

## Hype Cycle for Data Management, 2021

Published 27 July 2021 - ID G00747543 - 141 min read

By Analyst(s): Philip Russom, Donald Feinberg

Initiatives: [Data Management Solutions](#)

Innovations continue to enter the field of data management. Their potential benefits are numerous, but challenging to understand and track. This Hype Cycle helps data and analytics leaders plan ahead and make informed decisions relative to emerging and maturing data management technologies.

### Additional Perspectives

- [Summary Translation + Localization: Hype Cycle for Data Management, 2021](#)  
(29 September 2021)

## Analysis

### What You Need to Know

Data and analytics leaders must cope with the evolving requirements of digital business, the deepening impact of cloud and increasing ecosystem complexity. Today, data is even more distributed than ever in multicloud, intercloud and hybrid architectures. As a result, the technologies that support these ecosystems and their integration must evolve. Furthermore, there is a need for new solutions that address current data management issues in innovative and unprecedented ways.

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

To succeed with data and analytics initiatives, enterprises need a holistic view of critical technology capabilities. For that purpose, Gartner provides several Hype Cycles that cover data management, analytics and related fields. Together, these Hype Cycles draw a holistic view of the technologies that can make data and analytics programs modern, broadly applied, richer in functionality and successful. In addition, all Hype Cycles support a common method for tracking the progress of these technologies.

These additional data and analytics Hype Cycles for 2021 will help data and analytics (D&A) leaders to form a holistic view of D&A:

- [Hype Cycle for Analytics and Business Intelligence, 2021](#)
- [Hype Cycle for Data Science and Machine Learning, 2021](#)
- [Hype Cycle for Data and Analytics Governance and Master Data Management, 2021](#)
- [Hype Cycle for Enterprise Information Management, 2021](#)
- [Hype Cycle for Artificial Intelligence, 2021](#)

### The Hype Cycle

A Hype Cycle quantifies maturity in two senses: feature richness and mainstream adoption. With those in mind, we can see various levels of maturity (and various kinds of change) in this year's edition of the Hype Cycle for Data Management. Here are some highlights.

## Changes in This Hype Cycle

**New entrants:** Innovation is alive and well in data management. The 2021 Hype Cycle for Data Management includes five new entrants, all entering in the Innovation Trigger area of the cycle. (See the left side of Figure 1.) Among these, Gartner anticipates that edge data management (see [Top Trends in Data and Analytics for 2021: Data and Analytics at the Edge](#)) will be fastest to mature, progressing to the Plateau of Productivity in less than two years. However, be forewarned that D&A governance platforms may take over 10 years for that same journey. Five to 10 years are projected for the other new entrants, namely intercloud data management, active metadata and the lakehouse.

**Climbing entrants:** Technologies that entered in recent years have progressed through the Innovation Trigger and are nearing the Peak of Inflated Expectations, including distributed transactional databases, augmented data quality, and the ledger database management systems (DBMS). DataOps – though increasingly a popular development method – has not yet reached the peak.

**Maximum hype:** At or near the Peak of Inflated Expectations are some of the most hyped (and most promising) technologies in data management today. These include cloud data ecosystems, private cloud database platform as a service (dbPaaS) and data fabric.

**Transition from hype to reality:** A number of “augmented” technologies are currently passing through the Trough of Disillusionment, namely augmented transactions, augmented data cataloging and metadata management, and augmented data management. Given that data management professionals need automation and interoperability – and the technologies are being built out by vendors and open-source communities – augmented technologies will likely climb the Slope of Enlightenment to attain maturity within two to five years. Ahead of them, data lake and master data management are poised to exit the trough in 2022.

## The Current State of Data Management Maturity

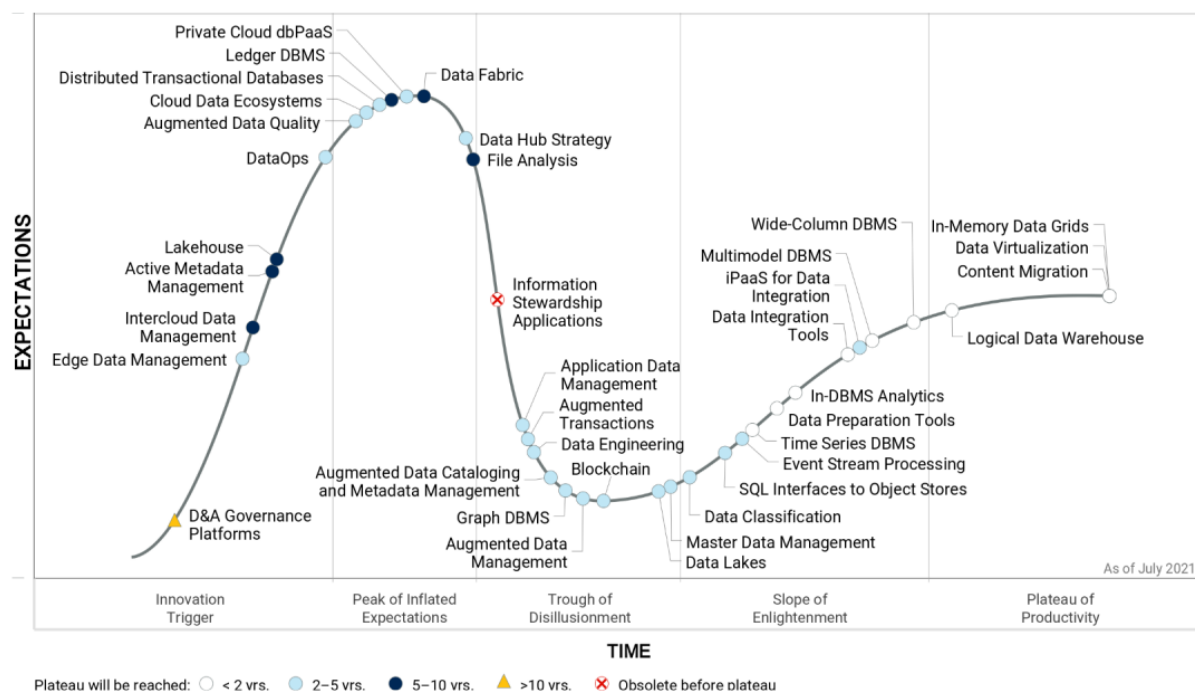
**Many hyped data management technologies are almost mature:** A prominent characteristic of the Hype Cycle for Data Management is that roughly half of its tracked technologies are lined up on the Slope of Enlightenment. Furthermore, most of these entrants will achieve full maturity on the Plateau of Productivity within two years.

This means that data and analytics leaders who are seeking new data management technology can adopt several emerging innovations, but with high business value and minimal technical risks, due to their relatively high maturity. These technologies are available for key tool, platform and practice areas within data management, as discussed below:

- **Database management systems (DBMS).** For many years now, one of the hottest areas for innovation in data management has been the DBMS. Many new DBMS types have emerged, and some of those are now approaching maturity via the Slope of Enlightenment on the Hype Cycle for Data Management. This includes wide-column DBMS, multimodel DBMS and time series DBMS. Also in this category are functions associated with DBMS, such as in-memory data grids, in-DBMS analytics and SQL interfaces to object stores. All those will attain full maturity within two years, whereas other DBMS will require five or more years for that journey, namely graph DBMS, private cloud dbPaaS, ledger DBMS and distributed transactional databases.
- **Data warehouse architectures.** The Gartner definition of the logical data warehouse (LDW) has become the most common approach to data warehouse architecture. This is because it flexes to accommodate many architectural variations, while still deploying its logical layer to unify even the most complex data warehouse environments. The design practices and DBMS platforms for the LDW concept are now quite mature. However, other major components of data warehouse and analytics architectures still have a way to go. These include the data lake (poised to exit the Trough of Disillusionment), data hub strategy (poised to enter the trough) and the lakehouse (a new entrant, still on the Innovation Trigger).
- **Data integration.** Another hot area for data management innovation has centered around tools and platforms for data integration and related disciplines. Many of these are on the cusp of maturity in this Hype Cycle, namely data virtualization, data integration tools, iPaaS for data integration, data preparation tools and event stream processing. In a related area, a new wave of metadata innovation has begun recently and will require five or more years for maturity. This involves augmented data cataloging and metadata management solutions (which includes active metadata), and these are required for the metadata-driven data fabric.

Figure 1: Hype Cycle for Data Management, 2021

## Hype Cycle for Data Management, 2021



Gartner

Source: Gartner (July 2021)

## Downloadable graphic: Hype Cycle for Data Management, 2021

## The Priority Matrix

The Priority Matrix for Data Management, 2021 positions each technology tracked in this Hype Cycle according to two dimensions — business benefit, and years to mainstream adoption and maturity. In the matrix, benefit is roughly quantified as transformational, high, medium or low.

**Transformational benefit:** The adoption of some hyped technologies could dramatically change how data management is performed, as well as transform organizations that depend on data management. For example, the adoption of event stream processing can extend data management beyond its stagnant focus on batch and bulk processing, thus transforming business practices that have been too slow to seize operational or competitive opportunities.

Similarly, data fabric promises to elevate data management out of its bucket of disconnected tools and siloed datasets to an unprecedented level of integration, interoperability and innovative business applications. Data fabric maturation will take five to 10 years, due to the enormous effort required of the vendor and open-source communities to integrate siloed tools and to enrich them with intelligent automation. For example, reengineering passive metadata management to become active metadata will be very difficult and time-consuming. However, data fabric's disruptive level of business transformation will be worth the wait.

**High benefit:** Many of the data management technologies tracked in this Hype Cycle promise a high level of business benefit upon their maturation. In particular, the many technologies described as “augmented” or “active” will:

- Speed up development
- Simplify maintenance
- Discover and track data assets with unprecedented breadth and accuracy
- Embed data management and analytics in a wide range of applications
- Intelligently automate just about every task performed in data management

Even better, most of these high-impact enhancements will be ready for low-risk adoption within two to five years. The logical data warehouse and in-memory data grids are mature now, and they promise high-impact benefits, so they should be considered for immediate adoption.

**Moderate benefit:** As discussed earlier, recent years have experienced numerous new brands of DBMS and types of DBMS. Some of these offer moderate benefits, and they will be on the Plateau of Productivity within two to five years, which makes them candidates for safe adoption. This includes multimodel, time series, wide-column, distributed and private cloud DBMS. Graph and ledger DBMS are in a similar position, as are database design best practices like the data lake and lakehouse.

**Table 1: Priority Matrix for Data 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		Blockchain Event Stream Processing	Active Metadata Management Data Fabric	
High	Data Integration Tools Data Preparation Tools In-DBMS Analytics In-Memory Data Grids Logical Data Warehouse	Augmented Data Cataloging and Metadata Management Augmented Data Management Augmented Data Quality Augmented Transactions Cloud Data Ecosystems Data Classification Data Engineering Data Hub Strategy DataOps Edge Data Management Graph DBMS IPaaS for Data Integration Master Data Management	Intercloud Data Management	D&A Governance Platforms
Moderate	Content Migration Data Virtualization Multimodel DBMS Time Series DBMS Wide-Column DBMS	Application Data Management Data Lakes Distributed Transactional Databases Private Cloud dbPaaS SQL Interfaces to Object Stores	File Analysis Lakehouse Ledger DBMS	
Low				

Source: Gartner (July 2021)

## Off the Hype Cycle

In 2021, a few innovations left the Hype Cycle for Data Management:

- **Fully mature innovations:** Operational In-Memory DBMS, Database Audit and Protection, Document DBMS and Apache Spark attained maturity on the Plateau of Productivity, and they have now moved off this Hype Cycle into mainstream adoption.
- **Obsolete before plateau:** SQL Interfaces to Hadoop has been subsumed into the broader profile SQL Interfaces to Object Stores.

- Data Catalog has been replaced by Augmented Data Cataloging and Metadata Management.



## On the Rise

### D&A Governance Platforms

Analysis By: Guido De Simoni

Benefit Rating: High

Market Penetration: Less than 1% of target audience

Maturity: Embryonic

#### Definition:

A data analytics governance platform represents a set of integrated technology capabilities that leverage automated data curation services (e.g., auto discovery, auto profile, auto data lineage, auto data quality analysis). It further exposes an extended user experience for decision management (including policy management) for all relevant participants (e.g., data steward, business analyst, LOB users, data scientist, and governance board members).

#### Why This Is Important

Modern D&A use cases need a portfolio of governance capabilities that cannot be fulfilled by existing, stand-alone products. This convergence is mandatory. Today the execution of data and analytics governance is inconsistent, with different organizations using different types of technology. This impedes the success of digital business initiatives. Governance needs have grown more diverse and complex; all aspects of governance for all types of policies can benefit from technology support.

#### Business Impact

- Emerging augmented data management services to match requirements across the enterprise data ecosystem.
- Analytics enablement for data at scale, critical and exposed linkages of data relationships, effective support of regulatory requirements.
- Enhanced productivity and efficiency in governance processes, more rigor in enforcement of policies, and therefore more control and trust in data and analytics.

#### Drivers

- Organizations want to have automated, synchronized, integrated, cost-effective and efficient solutions with a central design, yet a distributed deployment. This requirement is driven by the growing recognition that the work of data and analytics governance is different from the work of data management, but as such augmented data management supports the growth of these platforms of convergence.
- In particular, data and analytics governance, when implemented, provides a decision support system that leverages decision rights, accountabilities and behaviors for the valuation, creation, consumption and control of data and analytics.
- All these aspects are operationalized and more efficiency is gained when identification of data sources, curation of data, application of workflow, harmonization, and reporting and visualization are provided in a coherent platform and with automation. For example, you can address a platform auto generation of data quality rules from rule definitions and automated execution of data quality checks, AI-assisted data curation and association of business terms to technical artifacts, automated classification of sensitive data and build subject registry.

