as IT tools. Though business users receive important insights from data observability tools, they may find them less user-friendly.

- Organizations are embracing the concept of “observability.” But the actual adoption of the tools is not straightforward. The consideration of how they connect to the overall ecosystem and connecting this to data governance strategy is still a concern.

User Recommendations

- Identify the data elements or data pipelines which require high standards or SLA in quality, uptime, latency and performance. Pinpoint the gap of current monitoring capabilities vs. desired capabilities to support the requirements.
- Evaluate data observability tools available in the market that can enhance your observability based on priority of business requirements, primary users and interoperability with the enterprise data ecosystems.
- Pilot data observability program by building a monitoring mechanism as a starting point to increase visibility over the health of data. Invest in observability capabilities in a cloud environment first, as it's commonly supported by vendors and is faster and easier to demonstrate value.
- Include both business and IT perspectives when evaluating data observability tools by engaging with both personas early on in the evaluation process.
- Partner with business stakeholders to evaluate and demonstrate business value of data observation practices by tracking improvement of data quality, reduction in downtime and ability to meet SLAs to show tangible benefits.

Sample Vendors

Acceldata; Ataccama; Bigeye; Collibra; IBM; Kensu; Monte Carlo; Soda; Unravel

Gartner Recommended Reading

Data and Analytics Essentials: Data Observability

Quick Answer: What Is Data Observability?

The State of Data Quality Solutions: Augment, Automate and Simplify

Market Guide for DataOps Tools

At the Peak

Data Mesh

Analysis By: Roxane Edjlali, Ehtisham Zaidi, Mark Beyer, Michele Launi

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Emerging

Definition:

Data mesh is a data management approach. Though not an established best practice, it supports a domain-led practice for defining, delivering, maintaining and governing data products. Data products are a packaging and delivery mechanism for data that must be easy to find and use by data consumers (business users, data analysts, data engineers or other systems). Data products must also fulfill a contract (terms of service and SLAs) between the provider and the consumer.

Why This Is Important

The definition of data mesh is evolving as the market explores the approach. Data mesh represents a potential alternative or complement to centralization-only data management strategies for analytics and other use cases. Organizations continuously seek a means to balance data requirements for repeatability, reusability, governance, authority, provenance and optimized delivery of data. Data mesh is a skills- and resource-intensive approach that shifts responsibility and authority back to subject matter experts (SMEs) in each data domain.

Business Impact

From a governance and authority perspective, data mesh relies on a federated governance approach that can delegate authority to the business and data domain SMEs. SMEs are assumed to exhibit the greatest experience in capturing and using data within their domain of expertise. They are responsible for determining guidance and processes for creating, managing and preventing unnecessary proliferation of data products. The goal of the mesh is to provide ready access to data products.

Drivers

- Data mesh provides a model that allows for decentralized data management, which aligns to organizational needs.

- Data mesh gives domains the flexibility they need to build data products that meet their required use cases. It also gives domains more control over the use of those data products across the enterprise.
- By leveraging existing assets instead of centralizing the data architecture, data mesh can reduce the time and effort required to enable data reuse throughout the enterprise. Data mesh asserts remediation for flexibility, scalability and accountability issues in approaches like centralized data warehouses, data lakes and data hubs.
- In Gartner client interactions, delays in data access and utilization are the most frequently reported issues from organizations seeking to deploy data mesh. Organizations question the success of data centralization, which can't meet all analytical use cases.
- Data mesh emerges as a compromise to respond to delivery issues, budget constraints, and misunderstandings between central teams and lines of business. Centralized approaches are often detached from the broader business domain requirements.

Obstacles

- Data management maturity and skills are required for data governance at the domain level, data completeness, application design and deployment, data quality, data provenance, systems architecture, and analytics data management.
- Data products must be able to meet the SLAs of the other groups sharing, reusing or accessing them. The associated skills may not be present in the BUs.
- Inappropriate identification of either data details or correct integrity for combining them may cause data product proliferation, thus increasing management and maintenance and necessitating reengineering to reconcile different interpretations of the data.
- Data mesh implementations and practices do not follow any specific guidelines. They cannot be vetted against standardized, or even competing, approaches. Implementations vary and may incorporate multiple approaches (e.g., marketplace experiences, virtualized views or subject-specific data marts).
- Data mesh will be obsolete before the plateau. The practice and supporting technology will evolve toward data fabric as organizations start collecting passive metadata.

User Recommendations

- Commit to building a distributed data management team, as the data mesh concept is highly dependent on the organizational model and the distribution of skills across central IT and LOBs.
- Assess data products for business domain alignment and efficiency gains upon delivery. Data strongly aligned to a single domain with broad utility across the enterprise may provide lower risk for initial data product efforts.
- Control data product proliferation by monitoring technical debt and ensuring that data products continuously evolve to meet changes in usage and scope.
- Start a metadata management program in parallel with data mesh, and collect passive metadata. This approach will allow data mesh to evolve toward metadata activation over time, guiding data product operationalization, value justification and greater transparency in the uses of data.
- Mitigate irresponsible data management by addressing management and governance contention issues in data product design within the data domains.

Gartner Recommended Reading

Quick Answer: What Is Data Mesh?

Quick Answer: Comparing Data Fabric and Data Mesh

Data and Analytics Essentials: Data Fabric and Data Mesh

Quick Answer: How Are Organizations Overcoming Issues to Start Their Data Fabric or Mesh?

2023 Planning Guide for Data Management

Intercloud Data Management

Analysis By: Adam Ronthal, Donald Feinberg

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

Intercloud data management is the process of actively managing data in multiple cloud providers as part of a cohesive application and data management strategy. It builds on the foundation of multicloud capabilities, but adds the ability to access and use data across clouds in an operational context. It can be done at the cloud object store (COS), database management system (DBMS) or application tiers.

Why This Is Important

The vast majority of organizations using the public cloud are storing data on more than one cloud. Today, most of that data remains siloed — accessed and managed in the context of a single cloud environment. As data's center of gravity shifts to the cloud (or multiple clouds), data and analytics leaders will seek out means to unite that data in a logical and cohesive consumption tier to improve optimization, efficiency, flexibility and insight.

Business Impact

The ability to access data — regardless of where it is located — is a potentially transformational capability that will serve to break down barriers to access for end users and applications. Intercloud data management is cloud-agnostic and will allow enterprises to access their data in any cloud at any time, and by any means. It will enable globally distributed applications that span cloud providers and geographies, providing resilience and avoiding lock-in to any single cloud service provider.

Drivers

Though actual market penetration remains minimal, several drivers affect adoption, including:

- **Regulatory requirements:** Some industries are starting to mandate the use of multiple clouds for resilience and availability reasons, and some countries have strict data sovereignty laws.
- **Global applications:** Applications that operate on a global basis may require the use of multiple clouds to meet latency and performance requirements.
- **Distributed teams:** Organizations using more than one cloud may need to work with data in more than one cloud and provide continuity in their multicloud environments.
- **Integration:** Distributed data must be integrated (in storage and/or logically) to achieve maximum business value.

Obstacles

- Enterprises seeking to adopt intercloud data management approaches tend to be large, global enterprises with specific application requirements; most small to midsize organizations do not have these requirements. Thus, market penetration is minimal.
- Intercloud data management requires technologies built for distributed data management. Some established cloud service providers — most notably Google Cloud Platform and Microsoft Azure — are entering the space with recent offerings. However, most of the leading data management technologies are not designed to manage data across multiple clouds.
- Intercloud data management requires management of traffic between the clouds. Multiple methods exist, each with different cost, performance and security considerations.
- Any intercloud data management application will also be subject to the laws of physics. Practitioners will need to be aware of both performance implications and trade-offs in consistency and availability, and ensure that their applications are both aware of and designed with these trade-offs in mind.
- As with any data, intercloud data must be governed and integrated in storage and/or logically to achieve maximum business value.

User Recommendations

Weigh trade-offs in optimization and flexibility when making design decisions, as intercloud data management can occur at three different levels:

- **Object store:** Distribute data at the COS layer and replicate between clouds when flexibility and diversity of choice are required. Data science teams, for example, can choose whatever optimization layer they would like for last-mile delivery, as all can read and write to the local COS.
- **DBMS:** Select a DBMS that distributes and manages data in a geodistributed cluster when global operational efficiency and resilience are required. Nodes can reside in multiple clouds and/or on-premises. This approach can not only support local, low-latency read/write applications with global read capabilities, but also enforce data sovereignty requirements.

- **Application:** Use the application tier when data is already in multiple clouds and you require a means of bringing it together. This approach essentially defers data integration to the point of consumption.

Sample Vendors

Cockroach Labs; Couchbase; DataStax; Google; MongoDB; Oracle; PingCAP; Snowflake; WANdisco; Yugabyte

Gartner Recommended Reading

How to Plan for Optimal Multicloud and Intercloud Data Management

What Are the Key Factors to Consider When Choosing a Cloud Data Management Architecture?

6 Best Practices to Create a Cloud Management Services Offering in the World of Multicloud and Hybrid Cloud

Building an Edge Computing Strategy

Distributed Cloud: Does the Hype Live Up to Reality?

Augmented Data Quality

Analysis By: Ankush Jain

Benefit Rating: Transformational

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

Augmented data quality (ADQ) solutions provide the capabilities for enhanced experience aimed at improved insight discovery; next-best-action suggestions; and automation by leveraging artificial intelligence (AI)/machine learning (ML) features, graph analysis and metadata analytics. Each technology can work both independently and cooperatively to create network effects, which can then be used to increase data quality automation and effectiveness across a wide range of data quality use cases.

Why This Is Important

Ensuring high-quality data is important to data and analytics endeavors. Based on rapid expansion of contemporary data environments, a multitude of data types and pressing demands of businesses, organizations are searching for innovative approaches that are fast, affordable, scalable and easy to implement to tackle data quality issues. ADQ technologies revolutionize conventional and time-consuming manual procedures by increasing automation and enhancing insights.

Business Impact

- Automation/augmentation enhance data quality, reduce manual effort and improve efficiency.
- Multipersona usability enables nontechnical users to run processes via natural language, eliminating skill barriers.
- AI/ML techniques and metadata analytics enhance multiple data quality processes.
- Semantic connections, lineage tracing and domain data mapping enable impacts/solutions to be identified by knowledge graphs.
- Support for data engineers includes monitoring/observability across complex landscapes.

Drivers

- Traditional data quality practices that rely on manual efforts and subject matter experts struggle to address complex and exception-prone data quality problems.
- Data quality across various use cases offers accelerated time to value, reduced risk and increased competitive advantage across all business activities and user groups.
- Augmented data quality solutions are essential for emerging and future data ecosystems, integrating seamlessly with cohesive designs, such as data fabrics, supporting operational excellence and enhancing financial governance.
- Organizations need seamless integration, agile deployments and bidirectional exchange of intelligence with adjacent data management functions, which is core to ADQ.
- ADQ enables organizations to scale and unify data quality efforts for enterprisewide success, which is often a challenge, due to limited internal capabilities and strategies.
- ADQ makes use of advanced techniques including ML, natural language processing (NLP), large language models and GenAI, active metadata and knowledge graphs. This enables augmentation across several data quality capabilities, such as profiling and monitoring/observability; data transformation; rule discovery and creation; matching, linking and merging; data quality remediation; and role-based usability.
- Embracing augmented data quality solutions and leveraging emerging technologies is crucial to improve data integrity, governance and overall success in the data ecosystem.

Obstacles

- **Limited awareness and understanding of benefits** of ADQ solutions can impede adoption. Organizations should actively educate their teams about the value and potential impact of these tools, fostering a culture that embraces and leverages advanced data quality technologies.
- **Lack of scalability and integration with existing data infrastructure** can be an obstacle to the adoption of ADQ tools.
- **The lack of explainability and traceability of AI/ML algorithms** could lead to reluctance to adopt these tools.
- **The inclusion of data and analytics governance is crucial** when implementing ADQ tools. AI-driven automation provides users with independence, but it is essential to embed governance requirements into the AI models to mitigate data-related risks.

User Recommendations

- **Evaluate data quality capabilities:** Assess manual efforts/complexity needed to support use cases. Identify improvement areas ADQ can address. This will help determine requirements for adopting ADQ.
- **Explore ADQ capabilities:** Investigate the features, setup process, required skills and constraints associated with ADQ solutions. Assess offerings from incumbent data quality vendors and explore product roadmaps for enhancement.
- **Establish data cataloging:** Implement practices to collect/analyze metadata for automation and efficient data quality processes. Enhance management of data assets and facilitate ADQ integration.
- **Align with data governance:** Partner with stakeholders to monitor ADQ solutions. Ensure to governance requirements and framework adherence. Establish metrics to show benefits/business value.
- **Collaborate with solution providers:** Provide feedback, share experiences and suggest enhancements. Engage in user forums, contribute to improvements and shape development roadmap to meet needs.

Sample Vendors

Ataccama; Collibra; DQLabs; Experian; IBM; Informatica; Precisely; Syniti; Talend

Gartner Recommended Reading

The State of Data Quality Solutions: Augment, Automate and Simplify

Augmented Data Quality Represents a New Option for Upscaling Data Quality Capabilities

Building Automation Into Your Data Quality Initiatives

Magic Quadrant for Data Quality Solutions

Critical Capabilities for Data Quality Solutions

Distributed Transactional Databases

Analysis By: Rick Greenwald

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Early mainstream

Definition:

A distributed transactional database allows transactions to be performed on any of a distributed set of database instance nodes. The ability to accept writes from a geographically distributed set of nodes while maintaining data integrity and consistency, and delivering acceptable performance distinguishes this technology.

Why This Is Important

Transactional databases are the heart of operational systems, used in all lines of business. A distributed transactional DBMS allows for transactions to be performed on the distributed nodes of the database instance, including those separated by geographic distance, simultaneously and without a loss of data integrity.

Distributed transactional databases have to choose between consistency and availability, and typically use specialized technology and hardware to minimize the impact of that choice.

Business Impact

Organizations that need to implement applications whose transactions span geographic distances will need systems that don't compromise data integrity. Systems such as global trading applications will be able to use distributed transactional systems to expand their scope and reach, and may provide competitive advantage by allowing organizations to deliver unique capabilities, but will include significant development and deployment demands.

Drivers

- Global business operations increasingly require widely distributed data models and active transactional capabilities across the distributed nodes.
- Cloud platforms that provide underlying infrastructure for data distribution make it easier to deploy widely distributed platforms for globally active transactional systems.
- Distributed data allows for more local data access and lower latency worldwide for systems that need reduced latency in operations spread over a distributed set of data and that are still subject to write transactions.
- Systems that can support distributed transactions can also offer the highest levels of resilience and fault tolerance.

