Hype Cycle for Finance Data and Analytics Governance, 2023

Published 20 July 2023 - ID G00785765 - 106 min read

By Analyst(s): Valeria Di Maso

Initiatives: Finance Data and Analytics

Legacy approaches to D&A governance are no longer practical in today's increasingly digital data world. This Hype Cycle will help FP&A leaders to make informed decisions about the latest finance D&A governance trends, explore the newest innovations and identify appropriate investment opportunities.

Strategic Planning Assumptions

- Through 2026, 80% of finance organizations' advanced analytics investments will fall short of expected ROI because they failed to adapt and modernize their enterprise data governance and management.
- By 2026, 90% of data governance will be enabled via active, semiautonomous technologies that reduce integrated data delivery time by 50%.
- By 2026, 20% of large enterprises will use a single data and analytics governance platform to unify and automate discrete governance programs.
- Through 2025, poor data quality will persist as one of the most frequently mentioned challenges — if not *the* most frequently mentioned challenge — prohibiting advanced analytics (e.g., AI) deployment.

Analysis

What You Need to Know

As AI technology emerges, financial planning and analysis (FP&A) leaders must provide decision makers with more enriched and real-time analytics. They must navigate and collect data from multiple, disconnected and siloed sources, while reacting to dynamic market conditions. For many, data quality is too insufficient to provide decision makers with actionable insights, and enterprise stakeholders often lack the data skills or data literacy to resolve critical business problems.

Many vendors offer promising data governance technologies, but assessing the technologies' potential value is difficult due to the rapidly changing governance landscape. Some technologies are less mature but offer significant innovation or differentiation. Others have recently matured and are ready for general mainstream use.

FP&A leaders should use this Hype Cycle to:

- Stay current with the changing governance landscape.
- Disentangle innovations' hyped benefits from their real value.
- Determine which innovations may require supporting investments to fully recognize benefits.
- Decide when it is appropriate to evaluate technologies for adoption.

The Hype Cycle

The D&A governance landscape is growing, and there is no shortage of data governance technologies available. FP&A leaders must address four key challenges:

- **Data integrity —** Increased automation creates new mandates for methods and solutions supporting strong data governance and data reliability. Technical solutions used to prepare, store and secure data quality can help FP&A leaders to ensure uniformity, accuracy and integrity of enterprise's data, facilitating data democratization.

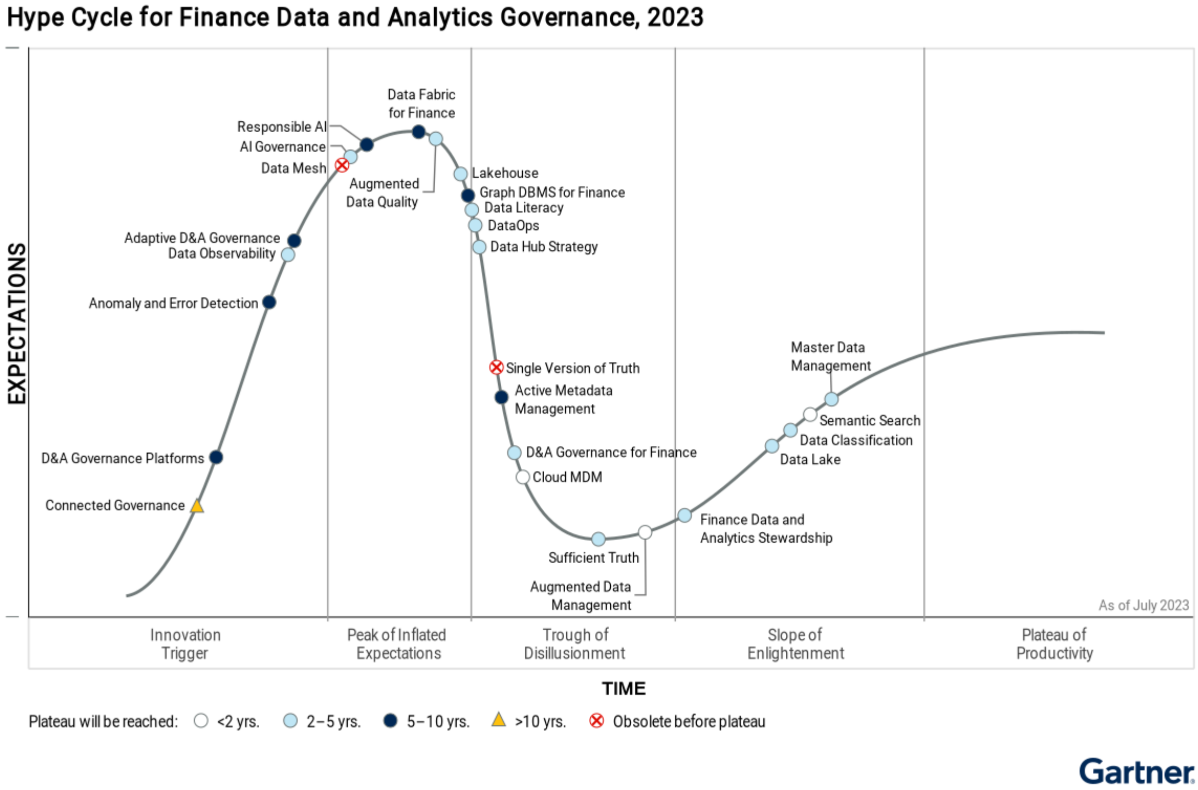
- **Adoption of advanced analytics/AI —** The emergence of AI and machine learning (ML) solutions that automate data management and assess data's interconnectedness generates the need for new governance frameworks. Solutions like augmented data management, AI governance and D&A governance platforms enable FP&A leaders to create and enforce governance rules consistently for building and deploying AI safely.

- **Decision support —** FP&A leaders support a growing number of decision makers who demand broader analytical contexts, supported by more granular financial and nonfinancial data. However, incorporating ungoverned nonfinancial data in decision support can introduce contextual ambiguity. The same ambiguity arises when AI and ML replace or support human decisions. Innovations like semantic search or responsible AI help FP&A leaders add context to reduce bias, even if a certain level of bias is present in the data.

- **Data-driven culture —** Every organization's investment in technology should aim to develop a data-driven culture within the enterprise. FP&A leaders can leverage data literacy assessments and stewardship programs to address the technological gaps and their teams' skills gaps to secure quality D&A, minimize investment risk and avoid eroding partners' trust.

This Hype Cycle enables FP&A leaders to explore the real risks and opportunities of finance D&A governance technologies. This information helps leaders avoid adopting an innovation too early or too late, avoid giving up on an innovation too soon or avoid hanging on to it for too long.

Figure 1: Hype Cycle for Finance Data and Analytics Governance, 2023



The Priority Matrix

FP&A leaders should evaluate their enterprises' specific use cases against their existing business practices and leverage this guide to analyze potential benefits when prioritizing their investment decisions.

When using this Hype Cycle, FP&A leaders should:

- Focus on the transformational innovations that are likely to reach the plateau within two to five years. These technologies may have a significant impact on an organization's processes. Sufficient truth is an evolving innovation that tailors data quality standards to specific business needs. It helps finance leaders preserve data richness, enabling data storage in multiple repositories with accepted, but documented, variability.

- Pay attention to innovations that provide high benefits. These technologies are less likely to change an organization's business model but will affect data governance implementation. They will progress toward the plateau over the next two to five years. Innovations like master data management (MDM) are approaching the plateau, while AI governance and DataOps are perhaps overhyped and will experience periods of inflated expectations and disillusionment before achieving consistent productivity.

- Explore complex innovations that promise transformational benefits but on a longer time horizon. Data fabric and responsible AI offer solutions to increase the connectedness and automation of data governance and deliver more enriched, real-time and accurate data.

Table 1: Priority Matrix for Finance Data and Analytics Governance, 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		Augmented Data Quality Data Literacy Sufficient Truth	Active Metadata Management Adaptive D&A Governance Data Fabric for Finance Responsible AI	
High	Augmented Data Management Cloud MDM Semantic Search	AI Governance D&A Governance for Finance Data Classification Data Hub Strategy Data Observability DataOps Finance Data and Analytics Stewardship Lakehouse Master Data Management	Anomaly and Error Detection D&A Governance Platforms Graph DBMS for Finance	Connected Governance
Moderate		Data Lake		
Low				

Source: Gartner (July 2023)

Off the Hype Cycle

- Data integration tool was removed; it has mainstream adoption in finance.
- Data preparation tool was removed; it has mainstream adoption in finance.
- Logical data warehouse was removed; it has mainstream adoption in finance.