## Obstacles

- The current convergence across data management may or may not satisfy what is needed by organizations across data and analytics governance.
- The data management platforms we are talking about today may be different to the platforms that will emerge for data and analytics governance, to service different uses and use cases (leveraging control, outcomes, agility and autonomy in the required combination).
- Other obstacles reside in the cultural shift that many organizations must address in leveraging the inherent value of data and analytics governance. Such obstacles can jeopardize the adoption of these platforms not as yet as another piece of technology but as enablers for continuous improvement when committing to data and analytics initiatives aligned to critical mission priorities. At the time of writing, we estimate this innovation to reach the Plateau of Productivity in more than 10 years.

## User Recommendations

- Design proofs of concept that will capitalize on the required critical technology capabilities. Identify the relevance of these technologies and their connection to business outcomes as a first step. Then look into their ability to support specific use cases (such as, risk management and compliance).

- Minimize the number of tools and solutions deployed by analyzing your strategic approach to data and analytics governance and by using available market technology capabilities in end-to-end scenarios supported by emerging data and analytics governance platforms.

## Sample Vendors

Alex Solutions; Collibra; Global Data Excellence; IBM; Informatica; OneTrust

## Gartner Recommended Reading

[Implement Your Data and Analytics Governance Through 5 Pragmatic Steps](#)

[The Role of Technology in Data and Analytics Governance Policy Management](#)

## Edge Data Management

Analysis By: Ted Friedman, Bob Gill

**Benefit Rating:** High

**Market Penetration:** 1% to 5% of target audience

**Maturity:** Emerging

### Definition:

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

### Why This Is Important

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

## Business Impact

Edge data management creates value in various ways:

- By distributing data management to edge environments, data-centric solutions can enable more real-time value
- More solutions, such as for IoT use cases, must operate in disconnected (or intermittently connected and low-bandwidth) scenarios
- Enabling smarter physical assets and collections of assets, including autonomous behavior via onboard (edge) data
- Addressing inconsistencies and other governance issues that arise from siloed edge environments

## Drivers

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

## Obstacles

- **Management of distributed data architectures:** Data management has been largely based on principles of centralization — bringing data to central data stores (e.g., data warehouses), then processing that data to create value. Edge environments break that model via distributed data architectures, raising complex choices of where to locate and aggregate data on the continuum of cloud/data center to edge. This includes finding the right balance of latency and consistency. Poor choices will lead to failures in edge deployments.
- **Governance:** With the distribution and complexity of edge environments, data governance becomes challenging. Organizations should extend their governance practices to address edge-resident data stores and processing capabilities, including policies for disposal of ephemeral or nonvalue event data.
- **Skills:** Edge computing is an area of weakness for many organizations. This gap is intensified when the focus is on managing data on the edge — ensuring the reliable persistence, integration and governance on platforms outside the traditional data center or cloud environment.

## User Recommendations

To capture opportunities and minimize risk from edge data management, IT leaders and their teams should:

- Identify use cases where data management capabilities in edge environments can enable differentiated products and services by collaborating with OT and IT personnel working in edge locations.
- Expand the skill sets of your teams to include edge platforms and the technologies required to manage data and data-intensive workloads on them.
- Augment existing data management infrastructure to support edge deployment by partnering with product teams that are implementing IoT platforms and similar distributed computing architectures.
- Place a greater emphasis on end-to-end system design. Understanding the dependencies between all components of distributed data pipelines, analytic workloads and AI models will be crucial to success.
- Ensure safety and control by extending existing governance capabilities to apply to edge data environments.

## Sample Vendors

EdgeDB; FairCom; Microsoft; Xencia

## Gartner Recommended Reading

[Top Trends in Data and Analytics for 2021: Data and Analytics at the Edge](#)

[How to Overcome Four Major Challenges in Edge Computing](#)

[How to Enable Edge Data in Support of IoT Analytics](#)

[Forecast: Internet of Things, Endpoints and Communications, Worldwide, 2019-2029](#)

[Leading the Edge: Gartner's Initial Edge Hardware Infrastructure Forecast](#)

## Intercloud Data Management

Analysis By: Adam Ronthal, Donald Feinberg

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

### Definition:

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

### Why This Is Important

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

## Business Impact

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

## Drivers

Though actual market penetration remains minimal, there are several drivers for adoption, including:

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

Enterprises seeking to adopt intercloud data management approaches tend to be large, global enterprises with specific application requirements; most small to midsize organizations do not have these requirements.

## Obstacles

- Intercloud data management typically involves investment in highly specialized technologies from relatively niche vendors. This must be weighed against existing strategic relationships with large cloud service providers, which generally do not support intercloud capabilities broadly at this time. Because the supporting vendors tend to be specialized, most applications will need to be custom-built and written with specific intercloud vendor capabilities in mind. This adds risk because switching costs may be high and technically difficult.

- Any intercloud data management application will also be subject to the laws of physics. They will need to be aware of both performance implications and trade-offs in consistency and availability, and ensure that their applications are both aware of and designed with these trade-offs in mind.
- As with any data, intercloud data must be integrated in storage and/or logically to achieve maximum business value.

## User Recommendations

Weigh trade-offs in optimization and flexibility in your design decisions. There are three primary deployment options for intercloud data management:

- **Object store:** If data is distributed at the cloud object store (COS) layer, there is not a single service — native or third-party — that cannot read and write from the local COS. This enables diversity of choice in selecting a data management offering for the last-mile delivery.
- **DBMS:** Also called “distributed SQL.” The database management system distributes and manages data in a geodistributed cluster. Nodes can reside in multiple clouds and/or on-premises. This approach can support local, low-latency read/write apps with global read capabilities, and data sovereignty enforcements.
- **Application:** This can be thought of as an after-the-fact solution. Data may already be in multiple clouds and require a means of bringing it together. This approach essentially defers data integration to the point of consumption.

## Sample Vendors

Cockroach Labs; DataStax; Denodo; HPE; MariaDB; MongoDB; NuoDB; WANdisco; Yugabyte

## Gartner Recommended Reading

[The Future of Cloud Data Management Is Multicloud](#)

[6 Best Practices to Create a Cloud Management Services Offering in the World of Multicloud and Hybrid Cloud](#)

[Understanding Cloud Data Management Architectures: Hybrid Cloud, Multicloud and Intercloud](#)



## Data and Analytics Essentials: Cloud

### 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; Datalytx; 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](#)

## Lakehouse

**Analysis By:** Adam Ronthal, Donald Feinberg

**Benefit Rating:** Moderate

**Market Penetration:** Less than 1% of target audience

**Maturity:** Emerging

### Definition:

A “lakehouse” combines the semantic flexibility of a data lake with the production optimization and delivery of a data warehouse. It is a converged infrastructure environment supporting the full progression of data from raw, unrefined data, through the steps of refining that data, and ultimately delivering optimized data for consumption.

### Why This Is Important

Data and analytics leaders continue to struggle with getting value from their data lake initiatives. Using a data lake in conjunction with a data warehouse adds complexity to the data and analytics landscape. A lakehouse aims to unify the two architectures and improve efficiency while minimizing the need to move data and artificial intelligence (AI)/machine learning (ML) models between the two. A simplified environment with a lower operational footprint is the potential result.

## Business Impact

- Business will benefit from streamlined delivery, rapid access to data and a consolidated data management platform that can support a broad range of personas. This includes the highly skilled data scientist or data engineer, the business analyst, and even the casual user who consumes data through prebuilt, preoptimized dashboards.
- A lakehouse inherently provides a well-defined path to production for data science initiatives.

## Drivers

- Data and analytics leaders frequently struggle with getting data science projects into production. A lakehouse promises to unify the exploratory and production environments, thus mitigating these struggles.
- Enterprises consistently seek rapid and unencumbered access to data and struggle with the processes and perception of delayed delivery associated with the data warehouse. The lakehouse is often positioned as a “silver bullet” to solve this problem.
- Data lakes and data warehouses are optimized for different things. There is usually a trade-off between flexibility and optimization. When insufficient optimization is provided, the difference must be made up in skills. Lakehouses blur these lines, and enterprises may struggle to find the sweet spot for their use cases.
- Many cloud data warehouse solutions and almost all cloud data lakes already leverage semantically flexible cloud object storage as their storage of record. It is a natural progression to unify these environments, thus reducing the disparate and duplicate infrastructures.
- Most cloud data warehouses today do not have strong and integrated in-DBMS analytics capabilities to support data science, predictive modeling and machine learning capabilities. A lakehouse approach provides these.

## Obstacles

- The maturity of lakehouse solutions is still developing. Many of them do not support the full range of transaction consistency or robust workload management capabilities that data and analytics leaders have come to expect in their data warehouse solutions.
- The most complex data warehousing workloads are still likely to be beyond scope for most lakehouse solutions that do not incorporate already-developed functionality.
- Immaturity of users' ability to design, deploy and maintain complex data architectures is a more prominent obstacle than "solutions."
- The full scope of optimization includes data quality, security, governance and performance. Many lakehouse approaches do not address all of these.

## User Recommendations

- Employ a targeted use-case approach that solves specific problems and expands from there for long-term success. Early adopters should tread carefully to avoid the disappointment of "over promise and under deliver," which plagued prior attempts at solving the problem of production delivery of workloads on data lake infrastructure.
- Assess your performance line of "good enough" by running your most complex workloads on the evaluated target platform in a proof of concept (POC) to make decisions on when a lakehouse approach is sufficient and when a dedicated data warehouse may still be required.
- Consider a logical data warehouse (LDW) approach when addressing a broader data and analytics scope. A lakehouse is a subset of the LDW built opportunistically. The LDW remains a mature and best practice.
- Evaluate security and governance capabilities to ensure they meet your enterprise standards and requirements by establishing clear governance and security "must haves."

## Sample Vendors

Amazon Web Services; ChaosSearch; Databricks; Dremio; Incorta; Microsoft; RAW Labs; Starburst; Varada

## Gartner Recommended Reading

[Market Guide for Analytics Query Accelerators](#)

[The Practical Logical Data Warehouse](#)

[Best Practices for Designing Your Data Lake](#)

## 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

### 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](#)

## Cloud Data Ecosystems

Analysis By: Adam Ronthal

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

**Maturity:** Emerging

**Definition:**

Cloud data ecosystems provide a cohesive data management environment that ably supports the whole range of data workloads, from exploratory data science to production data warehousing. They have a common governance and metadata management framework, unified access management, and integrate augmented data management capabilities with a set of services accessible by the business user. Operational data sources also participate in data ecosystems.

**Why This Is Important**

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

**Business Impact**

The integration of augmented data management capabilities to streamline the delivery of data and analytics to the business in a unified offering is an attractive proposition. These offerings promise to unify the exploratory world of data science with the production delivery of data warehouses. They promise to unify operational and analytics systems with a holistic management framework. They address key data management disciplines such as data integration, data sharing, governance and metadata.

## Drivers

- Data and analytics architectures are under significant stress on two fronts: hybrid and multicloud deployment environments, and the diversity of data persistence models required to meet the increasing demands of data and analytics.
- Data and analytics leaders are seeking to resolve data silos, which now span multiple deployment environments and frequently require different and potentially conflicting operating models.
- At the same time, enterprises are looking to unify the way they engage with different data models, platforms and use cases.
- Data ecosystems serve as a unifying approach to resolving these pressure points. Built on a common foundation of governance and metadata, they enable new practices like DataOps, and new architectures like the emerging data fabric. Infused with AI and ML capabilities, they will become self-optimizing and self-tuning, and support financial governance efforts through cost optimization.

## Obstacles

- While cloud data ecosystems have a vision of unifying data and analytics environments (exploratory analytics and production delivery, operational systems, and analytics systems) with common governance, security and metadata, there is still significant work needed to make this a reality. Gaps exist in data integration, data quality, metadata and governance. These will need to be addressed either through native CSP offerings or partnerships with ISVs to fully realize the vision of the cloud data ecosystem.
- When combining native CSP offerings with third-party ISV offerings, data and analytics leaders may find additional effort is required to integrate these components. This undermines the core concept of a unified, holistic data ecosystem. While CSPs are working with third-party ISVs to provide open ecosystems, their initial focus remains on ensuring that their own components are working well together, and a more polished experience for both ISV and CSP offerings may be delayed.

## User Recommendations

Data and analytics leaders will need to assess the maturity of these offerings, and the degree to which they deliver on the promise of a unified environment. End users report that cloud data ecosystems are still early in their maturity, with new features and capabilities emerging on a regular basis. Early adopters should be sure to:

- Assess points of integration between various components to determine how cohesive the resulting ecosystem is. A less cohesive ecosystem will require significantly more integration time and effort.
- Ensure that your cloud data ecosystem has a well-articulated path to production for the full data life cycle (from discovery to production-optimized delivery).
- Define what you expect CSPs to deliver as part of the solution, and what capabilities you expect to obtain from third-party ISVs; expect to spend more time on integration efforts when combining CSP and ISV offerings.

## Sample Vendors

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

## Gartner Recommended Reading

[Cloud Data Ecosystems Emerge as the New Data and Analytics Battleground](#)

## Distributed Transactional Databases

Analysis By: Rick Greenwald

Benefit Rating: Moderate

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

### Definition:

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

### Why This Is Important

Transactional databases are the heart of OLTP systems and are used in all lines of business. A distributed transactional database allows for transactions to be performed on any of the distributed nodes of the database instance without a loss of data integrity.

Distributed databases have to choose between consistency and availability, so distributed transaction technology must use specialized technology and hardware to reduce the downsides of that choice.

## Business Impact

Organizations who need to implement transactional applications that span geographic distances will be able to implement systems without compromise or workarounds. Systems such as global trading applications will be able to use distributed transactional systems to expand their scope and reach, and potentially create new markets. Globally implemented transactional systems will allow organizations to expand their use of transactional systems and provide competitive advantage.

