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

Published: 11 September 2019 **ID:** G00378476

Analyst(s): Guido De Simoni, Melody Chien, Ankush Jain, Ehtisham Zaidi

Due to changing requirements, modern data and analytics use cases need a portfolio of capabilities that cannot be fulfilled by existing, stand-alone products. Data and analytics leaders must invest in new data management solutions that leverage aggregated and integrated capabilities.

Impacts

- 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.
- Core data management functionalities (such as data profiling, data cataloging, metadata management and data integration) now appear in many individual data management applications and tools. The distinction among them is getting blurry, driving confusion in the market.
- Organizations now need a portfolio of capabilities driven by metadata and the ability to manage this across a range of use cases, to offer the capabilities that fulfill broader use-case-specific requirements in data and analytics.

Recommendations

For data and analytics leaders focused on modernizing data management solutions:

- Use business realization of the need to align critical data management capabilities as a driver to improve collaboration and communication between business stakeholders and IT.
- Define what you expect vendors to deliver today by using available market capabilities in end-to-end scenarios, to minimize the number of applications/tools deployed.
- Capture how evolving and modern data management platforms are using metadata by learning from and executing on innovative digital initiatives with leading vendors in this space.

Analysis

Modern data and analytics use cases need a portfolio of capabilities driven by their constantly changing requirements, which will not be fulfilled well enough through point-based tools. Digital transformation is a driver of this convergence. Organizations want to have automated, synchronized, integrated, cost-effective and efficient solutions with a central design but distributed deployment. Many of this convergence is happening in the cloud as well.

Demand for data quality tools, data catalogs, metadata management solutions and data integration in one comprehensive solution is growing from both business and IT perspectives. Some core functionalities (such as data profiling and data cataloging) appear in many individual applications. The distinction among them is getting blurred.

Additionally, the need for such aggregation of capabilities or platforms is driven by the growing recognition that the work of data and analytics governance is different to the work of data management. Although the capabilities that serve both are similar, the context in which those same capabilities are used differs between governance and management:

- **Data and analytics governance —** Converged capabilities are used more by business users to set and enforce policy, for example, whereas data management is used to execute that policy.
- **Data management —** Converged capabilities are more aligned to aspects such as integrating business applications or building infrastructure in support of analytical and operational uses cases.

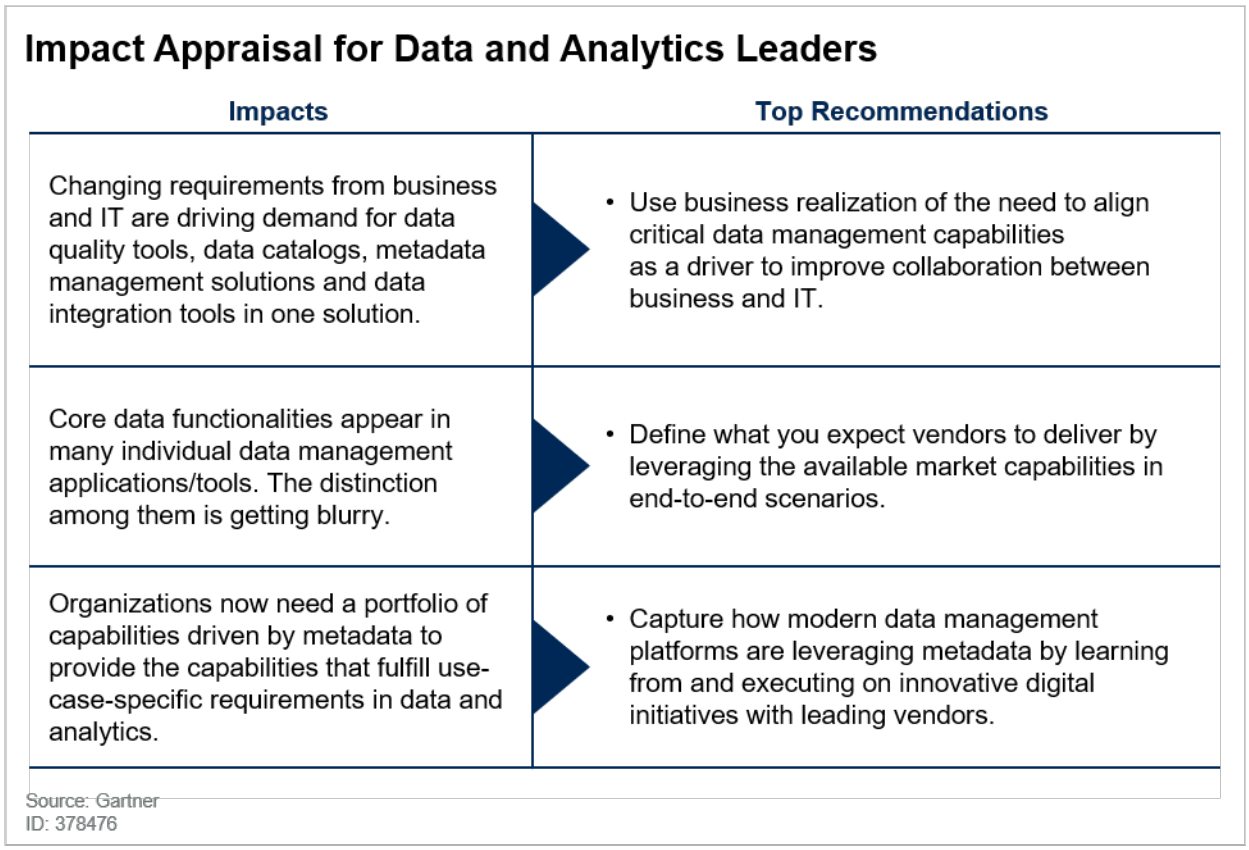
Organizations often need a portfolio of capabilities driven by managed metadata to provide the right capabilities to fulfill broader use-case-specific requirements in data and analytics. Capability-based solutions will likely come together into a modern data management platform serving different and, eventually, all or most use cases. We already see signs of this happening:

