Hype Cycle for Enterprise Information Management, 2021

Published 2 August 2021 - ID G00747537 - 145 min read

By Analyst(s): Donna Medeiros, Andrew White

Initiatives: Data and Analytics Strategies

Hype around data and analytics continues to be driven by digital acceleration and the demand for organizations to be data-driven. Data and analytics leaders must seize their opportunity to maximize the business value of data by aligning people, process, data and technology.

Analysis

What You Need to Know

The continued reset, recovery and renewal post-COVID-19 has refocused data and analytics (D&A) strategies away from business-as-usual toward data-driven real-time decision making. Deliberate coordination of the most critical business information across numerous initiatives has become key. Enterprise information management (EIM) has thus come back into focus as a core discipline to support the D&A strategy and operating model.

Digital transformations have been accelerated, with many markets rallying, and recoveries are in progress as COVID-19 vaccines roll out. Investments and data and analytics initiatives have gained further momentum, which in turn drives hype in many data and analytics areas (see Survey Analysis: Executive Leaders Should Align to Board Priorities for 2021). Hype leads to research and further investments, but acquiring tools, technologies and skilled resources is not enough. Data and analytics leaders, including chief data officers, need to maximize the business value of data by adopting integrative disciplines such as EIM for structuring, describing and governing information assets to drive business outcomes.

Enterprises are drowning in siloed data and a broad array of tools, yet so little is actually used and challenges abound with long lists of pain, needing processes and methods to derive value from data and improve business outcomes (see CDO Agenda 2021: Influence and Impact of Successful CDOs in the Sixth Annual CDO Survey). Uncertainties abound regarding what the post pandemic business environment will look like, however the new normal will be heavily reliant on new data, models, insights, tools and systems to fuel organization's digital acceleration. At the same time, organizations are navigating economic, political, business and technology risk, and seeking to prioritize investments to provide superior data-driven decision making across the organization.

Data and analytics leaders must now seize the opportunity that digital acceleration has provided, and increase their influence and impact in their enterprise's data-driven ambitions by:

- Clarifying executive accountability to drive digital and data mandates
- Focusing on ROI from D&A investments
- Prioritizing and pursuing modern change management

Additional Hype Cycles for 2021 that are related to this EIM Hype Cycle will help to form a holistic view of D&A:

- Hype Cycle for Analytics and Business Intelligence, 2021
- Hype Cycle for Data Security, 2021
- Hype Cycle for Privacy, 2021
- Hype Cycle for Artificial Intelligence, 2021
- Hype Cycle for Data Management, 2021
- Hype Cycle for the Digital Workplace, 2021
- Hype Cycle for Analytics Governance and Master Data Management, 2021

The Hype Cycle

EIM operates as a core discipline at the heart of any effective data and analytics strategy. It allows data and analytics leaders to maximize the impact of the most critical information assets on enterprise outcomes — linking data management and governance to operations and analytics use cases in a cohesive data ecosystem. This results in effective organization, optimized design and efficient deployment.

“Enterprise” in “enterprise information management” does not mean “enterprisewide”; it means “enterprise class,” whether scaling big or small. Its scope is “information,” meaning all forms of data, content, insights and analytics, encompassing the entire global data ecosystem — data creation, data rights, data acquisition, integration and data sharing. The required disciplines span capabilities across people, process and technology, while linking all together on the most important information assets.

Organizations are starting to plan for a data-driven future where EIM helps them achieve their goals. D&A leaders play a key role in every organization as they seek better business decisions in relation to a “new normal.”

Over the past year, enterprises experimented, piloted, and prioritized. They continue to invest in analytics, AI, ML and other advanced techniques relying on the use of synthetic data and trying new techniques such as developing their own data fabric. They are now forming fusion teams, embedding lean and agile techniques to speed delivery, and so enterprise information hype is as high as ever (see Accelerate Digital Business Aspirations by Becoming a Data-Driven Enterprise).

COVID-19-related innovations that enterprises put in place (D&A-specific technology and practices) bypassed the typical Hype Cycle time to value due to extreme pressures related to health, safety and reopening. Such was the case for vaccine management, health risk mitigation technologies and health passes, which had undergone innovation and adoption in fast turnaround.

Organizations need to enable D&A governance that capitalizes on lessons learned from existing use cases that have demanded adaptive approaches for over a decade. Decision rights in the context of business and information value are one example — for others, see Toolkit: Creating a Modern Data and Analytics Strategy and Operating Model.

This Hype Cycle focuses on EIM across all data and analytics initiatives, with less emphasis on discrete technologies than other Hype Cycles. It focuses on the most critical EIM concepts, disciplines and practices that will drive alignment within organizations — supporting the necessary transformative business outcomes from data and analytics programs, and demonstrating global enterprise value. This Hype Cycle captures the most hyped aspects driving interest in data and analytics strategy.

We highlight the following practices and processes in the 2021 Hype Cycle for EIM, noting many have moved up and are on the rise since 2020, and there are several additions.

New entrants to this year's Hype Cycle include:

- Analytics governance
- Connected governance
- COVID-19 health risk mitigation
- Data fabric
- Augmented data quality
- Health pass
- Vaccine management
- Graph analytics

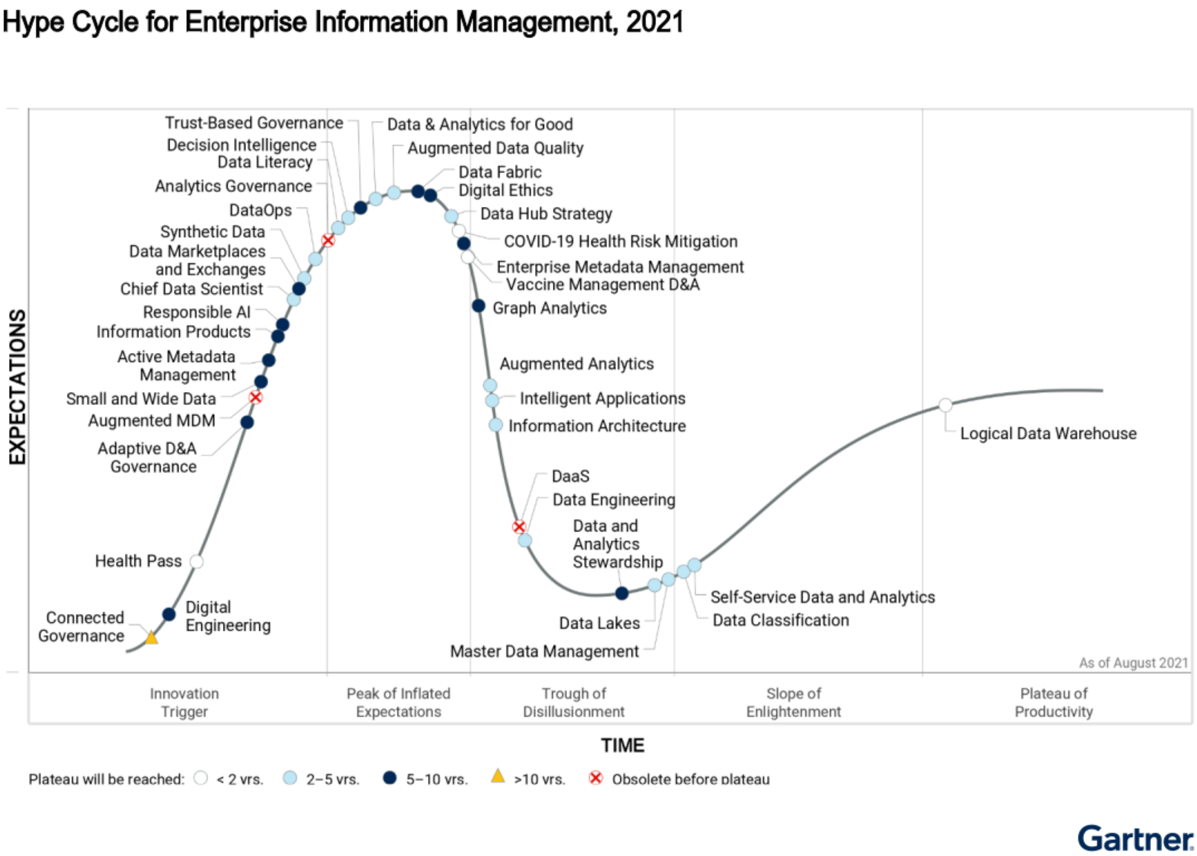
Those on the rise:

- Adaptive data and analytics governance

- Data literacy
- Data sharing
- Trust-based governance

The shape of the Hype Cycle reflects the fact that most disciplines and practices can take a long time to mature to the point where the mass market can gain repeatable value from them. Notable exceptions are practices related to COVID-19 response and risk mitigation, which have sprung onto the Hype Cycle and bypassed the typical time to value due to extreme pressures related to health, safety and reopening driving fever-pitch adoption. Data fabric burst onto the scene in part to fill a void for a connected data ecosystem and to support advanced analytics and near-real-time decision making. Self-service analytics leaped to the Slope of Enlightenment with heightened hype around empowering analytics use across the enterprise. Data engineering hype has continued its progression as it approaches the bottom of the Trough of Disillusionment. Otherwise, there is continuous innovation in and around the work of D&A. Opportunities to achieve value through EIM utilization remain abundant, so we see a large number of innovations grouped together closely either just before the Peak of Inflated Expectations, or just after.

Figure 1: Hype Cycle for Enterprise Information Management, 2021



Source: Gartner (August 2021)

Downloadable graphic: Hype Cycle for Enterprise Information Management, 2021

The Priority Matrix

EIM can be used to organize, implement and deploy any enterprise-level or business unit data and analytics initiatives. It can also be used to align and link people, roles, business processes, data and technology across these initiatives. Strong programs will therefore be transformational in nature, and it will take several years before most of the practices reach the desired level of maturity.

A significant number of initiatives and innovations will offer transformational or high-value benefits during the next two to 10 years (as shown in our Priority Matrix for EIM). These include:

- Data and analytics stewardship
- Data classification
- Data marketplaces and exchanges
- DataOps
- Digital ethics
- Enterprise metadata management
- Graph analytics
- Information products
- Responsible AI
- Small and wide data
- Synthetic data
- Trust-based governance

In addition, we foresee two notable innovations: adaptive data and analytics governance, and decision intelligence.

Adaptive data and analytics governance is an organizational capability that enables context-appropriate governance styles and mechanisms to be applied to different data and analytics scenarios in order to achieve desired business outcomes. Although there is now greater diversity and complexity in business scenarios than ever before, data and analytics governance has typically continued to adopt a single, control-oriented approach.

Decision intelligence provides several new tactics that, when taken together, will help organizations reengineer how they make decisions. First is the idea that decisions can be thought of as connected across business processes and value streams. Any effort that helps shed light on how people — and systems — make decisions can help identify opportunities to improve. Second, explicitly modeling some decision making will also help improve practices and apply new techniques. It will take some time for decision intelligence to become standard practice everywhere. To further explore the boundaries of decision intelligence, analyze the following technologies:

- Influence engineering
- Active metadata management
- Genomics and epigenetics
- AI-augmented software engineering
- Real-time incident center
- Digital platform conductor tools

Table 1: Priority Matrix for Enterprise Information Management, 2021

(Enlarged table in Appendix)

Benefit	*Years to Mainstream Adoption*			
↓	*Less Than 2 Years* ↓	*2 - 5 Years* ↓	*5 - 10 Years* ↓	*More Than 10 Years* ↓
Transformational	Health Pass	Augmented Analytics Decision Intelligence Intelligent Applications	Active Metadata Management Adaptive D&A Governance Data Fabric Digital Engineering Responsible AI	
High	COVID-19 Health Risk Mitigation Logical Data Warehouse Vaccine Management D&A	Augmented Data Quality Chief Data Scientist Data Classification Data Engineering Data Hub Strategy Data Literacy DataOps Information Architecture Master Data Management Self-Service Data and Analytics Synthetic Data	Data and Analytics Stewardship Data Marketplaces and Exchanges Digital Ethics Enterprise Metadata Management Graph Analytics Information Products Small and Wide Data Trust-Based Governance	Connected Governance
Moderate		Data & Analytics for Good Data Lakes		
Low				

Source: Gartner (August 2021)

Off the Hype Cycle

In 2021, three innovations left the Hype Cycle for Enterprise Information Management. These are:

- Business continuity management (BCM) D&A: This technology was short-lived and has disappeared as fast as it came onto the 2020 Hype Cycle.
- Infonomics: Innovations related to data marketplaces and exchanges, together with a renewed focus on data monetization, replaced infonomics.

- Data catalog: The hype concerning data catalogs waned and has been replaced by innovations concerning data classification.

Several disciplines have expanded in scope to include a data focus along with analytics:

- “Analytics Stewardship” has been renamed “Data and Analytics Stewardship,” and “Data for Good” has been renamed “Data and Analytics for Good.” The former changes to recognize the term most often used by clients; the latter to recognize the broader applicability of data and analytics.
- “Citizen Data Science” has been renamed “Chief Data Scientist,” as that role is getting mode hype today.
- “Small Data” has been renamed “Small and Wide Data” to help frame the scope more appropriately with the hype in the market.

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, spanning one or several legal entities and 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, and provides strategic oversight and accountability management across them through a virtual framework.

Business Impact

Senior business executives and board members in organizations spanning multiple legal entities and geographies will find value in exploring connected governance to address cross-enterprise strategic issues and opportunities, without setting up new governance boards. Organizations anticipating or undergoing M&As will also find value in connected governance, enabling risk management to be addressed earlier and allowing experimentation with governance bodies prior to their formal adoption.

Drivers

- The pace of digitalization is putting pressure on senior leaders across multiple business functions to respond to business 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 the case 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.
- While many responses to strategic, cross-enterprise business challenges have been to establish another layer of governance, this adds a greater overhead cost, creates another layer of bureaucracy and is often inflexible. Some strategic challenges (e.g., 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.

Obstacles

- Connected governance expects to leverage existing governance bodies, but some of these governance 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 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 requires support and facilitation from a strategic governance office, which may not yet have the skills needed.

User Recommendations

- Evaluate whether connected governance would be of value to your organization. If you operate in a complex environment, across multiple legal entities and geographies, there may be challenges that are currently difficult to address. In such situations, raise an agenda item at your executive committee to initiate a cost-benefit assessment and report its findings. If this is not the scenario in your organization, connected governance may not be relevant for you.

- Connected governance needs the support of a 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 governance office may be sufficient for now. However, large enterprises operating in diverse and complex ecosystems and expecting to address many strategic scenarios may need to establish a dedicated strategic governance office to support connected governance.

Gartner Recommended Reading

Dynamic Risk Governance Works Better Than the 3 Lines of Defense Model

Executive Leadership: Strategic Risk Management Primer for 2021

How to Compose a Board of Directors for Strong Governance, Awareness and Success

Digital Engineering

Analysis By: Carlton Sapp, Soyeb Barot, W. Roy Schulte

Benefit Rating: Transformational

Market Penetration: Less than 1% of target audience

Maturity: Embryonic

Definition:

Digital engineering is a digital transformation approach that integrates various physical and digital life cycle activities from conception to decommissioning of digital assets. It primarily consists of a platform for fusing models through model-based system engineering (MBSE) and heterogeneous data pipelines across the entire life cycle of all physical, digital or virtual assets.

Why This Is Important

Digital engineering is a departure from conventional software engineering life cycle management practices because it breaks down barriers between engineering disciplines, promotes collaboration, and reduces integration work and duplication of effort. It involves major changes to IT organizational structure, governance and development methodologies. Digital engineering produces traceability throughout a system's life cycle, better stakeholder alignment, lower costs and faster enhancements.

Business Impact

- Increases visibility into engineering disciplines.
- Promotes consistency across the entire asset life cycle.
- Reduces time required for all aspects of engineering.
- Increases performance across the entire life cycle of all digital, physical and virtual assets.
- Provides a consistent view of models across conception, development and production environments to reduce operational delays.
- Dramatically reduces the time to deliver a concept into operations.

Drivers

- Organizations need increasingly large and complex cyber-physical systems, but conventional software engineering practices and organizational structures are unable to produce and manage them.
- Organizations want increased transparency and visibility into and across all engineering life cycles and processes when building cyber-physical systems.
- Organizations want enhanced communication from conception to operations of physical, digital and virtual assets for greater flexibility/adaptability in design.
- Organizations want increased efficiencies in engineering practices in support of configuration management, but many lack effective discipline and a platform to enable more efficient engineering practices.
- Sharing data and information across engineering disciplines is often a downstream task that occurs too late to capture changes early in the development life cycle.
- When dealing with large cyber-physical systems, organizations want greater reliability and productivity in all aspects of their production line and manufacturing process.
- Organizations find it difficult to harmonize architectural patterns from various subdisciplines and to reduce the overlap in system engineering, software engineering and enterprise architecture due to a lack of contextualization of data and models.
- Organizations are increasingly concerned with reproducibility, explainability, audit and governance of digital assets. Siloed engineering efforts make it very difficult to capture changes in design data and other metadata that would enable these efforts.
- Organizations want greater interoperability: the ability to manage and exchange standardized data between systems owned by different engineering disciplines. Organizations understand the importance of a fully integrated environment to perform activities, collaborate and communicate across stakeholders.

Obstacles

- Digital engineering lacks a formal definition. It has been defined differently by industry leaders, creating confusion.
- Digital engineering principles are yet to be standardized. Some organizations focus on digital technologies and others on well-established practices.
- Vendor solutions have not reached the maturity of plug-and-go tools that can be downloaded and easily installed. Some digital engineering offerings require the enterprise to have certain dedicated infrastructure in place and the resources to manage it. In other cases, solution integrators will be heavily involved in the initial setup and in integrating the solutions into the existing technology stack.

User Recommendations

- Deliver a consistent view of engineering data across various stages of a life cycle by storing models in an integrated environment that is accessible across the entire life cycle of a digital asset.
- Increase the flexibility of model development, deployment and maintenance by providing engineers with an infrastructure to support models and tools.
- Shift from managing explicit knowledge and information with documents to managing knowledge and information via graph models.
- Extend existing data and analytics architectures to support data exchanges that define standards, data and architectural patterns for all engineering disciplines across the entire life cycle of the digital asset.