## Drivers

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

## Obstacles

- Geographically distributed transactions are still not requirements in many transactional systems, so these systems do not require the extra functionality offered by distributed transactional databases. Implementers can design systems and data schemas to reduce or eliminate the impact of potential losses of data integrity with distributed systems.
- Data that is not required for distributed transactional scenarios can be distributed and still be implemented with a much wider variety of systems while still delivering scalability and performance advantages.



## User Recommendations

- Select a distributed transactional database if your use case requires data integrity implemented over a widely distributed database instance.
- Examine the use of design compromises, rather than a distributed transactional system, if your need for distributed integrity can accommodate these compromises without excessive development and maintenance.

## Sample Vendors

Cockroach Labs; Google Cloud Platform (Cloud Spanner); MariaDB Foundation; NuoDB; Yugabyte

## Gartner Recommended Reading

[Data Consistency Flaws Can Destroy the Value of Your Data](#)

## Ledger DBMS

Analysis By: Donald Feinberg

Benefit Rating: Moderate

Market Penetration: Less than 1% of target audience

Maturity: Emerging

### Definition:

A ledger DBMS is an append-only, immutable DBMS with an embedded cryptographically verifiable audit trail. A ledger DBMS is useful for private and permissioned “blockchain-like” applications where distributed consensus is not required and one entity has control over the ledger and designates which parties, other than the owner, have access to and may add to it, i.e., the users trust the provider.

### Why This Is Important

Ledger DBMSs provide many of the benefits of a blockchain platform, like data tampering detection and auditing, without the complexity of configuring and managing a decentralized environment. They are managed by a single entity, which makes their implementation and management far easier and more secure.

## Business Impact

Today, many blockchain projects are forced to use public blockchain technologies when a DBMS would suffice. Ledger DBMSs represent a choice that is more manageable and easier to implement. This will enable businesses to use ledger technology where immutability is required in use cases such as audit trails, data lineage, digital assets and sharing data. Ledger DBMSs will also have better performance, as there is no need for massive public participation in the consensus process.

## Drivers

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

## Obstacles

- Most ledger DBMSs are new and relatively immature.
- Despite their utility, adoption of ledger DBMSs is evolving slowly as the benefits and potential use cases are being understood by the broader market.
- This technology has just reached the Peak of Inflated Expectations. It will take demonstrable market successes and additional product introductions to diminish the hype.

## User Recommendations

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

## Sample Vendors

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

## Gartner Recommended Reading

[Amazon QLDB Challenges Permissioned Blockchains](#)

## Private Cloud dbPaaS

Analysis By: Adam Ronthal

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

### Definition:

Private cloud database platform as a service (dbPaaS) offerings bring the self-service and scalability of public cloud dbPaaS to a private cloud infrastructure, without external exposure. Private cloud dbPaaS offerings should provide similar benefits to their public cloud counterparts — a database management system or a data store engineered as a scalable, elastic, multitenant service, ideally with subscription or chargeback pricing models.

## Why This Is Important

Private cloud database platform as a service (dbPaaS) offerings continue to emerge in the marketplace, with a number of offerings from both database management system (DBMS) and infrastructure vendors. These offerings may leverage the existing container-based infrastructure common to many private cloud offerings, either as a private IaaS or as PaaS frameworks. They may also be self-contained products in an appliance form factor or extensions of existing cloud service provider offerings.

## Business Impact

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

## Drivers

- Private cloud dbPaaS will primarily appeal to enterprises that are not yet ready or are unable to embrace public cloud alternatives. Public cloud inhibitors (and private cloud dbPaaS drivers) may include: (1) regulatory, governance or security requirements, or the need to operate in an “air-gapped” environment; (2) significant on-premises centers of gravity that are not yet able to move to public cloud; (3) concerns with network connectivity, latency or performance issues in a hybrid cloud environment; (4) data sovereignty requirements that cannot be met by public cloud data centers; (5) compatibility concerns with on-premises environments associated with native public cloud offerings.
- Many of these public cloud inhibitors are transient in nature and may be addressed as organizational culture adapts to the cloud, comfort levels and regulatory best practices associated with operating in the cloud improve. Cloud service provider presence will also expand to additional regions and availability zones. As such, private cloud dbPaaS will be a transitional stage — albeit a potentially long-term one — for many adopters.
- While the initial offerings in the space have been associated with more traditional vendors with strong on-premises presence, the cloud service providers are now engaging with private cloud dbPaaS offerings as well, almost always based on a container strategy that reaches into on-premises data centers.

## Obstacles

- “Cloud first” strategies are being broadly adopted by both vendors and end-users alike, which may inhibit choice and breadth of offerings as vendors increasingly focus on public cloud approaches.
- These cloud-first strategies will eventually become “cloud only,” further limiting options for those looking to deploy dbPaaS in on-premises environments.
- Any new implementations are likely to skip private cloud dbPaaS in favor of a straight to public cloud strategy.
- Adoption will be limited to established enterprises with existing on-premises data center footprints.

## User Recommendations

Focus on the following capabilities when evaluating private cloud dbPaaS offerings:

- **Breadth of DBMS services offered** — Not all offerings support a full range of database types.
- **Storage model** — For container-based services, a scalable, persistent data storage tier will be required to effectively use these offerings.
- **Pricing model** — Flexibility of pricing models is beneficial to accommodate both capex and opex approaches.
- **Production capabilities** — Evaluate the high availability and disaster recovery capabilities to ensure they can meet your requirements.
- **Disconnected operations** — Many of these offerings have a cloud-based control plane that is part of the management stack. If connectivity to the cloud is unreliable, ensure that the selected offering meets any requirements for disconnected operations.
- **Path to the public cloud** — Private cloud dbPaaS is often a transitional technology. These offerings should be evaluated for continuity from on-premises operations into the public cloud.

## Sample Vendors

Alibaba Cloud; Amazon Web Services (AWS); Google Cloud Platform (GCP); IBM; Microsoft; Oracle; Robin; VMware

## Gartner Recommended Reading

[Market Guide for Container Management](#)

[Solution Path for Implementing Containers and Kubernetes](#)

[Private Cloud: Divide to Conquer](#)

[Should My Cloud Be Public or Private?](#)

[Magic Quadrant for Cloud Database Management Systems](#)

## 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](#)

## 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-centric (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](#)

## File Analysis

Analysis By: Michael Hoeck

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

### Definition:

File analysis (FA) software analyzes, indexes, searches, tracks and reports across multiple file and database sources. FA software reports on detailed metadata and contextual information to enable better information governance, risk management, data management actions, and the analytical assessment of unstructured and structured data.

## Why This Is Important

FA solutions improve organizations' ability to manage ever-expanding repositories of unstructured/structured data. They increase visibility to disparate, unorganized sources of information, allowing IT teams to establish qualified operational efficiencies; compliance teams to improve insight to sensitive information, including personal information (PI); and security team exposure to areas of data access risk.

## Business Impact

- FA solutions reduce business risk and inefficiencies by identifying access permission issues, locating and protecting intellectual property, and eliminating or quarantining sensitive data.
- Users gain actionable insights to optimize storage efficiency by identifying redundant, outdated and trivial (ROT) data.
- Management of information governance is improved, as FA solutions feed data insights into corporate retention initiatives and classify valuable business data.
- Lower business risk and storage utilization lead to savings.

## Drivers

- The desire to mitigate business risks (including security, breach and privacy risks); identify sensitive data, optimize storage costs; and implement information governance
- The hype associated with the growing trend of privacy regulations, such as the EU's General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA), which has greatly increased interest in and awareness of FA software
- The potential value of contextually rich data, which is capturing the interest of data and analytics teams

## Obstacles

- Successful results from using file analysis software may be affected by a lack of data policy buy-in or consensus from key internal constituencies, including executive sponsorship.
- Establishment of action-oriented retention policies is required to defensibly delete redundant, outdated and trivial data identified by FA software.
- Although FA solutions and corresponding budgets resonate highest with compliance and efficiency use cases, budgeting aligned to data risk and analytics use cases may be challenged, requiring additional sponsorship.

## User Recommendations

- Use FA software to better grasp the risks of an unstructured data footprint, including where it resides and who has access to it, and to expose another rich dataset for driving business decisions.
- Develop strong information governance principles by establishing, updating and enforcing retention policies, using the information gathered and remediation actions from FA software.
- Identify the potential risks of unknown data stored in structured database repositories often associated with applications.
- Clean up old file shares containing ROT data that can be defensively disposed of or relocated to optimize data infrastructure.
- Create data visualization maps to better identify the value and risk of the data, including the data owner.
- Use FA software to enable IT, line of business (LOB) and compliance teams to make better-informed decisions regarding classification, information governance, storage management and content migration.

## Sample Vendors

ActiveNav; Data Dynamics; Ground Labs; Index Engines; Netwrix (Stealthbits); SailPoint; Spirion; Titus; Varonis; Veritas

## Gartner Recommended Reading

[Market Guide for File Analysis Software](#)

## Sliding into the Trough

### Information Stewardship Applications

Analysis By: Guido De Simoni, Andrew White

**Benefit Rating:** High

**Market Penetration:** 1% to 5% of target audience

**Maturity:** Obsolete

#### Definition:

Information stewardship applications support the work of information stewards, by providing application capabilities such as policy performance monitoring, data quality analysis, business glossary, workflow and exception management. These applications may also include playbooks and preloaded templates to help make this business role more effective.

#### Why This Is Important

Information stewardship applications emerged as a new technology to support effective operationalization of data and analytics governance policy management — from the perspective of data stewards — in use cases spanning business applications, operational data and analytics in data warehouses and data lakes.

#### Business Impact

Data and analytics governance cannot be sustained and scaled without an operational data and analytics stewardship role and function. A successful stewardship routine will lead to sustainable and persistent benefits in support of programs and projects such as EIM, MDM, application data management, analytics and business intelligence. These benefits include increased revenue, lower IT and business costs, reduced cycle times, improved trust in organizational data and increased business agility.

#### Drivers

- Encouragement to share data, increased data reuse, improved consistency and accelerated time to value because of the use of existing data dictionaries to identify areas of synergy between data used for different business initiatives (both data content and meaning)

- More effective understanding and communication of the semantic meaning of data will facilitate resolution of contention between business teams when inconsistency arises and reduce the amount of time and effort wasted on reconciliation, so that efforts can focus on new business actions
- Intelligent decisions about the information life cycle, from data interoperability and standards to archiving, disposal and deletion

## Obstacles

- The overlap between the data governance board and the analytics center of excellence, which is now discovering that it needs to comply with and respect policy set by others, has not been captured yet in the market.
- The variety of requirements affects vendors' experimentation with and assessment of information stewardship applications. In particular, we observe clear market disruptions related to the adoption of data catalogs and organizations scrambling to work within the privacy management requirements of regulations such as the EU's General Data Protection Regulation (GDPR).
- The potential convergence of capabilities in the context of data and analytics governance. This is leading to the emergent market of data and analytics governance platforms that will consume information stewardship applications. For this reason they are becoming obsolete before the Plateau of Productivity and in a state of stall on the Hype Cycle.

## User Recommendations

- Evaluate the capabilities needed from fit-for-purpose, business-user-oriented information stewardship and other solutions, as compared with IT-centric data management tools, including data quality, metadata management and federation/integration capabilities.
- Run a proof-of-concept for vendor solutions involving all contributing roles, such as business users, information governance board members, information architects, information stewards and business analysts.
- Focus on all dimensions (people, process, technology and data) when addressing the data and analytics stewardship use case. These dimensions are relevant for effective use of a solution to maximize your ROI through reuse, while also minimizing administrative costs and errors due to inconsistencies across technologies.



- Explore the emerging data and analytics governance platforms for all their data and analytics stewardship operational requirements

## Sample Vendors

Alation; Alex Solutions; Collibra; Global Data Excellence; Infogix; Informatica; OvalEdge; Protago

## Gartner Recommended Reading

[Predicts 2021: Data and Analytics Strategies to Govern, Scale and Transform Digital Business](#)

[The Role of Technology in Data and Analytics Governance Policy Management](#)

[Implement Your Data and Analytics Governance Through 5 Pragmatic Steps](#)

## Application Data Management

Analysis By: Andrew White, Malcolm Hawker

Benefit Rating: Moderate

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

### Definition:

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

## Why This Is Important

Clients are often shocked to find their vendor offers modern business applications that take scant care of governance of the data used in them. Application vendors often fail to offer governance and stewardship solutions as part of the application service. As a result, quality and trust in such data falls over time and eventually business process integrity falls and outcomes start to suffer. Note that application data may include reference to or copies of master data.

## Business Impact

The primary benefactors of this discipline are:

- Application data once identified ensures the right level of governance effort is aligned to the right kind of business impact the data has.
- The application of stewardship roles, in operational and analytical use cases, can be determined more effectively.
- Business goals for overall D&A governance is more likely assured with a more organized approach that now includes application data.

## Drivers

- The vast majority of “successful” go-lives of business applications such as ERP, CRM or custom-built applications do not include any qualification of data and analytics governance. The result, very often observed in client inquiry, is that on average 7 months after the go-live, organizations spot the vast array of small but noticeable business issues held hostage to lack of governed data. Business performance and process integrity fails and business outcomes start to be negatively impacted.
- MDM was misunderstood. MDM was meant to have a laser-like focus on the minimal number of most widely shared attributes describing things like customer and product. This was confused with idealistic technical jargon that ended up trying to master all data for every use equally. This has now been exposed by too many organizations that still don't see the difference and so MDM continues to fail and the need for ADM continues to grow.
- Digital business success hinges not on the quality and governance of all data equally, but a graduated, efficient means to classify data and apply only the needed level of governance. Such growing demand on scaling digital business will, of necessity, drive increased need to recognize and adopt ADM.