On the Rise

Connected Governance

Analysis By: Saul Judah, Malcolm Murray, Andrew White

Benefit Rating: High

Market Penetration: Less than 1% of target audience

Maturity: Embryonic

Definition:

Connected governance is a framework for establishing a virtual governance layer across organizations and business functions, or legal entities, in multiple geographies to achieve cross-enterprise business outcomes. By connecting existing governance bodies within and across enterprises, its component-based approach enables complex business challenges to be addressed without adding further layers of bureaucracy.

Why This Is Important

Governance bodies for enterprise functions such as HR, risk, and data and analytics are typically adequate for addressing their individual domain areas. However, cross-enterprise and interenterprise governance challenges are increasingly difficult to overcome. Rather than creating yet another permanent governance body, connected governance leverages existing governance bodies through a virtual framework, providing strategic oversight and accountability management across them with minimum additional overhead.

Business Impact

Senior business executives and board members spanning multiple organizations, legal entities and geographies will find value in exploring connected governance to address cross-enterprise strategic issues and opportunities. Organizations anticipating mergers and acquisitions (M&As) will find value in connected governance, enabling both value and risk management to be addressed earlier and allowing experimentation with governance bodies prior to their formal adoption.

Drivers

- The fast pace of deglobalization and digitalization is putting pressure on senior leaders across multiple business functions to respond to business and regulatory demands at greater effectiveness and speed than they are able to with their existing capabilities. Existing governance bodies are designed to address their functional areas, but understanding accountability and decision rights across these proves very difficult. This is especially relevant when some of the functional areas exist in different legal entities and different countries, and the same business asset is subject to potentially conflicting governance policies.
- A key driver for adoption of connected governance stems from the limitations of existing approaches. Traditional approaches to cross-enterprise governance challenges have been to establish another layer of governance, which adds a greater overhead cost, creates another layer of bureaucracy and is often inflexible. Furthermore, some strategic challenges (such as M&A and business model changes) require a one-off response for governance, and creation of additional governance layers in these circumstances is an excessive drain on executives' time without accrued benefit. Consequently, adoption of connected governance becomes an attractive option.

Obstacles

- Connected governance leverages existing governance bodies, but some of these bodies may operate poorly. As a result, the value that connected governance offers may be depleted in organizations that are not already mature in their governance.
- Siloed governance efforts might reinforce those silos and prevent the benefits of connected governance without disruptive organizational change. Either way, inertia and local success of siloed governance will slow down the adoption of connected governance.
- Once the board of directors or executive committee has approved the cross-governance initiative, an executive leader is expected to shape the cross-governance response. However, this needs support and facilitation from a strategic governance office, which requires skills that are currently in short supply.

User Recommendations

- Evaluate whether connected governance would benefit your organization. If you operate in a complex environment, across multiple legal entities and geographies, there may be challenges that are difficult to address now. In such situations, put on the agenda of your executive committee meeting to initiate a cost-benefit assessment and report its findings. If this does not apply at your organization, connected governance may not be for you.
- Connected governance needs the support of strategic, cross-enterprise governance. Analyze whether this needs a dedicated governance office or if operating as a virtual governance office will be sufficient. If your strategic challenge is a one-off situation, or if you are trialing this as a new initiative, a virtual office may be sufficient for now. However, large enterprises in diverse, complex ecosystems and expecting to address many strategic scenarios may necessitate a dedicated strategic governance office to support connected governance.

Gartner Recommended Reading

Connected Governance Orchestrates Complex Cross-Enterprise Decisions

Connected Governance Drives Adoption of Data and Analytics Governance Platforms

Quick Answer: What Kind of Governance Does Healthcare Data Interoperability Need?

Choose the Optimal Corporate Structure to Cope With Geopolitical Risks

Trends 2023: Rise and Risks From EU, U.S., China and Other Sovereign Data Strategies and Policies

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

Anomaly and Error Detection

Analysis By: Mark D. McDonald

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

Anomaly and error detection in finance leverages AI and ML to identify errors, mistakes or unusual activity, as well as violations of internal policies, compliance rules and accounting standards. Such tools may be on-premises or cloud-based, and may be stand-alone solutions or integrated with accounting and reporting systems (e.g., ERP). Solutions report anomalies and errors in real time or via periodic batch processing, allowing users to take investigative or corrective actions on findings.

Why This Is Important

Finding and correcting errors grows increasingly difficult for finance departments as data volume and complexity grow. New rules, regulations and accounting policies make compliance challenging and increase the chances of embarrassing violations and restatements. By leveraging AI and automating anomaly and error identification, finance teams spend time supporting business objectives rather than fixing problems and responding to audit findings.

Business Impact

Anomaly and error detection offers:

- **Early detection of problems** before they become time-consuming, frustrating, costly and embarrassing to correct
- **Error-free and compliant financial results** that build stakeholder trust and a solid company reputation
- **"Finding-free" audit reports** that require no punch list of time-consuming follow-up actions
- **Increased leadership confidence in the finance function** and an assurance that published financial statements are consistently and reliably accurate

Drivers

- **Increased complexity:** Consistent increases in data volume increase the complexity of managing that data. Increased volume and complexity lead to more errors that are difficult to uncover.
- **Cost and productivity pressure:** Increased pressure to reduce costs forces financial leaders to look for cost-cutting measures across all financial processes. By using AI-driven tools to detect errors early, fewer resources are needed to find and correct mistakes, allowing finance staff to spend more time on business support. Additionally, consistent adherence to statutory accounting guidelines and policies reduces the hourly cost of audit support.
- **Reputation:** Stakeholders demand financial integrity and financial statements they can rely on. By ensuring that results are integral, companies avoid reputation-damaging restatements.
- **Advancements in AI and machine learning (ML):** Increased accessibility of AI and ML is motivating vendors to leverage these techniques and develop effective and easy-to-use platforms that require minimal user training and implementation effort.
- **AI and ML curiosity:** CFOs are actively exploring the benefits of AI and ML in the finance workflow. Using anomaly and error detection software allows leaders to observe the benefits of AI without a large project or disrupting workflows.
- **Increasing effectiveness of external auditors:** Some larger audit firms are using AI to lower costs and find a larger number of detailed anomalies and errors. This forces audit clients to implement commensurate tools to keep up.

Obstacles

- **AI skepticism:** Negative AI media coverage prevents widespread acceptance of AI-driven tools and processes.
- **Establishment resistance:** Tools that impact the revenue stream of large audit firms will encounter certain resistance. Regulatory agencies will also challenge changes to established control processes.
- **Lack of AI experience:** Without AI experience, leaders struggle to relate AI's advantages with business benefits and to integrate these tools into legacy workflows.
- **Immature market**: A lack of clear market segments and benchmarked capabilities leaves this new era of AI-driven solutions off company radars.
- **Legacy system investment:** Years of sunk capital in tailored, rules-driven systems drive a falsely optimistic sense of security that these systems will adequately handle growing data volumes and complexity.
- **Lack of ROI commitment:** An inability to promise ROI for AI-driven software solutions drives hesitation when making investments.

User Recommendations

- Isolate the areas of finance processes with the largest number of and most costly errors to identify where anomaly and error detection has the largest potential impact.
- Use the cost of errors as a baseline for potential savings. Choose recent examples and assess their cost. Consider costs like late delivery penalties and lost revenue from internal mistakes. Other direct costs may include the time charged by external auditors to follow up on audit findings.
- Search for external vendors that can help. Consult a Gartner analyst to help narrow your search.
- Engage a shortlist of possible vendors in a live demonstration of their tools to evaluate them on the efforts needed to provide the tool with sufficient data and whether the tool addresses your process needs. Include an assessment of whether the software's output provides actionable information.
- If no vendors can help with your specific use case, engage your internal advanced analytics department to build custom solutions.

Sample Vendors

Ai XPRT; AppZen; AuditMap; MindBridge; Oversight

Adaptive D&A Governance

Analysis By: Saul Judah

Benefit Rating: Transformational

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

Adaptive data and analytics (D&A) governance is an organizational capability that enables context-appropriate governance styles and mechanisms to be applied to different D&A scenarios to achieve desired business outcomes.