Sample Vendors

Arena Solutions; IBM; Maplesoft; NI; Visual Paradigm

Gartner Recommended Reading

What Data and Analytics Leaders Need to Know and Do About Digital Twins

What to Expect When You're Expecting Digital Twins

Innovation Insight for Model-Based System Engineering

Health Pass

Analysis By: Ben Kaner, Sharon Hakkennes, Donna Medeiros

Benefit Rating: Transformational

Market Penetration: 20% to 50% of target audience

Maturity: Emerging

Definition:

A health pass provides a digital means to verify an individual's health status against set criteria defined by the verifying enterprise. These solutions access trusted sources to verify credentials, the output from which is an indicator of adherence to the criteria. Capabilities vary between solutions and include verification of vaccination status, laboratory test results, temperature and self-reported health declarations.

Why This Is Important

Governments and enterprises need to reopen offices, venues and travel locally and internationally without creating a resultant surge of COVID-19 disease. This requires an ability to control the risks of doing so. As the virus continues to evolve, this is likely to be required for some time. In addition, the ability to easily present and act on health information in a health pass will improve the ability to respond to individual and societal health challenges as the situation continues to evolve.

Business Impact

By exploiting a health pass, governments and enterprises should be able to reopen more efficiently while still complying with controls intended to reduce the risks from COVID-19 to their citizens and staff. They are also more likely to be able to control a subsequent surge. Over time, public health could be improved by exploiting the infrastructure to address other health and disease threats as they continue to emerge.

Drivers

- A widespread need for a health pass is driven by COVID-19 — a prime example of a contagious virus that is sufficiently infectious and virulent to spread worldwide and cause the near-collapse of multiple health systems and economies.

- Multiple vaccines against COVID-19 have been created and distributed across multiple countries. Until enough people worldwide are successfully vaccinated, there remain significant risks — and this takes time. As the virus continues to evolve, populations are likely to require revaccination before the first vaccination cycle is complete.

- To prevent new major outbreaks, governments need to exploit multiple tools — including detection (testing); breaking the infection chains (isolation); treatment (e.g., antivirals, antibodies); and of course, prevention (social distancing, vaccination).

- While health passes and the narrower capability of digital vaccine certificates cannot fully resolve this problem, they have the potential to enable economies to operate more normally, and reopen earlier than would otherwise be the case. As the impacts have been so high on many industries and human activities, this is critical to economies, government revenue, public health and civic behavior (e.g., outbreaks of unrest), and provides multiple drivers. Drivers include businesses needing to contain the risks both of health and of financial liability associated with customers, visitors and staff coming into contact with each other — and do so with as low a cost as possible. Government facilities need to support and deal with citizens without discrimination. Also, citizens wish to reestablish their ability to engage socially and economically as much as possible.

Obstacles

- International/cross-jurisdiction standards and acceptance, particularly variable privacy and security requirements, are challenging technically and politically.
- Availability of vaccination registries — some countries have the infrastructure in place, others do not.
- Differing interpretations about what is needed and at what level (WHO vs. EU Digital Green vs. Chinese International Travel Certificate).
- Who will be liable/responsible at point of entry?
- Cost of implementation and infrastructure availability, particularly for low- and middle-income countries.
- Cost and resource to operate the system, e.g., additional queue-handling capacity.
- Significant political and ethical concerns about creation of a mechanism of social exclusion.
- Reluctance from citizens to have to share health information for any purpose.

User Recommendations

- Ensure adaptability to varying standards and evolving needs by separating identity, event recording and verification into different capabilities.
- Allow for broad usage by considering how people with externally provided information can be enabled to operate within your jurisdiction, either by enabling local verification of other health passes or by allowing a registration into your health pass ecosystem.
- Ensure that the system can be used even under stress by ensuring there is a viable option for offline checking of the pass.
- Minimize privacy and security challenges by designing around minimal transfer of information.

Sample Vendors

Clear; Covid-19 Credentials Initiative; Good Health Pass Collaborative; IBM; International Air Transport Association; International Chamber of Commerce; Lumedic; The Commons Project

Gartner Recommended Reading

Demystifying the Relationship Between a Digital Vaccine Certificate and Health Pass

Data Ethics of Tracking Employee Vaccination Status

Now Is the Time to Make Digital Identity Work for Citizens and Governments

Adaptive D&A Governance

Analysis By: Saul Judah

Benefit Rating: Transformational

Market Penetration: Less than 1% 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 data and analytics scenarios in order to achieve desired business outcomes.

Why This Is Important

As organizations accelerate 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 leads to 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 the enterprise to respond faster to business opportunities and become more competitive, resilient and risk-aware.

Drivers

The hype relating to D&A governance has gradually increased over the past year, primarily due to the following factors:

- As organizations have largely addressed their more immediate needs to support a distributed workforce during the COVID-19 pandemic, their attention is again returning to the need for better accountability and a decision framework for D&A to drive organizational outcomes.
- Digital business demand has not only resumed but in some cases exceeded prepandemic levels (e.g., M&As), and D&A leaders are increasing their efforts to improve their governance practices to the levels of performance needed.
- Recognition by both vendors and organizational leaders that increased investment in infrastructure e.g., D&A platforms, cannot yield the ROI as expected, 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, as a response.

Obstacles

- In most organizations, maturity levels for D&A governance practices are lower than in other areas, such as data management and analytics. Governance is typically IT-oriented, center-out and control-oriented. Furthermore, current approaches resemble compliance rather than governance.
- 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.
- The COVID-19 pandemic has slowed down the adoption rate of D&A governance, with organizations changing focus to operational issues as a short-term response. However, as organizations have moved into recovery, the interest and hype for adaptive D&A governance is resuming due to its ability to promise flexible responses for changeable situations.

User Recommendations

- Use IT Score for D&A to evaluate your maturity and readiness to enhance governance capabilities. Don't attempt to establish agility and autonomous governance if the foundations for control- and/or outcome-based governance are missing.
- Create a proof-of-concept initiative to test the applicability of one of the more advanced governance styles (e.g., autonomous) in your environment, and 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 and develop a business case and strategic roadmap to establish adaptive D&A governance.
- Establish the control and outcome styles of adaptive governance first. Then evolve to the agile and autonomous styles. Proceed on the basis of "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

7 Must-Have Foundations for Modern Data and Analytics Governance

Reset Your Information Governance Approach by Moving From Truth to Trust

IT Score for Data & Analytics

Augmented MDM

Analysis By: Malcolm Hawker

Benefit Rating: Moderate

Market Penetration: Less than 1% of target audience

Maturity: Embryonic

Definition:

Augmented master data management (MDM) is the application of graph, machine learning and similar advanced technologies to MDM. Augmented MDM extends traditional MDM capabilities to reduce manual data management and governance tasks. It generates insights on complex relationships within and across both application and master data, allowing for technology to play an active role in enabling more adaptive and context-centric approaches to master data management.

Why This Is Important

Organizations use MDM to accelerate and differentiate digital transformations, particularly around customer/citizen and product/service experiences. This includes traditional as well as newly evolving MDM value propositions, like the use of application data to uncover previously unknown relationships across master data entities. Data and analytics leaders' need for such insights requires them to expand the scale and efficiency of their MDM programs, thus driving investment in augmented MDM.

Business Impact

Augmented MDM provides three primary benefits:

- Increased revenue associated with enhanced digital experiences, such as the use of augmented MDM to uncover a previously unknown sales opportunity.
- Reduced infrastructure costs by enabling more cost-efficient storage of, and compute of, increasing volumes of master data for entity relationship and discovery.
- Increased operational efficiencies and lower MDM program operating costs via the automation of several data management and governance tasks.

Drivers

- Increasing business focus on digital transformation and the critical role MDM plays in the enablement of a data-driven business strategy.
- Businesses' focus on creating "360-degree views" of their master data domains, fueling an expansion in the number of data types (such as application data) and data sources included in the scope of an MDM program. This is driving a need for increased scale, speed and efficiencies provided by augmented MDM.
- Business benefits realized by taking more context-centric views and governance of core master data objects.
- Improved operational efficiencies realized by IT organizations from the automation and scale related to the management and governance of master data.
- The value generated by exposing previously unknown relationships between master and nonmaster data within large, unstructured datasets.

Obstacles

- Augmented MDM is still in the early stages of development, both from the perspectives of customer demand and vendor focus.
- Augmented MDM suffers a lack of consistent definition and many MDM vendors claim to provide the functionality, but all in varying degrees.
- Individual capabilities described by augmented MDM are also available across other data and analytics solutions.
- The primary obstacles to broader market availability are technical. First, MDM software vendors must upgrade their platforms to integrate new features into existing infrastructures — which for many means a revamping of their underlying architectures. Second, companies seeking to reap the full benefit of augmented MDM will typically adopt cloud-based MDM deployments, a migration many have yet to make. We expect both of these obstacles to be largely overcome in the next two years, with broader availability from MDM vendors and wide-scale adoption shortly thereafter.

User Recommendations

- Be skeptical of vendors who focus their value propositions primarily on augmented MDM. This is because many MDM requirements, especially legal, finance or compliance-driven use cases, remain firmly rooted in traditional approaches to MDM.
- Evaluate your expected business outcomes and use cases — as some may not align well with augmented MDM, regardless of vendor.
- Be wary of augmented MDM vendors that lack the ability to integrate the use of graph and AI for data discovery, profiling and visualization with governance processes and operational MDM use cases.
- Contemplate a "best of breed" approach to solving for your needs when incumbent solutions lack augmented MDM capabilities. The combination of MDM, customer data platforms, analytics platforms, data quality and metadata management tools may support augmented MDM requirements for specific use cases.

Sample Vendors

CluedIn; Informatica; Reltio; Riversand Technologies

Gartner Recommended Reading

Top 10 Trends in Data and Analytics, 2020

Modern Data and Analytics Requirements Demand a Convergence of Data Management Capabilities

Small and Wide Data

Analysis By: Jim Hare, Pieter den Hamer, Svetlana Sicular

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

Small data is about the application of analytical techniques that require less data but still offer useful insights. Wide data enables the analysis and synergy of a variety of small, large, unstructured and structured data sources. Together these approaches apply a variety of data augmentation techniques such as X analytics, simulation, synthetic data, transfer learning, federated learning, self-supervised learning, few-shot learning and knowledge graphs.

Why This Is Important

Analytics and AI need to be able to use available data more effectively, either by reducing the required volume or by extracting more value from unstructured and diverse data sources. Small and wide data approaches enable more robust analytics, reducing dependency on big data and helping attain a more complete situational awareness or 360-degree view. This enables organizations to make analytics more resilient in the increasingly complex context of disruptions and evolving customer demands.

Business Impact

Small and wide data techniques help make analytical and machine learning models more resilient to disruptions. Small data is helpful for AI problems, where big datasets are not available, by applying less data-hungry techniques, synthetic approaches or by augmenting data through the synergy with unstructured, external or synthetic data sources. Wide data uses a broader variety of data sources to increase context and situational awareness for both human decision makers and AI applications.

Drivers

- Disruptions such as the COVID-19 pandemic have resulted in many production AI models across different industry verticals losing accuracy and relevance because they were trained using past big data that reflected how the world worked before the pandemic hit. Retraining models using the same approach was not feasible, because more recent data, just a few weeks old, was too limited to reflect the patterns of the new market circumstances.
- Organizations continue to struggle when getting started with AI projects if there is not enough volume or variety of data available to find the relevant model features or training datasets for complex models.
- Decision making is also becoming more complex and demanding, requiring a greater variety of data for better situational awareness and/or detecting rare events.
- More mature organizations that already implemented solutions for which they have enough data want to solve unique, differentiating problems where they need to overcome data size and variety limitations.

Obstacles

- Lack of tools and user skills needed to link disparate datasets across different data formats to uncover new insights or add more context to existing business decision making.
- Misperception that AI projects require large datasets before organizations can get started, resulting in lost productivity and delayed deployments.
- Organizations waiting until production AI models encounter accuracy issues rather than proactively incorporating small and wide data techniques early on as part of the model life cycle process.
- Confusion or lack of understanding about which small and wide data techniques are best for specific classes of AI problems.
- Nascent emerging techniques such as zero-, one- and few-shot learning that require specialized skills.
- Small data techniques are fragmented; they address specific challenges, such as only image analysis, only tabular data or only exclusively algorithmic side, rather than addressing the small and wide data issues overall.

User Recommendations

- Explore small data and algorithmic approaches to increase model resilience and lower the barrier to entry for AI. Data techniques include data and feature enrichment and expansion via synthetic data, external data sources, metadata, graphs, etc. Algorithmic techniques include generative AI (GANs, few-shot learning) and composite AI.
- Enrich and improve the predictive power of data by incorporating a greater variety of structured and unstructured data sources. These formats include tabular, text, image, video, audio, voice, temperature, or even smell and vibration. Wide data comes from an increasing range of internal and external data sources, such as data marketplaces and exchanges, brokers/aggregators, industry consortia, open data, social media, IoT sensors and digital twins.

Sample Vendors

Diveplane; Google; iOmniscient; Landing AI; MOSTLY AI; MyDataModels; Owkin; Veritone

Gartner Recommended Reading

Top Trends in Data and Analytics for 2021: From Big to Small and Wide Data

Tech Providers 2025: Why Small Data Is the Future of AI

3 Types of Machine Learning for the Enterprise

Understanding When Graph Technologies Are Best for Your Business Use Case

Working With Semistructured and Unstructured Datasets

Active Metadata Management

Analysis By: Mark Beyer, Guido De Simoni, Alan Dayley

Benefit Rating: Transformational

Market Penetration: 1% to 5% of target audience

Maturity: Embryonic

Definition:

Active metadata is the continuous analysis of user, data management, systems, infrastructure and data governance experience to determine the alignment and exceptions between data as designed versus operational experience. Its utilization includes operationalizing analytic outputs, operational alerts and recommendations. It identifies the nature and extent of patterns in data operations, resulting in AI-assisted reconfiguration of data and operations and use cases.

Why This Is Important

Active metadata management uses machine learning, data profiling and graph analytics to help determine data's relevance and validity. 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 even new data. At mature levels, it supports the evaluation of analytic and data biases (unintentional or intentional) as well as transparency, auditing and DataOps.

Business Impact

- Support self-service and application development by automating data content and structures and the availability discovery of data assets.
- Identify similarities among users, use cases, and reporting and analysis models across an organization to build social networks of users based on common data needs and operational and analytic requirements.
- Automate orchestration for data access, locations, processing requirements and resource allocation by enabling a balance between optimization and cost.

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.
- Human-driven data utilization cannot adapt quickly enough to the demand for the rapid discovery, access and incorporation of new data assets throughout an enterprise or organization.
- Data management is further complicated by third-party data and data utilized from adjacent and distantly removed industries.

- Organizations need a portfolio of capabilities and the ability to manage them across a range of use cases.
- The large-scale capabilities in cloud-based deployments have enabled the broadest diversity of data structures, processes and use cases to date. Intercloud demands determine the best data management approaches based on statistical analysis of the data.
- Demands are newly emerging to continuously compare experience statistics with design expectations to separate data into zones of concern: harmonic, dissonant or discordant.

Obstacles

- Active metadata management requires access to design and runtime metadata, usage and utilization statistics, user or user-group identification, graph analytics, continuous data profiling and machine learning. Prohibited or limited access to any of these assets or capabilities can inhibit the implementation of active metadata approaches.
- Automated cross-platform and tools orchestration will inhibit growth due to a reluctance among data management solution providers to make their metadata assets available to — much less accept metainstructions from — external or third-party optimization and resource allocation platforms.
- Human designers, implementers and users might resist AI-based data management approaches based on the concept that data is a valuable resource. However, the use of data is what makes it valuable. It is the combination of AI-based data management and human modulation that makes the active metadata management approach valuable.

User Recommendations

- Begin accumulating runtime logs from as many tools as possible. Analyze the logs for patterns of data used together, frequency of use, user or connection strings, queries and views executed, and even resource allocation to create a graph of which data is used, how often, by whom, for what purpose and on which platform.
- Introduce a data catalog strategy, and expand it to ingest metadata from master data management, data quality, data integration and data preparation tools and attach it to the catalog entries.
- Acquire or deploy at least one prototype combining at least three disciplines from data management to enable metadata notification between tools to be added to the runtime logs and design metadata repositories as notes. Deploy a user interface to reconfigure metadata repositories for analysis by data engineers and architects.

Sample Vendors

Alation; Alex Solutions; Ataccama; Atlan; Datalytyx; data.world; Informatica; OneTrust; Orion Governance; Semantic Web Company

Gartner Recommended Reading

Modern Data and Analytics Requirements Demand a Convergence of Data Management Capabilities

The State of Metadata Management: Data Management Solutions Must Become Augmented Metadata Platforms

5 Ways to Use Metadata Management to Deliver Business Value From Data

Information Products

Analysis By: Mike Rollings, Lydia Clougherty Jones

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

Definition:

An information product, also known as a data product, is an offering that directly monetizes data by generating value, revenue or other financial benefits. It includes licensing, bartering or sharing of data and/or insights that can be proprietary or a byproduct of business operations; provided by IoT or other instrumentation of physical products and services; or a combination of proprietary, public and exogenous data.

Why This Is Important

Providers and consumers of data are increasingly interested in developing their own information products, similar to the way firms such as data brokers have been licensing and selling information. This interest is driven by expanding recognition of the market value of enterprise data (such as an organization’s existing data assets or data collected from IoT devices). It is also driven by the desire to develop innovative offerings that generate new revenue streams or enable new market entries.

Business Impact

- Public-sector open data and data-sharing initiatives yield indirect benefits, such as increased public confidence, economic development or transparency.
- Private-sector opportunities range from selling or licensing raw data to creating information-enabled products and services, and data-sharing platforms.
- Data marketplaces and exchanges can enable publishing, curation, analysis and aggregation of data, revenue sharing and matchmaking.

Drivers

- Digital business thrives on data and insights. In a data-driven enterprise, data and analytics (D&A) is no longer afterthoughts — they are fundamental to digital business transformation and can be monetized internally and externally.

- Business executives and boards of directors are increasingly familiar with data monetization — the process of using data to obtain quantifiable economic benefit. Internal or indirect methods include using data to make measurable business performance improvements and informed decisions. External/direct methods include data sharing to gain beneficial terms or conditions from business partners, information bartering, selling data outright (via a data broker/independently), or offering information products and services (e.g., including information as a value-added component of an existing offering).

- Manufacturing and other industries have invested in IoT-enabled products and data analytics. Their companies amass a growing amount of data from those connections and some have created entirely new revenue streams from the sale of data and insights to their customers.