## Obstacles

- The half-life of a successful “go live” of business application success is 7 months. After that we get phone calls from clients noting, “we have lost control of our data.” The reason this has become acceptable is that, for the most part, most organizations don’t fail.
- The ability for the organization to change is held back by budgeted-for mediocrity of meddling with data while the customer sits and waits for service. This is not an acceptable way to run an organization but too few D&A leaders stand up and say this.
- Poorly scoped MDM programs prevent successful MDM and ADM from taking shape since the result is that too much data is treated equally as master data (when it clearly is not).
- Traditional top-down governance programs lead to the same misunderstanding and poorly scoped initiatives.

## User Recommendations

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

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

## Sample Vendors

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

## Gartner Recommended Reading

[Design an Effective Information Governance Strategy](#)

[Create a Master Data Roadmap With Gartner's MDM Maturity Model](#)

[How Augmented Data Management Capabilities Are Impacting MDM and Data Governance](#)

[Think Big, Start Small, Be Prepared – Master Data Management](#)

## Augmented Transactions

Analysis By: Donald Feinberg

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

### Definition:

Augmented transactions use various forms of augmented analytics – advanced analytics, artificial intelligence (AI) and machine learning (ML) – to enable concurrent analytical and transaction processing. In-memory computing (IMC) technology is a key enabler for augmented transactions.

### Why This Is Important

Augmented transactions will redefine the way some business processes are executed, in real time, assisted by analytics, AI and ML (for example, planning, dynamic repricing, forecasting and what-if analysis). The augmented transaction becomes an integral part of the business process itself, rather than a separate activity performed after the fact.

### Business Impact

Augmented transactions enable business users to make more-informed operational and tactical decisions in real time. They open up the possibility of driving prescriptive decisions without human intervention and improve business leaders' situation awareness in operations. Augmented transactions also provide constantly updated forecasts and simulations of future business outcomes.

## Drivers

- Augmented transactions were formerly called in-process hybrid transactional/analytical processing (HTAP), but the definition has broadened to include AI and ML, along with augmented analytics. The transactional and analytical processing dimensions are designed together in the context of an individual transaction or process and are delivered as a single application. This helps streamline the overall technology infrastructure.
- Decisions based on analysis can be improved with automated recommendations based on real-time analysis of the impact of different options on key business indicators. Performance and scalability limitations have historically prevented advanced analytics from running concurrently with or within transaction processing. However, IMC (specifically in-memory DBMS [IMDBMS]) makes it possible to relax the restriction of using external or after-the-fact analytics and implement even the most sophisticated augmented transactions.
- Many DBMSs are enabling ML within the DBMS, increasing the available functionality for augmented transactions. Further, the growth of IMDBMS, in general, is enabling the use of augmented transactions.

## Obstacles

- Skills for augmented transactions are still limited to the most leading-edge organizations. Best practices have not yet crystallized and skills are hard to find.
- When retrofitting existing (custom or packaged) applications at the core of business operations, augmented transactions require massive, expensive and risky reengineering efforts.

## User Recommendations

- Pilot augmented transactions in individual “system of innovation” projects.
- Discuss with strategic information management and business application providers their vision, roadmap and technology for augmented transactions in their products.
- Use augmented transactions for use cases involving observation data (such as IoT) or interaction data (such as log data) for which real-time operational analytics is required.
- Educate business leaders about augmented transactions and IMC concepts and their importance. Brainstorm with them to identify concrete opportunities to rethink business processes and create applications that could not be implemented using traditional architectures.

## Sample Vendors

Aerospike; Microsoft; Oracle; Redis Labs; SAP; SingleStore

## Gartner Recommended Reading

[Magic Quadrant for Cloud Database Management Systems](#)

[Critical Capabilities for Cloud Database Management Systems for Operational Use Cases](#)

## 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](#)

## **Augmented Data Cataloging and Metadata Management**

**Analysis By:** Guido De Simoni

**Benefit Rating:** High

**Market Penetration:** 5% to 20% of target audience

**Maturity:** Early mainstream

### **Definition:**

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

### **Why This Is Important**

Augmented data cataloging and MMSs support organizations that manage varied data assets. Moreover, demands for accessing and using data are no longer limited to IT, and data-oriented citizen roles are emerging in the business. Also, data and analytics leaders are facing privacy requirements that force new approaches to data management. The pervasive use of metadata is across the data management landscape and results in the automation of many activities.

## Business Impact

- Management of complexity: Augmented data cataloging and MMSs help to break down and reduce the complexity often inherent in data.
- Automation of processes: Because data is subject to change, there are numerous recurring activities that MMSs may enable or streamline by (partial) automation.
- Collaboration: Metadata requires the contribution of numerous people from different divisions and countries. An MMS can provide a multiuser environment able to address rich collaboration requirements.

## Drivers

- Augmented data cataloging and MMSs are accelerating due to innovation generated by active metadata that leverages AI and machine learning.
- Active metadata enables real-time analysis of the applicability of data, checks on the veracity of data sources used and monitors the ways that users act.
- Informal and formal teams emerge and convert to community participation with as much automation as possible when supported by augmented data cataloging and MMSs. These demands are only starting to be addressed by vendors, with modern metadata management practices slowly being established within organizations.

## Obstacles

- The lack of maturity of strategic business conversations about metadata.
- The expensive, but required, effort to integrate metadata management solutions in multivendor environments. This inhibitor has started to be addressed by new vendors' initiatives relating to openness and interoperability (see, for example, [ODPi](#)).
- The lack of identification of metadata management solutions with capabilities that meet the current and future requirements of specific use cases.

## User Recommendations

- Data and analytics leaders who have already invested in data management technologies should first evaluate the metadata management capabilities of their existing data management tools, including data integration, data quality and even master data capabilities, before buying a modern MMS.
- If dealing with emerging use cases, including data and analytics governance, security and risk, and support for analytics and augmented data value, they should learn about these use cases, and build pilot implementations using MMSs.
- The introduction of “active metadata” concepts means that some of the more basic catalog capabilities no longer differentiate solutions in the market. We expect augmented data cataloging and MMSs to take two to five years to reach the Plateau of Productivity as the technology continues to expand in terms of both capabilities and support for both existing and emerging use cases.
- Innovation continues to leverage this market that is in the early mainstream phase.

## Sample Vendors

Alation; Alex Solutions; Collibra; data.world; erwin; IBM; Infogix; Informatica; Semantic Web Company; SAP

## Gartner Recommended Reading

[Magic Quadrant for Metadata Management Solutions](#)

[Critical Capabilities for Metadata Management Solutions](#)

## Graph DBMS

Analysis By: Merv Adrian

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

**Definition:**

Graph DBMSs store data elements and their relationships as equal, first-class objects, optimizing for the connections among elements. Most graph DBMSs use basic graph theory and are suitable for multiple types of interactions ranging from simple node, edge traversal and triple pattern matching for transactional uses, to complex multihop queries, reasoning and inference, and algorithms for analytical workloads.

**Why This Is Important**

Processes such as complex traversal operation of data networks are difficult to do at scale with traditional, relational DBMSs, as is support for highly dimensional data. Graph analysis enables the use of multiple, simultaneous interpretive models, in parallel, over the same data, allowing business users to describe dynamic and complex domains more naturally and efficiently.

**Business Impact**

Graph support has widely demonstrated value for supply chain, transportation/logistics, epidemiology, financial fraud detection and network analysis. The COVID-19 pandemic has increased awareness of the value of graph. The power of correlations across multiple sensors in asset-intensive IoT use cases also fuels increasing marketplace interest. In addition, embedded graph support of metadata use cases is becoming increasingly important to a variety of IT activities and tools.

## Drivers

- The challenge of connecting a broader variety of data (high dimensionality) to uncover new insights, and hitting the wall with relational DBMS because of performance challenges, have accelerated experimentation.
- New vendors continue to enter the market, including cloud providers who offer both their own and third-party solutions.
- Cloud-first development and experimentation are increasing, facilitated by ingest and conversion of existing data in a low-cost, low-risk environment.
- The increasing importance of semantics in organizing and operating larger and more complex data fabrics, is accelerating experimentation, adoption and training inside IT organizations.
- Mature, widely distributed multimodel DBMS products are embedding graph DBMS, and both open-source and commercial graph DBMS offerings are themselves maturing and gaining adherents. The resultant increasing visibility of use cases that exploit graph and achieve successes in solving previously intractable operational problems is raising interest even further.
- Funding has been increasingly available to new and growing vendors in the space, sending another signal to potential adopters to test the value of the technology.

## Obstacles

- Modeling, loading, processing and analyzing graph data require new and uncommon skills. More user-friendly tools and interfaces are required due to the difficulty of modeling and evaluating graphs versus tables. Dueling architectures (triplets vs. property graphs), languages and APIs, and varying support among vendors for libraries of machine learning algorithms make choosing a product complex.
- Existing analytical tools can often be successful for graph use cases without requiring a dedicated DBMS. If the requirement is not too complex or the scale is not too great, these alternatives may be adequate, making graph DBMS unnecessary.
- The inability of relatively more technical practitioners to adequately explain the benefits of graph to business sponsors continues. Moreover, many use cases are now moving to embedded functionality in other tools (AI, metadata, data integration lead the way), making the decision to plunge into an unfamiliar type of product less attractive.

## User Recommendations

- Assess graph DBMSs' capabilities when relational DBMS performance requirements for highly nested or relational data fall outside current processing capabilities. The core advantage of graph DBMSs is the relationships they support and discover in data, which they then persist in models.
- Use open-source graph DBMS projects, or community editions of commercial ones, to experiment and gain experience.
- Ensure any open-source products you use in production environments are commercially supported. Graph DBMSs can render the relationships and traversal of data (as discovered by a data scientist or data science team) into a reusable form for data miners, data engineers and business analysts. Scaling and reliability are typically characteristics of commercial implementations and cannot be relied on with open-source offerings.
- Select graph DBMSs based on your analytics requirements. RDF- and property graph-based architectures have capabilities that should be factored in when making your selection.

## Sample Vendors

Amazon Web Services (AWS); Cambridge Semantics; Dgraph Labs; Franz; Neo4j; Ontotext; OpenLink Software; Redis Labs; Stardog; TigerGraph

## Gartner Recommended Reading

[Graph Steps Onto the Main Stage of Data and Analytics: A Gartner Trend Insight Report](#)

[Working With Graph Data Stores](#)

[Understanding When Graph Technologies Are Best for Your Business Use Case](#)

[How to Build Knowledge Graphs That Enable AI-Driven Enterprise Applications](#)

[Market Guide for Graph Database Management Solutions](#)

## Augmented Data Management

Analysis By: Donald Feinberg

Benefit Rating: High

**Market Penetration:** 5% to 20% of target audience

**Maturity:** Early mainstream

**Definition:**

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

**Why This Is Important**

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

**Business Impact**

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

**Drivers**

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

- Metadata management — Increasingly, AI and ML are used to explore and define metadata from the data, helping the analysts to evaluate metadata more rapidly, accurately and with reduced redundancy. Similarly, augmented data management functions can automatically catalog data elements during data extraction, access, and processing.

- Data integration — To automate the integration development process, by recommending or deploying repetitive integration flows, such as source-to-target mappings.
- MDM — MDM solution vendors will increasingly focus on offering AI- and ML-driven configuration and optimization of record matching and merging algorithms as a part of their information quality and semantics capabilities.
- Data quality — AI and ML will be used to extend profiling, cleansing, linking, identifying and semantically reconciling master data in different data sources.
- DBMS — In addition to enhancing cost-based query optimization, AI and ML are being used to automate many current manual management operations, including the management of configurations, elastic scaling, storage, indexes and partitions, and database tuning.

Products such as Amazon Aurora and Oracle Autonomous Database, and data management software that has made the transition to the cloud, are steadily gaining more users. In these platforms, enormous volumes of user data on a consistent infrastructure improve the applicability of the results and offer opportunities for the continuous training and retraining of models. As a result, they are being aggressively used to drive competitive improvements and some of the features are making their way into on-premises, private cloud deployments as well. We believe the improvements, initially available in cloud platforms, will neutralize the distinction between cloud and on-premises over the next few years.

## Obstacles

- The adoption of these capabilities is gated by the movement of the product categories to the cloud, where they are delivered first. Offerings in data integration, data quality, master data management (MDM), metadata management and DBMS software are proliferating and maturing rapidly but at different speeds.
- This profile therefore represents an aggregate view and position of what is happening across data management, some areas more mature in augmented data management than others.
- The lack of clear use cases and necessary critical capabilities also create an obstacle to adoption.



## User Recommendations

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

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

## Sample Vendors

Amazon Web Services (AWS); Cinchy; CluedIn; IBM; Informatica; Microsoft; Oracle; SAP; SnapLogic; Teradata

## Gartner Recommended Reading

[Data Fabrics Add Augmented Intelligence to Modernize Your Data Integration](#)

[Augmented Data Catalogs: Now an Enterprise Must-Have for Data and Analytics Leaders](#)

[Automating Data Warehouse Development](#)

## Blockchain

Analysis By: David Furlonger, Christophe Uzureau, Rajesh Kandaswamy

Benefit Rating: Transformational

Market Penetration: 5% to 20% of target audience

**Maturity:** Adolescent

**Definition:**

A blockchain is an expanding list of cryptographically signed, irrevocable blocks of records shared by all participants in a peer-to-peer (P2P) network. Each block of records is time-stamped and references links to previous data blocks. Anyone with access rights can historically trace a state change in data or an event belonging to any participant.

Distributed ledgers are design-limited and lack decentralized and tokenized elements.