Obstacles

- Geographically distributed transactions are still not requirements in most transactional systems, so these systems do not require the extra functionality offered by distributed transactional databases. Implementers can also design systems and data schemas to reduce or eliminate the impact of potential losses of data integrity with distributed systems.
- Data that is not required for distributed transactional scenarios can be distributed and still be implemented with a much wider variety of systems while still delivering scalability and performance advantages without the use of these specialized engines.

User Recommendations

- Select a distributed transactional DBMS if your use case requires data integrity implemented over a geographically distributed database instance.
- Examine the use of design compromises, rather than a distributed transactional system, if your need for distributed integrity can accommodate these compromises without excessive development and maintenance.
- Ensure that a distributed transactional DBMS does not include policies for update contention resolution that are not appropriate for your business use cases, as some distributed transactional DBMS have built in contention resolution.

Sample Vendors

Aerospike; Cockroach Labs; Couchbase; DataStax; Google; MariaDB; MongoDB; NuoDB; Oracle; Yugabyte

Gartner Recommended Reading

Data Consistency Flaws Can Destroy the Value of Your Data

Lakehouse

Analysis By: Roxane Edjlali, Adam Ronthal

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

A lakehouse is a converged infrastructure environment that combines the semantic flexibility of a data lake with the production optimization and delivery of a data warehouse. It supports the full progression of data from its raw, unrefined state, through the steps of refining it, to ultimately deliver optimized data for consumption.

Why This Is Important

Emerging practices for designing data lakes and integrating them with related systems is still an ongoing challenge for organizations. For example, using a data lake with a data warehouse adds complexity to the data and analytics landscape. A lakehouse aims to unify the two to simplify architecture and improve efficiency while minimizing the need to move data and analytic model scores between the two. A more efficient environment with a smaller operational footprint is the potential result.

Business Impact

Businesses will benefit from streamlined delivery, rapid access to data and a consolidated data management platform that supports highly skilled data scientists, engineers and analysts, plus casual users who consume data via prebuilt reports or dashboards. A lakehouse provides a well-defined path from discovery-oriented analytics and analytic model development (via the lake portion of the lakehouse) to the delivery of analytic insights and quantification to end users (via the warehouse portion).

Drivers

- Operationalizing data science projects so their insights are shared broadly continues to be a challenge. A lakehouse unifies the exploratory and production environments, for maximum business use and value.

- Enterprises consistently seek rapid and unencumbered access to data and struggle with the processes and perception of delayed delivery associated with the data warehouse. The lakehouse is often positioned as a “silver bullet” to solve this problem.

- Data lakes and data warehouses are optimized for different things. Lakes enable data science, other analytics and the management of any data structure, latency or container. Warehouses excel with refined data that requires an audit trail, high quality, and accuracy or special data structures (dimensions, time series and hierarchies). Between the two, the lakehouse supports a long list of data requirements and business use cases.

- Many cloud data warehouse solutions and almost all cloud data lakes already leverage semantically flexible cloud object storage as their storage of record. It is a natural progression to unify these storage environments, thus reducing the disparate and duplicate infrastructures.

- Data lakehouse concept is maturing fast as it benefits from the market understanding of data warehousing and data lakes, and their respective pros and cons.

Obstacles

- The maturity of vendor-built lakehouse platforms is developing. Some are strong with data lakes, but do not support the full range of transaction consistency or robust workload management capabilities that data warehouse solutions can support Other lakehouse platforms are strong with data warehousing, but lack the broad data model support and data science or data engineering features of a data lake.
- The most complex data warehousing workloads are still likely to be beyond the scope for most lakehouse solutions that do not incorporate already-developed functionality.
- Given that data lakes and the lakehouse concept are both still new, a common obstacle is the immaturity of users' ability to design, deploy and maintain complex data architectures.
- The full scope of optimization includes data quality, security, governance and performance and most importantly, good metadata management and data integration. Only a few lakehouse platforms address all of these.

User Recommendations

- Employ a targeted use-case approach that solves specific problems and expands from there for long-term success. Expect your lakehouse to grow into many more use cases over time, just as lakes and warehouses do.
- Avoid "overpromising and underdelivering" by testing candidate solutions thoroughly to ensure that you can actually deliver reliable and high-performance workloads on data lake infrastructure.
- Run your most complex workloads on the evaluated target platform in a proof of concept (POC) to make better-informed decisions about when a lakehouse approach is sufficient and when a dedicated data warehouse may still be required.
- Choose a logical data warehouse (LDW) approach when addressing a broader data and analytics scope. A lakehouse is a subset of the LDW built opportunistically. The LDW remains a mature and best practice.
- Evaluate metadata management, security and governance capabilities to ensure they meet your enterprise standards and data requirements.

Sample Vendors

Amazon Web Services; ChaosSearch; Databricks; Dremio; Google; IBM; Incorta; Microsoft; Snowflake; Starburst

Gartner Recommended Reading

Exploring Lakehouse Architecture and Use Cases

Market Guide for Analytics Query Accelerators

Ledger DBMS

Analysis By: Aaron Rosenbaum, Henry Cook

Benefit Rating: Moderate

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

Definition:

A ledger DBMS is an append-only, immutable DBMS with an embedded, cryptographically verifiable audit trail. Ledger DBMS also includes multimodel DBMS products with a ledger or blockchain data type. A ledger DBMS is useful for private and permissioned "blockchain-like" applications where distributed consensus is not required and one entity has control over the ledger and designates which parties, other than the owner, have access to and may add to it (i.e., the users trust the provider).

Why This Is Important

Ledger DBMSs provide many of the benefits of a blockchain platform, like data tampering detection and auditing, without the complexity of configuring and managing a decentralized environment. Many offer SQL query capabilities to the ledger data. They are managed by a single entity, which makes their implementation and management easy and secure.

Business Impact

Today, many blockchain projects are forced to use public blockchain technologies when a DBMS would suffice. Ledger DBMSs represent a choice, stand-alone or multimodel, that is more manageable and easier to implement. This will enable businesses to use ledger technology where performance and immutability is required — in use cases such as audit trails, data lineage, digital assets and sharing data.

Drivers

- Ledger DBMSs provide many of the benefits of a blockchain platform, like data tampering detection and auditing, without the complexity of configuring and managing a decentralized environment. They are managed by a single entity, which makes their implementation and management easy and secure. Because it is a DBMS and is managed by a single entity, it also has far better performance than a public blockchain system.
- Many vendors have services or products that fall into this category. IBM and Oracle support blockchain tables, Microsoft has added a ledger data type to CosmosDB, and AWS has the Amazon Quantum Ledger Database (QLDB), which is more mature because it has been used internally by Amazon for many years.
- During the next few years, we believe there will be a number of additional ledger DBMS products (mostly dbPaaS), which will increase the choices available.

Obstacles

- Distributed ledgers have become simpler to implement and are becoming more mature. At the same time, readily available immutable storage is available from many major cloud vendors. The combination has led to a migration away from ledger DBMSs toward either distributed ledgers or immutable storage. Additionally, some DBMSs are supporting immutability within the DBMS rather than requiring a specific ledger DBMS. Although some use cases may be best run using a ledger DBMS, most use cases requiring the benefits of ledger DBMSs will be built and operated with other technologies. As such, Gartner has marked this technology as Obsolete.
- Most ledger DBMSs are new and relatively immature.
- The broader market is only beginning to understand the benefits and potential use cases of ledger DBMSs causing slower adoption.
- This technology has just passed the Peak of Inflated Expectations. It will take demonstrable market successes and additional product introductions to diminish the hype.

User Recommendations

- Review your organization's tactical and strategic requirements, and be cognizant of the benefits and challenges of centralized, decentralized and distributed systems so that you select the most business-appropriate platforms.
- Compare the suitability of ledger DBMSs against permissioned blockchain technologies by carefully considering your need for: a centrally administered data-auditing capability versus a decentralized network of peers; a single organization as the "source of truth" versus multiple potential validators; and a ledger DBMS versus a normal RDBMS.
- Decide whether to keep business processes, and their associated data, private to a single organization versus sharing business processes executed through smart contracts.

Sample Vendors

Amazon Web Services; Fauna; Fluree; IBM; immudb; Microsoft; Oracle; TerminusDB

Gartner Recommended Reading

Amazon QLDB Challenges Permissioned Blockchains

Sliding into the Trough

DataOps

Analysis By: Robert Thanaraj, Ehtisham Zaidi, Sharat Menon, Aaron Rosenbaum

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Emerging

Definition:

DataOps is an agile and collaborative data management practice focused on improving the communication, integration, automation, observability and operations of data flows between data engineers and data consumers. The goal is to assist data and analytics (D&A) leaders in driving operational excellence in data delivery in support of their data management solutions.

Why This Is Important

DataOps eliminates various inefficiencies and misalignments between data management and consumption use cases by streamlining data delivery processes and operationalizing data workloads. DataOps practices:

- Improve organizational speed and trust in delivering data
- Manage interdependencies across business processes
- Increase reusability of data engineering work product
- Provide reliable data delivery service levels
- Govern data, leading to trust and use among consumers

Business Impact

- Data engineers benefit from increased productivity and robust change management, ensuring data delivery service levels like quality, lineage and security.
- Data consumers benefit from reduced cycle time of accessing ready-to-use data and improved data trust.

- Organizations thrive on data literacy, productivity gains, self-service enablement and a collaborative culture.
- Eliminate unwanted data delivery efforts by focusing on value flows tied directly to business impact.

Drivers

- Organizations strive to improve speed and efficiency of producing trusted and usable data. DataOps practices reveal bottlenecks in the current D&A delivery process and guide toward improving the lead time (process efficiencies) and cycle time (technical efficiencies).
- DataOps improves the shareability and reusability of the data across the organization. It involves formal processes around data architecture, quality and modeling, and ensures that the data governance requirements are being applied as part of the operational processes. Otherwise, the initial data pipeline would only be designed for a narrow use.
- DataOps practices enable reorganization of teams, which helps overcome the challenges caused by fragmented teams/processes and delays in delivering data in consumable forms.
- DataOps tools eliminate the various inefficiencies and misalignments across data management technologies by streamlining data delivery processes and operationalizing data workloads. It is an emerging technology market.

Obstacles

- Setting up DataOps is a challenge as it needs efforts on justifying operating model shifts, focus on metadata management practices (which are currently nascent) and continuing to provide effort or cost over value justifications which require aligning business outcomes to DataOps activities.
- Organizations have substantial domain expertise and siloed functional capabilities. It is challenging to retain the advantages of the legacy approaches to data management and the people holding those skills, while also aggressively pursuing DataOps.
- Organizations lack a holistic view of various stand-alone technologies that are often managed by multiple teams with varying levels of operational maturity.

- Many strive for end-to-end automation of pipelines driven by code and integrating diverse technologies to make them work together is complex as it involves diverse skill sets.

User Recommendations

- When introducing DataOps, target projects that are struggling due to lack of collaboration, overburdened by the pace of change, or where service tickets from data consumers are piling up.
- Apply the core DevOps approaches to data management such as automating deployment to test environments continuously and managing schema drifts in pipelines. Reach out to your application leaders who have successfully applied DevOps practices to application development.
- Track metrics such as time to deploy changes, degree of automation, developer productivity, code quality, failure rates in production, cost-efficiencies and business impact in dollar amount.
- Plan for data pipeline operations ownership and service-level management. DataOps teams generally own the full development life cycle from inception to production. They must be cross-functional teams that combine data management, software engineering and I&O expertise — some organizations even hire for new roles such as data product manager.

Sample Vendors

Astronomer; BMC; DataKitchen; DataOps.live; GitHub; Kensu; StreamSets; Tengu; Torana (iCEDQ); Unravel

Gartner Recommended Reading

Data and Analytics Essentials: DataOps

Market Guide for DataOps Tools

5 Ways to Enhance Your Data Engineering Practices

How to Apply DevOps and Value Stream Mapping to Data, Analytics and AI

Toolkit: Delivery Metrics for DataOps, Self-Service Analytics, ModelOps and MLOps

Data Ecosystems

Analysis By: Adam Ronthal, Robert Thanaraj, Aaron Rosenbaum

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

Data ecosystems provide a cohesive data management environment that supports all data and analytics workloads. They have a common governance and metadata management framework and unified access management. They integrate augmented data management capabilities with a set of services accessible by the business user. They provide streamlined delivery and comprehensive functionality that is straightforward to deploy, optimize and maintain. Third-party vendors also participate in data ecosystems.

Why This Is Important

Data and analytics leaders report that the cloud experience today requires a significant integration effort to ensure that components work well together. Cloud service providers (CSPs) and independent software vendors (ISVs) are responding with more refined data ecosystems as the market moves from “some assembly required” to a “packaged platform experience.”

Business Impact

Data ecosystems unify data management and associated use cases with streamlined delivery and easy integration within a holistic management framework. They address key data management disciplines, such as data integration, data quality, data sharing, governance, metadata and observability — via augmentation — and provide the basis for operational and analytics capabilities. They are delivered both by a combination of native CSP offerings, as well as ISV components that provide enhanced capabilities.

Drivers

- Data and analytics architectures are under significant stress on two fronts: hybrid and multicloud deployment environments, and the diversity of data persistence models required to meet the increasing demands of data and analytics.
- Cloud practitioners need to rationalize data silos, which span multiple deployment environments and frequently require different and potentially conflicting operating models.
- Enterprises are looking to unify the way they engage with different data models, platforms and use cases to improve efficiency and time to value for data-driven initiatives.
- Data ecosystems serve as a unifying approach to resolve these pressure points. Built on a common foundation of governance, metadata and emerging data fabric design, they enable new practices like DataOps, FinOps and PlatformOps. They will become self-optimizing and self-tuning, and support financial governance efforts through cost optimization.
- Data ecosystems promise improved productivity and ROI based on the value of not having to do explicit data and application integration as they are based on a common set of services.

Obstacles

- While data ecosystems have a vision of unifying data management with common governance, security and metadata, significant work is still needed to make this a reality. Gaps exist in data integration, data quality, metadata and governance, which need to be addressed either through native CSP offerings or partnerships with ISVs to fully realize the vision of the cloud data ecosystem.
- When combining native CSP offerings with third-party ISV offerings, end users may find that additional effort is required to integrate these components. This undermines the core concept of a unified, holistic data ecosystem, though the end result leveraging CSP and ISV offerings is likely to be more capable.
- While CSPs are working with third-party ISVs to provide open ecosystems, their initial focus remains on ensuring that their own components are working well within their own cloud and addressing the basic needs of their customers.
- For a cloud ecosystem (that encompasses a CSP and ISVs) to function well, CSP and ISV components must have standard interfaces that exchange metadata bidirectionally. It is essential that participating ISVs and CSP agree on common metadata sharing standards. This remains slow to emerge in the market today.