- Public and private sector enterprises of all sizes are realizing the power of data and insights for value creation, and are pursuing external monetization opportunities.

- Chief data officers (CDOs) and other D&A leaders are increasingly leading digital transformation initiatives where data monetization is a central part of value creation. Data monetization success is a significant driver in delivering business value to customers and business ecosystem partners.

- Enterprises may imagine that their data is quite valuable, but realize they do not have a clear strategy for how they will use or monetize it. Changing the conversation about data, defining a D&A strategy, and exploring customer outcomes, product offerings, and how data and insights could be used to achieve them lead to the identification of potential information products and services.

Obstacles

- The desire for new information-based revenue streams requires advances in treating information as an asset, changing the culture and improving data literacy to employ data-driven concepts, developing product management competencies. It also requires learning to ethically leverage data rights, while adhering to data protection, sovereignty and other restrictions.
- Most organizations are ill-prepared to skillfully collect, curate, manage, measure and monetize D&A assets.
- The required skills and competencies to develop information products differ significantly from the products and services that most enterprises currently sell. This requires advancing D&A maturity while developing commercial-level product management skills that most organizations do not possess.
- Developing offerings and pricing may require more analysis, market definition studies, insight and creativity than expected due to lack of defined markets with established pricing guidance.

User Recommendations

- Identify potential information products by identifying unmet data needs for current and potential stakeholders that support their business outcomes. Map those needs against available, IoT-derived and sourced data to conceive products.
- Apply infonomics calculations to quantify the potential and realized market value and liabilities of data assets with rigor like that for physical assets. Use that as an input for pricing.
- Co-create information products development and deployment with a third-party services provider. Evaluate data marketplaces and exchanges that facilitate selling, licensing and sharing data.
- Underpin success by fostering a data literate workforce, a data-driven strategy and a D&A operating model oriented by business outcomes.
- Establish commercial style product management and agile delivery practices. Hire a professional information product manager responsible for information procurement, product pricing and licensing, market definition, and product development.

Gartner Recommended Reading

Applied Infonomics: 7 Practices for Chief Data Officers to Monetize Information Assets

Essential Product Management Practices to Monetize Data and Analytics Assets

How to Monetize Data Assets With Your Data and Analytics Service Provider

Data Monetization Will Be the Next Frontier for Digital Insurance

Magic Quadrant for Managed IoT Connectivity Services, Worldwide

Case Study: Data Monetization Through Data Product Development (ZF Group)

Responsible AI

Analysis By: Svetlana Sicular

Benefit Rating: Transformational

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

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

Why This Is Important

Responsible AI has emerged as the top AI topic for Gartner clients. The more AI replaces human decisions at scale, the more it amplifies the positive and negative impacts of such decisions. Responsible AI pursues positive outcomes and prevents negative results by resolving dilemmas rooted in delivering value versus tolerating risks. Recently, many jurisdictions globally introduced new and pending AI regulations that challenge data and analytics leaders to respond in meaningful ways.

Business Impact

Responsible AI signifies the move toward 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, exploit and resolve the dilemmas they face. Ideally, it enhances both sides at the following levels:

- **Organizational** — Resolving AI's business value versus risk in regulatory, business and ethical constraints. It could also include employee reskilling and intellectual property protection.
- **Societal** — Resolving AI effectiveness for societal well-being versus limiting human freedoms. Existing and pending legal guidelines and regulations, such the EU's Artificial Intelligence Act, make responsible AI a necessity.
- **Customer**, **citizen** — Resolving privacy versus convenience involves a thin line between customers' readiness to give their data in exchange for goods or benefits and customer/citizen concerns about their privacy. Fairness and ethics are the greatest drivers in this space. Regulations shed light on the necessary steps — for example, the U.S. Federal Trade Committee's "Using Artificial Intelligence and Algorithms" for consumer protection. However, this does not relieve organizations of deliberation specific to their constituents.
- **Workplace** — Resolving work efficiency versus employer "creepiness" includes concerns about AI's effect on jobs and employee morale, as well as change management.

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 generative AI for creating deepfakes. Most organizations combine some of the following drivers under the umbrella of responsible AI:

- Accountability
- Diversity
- Ethics
- Explainability
- Fairness
- Human centricity
- Operational responsibility
- Privacy
- Regulatory compliance
- Risk management
- Safety
- Transparency
- Trustworthiness

Obstacles

- Unawareness of AI's unintended consequences prevails. Many organizations turn to responsible AI only after they hit AI's negative effects, whereas prevention is easier and less stressful.
- Legislative pace, uncertainty and complexity puts responsible AI on hold in many firms. It also leads to one-sided efforts for regulatory compliance, while ignoring other responsible AI drivers.
- 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 that brings together multiple perspectives and diversity of opinions is necessary, in addition to technology.
- Poorly defined accountability and incentives for responsible AI practices make responsible AI look good on paper, but ineffective in reality.

User Recommendations

Data and analytics leaders, take responsibility — it's not AI, it's you who are liable for the results and impacts, either intended or unintended.

- Combine the responsible AI aspects you currently address independently to promulgate consistent approaches across all focus areas. The most typical areas of responsible AI in the enterprise are fairness, bias mitigation, ethics, risk management, privacy and regulatory compliance.
- Designate a champion accountable for the responsible development of AI, for each use case.
- Raise awareness of AI differences from the familiar concepts continuously. Provide training and education on responsible AI, first to most critical personnel, and then to your entire AI audience.
- Establish an AI ethics board to resolve AI dilemmas. Ensure diversity of participants and the ease to voice AI concerns.
- Participate in industry or societal responsible AI groups. Learn best practices and contribute your own, because everybody will benefit from this.

Sample Vendors

Google, H2O.ai, IBM, Microsoft, SAS, Tazi.ai

Gartner Recommended Reading

Predicts 2021: Artificial Intelligence and Its Impact on People and Society

Top Trends in Data and Analytics for 2021: Smarter, More Responsible and Scalable AI

Cool Vendors in AI Governance and Ethical Response

Case Study: Ethical AI With an External Board (Axon)

What Non-Technology Executives Should Do in Support of Responsible AI Initiatives

Financial Services CIOs Must Focus AI Investments on 'Responsible AI' in 2021

Chief Data Scientist

Analysis By: Carlie Idoine, Farhan Choudhary, Erick Brethenoux

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

Chief data scientist is an emerging, leadership role responsible for translating data and analytics strategy into efficient and effective implementations of advanced D&A products and services. The role is typically the most senior data science position within an organization and has a specific focus on applied data science approaches.

Why This Is Important

The complexity, pervasiveness and criticality of advanced analytics and AI require dedicated attention to coordinate use and support of data science, ML and AI for businesses' top priorities. Chief data scientists are responsible for enterprise data science teams and for execution of the D&A vision. They must think tactically, assess the current situation and deliver value today while planning strategically, in parallel, to coordinate and maximize the future use of advanced analytics.

Business Impact

The business value derived from AI done right is substantially more than what can be accomplished by other means. Chief data scientists are key for unlocking this value. They propel the use of advanced analytics forward by:

- Translating business vision to data and advanced analytics priorities
- Coordinating all advanced analytics initiatives across the organization, minimizing silos while maximizing adoption
- Delivering measurable results and value to the organization

Drivers

- Organizations lacking the ability to understand and communicate the value of advanced analytics, ML and AI initiatives are making efforts to manage and coordinate both centralized and decentralized teams to deliver measurable business outcomes.
- Organizations are digitizing and automating more of their processes with AI and analytics at the core. As AI becomes a critical function in processes underlying digital businesses, it requires leadership skills and oversight.
- Siloed, unstructured approaches to advanced analytics and AI not only consume significant time and resources but also increase the risk and minimize return on investment and overall trust in these techniques. The role of chief data scientist aids breaking down and eliminating silos.
- There is increased commitment to driving advanced analytics throughout the organization but the aim is to do so using a consistent and managed approach. Also, organizations face huge analytical and AI debt as projects grow and become more complicated. A chief data scientist leads the charge of ensuring an efficient path to agile delivery of D&A initiatives.
- The democratization of data science boosting the adoption of ML techniques has generated an increasing amount of ML models that are often not operationalized. The need for better coordination with the lines of business and IT, and the harmonization of DSML practices, require a chief data scientist role.
- The proliferation of “shadow IT” practices across organizations is often the source of MLOps inefficiencies and the inefficient use of data science talents. Chief data scientists have a unique opportunity to federate those talents, physically or virtually.

Obstacles

- D&A leaders and their IT and business partners often lack effective influence, organization, process and practice to deliver, operationalize and scale DSML solutions and approaches.
- The chief data scientist role may not have enough organizational clout and defined authority to drive the enterprisewide changes required to reap the benefits of AI.
- The lack of business recognition often leaves the organization open to data science poaching, weakening the role of the chief data scientist.
- Recruiting and retaining an experienced chief data scientist, with the right blend of management, technical, business and communication skills, is challenging.

User Recommendations

- Define the chief data scientist role as a complement to other CxO roles, recognizing that alignment between these roles is critical.
- Work both within (the internal IT team and the broader organization) as well as outside the organization to orient the chief data scientist within the broader community and identify opportunities for learning and partnership.
- Empower chief data scientists to build a diverse team, develop processes and procure tools to deliver models in a way that builds trust while tracking the impact on key business priorities and value generated.
- Leverage the chief data scientist role to drive and coordinate application and exploration of advanced analytics methods and techniques to align those methods with real, prioritized business problems.

Gartner Recommended Reading

The Chief Data Scientist Role Is Key to Evolving Advanced Analytics and AI

The Current State of Demand for the Chief Data Scientist Role: Q1 2021 Report

Data Marketplaces and Exchanges

Analysis By: Eric Hunter, Jim Hare, Lydia Clougherty Jones

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Emerging

Definition:

Data marketplaces and exchanges are ecosystems centered around data assets that provide infrastructure, transactional capabilities and services for participants. Marketplaces support data monetization via one-time or recurring subscription transactions while exchanges prioritize sharing of assets over monetization objectives. Both can incorporate public or private-sector participants and commercial or no-cost assets, such as open data.

Why This Is Important

Public cloud adoption has largely minimized data gravity as the primary barrier to data marketplaces and exchanges as the balance of data has shifted to CSPs in many organizations. Adoption of data marketplaces and exchanges remain in the early phases, but providers are bringing together stakeholders with mutual interests. Participants are attracted to diverse third-party asset selection, simplified data access/integration, simplified procurement, and reduced operating and transaction costs.

Business Impact

- Increase the prevalence of third-party data assets for data science models, data enrichment and line of business operational demands (marketing campaigns, etc.).
- Reduce the complexity, time and cost demands for sourcing third-party data assets.
- Increase data monetization opportunities and market reach via marketplaces.
- Data exchanges reduce the barrier to entry for nonmonetized data sharing.
- Expand the visibility, findability, variety and availability of data products.

Drivers

Marketplaces and exchanges remove barriers to the acquisition of third-party data in support of increasingly data-driven business outcomes and underlying models. They also reduce the exclusivity of specific types of data products through increased competition that will reduce price points for similar data products over the long term. Marketplaces and exchanges continue to increase in adoption through the following drivers:

- Digital business transformation has positioned data products as a key enabler of emerging business outcomes — increasing the demand for third party data, the growing role of business ecosystems, and the growing awareness and need for companies about partnering

- There is increased awareness across both public and private sectors in terms of the value associated with both internal and external data assets and products.

- COVID-19 has driven change across the behavior of many customers, resulting in model accuracy drift for data science models that rely heavily upon first-party customer data for key model features. Third-party data acquired from marketplaces and exchanges has become more attractive as an enabler for remediating these model accuracy issues and in the creation of new model features.

- There is increased adoption of public cloud which has reduced on-premises data gravity limitations that slow the physical movement and integration of data across parties.

- The number of public and private data providers for data marketplaces and exchanges continues to increase — providing both an increased level of specialization and breadth in terms of available third-party data product offerings.

- There is rising awareness of internal and external data sharing benefits through increased virtual work environments, COVID-19-centric use cases and increased enterprise reliance on public cloud.

Obstacles

- Data privacy legislation and risks of sharing data impede the pursuit of monetizing and productizing specific types of data, which reduces participation within data marketplaces and exchanges.
- It is a challenge to have mutually acceptable standards for governance of data-sharing scenarios without a balance of common cause and enlightened self-interest.
- Evolving organizational data ethics and sharing standards can prevent the adoption of third party data and/or creation of third-party data products that drive data marketplaces and exchanges.
- Enterprise procurement processes and public cloud account structures provide friction for lines of business user spend within public cloud data marketplace and ecosystems.
- Absent the specialized capabilities for evaluating relevant data products for a given use case, the volume of available marketplace and exchange data product offerings and lacking metadata can overwhelm buyers evaluating the ability of specific data products for new model features.

User Recommendations

- Promote organizational participation in your marketplace of choice to accelerate the time to business value over the use of independent data asset providers or consumers.
- When seeking to monetize data products, look to marketplaces for the transactional infrastructure to allow internal efforts to focus on building unique and differentiated data products.
- Leverage providers that operate within or are optimized for your cloud providers of choice to reduce data movement complexity and improve integration consistency.
- Evaluate prospective data providers beyond their data product selection and coverage. Examples are integration/access (such as APIs), value-added capabilities and exception alerting.
- Adapt data management policies and standards to account for the realities presented by data marketplaces and exchanges.
- Explore the value of third-party data to increase analytic insights by either adding context as new attributes or through additional data science model features.

Gartner Recommended Reading

Smart Data Sharing Requires Mapping Use Cases to Architectures and Vendor Solutions

How to Monetize Data Assets With Your Data and Analytics Service Provider

Flip 'Don't Share Data' Mantras — Introducing Gartner's 'Must Share Data Unless' Data Sharing Model

Data Sharing Is a Business Necessity to Accelerate Digital Business

Top Trends in Data and Analytics for 2021: From Big to Small and Wide Data

Synthetic Data

Analysis By: Anthony Mullen, Alexander Linden

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

Synthetic data is a class of data that is artificially generated, i.e., not obtained from direct observations of the real world. Data can be generated using different methods such as statistically rigorous sampling from real data, semantic approaches, generative adversarial networks or by creating simulation scenarios where models and processes interact to create completely new datasets of events.

Why This Is Important

One of the major problems with AI development today is the burden in obtaining real-world data and labeling it so AI models can be trained effectively. This is remedied by synthetic data. Furthermore, synthetic data is critical in removing personally identifiable information (PII).

Business Impact

Adoption is increasing across various industries, along with use in natural language processing (NLP) applications. We predict massive increase in adoption as synthetic data:

- Avoids using PII when training machine learning (ML) models via synthetic variations of original data or synthetic replacement of parts of data.
- Reduces cost and saves time in ML development as it is cheaper and faster to obtain.
- Improves ML performance as more training data leads to better training outcomes.

Drivers

- In healthcare and finance, buyers' interest is growing as synthetic data can be used to preserve privacy in AI training data.
- To meet increasing demand for synthetic data for natural language automation training, especially chatbots and speech applications, new and existing vendors are bringing offerings to market. This is expanding the vendor landscape and driving synthetic data adoption.
- Synthetic data applications have expanded beyond automotive and computer vision use cases to include data monetization, external analytics support, platform evaluation and the development of test data.
- Increasing adoption of simulation techniques is accelerating synthetic data.
- While row/record, image/video, text and speech applications are common, R&D labs are expanding the concept of synthetic data to graphs. Synthetically generated graphs will resemble but not overlap the original. As organizations begin to use graph technology more, we expect this method to mature and drive adoption.

Obstacles

- Synthetic data still has significant flaws. It can have bias problems, miss natural anomalies, be complicated to develop or may not contribute any new information to existing, real-world data.
- Data quality is tied to the model that develops the data.
- Buyers are still confused over when and how to use the technology with other data pipeline tools. As the number of techniques in data and model pipeline increases, buyers struggle to determine what techniques to use to achieve their aims (e.g., synthetic data, federated learning, differential privacy) and how to use them together.
- Synthetic data can still reveal a lot of sensitive details about an organization so security is a concern. An ML model could be reverse-engineered via active learning. With active learning, a learning algorithm can interactively query a user (or other information sources) to label new data points with the desired outputs, meaning learning algorithms can actively query the user/teacher for labels.

User Recommendations

- Identify areas in your organization where data is missing, incomplete or expensive to obtain, and thus, currently blocking AI initiatives. In regulated industries, such as pharma or finance, exercise caution and adhere to rules.
- Use synthetic variations of the original data or synthetic replacement of parts of data, when personal data is required but data privacy is a requirement.
- Begin with a sampling approach and leverage data scientists to ensure statistical validity of the sample and distribution of the synthetic data.
- Leverage specialist vendors while the technology matures.
- Mature toward the simulation-driven approach, emphasizing creating agents and processes within a simulation framework to generate permutations of interactions that result in synthetic data.

Sample Vendors

AI.Reverie; Bitext; MOSTLY AI; Neuromation; Tonic; Twenty Billion Neurons (TwentyBN)

Gartner Recommended Reading

Top Trends in Data and Analytics for 2021: From Big to Small and Wide Data

2021 Strategic Roadmap for Enterprise AI: Natural Language Architecture

Cool Vendors in AI Core Technologies

DataOps

Analysis By: Robert Thanaraj, Alan Dayley, Ted Friedman

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

DataOps is a collaborative data management practice focused on improving the communication, integration, automation, observability and operations of data flows between data managers and data consumers across an organization.

Why This Is Important

DataOps is emerging as a response to frictions around the consumption and use of data across the organization. Organizations need DataOps practices for:

- Improving the communication among data managers and data consumers to bring both parties on the same page
- Integrating data flows across the enterprise
- Reducing the cost of operations through data pipeline automation
- Providing transparency and reliability through good monitoring and observability, increasing the use/reuse of data

Business Impact

DataOps focuses on improving organizational speed and trust in delivering data pipelines and related artifacts by co-creating "decision quality" data with the consumers. By automating data pipelines, it reduces risk and increases opportunities for data to be used across the organization. It improves timeliness and effectiveness of various data-reliant activities such as BI, data science and operational data uses, while also enhancing productivity of data engineers facing repetitive data tasks.