Why This Is Important

As organizations accelerate or scale out their digital business initiatives, ecosystems and platforms, their ability to deliver expected business value is limited by their current business practices — in particular, their governance of D&A assets. Despite greater diversity and complexity in business scenarios than ever before, D&A governance has typically continued to adopt a single, control-oriented approach, which is often unresponsive to business needs and causes or reinforces data silos.

Business Impact

Adaptive D&A governance has the potential to be a transformational change agent for digital business. It enables application of different governance styles (control, outcome, agility and autonomous) to different D&A scenarios, depending on business context. This allows better enterprise collaboration in D&A initiatives, allowing enterprises to respond faster to opportunities and become more competitive, resilient and risk-aware.

Drivers

- As levels of risk appetite and demands for growth have risen in organizations, so have expectations for flexibility and agility from D&A initiatives to meet these needs. As a result, chief data and analytics officers are increasingly turning to adaptive D&A governance practices that enable the greater flexibility, scale and resilience needed in D&A initiatives to deliver dynamic business outcomes.

- Both D&A and business leaders recognize that increased investment in infrastructure, such as D&A platforms, cannot yield the expected ROI without corresponding improvement in D&A governance practices.

- Organizations maturing in D&A increasingly recognize the key role that business leaders play in driving their governance initiatives. Business demand for greater flexibility, agility, responsiveness and interconnectedness of D&A requires better governance practices than currently exist. This, in turn, is leading D&A leaders to explore adaptive D&A governance.

Obstacles

- Although D&A governance practices are maturing in many organizations, maturity is still lower than in other areas, such as data management and analytics. Many organizations still take an IT-oriented, center-out, single-style approach to governance, which resembles compliance rather than governance. This is outdated and wrong and needs to change.

- Poor data literacy is prevalent in organizations. Business leaders often fail to understand or accept accountability for the information assets they create, instead expecting their data office (typically residing in IT) to sort out their data. When data offices initiate governance initiatives, business leaders fail to engage effectively, or at all.

User Recommendations

- Use the IT Score for Data & Analytics to evaluate your maturity and readiness to enhance governance capabilities. Don't establish agility and autonomous governance without foundations for control- and/or outcome-based governance.
- Create a proof-of-concept (POC) initiative to test the applicability of one of the advanced governance styles (like an autonomous governance style) in your environment; evaluate the business outcomes and value, emerging risks, technological limitations and cultural barriers to wider adoption.
- Engage senior business executive leadership to discuss the results of the POC initiative. Create a business case and strategic roadmap to establish adaptive D&A governance.
- Establish the control and outcome styles of adaptive governance first; evolve to the agile and autonomous styles subsequently. Use minimum governance, focusing on limiting the scope of data, analytics and business processes to those that deliver greatest business value and organizational outcomes.

Gartner Recommended Reading

Data and Analytics Leaders Must Use Adaptive Governance to Succeed in Digital Business

2022 Strategic Roadmap for Data and Analytics Governance

Adopt SMART Principles for Adaptive Data and Analytics Governance

Next Best Actions to Improve Your Data and Analytics Governance

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

AI Governance

Analysis By: Svetlana Sicular

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

AI governance is the process of creating policies, assigning decision rights, and ensuring organizational accountability for risks and investment decisions for the application and use of artificial intelligence techniques. AI governance is part of adaptive data and analytics governance, addressing the predictive and generative nature of AI.

Why This Is Important

With AI now delivering value in the enterprise, data and analytics leaders observe that scaling AI without governance is ineffective and dangerous. Generative AI and applications, like OpenAI's ChatGPT, make AI governance a necessity, as using pretrained AI models billions of times sharpens risk concerns. The leaders want to balance AI's business value and the need for appropriate oversight. AI draws the attention of legislators worldwide, who mandate actions by clarifying AI governance priorities.

Business Impact

AI governance, as part of the organizational governance structure, enacts responsible AI, and provides common implementation and adherence mechanisms across the business ecosystem when it comes to:

- Ethics, fairness, and safety to protect the business and its reputation,
- Trust and transparency to support AI adoption via explainability, bias mitigation, model governance, operationalization, and collaboration norms and capabilities.
- Diversity to ensure the right technology and roles for each AI project.

Drivers

- AI governance is in the peak area of the Hype Cycle. Enterprise practitioners are taking steps toward establishing AI governance. Leading organizations in various industries establish AI governance by addressing standards for AI development and operations, providing best practices, guidelines for model management and monitoring, data labeling and interpretation, explainability, fairness, bias mitigation, security, and legal.

- Regulations around the globe target AI directly and affect AI practices indirectly, making AI governance goals more concrete. The U.S. Blueprint for an AI Bill of Rights provides governance pathways, from principles to practice. The objective of the EU AI Act is to "enhance governance and effective enforcement of existing law on fundamental rights and safety requirements applicable to AI systems." The Algorithmic Impact Assessment is a mandatory risk assessment tool intended to support the Treasury Board of Canada. Singapore's Model AI Governance Framework guides organizations in developing appropriate governance structures and mechanisms.

- Trust and transparency of AI solutions are crucial for AI adoption. The probabilistic and opaque nature of AI is new to audiences familiar with deterministic outcomes. AI governance can minimize misinterpretations of AI results by scrutinizing trust in data sources and the explainability of AI decisions. It provides specific testing and validation guidelines, differentiating "life-critical AI."

- AI governance is necessary to establish AI accountability. It is difficult to achieve because use cases differ in terms of their data, solution and outcome requirements. It outlines reactive responsibilities, actions and procedures in the case of unanticipated and unintended consequences. It ensures that ethics are considered for each use case.

Obstacles

- Often, AI governance is stand-alone from mainstream governance initiatives, which stalls its progress. The best method is to extend existing governance mechanisms to take advantage of recognizable policies and methods, such as in data governance. AI governance benefits from a conversation with the security, legal and customer experience functions.

- Many governance initiatives assume command and control. Instead, adaptive governance supports freedom and creativity in AI teams but also protects the organization from reputational and regulatory risks. Little or no governance in AI teams to facilitate freedom and creativity is an acceptable approach if this is a conscious governance decision.
- AI value assurance and model risk management are new in AI. While methods exist — for example, in the financial industry — they are largely unknown to others, and every governance organization is inventing its own.
- Technologies to support AI governance are fragmented and are often designed for a single industry.

User Recommendations

- Extend to AI your existing governance mechanisms, such as risk management or data and analytics governance.
- Establish and refine processes for handling AI-related business decisions. Blend processes, people and technology to succeed.
- Aim to align your AI governance framework with the laws and regulations in your jurisdiction(s) to directionally assure your efforts amid evolving AI-specific considerations. Gain agreement on AI risk guidelines that are driven by the business risk appetite and regulations.
- Decide on the organizational structure and accountability for propagating responsible AI — for example, what to centralize and what to do locally.
- Implement tools for AI review and validation. For each AI use case, require an independent AI model validator, a data scientist whose job is to assure model explainability and robustness. Have all parties in the process defend their decisions in front of their peers and validators.
- Ensure that humans are in the loop to mitigate AI deficiencies.

Sample Vendors

Arthur; Chatterbox Labs; Credo AI; DarwinAI; FICO; Google; IBM; Prodago; SAS; Weights & Biases

Gartner Recommended Reading

Applying AI — Governance and Risk Management

[4 AI Governance Actions to Make a Swift Business Impact]

[Artificial Intelligence Primer for 2023]

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

Responsible AI

Analysis By: Svetlana Sicular

Benefit Rating: Transformational

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

Responsible artificial intelligence (AI) is an umbrella term for aspects of making appropriate business and ethical choices when adopting AI. These include business and societal value, risk, trust, transparency, fairness, bias mitigation, explainability, sustainability, accountability, safety, privacy, and regulatory compliance. Responsible AI encompasses organizational responsibilities and practices that ensure positive, accountable, and ethical AI development and operation.

Why This Is Important

Responsible AI has emerged as the key AI topic for Gartner clients. When AI replaces human decisions and generates brand-new artifacts, it amplifies both good and bad outcomes. Responsible AI enables the right outcomes by ensuring business value while mitigating risks. This requires a set of tools and approaches, including industry-specific methods, adopted by vendors and enterprises. More jurisdictions introduce new regulations that challenge organizations to respond in meaningful ways.

Business Impact

Responsible AI assumes accountability for AI development and use at the individual, organizational and societal levels. If AI governance is practiced by designated groups, responsible AI applies to everyone involved in the AI process. Responsible AI helps achieve fairness, even though biases are baked into the data; gain trust, although transparency and explainability methods are evolving; and ensure regulatory compliance, despite the AI's probabilistic nature.