**Why This Is Important**

Blockchain fundamentally changes how commerce is conducted and value is exchanged by enabling:

- The creation, use and representation of assets in new forms
- Different kinds of autonomous, machine-based interlocutors to make decisions
- The redesign and automation of rules and processes governing transactions and interactions, changing the competitive landscape
- The development and deployment of new digital infrastructures that will redefine citizen, enterprise, industry and geopolitical relationships

**Business Impact**

Gartner's blockchain spectrum anticipates disruptions evolving over this decade.

Blockchain-complete solutions start to enter the mainstream over the next two years via developments in areas such as:

- Tokenization of digital and physical assets
- The decentralization of finance and business ecosystems
- Software-designed and -defined business process interaction and execution
- Partnerships and collaborations based on a more decentralized governance
- Self-sovereign identity management and portability

## Drivers

- Opportunities are becoming clearer and experiments continue, especially for blockchain-inspired or enterprise distributed ledger solutions that focus on process efficiency and cost management.
- The COVID-19 pandemic has caused executive leaders to accelerate digital business activities, and this is propelling a renewed focus on blockchain and growth opportunities afforded by a more programmable economy.
- Blockchain is seen as a way to address multiple problems that other technologies cannot easily address, such as audit and compliance, oversight of public fund distribution, healthcare passporting, food security and financial inclusion — via CBDC and ESG.
- Blockchain is being perceived as a potential foundational infrastructure for capturing new growth opportunities via fractional ownership via decentralized finance (DeFi) and NFT, customer engagement through enhanced loyalty, and rewards.
- The emergence of new digital infrastructures, such as in China with the development of the BSN and DCEP, is now more directly linking supply chains to payments and financing. The integration of synergistic technologies such as AI, NFC, 5G and IoT with blockchain could also fuel digital infrastructure development (which will heighten geopolitical tensions), as well as enterprise project investments and vendor solutions.

## Obstacles

- Immature standards
- Establishing effective governance for consortia and alliances
- Geopolitical tensions surrounding emerging technologies
- Cybercrime/warfare
- Immature UX
- Executive leader education and awareness gaps
- Demand-side intransigence/apathy
- Shifts in government regulations
- Shortcomings in the evolution of some of the core technologies and proliferation of scalability problems
- Lack of interoperability technically and from siloed business projects
- Difficulty integrating with and retiring legacy systems
- Data management complexity
- Lack of blockchain talent
- Organizational obstinacy and lack of user experience and education
- Security and privacy challenges — especially relating to cross-border activities
- Organizational concerns about decentralized operations
- Negative enterprise perceptions concerning digital assets, tokens and cryptocurrency generally
- Shift in investments due to COVID-19
- Short-termism created by cryptocurrency valuations

## User Recommendations

- Educate executive leaders about the opportunities and threats that blockchain capabilities introduce by using workflow models of value exchange.
- Use clear language and definitions in internal discussions about how distributed ledgers may or may not improve existing systems and processes.
- Continue to develop proofs of concept (POCs) – especially in the context of market ecosystems.
- Identify integration points with existing infrastructures including: digital wallets, core systems of record, customer service applications and security systems, artificial intelligence (AI) and Internet of Things (IoT).
- Ensure sufficient innovation capacity is applied to the evolution of distributed ledgers and blockchains outside of your immediate industry.
- Read [The Real Business of Blockchain: How Leaders Can Create Value in a New Digital Age](#).

## Gartner Recommended Reading

[Non-Fungible Tokens \(NFTs\) Create New Digital Products and Business Models](#)

[Shape Your Digital Strategy With Central Banks' Intentions Toward Digital Currencies](#)

[Accelerate Financial Ecosystems to Keep Up With Digital Giants](#)

[What Is Ethereum 2.0 and How Does it Relate to Digital Business Acceleration and a New Programmable Economy?](#)

[Take Control of Your Digital Acceleration by Focusing on How Value Flows Through Ecosystems](#)

[Use 4 Business Currencies and 5 Archetypes to Evaluate Blockchain Initiatives](#)

[Executive Leaders Should Embrace Social and Economic Decentralization](#)

[Understanding the Gartner Blockchain Spectrum and the Evolution of Technology Solutions](#)

## 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](#)

## SQL Interfaces to Object Stores

Analysis By: Adam Ronthal, Merv Adrian

Benefit Rating: Moderate

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

### Definition:

SQL interfaces to object and file stores provide the ability for enterprises to interact with data residing in cloud object stores and on-premises object and file stores such as HDFS using familiar SQL syntax. Object stores are increasingly used for storing large volumes of multistructured data, and are often used in support of data lakes. This enables familiar tools for analysis — most of which use SQL as their query language.

## Why This Is Important

Cloud object stores like Amazon Simple Storage Service (Amazon S3), Microsoft Azure Blob Storage and Azure Data Lake Storage, and Google Cloud Storage have become the foundation for data lakes and a common clearinghouse for data between various cloud services. On-premises, HDFS is used for similar purposes. SQL interfaces to these stores offer easier access to a broader base of business analysts and data scientists than BI tools have hitherto delivered.

## Business Impact

Object and file stores increasingly serve as the basis for data ingest. Some data will pass through to other services, such as traditional data warehouses. Other data will remain and serve as the basis for a data lake for later exploration. The stores are becoming the standard for cloud data lakes. SQL interfaces allow ready access to this data from a standard interface, sometimes integrated with existing relational offerings. They play a key role in democratizing access to the data lake.

## Drivers

- As more data accumulates, demand will grow for a means of effectively accessing that data in familiar, standardized ways.
- SQL interface functionality may also be embedded within analytics products that use SQL for multiple targets, such as Amazon Athena, Microsoft Azure Data Lake Analytics and Google Cloud BigQuery — all available for cloud-specific targets.
- Increasing use of open standards, including Apache Arrow, Apache Drill, Apache Druid, Apache Hive, Apache Spark and Presto, by new and existing tools in multiple categories will drive continued development and enhancement of SQL interfaces in open and proprietary versions.
- With improving support for the SQL standard, the data becomes accessible via a broader range of BI and visualization tools and applications, opening increasing numbers of use cases.
- Increasing integration with relational offerings will also expand the range of use cases and provide more robust functionality in some cases.
- As these tools continue to mature, they will add additional layers of optimization, including performance, data quality, data preparation, governance and security capabilities — although few, if any, implement many of these today.



## Obstacles

- The multiple marketplace offerings that incorporate SQL interfaces into object stores create additional choice and complexity for data and analytics professionals.
- The interfaces do not offer the same level of performance optimization that is available in a DBMS. Newer stand-alone tools have less mature capabilities for processing complex SQL statements, even for relatively modest multiway JOINS.
- Since the storage does not support update, intermediate results will often not be persisted unless the tool in use provides its own persistence layer or has a mechanism to provide this functionality. This will limit their usefulness.
- Even tools that store their own derived indexes will often lack robust security and governance capabilities, including the ability to support large numbers of simultaneous users.
- Most leading DBMSs already in place offer their own means to access external object and file data, making the complexity, effort and cost of deploying a stand-alone tool less attractive.

## User Recommendations

- Evaluate these capabilities in incumbent products to see what benefits they can offer, while exploring synergies to related complementary technologies already in use, such as relational data warehouse offerings.
- Attempt to optimize multiple workload types, with the understanding that they may require several SQL interfaces, or a more broadly capable multitarget layer. SQL interfaces aren't general-purpose query engines.
- Test core functionality and integration with third-party BI and analytics tools to ensure that functional support and performance meet your needs.
- Deploy these offerings initially for exploratory and data discovery use cases, and the ability to access "cooler" data via SQL at lower cost and performance.
- Assess production applicability for operational use cases that may have specific performance and concurrency requirements only once these offerings have proven their value in the cases described above.

## Sample Vendors

Ahana; Amazon Web Services; Cloudera; Databricks; Dremio; Incorta; Jethro; Kyligence; RAW Labs; Starburst; Varada

## Gartner Recommended Reading

[Market Guide for Analytics Query Accelerators](#)

[Choosing the Right Path When Exploring Hadoop's Future](#)

[The Practical Logical Data Warehouse](#)

[Assessing the Optimal Data Stores for Modern Architectures](#)

## Event Stream Processing

Analysis By: W. Roy Schulte, Pieter den Hamer

**Benefit Rating:** Transformational

**Market Penetration:** 20% to 50% of target audience

**Maturity:** Early mainstream

### Definition:

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

### Why This Is Important

ESP is a key enabler of continuous intelligence and related real-time aspects of digital business. ESP's data-in-motion architecture is a radical departure from conventional data-at-rest approaches that historically dominated computing. ESP products have progressed from niche innovation to proven technology and now reach into the early majority of users. ESP will reach the Plateau of Productivity within several years and eventually be adopted by multiple departments within every large company.

## Business Impact

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

## Drivers

Five factors are driving ESP growth:

- Companies have ever-increasing amounts of streaming data from sensors, meters, digital control systems, corporate websites, transactional applications, social computing platforms, news and weather feeds, data brokers, government agencies and business partners.
- Business is demanding more real-time, continuous intelligence for better situation awareness and faster, more-precise and nuanced decisions.
- ESP products have become widely available, in part because open-source ESP technology has made it less expensive for more vendors to offer ESP. More than 40 ESP platforms or cloud ESP services are available. All software megavendors offer at least one ESP product and numerous small-to-midsize specialists also compete in this market.
- ESP products have matured into stable, well-rounded products with many thousands of applications (overall) in reliable production.
- Vendors are adding expressive, easy-to-use development interfaces that enable faster application development. Power users can build some kinds of ESP applications through the use of low-code techniques and off-the-shelf templates.

## Obstacles

- ESP platforms are overkill for most applications that process low or moderate volumes of streaming data (e.g., under 1000 events per second), or do not require fast response times (e.g., less than a minute).
- Many ESP products required low-level programming in Java, Scala or proprietary event processing languages until fairly recently. The spread of SQL as a popular ESP development language has ameliorated this concern for some applications, although SQL has limitations. A new generation of low-code development paradigms has emerged to further enhance developer productivity but is still limited to a minority of ESP products.
- Many architects and software engineers are still unfamiliar with the design techniques and products that enable ESP on data in motion. They are more familiar with processing data at rest in databases and other data stores, so they use those techniques by default unless business requirements force them to use ESP.

## User Recommendations

- Use ESP platforms when conventional data-at-rest architectures cannot process high-volume event streams fast enough to meet business requirements.
- Acquire ESP functionality by using a SaaS offering, IoT platform or an off-the-shelf application that has embedded CEP logic if a product that targets their specific business requirements is available.
- Use vendor-supported closed-source platforms or open-core products that mix open-source with value-added closed-source extensions for mainstream applications that require enterprise-level support and a full set of features. Use free, community-supported, open-source ESP platforms if their developers are familiar with open-source software and license fees are more important than staff costs.
- Use ESP products that are optimized for stream data integration to ingest, filter, enrich, transform and store event streams in a file or database for later use.

## Sample Vendors

Amazon; Confluent; Google; IBM; Informatica; Microsoft; Oracle; SAS; Software AG; TIBCO Software

## Gartner Recommended Reading

[Market Guide for Event Stream Processing](#)

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

### Market Share Analysis: Event Stream Processing (ESP) Platforms, Worldwide, 2020

#### Time Series DBMS

Analysis By: Rick Greenwald

Benefit Rating: Moderate

Market Penetration: 1% to 5% of target audience

Maturity: Early mainstream

#### Definition:

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

#### Why This Is Important

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

#### Business Impact

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

## Drivers

- The expansion of IoT data sources has sparked new interest in this technology, as organizations look to both perform more sophisticated analytics on IoT data and combine that data with more traditional sources.
- Leading cloud vendors, such as Microsoft and Amazon Web Services, are adding time series capabilities or services to their platforms.

## Obstacles

- Time series capabilities are being added to mainstream database offerings, and time series database services are being offered by major cloud providers. These dual developments place pressure on DBMSs whose main benefit is providing time series capability.

## User Recommendations

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

## Sample Vendors

Amazon Web Services; CrateDB; IBM; EraDB; InfluxData; Microsoft; OSIsoft; Redis Labs

## Gartner Recommended Reading

[Time Series Database Architectures and Use Cases](#)

## Data Preparation Tools

Analysis By: Ehtisham Zaidi, Sharat Menon

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

**Maturity:** Early mainstream

**Definition:**

Data preparation is an iterative and agile process for cleaning and transforming raw data into curated datasets for data integration, data science/ML and analytics/BI. Data preparation tools promise faster data delivery time by letting business users integrate datasets for their use cases. They allow users to identify anomalies and patterns, and review their findings in a repeatable fashion. Some tools embed machine learning algorithms that automate and augment repeatable data preparation tasks.

**Why This Is Important**

Data preparation tools are important because they allow less-skilled business analysts, citizen integrators and LOB users to find, enrich and transform their data with minimal coding and data management knowledge. These tools enable self-service data management and integration and use automation to reduce the manual (and often error-prone) interventions needed by business teams before they can make their data fit for purpose prior to analysis in their analytics/BI or data science tools.

**Business Impact**

Data preparation addresses a major challenge that most analysts, citizen integrators and line-of-business users face, i.e., the time, skills and effort needed to find, transform and deliver data for data and analytics use cases. Modern data preparation tools embed ML algorithms to significantly automate tedious data integration tasks. This allows business teams to prepare their data for faster time to analytics, and, therefore, spend more time doing high-value tasks such as data analysis.