Drivers

- The primary driver for DataOps is organizations' desire to improve speed and efficiency while producing trusted data, and increase data literacy and self-service while consuming decision-quality data.
- DataOps practices help organizations overcome challenges caused by fragmented teams/processes and delays in delivering data in consumable forms. DataOps does this by setting up cross-functional teams with data managers and consumers, positioning these teams as satellites within lines of business (supported by a central data and analytics team).
- DataOps can be supported by a diverse range of technologies, which helps organizations further improve the use and value of data by automating and orchestrating data delivery with appropriate levels of security, quality, observability and governance.

Obstacles

- While end-user customers are starting to explore DataOps for material impact, there are as yet no standard practices or industrywide guardrails that organizations can readily adapt to.
- Organizations have too much baggage from legacy approaches to data delivery, which is hard to change. For example, many are not organized in the right way and are not clear on the fact that DataOps is not a technology issue.
- The technology spectrum that supports DataOps is broad and varied — ranging from full-portfolio players, to specific, point solutions (such as integration, metadata, testing, monitoring, governance), to orchestration-focused. There are plenty of choices, but the technology investment decisions are complex due to overlapping and/or missing capabilities.

User Recommendations

- Choose data and analytics projects that are struggling due to lack of collaboration, are overburdened by the pace of change to meet business demands or where service tickets from data consumers are piling up. Such projects create the best opportunity to show value as DataOps is most successful on projects targeting a small scope with some level of executive sponsorship.
- Pilot DataOps principles by applying the core DevOps approaches to data management. Reach out to application development and deployment teams within your enterprise that have successfully applied DevOps practices in managing software.
- Evaluate technology investment to address your weakest delivery stages (like testing, monitoring, governance) or for end-to-end data pipeline orchestration.
- Track metrics such as time to market, degree of automation, code quality, cost-efficiencies and business impact in dollar amount. Metrics demonstrate business value to stakeholders.

Sample Vendors

Atlan; Cognite; Composable Analytics; DataKitchen; Hitachi Vantara; IBM; SecuPi; Tamr

Gartner Recommended Reading

Data and Analytics Essentials: DataOps

Introducing DataOps Into Your Data Management Discipline

3 Ways to Deliver Customer Value Faster Using DataOps

Operational AI Requires Data Engineering, DataOps and Data-AI Role Alignment

How to Build a Data Engineering Practice That Delivers Great Consumer Experiences

At the Peak

Analytics Governance

Analysis By: Andrew White, Kurt Schlegel

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

Analytics governance is the enforcement of D&A governance policy along the analytics pipeline from data discovery through to analytics model deployment, and access to the analysis and insight. Though the markets use the term "governance" here, the reality is that what is being sought is not related to policy setting but actually executing and enforcing policy along the analytics pipeline. A more appropriate name would be analytics stewardship.

Why This Is Important

For many years, organizations have limped along with semitrusted data in their analytics pipeline. With the increased adoption of data lakes and data sharing in the last few years, the gap between expectations and reality is starting to hit home. With analytical and forecasting models breaking in 2020 due to the COVID-19 pandemic, business leaders are finally grasping the reality: they, their decisions and even their organizational survival is held hostage to bad data and analytics.

Business Impact

If organizations do not implement data and analytics governance through their analytics pipeline, no amount of spending on the latest analytics tool or technology will survive scrutiny. Worse, business decisions may backfire and organizational performance may suffer as a result. With the right business outcome and adaptive governance focus, the least amount of mission-critical data and analytics will be governed, thus assisting with trusted and reliable analysis and insight leverage.

Drivers

- 2021 is marked with small and wide data, not big data. It is also marked with data and analytics everywhere and at the edge. With these new trends, additional pressures are being put on your organization.

- A shift in focus from truth to trust in governing data and analytics assets due to the vastness of data now to hand for analysis and the lack of accountability in third party sources.
- Protection and provenance of the inbound data to the analytics pipeline and at-rest data in the warehouse or lake.
- Need for enhanced integrity of the analytical model being developed.
- Guidance for ethical consideration.
- Continually evolving permissions for access to the data for model development or for consumption of the analysis output even as organizational boundaries shift almost daily.
- Often third-party-driven retention requirements for risk mitigation.
- Preservation of privacy that may even conflict when operating across multiple jurisdictions.
- No amount of technology can help; though it is with technology that D&A governance policies are applied and enforced (stewardship) along your analytics pipeline.
- While the hype is firmly placed on analytics governance, the reality is that organizations need to focus on extending their D&A governance program along the data and analytics pipeline.

Obstacles

- The biggest obstacle is the lack of a clear line-of-sight between a piece of rogue or untrusted data in a data warehouse or dashboard and its impact on a business decision or outcome. This lack of visibility between data and outcome helps explain why business leaders seem disinterested in the work of governance and stewardship.
- The second is that many organizations think that “analytics governance” is actually something different and distinct to data and analytics governance. This is just natural forces looking at the boundaries in front, and not visionaries looking beyond to see the same patterns and solutions emerging.

User Recommendations

- Recognize the work of policy setting (i.e., governance); policy enforcement (i.e., stewardship) and policy execution (i.e., management). Apply your response to your analytic pipeline.

- Extend or connect your data and analytics governance work so that the policy setting and enforcement efforts can be aligned — this will reduce redundancy and save money, and lead to improved outcomes

- Note also that most cloud analytics and cloud infrastructure vendors really don't understand what your needs are in this market. They mostly think it all hinges on tracking data lineage. That is nice, but not sufficient.

- Don't assume your analytics, business intelligence, data science or artificial intelligence solutions support your requirements for analytics stewardship (or governance). At most, they might respect the odd rule and follow it (i.e., management/execution). You may need to build your own capability outside of those solutions, until the vendors wake up and build what you need.

Sample Vendors

Alation; Collibra; ZenOptics

Gartner Recommended Reading

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

Use Enterprise Metadata Management to Extend Information Governance to Analytics

The State of Data and Analytics Governance Is Worse Than You Think

Data Literacy

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

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

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. It is the ability to describe the use-case application and resulting business value or outcome.

Why This Is Important

Data and analytics are pervasive in all aspects of all businesses, in communities and in our personal lives. The ability to understand, interpret and act upon data — data literacy — is increasingly foundational to the digital economy and society. Data literacy helps explain to the board how data and analytics manifest in a company's use cases, explain how to identify, access, integrate and manage internal and external datasets, and describe advanced analytics techniques and enabling AI.

Business Impact

Data-driven enterprises require explicit and persistent organizational change to achieve measurable business outcomes. Employees know their organization is serious about change only when they see their leaders changing their own behavior. CDOs need to promote and orchestrate "leadership moments" where they act as role models, exemplifying new cultural traits at critical points. Central to success will be the ability to guide the workforce by addressing both data literacy and data-driven culture.

Drivers

- With the steady rise of the digital economy, and the need for businesses to be digitally literate, there is growing recognition of the role that employees' data literacy plays within an organization's overall digital dexterity.
- The role of the data and analytics function has changed. It is now at the core of an organization's business model and digital platforms.
- CDOs can emulate their higher-performing peers by putting much more emphasis, energy and effort into meeting the change management requirements of their data and analytics strategies.
- 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 data and analytics skills, knowledge and competencies.
- CDOs need to take immediate action to create and sustain data literacy. Quick wins build momentum, but lasting and meaningful change takes time because it requires people to learn new skills and behave in new ways.

Obstacles

- Lack of common data literacy models/frameworks/standards
- A piecemeal approach to training and certification
- Aversion to change
- Lack of talent and poor data literacy
- Lack of initiatives to address cultural and data literacy challenges within strategies and programs
- "Data literacy" means different things to different providers: from enhanced data visualization skills to fostering curiosity about data more broadly
- Overall adoption will still take years

User Recommendations

- Create a strong narrative vision of desired business outcomes, particularly with respect to innovation. Raise awareness through storytelling.
- Call out examples of “good” and “bad” data literacy to promote desired behaviors.
- Work with stakeholders who have enthusiasm and appetite, and who recognize that improved data literacy is a factor for success.
- Partner with HR and business leaders to identify the level of data literacy, learning goals and outcomes for various job roles and personas. Use data literacy assessments to evaluate current data literacy levels and desire to participate.
- Go beyond vendor product training to focus on people’s other role-related skills. Use a mix of training delivery methods (classroom, online, community, on the job) to improve overall learning effectiveness.
- Align training and self-service solutions with a broader data literacy portfolio to meet the data literacy needs of both data consumers and creators.

Sample Vendors

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

Gartner Recommended Reading

Roadmap for Data Literacy and Data-Driven Business Transformation: A Gartner Trend Insight Report

Tool: Communicating the Need for Data Literacy Improvement

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

Tool: Data Literacy Personas

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

Decision Intelligence

Analysis By: Erick Brethenoux

Benefit Rating: Transformational

Market Penetration: 5% to 20% of target audience

Maturity: Emerging

Definition:

Decision intelligence (DI) is a practical discipline used to improve decision making by explicitly understanding and engineering how decisions are made and how outcomes are evaluated, managed and improved via feedback.

Why This Is Important

The current hype around automated decision making and augmented intelligence, fueled by AI techniques in decision making, is pushing DI toward the Peak of Inflated Expectations. The COVID-19 pandemic has revealed the brittleness of decision models; rebuilding those models to be resilient, adaptable and flexible will require the discipline brought by DI methods and techniques. A fast-emerging market around various software disciplines is starting to provide sensible answers for decision makers.

Business Impact

Decision intelligence helps organizations:

- Improve the impact of business processes by materially enhancing the sustainability of organizations' decision models based on the power of their relevance, the quality of their transparency and the strength of their resilience, thus making decisions more transparent and auditable.
- Reduce the unpredictability of decision outcomes by properly capturing and accounting for the uncertainty factors in the business context.

Drivers

- A dynamic and complex business environment, with an increasingly unpredictable and uncertain pace of business. The combination of AI techniques (such as NLP, knowledge graphs, machine learning), and the confluence of several technology clusters around composite AI, smart business process, decision management and advanced personalization platforms, are creating a new market around decision systems platforms supporting the DI discipline.

- Need to curtail unstructured, ad hoc decisions that are siloed and disjointed. Often uncoordinated, such decisions promote local optimizations at the expense of global efficiency.

- Expanding collaboration between humans and machines, supplemented by a lack of trust in technologies (such as AI) increasingly replacing tasks and promoting uneasiness from a human perspective. DI practices promote transparency, interpretability, fairness, reliability and accountability of decision models critical for the adoption of business-differentiating techniques.

- Tighter regulations that are making risk management more prevalent. From privacy and ethical guidelines to new laws and government mandates, it is becoming difficult for organizations to fully understand the risk impacts of their decisions. DI enables an explicit representation of decision models, reducing this risk.

- Uncertainty regarding decision consistency across the organization. Lack of explicit representation of decisions prevents proper harmonization of collective decision outcomes. This is remedied by DI.

Obstacles

- Decision-making silos have created data, competencies and technology clusters that are difficult to reconcile and could slow down the implementation of decision models.
- An inadequate organizational structure around advanced techniques, such as the lack of an AI center of excellence, could impair DI progress.
- Reputation-damaging outcomes from autonomous decision models (from embedded analytical assets to self-contained machine agents) and the failure to understand their collective impact impede DI adoption.
- Lack of proper coordination between business units and inability to impartially reconsider critical decision flows within and across departments diminish the effectiveness of early DI efforts.

User Recommendations

- Improve the outcomes of decision models and accommodate uncertainty factors by evaluating the contributing decision-modeling techniques.
- Promote the sustainability of cross-organizational decisions by building models using principles aimed at enhancing traceability, replicability, pertinence and trustworthiness.
- Improve the predictability and alignment of decision agents, by simulating their collective behavior while also estimating their global contribution versus local optimization.
- Develop staff expertise in traditional and emerging decision augmentation and decision automation techniques, including descriptive, diagnostic (interactive data exploration tools), predictive (machine learning) and prescriptive (optimization, business rule processing and simulation) analytics.
- Tailor the choice of decision-making technique to the particular requirements of each decision situation by collaborating with subject matter experts, AI experts and business process analysts.

Gartner Recommended Reading

Improve Decision Making Using Decision Intelligence Models

How to Manage the Risks of Decision Automation

How to Choose Your Best-Fit Decision Management Suite Vendor

AI Security: How to Make AI Trustworthy

Top Strategic Technology Trends for 2021: AI Engineering

Trust-Based Governance

Analysis By: Andrew White, Saul Judah

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

Trust-based governance prescribes and reports the desired and actual trust, reliability and efficacy observed in data, analytics, systems, partners and organizations, so that asset usage and risk mitigation are appropriate. Traditional D&A governance (including data quality, MDM and ADM) focuses on single dimensions (e.g., yes/no compliance or single version of the truth) for policy or definitional rules and policies and leads to excessive effort and high risk mitigation costs.

Why This Is Important

Trust-based governance alleviates the demands put on people, processes and data that use distinct and precise definitions of boundaries for policies such as quality and standard definitions. It supersedes traditional approaches to governance, which use simplistic yes/no conclusions (e.g., the attribute is or is not of the right quality) that may not serve business needs adequately. This reinforces a view that governance is slowing work down and preventing the business from achieving its goals.

Business Impact

Where business complexity is held back by traditional or trust-based data quality and entity resolution definitions (e.g., yes/no or single version), a trust-based approach (e.g., graduated) will help reduce and align the governance effort to the value such data offers:

- Truth-based efforts lead to excessive costs, controls and delays in achieving business goals.

- Trust-based approach will reduce costs, mitigate risks and help businesses achieve their expected goals more effectively.

Drivers

- Hype around trust and trust-based governance models is increasing. This is because the use of data to drive decisions and outcomes is at fever pitch. Boards of directors for the last two years have highlighted how analytics and AI are critical game-changing technologies (see Survey Analysis: Board Directors Say Pandemic Drives Increased Investments in IT). Trust is at the center of effective use of analytics since it aligns with the vagaries and contexts of complex situations and models. Absolute forms of data quality and definitions do not work well in these new environments, which are predicated on openness, shareability and exploration.

- Trust cannot be assumed; it needs to be evaluated, and is hard to earn when people or relationships are involved. New technologies like ML-augmented knowledge graphs and entity resolution/data quality are helping discover relationships in data to help infer or inform additional insights on data use, which can help reinforce trust. For example, data logs can be used to discover the most often used version of "customer" data.

- We have seen some trust-based governance in a few key situations, often related to intelligence work, but they are very manual. Technology vendors have yet to operationalize support for the need. In 2021, we are seeing more commercial and private organizations realize they need to develop their traditional data quality, MDM and D&A governance programs with a more flexible, trust-based approach. More pilots are underway, and there is greater likelihood of vendors offering more operational capabilities.

Obstacles

- Lack of maturity is the biggest obstacle to trust-based governance. Data and analytics teams often spend much of their time firefighting operational issues (e.g., data quality issues that prevent a business transaction or a report with “bad” data) and don’t have the opportunity to step back and assess their landscape and understand lineage, curation and usage of their ecosystem by their organizational users. As a result, they are unable to put in place the framework that will help them reduce the issues that they face on a daily basis.

- Numerous widely deployed technologies, such as data dictionaries, glossaries, catalogs, data quality and entity resolution, and business rule engines, all need to evolve to support the enhanced trust-based models.

- Intelligence agencies have used trust-based governance in the last few years, but widespread adoption across industries is probably held back due to lack of investment in supporting or enabling tools.

User Recommendations

- Pilot a trust framework to some critical data and its source where DQ efforts appear costly or onerous. Explore how it can help users of the data align their risk mitigation efforts to the value and use of the (trusted) data.

- Where data quality only has a single dimension (e.g., yes/no), use a simple three-tier framework (untrusted, unknown, trusted), and test the boundaries and the savings in time and effort to govern such data.

- Align trust-based governance to your data use case and enterprise goals. Many use cases will still work very effectively with traditional rules and policies, such as those defined by yes/no qualification or definitions — these might be central MDM. But for departmental use, that data might be treated differently, especially if enriched with third-party data for analytics, such as a customer data platform (CDP). It is in the CDP that trust may be more useful than a truth-based approach.

Gartner Recommended Reading

7 Must-Have Foundations for Modern Data and Analytics Governance

Reset Your Information Governance Approach by Moving From Truth to Trust

Data & Analytics for Good

Analysis By: Carlie Idoine, Jorgen Heizenberg

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

"Data & Analytics for Good" is a movement in which people and organizations transcend organizational boundaries to use data and data-driven insights for social impact. The usage may be within a data sharing, analytics and BI context or in more sophisticated data science and machine learning use cases and specifically does not include leveraging social causes to market products or services.

Why This Is Important

NGOs and public-sector organizations try to be more data-driven, but lack knowledge, skills, expertise and access to tools to leverage data to fulfill their missions. Meanwhile, commercial organizations (vendors and end-user organizations) have D&A expertise that can be used for the good of society. By crossing traditional organizational boundaries, these stakeholders unite in their efforts to leverage data for good and to provide more meaningful work to the most sought-after employees.

Business Impact

The "Data & Analytics for Good" movement gives resources to the public sector and NGOs through free or reduced-cost technology, data and people. In the commercial sector, participation in "D&A for Good" initiatives can be through philanthropic benefits that attract and retain workers, and provide resources. "D&A for Good" initiatives can signal social responsibility to investors. From a technology and vendor perspective, they can drive a positive impact on both marketing and sales.

Drivers

Focus on "Data & Analytics for Good" initiatives is growing with increased visibility and understanding of the value provided from these efforts.

- "Data & Analytics for Good" efforts initially focused on educational enhancement, providing clean water, reliable food, ecological management and arts and science community support.
- The number of organizations — from universities and communities to vendors — having a focus on "Data & Analytics for Good" has increased.
- D&A has the transformative power and availability to gain insights into the descriptive existence and diagnostics of the root causes of human suffering, and predict and prescribe how to action changes which will mitigate it.
- Significant market momentum for "Data & Analytics for Good" comes from vendors (both software and services), and has been especially evident in relation to COVID-19 response efforts.
- Many organizations' COVID-19 responses have spotlighted exceptional "D&A for Good" initiatives as organizations teamed together to share data and analytic resources to combat and manage the crisis.