Drivers

- Responsible AI means a deliberate approach in many directions at once. Data science's responsibility to deliver unbiased, trusted and ethical AI is just the tip of the iceberg. Responsible AI helps AI participants develop, implement, utilize and address the various drivers they face.

- Organizational driver assumes that AI's business value versus risk in regulatory, business and ethical constraints should be balanced, including employee reskilling and intellectual property protection.

- Societal driver includes resolving AI safety for societal well-being versus limiting human freedoms. Existing and pending legal guidelines and regulations, such as the EU's Artificial Intelligence Act, make responsible AI a necessity.

- Customer/citizen driver is based on fairness and ethics and requires resolving privacy versus convenience. Customers should exhibit readiness to give their data in exchange for benefits. Consumer and citizen protection regulations provide the necessary steps, but do not relieve organizations of deliberation specific to their constituents.

- With further AI adoption, the responsible AI framework is becoming more important and is better understood by vendors, buyers, society and legislators.

- AI affects all ways of life and touches all societal strata; hence, the responsible AI challenges are multifaceted and cannot be easily generalized. New problems constantly arise with rapidly evolving technologies and their uses, such as using OpenAI's ChatGPT or detecting deepfakes. Most organizations combine some of the drivers under the umbrella of responsible AI, namely, accountability, diversity, ethics, explainability, fairness, human centricity, operational responsibility, privacy, regulatory compliance, risk management, safety, transparency and trustworthiness.

Obstacles

- Poorly defined accountability for responsible AI makes it look good on paper but is ineffective in reality.
- Unawareness of AI's unintended consequences persists. Forty percent of organizations had an AI privacy breach or security incident. Many organizations turn to responsible AI only after they experience AI's negative effects, whereas prevention is easier and less stressful.
- Legislative challenges lead to efforts for regulatory compliance, while most AI regulations are still in draft. AI products' adoption of regulations for privacy and intellectual property makes it challenging for organizations to ensure compliance and avoid all possible liability risks.
- Rapidly evolving AI technologies, including tools for explainability, bias detection, privacy protection and some regulatory compliance, lull organizations into a false sense of responsibility, while mere technology is not enough. A disciplined AI ethics and governance approach is necessary, in addition to technology.

User Recommendations

- Publicize consistent approaches across all focus areas. The most typical areas of responsible AI in the enterprise are fairness, bias mitigation, ethics, risk management, privacy, sustainability and regulatory compliance.
- Designate a champion accountable for the responsible development and use of AI for each use case.
- Define model design and exploitation principles. Address responsible AI in all phases of model development and implementation cycles. Go for hard trade-off questions. Provide responsible AI training to personnel.
- Establish operationalize responsible AI principles. Ensure diversity of participants and the ease to voice AI concerns.
- Participate in industry or societal AI groups. Learn best practices and contribute your own, because everybody will benefit from this. Ensure policies account for the needs of any internal or external stakeholders.

Sample Vendors

Amazon; Arthur; Fiddler; Google; H2O.ai; IBM; Microsoft; Responsible AI Institute; TAZI.AI; TruEra

Gartner Recommended Reading

A Comprehensive Guide to Responsible AI

Expert Insight Video: What Is Responsible AI and Why Should You Care About It?

Best Practices for the Responsible Use of Natural Language Technologies

Activate Responsible AI Principles Using Human-Centered Design Techniques

How to Ensure Your Vendors Are Accountable for Governance of Responsible AI

Data Fabric for Finance

Analysis By: Ash Mehta

Benefit Rating: Transformational

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

Data fabric for finance is a design framework that improves data access and speed by offering more flexible, reusable and automated data integration solutions. It leverages data integration, active metadata, knowledge graphs, semantics, machine learning (ML) and data cataloging, and supports various analytics and operational use cases across multiple platforms. It leverages traditional approaches and enables the enterprise to adopt technology advances — and avoid a "rip and replace" experience.

Why This Is Important

Data fabric for finance enables enterprises to leverage and access all data and analytics (D&A), regardless of source, format or structure. It capitalizes on past D&A investments while concurrently providing prioritization, cost control and new D&A management spending guidance. By enhancing human and technology interactions, the fabric ensures greater flexibility, composability, scalability and orchestration of consumable data that spans use cases, sources, models, or other hybrid D&A forms.

Business Impact

The data fabric for finance has many benefits, including:

- Monitoring data assets on allocated resources for context, optimization and cost control.
- Allowing access to connected D&A types: sources or locations.
- Automation of repeatable D&A tasks like integration, quality or delivery.
- Access via natural language rather than technical labels, enriching usability and meaning.
- Enabling machine learning to create self-learning models capable of detecting new or similar content, while increasing accuracy.

Drivers

- Increased growth in data volume stored in numerous systems without a consistent, contextual labeling approach is preventing companies from leveraging the investments in collecting and storing the data.
- A shortage of data management professionals is increasing the demand for automation. Data fabric for finance's active metadata aspects are powerful solutions to this shortage.
- Growth in ML and intensive D&A tasks increase the need for more data. This includes demand for data tracking, auditing, monitoring, reporting and evaluating use and utilization, and data analysis for content, values and veracity of data assets.
- New sources and types of data make cataloging and labeling the data nearly impossible without the help of an assistive technology. Demand for rapid contextual comprehension of new data assets increased sharply and continues to accelerate.
- Catalogs alone are insufficient in assisting with data self-service. Data fabrics for finance capitalize on ML recommendation engines for integration design and delivery, reducing the need for manual human labor.
- The data fabric for finance actively alerts developers to D&A performance issues, identifying when user experiences vary from expectations depicted in system designs.
- New technology advancements enable data fabrics for finance to assist with graph data modeling capabilities (preserving data's context and its complex relationships) and also allow business users greater access to enrich analytic models with agreed-upon semantics.
- Significant growth in demand and use of knowledge graphs of linked data and ML algorithms can be supported in data fabric for finance.
- Companies have found that one or two approaches to data acquisition and integration are insufficient. Data fabrics for finance provide capabilities to deliver integrated data through a broad range of combined data delivery styles including bulk/batch (extraction, transformation and loading), data virtualization, message queues, use of APIs and microservices.

Obstacles

- Proprietary metadata restrictions hamper the data fabric for finance as it is wholly dependent upon acquiring metadata from a variety of data management platforms. Fabric requires analytical and ML capabilities to infer missing metadata — this will be error-prone at first.
- Diversity of skills and platforms to build a data fabric for finance presents technical and cultural barriers. A shift is required from data management based on analysis, requirements and design, to discovery, response and recommendation.
- Market confusion about "data mesh" has created distractions, and will slow data fabric for finance adoption. Market providers and services organizations purporting complete data fabric for finance delivery are adding to market cynicism.
- Misunderstanding and lack of knowledge in how to reconcile a data fabric for finance with legacy D&A governance programs will complicate implementations.

User Recommendations

- Begin and end your efforts with understanding "active metadata."
- Invest in an augmented data catalog that assists with creating a flexible data model. Enrich the model through semantics and ontologies for the business to use the catalog.
- Deploy data fabrics for finance that populate and utilize knowledge graphs.
- 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 platform as a service solutions that orchestrate across solutions.

Sample Vendors

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

Gartner Recommended Reading

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

From Logical Data Warehouse to Data Fabric

Quick Answer: What Is Data Fabric Design?

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

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

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

Graph DBMS for Finance

Analysis By: Valeria Di Maso

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

Graph is a purpose-built database management system (DBMS) targeted at broad analytics, operational, data science and AI use cases. Graph DBMS provides a facility to query data and their relationships with other data. They are suitable for multiple types of interactions including simple node, edge traversal and triple pattern matching for transactional uses, complex multihop queries, reasoning and inference, and algorithms for analytical workloads.

Why This Is Important

Few in FP&A know of graph DBMS benefits, and even where in use, it is underutilized. Relational DBMS (RDBMS) struggle to traverse complex networks with point-to-point (SQL) as they use tables to store and connect data. With a graph, as long as data and a path exists, users can explore and query real-time data with complex relationships across domains. For example the query — show clients who bought our shoes because they liked our socks — is complicated in SQL but is easy in graphs.