## Drivers

- The primary driver for data preparation tools is organizations' desire to improve speed, time to analytics, data sharing and user collaboration, and data science/ML and efficiency while producing trusted and integrated datasets ready for analysis.
- Data preparation tools allow less-skilled users such as citizen integrators, analysts and line-of-business users to perform last-mile data wrangling and transformation without any coding knowledge and minimal IT intervention before they can run analytics.
- Data preparation tools are necessary for data engineering teams to significantly reduce and even automate repetitive data preparation, integration and enrichment tasks so that they can focus on more strategic initiatives and allow business teams to prepare their own data for analysis. This greatly improves their productivity and allows them to focus on more important tasks such as assisting business teams with creating data products and focusing on automation.
- Data science teams demand tools that can significantly reduce the barriers to experimenting with new data sources and application data. Data preparation tools allow data scientists and data engineers to be more flexible with integrating and testing new data sources and targets for their experimental use cases, where data replication with ETL technology will lead to more costs and efforts to manage data pipelines and store data.
- Automating data preparation and integration is frequently cited as one of the major investment areas for data and analytics teams. Data preparation tools that embed AI/ML algorithms drive automation of various repetitive data engineering tasks, which thereby significantly reduces the efforts in integration and change management for data engineering teams.



## Obstacles

- Data preparation tools have proved to be successful in enabling business teams to find and transform their own data, but most organizations struggle with “operationalization” of data preparation tasks, i.e., moving the prepared data from sandbox instances to production. This is because most data preparation tools don’t provide the necessary capabilities to allow IT teams to regulate the prepared data on acceptable data quality and/or enforce data governance on the prepared data.
- Data preparation tools are sometimes wrongly marketed as a replacement to ETL tools. This leads to wrong expectations and less than optimal performance. Data preparation tools need to be used as a complement to data integration tools and not as a replacement.
- Some data preparation tools don’t allow bidirectional exchange of metadata with traditional data management tools like data quality and metadata management tools. This leads to poor scaling of the models created with these tools or even governance chaos.

## User Recommendations

- Evaluate data preparation tools that help analysts, data engineers, citizen integrators, data stewards and other analytics authors to enhance and streamline their data preparation time and effort.
- Evaluate data preparation tools for their ability to scale from self-service models to enterprise-level projects.
- Evaluate stand-alone data preparation tools when your use case is a general-purpose one, needing data integration for different analytics and data science tools. By contrast, evaluate the embedded data preparation capability of incumbent tools if you need data preparation only in the context of those tools.
- Give preference to tools that can coexist with other data management tools (such as data quality or data governance) and can capture, analyze and share metadata with them, to ensure governance and compliance.
- Deploy data preparation tools to complement traditional data integration approaches. Do not look to replace traditional data integration tools with data preparation tools.

## Sample Vendors

Alteryx; DataRobot; IBM; Informatica; SAS; Talend; Trifacta

## Gartner Recommended Reading

[Market Guide for Data Preparation Tools](#)

[Tool: Evaluate Data Preparation Tools Across Key Capabilities](#)

[How to Build a Data Engineering Practice That Delivers Great Consumer Experiences](#)

[Magic Quadrant for Data Integration Tools](#)

[Critical Capabilities for Data Integration Tools](#)

[Toolkit: Job Description for the Role of a Data Engineer](#)

## In-DBMS Analytics

Analysis By: Henry Cook

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

### Definition:

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

### Why This Is Important

In-DBMS analytics provides agility, productivity and a robust way of getting ML results into production — all of which are desirable. They are also relatively new to cloud DBMS systems; once organizations get up to speed on the new in-DBMS analytics capabilities, their use will increase.

## Business Impact

In-DBMS analytics provides a robust way of developing advanced analytics, such as machine learning and artificial intelligence. They provide an ideal vehicle for moving machine learning models into production and monitoring their effectiveness. This increases the productivity of the developers, makes them more agile and means that machine learning can more readily be productionized. This makes the data science process more efficient and delivers increased business benefits.

## Drivers

- In-DBMS analytics offerings have been available from on-premises data warehouse software vendors for many years and are now increasingly featured in cloud DBMS which is increasing acceptance and adoption.
- Machine learning is becoming more commoditized as its use spreads beyond specialist data scientists, in-DBMS machine learning is an excellent enabler for this wider group of developers and users.
- Most of today's DBMS vendors are offering in-DBMS analytics capabilities with various ML libraries. Some analytics vendors such as SAS and IBM (SPSS software) can push their analytics processing down into a suitable DBMS. Also, some vendors such as Alteryx and Fuzzy Logix provide analytics libraries that can be used with DBMS from more than one vendor.
- The drive for greater productivity in the use of machine learning, ease of administration and the need to reliably move machine learning into production is encouraging adoption. Adopting in-DBMS analytics provides a very good solution for moving analytic models to production with model generation, administration and execution all in the same environment.

## Obstacles

- There has been a lack of familiarity with data management tools, including DBMSs, among data scientists plus a lack of familiarity with ML among DBMS professionals. Data scientists have tended to prefer R, Python, and notebooks (Jupyter, Apache Zeppelin), DBMS practitioners SQL.
- In-DBMS analytics requires a sufficient range of analytical algorithms. This is now much easier than was previously possible, and in fact is becoming the norm. However, using organizations still need to validate how they will fit into their overall estate and most importantly how the analytics will be monitored and controlled.
- Some implementations are restricted in performance and their ability to scale. To be used at scale the algorithms do not just need to be made available but to be modified to take advantage of parallel processing. This is not a problem with most offerings, but needs to be checked in a proof of concept prior to adoption.

## User Recommendations

Data and analytics leaders should:

- Contemplate in-DBMS analytics as a viable option for making large-scale business analytics available to a wider audience. In-DBMS analytics embeds machine learning capabilities in familiar platforms that can deliver rapid insights on both historical and incoming data. By avoiding the need to move data out of the DBMS to build analytic models, in-DBMS analytics allows for more flexible experimentation and efficient development.
- Review your data science development process. Evaluate whether it can be better enabled through in-DBMS analytics, especially for deployment which can be much easier with in-DBMS analytics.
- Check whether in-DBMS analytics is supported when evaluating DBMS systems and, if so, the range of algorithms offered. Experiment with use cases where it is more efficient bringing ML algorithms to the data at scale.

## Sample Vendors

Fuzzy Logix; Google; IBM; Micro Focus; Oracle; SAP; Teradata; VMware

## Gartner Recommended Reading

[5 Useful Ways to Use Artificial Intelligence and Machine Learning With Your Logical Data Warehouse](#)

[Magic Quadrant for Data Science and Machine Learning Platforms](#)

[Magic Quadrant for Cloud Database Management Systems](#)

[Why In-DBMS Analytics Deserves a Fresh Look](#)

## Data Integration Tools

Analysis By: Ehtisham Zaidi, Mark Beyer, Robert Thanaraj

Benefit Rating: High

Market Penetration: More than 50% of target audience

Maturity: Early mainstream

### Definition:

Data integration tools enable the design and implementation of data access and delivery capabilities that allow independently designed data structures to be leveraged together. Data integration tools have matured from supporting traditional bulk/batch integration scenarios to now supporting a combination of modern delivery styles (such as data virtualization and stream data integration). Data integration tools are now expected to support hybrid and multicloud integration scenarios.

### Why This Is Important

Data integration tools are needed by organizations to support distributed data management and deliver data at all latencies across a range of use cases. These include MDM, analytics, data science, data warehousing, and multicloud and hybrid cloud integration.

Data integration tool suites are expected to deliver simpler interfaces to support less-skilled roles like citizen integrators. Growing requirements for automated data integration also require support for data fabric architectures.

## Business Impact

Organizations adopting mature data integration tools increasingly exploit comprehensive data access and delivery capabilities. They get immediate benefits in the form of:

- Reduced time to integrated data delivery
- Cost savings (by reduced integration technical debt)
- Quality enhancements (for analytics/data science products)
- Flexibility (to access new data sources)

Integration tools that support data fabric designs will increase the productivity of data engineering and data science teams.

## Drivers

- Ability to execute data integration in a hyperconnected infrastructure (irrespective of structure and origins) and the ability to automate transformations through embedded ML capabilities are the most important drivers for organizations investing in modern data integration tools.
- Traditional data integration architectures and tools (which focused solely on replicating data) are slow in delivering semantically enriched and integrated datasets that are ready for analytics. This exacerbates data integration tools to provision a mix of variable latency, granularity, physical and virtualized data delivery. This is another major reason to invest in these tools.
- Activities for self-service data access and data preparation by skilled data engineers, citizen integrators and other non-IT roles spur requirements for new data integration tools.
- While traditional data integration tools have now become mature on technical metadata ingestion and analysis to support data integration activities, there is still room for maturity for data integration vendors to introduce capabilities to harness and leverage “active” metadata. Organizations must therefore investigate and adopt data integration tools that can not only work with all forms of metadata, but also share it bidirectionally with other data management tools (e.g., data quality tools) to support data fabric architectures for automation.
- Dynamic data fabric designs bring together physical infrastructure design, semantic tiers, prebuilt services, APIs, microservices and integration processes to connect to reusable integrated data. Vendors will continue to add data integration functionality or acquire technology in these areas.

## Obstacles

- Tightly integrated data integration tool suites in which all components share metadata (both active and passive), design environment, administration and data quality support remain an area for improvement in the data integration tools market.
- The popularity of data preparation (and other self-service ingestion tools), with the sole focus on analytics use cases demonstrated, will create some confusion in the market, slowing the advance of data integration tool suites.
- The demand for a seamless integration platform that spans and combines multiple data delivery styles (batch with data virtualization, for example), multiple deployment options (hybrid and multicloud) and multiple personas currently exceeds the capabilities of most offerings.
- Most existing data integration tools are limited in their ability to collect and analyze all forms of metadata to provide actionable insights to data engineering teams to support automation.

## User Recommendations

- Assess your data integration capability needs to identify gaps in critical skill sets, tools, techniques and architecture needed to position data integration as a strategic discipline at the core of your data management strategy.
- Review current data integration tools to determine if you are leveraging the capabilities they offer. These may include the ability to deploy core elements (including connectivity, transformation and movement) in a range of different data delivery styles driven by common metadata, modeling, design and administration environments.
- Identify and implement a portfolio-based approach to your integration strategy that extends beyond consolidating data via ETL to include stream data integration, event recognition and data virtualization.
- Make automation of data integration, ingestion and orchestration activities your primary goal for the year, and focus on those data integration tools that can support data fabric designs.

## Sample Vendors

Denodo; Fivetran; HVR; IBM; Informatica; Matillion; Precisely (Syncsort); Qlik; Talend; TIBCO



## Gartner Recommended Reading

[Magic Quadrant for Data Integration Tools](#)

[Critical Capabilities for Data Integration Tools](#)

[Market Share Analysis: Data Integration Tools, Worldwide, 2019](#)

[Position Your Product to Benefit From the Rise of Intercloud Data Integration](#)

[Data Fabrics Add Augmented Intelligence to Modernize Your Data Integration](#)

## iPaaS for Data Integration

Analysis By: Eric Thoo

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

### Definition:

Integration platform as a service (iPaaS) technologies provide data integration capabilities such as extraction, transformation and loading (ETL), data replication and data virtualization as cloud services. The use of iPaaS technology for data integration applies to diverse scenarios, including the integration of data in cloud, multicloud, on-premises, between enterprises, at the edge and hybrid deployments.

### Why This Is Important

Increasingly, iPaaS acts as a strategic component, or as an extension to data integration infrastructure, to enable a hybrid integration platform strategy. Organizations experienced in dealing with SaaS and other cloud services are simplifying deployment and elastically scaling computing environments for workloads often not well-handled by existing data integration infrastructure.

### Business Impact

The use of iPaaS for data integration offers benefits for most organizations.

- iPaaS relieves some of the common challenges concerning flexibility, developer productivity, skills availability and entry cost.
- Resource-constrained businesses can apply data integration capabilities offered in iPaaS to targeted issues — such as the synchronization of data between on-premises and off-premises applications and the composition of data sources in cloud and on-premises together.

## Drivers

- The increasing interest in cloud-based solutions for data warehousing and analytics, requirements for access to, and delivery of datasets using cloud environments presents growing opportunities to leverage iPaaS for data integration.
- The adoption of hybrid and multicloud deployment models, cloud data stores, and cloud data ecosystems in general, as well as the movement of enterprise data into cloud applications, fuels demand for using cloud services for data integration.
- On-premises data integration tool vendors are actively competing in the iPaaS market. These providers target specific data integration opportunities, such as integrating SaaS endpoints and supporting IoT and multicloud solutions via a cloud-native platform or by offering an iPaaS rendition of existing on-premises integration platforms. In many cases, users of iPaaS favor offerings that support both data and application integration within a single toolset.
- Midsize organizations — plus business roles and citizen integrators outside of IT in larger organizations and teams focused on event-driven IT capabilities — are using these capabilities to move and synchronize data that includes cloud endpoints.
- Some SaaS providers embed an iPaaS to make it easier and faster to integrate their services with the rest of the application portfolio customers use. As the use of hybrid delivery models continues to grow, organizations are increasingly considering iPaaS as either a strategic component or an extension of their data integration infrastructure to enable a hybrid integration platform strategy.

## Obstacles

- iPaaS may not address the full scope and complexity of broader data integration requirements, and adoptions often reflect a selective, targeted scope.

- Many deployments are for purpose-built integration problems involving cloud data, where sources and targets are mainly common structures. This addresses data residing in popular SaaS, cloud object stores, cloud data warehouses or operational databases, rather than extremely large-scale bulk/batch workloads such as on-premises enterprise data warehouses.
- Some iPaaS providers are limited in their partnering and metadata extensibility to operate with broader data management capabilities. As a result, many iPaaS deployments are not applied to data integration processes that must be governed and traceable via data lineage or metadata management.
- Migrating data from one part of the cloud to another, or into internal applications, may elicit data integrity or quality issues that aren't necessarily addressed by an iPaaS.

## User Recommendations

- Use iPaaS for data integration as an extension of the organization's data integration infrastructure and as an enabling technology for hybrid integration platform capabilities.
- Use iPaaS as a way of supporting data integration workload involving cloud integration, accelerating time to value, minimizing costs and resource requirements relative to on-premises models.
- Empower less-technical roles, such as citizen integrators or business analysts, to perform data integration using iPaaS.