User Recommendations

- Assess the maturity of these ecosystems and the degree to which they deliver on the promise of a unified environment.
- Assess points of integration between various components (data persistence, use cases, data integration, observability, governance and metadata capabilities) to determine how cohesive the resulting ecosystem is. A less cohesive ecosystem will require significantly more integration time and effort.
- Ensure that the data ecosystem has a well-articulated path to production for a full data life cycle (from discovery to production-optimized delivery).
- Define what CSPs need to deliver as part of the solution and what capabilities to obtain from third-party ISVs; expect to spend more time on integration efforts when combining CSP and ISV offerings.

Sample Vendors

Amazon Web Services (AWS); Cloudera; Databricks; Google Cloud Platform; IBM; Microsoft; Oracle; SAP

Gartner Recommended Reading

The Impacts of Data Ecosystems: A Cloud Architectural Perspective

Innovation Insight: Data Ecosystems Will Reshape the Data Management Market

Strategic Roadmap for Migrating Data Management to the Cloud

Data Hub Strategy

Analysis By: Andrew White, Thornton Craig

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

A data hub strategy effectively determines where, when and how data needs to be mediated, governed and shared in an enterprise. It layers data and analytics governance requirements atop sharing demands to establish the patterns for data flow. The strategy drives the implementation of one or more data hubs — architectures that enable data sharing and access by connecting data producers (applications, processes and teams) with data consumers (other applications, processes and teams).

Why This Is Important

Digital business cannot scale by continuing to piecemeal all the programs and practices that have evolved in the last 10 years. A connected, aligned approach is needed. A data hub strategy provides that connected approach to streamline and simplify how all programs related to D&A governance operate: data quality, MDM, ADM, metadata management, data catalogs and so on. You end up getting more from less effort and investment.

Business Impact

- Increased operational efficiency by aligning and integrating previously siloed governance programs such as data quality, MDM, ADM, metadata management, data catalogs and so on.
- Increased return on all D&A investments through more effective and targeted efforts on implementing governance of D&A information assets such as data, analytics, models, etc.
- Reduced complexity and cost across overall information infrastructure and data fabric or mesh.

Drivers

- Demands for seamless data flow across teams, processes and systems in the enterprise, which have increased dramatically in complexity and mission-criticality.
- New demands for consistent and reliable sharing of critical data between the organizations and things that comprise the extended enterprise — for example, in support of Internet of Things (IoT) solutions and new digital products.
- Better collaboration across business-oriented (governance) and IT-centric (integration) roles concerned with delivering data to points of need across the enterprise.
- Longtime and continued frustration of business stakeholders over the lack of consistency and trust of data driving strategic business outcomes — a data hub strategy enables more-focused application of governance controls, as compared with trying to align governance approaches inside many endpoint systems.
- Emerging data fabric design patterns that both need and leverage trusted sources of data and can inform what data should be governed more importantly.
- Growing need for a flexible and governable architecture that complements centralized data stores such as data lakes and data warehouses.
- Desire of many organizations to leverage the concepts and successes of MDM programs toward governance and sharing of other types of critical data. Includes coupling MDM and ADM across the enterprise.

Obstacles

- Inability to modernize D&A governance programs and shift away from legacy domain and data-centric or IT focused efforts to an outcome-based program.
- Resistance from teams or business units that prefer to retain control over their choices regarding how data is shared and governed.
- Inability to enable collaboration and agreement of critical stakeholders on data sharing and governance requirements across boundaries in the enterprise.
- Overreliance on technology and viewing governance and sharing of data as purely an implementation issue.

User Recommendations

- Identify the data that is most frequently used or is most important with most business value, and that requires effective governance and sharing. This might be a lean MDM or ADM/ERP program.
- Design a data hub strategy to understand data and analytics governance and sharing requirements, and to drive integration efforts across multiple use cases.
- Include any master data, application data, reference data, analytics data hubs or other intermediaries (e.g., customer data platforms) in your overall data hub strategy.
- Iterate changes to your data hub strategy as requirements for governance, sharing and integration change.

Sample Vendors

IBM; Informatica; MarkLogic; Profisee

Gartner Recommended Reading

Data and Analytics Essentials: Data Hubs

Data Sharing Is a Business Necessity to Accelerate Digital Business

Use a Data Hub Strategy to Meet Your Data and Analytics Governance and Sharing Requirements

Data Hubs: Understanding the Types, Characteristics and Use Cases

Private Cloud dbPaaS

Analysis By: Adam Ronthal

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Early mainstream

Definition:

Private cloud database platform as a service (dbPaaS) offerings bring self-service and scalability to a private cloud infrastructure, which minimizes external exposure. Private cloud dbPaaS offerings should provide similar benefits to their public cloud counterparts — a database management system or a data store engineered as a scalable, elastic, multitenant service, ideally with subscription of chargeback pricing models.

Why This Is Important

Private cloud dbPaaS offerings primarily appeal to enterprises that cannot embrace public cloud offerings (usually due to regulatory or governance concerns), but want cloud-like service delivery. These offerings may leverage the existing container-based infrastructure common to many private cloud offerings. They may also be self-contained products in an appliance form factor or extensions of existing cloud service provider offerings.

Business Impact

Private cloud dbPaaS offerings promise a marketplace-like experience for a range of DBMS offerings: commercial and open-source, relational and nonrelational. Many are still maturing to offer services that go beyond self-provisioned developer environments to true production-class environments. These offer high availability, elastic scalability and solid performance. A number of offerings from established vendors like IBM (Cloud Pak for Data System) and Oracle (Cloud@Customer) are available.

Drivers

- Private cloud dbPaaS will primarily appeal to enterprises that are not yet ready or are unable to embrace public cloud alternatives. Public cloud inhibitors (and private cloud dbPaaS drivers) may include: (1) regulatory, governance or security requirements, or the need to operate in an "air gapped" environment; (2) significant on-premises centers of gravity that are not yet able to move to public cloud; (3) concerns with network connectivity, latency or performance issues in a hybrid cloud environment; (4) data sovereignty requirements that cannot be met by public cloud data centers; and (5) compatibility concerns with on-premises environments associated with native public cloud offerings.

- Initial offerings in the space have come from traditional vendors with a strong on-premises presence. But cloud service providers are now engaging with private cloud dbPaaS offerings as well, almost always based on a container strategy that reaches into on-premises data centers.

Obstacles

- "Public cloud first" strategies remain a primary focus for vendors and end users alike, which may inhibit breadth of offerings as vendors increasingly focus on public cloud approaches.
- Many public cloud inhibitors are transient in nature, and may be addressed as organizational culture adapts to the cloud, regional availability improves and regulatory best practices associated with operating in the cloud improve. Private cloud dbPaaS may be a transitional stage for many adopters.
- Adoption will be limited to established enterprises with existing on-premises data center footprints, or via trusted partners or systems integrator.
- The requirement to leverage governance, security and operational updates by connecting to a control plane raises questions about how truly separated some private cloud offerings are.

User Recommendations

Data and analytics leaders considering private cloud DbPaas must consider:

- **Breadth of DBMS services offered:** Not all offerings support a full range of database types and features.

- **Storage model:** Container-based services require a scalable and persistent data storage tier for effective usage.
- **Pricing model:** Flexibility of pricing models is beneficial to accommodate both capital expenditure (capex) and operating expenditure (opex) approaches. Cloud pricing models (including private cloud offerings) can be complex.
- **Production capabilities:** High-availability and disaster-recovery capabilities must meet your requirements.
- **Disconnected operations:** Many of these offerings have a cloud-based management control plane. They must meet requirements for disconnected operations, if connectivity to the cloud is unreliable.
- **Path to the public cloud:** Private cloud dbPass will be a transitional technology. Continuity from on-premises operations into the public cloud is an important consideration.

Sample Vendors

Aiven; Alibaba Cloud; Amazon Web Services; IBM; Microsoft; Nutanix; Oracle; Pure Storage (Portworx); Robin; VMware

Gartner Recommended Reading

Market Guide for Container Management

What Are the Key Factors to Consider When Choosing a Cloud Data Management Architecture?

Quick Answer: How Do I Obtain Isolated Private Cloud Services?

Differentiate Hosted Private Cloud Offerings Using These 7 Dimensions

Magic Quadrant for Cloud Database Management Systems

Data Fabric

Analysis By: Mark Beyer, Ehtisham Zaidi, Roxane Edjlali, Sharat Menon, Robert Thanaraj

Benefit Rating: Transformational

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

Definition:

A data fabric is a design framework for attaining flexible and reusable data pipelines, services and semantics. The fabric leverages data integration, active metadata, knowledge graphs, profiling, ML and data cataloging. Fabric overturns the dominant approach to data management which is "build to suit" for data and use cases and replaces it with "observe and leverage."

Why This Is Important

Data fabric leverages traditional approaches while enabling the enterprise to adopt technology advances and avoids "rip and replace." It capitalizes on sunk costs and simultaneously provides prioritization and cost control guidance for new spending for data management. It leverages concepts and existing platforms/tools or implementation approaches. It offers flexibility, scalability and extensibility in infrastructure for humans or machines to assure data is consumable across multiple use and reuse cases on-premises, multicloud or hybrid deployments.

Business Impact

Data fabric:

- Increases identification, deployment and availability of data for reuse at scale.
- Provides insights to data engineers by standardizing repeatable integration tasks, improving quality, and more.
- Adds semantic knowledge for context and meaning, and enriched data models.
- Evolves into a self-learning model that recognizes similar data content regardless of form and structure, enabling connectivity to new assets.
- Enables observability across the data ecosystem.
- Reduces maintenance, support and optimization costs associated with managing data.

Drivers

- The dearth of new staffing or personnel seeking data management roles and the attrition of experienced professionals leaving the practice area has increased the demand for more efficient data reuse.

- Demand for rapid comprehension of new data assets has risen sharply and continues to accelerate, regardless of the deployed structure and format.

- Increased demand for data tracking, auditing, monitoring, reporting and evaluating use and utilization, and data analysis for content, values and veracity of data assets in a business unit, department or organization.

- Catalogs alone are insufficient in assisting with data self-service. Data fabrics capitalize on machine learning (ML) to provide recommendations for integration design and delivery, reducing the amount of manual human labor that is required.

- Significant growth in demand and utilization of knowledge graphs of linked data, as well as ML algorithms, can be supported in a data fabric to assist with graph data modeling capabilities and use-case generic semantics.

- Organizations have found that one or two approaches to data acquisition and integration are insufficient. Data fabrics provide capabilities to deliver integrated data through a broad range of combined data delivery styles including bulk/batch (ETL), data virtualization, message queues, use of APIs, microservices and more.

Obstacles

- Organizations will keep applying budget or staff to one-off and point-to-point integration solutions.
- Differing design and semantic standards used by various vendors to document and share metadata create challenges in its integration and effective analysis to support a data fabric design.
- Fabric needs analytic and ML capabilities to infer missing metadata. This will be error-prone at first with staffing and resources assigned to competing demands in advanced analytics, data science and AI near the data consumption layer.
- Active metadata management practices lag behind data fabric adoption but are critical to its implementation.
- Diverse skills and platforms demand a cultural and organizational change from data management based upon analysis, requirements and “design then build” to discovery, response and recommendation based upon “observability and leveraging.”
- Improper split from data mesh implies choosing one approach over another and not a complementary relationship.
- Inexperience in reconciling a data fabric with legacy data and analytics governance programs will confound implementers.

User Recommendations

- "Active metadata" and leveraging the inherent practices to it is mandatory in a data fabric (covered separately).
- Invest in an augmented data catalog that permits multiple ontologies over top of business data taxonomies and is alerted to new use cases for data and the related business units utilizing data.
- Deploy data fabrics that populate and utilize knowledge graphs in targeted areas where adequate metadata and metadata management practices already exist.
- Ensure business process experts can support the fabric by enriching knowledge graph capabilities with business semantics.
- Evaluate all existing data management tools to determine the availability of three classes of metadata: design/run, administration/deployment and optimization/algorithmic metadata. When adopting new tools, favor those that share the most metadata.
- Do not permit SaaS solutions to isolate their metadata from access by PaaS solutions that orchestrate across solutions.

Sample Vendors

Cambridge Semantics; Cinchy; CluedIn; Denodo Technologies; IBM; Informatica; Semantic Web Company; Stardog; Talend

Gartner Recommended Reading

Data and Analytics Essentials: How to Define, Build and Operationalize a Data Fabric

Quick Answer: What Is Data Fabric Design?

Emerging Technologies: Critical Insights on Data Fabric

Active Metadata Management

Analysis By: Mark Beyer, Guido De Simoni, Ehtisham Zaidi

Benefit Rating: Transformational

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

Active metadata management (AMM) is the continuous analysis of all types of metadata to determine the alignment and exceptions between “data as designed” and “operational experiences.” Its utilization includes operationalizing analytic outputs, operational alerts and recommendations. It identifies the nature and extent of patterns in data operations, which can result in AI-assisted reconfiguration of data processes and use cases.

Why This Is Important

Active metadata management alters/changes the behavior of data tools, platforms and systems in how they respond to instructions from each other in a continuous learning loop. It enables cross-platform orchestration of data tools, cross-industry validation and verification processes, and the identification of flawed data capture, inappropriate usage, logical fallacies and newly encountered data. It leverages machine learning, data profiling and graph analytics to determine data relevance and validity relative to use cases. At mature levels, it can support the evaluation of analytic and data biases, as well as transparency, auditing and DataOps.

Business Impact

Active metadata management can help businesses:

- Support self-service analytics and application development by automating data content, structures, availability and discovery of data assets.
- Identify commonalities among users, use cases, and reporting and analysis models across an organization, and social networks of users based on data needs and operational requirements.
- Automate orchestration for data access, locations, performance, processing requirements and resource allocation.