Business Impact

The growing operational complexities and the demand for predictive insights are driving awareness, adoption and value of graph DBMS. Graph has proven value in fraud detection, customer 360, optimizing supply chains and other areas where connecting multiple internal, fringe and external data is required. It also accelerates the path to predictive analytics, scenario optimization (e.g., budget or capex), cluster identification, and other correlations of asset-intensive use cases.

Drivers

- Generating actionable and predictive insights amid broader and growing data varieties, is becoming increasingly burdensome with relational DBMS and query. Graph DBMS offers powerful data modeling and analysis capabilities that allow teams to model complex business scenarios in a more flexible, representational and intuitive way.
- FP&A teams are increasing experimentation, adoption and training in graph, machine learning (ML) and artificial intelligence (AI) to enhance semantic search and position graph DBMS as a tool to develop more complex data fabrics.
- Mature, multimodal, widely distributed DBMS products are embedding graph DBMS approaches. Both open-source and commercial graph DBMS offerings are maturing quickly and gaining users. The resulting visibility of graph DBMS use cases in previously inaccessible data domains drives adoption.
- Funding has been increasingly available to new vendors, including cloud providers who offer both their own and third-party solutions, sending another signal to potential adopters to test graph's value. In addition, some graph vendors have begun to offer graph connectors to BI tools (like Looker, Microsoft Power BI and Tableau) making graph data available alongside RDBMS data.
- Demand is quickly increasing on FP&A teams to reduce data development and management lead time (from preparation to insight discovery to productization).
- Graph DBMS coupled with AI/ML improves data quality, by reconciling differences in customer records and products' attributes.
- FP&A's evolution toward cloud-based solutions and extended planning and analysis (xP&A) has accelerated the need to associate the connectedness and meaning of data's relationships.

Obstacles

- Graphs require uncommon modeling, loading, processing and analytical skills. It suffers from lack of standardization and programming ease that slows down enterprise adoption. More user-friendly tools and interfaces are required due to the difficulty of modeling and evaluating graphs versus tables.
- Nonstandardized query languages (like Gremlin), APIs and varying quality and support of vendor ML algorithm libraries, make choosing a product and staffing increasingly complex.
- Existing graph analytical tools or graph analytics functions built in existing applications can function without a dedicated DBMS. If the requirement or scale is not too great, then alternatives may suffice, making graph DBMS unnecessary.
- Many technical graph practitioners are unable to effectively explain the business benefits to sponsors. Moreover, use cases are now moving to embedded functionality into other tools (AI, metadata, data integration), making the decision to adopt new-to-FP&A solutions less attractive.

User Recommendations

- Evaluate graph DBMS wherever RDBMS and query are unable to connect and/or capture data relationships or where data elements are extensive. Start with a small pilot project to demonstrate value to the FP&A community.
- Explore whether current DBMS includes embedded, unused graph functionality.
- To ensure scaling and reliability, use commercially-supported, open-source graph DBMS projects, or community editions of commercial solutions, to experiment and gain experience.
- Use graph DBMS 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 FP&A teams.
- Select graph DBMS based on your analytics requirements. Resource description framework (RDF)- and property graph-based architectures have capabilities that should be factored in when making your selection.

Sample Vendors

Amazon Web Services (AWS); Cambridge Semantics; Dgraph; MarkLogic; Microsoft; Neo4j; Ontotext; OpenLink Software; TigerGraph

Gartner Recommended Reading

CFOs' Autonomous Future Podcast: Graph, the Future of Databases with Grant Nelson

Understanding When Graph Analytics Are Best for Your Business Use Case

Quick Answer: When to Use Graph Analytics in Finance

Market Guide for Graph Database Management Systems

Working With Graph Data Stores

Sliding into the Trough

Data Literacy

Analysis By: Alan D. Duncan, Donna Medeiros, Sally Parker

Benefit Rating: Transformational

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

Data literacy is the ability to read, write and communicate data in context, with an understanding of the data sources and constructs, analytical methods and techniques applied. Data-literate individuals have the ability to identify, understand, interpret and act upon data within business context and influence the resulting business value or outcomes.

Why This Is Important

Data and analytics (D&A) are pervasive in all aspects of businesses, communities and our personal lives. Thus, data literacy is foundational to the digital economy and society. It helps stakeholders:

- Draw a directlink between D&A and desired outcomes
- Unlock knowledge workers' business acumen
- Explain how to identify, access, integrate and manage datasets
- Draw insights relevant to specific use cases
- Describe advanced analytics techniques and enable AI
- Reduce risk through improved decision making

Business Impact

To become data-driven and equipped to use data and analytics to their competitive advantage, enterprises require explicit and lasting organizational change. Chief data and analytics officers (CDAOs) need to promote and orchestrate “leadership moments” where they act as role models, exemplifying new cultural traits at critical points. To be successful, they will need to guide the workforce by addressing both data literacy and data-driven culture.

Drivers

- The continued growth in digital transformation is amplifying a focus on D&A best practices. Employee data literacy is becoming increasingly recognized as an important factor in an organization’s overall digital dexterity.
- The role of the D&A function has evolved. It is now at the core of an organization’s business model and digital platforms, and with everyone being an information worker, the footprint of business use of data and analytics is broader than ever before.
- Effective D&A strategies require an increased focus on change management. Higher-performing CDAOs prioritize their emphasis, energy and effort on change management requirements, including data literacy.
- Defining what data-driven behaviors are expected — using a “from/to/because” approach — is central to employee development plans. It ensures that creators, consumers and intermediaries have the necessary D&A skills, knowledge and competencies.
- Data literacy is not a one-off project. CDAOs need to take immediate action to create and sustain data literacy through assessment of maturity, awareness, and education. Quick wins build momentum, but lasting and meaningful change takes time because it requires people to learn new skills and behave in new ways. (For example, there is a hunger for this type of skills development within Gen Z, especially in order to future-proof their careers.)

Obstacles

- Lack of common data literacy models/frameworks/standards and terminology.
- Varying interpretations of the term “data literacy” in terms of training, curriculum and understanding, ranging from enhanced data visualization skills to fostering business curiosity about data.
- Failure to measure contribution of data to business outcomes.

- A sporadic and inconsistent approach to training and certification.
- Not recognizing that data use is a behavioral change or change management initiative.
- Lack of talent and poor data literacy within the current workforce.
- Lack of initiatives to address cultural and data literacy challenges within strategies and programs.
- Overall adoption will still take years, due to the complexity of upskilling entire workforces.
- Data literacy is treated as a checkbox activity, especially when delegated to more junior (and unempowered) resources.
- Lack of a designated leader accountable for the development and execution of the program, roadmap and communication plan.

User Recommendations

- Make the business case for data literacy by identifying stakeholder outcomes and linking these to underlying learning needs.
- Designate a leader who will be accountable for developing and executing the roadmap.
- Foster data literacy during D&A requirements gathering by bringing data and business experts together around the problem to be solved.
- Call out examples of "good" and "bad" data literacy to promote desired behaviors.
- Nurture data literacy by rewarding stakeholders who recognize this as a factor for success and sharing their stories.
- Partner with HR and business leaders to incorporate data literacy learning outcomes into job descriptions, career paths and employee value proposition.
- Use data literacy assessments to evaluate current skill levels and desire to participate.
- Go beyond vendor product training to focus on people's role- and industry-related D&A skills. Improve learning effectiveness by using a mix of training delivery methods (classroom, online, community, on the job).

Sample Vendors

Avado; The Center of Applied Data Science (CADS); Coursera; The Data Lodge; Data To The People; Pluralsight; Skillsoft; Udacity; Udemy

Gartner Recommended Reading

How CDAOs Must Lead Data Literacy and Data-Driven Culture

Address Both ‘Skill’ and ‘Will’ to Deliver Data-Driven Business Change

Drive Business Outcomes by Measuring the Value of Data Literacy

Tackle Data Literacy Head-On to Avoid Data and Analytics Program Failure

Partner With Data Literacy Providers to Accelerate the Time to Value for Data-Driven Enterprises

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 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

Data Hubs, Data Lakes and Data Warehouses: How They Are Different and Why They Are Better Together

Single Version of Truth

Analysis By: Nick Duffy

Benefit Rating: Low

Market Penetration: 20% to 50% of target audience

Maturity: Obsolete

Definition:

Single version of truth (SVOT) in data and analytics (D&A) is the approach of creating a single, clear enterprisewide data standard in a consistent and nonredundant form. This is often achieved by storing data in a centralized and harmonized database fed by multiple sources. SVOT enables one view of data that everyone within the organization trusts and that best describes the company's position and objectives.