Obstacles

- Justification for "Data & Analytics for Good" is difficult to initiate and maintain because the goals and objectives are considered altruistic, and lose influence relative to business delivery-driven efforts.
- "D&A for Good" programs are often dismissed because the funding stream is considered temporary or at least inconsistent.
- Some "D&A for Good" programs are rejected when they seek to qualify inclusion or delivery based upon personally identifiable information or data that is considered ethically dangerous.
- Lack of transparency can lead to negative unintended consequences; data ethics and trust must be incorporated to avoid data misuse.
- Low levels of data literacy get in the way of effectively using the contributed data and analytics to achieve the desired social impact.
- Legal impediments, technical data format standards and the practical issue of data cataloging and aggregation have also hindered efforts.

User Recommendations

- Leverage free resources (people/services, software, technology, data) from organizations that support “D&A for Good” projects.
- Participate in community events such as those hosted by DataKind, Kaggle and universities to collaborate on “D&A for Good.” Contribute to, and explore, open data in support.
- Allow employees time to work on philanthropic initiatives as part of social responsibility. Use this HR benefit as a differentiator in recruiting and skills’ enhancement.
- Evaluate internal, external and open data to assess its usefulness for social purpose, while also adhering to privacy and security policies. Instill data ethics considerations in data use and sharing.
- Drive data literacy to help identify, understand and recommend controls for “D&A for Good” use cases in an effort to provide transparency without endangering individuals’ privacy or sensitivities.
- Grow awareness about “D&A for Good.” Share internal and external case studies as well as resources that demonstrate what “D&A for Good” is and its potential impact.

Sample Vendors

Alteryx; DataKind; DataRobot; Esri; Google; IBM; Salesforce; SAS; Tableau; Teradata

Gartner Recommended Reading

How to Use Data for Good to Impact Society

Modernize Your MDM Program With External Master Data Sharing

Dare to Dream! Give Your Data and Analytics Initiatives a Purposeful Mission to Improve the World

Magic Quadrant for Data and Analytics Service Providers

Smart Data Sharing — Five Insights to Get It Right

Augmented Data Quality

Analysis By: Melody Chien

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Early mainstream

Definition:

Augmented data quality represents an enhanced capability to evolve data quality processes — for improved insight discovery, next-best-action suggestions and accuracy — through the use of metadata, knowledge graphs and AI-related technologies. This capability exists in data quality solutions and is aimed at increasing efficiency and productivity by automating process workflows, minimizing dependency on humans and reducing time to value by means of data quality improvement.

Why This Is Important

Data quality is a key enabler for digital business initiatives. As organizations accelerate their pace in the digital transformation journey, the challenge of managing trusted, high-quality data at scale has increasingly become a limiting factor. With rapid growth of distributed data landscapes, the diversity of data and the number of new business requirements, augmented data quality technologies facilitate and even automate manually intensive data quality processes.

Business Impact

Data quality vendors build AI models to improve discovery of data characteristics, suggest next best actions and automate data quality processes by using metadata, reference data, application logs, users' actions, best practices and AI algorithms. The discovery of certain data patterns or outliers can inspire corrective actions. Automation helps organizations solve complex data quality requirements quickly and effectively. It also leads to higher productivity, greater accuracy and quicker ROI.

Drivers

- Traditional data quality applications provide tooling for common data quality practices such as profiling, matching, cleansing and monitoring. These applications, however, depend to a large degree on SMEs to troubleshoot and remediate data quality problems. Complex and exception-prone issues are difficult to solve with existing practices and technologies.
- Augmented data quality provides transformational means of enabling organizations to process data quality tasks with deeper data insights, next-best-action suggestions and higher degree of automation.
- Augmented data quality aids discovery and classification of sensitive PII data, pattern detection and correlation identification among data entities.
- By using active learning and collective knowledge, augmented data quality suggests matching proposals. It also proposes mapping of data quality rules to data elements based on previous user actions and data similarities, and automatically infers, creates or curates data quality rules to apply fixes to data following patterns previously identified.
- As underlying technologies (ML, NLP, active metadata, knowledge graphs, predictive analytics) mature over time and become more widely adopted, we expect augmented data quality support broadening to the entire spectrum of data quality tasks.

Obstacles

- Trustworthiness of AI models: The degree of accuracy depends on the accuracy and consistency of the metadata controlling the process and the data used to train models. “Data drift” may occur over time and affect supervised models.
- Inclusion of data and analytics governance: AI-driven automation enables users to be independent, but existing requirements for governance need to be embedded into AI models to avoid data-related risks.
- Requirement for experts to maintain AI models: Vendors produce augmented data quality differently using supervised or unsupervised techniques. Continuous model improvement may be required through active learning. Business users may need to learn how to “talk to the machine” to receive correct results.

User Recommendations

- Identify data quality use cases that could benefit from augmented data quality capabilities by focusing on solving specific data quality problems with well-defined business outcomes.
- Start with data quality problems that are currently tackled manually and are time-consuming or prone to exceptions.
- Explore the augmented data quality capabilities that are available in the market by investigating their features, upfront setup, required skills and possible constraints. Depending on vendors' technology maturity, it's very likely that some degree of custom development may be required to fully leverage the features.
- Assess incumbent data quality vendors' existing offering and future product roadmap for enhancement.
- Partner with business stakeholders to evaluate and monitor solutions supported with augmented data quality by checking for adherence to existing governance requirements and establishing metrics to show tangible benefits.

Sample Vendors

Ataccama; Collibra (OwlDQ); DQ Labs; IBM; Informatica; MIOsoft; Precisely; SAP; SAS; Talend

Gartner Recommended Reading

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

Predicts 2021: Data Management Solutions — Operational Efficiency Rises to the Top

Data Fabric

Analysis By: Ehtisham Zaidi, Robert Thanaraj, Mark Beyer

Benefit Rating: Transformational

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

A data fabric is an emerging data management design for attaining flexible and reusable data integration pipelines, services and semantics. A data fabric supports various operational and analytics use cases delivered across multiple deployment and orchestration platforms. Data fabrics support a combination of different data integration styles and leverage active metadata, knowledge graphs, semantics and ML to automate and enhance data integration design and delivery.

Why This Is Important

A data fabric leverages both traditional and emerging technologies in enterprise architectural design and evolution. It is composable and supports flexibility, scalability and extensibility in an infrastructure used by humans or machines across multiple data and analytics use cases. It abstracts data management infrastructure to disintermediate any incumbent platforms, and enables data integration and delivery regardless of the number of on-premises or CSP-based data assets in use.

Business Impact

Organizations benefit as data fabric:

- Provides insights to data engineers and ultimately automates repeatable tasks in data integration, quality, data delivery, access enablement and more.
- Adds semantic knowledge for context and meaning, and provides enriched data models.
- Evolves into a self-learning model that recognizes similar data content regardless of form and structure, enabling broader connectivity to new assets.
- Monitors data assets on allocated resources for optimization and cost control.

Drivers

- A data fabric enables tracking, auditing, monitoring, reporting and evaluating data use and utilization, and data analysis for content, values, veracity of data assets in a business unit, department or organization. This results in a trusted asset capability.

- Demand for rapid comprehension and adaptation of new data assets has risen sharply and continues to accelerate — regardless of the deployed structure and format. The data fabric provides an operational model that permits use cases, users and developers to identify when data experience varies from the data expectations depicted in system designs.

- A shortage of data management professionals is increasing the demand for accurate and actively utilized metadata to make system design, data availability and data trust decisions.

- Catalogs alone are insufficient in assisting with data self-service. Data fabrics capitalize on machine learning to resolve what has been a primarily human labor effort using metadata to provide recommendations for integration design and delivery.

- Business delivery and management professionals find it difficult to identify adjacent, parallel and complementary data assets to expand their analytical models. Data fabrics have the capability to assist with graph data modeling capabilities (which is useful to preserve the context of the data along with its complex relationships), and allow the business to enrich the models with agreed upon semantics.

- Significant growth in demand and utilization of knowledge graphs of linked data as well as ML algorithms to provide actionable recommendations and insights to developers and consumers of data can be supported in a data fabric.

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

Obstacles

Data fabrics are just past the Peak of Inflated Expectations. The main challenges surrounding broad adoption are:

- Diversity of skills and platforms to build a data fabric present both technical and cultural barriers. It requires a shift from data management based upon analysis, requirements and design to one of discovery, response and recommendation.

- Intentional market hype by providers and services organizations purporting a data fabric delivery is adding to market cynicism.

- Misunderstanding and lack of knowledge in how to reconcile and manage a data fabric and a legacy data and analytics governance program that assumes all data is equal will lead to failure.
- Proprietary metadata restrictions will hamper the data fabric, which is wholly dependent upon acquiring metadata from a wide variety of data management platforms. Without metadata, the fabric requires analytic and machine learning capabilities to infer missing metadata, and while possible, will be error prone.

User Recommendations

Data and analytics leaders looking to modernize their data management with a data fabric should:

- 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 understand and contribute to the catalog.
- Invest in data fabrics that can utilize knowledge graph constructs.
- Ensure subject matter expert support by selecting enabling technologies that allow them to enrich knowledge graphs with business semantics.
- Combine different data integration styles into your strategy (bulk/batch, message, virtualization, event, stream, replication and synchronization).
- Evaluate existing tools to determine the availability of three classes of metadata: design/run, administration/deployment and optimization/algorithmic metadata. Rate existing and candidate platforms and favor those that share the most metadata.
- Focus on a similar transparency and availability of metadata between PaaS and SaaS solutions.

Sample Vendors

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

Gartner Recommended Reading

Top Trends in Data and Analytics for 2021: Data Fabric Is the Foundation

What Is Data Fabric Design?

Top Trends in Data and Analytics for 2021: Data Fabric Is the Foundation

Emerging Technologies: Data Fabric Is the Future of Data Management

Digital Ethics

Analysis By: Pieter den Hamer, Frank Buytendijk, Svetlana Sicular, Bart Willemsen

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

Digital ethics comprise the systems of values and moral principles for the conduct of electronic interactions among people, organizations and things.

Why This Is Important

Digital ethics, and in particular privacy and bias, remain a growing concern. The voice of society and AI-specific ethical considerations are rapidly coming into focus for individuals, organizations and governments. People are increasingly aware that their personal information is valuable; they're frustrated by lack of transparency and continuing misuses and breaches. Organizations act to mitigate the risks involved in securing and managing personal data, and governments are implementing strict legislation in this area.

Business Impact

Digital ethics strengthens the organization's positive influence and reputation among customers, employees, partners and society. Areas of business impact include influencing innovation, product development, customer engagement, corporate strategy and go-to-market. Intention is key. If ethics is simply a way to achieve business performance, it leads to window dressing. The goal to be an ethical company serves all parties and society more broadly and leads to better business trust and performance.

Drivers

- Despite the hype around digital ethics, many organizations are still ignoring it. They think it doesn't apply to their industry or domain without giving it a deliberate consideration.
- Board members and other executives are sharing concerns about the unintended consequences that the innovative use of technology can have.
- There is frequent, high-profile press coverage of stories that concern the impact of data and technology on business and society more broadly.
- With the emergence of artificial intelligence, for the first time the ethical discussion is taking place before — and during — a technology's widespread implementation. AI ethics and other responsible AI steps are a foundation to reverse the negative popular sentiment around AI and lead to a more responsible use of its powers.
- Government commissions and industry consortia are actively developing guidelines for ethical use of AI. Examples include Ethical Framework for Artificial Intelligence In Colombia, New Artificial Intelligence Regulation in the EU and Using Artificial Intelligence and Algorithms in the U.S.
- Over the past year, a quickly growing number of organizations declared their AI ethics principles, frameworks and guidelines. They have a long way to go from declaration to execution, although some organizations already have digital ethics practices.
- Gartner predicts that by 2024, 30% of major organizations will use a new "voice of society" metric to act on societal issues and assess the impact on their business performance. The voice of society will put more pressure on governments and public and private organizations alike to ethically use technology. "Big tech" is already a negative stereotype in societal jargon.
- More universities across the globe are adding digital ethics courses and launching programs and centers to address ethical, policy and legal challenges posed by new technologies.

Obstacles

- Digital ethics is seen as a moving target because of confusion on what society expects. It might even lead to opposing the majority's opinion, based on an organization's position and beliefs.
- Digital ethics is too often reactive and narrowly interpreted as compliance, or confined to the technical support of privacy protection or viewed as explainable AI only.
- AI ethics is an emerging area in overall digital ethics. Early high-level guidelines are inconsistent and will evolve over time.
- The voice of society is a new metric where digital ethics should be present, but its weight is still to be understood. Insufficient attention leaves organizations exposed to lost business, higher costs and increased risk.
- Opinions differ across people, regions and cultures on what constitutes "good" and "bad." Even in organizations where ethics have been recognized as an important issue, consensus between internal and external stakeholders (such as customers) remains sometimes difficult to achieve.

User Recommendations

Business and IT leaders responsible for digital transformation in their organizations:

- Identify specific digital ethics issues and opportunities to turn awareness into action.
- Discuss ethical dilemmas from diverse points of moral reasoning. Ensure that the ethical consequences have been accounted for and that you are comfortable defending the use of that technology, including unintended negative outcomes.
- Elevate the conversation by focusing on digital ethics as a source of societal and business value, rather than simply focusing on compliance and risk. Link digital ethics to concrete business performance metrics.
- Ensure that digital ethics is leading and not following digital transformation. Address digital ethics early "by design" to move faster by knowing methods to resolve ethical dilemmas.
- Organize training in ethics and run workshops to create awareness within all AI initiatives about the importance that AI design and implementation require an ethical mindset and clear accountability.

Gartner Recommended Reading

Digital Ethics: From Compliance Duty to Competitive Differentiator

AI Ethics: Use 5 Common Guidelines as Your Starting Point

Every Executive Leader Should Challenge Their Teams on Digital Ethics

Digital Ethics by Design: A Framework for Better Digital Business

Data Ethics and COVID-19: Making the Right Decisions for Data Collection, Use and Sharing

Use Privacy to Build Trust and Personalize Customer Experiences

Data Hub Strategy

Analysis By: Ted Friedman, Andrew White

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

A data hub strategy effectively determines where, when and how data needs to be mediated and shared in the 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 — logical architectures that enable data sharing by connecting data producers (e.g., applications, processes and teams) with data consumers (e.g., other applications, processes and teams).

Why This Is Important

Digital business demands an emphasis on the connections among systems, people and things as a source of competitive advantage. Enterprises often meet this need by two approaches:

- Connecting applications and data sources via point-to-point interfaces
- Centralizing as much data as possible in a single system

Both approaches always become costly and inflexible. Executing a data hub strategy enables improved data sharing via more consistent, scalable and well-governed data flow.

Business Impact

A data hub strategy can capture benefits including:

- Increased operational effectiveness through consistent governance of data and analytics across sets of endpoints that need to share data
- Improvements in understanding and trust of critical data across process and organization boundaries
- Reductions in cost and complexity, compared with point-to-point integration
- Alignment of data and analytics initiatives focused on governance and sharing of critical data, such as master data management (MDM)

Drivers

Data and analytics leaders and their teams continue to expend significant resources to support and expand the reach and scale of various types of data flows across the enterprise. The interest in data hub strategy concepts and data hub architectures continues to grow as a result of:

- 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
- 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 changing governance approaches inside numerous endpoint systems
- The high cost, complexity and fragility of traditional architectures involving only point-to-point interfaces or centralized data stores
- Desire of many organizations to leverage the concepts of MDM programs toward governance and sharing of other types of data

- Increasing numbers of technology providers adopting data hub messaging and delivering product offerings that enable data hub architectures
- Better collaboration across business-oriented (governance) and IT-centrc (integration) roles concerned with delivering data to points of need across the enterprise

Obstacles

Although the benefits of a data hub strategy are compelling, many data and analytics leaders find challenges in realizing them, including:

- Substantial investment in traditional governance and integration approaches, which have created complexity and make change difficult and risky
- 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
- Over-reliance on technology and viewing governance and sharing of data as purely an implementation issue

User Recommendations

Data and analytics leaders and their teams should work with stakeholders to craft a data hub strategy that will align initiatives involving governance and sharing of critical data. Specifically, they must:

- Focus on the most high-value or complex areas, first to gain a significant business benefit impact through the initial deployment of data hubs.
- Design a data hub strategy to understand data and analytics governance and sharing requirements, and to drive integration efforts.
- Identify the data that is most frequently used or is most important, with most business value, and that requires effective governance and sharing.
- 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.

Gartner Recommended Reading

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

COVID-19 Health Risk Mitigation

Analysis By: Sam Grinter, Donna Medeiros

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

COVID-19 risk mitigation is made up of an array of functions designed to reduce the health risks COVID-19 poses. Functionality typically includes health screening, contact tracing, office booking, test data management and vaccine tracking. Additional functions such as corporate communications and a knowledge base may also be delivered.

Why This Is Important

Employers in many countries must comply with "duty of care" requirements to ensure a safe working environment. With more than 150 million confirmed COVID-19 cases and more than 3 million deaths at the time of this writing, the pandemic is the challenge of our time. Despite the development of multiple vaccines and improved therapies, it is likely we will be living with it for the foreseeable future due to differing vaccine rollout timelines and potential virus variants.

Business Impact

- COVID-19 risk mitigation aims to deliver a safer working environment for employees, their families, customers, partners and the wider community.
- COVID-19 risk mitigation can deliver the option for those working from home to safely return to the office.
- COVID-19 risk mitigation may be a compliance requirement (depending on country/state regulations), but also provides employees with the comfort and reassurance that their employer is acting with proper caution and responsibility.

Drivers

- There is a desire for employees currently working from home due to COVID-19 to have the option to return safely to the office at least part of the time. In Gartner's 2021 Digital Worker Experience Survey, employees were asked: "If you could make the schedule yourself, what proportion of time would you like to spend working?" Forty percent indicated they prefer to work from a combination of locations; 17% prefer to mostly work from the corporate office; 14% prefer to only work in the corporate office; 14% prefer to mostly work from home; 10% prefer to only work from home; and 5% prefer to only/mostly work from other locations (n = 9,725).
- COVID-19 risk mitigation technology enables employees to return safely to the office. It is anticipated that this option to return safely to the office will offer more flexible working options to employees and may also foster improved work quality and organizational culture.
- For those employees not able to work remotely, processes must be developed and maintained to ensure there is no unnecessary risk of infection while at work. Even when COVID-19 infections subside, as we have seen repeatedly, the infection rate can rise again very quickly. Furthermore, as the virus mutates, the nature of the risk changes. As such, the provision of COVID-19 risk mitigation technology for frontline and key workers not able to work from home must be continually monitored and adjusted when necessary to ensure safe working conditions for these employees.
- A longer-term driver for investment in such technology is the potential for future global pandemics and other health-related crises, which some epidemiologists have suggested is likely, as well as part of a flexible hybrid working strategy and for business continuity.