Drivers

- Changing requirements from both business and IT are driving demand for data quality tools, data catalogs, metadata management solutions and data integration tools in one comprehensive solution while also recognizing and supporting data privacy in the midst of data sharing.
- Human-driven data utilization must be augmented to adapt quickly to the demand for the rapid discovery, access and incorporation of new data assets throughout an enterprise or organization.
- Active metadata experienced a significant acceleration in the last 24 months. This occurred in the midst of real progress with regard to quality, master data, integration, governance and even security tools sharing more metadata.
- Data integration is required even from sometimes distantly removed vertical industries but complicated by third-party data and data utilized from enterprise partners.
- Intercloud data demands are increasing rapidly. The large-scale capabilities in cloud-based deployments have enabled the broadest diversity of data structures, processes and use cases to date.
- Demands are emerging for organizations to be able to isolate data anomalies and classify them as errors, outliers or actual data design changes that are undocumented.

Obstacles

- Active metadata is not automated accumulation of metadata for passive utilization. Vendors/suppliers have begun using the terminology to describe any periodic update to passive metadata as “active” (active metadata “washing,” sometimes deliberately). This may discredit the approach, forming resistance and barriers that will slow adoption due to skepticism in the market and push active metadata into the Trough.
- Access to all available metadata (such as social, operational, technical and business) is required for active metadata practices to reach full potential. Many existing platforms don’t make internal metadata available. Automated cross-platform and tools orchestration will be inhibited as a result.
- Data management solution providers are reluctant to make their metadata assets available to — much less accept — metainstructions.
- Human designers, implementers and users might resist this approach assuming humans are the best interpreters of data value, which also deters metadata management maturity.
- There is a significant lack of metadata management standards in today’s market.

User Recommendations

- Introduce an enterprise data catalog strategy, and expand it to ingest metadata beyond your data warehouses/lakes, such as from master data management, data quality, data integration, data preparation tools and analytical tools. Attach it to the catalog entries to begin accumulating metadata for analysis.
- Begin accumulating operational metadata (such as runtime logs and system workloads) from the wider D&A ecosystem. Analyze the logs for patterns of data used together and the frequency of use. Then examine user or connection strings, queries and views executed, and even resource allocation. This creates an operational knowledge graph of which data is used, how often, by whom, for what purpose and on which platform.
- Acquire or deploy at least one prototype combining at least three disciplines from data management to enable metadata notification between tools. Deploy a user interface to reconfigure metadata repositories for analysis by data engineers. This will change the data architecture and design culture to one that is observing and analyzing instead of designing from scratch — over and over again.

Gartner Recommended Reading

Tool: Vendor Identification for Active Metadata Management

Market Guide for Active Metadata Management

Quick Answer: What Is Active Metadata?

Case Study: An Active Metadata Augmented Data Classification System to Boost Analytics Efficiency

Deploying Effective Metadata Management Solutions

Information Stewardship Applications

Analysis By: Andrew White, Guido De Simoni

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Obsolete

Definition:

Information stewardship applications support the work of the data and analytics (D&A) steward role in a business (not IT) context. They provide application capabilities such as policy performance monitoring, data policy analysis, business glossary, workflow, exception management and root cause analysis.

Why This Is Important

Early master data management (MDM) solutions did not support policy monitoring and enforcement capabilities, so organizations started to build their own. Then, the General Data Protection Regulation (GDPR) came along, and the entire governance market pivoted to more basic needs of risk mitigation and compliance. The industry has now come full circle. Organizations are once again building their own distinct solutions for business stewards, and vendors in the D&A governance market are back to packaging up capabilities to meet those needs.

Business Impact

Without a stewardship function, D&A governance is not sustainable or scalable. Stewards need technology to monitor and enforce the condition of data they are responsible for. Business roles have specific requirements that are not met by IT-centric capabilities. Moreover, business roles use these tools infrequently, placing a much greater importance on developing tools that are easy to onboard and intuitive to use. Until vendors deliver such tools, organizations will build their own, or their governance programs will continue to fail.

Drivers

- Ultimately, the work of the steward, in a business role, requires an exception-oriented experience that maximizes impact and ease of use. If this capability is not met, D&A governance as a business capability will continue to fail.
- Businesses are under increasing pressure to share data for better data reuse, improved consistency and accelerated time to value. The focus is on using existing data dictionaries to identify areas of synergy between data used for different business initiatives (both data content and meaning).
- More effective understanding and communication of the semantic meaning of data will help resolve contention between business teams when inconsistency arises. It will reduce time and effort wasted on reconciliation, so that resources can focus on new business actions.
- Organizations need to make intelligent decisions about the information life cycle, from data interoperability and standards to archiving, disposal and deletion.
- Compliance, data sovereignty and digital strategy requirements are growing ever complex, increasing demands on D&A governance. These demands will push governance programs over the edge, unless effective stewardship solutions are provided.

Obstacles

- Organizations are struggling to understand the unique needs of governance for analytics and business intelligence (BI). IT is well aware of what needs to take place in BI solutions, data warehouses and so on. By contrast, businesses have little interest in that work, but want to exploit key data for their most important initiatives.
- The market has moved beyond privacy as a short-term driver, and has reverted to a broad-based approach. Vendors are once again looking to be visionary with various forms of consolidation of capabilities. This shift will both hinder and speed up formation of the expected D&A governance platform on which stewardship will be supported for all policy categories. This explains why this innovation will be obsolete before plateau.
- Some D&A governance vendors remain fixated on policy execution (i.e., data management), perceiving it as an "easier" sell. Others remain fixated on IT roles and do not focus on the needs of business roles.

User Recommendations

- Evaluate the capabilities needed from fit-for-purpose, business-user-oriented information stewardship solutions, as compared with IT-centric data management tools, including data quality, metadata management and federation/integration capabilities.
- Run a proof of concept for vendor solutions involving all contributing roles, such as business users, information governance board members, information architects, information stewards and business analysts.
- Focus on all dimensions (people, process, technology and data) when addressing the D&A stewardship use case. This holistic focus will help you maximize ROI through reuse, while minimizing administrative costs and errors due to inconsistencies across technologies.
- Explore the emerging D&A governance platforms for all their data and analytics stewardship operational requirements.

Sample Vendors

Alation; Alex Solutions; Collibra; Global Data Excellence; Informatica; OvalEdge; Prodago; SAP

Gartner Recommended Reading

A Day in the Life of a Data and Analytics Steward

Market Guide for Data and Analytics Governance Platforms

Tool: Vendor Identification for Data and Analytics Governance Platforms

Tool: The Gartner Data and Analytics Governance Technology Atlas

Knowledge Graphs

Analysis By: Afraz Jaffri

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

Knowledge graphs are machine-readable representations of the physical and digital worlds. They include entities (people, companies, digital assets) and their relationships, which adhere to a graph data model — a network of nodes (vertices) and links (edges/arcs).

Why This Is Important

Knowledge graphs capture information about the world in an intuitive way yet are still able to represent complex relationships. Knowledge graphs act as the backbone of a number of products, including search, smart assistants and recommendation engines. Knowledge graphs support collaboration and sharing, exploration and discovery, and the extraction of insights through analysis. Generative AI models can be combined with knowledge graphs to add trusted and verified facts to their outputs.

Business Impact

Knowledge graphs can drive business impact in a variety of different settings, including:

- Digital workplace (e.g., collaboration, sharing and search)
- Automation (e.g., ingestion of data from content to robotic process automation)

- Machine learning (e.g., augmenting training data)
- Investigative analysis (e.g., law enforcement, cybersecurity or financial transactions)
- Digital commerce (e.g., product information management and recommendations)
- Data management (e.g., metadata management, data cataloging and data fabric)

Drivers

- The need to complement AI/ML methods that detect only patterns in data (such as the current generation of foundation models) with the explicit knowledge, rules and semantics provided by knowledge graphs.
- Increasing awareness of the use of knowledge graphs in consumer products and services, such as smart devices and voice assistants, chatbots, search engines, recommendation engines, and route planning.
- The emerging landscape of Web3 applications and the need for data access across trust networks, leading to the creation of decentralized knowledge graphs to build immutable and queryable data structures.
- Improvements in graph DBMS technology that can handle the storage and manipulation of graph data structures at scale. These include PaaS offerings that take away the complexity of provisioning and optimizing hardware and infrastructure.
- The desire to make better use of unstructured data held in documents, correspondence, images and videos, using standardized metadata that can be related and managed.
- The need to manage the increasing number of data silos where data is often duplicated, and where meaning, usage and consumption patterns are not well-defined.
- The use of graph algorithms and machine learning to identify influencers, customer segments, fraudulent activity and critical bottlenecks in complex networks.

Obstacles

- Awareness of knowledge graph use cases is increasing, but business value and relevance are difficult to capture in the early implementation stages.
- Moving knowledge graph models from prototype to production requires engineering and system integration expertise. Methods to maintain knowledge graphs as they scale — to ensure reliable performance, handle duplication and preserve data quality — remain immature.
- The graph DBMS market is fragmented along three properties: type of data model (RDF or property), implementation architecture (native or multimodal) and optimal workload (operational or analytical). This fragmentation continues to cause confusion and hesitation among adopters.
- Organizations want to enable the ingestion, validation and sharing of ontologies and data relating to entities (such as geography, people, events). However, making internal data interoperable with external knowledge graphs is a challenge.
- In-house expertise, especially among SMEs, is lacking, and identifying third-party providers is difficult. Often, expertise resides with vendors of graph technologies.

User Recommendations

- **Create a working group of knowledge graph practitioners and sponsors** by assessing the skills of D&A leaders and practitioners and business domain experts. Highlight the obstacles to dependable and efficient data delivery for analytics and AI, and articulate how knowledge graphs can remove them.
- **Run a pilot to identify use cases that need custom-made knowledge graphs.** The pilot should deliver not only tangible value for the business, but also learning and development for D&A staff.
- **Create a minimum viable subset that can capture the information of a business domain to decrease time to value.** Assess the data, both structured and unstructured, needed to feed a knowledge graph, and follow Agile development principles.
- **Utilize vendor and service provider expertise** to validate use cases, educate stakeholders and provide an initial knowledge graph implementation.
- **Include knowledge graphs within the scope of D&A governance and management.** To avoid perpetuating data silos, investigate and establish ways for multiple knowledge graphs to interoperate and extend toward a data fabric.

Sample Vendors

Cambridge Semantics; Diffbot; eccenca; Neo4j; Ontotext; Stardog; TigerGraph; TopQuadrant; Trace Labs

Gartner Recommended Reading

How to Build Knowledge Graphs That Enable AI-Driven Enterprise Applications

3 Ways to Enhance AI With Graph Analytics and Machine Learning

Working With Graph Data Stores

How Large Language Models and Knowledge Graphs Can Transform Enterprise Search

Augmented Data Catalog/Metadata Management

Analysis By: Guido De Simoni, Brian Lowans, Robert Thanaraj

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Early mainstream

Definition:

Metadata management solutions are software that includes one or more of the following: metadata repositories, a business glossary, data lineage, impact analysis, rule management, semantic frameworks, and metadata ingestion and translation from different data sources. Modern AI-driven augmented data cataloging is part of the solutions automating metadata discovery, ingestion, translation, enrichment, and the creation of semantic relationships between both business and security metadata.

Why This Is Important

Augmented data catalogs and metadata management solutions support organizations managing varied data assets. Demands for accessing, using and sharing data are not limited to IT as data-oriented citizen roles emerge. Data and analytics (D&A) and security leaders face growing security and privacy risks, necessitating new data management approaches. With the pervasive use of data across a distributed data landscape by citizen users, augmented data catalogs can support metadata discovery and inventory management automation.

Business Impact

D&A leaders investing in augmented data cataloging (ADC) and metadata management solutions will see benefits from:

- **Collaboration:** Metadata requires the contribution of many people and ADC and metadata management solutions can provide a multiuser environment to address the requirements.
- **Automation of processes:** As data changes, metadata management solutions can streamline many recurring activities by (partial) automation.
- **Cost optimization:** Metadata management solution ensures that organizations understand the datasets, workloads, queries and tools being utilized, and highlights redundant tools and technologies.

Drivers

- Augmented data cataloging and metadata management solutions can remediate suboptimal results in an organization's use of data due to improperly managed metadata. This saves time, effort and money, while ensuring organizations are not exposed to unnecessary risks.
- Innovation generated by active metadata, reducing human effort in inventorying and managing data assets, is accelerating augmented data cataloging and metadata management solutions. Humans are primarily validators as opposed to doers of the operational tasks associated with the metadata management solutions.
- Informal and formal teams emerge and convert to community participation with as much automation as possible when supported by augmented data cataloging and metadata management solutions. These demands are only starting to be addressed by vendors, with modern metadata management practices slowly being established within organizations.
- Enterprise data cataloging techniques are emerging from several vendors that combine business, security and privacy metadata. This enables cross-functional operations for both D&A and security products to leverage the same metadata management solutions.

Obstacles

- The lack of maturity of strategic business conversations about metadata management solutions, as historically, enterprises have struggled to understand, define and use metadata showing business value.
- The expensive, but required, effort to integrate metadata management solutions in multivendor environments. This inhibitor has started to be addressed by vendors and community initiatives relating to openness and interoperability (see, for example, the open-source Egeria).
- The lack of identification of metadata management solutions with capabilities that meet the current and future requirements of specific use cases.

User Recommendations

- Recognize that the augmented data cataloging and metadata management solutions market will take two to five years to reach the Plateau of Productivity as the technology continues to expand both capabilities and support for existing and emerging use cases.
- Evaluate the metadata management capabilities of your company's existing tools, including data integration, data quality and master data capabilities, before buying a modern metadata management solution.
- Pilot the use of metadata management solutions for emergent use cases, including D&A governance, security and risk, and augmented data value for analytics.
- Invest in augmented data cataloging and metadata management solutions to augment manual efforts needed to access all types of metadata and analyze the data to support your company's data fabric designs for automation.
- Drive metadata insights to mitigate risks by informing your business leaders and their teams to ask more relevant questions of the data around them.
- Cooperate with the security and privacy teams to establish if enterprise data cataloging techniques would be beneficial.

Sample Vendors

Alation; Alex Solutions; Atlan; BigID; Collibra; data.world; IBM; Informatica; Zeenea

Gartner Recommended Reading

Market Guide for Active Metadata Management

How to Succeed With Data Classification Using Modern Approaches

Quick Answer: What Is Active Metadata?

Quick Answer: What Is Data Fabric Design?

Operational Intelligence

Analysis By: Henry Cook

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

Operational intelligence provides analytical processing within transactionlike workloads. Two forms of the technology exist. In the first, transactional systems extend to include advanced analytics, artificial intelligence (AI) and machine learning (ML). In the second, analytical databases extend to enable highly concurrent transactionlike queries. In-memory computing (IMC) technology is a key enabler of operational intelligence.