Why This Is Important

SVOT promises a single, enterprisewide data quality standard using corporate-led governance principles. It is useful for highly accurate and nonfragmented data. It is favored by organizations aiming to create standard data-to-outcome relationships. However, it is difficult to scale or sustain SVOT in complex and changing data ecosystems. Its focus on risk aversion and accuracy may also inhibit innovation. We predict it will soon become obsolete in the context of finance functions.

Business Impact

SVOT delivers accurate and clear data to decision makers to help them answer strategic questions. Its aim of a single, enterprisewide data quality standard inspires consistency and trust in data, analytics and decision making. By drawing standardized causal links between the D&A processes used across functions, it promises to enable a corporate-sanctioned and accepted D&A vocabulary to support complex and cross-functional enterprise decisions.

Drivers

- Having multiple versions of the truth can lead to different interpretations of data and reports, thereby creating confusion, paralysis and bad decision making.
- Unbridled growth in data and organizational complexity often creates ambiguous or contrary insights, leading to suboptimal decisions and unproductive discussions.
- Organizations need to empower self-service D&A users to make (near-) real-time decisions, which is easier with a singular, trusted data standard.
- With SVOT, team members no longer have to pull data from multiple sources to build reports. Analysts, therefore, spend less time collecting data and more time finding new insights and drawing conclusions.
- Leaders want to share and reuse data, improve data consistency and accelerate time to value.
- Having data centralized and accessible in a standard way enables the use of artificial intelligence and machine learning as well as other data science innovations.

Obstacles

- SVOT is a nearly impossible, complex and time-consuming exercise. The “truth” may change over time to meet evolving needs. So even if achieved, the pace of change in organizational data uses makes SVOT extremely difficult to maintain.
- SVOT implies that other truths exist. Therefore, it is fundamentally an accepted or sanctioned version of truth. However, the growth of digital and data complexity makes SVOT inefficient, as data is often reviewed outside its original context and potentially misinterpreted.
- SVOT requires one entity, person, group or department to be the arbiter or the final say, centralizing a level of control or power into that entity, which is often very distant from the decision makers using the data.
- SVOT requires an ideal balance of business acumen, technical proficiency, D&A architecture, domain understanding and data governance. Companies often struggle to acquire or develop the appropriate breadth of skills required.
- The advent of the data lake encourages the proliferation of big data and pushes SVOT beyond its capabilities.

User Recommendations

SVOT is descending through the Trough of Disillusionment, as many organizations have invested significant time and resources in SVOT projects without achieving their aims. To effectively apply or implement SVOT, companies should:

- Cross-functionally align functions to develop SVOT and tailor it to the desired outcome by enabling it with available technology.
- Differentiate between a "single source of truth" (SSOT) approach and SVOT. SSOT is an architectural concept where source system data may be engineered to meet design criteria; it should not be combined with SVOT.
- While developing SVOT, prioritize the most actionable, relevant and reliable data through single-sourced master data to avoid taking on too much at once.
- Create a data governance committee dedicated to meeting regularly (e.g., monthly, quarterly, biannually) to discuss and deliberate SVOT.
- Gradually move to a "sufficient version of the truth" approach, balancing the costs of imperfect data and data governance, including context-appropriate style.

Gartner Recommended Reading

Data Management: A Single Version of the Truth No Longer Works

Case Study: Decision-Focused Data Maps for Finance (General Mills)

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

D&A Governance for Finance

Analysis By: Ash Mehta, Andrew White, Debra Logan

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

Data and analytics (D&A) governance for finance is the framework for the specification of decision rights and accountability to ensure appropriate behavior in the valuation, creation, consumption and control of D&A, in the context of the finance function. It includes the processes, roles, policies, standards, and metrics that ensure the effective and efficient use of D&A for finance.

Why This Is Important

D&A governance is critical for managing finance data assets. Though traditionally the domain of CIOs or CDOs, governance of financial data and analytics assets is shifting to finance department leaders who have better visibility into how decision makers use data in practice. Finance leaders responsible for enterprise analytics require good governance practices to ensure analytical consistency across business lines and departments and to deliver analytical insights that support key outcomes.

Business Impact

Finance leaders should anticipate the following impacts:

- Stronger governance over D&A decision rights across the enterprise and within business areas.
- Tighter regulatory compliance leading to lower risk of infractions and fines.
- Increased levels of business collaboration with data stewards to engage, orchestrate, innovate and drive governance to support enterprise business outcomes.
- Greater understanding of how key business outcomes relate to data quality metrics.
- Governance enabling data literacy.

Drivers

- As organizational D&A capabilities mature, organizations are recognizing that enterprise — as opposed to siloed — governance yields better business results.
- Accelerating investment in advanced (artificial intelligence, machine learning) analytics requires an enterprise governance approach.
- The level of data quality required for FP&A leaders to deploy advanced analytics requires an active role in governance.
- Decision-ready data required to deliver actionable insights in the decision making requires sufficient version of truth-based governance.
- Multiple business units within organizations are building and investing in their own D&A teams leading to conflicting metrics and KPIs.
- Organizations are investing heavily in technologies that are based on loose foundations without proper data governance leading to lower-than-anticipated ROIs.
- Shadow data and metrics continue to proliferate across organizations even as technology investments and digital transformation initiatives accelerate.

Obstacles

- Enterprise data governance has long been thought of as the domain of IT and CDOs.
- Overly restrictive data governance principles are created when policymakers write data governance principles as fixed rules with specific do's and don'ts for employee actions, burdening (rather than enabling) productive practices and behaviors.
- Top-down evaluation process where policymakers seek policy approval from the data governance steering committee and do not address end-user concerns regarding policy language and implementation.
- Control-oriented data governance approaches have proven ineffective at achieving the data quality level desired by business leaders, creating a one-size-fits-all approach that is ill-suited for advanced analytics.
- Centralized data governance results in data governance that causes data owners and stewards to disengage, casting them as passive recipients of data governance mandates rather than active participants in data quality improvement efforts.

User Recommendations

- Define data governance policy objectives and establish a finance-led steering committee to implement enterprise data governance.
- Detect whether a CxO leads any existing governance efforts; if so, then we recommend a finance D&A leader serve as co-chair.
- Appoint a member of the leading governance committee to oversee the policy creation.
- Strengthen governance accountability by establishing roles that encourage a collaborative, enterprisewide data governance model.
- Build a business case and appoint an executive champion to coordinate and arbitrate governance opportunities and business outcomes.
- Increase the consistency of interactions between data governance councils and data stewards by establishing a data governance charter.
- Establish guidelines, not rules, using natural business language (free from technical or other jargon). Draft broad principles that empower employees with the autonomy to make decisions and value judgments where appropriate.

Gartner Recommended Reading

Finance's Role in Enterprise Data Governance: Executive Champion and Lead Steward

Ignition Guide to Creating a Finance Data Governance Policy

Ignition Guide to Building a Data and Analytics Governance Program for Finance

Use Value Streams to Create a Common Language for Finance Data Definition Governance

How to Align Finance's Data and Analytics Governance Strategy With a Metrics Cascade

Cloud MDM

Analysis By: Sally Parker, Helen Grimster

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

Cloud master data management (MDM) solutions ensure the uniformity, accuracy, stewardship, governance, semantic consistency and accountability of the enterprise's official shared master data assets. Available in the cloud across a spectrum of resource delivery models, these range from single-tenant shared nothing (IaaS) to multitenant shared something (PaaS) to multitenant shared everything (SaaS).

Why This Is Important

Trusted master data is a foundational requirement of digital business. As organizations' applications have transitioned to the cloud as part of their digital transformation programs, so too has their associated data. This has subsequently shifted the center of gravity in favor of cloud-based MDM solutions.

Business Impact

The MDM market was relatively late in transitioning from on-premises to cloud-based solutions. Cloud-based MDM solutions have lowered the barrier to entry for MDM with the provision of access to subscription-based licensing models, deployment flexibility and improved time to value. They have effectively enabled support for the MDM best practice of "think big, start small, deliver incremental business value." As the center of gravity shifts to the cloud, it opens up greater opportunities for data sharing and syndication ecosystems.