Obstacles

- A “set it and forget it” mindset will impair the ability for employers to adapt processes as the risks COVID-19 poses change. Many processes initially developed were self-built, with many organizations using spreadsheets or even paper. Such approaches may put employers at risk when it comes to data privacy regulations, and the tools in use do not deliver the agility needed as conditions change.
- Some employers and employees have not taken the matter as seriously as others, perhaps due to low local infection rates, low priority in investing in tools perceived as short term or other reasons. Such employers and employees are at severe risk.
- COVID-19 risk mitigation may not be universal to all types of workers and across multiple territories. Some workers have regular public contact, while others can perform their duties from home.
- Some countries will have higher/lower infection rates than others. This creates a complex, changing environment to develop and maintain an effective response.

User Recommendations

- Evaluate the current processes and tools in place to mitigate the health risks COVID-19 poses to ensure they are fit for purpose.
- Adapt processes and tools to ensure employees are safeguarded and that an appropriate response to the health risks is maintained as COVID-19 mutates and as infection rates rise and fall.
- Vet any manual processes initially deployed as a response to COVID-19 (Microsoft Excel, for example) for the utility as an effective/appropriate response to the pandemic, and the extent to which such processes are compliant with data privacy regulations. Replace any manually developed process with something off the shelf if the utility of the manual process is low and/or the compliance risk is moderate to high.

Sample Vendors

dev-id; Herta; Juvare; Qualtrics; Ramco Systems; Recovery Amped; ReturnSafe; SaferMe; Salesforce; ServiceNow

Gartner Recommended Reading

Plan for the Aftermath of COVID-19 for Your HCM Technology Portfolio

Peer Benchmarking on Health and Safety for a Postvaccine Return to the Workplace

The Business Continuity Management Software Ecosystem

Market Guide for Crisis/Emergency Management Solutions and COVID-19 Safe Return to Work

5 Board-Level Issues About COVID-19 That Executives Must Discuss With Directors

Tool: COVID-19 Long-Term Scenario Planning Workshop

Enterprise Metadata Management

Analysis By: Guido De Simoni

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

Enterprise metadata management (EMM) is a business discipline for governing shared metadata assets between and across analytics and operational projects and programs. For example, those for master data management (MDM), business intelligence (BI) and records management. The aim is to achieve the benefits of enterprise information management (EIM). EMM differs from project-specific metadata management that manages metadata for specific uses within that single project.

Why This Is Important

The need to align semantics across multiple data and analytics silos is the source of the increased demand for modern data management requirements. A Gartner survey indicates that 43% of the participants consider it difficult to identify data that delivers value, and 32% of them consider the management of new data sources as challenging. From hundreds of inquiries on metadata management, we see that these challenges are associated with a lack of a comprehensive metadata management approach.

Business Impact

- EMM enables coordinated efforts in data and analytics use cases. As a result, it provides an immediate benefit to business teams by allowing them to search, request and reuse governed metadata across siloed projects for their self-service use.
- EMM extends the benefits of other programs — such as those for analytics, MDM, data quality, data integration, business process management and service-oriented architecture — by supporting reconciled semantics in the information sources they use.

Drivers

- Most organizations cope with managing metadata within individual initiatives, that is, within the confines and needs of each data and analytics program, business initiative or system. For example, an MDM program, a BI initiative and a data warehouse implementation will include a specific metadata management focus. EMM supports the discipline of aligning and governing shared and common metadata among all such programs.
- Hype remains significant due to technological innovations, like augmented data catalogs that are sparking new interest in linking information silos to improve the value of information-based business outcomes. These innovations increase the need to govern information assets across multiple information management investments, which, in turn, creates fresh demand for EMM and EMM-enabled systems.

Obstacles

- When EMM is poorly planned, it can be prohibitively costly to implement technologies capable of managing the enterprise wide variety, volume, velocity and complexity of metadata about vital information assets.
- It is often poorly planned since organizations assume EMM is a wall-to-wall program; it is not meant to be. As with other informed and modern EIM efforts, not all metadata is equal, so not taking this view on board leads to bloated and costly programs that don't add business value.
- As a result, EMM is moving off the Peak of Inflated Expectations, but most organizations' adoption remains at an early phase. Concurrently, various technology innovations, while trying to fill EMM gaps, are disrupting the discipline's maturation, which means EMM will move slowly along the Hype Cycle.

User Recommendations

- Explore EMM when you have common corporate goals yet disparate information management programs (each with its own metadata) that are neither aligned nor sharing consistent information.
- Use EMM to govern the most important metadata and information assets between these discrete programs. EMM is valuable when your organization needs to incorporate its information management programs into a more mature enterprise information management framework.
- Grow the “connections” between the programs and datasets as needed, over time, if your goal is to align information across these metadata elements, and use EMM to govern the shared metadata.
- Adopt an EMM strategy to improve the situation by drawing on other planned initiatives, which may involve the participation of individuals from different organizational units.
- Account for people and process issues as well as technological issues and choices to create and sustain an EMM program.

Gartner Recommended Reading

Implement Enterprise Metadata Management to Drive Effective Enterprise Information Management

Vaccine Management D&A

Analysis By: Donna Medeiros, Rick Howard

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

Vaccine management provides data and analytics functionality and insight for data-driven decision making and operations support, taking into account security and risk. It enables the effective distribution, administration and surveillance of vaccines and immunization.

Why This Is Important

Vaccination on a global scale is necessary to end the COVID-19 pandemic. As pandemic waves continue and variants circulate, vaccine delivery offers a path to save lives and reopen economies. Information technology — vaccine management tools equipped with near-real-time data and analytics capabilities — provides the means to plan, coordinate and deliver mass vaccination programs. These systems provide for data-driven decision making and adaptive response to get doses to the people and places that need them.

Business Impact

Vaccine management plays a vital role in delivering on the promise of vaccines to save lives and safely reopen economies. Border closures and restrictions have had devastating economic consequences for millions of enterprises and workers. Vaccine management can provide end-to-end visibility of vaccine delivery via command center capabilities with analytics providing invaluable insights to target vaccine delivery via data sharing and generating artifacts such as consumer digital COVID-19 credentials.

Drivers

- There has been urgent pressure to reopen businesses and borders for workers, travelers and goods, and to lift restrictions to get commerce flowing again so the world can get back to some state of "new normal."
- Vaccine management burst onto the scene this year and is at the height of inflated expectations as the burgeoning solutions market has proven value in enabling the successful delivery of vaccines. Data and analytics systems and processes have been established to address immunization at a scale and pace not ever before attempted.
- Vaccine management encompasses (1) vaccine distribution — real-time visibility into vaccine delivery and the supply chain, includes shipment milestones via tracking data including traceability of supply and demand, cold chain status, quality assurance, storage, inventory management and data integration with supply chain stakeholders; (2) vaccine administration — command and control capabilities to manage the workflows, data flows and processes required for end-to-end administration of vaccines (e.g., registration, scheduling, tracking and monitoring, digital certificates, vaccine allocations and usage); and (3) vaccine monitoring and surveillance — collects person-level immunization to track individual vaccination status and population adherence.
- The need for public-private partnerships to address a complex set of interrelated issues, including health and safety; social and ethical; legal and risk; financial and operational; and consumer and employee experience is unprecedented.
- Vaccine management serves a risk-based approach to economic reopening by generating real-time data for decision making on the status of the vaccine rollout and for targeted vaccination campaigns.
- Consumer COVID-19 digital certificates that include vaccine, testing and recovery provide digital evidence of health status.
- Manual paper processes have been reduced, and registration and scheduling have been streamlined.
- Longer-term investments for future pandemic preparedness and other health-related crises have increased.

Obstacles

- Large and diverse sets of stakeholders that include pharmaceuticals, supply chain, government, technology, healthcare, businesses and various other privacy sectors pose complex governance challenges.
- Vaccine regulations and data privacy rules vary (local, national and international), making it difficult to stay aware of and navigate them.
- A new and dynamic vaccine management solutions market has evolved in the past year to meet urgent demand for vaccine delivery. Some systems lack maturity, and interoperability is a challenge with standards not well-specified by policy.
- Data sharing and exchange constraints lead to gaps in critical data.

User Recommendations

- Implement ongoing collaborative governance by business leaders and IT for requirements gathering, planning and coordination utilizing partnerships between government, employers, pharmacies, health departments, laboratories, technology providers and other relevant organizations.
- Stay current with regulations and consider data privacy and security implications.
- Establish an enterprise architecture of the necessary data assets with a holistic vaccine management data ecosystem architecture for now and the longer term.
- Fill gaps as feasible by leveraging vendors offering low- or no-code solutions.
- Seek out solutions with geointelligence capabilities that can help identify hard to reach, at-risk and not vaccinated populations.

Sample Vendors

Cardinality.ai; Epic; Esri; Everbridge; Guardtime; IBM; Qualtrics; Salesforce; ServiceNow; Tableau

Gartner Recommended Reading

Vaccine Management Resource Center

Sliding into the Trough

Graph Analytics

Analysis By: Mark Beyer, Rita Sallam, Jim Hare, Afraz Jaffri

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

Graph analytics techniques allow for the exploration of relationships between entities such as organizations, people or transactions. Graph analytics consist of models that determine the "connectedness" across data points. Graph analytics is typically portrayed via multicontext visualization for business user consumption.

Why This Is Important

- Graph analytics has proven value in specific use cases (disease tracking, supply tracing, crime prevention, anti-fraud, and more).
- Graph technology is now ready to expand into broader use cases that require path analysis, network coordination (of humans or machines), macro effects on direct market or services delivery, and microeffects on broader environments.
- The utilization of graph analytics is necessary in order to develop knowledge graphs, which are also accelerating in terms of market adoption.

Business Impact

Graph analytics:

- Analyzes underleveraged data for insights in complex connected data
- Is highly effective at assessing risk and responding to it to analyze fraud, route optimization, clustering, outlier detection, Markov chains, and more
- Identifies issues within an organization regarding liability and suggests proactive resolution
- Identifies peculiarly successful patterns in an organization

- Extends data discovery capabilities in modern business intelligence and analytics platforms

Drivers

Graph analysis is showing increased demand across all global regions, but not across all industry verticals:

- Graph analytics generally exhibits demand in 10% to 15% of the market.
- The COVID-19 pandemic has increased graph analytics over 90% in healthcare management, clinical research and healthcare supply chain use cases.
- Use cases that require analysis across highly complex models are developed and used within machine learning with the output stored in graph databases.
- Graph databases are ideal for storing, manipulating and analyzing the widely varied perspectives in the graph model due to their graph-specific processing languages, capabilities and computational power.
- Established AI techniques (such as Bayesian networks) are increasing the power of knowledge graphs and the usefulness of graph analytics through further nuance in representational power.
- Graph analytics also offer capabilities relative to contact tracing applications — showing significant advancement during the ongoing pandemic.

When graph analytics is used across data and metadata, metadata from unexpected sources adds to the graph analysis capabilities in the following ways:

- Certain combinatorial evaluations can build data “push” models that recommend new data assets to existing use cases by analyzing data access logs and analytical model development.
- Machine-enabled data profiling combined with graphs can evaluate brand new assets for similarities as compared to more familiar datasets — identifying certain characteristics of new data that are already aligned to AI techniques or ML features.
- Determines whether new and unfamiliar data is similar to training datasets already in use.

Specific industries are exhibiting adoption for vertical market requirements, and other use cases that span many industry verticals in a horizontal fashion are seeing early to moderate levels of adoption, such as:

- Law enforcement, epidemiology, genome research, anti-money laundering.
- Route optimization, market basket analysis, fraud detection, social network analysis or location intelligence.

Obstacles

- Graph analytics and the closely related graph databases are driving a demand for new skills related to graph-specific knowledge, which may limit growth in adoption. Some vendors have created graph analytic solutions that make it possible to execute graph analytics using SQL.
- New skills required include knowledge and experience with the Resource Description Framework (RDF), property graphs, SPARQL Protocol and RDF Query Language (SPARQL), as well as executing graph analysis in Python and R.

User Recommendations

- Test graph analytics to address use cases that exhibit development, coding and data models that are overly complex using traditional SQL-based queries and visualizations.
- Consider graph analytics to enhance pattern analysis — especially in the verticals and use cases noted above.
- Transition metadata analytics from simple catalog search and discovery into a graph analysis model to identify user communities that conduct statistical and logical processes that are applied to shared datasets.
- Implement interactive user interfaces with the graph elements to find insights and analytic results, and store the outputs/results for repeated use in a graph database.
- Train existing personnel how to align data assets, statistical processes, algorithms to create training datasets and building identification processes to detect data changes that will drive changes in the analytical models.

Sample Vendors

Cambridge Semantics; Digital Reasoning; Elastic; Maana; Siren; SynerScope

Gartner Recommended Reading

Graph Technology Applications and Use Cases

Connecting the Dots: Why Graph Analytics Are Key to Understanding Human and Machine Misbehavior

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

Augmented Analytics

Analysis By: Rita Sallam

Benefit Rating: Transformational

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

Augmented analytics uses AI and ML techniques to automate data preparation, insight discovery, data science, and machine learning model development and insight sharing for a broad range of business users, operational workers and citizen data scientists.

Why This Is Important

Many of the activities associated with preparing data, finding patterns in data, building models on complex combinations of data, and sharing insights with others, remain highly manual. This limits user adoption and potential business impact.

Business Impact

- Augmented analytics is transforming how and where users interact with analytics content as it has become a core component of most analytics and BI and data science platforms.
- Insights from advanced analytics — once available only to skilled analysts, citizen data scientists and data science specialists — are now in the hands of business analysts and a broad range of decision makers and operational workers across the enterprise — the augmented consumer — driving new sources of business value.

Drivers

- Organizations increasingly want to analyze more complex datasets combining diverse data from across the enterprise as well as from external sources. With an increasing number of variables to explore in data harmonized from many diverse datasets, it is practically impossible for users to explore every possible pattern combination, and even more difficult to determine whether their findings are the most relevant, significant and actionable. Expanding use of augmented analytics will reduce the time users spend on exploring data, while giving them more time to act on the most relevant insights.
- Augmented analytics capabilities are increasingly mainstream features of data preparation, analytics and BI platforms and data science and machine learning tools. They are also being embedded in enterprise applications and domain and industry specific solutions. This is delivering insights most relevant to a broad set of application users to improve decision-making and actions.
- Dynamic data stories are an example of a combination of augmented analytics features used to automate insights. This combines augmented analytics with natural language query (NLQ), natural language generation (NLG) and anomaly detection into dynamically generated data stories delivered to users in their context. This type of user experience will reduce the use of predefined dashboards for monitoring and analysis and increase the use of augmented analytics.

Obstacles

- Trust in autogenerated models. Organizations must ensure that the augmented approach is transparent and auditable for accuracy and bias, and that there is a process to review and certify analyses created.
- Training. With more automation comes greater user responsibility and the need for more, but different user training.
- Collaboration. Establishing a collaborative environment, pairing expert data scientists with nonexperts across the analytic life cycle will be essential to capitalize on the skills of all parties.
- User outreach. Using augmented analytics not only to support new and less expert analytic users, but also to shorten time to insight for more expert users.
- Ecosystem. It will be critical to build an ecosystem that includes not only tools but also data, people and processes to support the use of augmented analytics.

User Recommendations

Data and analytics leaders looking to make analytics more pervasive should:

- Identify the personas that will benefit most from augmented analytics capabilities.
- Ensure users can get value from new augmented analytics features by providing targeted and context-specific training. Invest in data literacy to ensure responsible adoption.
- Focus on explainability as a key feature to build trust in autogenerated models.
- Assess the augmented analytics capabilities and roadmaps of analytics and BI, data science, data preparation platforms, and startups as they mature. Look for upfront setup and data preparation required, the types of data and range of algorithms supported, integration with existing tools, explainability of models and the accuracy of the findings. Also evaluate emerging dynamic data storytelling capabilities.
- Provide incentives for citizen data scientists to collaborate with, and be coached by, specialist data scientists who still need to validate models, findings and applications.

Sample Vendors

Microsoft (Power BI); Oracle (Analytics Cloud); Qlik; SAP Analytics Cloud; SAS; Tableau; Tellius; ThoughtSpot; AnswerRocket; Igenius; Conversight.ai; TIBCO Spotfire

Gartner Recommended Reading

Magic Quadrant for Analytics and Business Intelligence Platforms

Critical Capabilities for Analytics and Business Intelligence Platforms

Tool: Visual Guide to Analytics and Business Intelligence Platform Capabilities

Top Trends in Data and Analytics for 2021

How Augmented Analytics Will Transform Your Organization: A Gartner Trend Insight Report

Intelligent Applications

Analysis By: Alys Woodward

Benefit Rating: Transformational

Market Penetration: 5% to 20% of target audience

Maturity: Emerging

Definition:

Intelligent applications (IAs) are enterprise business applications with embedded or integrated AI technologies, such as intelligent automation, data-driven insights and guided recommendations, that can deliver a more personalized interface, improve productivity and support decision making.

Why This Is Important

AI is the next major battleground for enterprise applications with many technology providers now incorporating AI and ML capabilities in their products. Enterprise application vendors are actively embedding AI technologies within their offerings — from ERP to CRM and HCM to productivity applications. Adding intelligence into applications, instead of more procedural features, allows applications to support decision-making processes alongside transactional processes.

Business Impact

Intelligent applications transform many core business processes by:

- Improving user experience through significant degrees of personalization and new interaction channels
- Creating efficiencies through automation
- Improving decision making through intelligent decision support within the application
- Scaling the ability to capture business value in core business processes
- Increasing the ability to work with many ecosystem partners to deliver value
- Decreasing risk through fraud or other risk detection

Drivers

- Organizations are demanding more and more functionalities from applications, whether built or bought, expecting them to enhance current processes for both transactions and decision-making with recommendations, insights and additional information.
- The trend toward composable business architectures has highlighted the possibilities around delivering advanced and flexible capabilities to support, augment and automate decisions. This has increased customers' appetite for intelligent augmentation within applications. However, to build this requires an underlying data fabric and packaged business capabilities.
- AI capabilities and features are increasingly being integrated into ERP, CRM, supply chain and knowledge management software within enterprise application suites.
- Embedded intelligence is typically a part of a core enterprise application, not a separate technology product. From the provider's perspective, including AI helps improve customer loyalty and customer lifetime value by extending and enhancing their applications with AI, rather than a separate revenue stream.
- The most important enabler for intelligent applications is ML, which allows features like recommendations, insights and personalization. Conversational UIs, while a part of AI, are more about improving the interface and less about adding intelligence to the application. As intelligent virtual assistants become more widely incorporated by application platforms, the line between interface and intelligence will blur.
- Emerging vendors in a number of enterprise domains are designed with AI first (known as AI-native applications), competing directly with suite-based incumbents in those same domains.
- Intelligent applications were the most popular technology in a survey of organizations implementing AI, and almost one-fifth of organizations surveyed intended to adopt intelligent applications in the next 12 months.