- Data quality and data integration tools are incorporating data catalogs.
- Metadata management is mandated in data quality and data integration solutions.
- Data integration tools are incorporating data quality features.
- Analytics and BI tools are incorporating data catalogs.
- Almost all types of data management tools are embedding data preparation capabilities.

We cannot yet say that this represents a market, but it is definitely starting to emerge.

Figure 1 summaries the impacts of this change and the related recommendations.

Figure 1. Impacts and Top Recommendations for Data and Analytics Leaders



Impacts and Recommendations

Demand for Disparate Data Management Tools in One Comprehensive Solution Is Growing

In the 2019 Gartner CEO and Senior Business Executive Survey, 82% of respondents said they now have a digital transformation or management initiative (up from 62% in 2018). Digital business transformation is an effort to create connected, platform and new industry revenue streams (see "Digital Business Ambition: Transform or Optimize?"). In the context of IT, the CEO survey shows that top priorities are innovation, improvement and integration, and data and analytics (see "2019 CEO Survey: The Year of Challenged Growth").

In Gartner's Fourth Annual CDO Survey, the CDO respondents who were already using data assets to generate economic benefit were statistically more likely to be reporting higher overall organizational performance and higher team effectiveness.

Half or more respondents within the survey reported that they are generating this economic benefit using available data assets and through indirect monetization, including:

- Using data assets to improve internal processes (60% of respondents)
- Increasing the value of products/services to customers by infusing data into existing products and services (57%)
- Improving or developing new offerings by innovating with data (50%)

The results also showed that a majority of respondents rated machine learning (76%) and AI (67%) as critical to an organization's data and analytics success.

The primary means by which data and analytics will power digital business is the data and analytics platform (see "How to Make Data and Analytics Central to Your Digital Transformative Initiative"). This supposes a level of interoperability and potential integration between numerous and disparate data and analytics capabilities. There are various levels and forms of consolidation and convergence across the different parts of data and analytics. Data management is one of the larger, more mature platforms in scope of the data and analytics platform.

As such, data management capabilities that are location-agnostic, able to reside and execute anywhere, are growing more valuable. This increases the importance of metadata management, which enables such distributed architecture by providing the knowledge of where data assets of interest reside and how they are related.

The bottom line is that data and analytics leaders are faced with diverse digital business requirements on the one hand and increasingly distributed data and analytics on the other. And the scope of governed data and analytics is enterprisewide, including application data, data lakes, data science labs, analytics and AI.

To succeed with data and analytics initiatives, enterprises must develop a holistic view of critical data management capabilities. These capabilities will power the drivers for a data management platform as well as support the growing and dynamic needs of the analytics/BI and data science/AI platforms, as well as data and analytics governance. Demand for data management capabilities has translated into demand for convergence of capabilities across data management tools and solutions.