Drivers

- **Stated vendor direction is cloud:** Vendors will ultimately pull the market to the cloud with product roadmap priorities supporting cloud-based platforms to streamline their own product management overheads.
- **Gravitational pull of their application and data ecosystems:** MDM creates a single source of truth for master data across the enterprise's heterogeneous application landscape. As the center of gravity for these applications and their data shifts to the cloud, MDM logically follows.
- **Acceptance of cloud for master data:** MDM has been slow to follow the broader software solutions market in transitioning to cloud deployment models. Vendors that previously delayed offering cloud-based solutions are responding to demand from end-user organizations now ready to embrace cloud for their most critical data — their master data.
- **Lower barrier to entry:** Cloud has lowered the barrier to entry for MDM programs, permitting expansion into a previously untapped and broader client base. SaaS also reduces some MDM skills pain points.
- **Increased availability of cloud-based offerings:** Few MDM vendors were cloud native from the outset. Through the end of 2022, MDM software solution vendors continued and completed their transitions to subscription and cloud-based solutions.
- **Scalability:** To handle compute intensive requirements such as ML/AI for matching.
- **Delivery of incremental business value:** Facilitates the best practice of a more granular and business outcome drive approach to MDM.

Obstacles

- **Migration complexity:** Not all MDM solutions are cloud-native. Some solutions rearchitected for cloud may lack functional parity in the near term as the products mature and may require a lift-and-shift migration from on-premises to cloud requiring external support services.

- **Installed base:** Although vendors are motivated to transition existing clients to cloud, clients will do this at their own pace and over time — in the absence of a hard trigger.

- **Governance:** As master data is heavily shared, a need for real-time integration into associated data sources and processes arises. Organizations in transition to cloud must optimize the business processes and more complex governance of a hybrid ecosystem.

- **Complexity of navigating the vendor landscape:** SaaS alleviates some MDM skills challenges. PaaS/IaaS offer greater configuration flexibility. Licensing constructs vary with MDM spend for some counting toward clients' cloud provider committed annual spend.

- **Best practices persist:** Cloud does not alleviate the business challenges related to being successful with MDM.

User Recommendations

- Take a "*think big, start small, deliver incremental value*" approach to MDM by leveraging cloud as the enabler for business value.
- Conduct a thorough review of current governance practices as a precursor to cloud readiness. Governance complexity increases in a hybrid ecosystem.
- Map and actively track the center of data gravity within your organization for each master data domain to identify and plan for prospective transition points for the cloud.
- Review and document integration complexity to provide a manageable integration scenario that does not negate any benefits of cloud-based MDM.
- Evaluate gaps in capability between candidate vendors' cloud-based and on-premises MDM solutions to determine when and whether a migration between the cloud and on-premises environments is viable.
- Cost should not be the driver for adoption of cloud MDM. Without appropriate capacity planning and cost modeling, cloud services may prove more expensive on a total cost of ownership (TCO) basis. Due diligence is required around capacity planning and TCO modeling.

Sample Vendors

Ataccama; CluedIn; Informatica; Profisee; Reltio; Semarchy

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

Sufficient Truth

Analysis By: Nick Duffy

Benefit Rating: Transformational

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

Sufficient truth promotes data governance and quality standards that are fit-for-purpose, not monolithic. It stems from a recognition that not all corporate data can be governed to the same degree without significant cost. It entails a hybrid governance model — embracing both centralized and distributed data owners — which enables data to be managed in multiple locations with scalable governance. This delivers highly accurate data where needed and rich, relevant data everywhere else.

Why This Is Important

Legacy data and analytics (D&A) governance often aims to build a single version of the truth, resulting in expensive, unsustainable, and unscalable governance efforts, as well as data silos and hidden data sources. Sufficient truth orchestrates governance trade-offs among the risks, opportunities and efforts of enterprise data governance. This improves the accessibility and relevance of data while enabling adaptive governance infrastructures that can support digital finance and advanced analytics.

Business Impact

Sufficient truth seeks to simplify D&A governance with flexible principles — spanning from controlling to autonomous — that promote enterprise data quality and directional accuracy. This builds trust and collaboration between data stakeholders, rather than creating silos. It also supports a scalable governance infrastructure (e.g., hybrid or federated) capable of sustaining D&A quality as data volumes, varieties and velocities accelerate.

Drivers

- Poor data quality is among the most frequently mentioned inhibitors of analytics, digital, and data science and machine learning (DSML). Sufficient truth distributes the work of D&A governance horizontally and vertically to ensure data is governed to the extent it should be, as determined by stakeholders and decision makers that use the data in their operations.
- Finance's traditional approach to governance is inefficient at ingesting new data types and sources for use in modern analytics. Democratization of data and analytics across many domains further drives the need for a more integrated, collaborative and democratized approach to D&A governance.
- Increasingly matrixed work environments demand a more flexible governance approach that doesn't overemphasize a single model (e.g., accounting) for all data.
- D&A governance is most successful as a collaborative effort, given the range of expertise needed to determine the necessary level of governance. Sufficient truth enables outcome owners (e.g., business leaders) to contribute to and improve D&A governance initiatives.
- Other data management and reference innovations (e.g., universal data catalogs, data glossaries, semantic data models) enable sufficient truth with self-service approaches. Data may now be stored in multiple repositories with clarification of acceptable inconsistencies across datasets. In turn, this enables more flexible data architectures that use a suite of data lakes, warehouses, marts and hubs.

Obstacles

- Sufficient truth requires an effective data stewardship model and outcome owners that will commit resources to data governance.
- Many outcome owners rely on D&A for decision making but lack the data literacy to inform governance trade-offs. Owners tend to defer governance to data professionals, failing to understand or accept accountability for the data their organization creates.
- Motivating diverse outcome owners to assume more governance responsibilities remains problematic. Few outcome owners are familiar with the fundamentals of governance, and many remain unconvinced it is their responsibility.
- Sufficient truth requires a collaborative approach to D&A governance, managed by lead data stewards capable of coordinating and negotiating enterprise data quality. Most organizations have yet to establish expectations of data stewards.
- Sufficient truth requires a mental shift of decision makers to become comfortable with imperfect data, even when the benefits of the trade are communicated.

User Recommendations

- Identify and prioritize key business outcomes and outcome owners. Speak with these owners to decide which data needs to be centrally governed and which is better suited for distributed governance.
- Create simple frameworks, such as driver maps or metrics cascades, to map data's relevance to outcomes. These identify relevant processes and key quantifiable metrics associated with outcomes, as well as the data sources used to derive key metrics and KPIs. Note any data sources that could be improved upon.
- Conduct an assessment to determine the quality of enterprise data assets and their relevance to decision making or other processes. Use this assessment to develop a rolling plan, working domain-by-domain, to determine necessary governance actions based on how the data will be used.
- Explicitly define D&A governance roles and responsibilities within finance by updating teams' job descriptions. Finance teams, at all levels of the organization, need to take ownership of D&A governance activities.

Gartner Recommended Reading

Data Management: A Single Version of the Truth No Longer Works

Digital Business Success Needs Adaptive Data and Analytics Governance

Fit-for-Purpose Data Watermarking (NFU Mutual)

Driver Mapping: A Cornerstone of Finance Data and Analytics

Overcoming Barriers to Advanced Analytics: 4 Lessons from UCB's Digital Finance Transformation

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

Climbing the Slope

Finance Data and Analytics Stewardship

Analysis By: Valeria Di Maso

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Early mainstream

Definition:

Finance data and analytics stewardship is the creation, analysis, management and control of finance processes, information and data to enforce governance policies and standards. Information in this context includes finance data, analytics, algorithms, documents, images and metadata — effectively, any and all data assets.

Why This Is Important

Financial planning and analysis (FP&A) leaders can lead stewardship efforts due to their understanding of key financial data and experience as producers and consumers of data. Organizations with established data and analytics (D&A) stewardship practices are better equipped to leverage technology that supports operationalization and automation of governance via AI/machine learning (ML). As FP&A leaders deploy predictive and prescriptive analytics, they must lead enterprisewide data stewardship.

Business Impact

D&A stewardship leads to better governance oversight, accountability and understanding of decision rights relating to D&A across the enterprise and within business areas. This increases the level of business collaboration to engage, facilitate trade-offs, innovate and drive governance to support mission-critical business outcomes. It also leads to better coordination of data governance efforts across the organization.

Drivers

- Most FP&A organizations are deploying advanced analytics, and leaders must recognize that effective D&A governance and advocacy are requisites for success.
- With the emergence of generative AI applications (such as ChatGPT), stewardship is required to ensure proper behavior and trust in how AI is applied in the business.