Why This Is Important

Operational intelligence provides analytical processing within transactionlike workloads. It enables transactional business processes to be steered in real time, assisted by analytics, AI and ML. Example use cases include planning, dynamic repricing, forecasting and what-if analysis. The analytics become an integral part of the business process itself, rather than a separate activity performed after the fact.

Business Impact

Operational intelligence enables business users to make more-informed operational and tactical decisions in real time. It opens up the possibility of driving prescriptive decisions without human intervention — that is, without relying solely on a business leader's intuitive situational awareness of operations. Operational intelligence also provides constantly updated forecasts and simulations of future business outcomes.

Drivers

- Operational intelligence subsumes “augmented transactions” (the previous name of this innovation). In operational intelligence, the transactional and analytical processing dimensions are designed together in the context of an individual transaction or process, and are delivered as a single application. This packaging helps streamline the overall technology infrastructure. Operational intelligence and augmented transactions were formerly the two flavors of hybrid transactional/analytical processing (HTAP) — named “in-process HTAP” and “point-of-decision HTAP.”

- Because of in-memory database technology and other techniques, organizations, for some types of analysis, no longer need to rely on external or after-the-fact analytics. These techniques implement significant analyses within a transaction. For example, they can execute predictive models or run significant OLAP queries. Likewise, analytical systems have improved, allowing large numbers of transactionlike queries to be executed concurrently with fast response times.

- Automated recommendations, based on real-time analysis of the impact of different options on KPIs, can improve decisions. Performance and scalability limitations have historically prevented advanced analytics from running concurrently with or within transaction processing.

- Many DBMSs are enabling ML within the database, increasing the functionality available for operational intelligence. Further, the growth of in-memory DBMSs, including the addition of in-memory functionality to general DBMSs, is enabling the use of operational intelligence.

Obstacles

- Operational intelligence may require significant migration of function when organizations are retrofitting existing (custom or packaged) applications at the core of business operations.
- Skills in operational intelligence are limited. This discipline requires a good understanding of designing for performance and the trade-offs involved. However, as the technology improves, the skills requirements are easing.
- Other than packaged applications such as Oracle and SAP, adoption of operational intelligence has been slow. However, it is now becoming more commonplace with, for example, reporting in smart cities and real-time reporting in manufacturing systems.
- Organizations may not be fully aware of the availability of mixed transactional and analytical processing features and the addition of some analytical functions to transactional systems. Not all use cases can or should be addressed through operational intelligence. Suitability must be determined on a use-case-by-use-case basis.

User Recommendations

- Pilot operational intelligence in individual “system of innovation” projects to get a sense of how much mixed transactional and analytical processing is possible.
- Ask strategic information management and business application providers about their vision, roadmap and technology for using operational intelligence within their products — that is, their support for real-time and near-real-time working.
- Employ operational intelligence for use cases involving observation data (such as IoT) or interaction data (such as log data) for which real-time operational analytics is required.
- Educate business leaders about operational intelligence concepts and their importance. Brainstorm with them to identify concrete opportunities to rethink business processes and create applications that could not be implemented using traditional architectures.

Sample Vendors

Aerospike; IBM; Microsoft; Oracle; Redis; SAP; SingleStore; Teradata

Gartner Recommended Reading

Magic Quadrant for Cloud Database Management Systems

Critical Capabilities for Cloud Database Management Systems for Operational Use Cases

Critical Capabilities for Cloud Database Management Systems for Analytical Use Cases

Application Data Management

Analysis By: Andrew White, Tad Travis

Benefit Rating: Moderate

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

Application data management (ADM) is a technology-enabled discipline where business and IT work together to ensure uniformity, accuracy, stewardship, governance, semantic consistency and accountability for data in an application or suite, such as ERP, customer data platform or custom-made app. Application data is the consistent and uniform set of identifiers and extended attributes used within an application or suite for items like customers, products or prices.

Why This Is Important

Clients continue to be shocked to find their cloud and application vendors offer modern SaaS and business applications that take scant care of governance of the data they use. The vast majority of business application implementations (including ERP, supply chain management [SCM] and even CRM) still lack holistic solutions for governance and stewardship of data. Whereas master data management (MDM) applies governance to shared data across all applications, ADM applies governance to data in a specific application.

Business Impact

ADM can offer the following benefits:

- Application data, once identified, ensures the correct governance effort is aligned to the right kind of business impact the application should have.

- Stewardship roles in the business, and in operational and analytical use cases, can be determined more effectively.
- Business goals for business applications are more likely assured with a more organized data and analytics governance approach that includes application data.
- MDM programs will help govern such application data that is shared with other applications.

Drivers

- The vast majority of “successful” go-lives of business applications, such as ERP, CRM or custom-built applications, do not include any qualification of data and analytics governance. The result, very often observed in client inquiry, is that, on average, seven months after the go-live, organizations spot the vast array of small but noticeable business issues held hostage to the lack of governed data. Business performance and process integrity fail, and business outcomes start to be negatively impacted.
- MDM was and still is misunderstood. An MDM program should have a laserlike focus on the minimal number of most widely shared attributes describing things like customer and product. Bloated MDM programs will continue to fail, leading to a greater need to split the effort up and create distinct ADM programs/requirements.
- Digital business success hinges not on the quality and governance of all data equally, but on a graduated, efficient means to classify data and apply only the needed level of governance. Such growing demand on scaling digital business will, of necessity, drive increased need to recognize and adopt ADM.

Obstacles

- Some MDM programs associated with large, global ERP, CRM and SCM implementations mistakenly centralize all the work related to governing application data. Others create a hybrid organization across business and IT, and call it all "MDM" (when it isn't). Put another way, these programs conflate MDM and ADM, making both too slow, expensive and unwieldy. As a result, neither program is a success.
- The half-life of a successful business application go-live is, anecdotally, seven months. After that, clients tell Gartner, "We have lost control of our data." This situation has become acceptable because, overall, most organizations don't fail.
- The organization's ability to change is held back, and consequently, budgets are set that even support mediocrity via poor governance practices. This is not an acceptable way to run an organization, but too few data and analytics leaders stand up and say so.
- Traditional top-down governance programs lead to the same misunderstandings and poorly scoped initiatives.

User Recommendations

Starting with a focus on business outcomes to identify what data matters most, organize, classify and govern data based on which data drives the most important business outcomes:

- Identify your application data to scope ADM. That is, identify the data that matters most to a specific set of use cases supported by one application or suite like ERP, e-commerce, product information management or customer data platform.
- Examine reusing MDM solutions to support your ADM implementation — even if in a distinct instance. The business requirements are very similar — but the value propositions are different.
- Demand from your business application provider (and those in the cloud) the necessary capability to set (that is, govern) and enforce (that is, steward) information policy pertaining to data used in the application or suite.
- Implement ADM alongside any MDM program so that they can operate at their own speeds and benefit. They do align and share metadata in support of a wider enterprise information management (EIM) program.

Sample Vendors

ChainSys; Epicor; Oracle; PiLog Group; Tealium; Utopia Global

Gartner Recommended Reading

4 Master Data Best Practices for ERP

Why CIOs in Midsize Enterprises Must Emphasize ERP Data Management

Create a Master Data Roadmap With Gartner's MDM Maturity Model

Graph DBMS

Analysis By: Aaron Rosenbaum

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

Graph DBMSs are designed to store any data in multiple relationships at the same time. Most graph DBMSs use basic graph theory and are suitable for multiple types of interactions, including complex multihop queries, reasoning and inference, and algorithms for analytical workloads.

Why This Is Important

Graph DBMSs have built-in commands, controls and structures to reduce thousands of lines of relational code into dozens of graph-based instructions. Relational DBMSs (RDBMSs) are best-suited to specific relationships and optimized to follow pathways that are preapproved. Graph DBMSs allow business users to describe and query dynamic and complex domains more naturally and efficiently. This reduces the need to create very complex looping controls that do not scale well.

Business Impact

Graph support has widely demonstrated value. Customer 360, supply chain, transportation/logistics, financial fraud detection and network analysis are typical use cases. The COVID-19 pandemic increased awareness of the value of graph for contact tracing: connections from one kind of data to another. Correlations across multiple sensors in asset-intensive IoT use cases are fueling increasing marketplace interest. Embedded graph support of metadata use cases is becoming increasingly important.

Drivers

- The challenges of connecting a broader variety of data to uncover new insights, and hitting the wall with relational DBMS because of performance challenges, have accelerated experimentation.
- Cloud-first development and experimentation are increasing, facilitated by ingest and conversion of existing data in a low-cost, low-risk environment, and graph DBMSs are widely available in cloud platforms.
- The increasing importance of semantics in organizing and operating larger and more complex data fabrics is accelerating experimentation, adoption and training.
- The increasing visibility of use cases that exploit graph and achieve successes in solving previously intractable operational problems is raising interest even further. Mature, widely distributed multimodel DBMS products are embedding graph DBMS engines and/or functions, and both open-source and commercial graph DBMS offerings are maturing and gaining adherents.

Obstacles

- Modeling, loading, processing and analyzing graph data require new and uncommon skills, including the sometimes steep learning curve of a graph query language like Gremlin.
- Dueling architectures (knowledge graphs versus property graphs), languages and APIs, and varying support among vendors for libraries of machine learning algorithms make choosing a product complex.
- Existing analytical tools can often be successful for graph use cases without requiring a dedicated DBMS, making graph DBMS unnecessary.
- The inability of relatively more technical practitioners to adequately explain the benefits of graph to business sponsors continues. Moreover, many use cases are now moving to embedded functionality in other tools (AI, metadata and data integration lead the way), making the decision to plunge into an unfamiliar type of product less attractive.

User Recommendations

- Assess graph DBMS capabilities when performance requirements for highly nested or connected data fall outside current RDBMS processing capabilities; be sure to consider the graph functions of your current DBMS.
- Use open-source graph DBMS projects, or community editions of commercial ones, to experiment and gain experience. Ensure any open-source products you use in production environments are commercially supported. Scaling and reliability cannot be relied on with open-source offerings.
- Use graph DBMSs to render the relationships and traversal of data (as discovered by a data scientist or data science team) into a reusable form for data miners, data engineers and business analysts.
- Select graph DBMSs based on your analytics requirements. RDF-based and property graph-based architectures have capabilities that should be factored in when making your selection.

Sample Vendors

Amazon Web Services; ArangoDB; Cambridge Semantics; Dgraph; Microsoft; Neo4j; Ontotext; OpenLink Software; TigerGraph

Gartner Recommended Reading

Graph Steps Onto the Main Stage of Data and Analytics: A Gartner Trend Insight Report

Working With Graph Data Stores

Understanding When Graph Analytics Are Best for Your Business Use Case

How to Build Knowledge Graphs That Enable AI-Driven Enterprise Applications

Market Guide for Graph Database Management Systems

Augmented Data Management

Analysis By: Roxane Edjlali

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Early mainstream

Definition:

Augmented data management is the application of AI and ML for optimization and improved operations. Based on existing metadata and usage data, AI and ML are applied to tune operations and to optimize configuration, security and performance. Augmented data management can automate some data management tasks and create, manage and apply policy rules within different products, such as metadata management, master data management, data integration, data quality and database management systems.

Why This Is Important

AI and machine learning (ML) can automate data management capabilities, altering job roles, product design and overall data management processes. These solutions are being used not only to tune and optimize the use of the products themselves based on actual usage, including failures and poor performance, but also to suggest and implement new designs, schemas and queries. These solutions can even infer the semantics of the data to recommend structural improvements.

Business Impact

Augmented data management offers benefits in metadata management, data integration, MDM, data quality and DBMS. This will assist those engaged in data management by automating many manual, repetitive tasks performed today, and increasing accuracy and reducing time spent on these tasks. This allows valuable resources to perform other tasks with far more business value.

Enormous volumes of user data on a consistent infrastructure can improve results and offer opportunities for continuous training and retraining of models. As a result, cloud-based data management is increasingly used to drive competitive improvements, and some features are making their way into on-premises, private cloud deployments.

Drivers

Data management is made up of several disciplines; augmentation will have a different impact in each:

- **Metadata management** — Increasingly, AI and ML are used to explore and define metadata from the data, helping the analysts to evaluate metadata more rapidly, accurately and with reduced redundancy. Similarly, augmented data management functions can automatically catalog data elements during data extraction, access and processing.
- **Data integration** — To automate the integration development process, by recommending or deploying repetitive integration flows, such as source-to-target mappings.
- **Master data management** — MDM solution vendors will increasingly focus on offering AI- and ML-driven configuration and optimization of record matching and merging algorithms as a part of their information quality and semantics capabilities.
- **Data quality** — AI and ML will be used to extend profiling, cleansing, linking, identifying and semantically reconciling master data in different data sources.
- **DBMS** — In addition to enhancing performance and cost-based query optimization, AI and ML are being used to automate many current manual management operations, including the management of configurations, elastic scaling, storage, indexes and partitions, and database tuning.

Obstacles

- The adoption of these capabilities is gated by the movement of the product categories to the cloud, where they are delivered first. Offerings in data integration, data quality, MDM, metadata management and DBMS software are proliferating and maturing rapidly but at different speeds.
- As this profile therefore represents an aggregate view and position of what is happening across data management, some areas are more mature in augmented data management than others.

User Recommendations

For data and analytics leaders focused on data management capabilities, we recommend you:

- Create a business case for using these new capabilities, and be sure to model and measure the benefits realized from the resources that will be released for other functions of greater business value.
- Question the vendors of your data management tools about their roadmap for the introduction of AI and ML into their products. Make augmented capabilities a “must have” selection criterion for new purchases of data management products.
- Begin testing the components of augmented data management products (where visible) to understand their capabilities and the validity of the automated functionality. Audit the results: With any new functionality, there is the risk of introducing errors and reduced performance.
- Plan for roles to change. Provide new skills training to add value as responsibilities evolve.
- Seek data management solutions that share design and performance metadata for external usage.

Sample Vendors

Amazon Web Services (AWS); Cinchy; CluedIn; Google Cloud Platform (GCP); IBM; Informatica; Microsoft; Oracle; SAP; SnapLogic

Gartner Recommended Reading

Market Guide for Active Metadata Management

Innovation Insight: State of Data Management Support for Self-Service

The State of Data Quality Solutions: Augment, Automate and Simplify

Understanding Modern MDM

Critical Capabilities for Data Integration Tools

Data Engineering

Analysis By: Robert Thanaraj, Ehtisham Zaidi, Roxane Edjlali

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

Data engineering is the discipline of translating raw data into usable forms by building and operationalizing data pipelines across various data and analytics systems meeting business requirements, data governance principles and SLAs. Data engineering is a team competency that brings together three different practices — data management, software engineering and I&O — to achieve frictionless, trustworthy data delivery with agreed-upon business and technical SLAs.