The following are two examples of convergence of these capabilities happening now in the market.

1. Data Integration and Data Catalog Tool Convergence

We are witnessing that data catalogs are an important source of enriched metadata that go beyond technical metadata to also provide enriched active metadata (see "Augmented Data Catalogs: Now an Enterprise Must-Have for Data and Analytics Leaders"). This active metadata (see Note 1) is best utilized when organizations can share it with data integration and data quality tools to inform and, in some cases, even automate integration design.

With the emergence of data catalogs and their usage in inventorying and scanning metadata — as well as often collecting important metrics on usage — data integration tooling can definitely be enriched with actual metadata-based design. This can occur through better allocation of schema and semantics from the business, making the overall design much more dynamic.

Hence, there is a definite case and market gap for convergence of data catalogs with data integration platforms through better sharing of metadata bidirectionally. This will allow future integration artifacts to be scanned and mapped in the data catalogs for reuse and effective dynamic integration, and vice versa.

2. Data Integration and Data Quality Tool Convergence

Finally, data integration and data quality have long been separate capability-based tools, but significant convergence is happening between them already. Seldom does Gartner see integration requirements where data quality is viewed separately. For most analytics or operational use cases (such as MDM), we already see data integration, metadata management and data quality tooling being sold together under the collective label of "integration." Without data transformations such as mapping, matching, linking, merging and deduplication, actionable data is not going to be provided for analysis.

Augmented data management will be a significant driver of this convergence, wherein machine learning (ML) will act as a way to automate certain repetitive integration and data quality tasks (see "Hype Cycle for Data Management, 2019"). It will do so through analysis of active metadata to secure data pipelines and make them more dynamic for design and runtime, as well as even automate certain integration tasks altogether.

Recommendations for data and analytics leaders:

- Use business realization of the need to align critical data management capabilities as a driver to improve collaboration and communication between business stakeholders and IT.
- Identify critical capabilities across your distributed data management tool portfolio (data integration, data catalogs, data quality and data preparation). Recognize existing and upcoming overlaps, then start working with existing or new vendors that have managed to converge tools to remove these redundancies. (See Gartner's Critical Capabilities research under the Gartner Recommended Reading section.)

The Data Management Market Often Presents Similar Capabilities Across Tools

Every data and analytics use case — regardless of approach and business goals — requires the following data management capabilities:

- **Describe —** Collecting knowledge about data assets, including location, format, quality and potential value.
- **Organize —** Aligning and structuring data assets so that they can be readily found and easily consumed by other use cases. Deciding if data should be structured in a way that conforms to the organization's standards of syntax (format), semantics (meaning) and terminology (use of common terms), or whether the use case allows for local standards. Opting to organize data locally may affect the ability to integrate with other sources or support other use cases.

- **Integrate —** Supporting accessing and ingesting of diverse data types, performing transformations (changing formats and semantics, or combining data, for example), and allowing independently designed data structures to be used together toward a common objective.
- **Share —** Making data available to consumption points. This can mean a single use case or a variety of use cases depending on the trade-offs made for organizing and integrating data.
- **Govern —** Supporting the execution of data and analytics governance. Providing risk assessment, control and compliance as related to data quality, security, privacy and retention. Data governance will need to take a trust-based approach that is no longer a one-size-fits-all, top-down approach, but that adapts to the situation and the level of central governance required.
- **Implement —** Supporting the deployment and execution of the other five capability types. The decision of collecting versus connecting to data only needs to be resolved at implementation. Changes in implementation can also occur over time as the level of usage (or the use case) evolves.

Data and analytics leaders should think of these six data management processes as "common capabilities" — common in the sense that they can be used and reused across any mix of data-related use cases. This is how organizations can achieve a modern data management infrastructure that is more efficient and effective, breaking the model of siloed and initiative-specific deployments of data management capabilities that most organizations have pursued for many years.

Once applied to real-life scenarios, however, these common capabilities are often expanded into more precise or specific capabilities driven by data infrastructure and data management program requirements.

Most organizations need to work on end-to-end scenarios, each requiring often different capabilities for their use-case-specific requirements. These differing requirements are enabled by different technologies, which organizations need in order to really leverage that effort to operationalize and automate (for data governance, and specifically policy management, see "The Role of Technology in Data and Analytics Governance").