## Sample Vendors

DBSync; Etlworks; Fivetran; Informatica; Matillion; Microsoft; Qlik; SnapLogic; Talend; TIBCO Software

## Gartner Recommended Reading

[Magic Quadrant for Enterprise Integration Platform as a Service](#)

[Magic Quadrant for Data Integration Tools](#)

## Multimodel DBMS

Analysis By: Merv Adrian

**Benefit Rating:** Moderate

**Market Penetration:** More than 50% of target audience

**Maturity:** Mature mainstream

**Definition:**

Multimodel database management systems (DBMS) incorporate several data engines, relational and/or nonrelational (e.g., document, graph, key value, time series, wide column), in a single DBMS. They provide a common access mechanism for different persistence types, each optimized for the nature of the data being used.

**Why This Is Important**

Data and analytics professionals are confronted with increasing choices among DBMS types, especially in the cloud. Choosing multimodel and specialty DBMS can influence the direction of your cloud architecture, staffing and vendor management. Multimodel DBMS can simplify development and integration, and make performance predictable. However, they create new skills requirements of their own, including adding the complexities of different modeling requirements.

**Business Impact**

Although multimodel is rarely used as a leading marketing message, it helps anchor incumbent products. Multimodel DBMS are in place in the overwhelming majority of organizations, even if the organization did not make an explicit choice to adopt them. Embracing them can accelerate the introduction of new technologies, reduce operational complexity and minimize management of a portfolio of vendors — although it can potentially expand dependence on a single vendor.

## Drivers

- Incumbency and familiarity are key drivers of further adoption; some of the most widely deployed DBMSs are multimodel. Every leading vendor now has a widely adopted flagship offering that supports two, three or more types.
- Managing fewer vendors is always an attractive outcome for most organizations.
- Commercial barriers to more usage of the same product are lower in a usage-based licensing environment where no new negotiations and contracts need to be considered.
- Fewer integration issues compared to multiple services-based DBMS also make multimodel an attractive candidate for new development. In addition, older multimodel products often provide more mature auditing, concurrency controls, versioning, distributed data complexity management, points of governance and security that specialty products lack.
- Newer multimodel offerings that are experiencing rapid growth in the cloud often are gaining a foothold because of developer preferences. Putting new capabilities closer to existing practices promotes experimentation and lowered barriers to rapid deployment.

## Obstacles

- Paradoxically, familiarity with a product does not always mean familiarity with its capabilities. Gartner inquiries regarding multimodel DBMS remain rare, reflecting minimal awareness — it is an analysts' abstraction, not a market label. Oracle's "Converged" message is one of the few explicit promotions of the idea by a leading vendor.
- The approachability of specialized products makes them attractive for different modes such as graph, with design and development tools more immediately productive for new users. Multimodel offerings may seem less friendly to new developers who are not already expert in their more traditional uses.
- Multimodel offerings do not often target specialized features and performance; they are marketed for a broader set of design criteria and thus may not have focused there. Thus, multimodel offerings may need to prove their ability to meet these requirements to skeptical advocates of "new" engines.

## User Recommendations

- Assess multimodel capabilities of your incumbent DBMSs first, especially when planning augmented transaction processing and logical data warehouse deployments due to their multipurpose capabilities, which may simplify development challenges.
- Leverage modest “different” requirements for a use case where it may be met with an incumbent multimodel product. Evaluate alternative products for a full blown use case for a specialty such as graph.
- Use multimodel DBMSs to create a stable, mature Mode-1-style platform that is suitable for responding to and stabilizing agile, Mode 2 explorations of alternative data paradigms.
- Plan for internal advocacy and training because skills in multimodel DBMSs are lacking. Demand that vendors provide assistance with the transition to new models of design and deployment.
- Choose vendors whose stability and track record are demonstrable, and whose roadmaps are consistent with your planned use cases.

## Sample Vendors

Amazon Web Services; Couchbase; DataStax; EDB; Google Cloud Platform; IBM; MarkLogic; Microsoft; MongoDB; Oracle; SAP

## Gartner Recommended Reading

[Choosing Between Multimodel DBMS and Multiple Specialized Engines](#)

[The Impact of Modern Data Solutions on IT Modernization Decisions](#)

[IT Market Clock for Database Management Systems, 2019](#)

## Wide-Column DBMS

Analysis By: Merv Adrian

Benefit Rating: Moderate

Market Penetration: 20% to 50% of target audience

Maturity: Mature mainstream

**Definition:**

Wide-column database management systems (DBMS) store rows of data in tables, similar to a relational DBMS, but in which not all fields in a row may be present, without referential integrity and often without support for JOINS. Their flexible schema definitions make them popular for storing semistructured data, like log and sensor data, and relaxed consistency makes them attractive for geographically distributed deployments with challenging SLAs.

**Why This Is Important**

Wide-column databases are increasingly used by enterprises processing growing amounts of structured and unstructured data in distributed data collection and transactional scenarios.

**Business Impact**

The current business impact of wide-column DBMS is moderate. The ability to distribute massive amounts of semistructured data addresses several existing and emerging use cases, particularly in support of more complex Internet of Things (IoT) initiatives. Although vendors offering products, training and support are driving adoption and accelerating development, multimodel DBMS may provide similar functions with more flexible opportunities — many include wide-column capability.

**Drivers**

- Cloud-based applications often require ingesting very large data volumes with a scalable and highly available architecture. Wide column stores are effective for use cases that require these characteristics.
- Significant efforts by hyperscalers to support and market wide-column offerings are increasing visibility beyond the sizable existing user base. These include Microsoft's addition of the Cassandra API to its CosmosDB, AWS's introduction of Keyspaces and Google's support for DataStax Astra in addition to its own Cloud Bigtable.
- The complexity of deployment and operation, formerly a challenge for most wide-column offerings, is being reduced in the cloud with improved and often automated, even serverless operations.
- The pace of functional enhancement will pick up as cloud vendors battle for differentiation and deliver new releases more rapidly.

## Obstacles

- Because wide-column DBMSs do not typically support relationships between rows or tables, they are unsuitable as a platform for data warehousing and complex analytics applications.
- The increasing pace of multimodel adoption is challenging products that cannot support both operational and analytical workloads. New projects will increasingly be deployed in more broadly capable offerings, reducing the expansion of wide-column DBMS into new use cases.

## User Recommendations

- Model expected workloads and data volume. The right hardware or cloud instance types for wide-column databases are critical for cost-effective delivery to your SLAs.
- Focus on operational use cases; the absence of JOINS degrades analytics performance.
- Develop the right data model to enable decisions between consistency and latency. Choose DBMS solutions that allow such trade-offs to be made; developers must incorporate this into applications.
- Eliminate network bottlenecks preventing efficient distributed operations, including replication. Note: conflict resolution requires time synchronization across the cluster and between applications interacting with individual nodes.
- Leverage open-source offerings for experiments and pilots. Engage with vendors offering commercial support before moving to the production stage.
- Investigate the security capabilities of wide-column DBMS before implementing applications with significant risk profiles, since these features still lag behind mature RDBMS.

## Sample Vendors

Alibaba Cloud; Amazon Web Services (AWS); Cloudera; DataStax; Google; Microsoft; Scylla

## Gartner Recommended Reading

[Assessing the Optimal Data Stores for Modern Architectures](#)

[Assessment of DataStax Enterprise With Cassandra](#)





## 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](#)

## Content Migration

Analysis By: Gavin Tay

Benefit Rating: Moderate

**Market Penetration:** More than 50% of target audience

**Maturity:** Mature mainstream

**Definition:**

Content migration refers to the process of consolidating and transferring unstructured content (files, documents, objects) stored permanently in one or more content repositories to a new environment (cloud content services), along with related metadata, permissions, certificates, structure and linked components. During the migration process, enterprises typically choose to cleanse their content repositories by archiving old and outdated content.

**Why This Is Important**

The demand for, and complexity of, content migration have increased sharply as organizations reevaluate their content services platform (CSP) investments due to market consolidation, new work hub and digital workplace initiatives. The associated complexities include content distributed across myriad repositories, on-premises and in the cloud. Common use cases are file-share elimination and merger and acquisition/divestiture activity (including multicloud tenant consolidation or splits).

**Business Impact**

Content migration technology isn't only about pushing information and users toward cloud platforms. Cloud office migrations have expanded their scope toward content classification, governance and compliance, to ensure such shifts offer an exit strategy subsequently. Companies in highly regulated industries, such as financial services and health sciences, will benefit from stringent, auditable and automated processing using content migration tools to account for all their content all the time.

## Drivers

- At present, migration tools (such as content collaboration tools [CCTs] or cloud content services) that support the movement of large volumes of content from end-of-life repositories or shared drives to newer technologies are a priority.
- Content is also increasingly becoming distributed and progressive; federation of content from CSPs to cloud alternatives and vice versa will become the norm going forward.
- While migration tools add immediate value, some organizations are choosing not to migrate content given the cost involved and the lack of available IT budget. In these cases, purpose-built content migrators will transform content in accessible repositories instead, and also allow for content integration and federation rather than migration.
- With content migration going mainstream, file analysis tools are taking center stage. File analysis tools analyze, index, search, track and report on detailed file metadata and file content. This enables organizations to take action on current and legacy files and objects according to what was identified, thereby improving information governance and organizational efficiency for unstructured data management.

## Obstacles

- During the past year, multicloud tenant consolidation and splits were a focus. Moreover, as cloud content becomes commonplace, cloud-to-cloud migration efforts are set to increase. Content migration tools have evolved from on-premises to cloud scenarios to enable such migration. However, absence of standards and uncertainty of service availability still render cloud-to-cloud migrations complex.
- The complexity of cloud-to-cloud migrations is further exacerbated by the fact that hybrid content architectures are left to inherently manage the security risks associated with the integration of public cloud or hosted repositories.

## User Recommendations

- Evaluate opportunities to employ in-house IT expertise with the acquisition of migration tools. Alternatively, hire system integrators with their team of experts that would use sanctioned migration tools and frameworks. Using in-house staff may absorb the cost involved, but external expertise should ease the rigorous efforts required to perform complex migrations and bring valuable experience to the process.
- Evaluate CSP and CCT vendors' built-in migration options for a common set of legacy repositories. Identify limiting factors of in-built tools which inhibit the fulfillment of content transformation expectations. Such expectations include improving and evolving content governance iteratively, or introducing an organizational taxonomy driven by business outcomes.
- Decide on your migration approach based on business needs. Many organizations have been highly successful by simply archiving old environments and moving to a new "greenfield" environment, without bringing across any baggage.

## Sample Vendors

AvePoint; Cloud FastPath; CloudM; Proventeq; Quest Software; ShareGate; Simflofy; SkySync; T-Systems (Vamosa Technologies); Xillio

## Gartner Recommended Reading

[Market Guide for Cloud Office Migration Tools](#)

[Toolkit: Data Slicer to Derive a Cloud Office Migration Tools Shortlist](#)

[Market Guide for File Analysis Software](#)

[Magic Quadrant for Enterprise Information Archiving](#)

[Critical Capabilities for Enterprise Information Archiving](#)

## Data Virtualization

Analysis By: Ehtisham Zaidi

Benefit Rating: Moderate

Market Penetration: More than 50% of target audience

**Maturity:** Mature mainstream

**Definition:**

Data virtualization is one style of data delivery in the broader data integration tools market. It is based on the execution of distributed queries against heterogeneous data sources, federation of query results into cached virtual views, and consumption of these views by consuming applications. It can be used to create integrated views of data in-memory (rather than physically storing views in a target data store) and provides a layer of abstraction above the physical implementation of data.

**Why This Is Important**

Data virtualization is an established data delivery style that enables federated and real-time data access, integration and sharing. It is important because it enables organizations to quickly connect, integrate and deliver data without further replication or creation of more data silos. Data virtualization, therefore, improves flexibility to integrate new data sources and targets, reduces costs by not having to copy and store data, and improves change management processes.

**Business Impact**

Businesses that identify a demand for rapid accessibility to data sources and targets for their data and analytics use cases will find data virtualization to be an important tool in their portfolio. Data virtualization will vastly reduce the time, effort and cost needed to enable faster time to integrated data delivery of siloed data sources. Data virtualization will be beneficial to data engineering teams who want to get quick access to data sources and need cost-effective change management.



## Drivers

- **Flexibility:** Data virtualization adds flexibility to enterprise data integration architectures by allowing users to virtualize and integrate inherently different data models (without physically moving data into a repository), as well as experiment with unstructured and diverse types of data (IoT data, time series data, etc.) much faster.
- **Data access for experimentation:** Less technical roles such as citizen integrators can participate in federating or incorporating data access and usage from a business perspective to ensure favorable application outcomes. A variety of data, federated as an integrated set and provisioned as a “sandbox,” can give citizen integrators the ability to do Mode 2-based experimental integration and get data for ad hoc queries.
- **Reusability:** Being increasingly relevant as a data service layer, data virtualization allows organizations to benefit from increased reusability of integration artifacts across a broad variety of business needs. These needs include logical data warehousing, virtualized operational data stores and registry-style master data management (MDM) scenarios. Data virtualization therefore enables cross-platform analysis of data, such as combining historical data from your enterprise data warehouse with data from an external third-party data source for answering a specific query.
- **Cost:** Data virtualization tools are not necessarily cheaper than ETL tools. However, it is the high cost of developing and maintaining integration flows using traditional ETL tools that makes data virtualization an enticing option to reduce ongoing costs. The key drivers of using data virtualization (regarding costs) include shorter delivery cycles as well as reduced ongoing maintenance and change management.
- **Data sharing:** Data virtualization assists organizations with integrated data access and sharing especially when there are regulatory constraints around physically moving data.