Obstacles

- **Lack of necessary data —** Intelligent applications require access to data from a range of systems, meaning application vendors need to think about data management technology and processes outside their own solutions.
- **Adding AI to the existing applications is often complex —** The models have to be trained and maintained, and the user has to understand the latency of the data. Insights generated with AI have to be contextualized for the process that the user is executing, which requires additional business metadata.
- **Overuse of "AI" in marketing —** Vendors have a tendency to overuse the term "AI" in marketing and neglect the focus on business impact, which can generate a cynical response in business buyers.
- **Trust in system-generated insights —** It takes time for business users to see the benefit and trust such insights, and they need some understanding of how the decision was taken in terms of inputs and logic.

User Recommendations

- Challenge your packaged software providers to outline in their product roadmaps how they are incorporating AI to add business value in the form of a range of AI technologies.
- Evaluate the architecture of your providers by considering that the best-in-class intelligent applications are built from the ground up to be constantly collecting data from other systems, with a solid data layer in the form of a data fabric.
- Prioritize investments in highly specialized and domain-specific intelligent applications delivered as individual point solutions, which help solve problem areas such as customer engagement and service, talent acquisition, collaboration, engagement and more.
- Bring AI components into your composable enterprise thinking to innovate faster and safer, to reduce costs, and to lay the foundation for business-IT partnerships. Remain aware of what makes AI different, particularly how to refresh and rebuild machine learning models, as this can cause implementation and usage challenges.

Sample Vendors

BizMerlinHR; Eightfold AI; JAGGAER; Salesforce; Sievo

Gartner Recommended Reading

Emerging Technologies: AI Survey Indicates That Vendors Must Take a Long-Term View

Predicts 2021: Artificial Intelligence in Enterprise Applications

Top Trends in Data and Analytics for 2021: Composable Data and Analytics

Information Architecture

Analysis By: Guido De Simoni

Benefit Rating: High

Market Penetration: More than 50% of target audience

Maturity: Early mainstream

Definition:

Information architecture (IA) describes the current/future state and guidance necessary to share and exchange data assets. Accomplished through requirements, principles and models, IA also formalizes the technology capabilities needed to analyze and organize data needed to deliver business value.

Why This Is Important

Information is the lifeblood of digital business. As pervasive sensing approaches drive the instrumentation to collect, share and develop insights into all facets of business activities, stand-alone information architectures are insufficient for the emerging data economy. Mastery of information architecture practices creates the potential to alter the competitive landscape and provide increased business insight.

Business Impact

- Understanding customer buying habits, purchasing behaviors, churn and missed opportunities can enable enhanced personalization and targeted promotions, along with service and product catalog refinements.
- Strategic decision making. Visibility across the enterprise to make important decisions, which requires investments in common data models and governance.
- Enabling innovation. Data presents an invaluable opportunity for firms to innovate, but only if they know what to do with it.

Drivers

- Stand-alone information architectures are insufficient for the emerging data economy. The challenge is how to plan, design and implement information-sharing environments, given a large number of information silos and the difficulties teams have in coordinating activities enterprisewide.
- Information architecture practices support the continuous analysis of requirements and enable significant assessment for the evolution of data and analytics capabilities map.

Obstacles

- Although there are case studies outlining the benefits of information architecture, those who pursue this approach face significant challenges and will take longer to reach maturity, although the best practices to achieve it may be well-known.
- The two challenges are that the supported analytics a) exclude unstructured information, and b) they cannot provide in-the-moment insight supporting intelligent responses to emerging situations or contextual versions of the truth.
- The penetration is higher than should justify its positioning, because these points have not been resolved. As such, we have positioned it at slow movement after peak-trough midpoint.

User Recommendations

- Map the information architecture to business strategy via business capability models.
- Leverage tools such as business capability modeling to understand the information impacts to the organization's critical business imperatives.
- Use quantifiable metrics linked to business key performance indicators (KPIs), whenever possible.
- Treat information as a strategic asset. Information warrants its own strategy to ensure that its economic benefits are fully maximized.
- Begin indirect and direct data monetization, making metadata and master data essential for business outcomes to capture value and minimize risk.

Sample Vendors

Caserta, Deloitte, Evolytics, ExistBI, TCS

Gartner Recommended Reading

Achieving the Business Value of Data and Analytics

Tool: Sample Job Description for the Role of Information Architect

DaaS

Analysis By: Mark Beyer, Ehtisham Zaidi

Benefit Rating: Moderate

Market Penetration: 20% to 50% of target audience

Maturity: Adolescent

Definition:

Data as a service (DaaS) is an architectural and design approach that delivers data on demand via consistent, prebuilt access with the aid of standard processing and connectivity protocols (e.g., APIs), and is one possible mode of delivery in the evolving data fabric (along with data virtualization, semantic tiers, bulk/batch, replication and others). Originating data resides on its current platform and is presented as output for use in a subsequent process or delivery endpoint.

Why This Is Important

DaaS will be subsumed by data integration or data fabric. DaaS enterprise is independent data access and management processes in a services-enabled approach. DaaS brokers provide free or fee-based aggregators utilize DaaS to deliver data products to clients. Marketplaces that sell data products use DaaS technology but are not DaaS. Data exchanges using DaaS are membership based consortiums that contribute, align and aggregate data for data sharing models.

Business Impact

In general, DaaS enables users to have faster access data:

- Enabling faster time to value for new business applications and new developers or consultants to deploy applications more quickly.

- Complementing complex hub-and-spoke data warehouses to eliminate data silos by federating existing structures.
- Consolidates the cost and delivery of multiple data integration efforts into a reusable infrastructure.
- Enables a monetized model for sharing data via marketplaces or open data hosting.

Drivers

The vendors and suppliers sometimes referring to DaaS are healthy and could easily migrate to participate in the data integration, the data fabric or the active metadata management practice areas. At this time, many have already changed their position away from DaaS. This is not only a correct change in positioning for those suppliers, but is highly beneficial to them, their customers and the data management markets across the spectrum of tools required.

It is extremely important to understand this market is largely a contrived practice area with multiple interpretations.

While DaaS is considered broadly in the market, it often has multiple definitions. Each of the widely divergent uses of the DaaS term have different drivers.

- DaaS enterprise offers a services-enabled approach that can help break the often very tight coupling between application design and data management in individual applications.
- DaaS brokers will persist in the delivery of data products to clients and remain a thriving market that demands curated data assets from expert integrators with built-in data trust models. These offerings may even continue to use the DaaS designation in their branding and go-to-market plans.
- Data marketplaces (as DaaS) will also increase in demand as organizations seek to augment their own data with otherwise unavailable data and will be willing to pay for new data assets. Marketplaces are not DaaS.
- Data exchanges (as DaaS) will continue to rise as vertical industries and horizontal practices (such as international accounting practices) — creating and adhering to collective data quality standards.

The DaaS architectural approach will not reach the Plateau of Productivity but will be replaced by the more advanced data fabric, which will use this approach as one option. Further, most of the example vendors already involved in this approach will expand to the broader data fabric.

Obstacles

- DaaS continues as a confused term with even more inappropriate synonym usage. The DaaS architecture will become obsolete before plateau and be absorbed by "data fabric." Data brokers, data marketplaces and data exchanges will evolve to the data fabric approach.
- Assertions that mechanical steps in data integration could be automated and repeated by leveraging machine learning over metadata carried DaaS over the peak and it is now heading into the Trough of Disillusionment. The difficulty of building DaaS in-house (the solution) has increased interest in data from more mature brokers, marketplaces and exchanges.
- The ingest and output sides of DaaS will be complex. All data is available at different levels of access and different speeds of acquisition, and refresh. Data is often sourced from disparate governance frameworks that need to be resolved, with varying levels of detail and in various integrated and nonintegrated models.

User Recommendations

- Design DaaS environments to provision data that is integrated, accurate, enriched, described and available for access. DaaS contrasts with self-service data integration. Users can acquire additional data using DaaS, and then integrate it themselves — with more data or not — as their needs require.
- Ensure that data sources made available are well-described and well-documented, are locatable, and that security, privacy and quality controls are well-embedded.
- Consider DaaS-style architecture as one option to temporarily expand and complement the existing data management strategy and infrastructure as they progress toward the wider data fabric.

Sample Vendors

Denodo; Domo; Intenda (Fraxses); Nexla; StreamSets; Zeenea; Zetaris

Gartner Recommended Reading

Emerging Technologies: Data Fabric Is the Future of Data Management

Use Gartner's Reference Model to Deliver Intelligent Composable Business Applications

Data Engineering

Analysis By: Robert Thanaraj, Ehtisham Zaidi

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Early mainstream

Definition:

Data engineering is the discipline of reconstituting data into usable forms, such as composable data and analytics applications, by building and operationalizing data pipelines across various data and analytics platforms. It's a combination of three different practices: data management, software engineering, and infrastructure and operations.

Why This Is Important

The essence of data engineering is translating data into readily consumable forms for the purpose of delivering targeted consumer experiences. It's an emerging data management discipline that enables the creation, operationalization and maintenance of data pipelines aimed at delivering integrated data to data consumers, for their data and analytics use cases, across heterogeneous environments, despite their infrastructure choices.

Business Impact

Organizations with a mature data engineering practice will experience benefits such as:

- Faster time to delivery when adding new data to existing analytics and data science models
- The ability to incorporate third-party data more quickly than their peers
- Easier fulfillment of regulatory requirements to meet data transparency expectations
- Business teams that are empowered with composable data and analytic applications

Drivers

- Data management teams spend more time on data preparation and data integration, and as a result, these are the primary candidates for automation.
- To relieve bottlenecks and barriers to delivering data and analytics solutions, organizations need to change the way they work, such as introducing DataOps and agile practices.
- Organizations seek successful consumer experiences, which is the last mile in the “data-insights-decisions” continuum.
- Cloud adoption is a major driver for data engineering, where the use of open-source and homegrown applications increases (following best-fit approach).
- Organizations are bound to fail if they launch data science initiatives without onboarding the necessary data engineering skills.
- Business teams seek increased autonomy, but their data maturity levels vary significantly. They need guardrails and established best practices to follow.

Obstacles

- Lack of skilled data engineers in the market is hurting organizations the most. Organizations turn to data and analytics service providers or upskill related roles, like ETL developers.
- Unicorn data engineers do not exist. Many think a data engineer can “do it all,” catering to the full spectrum of data engineering that includes data management, software engineering, and infrastructure and operations. Data engineering is a team competency.
- Legacy baggage around poor integration and operations practices hurt and/or delay data engineering practice adoption.

User Recommendations

- Catalog an inventory of data assets and make them searchable.
- Leverage metadata to drive automation of data pipelines and related artifacts. Study the data usage and utilization patterns of users and systems and employ this metadata to improve efficiency and optimize delivery.
- Evaluate your success measures on a regular basis. Examples include time to market, productivity, CI/CD automation of data pipelines, code quality and cost-efficiency of build and operations.
- Introduce targeted use-case-specific tooling to accelerate data pipeline builds and operations. Examples include data warehouse automation tools and data preparation tools.
- Create data engineers by upskilling your ETL developers, data analysts or similar roles. Train them on software engineering, DevOps tooling and product development.
- Expand the composition of your data engineering team by adding new roles like test engineers, automation engineers and infrastructure specialists.

Sample Vendors

Ascend.io; Databricks; Fishtown Analytics (DBT); Saagie; StreamSets; Unravel

Gartner Recommended Reading

How to Build a Data Engineering Practice That Delivers Great Consumer Experiences

Operational AI Requires Data Engineering, DataOps and Data-AI Role Alignment

Data and Analytics Essentials: DataOps

Market Guide for Data Preparation Tools

Magic Quadrant for DI Tools

Data and Analytics Stewardship

Analysis By: Guido De Simoni, Andrew White

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

Data and analytics stewardship is the analysis, management and control of the operational processes and data needed to enforce approved data and analytics governance policies and standards. Information in this context includes data, analytics, algorithms, documents, images and metadata — effectively, any and all data assets as needed.

Why This Is Important

Data and analytics stewardship enables better information behaviors to be achieved in the enterprise. Organizations with established data and analytics stewardship practices are equipped to evaluate and implement modern technology that supports operationalization and automation of data and analytics governance leveraging ML/AI.

Business Impact

- Data and analytics stewardship improves the level of trusted data for business operations, through adoption and enforcement of agreed-on data standards in the organization.

- Records management is similarly impacted by stewardship issues, where poor information handling and mismanagement of information classification and retention schedules can lead to increased costs, as well as greater exposure to risk and regulatory fines.

Drivers

- Recognizing that effective data and analytics governance and advocacy are critical for enterprise information management (EIM) programs, like master data management (MDM), application data management (ADM), or business intelligence and analytics, has resulted in wider (although still limited) acceptance of data and analytics stewardship.
- What data and analytics stewardship brings to a data and analytics governance initiative, when adopted correctly, is operational support in a day-to-day business context environment.
- The work of data and analytics stewardship is focused on problem solving, making it a critical business driver for continuous improvement when organizations drive strategic data and analytics programs.

Obstacles

- Despite the wider acceptance of information stewardship needs, many organizations have relied on the often reactive and heroic efforts of “citizen stewards” to solve data problems, holding outcomes and decisions back.
- Organizations are not yet, “en masse,” ready to invest the necessary time and money on the right solutions or training of their business users to deliver an operational function for stewardship. In the recent past, we have seen organizations maturing in this area still trying to shape data and analytics stewardship (even within business areas) by testing and validating the approach before committing to an established discipline.
- This has been largely affected by the lack of maturity in the overall discipline of data and analytics governance. The profile is moving slowly in the Hype Cycle.

User Recommendations

- If data stewardship exists in IT, movement of aligned roles within the jurisdiction of a business operational area should be investigated and appropriate action taken as the knowledge needed for business data work might not exist in IT.
- Where strategic programs such as MDM or compliance are underway, organizations should also commit to information stewardship that spans multiple business areas, and should potentially identify a lead information steward.
- Where a chief data officer is in place, the relationship with the business area information stewardship process should be made clear and the reporting lines for information stewards should be established for consistency with desired business outcomes. IT can execute the instructions and results of stewardship (for example, data maintenance or policy execution).
- Do not outsource the work of policy enforcement, because of the lack of context and limited business domain knowledge of the outsourcing partners.

Gartner Recommended Reading

What Are the Must-Have Roles for Data and Analytics?

Data Lakes

Analysis By: Philip Russom, Henry Cook

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

A data lake is a concept constituting a collection of storage instances of various data assets combined with one or more processing capabilities. Data assets are stored in a near-exact, or even exact, copy of the source format and in addition to the originating data stores. Structured and semistructured data may also be held.

Why This Is Important

Data lakes are important because they enable advanced analytics and complement traditional data warehouses. For example, the massive repository of source data that typifies a data lake supports broad, flexible and unbiased data exploration, which is required for data mining, statistics, machine learning and other analytics techniques. Furthermore, a data lake can provide scalable and high-performance data acquisition, landing and staging for data to be refined and loaded into a data warehouse.

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, structure-agnostic storage and processing options that advanced analytics require.

- **Data exploration 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. To achieve their productivity and discovery goals, each type of user needs massive volumes of broadly collected data that is in a condition suited to their skills and analytics techniques. A data lake, when designed properly, can provision data for the diverse exploration requirements of multiple user types and use cases.

- **Data warehouses continue to be relevant, but only when modernized.** Many legacy data warehouses were designed primarily for reporting, dashboards and OLAP. Instead of redesigning a warehouse to accommodate the massive stores of detailed source data that advanced analytics demands, many organizations prefer to build a data lake for advanced analytics. In these cases, the warehouse and lake are integrated by shared datasets, platform infrastructure (DBMS brands and storage, whether on-premises or cloud) and architecture components (data landing/staging). Hence, a data lake can modernize a data warehouse, to extend its investment, relevance and life cycle.

Obstacles

- **Data lake best practices are still evolving.** There is still much confusion about how to design and govern a data lake, as well as how to optimize a lake's data without losing its purpose as a repository for data science and advanced analytics. An emerging best practice is to design the internals of a data lake to include multiple 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).

- **Today's cloud data lake differs from the old Hadoop data lake.** The first data lakes were built on Hadoop, for data science only, and they lacked metadata, relational functionality and governance. If you build that kind of data lake today, it will fail. Today's data lake is on cloud, and it supports multiple analytics techniques (not just data science). For example, self-service data prep on a data lake requires business metadata, SQL for ad hoc queries and data curation.

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 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, customer 360, data warehousing and operational intelligence.

- 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 a data warehouse by extending it with an integrated data lake and/or a logical layer.

- 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); Cazena; ChaosSearch; Databricks; Dremio; Google Cloud Platform; 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

Metadata Is the Fish Finder in Data Lakes

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

Best Practices for Designing your Data Lake

Market Guide for Analytics Query Accelerators

Master Data Management

Analysis By: Sally Parker, Simon Walker

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Early 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 consistent and uniform set of identifiers and extended attributes that describes the core entities of an enterprise.

Why This Is Important

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

Business Impact

Trusted master data is a foundational requirement of digital business with a range of vested stakeholders across the organization. Leading organizations draw the causal link between master data and the business outcomes it supports across finance, sales, marketing and supply chain, as examples. Improvements include:

- Risk management and regulatory compliance
- Customer experience
- Cross-sell and upsell
- Supply chain optimization
- Accurate reporting
- End-to-end process optimization
- Reduced time to market

Drivers

Trusted master data is a foundational requirement of digital business:

- Organizations with complex or heterogeneous application and information landscapes typically suffer from inconsistent master data, which in turn weakens business-process integrity and outcomes. As a result, interest in MDM extends beyond the office of the CDO/CIO to business leaders across finance, marketing and supply chain who have drawn a causal link between trusted master data and the ability to optimize their business strategies.