Recommendations for data and analytics leaders:

- Define what you expect vendors to deliver today by leveraging available market capabilities in end-to-end scenarios, to prioritize essential data management capabilities while minimizing the number of applications and tools deployed.
- For new projects such as enabling data lakes or moving data to a dbPaaS solution, favor vendors that provide a rich set of capabilities in a metadata-driven platform over selecting many separate vendor tools that are difficult to integrate later.

Organizations Need a Set of Capabilities to Fulfill Broader Use-Case-Specific Requirements

The opportunity for any organization to fulfill its requirements in data and analytics is supported by market trends that exploit and leverage metadata (see "Top 10 Data and Analytics Technology Trends That Will Change Your Business"). Augmented data management, graph technology, AI/ML and data fabric drive the exploitation of metadata. In addition, data warehouse modernization, streaming analytics, 360 solutions (customer, reference, product and supplier), explainable AI and DataOps can benefit from applied metadata management.

This set of requirements is not just supported by metadata management solutions and their capabilities (see "Critical Capabilities for Metadata Management Solutions").

Enterprise Metadata Management

In many organizations, metadata management is still an intensive manual task that is performed by data architects and data modelers, and driven by enterprise architecture requirements for specific use cases. Enterprise metadata management (EMM) takes place when you need to share and govern that metadata between use cases and implementations. All EMM activities are also embedded within other data and analytics practices, such as data integration. Technical EMM such as lineage and impact analysis are common requirements for data integration developers who need to be able to easily develop and maintain the data transformation flows.

Moreover, the EMM practice could prove to be resource-intensive. Ingesting metadata, defining relationships between metadata assets, building business glossaries and documenting metadata were, in the recent past, mostly manual activities performed and consumed by IT audiences. As a result, EMM as a discipline has been invisible to data and analytics stakeholders and the business. The ability for teams to explain its benefits or demonstrate its business value was — and can still be — challenging (see "The State of Metadata Management").

Data Catalog

The term "data catalog" emerged, in its most recent iteration, in reaction to this old-fashioned perception of EMM. The hype around data catalogs has resulted from market demand for greater transparency and shared understanding around data, such as in data lake scenarios or hybrid/multicloud data management scenarios. This has converged with growing demand from users to autonomously assess the level of trust. This is critical to analytical use cases, though less useful to most data and analytics governance use cases.

These demands involve business roles such as data analysts, data engineers or data scientists needing to understand what data is available and what it could be used for. Similarly, on the data governance side, data stewards need to track the uses of data and be able to make informed decisions about the level of governance that should be applied to the data. However, there is much less priority in discovering data that might be useful to govern; the most important data that needs to be governed is already well known and very visible inside core, mission-critical business applications — no catalog is needed here.

However, these new audiences have neither the technical understanding nor the time to perform the manual activities required by traditional EMM approaches. To succeed, therefore, the EMM practice needs to evolve toward greater automation for the harvesting, inventory and assessment of the relationships among the various data assets.

The pervasive use of metadata that could result in automation of many activities (for example, automation of database optimization and tuning) will require an effective approach to active metadata. EMM solutions have a unique opportunity to play a key role in collecting, analyzing and sharing metadata from the overall data management landscape, and turning this metadata into actions. However, given the complexity and the distributed data landscape within organizations, it is unlikely that EMM solutions alone can succeed. Managing metadata is foundational to enable openness and interoperability with the wide set of solutions required to deliver on automation capabilities, which reach outside the scope of each technology in isolation.

Example Capabilities for Specific Use Cases

Here are some examples of required capabilities that could leverage metadata for automation for specific use cases. (Note that these examples are tracking only data preparation, analytics, integration and data governance.)

Data Preparation

- Using active metadata analysis to automate operationalization of self-service data preparation flows.
- Using ML over metadata to automate frequent data preparation tasks, such as joins, filters, transforms, anomaly detection and corrections.

Analytics

- End-to-end pipeline execution and lineage, including streaming, scripts and open APIs.
- Rapid discovery of governed data and automation of provisioning for analytics.
- Collaborative self-service data preparation with advanced transforms, data quality and masking.
- Automated domain discovery, transformation and dataset recommendations.

Integration

- Automated extraction, transformation and loading script generation.

- Graph analytics for content analysis of integrated data objects.
- Metadata analysis for optimization of application strategies.