Why This Is Important

Data engineering caters to the full data supply chain provisioning trusted, quality data to be used at the right business moments with agreed service levels. It enables the creation, operationalization and maintenance of data pipelines across heterogeneous environments, aimed at delivering integrated data to consumer needs. It also manages data delivery debt around reusability, governance, compliance and operational readiness through data cataloging, data quality and data observability efforts.

Business Impact

Data engineering has the following business impacts:

- Good-quality data to improve data trust among the final consumers.

- Faster time to onboard new data (own or third-party) to existing analytics and data science models with robust data pipeline change management.
- Easier fulfillment of regulatory requirements to meet data transparency expectations through cataloging efforts in tracking sensitive data and enabling data governance enforcement.

Drivers

- **Productivity through automation:** Data management teams spend most of their time on data preparation, data integration and operationalization, and as a result, these are the primary candidates for automation.
- **Agility:** Organizations are forced to change the way they traditionally work with waterfall ways of data pipeline delivery because they are unable to keep up with the demand increase and skills shortage. As a result, they must adopt DataOps approaches to be more agile.
- **Self-service and customer experience:** Organizations seek successful consumer experiences, the last mile in the “data insights decisions” continuum. Data engineering enables self-service data management among citizen users and domain experts; and also provides guardrails and established best practices to follow.
- **Cloud:** As organizations struggle to manage multiple data gravities across their on-premises and multicloud setup, the data engineering practice plays a major role in balancing the collect and connect approaches in distributed architectures.
- **Data science success:** Organizations are bound to fail if they launch data science initiatives without onboarding the necessary data engineering skills because of high-technical debt associated with data management (such as governance, compliance and operations readiness); which must be managed.

Obstacles

- **Inefficient legacy practices:** Baggage around poor integration and operations practices hurts and/or delays data engineering practice adoption.

- **Unrealistic expectations:** Many think a data engineer can “do it all,” catering to the full spectrum from data management to software engineering to infrastructure and operations. Data engineering is a team competency. This involves data engineering, data architecture, data modeling, data stewardship, software engineering skills (like DevOps engineer, Python developer, test engineer), domain knowledge and operationalization expertise. Sometimes, organizations might even need adjacent roles for test engineering and infrastructure automation.

- **Inability to scale:** Data engineer roles are a critical part of data engineering teams. However, adding more data engineers in response to increasing data demands is not sustainable. Organizations need DataOps practices to streamline and scale data engineering delivery.

User Recommendations

- Establish a data engineering discipline with roles that support end-to-end delivery of data pipelines aligned to business use cases, SLAs and governance requirements.

- Catalog an inventory of data assets and make them searchable. Use metadata to drive automation of data pipelines and related artifacts. Study data usage patterns among consumers and systems, and employ this metadata to improve efficiency and optimize delivery.

- Introduce targeted, use case-specific tools — such as data warehouse automation tools and data preparation tools — to accelerate data pipeline builds and operations.

- Evaluate progress measures quantitatively — such as time to market, productivity, CI/CD automation of data pipelines, code quality and cost-efficiency of build and operations — regularly and share them with your stakeholders. As a data engineering leader, add this to your communication plan.

Sample Vendors

Ascend.io; CloverDX; dbt Labs; Nexla; Prophecy; Upsolver

Gartner Recommended Reading

5 Ways to Enhance Your Data Engineering Practices

Critical Capabilities for Data Integration Tools

Data and Analytics Essentials: DataOps

Climbing the Slope

Data Lake

Analysis By: Roxane Edjlali, Michele Launi

Benefit Rating: Moderate

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

A data lake is a collection of data storage instances combined with one or more processing capabilities. Most data assets are copied from diverse enterprise sources and are stored in their diverse formats, so they can be refined and repurposed repeatedly for multiple use cases. Ideally, a data lake will store and process data of any structure, latency or container (files, documents, result sets, tables, formats, BLOBs, messages, etc.).

Why This Is Important

Data lakes enable advanced analytics and complement traditional data warehouses. For example, the massive repository of source data in a data lake supports broad, flexible and unbiased data exploration, which is required for data mining, statistics, ML and other analytics techniques. A data lake can also provide scalable and high-performance data acquisition, preparation and processing, either to be refined and loaded into a data warehouse, or for processing within the data lake.

Business Impact

A data lake can be a foundation for multiple forms of business analytics. For example, data science is a common first use case for a data lake, which leads to predictive analytics that help a business retain customers, execute impact analyses, and anticipate issues in maintenance, logistics, risk and fraud. Similarly, using a data lake for self-service data access is a growing business use case that contributes to programs for business transformation and digitization.

Drivers

- User organizations are increasingly driven by data and analytics. This is so they can achieve their goals in business transformation, digitization, data democracy, operational excellence and competitiveness. A data lake provides data and supports analytics for these high-value goals.

- Organizations need to expand their analytics programs. Established forms of analytics will continue to be relevant — namely reports, dashboards, online analytical processing (OLAP) and statistical analysis. Hence, organizations must maintain these while expanding into advanced forms of analytics, such as data mining, natural language processing (NLP), machine learning, artificial intelligence and predictive analytics. A data lake provides the scale, as well as the structure-agnostic storage and processing options, that advanced analytics require.

- Data exploration and data engineering has become a common practice. This is true for many user types, from data scientists and analysts to business end users who are capable of self-service data prep. A data lake, when designed properly, can provision data for the diverse exploration requirements of multiple user types and use cases.

- Data lakes can expand the data warehouse and address additional use cases, such as data exploration on data. In these cases, the warehouse and lake are integrated by shared refined datasets, platform infrastructure (DBMS brands and storage, whether on-premises or cloud) and architecture components (data landing/staging).

Obstacles

- Data lake best practices are still evolving. There is still much confusion about how to design and govern a data lake, plus how to optimize a lake's data without losing its purpose as a repository for data science and advanced analytics. An emerging practice clears this obstacle by designing the internals of a data lake as a series of data zones for business use cases (data science, exploration and self-service) and technology architectural components (data land/staging and special data structures or latencies).

- The first data lakes, built on Hadoop, were for data science only, and they lacked metadata, relational functionality and governance. Today's data lake is on cloud, it has different data storage types, and it supports multiple analytics techniques (not just data science). Data governance is crucial and cannot be neglected; it includes data quality, data catalog, data security and data life cycle management.

User Recommendations

- Build a competency in data science and advanced analytics by first building a data lake as a foundation.

- Staff the data lake for maximum value by hiring data scientists, data engineers and analysts who have the skills required to conduct data exploration and analytics with the lake's data.
- Create business value by designing a data lake that addresses multiple high-value business use cases, such as data science, analytics, self-service data access or customer 360.
- Enable broad data exploration, multiple analytics techniques, and machine learning by populating a data lake with broadly collected data in various structures, formats and containers.
- Modernize the whole data architecture to extend the data lake. Consider logical data warehouse and lakehouse concepts.
- Keep each data lake from becoming a data swamp by governing the use of data in the lake, curating the data allowed into the lake, and documenting data via metadata and other data semantics.

Sample Vendors

Amazon Web Services (AWS); ChaosSearch; Cloudera; Databricks; Dremio; Google Cloud Platform (GCP); Infoworks; Microsoft; Snowflake

Gartner Recommended Reading

Building Data Lakes Successfully — Part 1 — Architecture, Ingestion, Storage and Processing

Building Data Lakes Successfully — Part 2 — Consumption, Governance and Operationalization

Data and Analytics Essentials: Data Warehouses, Data Lakes and Data Hubs

Market Guide for Analytics Query Accelerators

Data Classification

Analysis By: Ravisha Chugh, Bart Willemsen, Andrew Bales

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Mature mainstream

Definition:

Data classification is the process of organizing information assets using an agreed-upon categorization, taxonomy or ontology. The result is typically a large repository of metadata useful for making further decisions. This can include the application of a tag or label to a data object to facilitate its use and governance, either through the application of controls during its life cycle, or the activation of metadata using data fabric.

Why This Is Important

Data classification facilitates effective and efficient prioritization of data within data governance and data security programs concerned with value, access, usage, privacy, storage, ethics, quality and retention. It is vital to security, privacy and data governance programs. Data classification helps organizations distinguish the sensitivity of the data that they process, promotes a risk-based approach and improves the effectiveness of data protection controls.

Business Impact

Data classification supports a wide range of use cases, such as:

- Implementation of data security controls
- Privacy compliance
- Enablement of purpose-based access controls
- Risk mitigation
- Master data and application data management
- Data stewardship
- Content and records management
- Data catalogs for operations and analytics
- Data discovery for analytics and application integration
- Efficiency and optimization of systems, including tools for individual DataOps

Drivers

- Data classification approaches — which include classification by type, owner, regulation, sensitivity and retention requirement — enable organizations to focus their security, privacy and analytics efforts on important datasets.
- When properly designed and executed, data classification serves as one of the foundations supporting ethical and compliant processing of data throughout an organization.
- Data classification is also an essential component of data governance, as by classifying the data, organizations can establish data retention, data access and data protection policies that can help reduce the risk related to data exfiltration.

Obstacles

- Data classification initiatives have often failed because they were dependent on manual efforts by users with insufficient training.
- Data classification adoption is typically a reflection of the security posture of the organization. If the purpose of data classification is not clearly defined for employees using natural language, engagement in the data classification program is minimized.
- Data classification often fails due to poor communication. Program objectives, policies and procedures should be effectively communicated to all necessary stakeholders to avoid resistance to data classification initiatives.
- Although many vendors offer automated data classification tools that can classify more data more accurately while minimizing user effort, they are not 100% accurate — especially if they use machine learning or artificial intelligence algorithms for which models require ongoing training.

User Recommendations

- To identify, tag and store all of their organization's data, security and risk management leaders and chief data officers should collaboratively architect and use classification capabilities.
- Implement data classification with user training as part of a data governance program.
- Use a combination of user-driven and automated data classification for success in a data classification program.
- Determine organizationwide classification use cases and efforts, and, at minimum, keep all stakeholders informed.
- Combine efforts to adhere to privacy regulations with security classification initiatives. Information can be classification-based by nature (i.e., personally identifiable information, protected health information or PCI information), or by type (i.e., contract, health record or invoice. Records should also be classified by risk category, so as to indicate the need for confidentiality, integrity and availability. Additionally, records can be classified to serve specific purposes.

Sample Vendors

BigID; Concentric AI; Congruity360; Microsoft; Netwrix; OneTrust; SecuritiAI; Spirion; Varonis

Gartner Recommended Reading

Building Effective Data Classification and Handling Documents

Improving Unstructured Data Security With Classification

How to Succeed With Data Classification Using Modern Approaches

Video: What Is Data Classification, and Why Do I Need It?

Time Series DBMS

Analysis By: Rick Greenwald

Benefit Rating: Moderate

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

Time series DBMSs are designed to provide rapid ingestion of data, manipulation of data based on its position in a time series and aggregation of data outside of an active window of data. Once data leaves the active window, downsampling can provide an aggregate view of historical data.

Why This Is Important

Time series DBMSs are well-suited for many types of Internet of Things (IoT) and financial services systems. The requirements of rapid ingestion and high throughput of data combined with real-time analysis of incoming data are typical tasks for time series DBMSs. These products can keep detailed data for a time and aggregations based on that data for a longer time. Time series DBMSs can compare and analyze data across multiple time streams, which is difficult to do outside of this technology.

Business Impact

Time series DBMSs are useful for some real time, IT infrastructure and operations, industrial operations and financial systems where time series measurements and rapid ingestion of data are critical. These DBMSs can be used with other DBMS offerings as part of an overall data solution. Cloud provides best-fit solutions to many groups of use cases without significantly increasing management overhead, so using time series cloud offerings for specific use cases can be a more viable approach.

Drivers

- The expansion of IoT data sources has sparked increasing interest in this technology, as organizations look to both perform more sophisticated analytics on IoT data and combine that data with more traditional sources.
- Leading cloud vendors, such as Microsoft and Amazon Web Services, are adding time series capabilities or services to their platforms.
- The emergence of cloud data ecosystems will allow offerings with a specific focus, like time series capabilities, to be more easily integrated with the overall data landscape.

Obstacles

- Time series capabilities are being added to mainstream database offerings, and time series database services are being offered by major cloud providers. These dual developments, coupled with the ability to do standard analysis without specific time series capabilities for many use cases, place pressure on DBMSs whose main benefit is providing time series capability.

User Recommendations

- Choose a time series DBMS when you need high levels of performance for specific time series tasks.
- Choose a time series DBMS if your use case fits the target profile — real-time or historical analysis of append-only event data that is in time order — to add increased flexibility, performance and agility for processing the data.
- Choose a time series DBMS if there is easy integration into your overall data ecosystem.
- Choose a more standard DBMS or newer cloud service if your use case does not have demanding requirements for ingestion or analysis, since the time series capabilities may be adequate for your needs.

Sample Vendors

Amazon Web Services; Crate.io; IBM; InfluxData; KX Systems; Microsoft; Redis; Schneider Electric (AVEVA); ServiceNow (Era Software); Timescale

Gartner Recommended Reading

Time Series Database Architectures and Use CasesHow to Use Real-Time and Right-Time Database Analytics

Master Data Management

Analysis By: Sally Parker

Benefit Rating: High

Market Penetration: More than 50% of target audience

Maturity: Mature mainstream

Definition:

Master data management (MDM) is a technology-enabled business discipline in which business and IT work together to ensure the uniformity, accuracy, stewardship, governance, semantic consistency and accountability of the enterprise's official shared master data assets. Master data is the least number of consistent and uniform sets of identifiers and extended attributes that describe the core entities of an enterprise.

Why This Is Important

MDM is a cross-organizational collaborative effort that focuses on the consistency, quality and ongoing stewardship of master data. Master data is the subset of data that describes the core entities an organization requires to function — customers, citizens, products, suppliers, assets and sites. Master data sits at the heart of the most important business decisions, driving a need for a consistent view across business silos.

Business Impact

MDM initiatives are progressing as a foundational component of digital transformation. Leading organizations draw a causal link between their master data (parties, things and places) and the business outcomes it supports, including customer retention, supply chain optimization, and risk and regulatory compliance.

Interest in MDM extends to a broad range of vested-interest stakeholders, including finance, marketing and supply chain. MDM is now mainstream. Organizations seeking a single view of their master data recognize it as a necessity.