## Obstacles

- Some organizations still think of data virtualization as a silver bullet to all data integration challenges. They use data virtualization as a replacement to ETL which leads to poor performance and thus poor adoption.
- Applications such as analytics and even embedded virtual tiers within data preparation tout their semantic virtual tiers (which have limitations to performance optimization) as a replacement to data virtualization. This hype around semantic virtual tiers has led to performance challenges.
- When data virtualization is not used in conjunction with other data delivery styles (e.g., bulk/batch with ETL), it might limit performance optimization and inhibit adoption.
- Data virtualization continues to be met with resistance by some data management teams because of an internal resistance to new ways to integrate data that challenge conventional wisdom. Organizations that are least open to “experimental and fail-fast” culture are likely to resist data virtualization.

## User Recommendations

- Investigate your existing data integration tools, BI tools or DBMSs for their embedded data virtualization capabilities. Approach a stand-alone data virtualization tool in support of advanced data virtualization capabilities.
- Make data virtualization a must-have data integration style within the broader data integration portfolio to assist with agility and cost optimization.
- Evaluate the current state of your data integration. Set proper expectations, select the right use cases and document the agreed-upon SLAs to separate out when to collect data versus connecting to it using data virtualization.
- Be prepared to respond to monitoring and auditing of data virtualization jobs to determine when they have evolved toward commonly shared models. Create a plan for moving that common virtual model to a more traditional physical data integration process for consolidation. For this purpose, evaluate if your data virtualization tool has the ability to materialize virtual views.

## Sample Vendors

Data Virtuality; Denodo; Dremio; Gluent; IBM; TIBCO Software

## Gartner Recommended Reading

[Magic Quadrant for Data Integration Tools](#)

[Critical Capabilities for Data Integration Tools](#)

[Market Guide for Data Virtualization](#)

[Toolkit: Assessing the Applicability of Data Virtualization to Your Data and Analytics Use Cases](#)

[6 Things to Get Right for the Logical Data Warehouse](#)

## In-Memory Data Grids

Analysis By: Massimo Pezzini

Benefit Rating: High

Market Penetration: More than 50% of target audience

Maturity: Mature mainstream

### Definition:

In-memory data grids (IMDGs) provide a distributed, reliable, scalable, and consistent in-memory object store — the data grid — that is shareable across multiple distributed applications. IMDG-based applications concurrently perform transactional and/or analytical operations in the low-latency data grid, thus drastically reducing the use of conventional, high-latency storage. IMDGs maintain data grid consistency, availability and durability via replication, partitioning, and persistent storage.

### Why This Is Important

The move toward digital is forcing organizations to update their architectures to sustain the escalating workloads stemming from mobile apps, customer portals and the Internet of Things. Although well known only by a minority of software engineering leaders, IMDGs are mature enablers for the scalable, high throughput and low latency data access needed to enable digital architectures. Software engineering leaders engaged in digital initiatives should put IMDG technology in their radar screen.

## Business Impact

By facilitating scale-out architectures, IMDGs (stand-alone or embedded in other software) enable users to implement innovative, hyper-scale, high-performance and low-latency applications based on architectures that cannot be supported by traditional tightly coupled platforms. These architectures include augmented transaction, event brokering and digital integration hub. IMDGs can also enhance traditional application scalability, performance, throughput and cross-application coordination.

## Drivers

- The requirements for scale-out architectures to support low-latency/high-throughput digital scenarios as well as new use cases, such as stream analytics and digital integration hubs, will continue to drive IMDG growth.
- IMDGs are increasingly bundled into software products and cloud services, which favors adoption, albeit often unrecorded.
- Technology is proven and mature in terms of functionality, manageability, high availability and support of complex use cases. During the 15+ years since when this technology has emerged, providers have accumulated several thousands of clients worldwide across multiple industry verticals. They successfully leverage IMDG to enable large-scale business-critical systems, typically hyper-scale, high-performance transaction processing applications, such as digital banking, financial trading, e-commerce and travel reservation systems.
- Atop the classic caching-oriented capabilities, IMDG products often provide in-memory DBMS, stream processing and real-time analytics capabilities, thus extending their use beyond the traditional advanced caching and high-performance transaction processing scenarios.
- Some of the most popular and widely used IMDG products are now available also as open source software, which encourages use not only by user organizations, but also by ISVs and cloud service providers.
- Technologies such as containers make IMDG scale-out deployments easier, thus further fostering adoption by mainstream organizations.

## Obstacles

- Issues such as small skills pool, little industry support, deployment and management complexity, data stewards' conservatism have historically inhibited and will continue to limit IMDG penetration in the less technically skilled and more risk averse organizations.
- Several general-purpose vendors either withdrew from the market or scaled down their commitment to IMDG, although some still enjoy a large and profitable installed base. The lack of their marketing push is reducing IMDG visibility among mainstream organizations and ISVs.
- Enthusiastic and passionate pure play vendors are the key IMDG players. However, albeit in the market for many years, they are considered by mainstream organizations only in very special cases because of the inevitable viability concerns stemming from these providers' small size.
- The competition from traditional DBMS that now support in memory techniques, which are good enough for many applications without the additional complexity of an IMDG.

## User Recommendations

Adopt IMDGs to:

- Boost established application performance and scalability. This is enabled by IMDG support for popular APIs such as REST, Hibernate, JPA, JDBC and, increasingly, also SQL.
- Enable high-scale/high-performance use cases such as the digital integration hub, which decouples APIs service layers from system or records via an intermediate high-performance data store.
- Develop native hyperscale systems, for example business platforms to support vast ecosystems.
- Enable scenarios that require high-performance collaboration between applications or microservices.
- Support low-latency event processing and real-time analytics applications.
- Implement augmented transaction applications that support stream analytics use cases.

Make sure you have access to IMDG architecture and technology skills, whether internal or from an external service provider. IMDG is sophisticated technology that requires advanced skills to be properly leveraged.

## **Sample Vendors**

GigaSpaces; GridGain Systems; Hazelcast; Oracle; Red Hat; ScaleOut Software; Software AG; TIBCO Software; VMware

## **Gartner Recommended Reading**

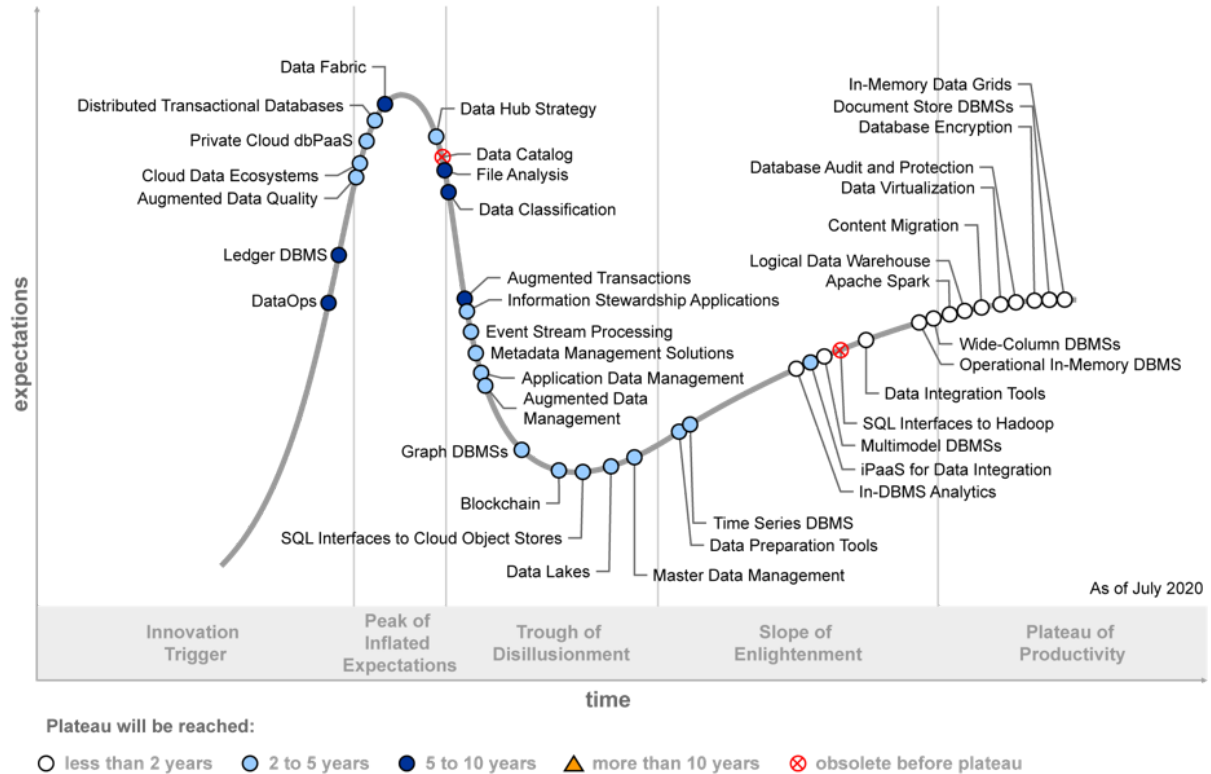
[Innovation Insight: Turbocharge Your API Platform With a Digital Integration Hub](#)

[Forecast Analysis: Application Infrastructure and Middleware Software, Worldwide](#)

## Appendixes

Figure 2. Hype Cycle for Data Management, 2020

### Hype Cycle for Data Management, 2020



Gartner

Source: Gartner (July 2020)

## Hype Cycle Phases, Benefit Ratings and Maturity Levels

**Table 2: Hype Cycle Phases**

(Enlarged table in Appendix)

Phase ↓	Definition ↓
<i>Innovation Trigger</i>	A breakthrough, public demonstration, product launch or other event generates significant media and industry interest.
<i>Peak of Inflated Expectations</i>	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.
<i>Trough of Disillusionment</i>	Because the innovation does not live up to its overinflated expectations, it rapidly becomes unfashionable. Media interest wanes, except for a few cautionary tales.
<i>Slope of Enlightenment</i>	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.
<i>Plateau of Productivity</i>	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.
<i>Years to Mainstream Adoption</i>	The time required for the innovation to reach the Plateau of Productivity.

Source: Gartner (July 2021)



Table 3: Benefit Ratings

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

Source: Gartner (July 2021)

**Table 4: Maturity Levels**

(Enlarged table in Appendix)

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

Source: Gartner (July 2021)

## Document Revision History

[Hype Cycle for Data Management, 2020 - 15 July 2020](#)

[Hype Cycle for Data Management, 2019 - 31 July 2019](#)

[Hype Cycle for Data Management, 2018 - 25 July 2018](#)

[Hype Cycle for Data Management, 2017 - 26 July 2017](#)

[Hype Cycle for Information Infrastructure, 2016 - 8 July 2016](#)

[Hype Cycle for Information Infrastructure, 2015 - 13 August 2015](#)

[Hype Cycle for Information Infrastructure, 2014 - 6 August 2014](#)

## Recommended by the 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](#)

Data Management Solutions Primer for 2021

2021 Planning Guide for Data Management

Emerging Technologies: Data Fabric Is the Future of Data Management

Magic Quadrant for Cloud Database Management Systems

---

© 2021 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates. This publication may not be reproduced or distributed in any form without Gartner's prior written permission. It consists of the opinions of Gartner's research organization, which should not be construed as statements of fact. While the information contained in this publication has been obtained from sources believed to be reliable, Gartner disclaims all warranties as to the accuracy, completeness or adequacy of such information. Although Gartner research may address legal and financial issues, Gartner does not provide legal or investment advice and its research should not be construed or used as such. Your access and use of this publication are governed by [Gartner's Usage Policy](#). Gartner prides itself on its reputation for independence and objectivity. Its research is produced independently by its research organization without input or influence from any third party. For further information, see "[Guiding Principles on Independence and Objectivity](#)."

Table 1: Priority Matrix for Data Management, 2021

Benefit ↓	Years to Mainstream Adoption			
	Less Than 2 Years ↓	2 - 5 Years ↓	5 - 10 Years ↓	More Than 10 Years ↓
Transformational		Blockchain Event Stream Processing	Active Metadata Management Data Fabric	
High	Data Integration Tools Data Preparation Tools In-DBMS Analytics In-Memory Data Grids Logical Data Warehouse	Augmented Data Cataloging and Metadata Management Augmented Data Management Augmented Data Quality Augmented Transactions Cloud Data Ecosystems Data Classification Data Engineering Data Hub Strategy DataOps Edge Data Management Graph DBMS iPaaS for Data Integration Master Data Management	Intercloud Data Management	D&A Governance Platforms

Benefit	Years to Mainstream Adoption			
↓	Less Than 2 Years ↓	2 - 5 Years ↓	5 - 10 Years ↓	More Than 10 Years ↓
Moderate	Content Migration Data Virtualization Multimodel DBMS Time Series DBMS Wide-Column DBMS	Application Data Management Data Lakes Distributed Transactional Databases Private Cloud dbPaaS SQL Interfaces to Object Stores	File Analysis Lakehouse Ledger DBMS	
Low				

Source: Gartner (July 2021)

Table 2: Hype Cycle Phases

Phase ↓	Definition ↓
<i>Innovation Trigger</i>	A breakthrough, public demonstration, product launch or other event generates significant media and industry interest.
<i>Peak of Inflated Expectations</i>	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.
<i>Trough of Disillusionment</i>	Because the innovation does not live up to its overinflated expectations, it rapidly becomes unfashionable. Media interest wanes, except for a few cautionary tales.
<i>Slope of Enlightenment</i>	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.
<i>Plateau of Productivity</i>	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.
<i>Years to Mainstream Adoption</i>	The time required for the innovation to reach the Plateau of Productivity.

Phase ↓	Definition ↓
---------	--------------

Source: Gartner (July 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 (July 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 (July 2021)