Data Governance

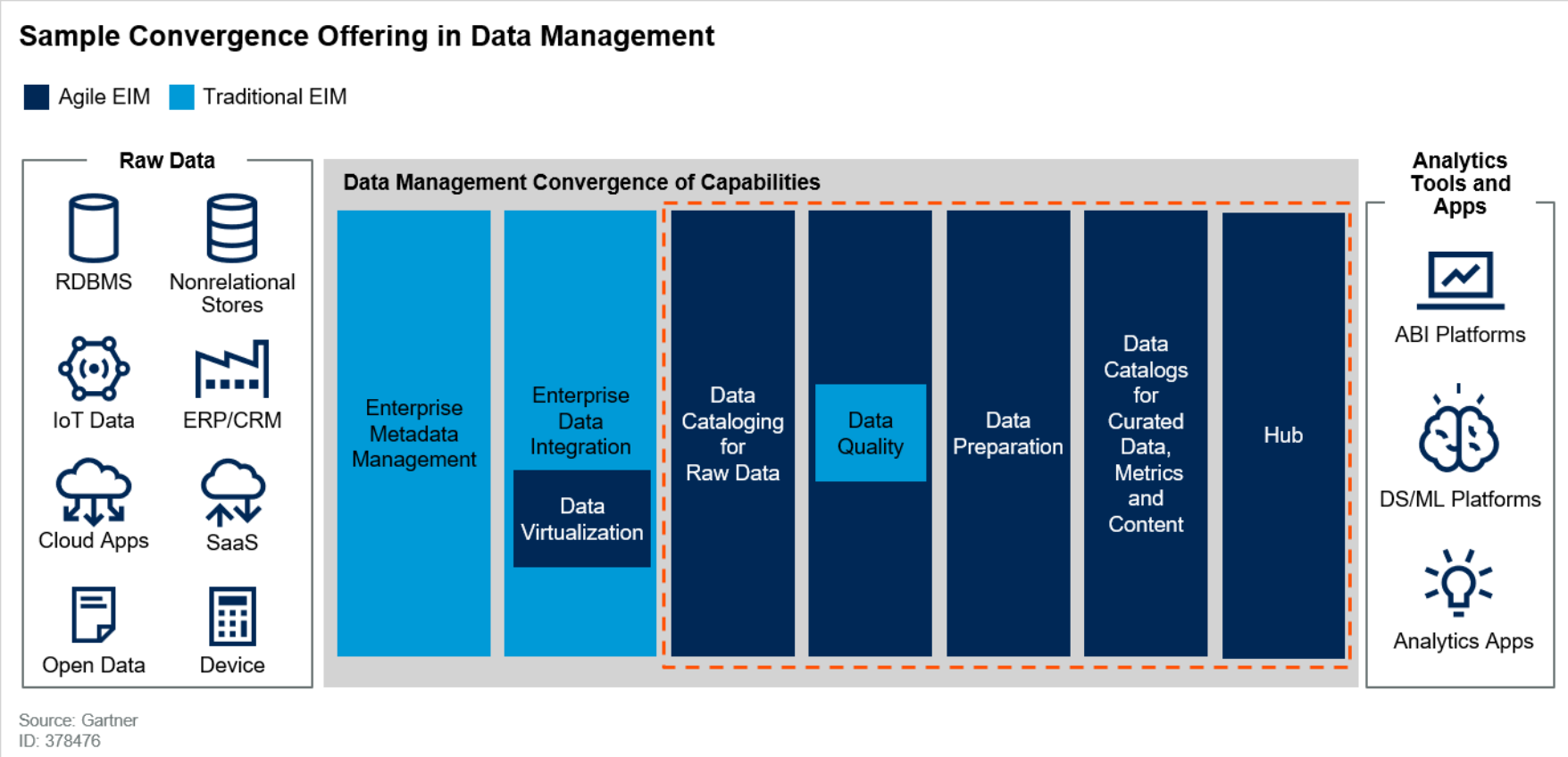
- Autogeneration of data quality rules from rule definitions and automated execution of data quality checks.
- AI-assisted data curation and association of business terms to technical artifacts.
- Automated indexing of personal data in structured and unstructured sources to specific identities.
- Synthesized data proliferation, user activity and anomaly detection to automate risk exposure analysis.
- Automated classification of sensitive data and build subject registry.