Drivers

- MDM is not a new concept, but adoption varies across geographic regions, with North America the most mature region, followed by Western Europe. The rest of the world is earlier in the maturity cycle and representative of markets primed for growth.
- Business process integrity eludes organizations with complex or heterogeneous application and data landscapes. Such organizations can suffer from inconsistent master data and/or a lack of trust in their master data. Organizations are increasingly recognizing the direct and causal link between this data and business outcomes, which MDM is designed to address.
- Rapidly evolving business needs, particularly in uncertain times, translate into greater demand for the benefits afforded by MDM — notability agility. The COVID-19 pandemic, which initially stalled projects, ultimately served to fast-track a broader realization of the causal link between trusted and connected master data and business resilience.
- Interest levels are increasing across a broader range of stakeholders (beyond technology), in both private and public sectors.
- A prior hesitance to embark upon MDM initiatives, due to complexity and cost, is easing.
- The barrier to entry has dropped significantly over the past two years with the broader availability of cloud-based and subscription-based MDM vendor offerings, which are now the most dominant offerings for net new clients. This lowering of the barrier to entry renders MDM viable for a broader target audience that comprises small and midsize organizations.
- A shift in mindset toward a more granular and business-outcome-led MDM program is reflected in the MDM vendors' “land and expand” strategies, where clients start small and progress toward incremental mastery of use cases and domains.
- Digital transformation requirements are forcing organizations to either start or modernize their MDM programs to leverage more recent cloud-based offerings and new augmented MDM capabilities.

Obstacles

- **Lack of consistent vendor presence:** Coverage is weaker outside North America and Europe.
- **Technology blinkers:** The prevailing pitfall remains the instinct to treat MDM as a technology initiative in isolation. Technology alone won't solve a challenge that traverses people, processes and technology.
- **Human factors:** Organizations that fail to proactively engage business stakeholders in scoping struggle to meet expectations of value and to establish an operational governance structure in service of MDM.
- **Goals:** MDM is still too often seen as an IT project. When MDM is a data or IT project that doesn't align to business outcomes, it fails.
- **Perceived complexity:** The MDM solutions market only recently shifted toward subscription pricing, cloud-based offerings and simpler products, which contribute to more approachable solutions and shorter deployment times.
- **Skills:** Successful MDM implementations require business acumen, technology and governance capabilities. Finding the right balance and availability of these skill sets remains problematic and is driving a need for third-party services as the norm.

User Recommendations

- Use business outcomes to identify the least amount of data with the greatest business impact.
- Approach MDM as a technology-enabled business-led initiative.
- Secure executive sponsorship to facilitate cross-organizational collaboration.
- Ensure that the causal link between the MDM initiative and the business outcomes it supports is clearly understood and articulated.
- Keep your master data attributes lean and focused.
- Leverage third-party services to fast-track time to value. The majority of organizations leverage external support with their MDM strategy and/or implementation. Third parties offering industry expertise and accelerators can greatly impact time to value.

Gartner Recommended Reading

3 Essentials for Starting and Supporting Master Data Management

Create a Master Data Roadmap With Gartner's MDM Maturity Model

Data and Analytics Essentials: Master Data Management — Presentation Materials

In-DBMS Analytics

Analysis By: Henry Cook

Benefit Rating: High

Market Penetration: More than 50% of target audience

Maturity: Mature mainstream

Definition:

In-DBMS analytics is also known as in-database analytics or in-database processing. This approach pushes data-intensive processing — such as data preparation, online analytical processing, predictive modeling, operations and model scoring — down into the DBMS platform, close to the data, in order to reduce data movement and support rapid analysis.

Why This Is Important

In-DBMS analytics enables agility and productivity. It is a robust way of getting machine learning (ML) and other analytics results into production. It can also have performance benefits. This technology is relatively new to cloud DBMS systems. Use of cloud in-DBMS analytics will likely increase.

Business Impact

In-DBMS analytics provides a robust way of developing advanced analytics, such as machine learning. It is an ideal vehicle for moving ML models into production and monitoring their effectiveness. This technology not only provides productivity and agility benefits, but also enables machine learning to be productionized more readily.

Drivers

- In-DBMS analytics eliminates the need for data copies and additional processing elements because those aspects can be managed within the database. Thus, in-DBMS analytics is cost-efficient from a data consumption standpoint. In-DBMS analytics offerings have been available from on-premises data warehouse software vendors for many years. They are now increasingly featured in cloud DBMS offerings, which are gaining acceptance and adoption.
- By avoiding the need to move data out of the DBMS in order to build analytical models, in-DBMS analytics enables more flexible experimentation and efficient development.
- Most DBMS vendors are offering in-DBMS analytics capabilities. The means of doing this are well-understood. Algorithms can be embedded within the DBMS, invoked using external libraries, or executed by integration with a separate data science service. In some DBMSs, the same facility also enables processing, such as load and transformation logic for ELT to assist with data preparation.
- ML is becoming more commoditized as its use spreads beyond specialist data scientists. In-DBMS ML is an excellent enabler for this wider group of developers and users.
- Three forces are encouraging adoption: the drive for greater productivity in the use of ML, the demand for easier ML administration and the need to reliably move ML into production. Adopting in-DBMS analytics provides a very good solution for moving analytics models to production — with model generation, administration and execution all in the same environment.

Obstacles

- Data scientists lack familiarity with DBMSs and tend to prefer R, Python and notebooks. DBMS professionals lack familiarity with ML and tend to prefer SQL.
- In-DBMS analytics requires a sufficient range of analytical algorithms, and this is becoming the norm. However, organizations still need to validate how in-DBMS analytics will fit into their overall estate and, most importantly, how the analytics will be monitored and controlled.
- Some implementations have performance and scalability restrictions. To be used at scale, the algorithms must not only be made available, but also be modified to take advantage of parallel processing. This is not a problem with most offerings, but needs to be checked in a proof of concept prior to adoption.
- Most ML capabilities are still handled through supplemental services outside the DBMS. Cloud providers can potentially charge more for separate services. In-DBMS capabilities also bring “stickiness” and dependencies between the DBMS and ML.

User Recommendations

- Use in-DBMS analytics for making large-scale business analytics available to a wider audience, and for embedding ML capabilities in applications to deliver rapid insights on historical and incoming data.
- Review your data science development process. It may be better-enabled through in-DBMS analytics, especially for deployment to production.
- Check whether in-DBMS analytics is supported when you’re evaluating DBMS systems. If so, verify the range of algorithms offered. Experiment with use cases where it is more efficient to bring ML algorithms to the data at scale.
- Be aware that support for programming languages, such as Python, Scala, Java and JavaScript, varies among products. Before committing to a DBMS brand, be sure it supports the languages you need.

Sample Vendors

Amazon Web Services; Fuzzy Logix; Google; IBM; OpenText (Micro Focus); Oracle; SAP; Snowflake, Teradata; VMware

Gartner Recommended Reading

5 Useful Ways to Use Artificial Intelligence and Machine Learning With Your Logical Data Warehouse

Market Guide for Multipersona Data Science and Machine Learning Platforms

Magic Quadrant for Cloud Database Management Systems

Event Stream Processing

Analysis By: W. Roy Schulte, Pieter den Hamer

Benefit Rating: Transformational

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

Event stream processing (ESP) is computing that is performed on streaming data (sequences of event objects) for the purpose of stream analytics or stream data integration. ESP is typically applied to data as it arrives (data “in motion”). It enables situation awareness and near-real-time responses to threats and opportunities as they emerge, or it stores data streams for use in subsequent applications.

Why This Is Important

ESP enables continuous intelligence and real-time aspects of digital business. ESP’s data-in-motion architecture is a radical alternative to the conventional data-at-rest approaches that have historically dominated computing. ESP platforms have progressed from niche innovation to proven technology, and now reach into the early majority of users. ESP will reach the Plateau of Productivity in less than two years and eventually be adopted by multiple departments within every large company.

Business Impact

ESP transformed financial markets and became essential to telecommunications networks, smart electrical grids, and some IoT, supply chain, fleet management and other transportation operations. However, most of the growth in ESP during the next 10 years will come from areas where it is already established, especially IoT and customer engagement. Stream analytics from ESP platforms provide situation awareness through dashboards and alerts, and detect anomalies and other significant patterns.

Drivers

Six factors are driving ESP growth:

- Organizations have access to ever-increasing amounts of low-cost streaming data from sensors, machines, smartphones, corporate websites, transactional applications, social computing platforms, news and weather feeds, and other data brokers. Many new AI and other analytical applications need this streaming data to satisfy business requirements for situation awareness and faster, more-accurate decisions.
- The wide use of Apache Kafka and similar streaming messaging systems is reducing the cost and complexity of ingesting, storing and using streaming data.
- Conventional data engineering pipelines take hours or days to prepare data for use in BI and analytics, causing delays that are unacceptable for some purposes. Therefore, an increasing number of data engineering pipelines are being reimplemented as real-time data flows (continuous ETL) in ESP platform products or stream data integration tools with embedded ESP. These real-time data flows filter, aggregate, enrich, and perform pattern detection and other transformations on streaming data as it arrives.
- ESP products have become widely available, in part because open-source ESP technology has made it less expensive for more vendors to offer ESP. More than 30 ESP platforms or cloud ESP services are available. All software megavendors offer at least one ESP product, and numerous small-to-midsize specialists also compete in this market. Cloud ESP platforms have lowered the cost of entry.
- Vendors are embedding ESP platforms into a wide variety of other software products, including industrial IoT platforms, stream data integration tools, unified real-time platforms (aka continuous intelligence platforms), insider threat detection tools and AI operations platforms.

- Vendors are adding highly productive development tools that enable faster ESP application development. Power users can build some kinds of ESP applications via low-code techniques and off-the-shelf templates.

Obstacles

- ESP platforms are overkill for many applications that process low volumes of streaming data (i.e., under 1,000 events per second), or that do not require fast response times (i.e., less than a minute). Conventional BI and analytics tools with data-at-rest architectures are appropriate for most stream analytics with these less-demanding requirements.
- Many architects and software engineers are still unfamiliar with the design techniques that enable ESP on data in motion. They are more familiar with processing data at rest in databases and other data stores, so they use those techniques by default unless business requirements force them to use ESP.
- Some streaming applications are better-implemented on unified real-time platforms that process both data in motion and data at rest. Some unified platforms use embedded open-source ESP platform products, while others get their ESP capabilities from custom internal code.

User Recommendations

- Use ESP platforms when conventional data-at-rest architectures cannot process high-volume streams fast enough to meet business requirements.
- Acquire ESP functionality through a SaaS offering, an IoT platform or an off-the-shelf application that has embedded ESP logic if a product that targets specific business requirements is available.
- Use vendor-supported closed-source platforms or open-core ESP products that mix open-source with closed-source extensions for applications that need enterprise-level support. Use free, community-supported, open-source ESP products if developers are familiar with open-source software, and license fees are more important than staff costs.
- Use ESP platforms or stream data integration tools to ingest, filter, enrich, transform and store event streams in a file or database for later use.
- Choose a unified real-time platform with embedded ESP capabilities over a plain ESP platform if the application uses both data at rest and data in motion.

Sample Vendors

Confluent; EsperTech; Google; Hazelcast; IBM; Microsoft; Oracle; SAS; Software AG; TIBCO Software

Gartner Recommended Reading

Market Guide for Event Stream Processing

5 Essential Practices for Real-Time Analytics

Create an Optimal IoT Architecture Using 5 Common Design Patterns

Adopt Stream Data Integration to Meet Your Real-Time Data Integration and Analytics Requirements

Entering the Plateau

Wide-Column DBMS

Analysis By: Aaron Rosenbaum

Benefit Rating: Moderate

Market Penetration: 20% to 50% of target audience

Maturity: Mature mainstream

Definition:

Wide-column database management systems (DBMSs) store rows of data in tables, similar to a relational DBMS. However, all fields in a row may not be present, and support for JOINs and referential integrity may be omitted. Flexible schema definitions make DBMSspopular for storing semistructured data, like log and sensor data. Relaxed consistency makes them attractive for geographically distributed deployments with challenging SLAs.

Why This Is Important

Wide-column DBMSs are widely used by enterprises processing growing amounts of structured and semistructured data in distributed operational and analytic data scenarios.

Business Impact

The current business impact of wide-column DBMSs is moderate. The ability to distribute massive amounts of semistructured data addresses numerous use cases, particularly in support of connected data and real-time operations. Vendors promoting specialized wide-column DBMS, training and support are driving adoption and accelerating development. In addition, multimodel DBMS collections and transactions provide similar functions with more flexible opportunities. Many include wide-column capability.

Drivers

- Cloud-based applications often require ingesting very large data volumes with a scalable and highly available architecture. Wide column stores are effective for use cases that require these characteristics.

- Significant efforts by hyperscalers to support and market wide-column offerings have increased visibility beyond the sizable existing user base. These include Microsoft's addition of the Cassandra API to Azure Cosmos DB, Amazon Web Services' (AWS's) introduction of Keyspaces (for Apache Cassandra) and Google's support for DataStax Astra DB in addition to its own Cloud Bigtable.
- Complexity of deployment and operation, formerly a challenge especially for open-source wide-column offerings, is being reduced by commercial providers. This is accelerating in the cloud with improved, often automated and even serverless operations.
- The pace of functional enhancement is picking up as cloud vendors battle for differentiation and deliver new releases more rapidly. An example is Cloudera's efforts to enhance the operational capabilities of Apache HBase by adding Apache Phoenix to its portfolio as a SQL front end.

Obstacles

- Multimodel databases, particularly those offered by hyperscalers, address many workloads' volume and scale requirements while supporting richer functionality.
- Because wide-column DBMSs do not typically support relationships between rows or tables, they are often unsuitable as a platform for data warehousing and complex analytics applications.
- The increasing pace of multimodel adoption is challenging products that cannot support both operational and analytical workloads. New projects will increasingly be deployed in more broadly capable offerings, reducing the expansion of pureplay wide-column DBMS into new use cases.

User Recommendations

- Model expected workloads and data volume. The right hardware or cloud instance types for wide-column DBMSs are critical for cost-effective delivery to your SLAs.
- Focus on operational use cases; the absence of JOINs degrades analytics performance.
- Develop the right data model to enable decisions between consistency and latency. Choose DBMS solutions that allow such trade-offs to be made. Developers must incorporate this into applications.