- Operational support in a day-to-day business context for D&A governance initiatives requires finance data and analytics stewardship.
- Lack of enterprise cohesion in data quality, standards, use and processes lead to decision paralysis and leadership confusion, when data stewardship is missing.
- D&A stewardship's work focuses on problem solving, making it a critical enabler for continuous improvement of strategic D&A programs.

Obstacles

- FP&A leaders have been slow to recognize their responsibility in guiding D&A governance, and have previously thought of this work as solely the domain of IT and chief data officers (CDOs).
- Despite the wider acceptance of information stewardship needs, many organizations have relied on reactive and often heroic efforts of “citizen stewards” to solve data problems. This approach is widespread but insufficient, as it holds back decision making.
- Most organizations are not yet ready to invest the necessary time and money on the right solutions or training their business users to deliver an operational stewardship function. Instead, they are trying to shape D&A stewardship by using a trial-and-error approach before committing to an established discipline.
- Data is power, and it's what gives business unit stakeholders their value to the business. This makes them reluctant to share data with finance.
- Decentralized decision making, based on siloed and shadow data, leads to inefficiencies.

User Recommendations

- Collaborate to create an approach that leverages the existing framework, if stewardship exists in IT.
- Clarify the stewardship process and establish information stewards' reporting lines for consistency with desired business outcomes. IT can execute the instructions and recommendations of stewardship (e.g., data maintenance or policy execution).
- Establish stewardship to guarantee AI's trustworthiness and explainability. .

- Commit to information stewardship that spans multiple business areas, when strategic programs such as MDM or compliance already exist or are underway. Also, identify a lead information steward in finance.
- Build a business case for adopting an enterprise data governance model by demonstrating how it helps deliver on key priorities.
- Ensure to not outsource the work of policy enforcement, due to the lack of context and limited business domain knowledge of the outsourcing partners.

Gartner Recommended Reading

Finance's Role in Enterprise Data Governance: Executive Champion and Lead Steward

Ignition Guide to Creating a Finance Data Governance Policy

Ignition Guide to Building a Data and Analytics Governance Program for Finance

Distilling Data Governance Essentials for FP&A Leaders

Use Value Streams to Create a Common Language for Finance Data Definition Governance

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?

Semantic Search

Analysis By: Stephen Emmott

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Early mainstream

Definition:

Semantic search processes natural language to capture the relationships between words in addition to words themselves. By processing “things, not strings,” entities such as people, organizations and products are revealed along with their attributes and relationships. This serves to better mediate between intent and outcome within a given context, improving relevance and accuracy.

Why This Is Important

Semantic search amplifies performance by indexing words and their relationships to one another. This enables the processing of what is represented by content and data, as well as the context and intent of natural language queries, thereby improving the accuracy and relevance of experiences compared to keyword-based search. Such an approach impacts all types of search, especially domain-specific, and it can represent a critical component of generative AI pipelines in a variety of design patterns.

Business Impact

Semantic search can:

- Elevate employee experience by connecting employees to one another, or to information, based on their expertise, activities or need for knowledge in the digital workplace.
- Enhance customer experience by supporting product selection and purchase, or self-help before or after sale in digital commerce.
- Decompose natural language into data for analysis and automated processing.
- Build context for prompt engineering, elevating relevance and accuracy, which facilitates the adoption of generative AI.

Drivers

The adoption of semantic search is driven by:

- The continued need for more relevant experiences in the context of digital commerce, enterprise applications, analytics and automation. For instance, semantic search is considered a necessary prerequisite for conversational search applications such as conversational commerce.
- The capabilities and expectations of generative AI models, and the application of these models, as demonstrated by ChatGPT. Semantic search can be a key component of mature and complex design patterns that leverage large language models through retrieval augmented generation. Semantic search enables meaning to be processed reliably in the workplace.
- The capability to extract data from content including audiovisual sources, thereby extending the reach of search to a wide range of data sources and formats, enabling multistructured analytics. Semantic search also enables natural language query in analytics and business intelligence platforms.
- The need to support discovery in the context of multiagent systems.

- Advances in vector-based embedding to provide an effective, math-based alternative to rule-based approaches to processing the semantics of natural language. Such approaches are delivering capability in a way that is easily attainable and scalable for enterprises. This is also leading to advances in knowledge graphs, enabling explicit representation of the entities and concepts that words represent, their attributes and relationships.
- The emergence of composite AI and the ability to marshal multiple AI approaches such as rules and machine learning to optimize capabilities, for example, analyzing semantics across languages and harmonizing search results for multilingual datasets.
- The availability of data exchanges and marketplaces for ontologies and other semantic assets. In addition, the entry barrier is low due to availability of open-source platforms and toolkits such as txtai.

Obstacles

- Vendors are working to accommodate foundational models, and their use through prompt engineering, into their products.
- While foundational models are flexible, rule-based approaches require further development — meaning those products that utilize a composite approach tend to be limited to specific languages.
- There is a requirement for continuous commitment to the development and optimization of models, rules and data structures. Specific use cases mean fewer customers, resulting in models and rules that are at lower levels of maturity. Use is typically related to private, confidential or commercial sources, which results in learning and adaptation that cannot be shared beyond individual customers.
- Semantic models need to be developed, whether from scratch or derived from existing models available externally. Professional services are often required to get started and continue, especially as insufficient data governance and management lead to poor-quality data.

Gartner

User Recommendations

- Use insight engines or digital commerce search as the platform for building applications that utilize semantic search. Where applicable, consider vector databases as alternative or complement. Start by reviewing your existing insight engine's capabilities to ensure they support semantic search, and factor it in as a requirement when selecting new products.
- Use semantic search as an integral part of your generative AI initiatives, to facilitate grounding in the context of prompt engineering.
- Link your search and insight activities with your other natural language technology initiatives and in the context of your organization's wider data fabric. Seek rationalization and consolidation with natural language query solutions for structured data where appropriate.
- Test the performance of semantic search in the business context. Engage subject matter experts to contribute — ideally independently and proactively — to the maintenance of semantic search capabilities.

Sample Vendors

Access Innovations; Cohere; expert.ai; Glean Technologies; IBM; Mindbreeze; Ontotext; Semantic Web Company; Yext

Gartner Recommended Reading

Magic Quadrant for Insight Engines

Critical Capabilities for Insight Engines

Quick Answer: What Impact Will Generative AI Have on Search?

AI Design Patterns for Large Language Models

How Large Language Models and Knowledge Graphs Can Transform Enterprise Search

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

Appendixes

See the previous Hype Cycle: Hype Cycle for Finance Data and Analytics Governance, 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)

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 (August 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)

Acronym Key and Glossary Terms

Data Governance	Gartner defines data governance as the specification of decision rights and an accountability framework to ensure appropriate behavior in the valuation, creation, consumption and control of D&A. Data governance, therefore, aims to ensure data quality, transparency, relevance and accountability.
Data Literacy	Gartner defines data literacy as the ability to read, write and communicate data in context, with an understanding of the data sources and constructs and the analytical methods and techniques applied, and the ability to describe the use-case application and resulting business value or outcome.

Document Revision History

Hype Cycle for Finance Data and Analytics Governance, 2022 - 27 June 2022

Hype Cycle for Finance Data and Analytics Governance, 2022 - 26 January 2022

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

Finance’s Role in Enterprise Data Governance: Executive Champion and Lead Steward

Hype Cycle for Finance Analytics, 2023

Hype Cycle for Data Management, 2023

Hype Cycle for Data and Analytics Governance, 2023

Hype Cycle for Emerging Technologies in Finance, 2023

Table 1: Priority Matrix for Finance Data and Analytics Governance, 2023

Benefit	*Years to Mainstream Adoption*			
↓	***Less Than 2 Years*** ↓	***2 - 5 Years*** ↓	***5 - 10 Years*** ↓	***More Than 10 Years*** ↓
Transformational		Augmented Data Quality Data Literacy Sufficient Truth	Active Metadata Management Adaptive D&A Governance Data Fabric for Finance Responsible AI	
High	Augmented Data Management Cloud MDM Semantic Search	AI Governance D&A Governance for Finance Data Classification Data Hub Strategy Data Observability DataOps Finance Data and Analytics Stewardship Lakehouse Master Data Management	Anomaly and Error Detection D&A Governance Platforms Graph DBMS for Finance	Connected Governance
Moderate		Data Lake		
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 (August 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)