Machine learning can leverage metadata for the following (not exhaustive):

- Business rule translation
- Data relationship inference
- Search ranking
- Self-tuning processing
- Smart data visualization
- Data anomaly detection
- Dataset similarity
- Data volume projections
- Dataset recommendations
- Operational anomaly detection
- Schema inference
- Cost of data breach
- Predictive operational analytics

The opportunity does exist, and today the data management market offers a promise of convergence that is presented in Figure 2.

Figure 2. Sample Convergence Offering in Data Management



More and more vendors are combining all these capabilities into one or more converged solutions or packages. They are doing so even if not yet addressing the semantic consistency and interoperability between the parts, nor all the capabilities needed for the data and analytics governance set of use cases.

To see examples, explore the following vendors and solutions:

- Cinchy
- Cambridge Semantics (Anzo)
- CluedIn
- SAS (Viya)
- SAP (Data Hub)
- Talend (Data Fabric)
- Informatica (packaging separate products but tightly integrated and leveraging CLAIRE)

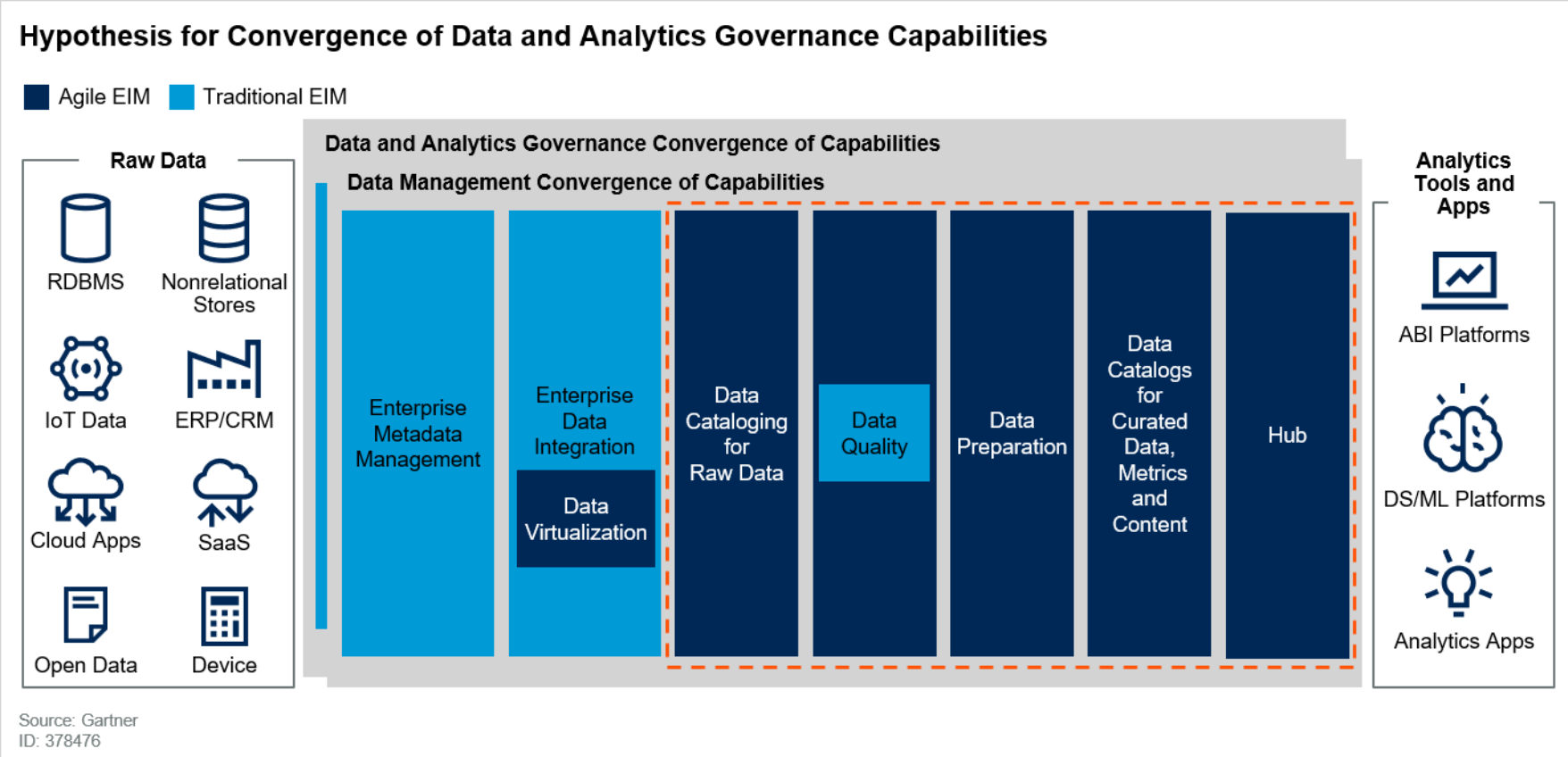
Many of these convergences are happening in the cloud as well. Take advantage of cloud computing and address the challenges presented by increasingly distributed data sources.