- Eliminate network bottlenecks preventing efficient distributed operations, including replication. Note that conflict resolution requires time synchronization across the cluster and between applications interacting with individual nodes.
- Leverage open-source offerings for experiments and pilots. Engage with vendors offering commercial support before moving to the production stage.
- Investigate the security capabilities of wide-column DBMS before implementing applications with significant risk profiles because these features still lag behind mature RDBMS.

Sample Vendors

Alibaba Cloud; Amazon Web Services; Cloudera; DataStax; Google; Microsoft; ScyllaDB

Gartner Recommended Reading

Comparison of Data Stores to Support Modern Use Cases

Decision Point for Selecting the Right DBMS

2023 Planning Guide for Data Management

The Impacts of Data Ecosystems: A Cloud Architectural Perspective

Multimodel DBMS

Analysis By: Aaron Rosenbaum

Benefit Rating: Moderate

Market Penetration: More than 50% of target audience

Maturity: Mature mainstream

Definition:

A multimodel database management system (DBMS) incorporates several data engines, relational and/or nonrelational (e.g., document, graph, key value, time series, wide column), in a single DBMS. It provides a common access mechanism for different persistence types, each optimized for the nature of the data being used.

Why This Is Important

Data and analytics professionals are confronted with increasing choices among DBMS data models, especially in the cloud. Choosing between multimodel and specialty DBMSs can influence the direction of your cloud architecture, use cases supported, staffing and vendor management. A multimodel DBMS can simplify operations, development and integration. However, it creates new skills requirements of its own, including adding the complexities of querying and different modeling requirements.

Business Impact

Multimodel capabilities help anchor incumbent products by making them suitable for additional use cases without introducing a new product and/or vendor. Most organizations will have them in their technology portfolio, even if they did not make an explicit choice to adopt multimodel capabilities. Embracing them can accelerate the introduction of new technologies, reduce operational complexity and reduce vendor management efforts — although it can potentially expand dependence on a single vendor.

Drivers

- Incumbency and familiarity are key drivers of further adoption; most widely deployed DBMSs are multimodel. Every leading vendor now has a widely adopted flagship offering that supports two, three or more models.
- Managing fewer vendors is always an attractive outcome for most organizations.
- Commercial barriers to more usage of the same product are lower in a usage-based licensing environment where no new negotiations and contracts need to be considered.
- Fewer integration issues compared to multiple services-based DBMSs make multimodel an attractive candidate for new development.
- Mature multimodel products often provide richer auditing, concurrency controls, versioning, distributed data complexity management, points of governance and security that specialty products lack.
- Developer preferences often drive multimodel usage, especially in the cloud. Putting new capabilities closer to existing practices promotes experimentation and lowered barriers to rapid deployment.
- Data lakes increasingly capture and manage multimodel data. As lakes proliferate, they drive increased adoption of multimodel DBMSs.

Obstacles

- Familiarity with incumbent product capabilities may be lacking. Gartner inquiries regarding multimodel DBMSs remain rare, reflecting minimal awareness — it is an analysts' abstraction, not a market label. That said, while some DBMS products do promote their multimodel capabilities, there are others that have extensive multimodel capabilities but do not market that functionality actively.

- Multimodel offerings may seem less friendly to new developers not already expert in their more traditional uses. The approachability of specialized products makes them attractive for different modes such as graph, with design and development tools more immediately productive for new users.

User Recommendations

- Assess multimodel capabilities of your incumbent DBMSs first, especially when planning new use cases. Extensions of familiar use cases, such as augmented transaction processing, native handling of file-based data (e.g., XML, JSON, Apache Parquet), and logical data warehouse deployments, may also benefit from simplified development.

- Evaluate requirements for usage of different models within the same query. Some products supporting multiple models do not support utilization of those models within the same database instance or query.

- Evaluate alternative products for a truly specialized use case such as generative AI, which might require use of a database capable of indexing vector embeddings.

- Choose vendors whose stability and track record are demonstrable, and whose roadmaps are consistent with your planned use cases.

Sample Vendors

Amazon Web Services; Couchbase; Datastax; EDB; Google; Intersystems; MarkLogic; Microsoft; MongoDB; Oracle

Gartner Recommended Reading

Choosing Between Multimodel DBMS and Multiple Specialized Engines

Comparison of Data Stores to Support Modern Use Cases

Decision Point for Selecting the Right DBMS

SQL Interfaces to Object Stores

Analysis By: Aaron Rosenbaum, Adam Ronthal

Benefit Rating: Moderate

Market Penetration: More than 50% of target audience

Maturity: Early mainstream

Definition:

SQL interfaces to object and file stores provide the ability for enterprises to interact with data residing in object and file stores (such as Apache HDFS, Apache Iceberg or cloud object stores) using familiar SQL syntax. The interfaces are often used in support of data lakes to accelerate data analysis and use SQL as the query language.

Why This Is Important

Cloud object stores like Amazon S3, Microsoft Azure Blob Storage and Azure Data Lake Storage, and Google Cloud Storage have become the common storage for data lakes and a clearinghouse for data between various cloud services. Apache HDFS and CEPH Object Storage are used for similar purposes. SQL interfaces to these stores allow data to be queried directly for many different use cases.

Business Impact

Object and file-based stores have become very important in the support of analytics including data warehouses, data marts, lake houses and ML platforms. The combination of lower cost storage together with standard SQL interface allows for some data to be queried at significantly lower cost vs. other alternatives. This has significant business impact for infrequently accessed data, very large datasets and often makes it simpler to share data by avoiding a loading process.

Drivers

- With improving support for the SQL standard, the different types of nonrelational data becomes accessible via a broader range of BI and visualization tools and applications, opening increasing numbers of use cases.
- Object stores have become a natural choice to store data cost-effectively without requiring a dedicated computer layer like DBMS to access the data.
- Open storage formats such as Apache Parquet and Apache Iceberg provide compression and improved performance through columnar format and support schema evolution.
- Increasing use of open standards, including Apache Arrow, Apache Drill, Apache Druid, Apache Hive, Apache Spark, Apache PrestoDB and Trino, by new and existing tools in multiple categories will drive continued development and enhancement of SQL interfaces in open and proprietary versions.

Obstacles

- The interfaces do not offer the same level of performance optimization that is available in a DBMS.
- Newer stand-alone tools have less mature capabilities for processing complex SQL statements, even for relatively modest multiway JOINs.
- Since most SQL interfaces to object storage do not support update, intermediate results will often not be persisted unless the tool in use provides its own persistence layer or has a mechanism to provide this functionality. This will limit their usefulness.

User Recommendations

- Assess production applicability for operational use cases that may have specific performance and concurrency requirements only once these offerings have proven their value in the cases described above.
- Seek these capabilities out in the products already being utilized in your architecture.
- Attempt to optimize multiple workload types, with the understanding that they may require several SQL interfaces, or a more broadly capable multitarget layer.
- Test core functionality and integration with adjacent D&A components like metadata management tools, BI tools and data governance solutions, to ensure that functional support and performance meet your needs.

Sample Vendors

Amazon; AtScale; Cloudera; Databricks; Google; Incorta; Kyligence; Microsoft; Oracle; Starburst

Gartner Recommended Reading

Market Guide for Analytics Query Accelerators

Exploring Lakehouse Architecture and Use Cases

Next-Generation SQL Engines

Appendixes

See the previous Hype Cycle: Hype Cycle for Data Management, 2022

Hype Cycle Phases, Benefit Ratings and Maturity Levels

Table 2: Hype Cycle Phases

(Enlarged table in Appendix)

Phase ↓	Definition ↓
Innovation Trigger	A breakthrough, public demonstration, product launch or other event generates significant media and industry interest.
Peak of Inflated Expectations	During this phase of overenthusiasm and unrealistic projections, a flurry of well-publicized activity by technology leaders results in some successes, but more failures, as the innovation is pushed to its limits. The only enterprises making money are conference organizers and content publishers.
Trough of Disillusionment	Because the innovation does not live up to its overinflated expectations, it rapidly becomes unfashionable. Media interest wanes, except for a few cautionary tales.
Slope of Enlightenment	Focused experimentation and solid hard work by an increasingly diverse range of organizations lead to a true understanding of the innovation's applicability, risks and benefits. Commercial off-the-shelf methodologies and tools ease the development process.
Plateau of Productivity	The real-world benefits of the innovation are demonstrated and accepted. Tools and methodologies are increasingly stable as they enter their second and third generations. Growing numbers of organizations feel comfortable with the reduced level of risk; the rapid growth phase of adoption begins. Approximately 20% of the technology's target audience has adopted or is adopting the technology as it enters this phase.
Years to Mainstream Adoption	The time required for the innovation to reach the Plateau of Productivity.

Source: Gartner (July 2023)

Gartner

Table 3: Benefit Ratings

Benefit Rating ↓	*Definition* ↓
Transformational	Enables new ways of doing business across industries that will result in major shifts in industry dynamics
High	Enables new ways of performing horizontal or vertical processes that will result in significantly increased revenue or cost savings for an enterprise
Moderate	Provides incremental improvements to established processes that will result in increased revenue or cost savings for an enterprise
Low	Slightly improves processes (for example, improved user experience) that will be difficult to translate into increased revenue or cost savings

Source: Gartner (July 2023)

Table 4: Maturity Levels

(Enlarged table in Appendix)

Maturity Levels ↓	*Status* ↓	*Products/Vendors* ↓
Embryonic	In labs	None
Emerging	Commercialization by vendors Pilots and deployments by industry leaders	First generation High price Much customization
Adolescent	Maturing technology capabilities and process understanding Uptake beyond early adopters	Second generation Less customization
Early mainstream	Proven technology Vendors, technology and adoption rapidly evolving	Third generation More out-of-box methodologies
Mature mainstream	Robust technology Not much evolution in vendors or technology	Several dominant vendors
Legacy	Not appropriate for new developments Cost of migration constrains replacement	Maintenance revenue focus
Obsolete	Rarely used	Used/resale market only

Source: Gartner (July 2023)

Document Revision History

Hype Cycle for Data Management, 2022 - 30 June 2022

Hype Cycle for Data Management, 2021 - 27 July 2021

Hype Cycle for Data Management, 2020 - 15 July 2020

Hype Cycle for Data Management, 2019 - 31 July 2019

Hype Cycle for Data Management, 2018 - 25 July 2018

Hype Cycle for Data Management, 2017 - 26 July 2017

Hype Cycle for Information Infrastructure, 2016 - 8 July 2016

Hype Cycle for Information Infrastructure, 2015 - 13 August 2015

Hype Cycle for Information Infrastructure, 2014 - 6 August 2014

Recommended by the Author

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

Understanding Gartner’s Hype Cycles

Tool: Create Your Own Hype Cycle With Gartner’s Hype Cycle Builder

Data Management Solutions Primer for 2023

Data and Analytics Essentials: Data Fabric and Data Mesh

Data and Analytics Essentials: DataOps

Data and Analytics Essentials: Data Quality

Data and Analytics Essentials: Metadata Management

2023 Technology Adoption Roadmap for Data and Analytics Functions in Large Enterprises

---

Table 1: Priority Matrix for Data Management, 2023

Benefit	*Years to Mainstream Adoption*			
↓	***Less Than 2 Years*** ↓	***2 - 5 Years*** ↓	***5 - 10 Years*** ↓	***More Than 10 Years*** ↓
Transformational	Event Stream Processing	Augmented Data Quality Data Product	Active Metadata Management Augmented FinOps Data Fabric Generative AI for Data Management	
High	Augmented Data Management In-DBMS Analytics Operational Intelligence	Augmented Data Catalog/Metadata Management Data Classification Data Engineering Data Hub Strategy Data Observability DataOps Edge Data Management Graph DBMS Knowledge Graphs Lakehouse Master Data Management	D&A Governance Platforms Data Ecosystems Intercloud Data Management Vector Databases	

Benefit	Years to Mainstream Adoption			
↓	Less Than 2 Years ↓	2 - 5 Years ↓	5 - 10 Years ↓	More Than 10 Years ↓
Moderate	Multimodel DBMS SQL Interfaces to Object Stores Time Series DBMS	Application Data Management Data Lake	Distributed Transactional Databases Self-Service Data Management	
Low				

Source: Gartner (July 2023)

Table 2: Hype Cycle Phases

Phase	*Definition*
Innovation Trigger	A breakthrough, public demonstration, product launch or other event generates significant media and industry interest.
Peak of Inflated Expectations	During this phase of overenthusiasm and unrealistic projections, a flurry of well-publicized activity by technology leaders results in some successes, but more failures, as the innovation is pushed to its limits. The only enterprises making money are conference organizers and content publishers.
Trough of Disillusionment	Because the innovation does not live up to its overinflated expectations, it rapidly becomes unfashionable. Media interest wanes, except for a few cautionary tales.
Slope of Enlightenment	Focused experimentation and solid hard work by an increasingly diverse range of organizations lead to a true understanding of the innovation's applicability, risks and benefits. Commercial off-the-shelf methodologies and tools ease the development process.
Plateau of Productivity	The real-world benefits of the innovation are demonstrated and accepted. Tools and methodologies are increasingly stable as they enter their second and third generations. Growing numbers of organizations feel comfortable with the reduced level of risk; the rapid growth phase of adoption begins. Approximately 20% of the technology's target audience has adopted or is adopting the technology as it enters this phase.
Years to Mainstream Adoption	The time required for the innovation to reach the Plateau of Productivity.

Phase ↓	*Definition* ↓

Source: Gartner (July 2023)

Table 3: Benefit Ratings

Benefit Rating ↓	*Definition* ↓
Transformational	Enables new ways of doing business across industries that will result in major shifts in industry dynamics
High	Enables new ways of performing horizontal or vertical processes that will result in significantly increased revenue or cost savings for an enterprise
Moderate	Provides incremental improvements to established processes that will result in increased revenue or cost savings for an enterprise
Low	Slightly improves processes (for example, improved user experience) that will be difficult to translate into increased revenue or cost savings

Source: Gartner (July 2023)

Table 4: Maturity Levels

Maturity Levels ↓	*Status* ↓	*Products/Vendors* ↓
Embryonic	In labs	None
Emerging	Commercialization by vendors Pilots and deployments by industry leaders	First generation High price Much customization
Adolescent	Maturing technology capabilities and process understanding Uptake beyond early adopters	Second generation Less customization
Early mainstream	Proven technology Vendors, technology and adoption rapidly evolving	Third generation More out-of-box methodologies
Mature mainstream	Robust technology Not much evolution in vendors or technology	Several dominant vendors
Legacy	Not appropriate for new developments Cost of migration constrains replacement	Maintenance revenue focus
Obsolete	Rarely used	Used/resale market only

Source: Gartner (July 2023)