- Organizations having invested in establishing a trusted enterprisewide view of their master data benefit from a greater agility to predict and respond to unexpected events — to pivot strategies in response to external factors such as COVID-19. Gartner inquiries on MDM rose 28% from March 2020 to December 2020 (n = 1,534) compared with the same period in 2019 as organizations scrambled to get their "data houses" in order.
- Although MDM is not a new concept, market penetration of MDM as a whole is led by North America, followed by Europe, then Asia/Pacific, with Latin America trailing.
- A prior hesitance to embark upon MDM initiatives due to complexity and cost is easing. This can be attributed to two contributing factors: increased recognition of the causal link between trusted master data and business agility/outcomes by a broader range of stakeholders; a lowering of the barrier to entry to adopt commercial MDM solutions. As the technological barrier to entry has lowered, the target audience has expanded beyond large enterprises with deep pockets.

Obstacles

In recent times technological barriers to MDM solutions have eased — but this addresses only part of the complexity.

- Slow to embrace cloud, the MDM solutions market has relatively recently shifted toward subscription pricing, cloud-based offerings and simpler (configure vs. code) products, which now contributes to a more approachable solution and shortening of deployment times.
- Technology alone is insufficient to solve a problem that traverses people, process and technology across the enterprise. Thus, MDM remains a complex and maturing undertaking.
- Successful MDM implementations require capabilities including business acumen, technical know-how, domain understanding and data governance. 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

If your business strategy depends on the consistency of data within your organization, you will likely consider MDM as an enabler of this strategy. MDM is leaving the Trough of Disillusionment as organizations better understand both the opportunity and the challenges — challenges many are often now unable to overcome without external guidance.

Organizations investigating MDM should:

- Approach MDM as a technology-enabled business initiative
- Secure executive sponsorship to facilitate cross-organizational collaboration.
- Ensure the causal link between the MDM initiative and the business outcomes it supports are clearly understood and articulated.
- Keep it lean and focused.
- Leverage third-party services to fast-track time to value. Over 90% 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

Magic Quadrant for Master Data Management Solutions

Critical Capabilities for Master Data Management Solutions

Three Essentials for Starting and Supporting Master Data Management

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

Climbing the Slope

Data Classification

Analysis By: Ravisha Chugh, Bart Willemsen, Bernard Woo

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Early 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, or the application of a “tag” to an object to facilitate the use and governance of the data, including application of controls during its life cycle.

Why This Is Important

Data classification enables an effective and efficient prioritization for data governance programs that span value, security, access, usage, privacy, storage, ethics, quality and retention. It is vital to security, privacy and data governance programs. It also allows organizations to have the required knowledge about the sensitivity of the data they process.

Business Impact

Data classification can be used to support a wide range of use cases; for example:

- Applying data security controls for example DLP and EDRM
- Privacy compliance
- 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

- The data classification approaches include categorization by data type, owner, regulation, or classification by data sensitivity or retention requirements. This enables organizations to focus security, privacy and analytics efforts primarily on their important datasets.
- When properly designed and executed, classification serves as one of the foundations supporting the ethical processing of data throughout the organization.

Obstacles

- Data classification initiatives have often failed in organizations because they are dependent on manual efforts by users with insufficient training involved in the process.
- Classification efforts often revolve around a security-centric mindset, which means the purposes are not explained to users using natural language, and results in low levels of engagement.
- Today many vendors provide automated classification products, which can offer more accurate results while minimizing user efforts. However, it is important to note that, while automatic classification tools can significantly improve the amount of data classified, they are not 100% accurate, especially if the tools have been created using ML/AI algorithms where models require ongoing training.

User Recommendations

- Data classification objectives can be difficult: To identify, tag and store all of an organization’s data, SRM leaders and chief data officers (CDOs) should collaboratively architect and use classification capabilities.
- Implement data classification as part of a data governance program.
- Use a combination of user-driven and automated data classification.
- Determine organizationwide classification use cases and efforts, and, at a minimum, keep all stakeholders informed.

- Combine privacy regulation adherence efforts with the security classification initiatives. As information can be categorized by nature (e.g., PII, PHI or PCI), or by type (e.g., contract, health record, invoice). Regardless, records should also be classified by risk categories as to indicate the need for confidentiality, integrity and availability. Finally, records can be indicated to serve specific purposes.

Sample Vendors

Dathena; HelpSystems; Informatica; Microsoft; Netwrix; Spirion; Varonis

Gartner Recommended Reading

Building Effective Data Classification and Handling Documents

Ignition Guide to Data Classification

How to Overcome Pitfalls in Data Classification Initiatives

Using Classification to Improve Unstructured Data Security

Self-Service Data and Analytics

Analysis By: Austin Kronz, Joao Tapadinhas, Sharat Menon, Alys Woodward

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

Self-service data and analytics refers to technology and processes that business users leverage with minimal involvement from IT departments. Enabled via low-code/no-code tools in areas such as analytics and business intelligence, data preparation and data catalogs, self-service is now moving into other areas of data and analytics as automation and augmentation impact all aspects of D&A.

Why This Is Important

Self-service is established in many organizations as an ideal best practice. It can be a way for organizations to create analytics prototypes and pilot them faster than relying solely on IT for analytics projects. However, organizations must recognize when prototypes require large or real-time data or the need for content to be accessed by many users, as IT support will likely still be required to help implement dedicated data integration pipelines or catalog data and analytics, for example.

Business Impact

Self-service data and analytics is critical to scaling the benefits of data-driven decision making. Emerging citizen analyst or citizen data scientist personas who understand the business context of the data are able to use powerful no-code/low-code analytic platforms to quickly discover insights.

Drivers

- Vendors are building self-service capabilities into their products, but need to make sure they match users' abilities, particularly at the less able end of the spectrum. Vendors have claimed self-service in analytics for decades, but self-service needs superior user interfaces and more data literacy programs to advance.
- As business users advance in terms of information requirements, they expect to be able to extend the use of self-service into data areas. Adding data sources to analytical environments, selecting data sources from data catalogs and integrating data sources from outside the organizations are all tasks that advanced business users (also referred to as power users or citizen developers) expect to be able to do.
- Budgets for purchasing analytics and BI tools are increasingly coming from business units and not just central IT/data teams.

Obstacles

- Governance over self-service tools is a common concern. Modern data and analytics platforms deployed without a plan around user enablement and training typically leads to governance challenges down the road. Aware of this possibility, many organizations overcorrect and do not take advantage of the power of self-service tools, overrestricting who can use them and what they have access to. Organizations need to achieve a balance of agility and control.

- The closer the self-service gets to the data; the greater involvement is required from the data engineering team. Data staff are less aware and understanding of the desire for self-service than analytics staff, because they are typically more concerned about governance and accuracy.

- High quality data is still a struggle for many organizations. Despite having powerful tools, poor data quality can lead to greater potential for misunderstanding or misuse of the data.

User Recommendations

- Segment your users by their ability and inclination to become self-servicing, and deliver to the most prepared users first. Success often snowballs and drives further successes, and aids in improving your D&A maturity over time.

- Evaluate data catalog and self-service data management to allow business users to add curated or external sources to their data landscapes.

- Form communities of self-service users and nonusers alike. Self-service should not be self-serving; communities where sharing, collaboration, education, project overviews and success evangelism become critical as analytics audiences grow.

- Build data literacy and certification programs to ensure users are best prepared to gain productivity improvements from self-service without mistakenly delivering bad or siloed information. Implement self-service for Mode 2 projects with dedicated staff members who have time to focus on learning how to use tools for data management, exploration and analysis.

Sample Vendors

Alteryx; DataRobot; Microsoft; Qlik; Tableau; Trifacta

Gartner Recommended Reading

How to Build a Data Engineering Practice That Delivers Great Consumer Experiences

How to Balance Control and Agility in Your Self-Service Analytics

Create a Hybrid Centralized and Decentralized Data and Analytics Organizational Model

Build a Comprehensive Ecosystem for Citizen Data Scientists to Drive Impactful Analytics

Market Guide for Data Preparation Tools

Tool: Evaluate Data Preparation Tools Across Key Capabilities

The 3 Pillars of Citizen-Driven Data Management

Entering the Plateau

Logical Data Warehouse

Analysis By: Henry Cook, Adam Ronthal

Benefit Rating: High

Market Penetration: More than 50% of target audience

Maturity: Early mainstream

Definition:

The logical data warehouse (LDW) is a best-practice analytics data management architecture design that combines multiple physical analytics engines into a logically integrated whole. Data and analytics leaders can use the LDW to cover the full range of modern analytic requirements with a logically unified system that provides a simple view of all their data without needing to copy the data.

Why This Is Important

The LDW architecture is a current best practice for analytical systems design. It enables users to accommodate a wide variety of user types, data types, data sources and analytical techniques: SQL, OLAP, Graph, Geospatial, machine learning, statistical and others. The LDW integrates the traditional data warehouse, data lake and other analytical systems into a cohesive whole. This allows users to be more agile and productive while meeting their demanding service levels.

Business Impact

By accommodating a wide variety of users, data and analytical processing, the LDW architecture enables organizations to maximize their return on investment in analytics. In addition, the modular LDW architecture builds in a variety of flexible choices that can address both current and future needs.

Drivers

- Modern analytics requirements need to support many types of data, analytical processing techniques, types and numbers of users, and service levels. Designers can meet these requirements by integrating multiple analytic servers and services using data virtualization, data transports and common metadata to achieve a single logical view of all data.

- The LDW enables enhanced enterprise agility and maximizes return on investment for both development and runtime. It does this by ensuring that there is a natural home for each requirement, in terms of data storage and processing. This minimizes the need to change the architecture to meet new requirements. This makes development and deployment easier and more productive.
- Using the right component for each requirement processing is done with maximum efficiency and minimum cost. This also contributes to maximum return on investment.

Obstacles

Architecture skills are needed when building a system based on the LDW architecture:

- The appropriate components need to be chosen and integrated, typically using common metadata, data virtualization and data transport mechanisms.
- Architects need to be able to identify the correct components and interfaces to meet functional, performance and scalability needs.

Obstacles to adoption have diminished recently due to the adoption of the principles of the LDW architecture by most vendors:

- It is now common for the data warehouse and data lake components of the LDW to be preintegrated.
- The architect's job is further aided by advancements in key enabling technologies such as active metadata, data management automation, and improved DBMS performance and scaling.
- Likewise, relevant best practices are now better understood for data architecture and logical modeling.

User Recommendations

Data and analytics leaders should:

- Adopt an LDW architectural approach — this has emerged as a best practice for data management in analytics environments.

- Expect to build the LDW incrementally. It is not necessary to build the entire system at once. All that is required is to anticipate likely future components and preposition the necessary interfaces.
- Use the LDW architecture to also resolve the tension between the need for agile experimentation and prototyping and the need to accommodate the more stringent acceptance criteria for more traditional querying and reporting.
- Leverage the LDW to address challenges that can be solved via a logical integration approach, such as distributed data, time-sensitive data, and volatile environments where sources, targets and datasets come and go frequently.

Sample Vendors

Amazon Web Services; Cloudera; Databricks; Denodo; IBM; Microsoft; Oracle; Snowflake; Teradata

Gartner Recommended Reading

Solve Your Data Challenges With the Data Management Infrastructure Model

The Practical Logical Data Warehouse

6 Things to Get Right for the Logical Data Warehouse

Market Guide for Analytics Query Accelerators

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

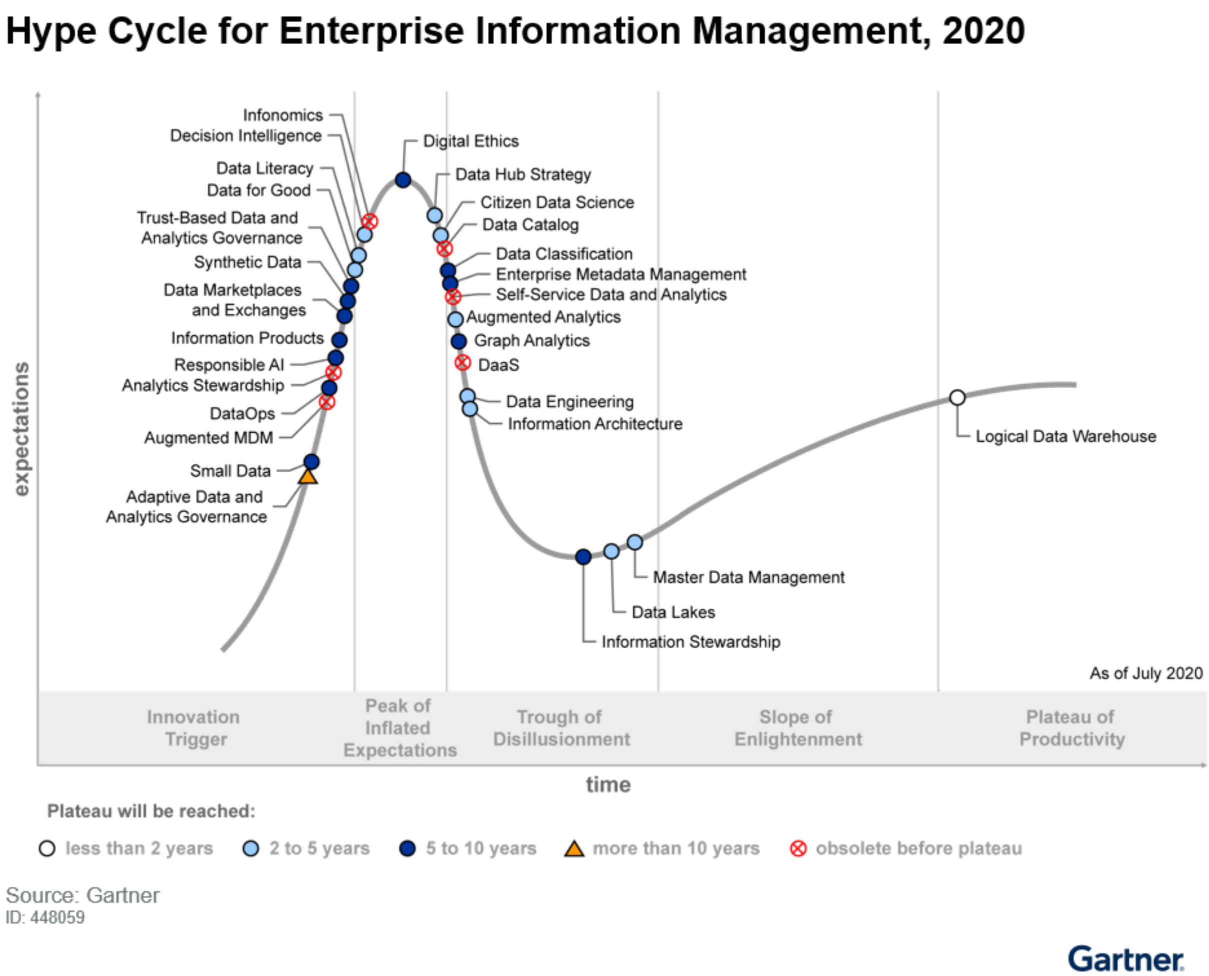
Organizing Your Teams for Modern Data and Analytics Deployment

Magic Quadrant for Cloud Database Management Systems

Critical Capabilities for Cloud Database Management Systems for Analytical Use Cases

Appendixes

Figure 2. Hype Cycle for Enterprise Information Management, 2020



Source: Gartner (July 2020)

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 (August 2021)

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 2021)

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 (August 2021)

Evidence

The 2021 Gartner View From the Board of Directors Survey: Survey Analysis: Executive Leaders Should Align to Board Priorities for 2021

Document Revision History

Hype Cycle for Enterprise Information Management, 2020 - 31 July 2020

Hype Cycle for Enterprise Information Management, 2019 - 25 July 2019

Hype Cycle for Enterprise Information Management, 2018 - 31 July 2018

Hype Cycle for Enterprise Information Management, 2017 - 3 August 2017

Hype Cycle for Enterprise Information Management, 2016 - 13 July 2016

Hype Cycle for Enterprise Information Management, 2015 - 17 July 2015

Hype Cycle for Enterprise Information Management, 2014 - 6 August 2014

Hype Cycle for Enterprise Information Management, 2013 - 9 August 2013

Hype Cycle for Enterprise Information Management, 2012 - 26 July 2012

Hype Cycle for Enterprise Information Management, 2011 - 29 July 2011

Hype Cycle for Enterprise Information Management, 2010 - 28 July 2010

Hype Cycle for Enterprise Information Management, 2009 - 27 July 2009

Recommended by the Authors

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

Understanding Gartner’s Hype Cycles

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

Predicts 2021: Operational AI Infrastructure and Enabling AI Orchestration Platforms

Top Trends in Data and Analytics for 2021: Data Fabric Is the Foundation

Tool: Communicating the Need for Data Literacy Improvement

Data Sharing Is a Business Necessity to Accelerate Digital Business

Coronavirus (COVID-19) Resource Center Primer for 2021

Vaccine Management Resource Center

Table 1: Priority Matrix for Enterprise Information Management, 2021

Benefit	*Years to Mainstream Adoption*			
↓	***Less Than 2 Years*** ↓	***2 - 5 Years*** ↓	***5 - 10 Years*** ↓	***More Than 10 Years*** ↓
Transformational	Health Pass	Augmented Analytics Decision Intelligence Intelligent Applications	Active Metadata Management Adaptive D&A Governance Data Fabric Digital Engineering Responsible AI	
High	COVID-19 Health Risk Mitigation Logical Data Warehouse Vaccine Management D&A	Augmented Data Quality Chief Data Scientist Data Classification Data Engineering Data Hub Strategy Data Literacy DataOps Information Architecture Master Data Management Self-Service Data and Analytics Synthetic Data	Data and Analytics Stewardship Data Marketplaces and Exchanges Digital Ethics Enterprise Metadata Management Graph Analytics Information Products Small and Wide Data Trust-Based Governance	Connected Governance
Moderate		Data & Analytics for Good Data Lakes		

Benefit	*Years to Mainstream Adoption*			
↓	***Less Than 2 Years*** ↓	***2 - 5 Years*** ↓	***5 - 10 Years*** ↓	***More Than 10 Years*** ↓
Low				

Source: Gartner (August 2021)

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 (August 2021)

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 2021)

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 (August 2021)