Data and Analytics Governance

We also need to clarify that data and analytics governance might drive the same convergence of capabilities but in a different context. This could bring into the picture capabilities offered today by ADM and MDM solutions as well as information stewardship applications (see “Critical Capabilities for Master Data Management Solutions” and “Market Guide for Information Stewardship Applications”).

Figure 3 represents this potential convergence. That “different context” we mentioned cannot be underestimated — the current convergence across data management may or may not satisfy what is needed by organizations across data and analytics governance. It is very possible — indeed likely — that 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 different use cases.

Figure 3. Hypothesis for Data and Analytics Governance Convergence of Capabilities (Overlap, but With Different Context)



Note that we do not yet observe this level of convergence for products and solutions in the market.

Recommendation for data and analytics leaders:

- Capture how evolving and modern data management platforms are leveraging metadata by learning from and executing on innovative digital initiatives with leading vendors in this space.

Acronym Key and Glossary Terms

ADM	application data management
EMM	enterprise metadata management
MDM	master data management

Gartner Recommended Reading

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

"Critical Capabilities for Data Quality Tools"

"Critical Capabilities for Metadata Management Solutions"

"Critical Capabilities for Master Data Management Solutions"

"Critical Capabilities for Data Integration Tools"

"Toolkit: How to Use the Gartner Data and Analytics Glossary to Boost Your Data Literacy"

"Market Guide for Information Stewardship Applications"

Evidence

The fact base for this research derives from users of the Gartner client inquiry service, surveys of IT leaders, workshops and case studies conducted with end-user organizations. The sources include, but are not limited to:

- **Gartner CEO and Senior Business Executive Survey —** Gartner conducted this research from September through December 2018 to examine CEO and senior business executive views on current business issues, as well as some areas of technology agenda impact. In total, 473 business leaders were qualified and surveyed. The research was conducted via an online survey (355); an additional 112 surveys were held through telephone interviews and six were self-administered paper surveys. All respondents were screened for active employment in organizations with $50 million or more in annual revenue. Sixty percent had $1 billion or more and 15% had $10 billion or more in revenue.

The survey was developed collaboratively by a team of Gartner analysts who examine IT's role in business, and was reviewed, tested and administered by the Gartner Research Data and Analytics (RDA) team. The results of this study are representative of the respondent base and not necessarily business as a whole.

- **Gartner's Fourth Annual Chief Data Officer Survey —** This research was conducted via an online survey from September through December 2018. Respondents were required to have the title of chief data officer (CDO) or chief analytics officer (CAO), or to have the responsibilities of an executive-level data and analytics leader in their organization (in the case of organizations without an official C-level data and analytics title). In total, 257 respondents participated in the survey.

  The survey was developed collaboratively by a team of Gartner analysts, and was reviewed, tested and administered by Gartner's RDA team.

- **Client inquiry —** More than 6,700 aggregated conversations with users of Gartner's client inquiry service were taken by the authors from January 2018 through July 2019, related to data quality, data catalogs, metadata management and data integration.

- **Interactive briefings —** Vendors provided Gartner with updates on their strategy, market positioning, recent key developments and product roadmaps.

- **Informal —** Discussions with Gartner colleagues with expertise in this area.

Note 1 Active Metadata

When data management capabilities make extensive use of metadata, leading to automation of many of the data management implementation and ongoing maintenance activities, they turn it into active metadata (that is, taking actions by leveraging metadata).

GARTNER HEADQUARTERS

Corporate Headquarters
56 Top Gallant Road
Stamford, CT 06902-7700
USA
+1 203 964 0096

Regional Headquarters
AUSTRALIA
BRAZIL
JAPAN
UNITED KINGDOM

For a complete list of worldwide locations,
visit http://www.gartner.com/technology/about.jsp