

Capacités critiques pour les solutions de qualité des données

Publié le 11 octobre 2021 - ID G00 734610 - 40 min de lecture

By Ankush Jain, Melody Chien

Le marché des solutions de qualité des données continue de mûrir, les fournisseurs essayant de se différencier en utilisant davantage les métadonnées et l'IA. Cette recherche aide les responsables des données et de l'analyse à évaluer 17 de ces solutions en fonction de l'importance variable de neuf capacités critiques dans sept cas d'utilisation.

Aperçu

Principales conclusions

- Les solutions de qualité des données continuent à évoluer d'un composant autonome vers un ensemble de fonctionnalités intégrées pour répondre aux besoins modernes de gestion des données. Les solutions de qualité des données incluent désormais des fonctionnalités connectées avancées et l'assistance du fournisseur, et elles répondent à une gamme de cas d'utilisation complexes et connectés.
- Les fournisseurs de solutions de qualité des données continuent de se concentrer sur les innovations pour augmenter les utilisateurs professionnels et améliorer la capacité de libre-service de l'organisation utilisatrice. Les fournisseurs augmentent leur utilisation des métadonnées, de l'IA et de l'apprentissage automatique pour améliorer l'automatisation de plusieurs capacités critiques de qualité des données, telles que le profilage, l'analyse, le workflow et la gestion des règles.
- End users have growing needs in data and analytics (D&A) governance and analytics and business intelligence (ABI) use cases. These use cases have extensive focus on business- and self-service-driven capabilities such as profiling, visualization and analytics, workflow, and role-based usability.

Recommandations

As a data and analytics leader assessing data quality solutions, you should:

- Use the Critical Capabilities research to evaluate feasibility of existing and potential future procurement of data quality solutions for use cases ranging from mission-critical data integration and migration to exploratory complex AI and machine learning.
- Explore how metadata, AI and machine learning impacts augmented data quality capabilities of these solutions. Solutions that provide augmented capabilities will be better placed to address needs for self-service, automation, reducing costs linked to data quality programs, and scale and distribution.
- Focus on solutions built not just for identification but also for easier analysis and remediation of data quality issues. This will allow you to unify your business users, augment your data engineers, improve data

literacy, reduce operational challenges and reduce your IT specialist costs. Business usage of data quality solutions determines the success of data quality programs in the long term.

- Combine this research with the companion Magic Quadrant to fully evaluate the vendors and their solutions. Besides functionality, it is important to assess your vendor preference, vendor market presence, availability of skills, support and deployment options, and pricing model.

Strategic Planning Assumptions

- By 2022, 60% of organizations will leverage machine-learning-enabled data quality (DQ) technology for suggestions to reduce manual tasks for data quality improvement.
- Through 2024, 50% of organizations will adopt modern data quality solutions to better support their digital business initiatives.

What You Need to Know

The data quality market, comprising both the demand side and the supply side, is witnessing maturity for many of the traditional requirements. However, there is a growing demand for a higher degree of automation across the data quality activities, as witnessed in some evolving and forward-looking data quality use cases such as data and analytics governance.

The majority of clients have achieved some level of maturity in their reactive approaches and accuracy-driven assessment of data quality (primarily focusing on objective data quality dimensions; see [Drive Data Quality Improvement From a Foundation of Metrics](#)). But organizations struggle when there is a need to evaluate data quality more subjectively, which depends on user opinions, data consumption patterns, understanding data using better context and business team engagement. Organizations also find it difficult to reuse and scale the existing data quality setup for large volumes of data, new datasets and pushing it upstream to be more proactive in their data quality approach.

Vendors are trying to improve the existing solutions to address these requirements by enhanced capabilities (using metadata and AI) for improved context discovery, focus on relevancy, best next suggestions and assistance with remediation of data quality issues. The result is an increase in efficiency and productivity by automating process workflows, minimizing dependency on humans and reducing time to value. That said, end-user organizations should consider the functional characteristics of data quality solutions as an increasingly important dimension of evaluation upon which data and analytics leaders — and other buyers such as D&A governance-related stakeholders — need to focus.

In this research, the ratings allocated to each vendor on the strength of its solutions across nine critical capabilities are without consideration of other vendor characteristics such as marketing strategy or sales execution, which are covered in [Magic Quadrant for Data Quality Solutions](#). The methodology used for the model and the approach to the findings that underpins it are shown at the end of this research (see Note 1: Critical Capabilities Methodology and the Evidence section).

Note that data quality solutions are generally deployed to support other data management processes. They do not exist in their own context like data integration or master data management solutions that have their own primary functions without dependence on secondary tools. So organizations should also take note of the relative importance of each capability to the use cases they are addressing. Some of these capabilities

may not be as important as others for a use case, but this doesn't mean it is not needed. By leveraging the Critical Capabilities ratings and use-case scores, buyers can identify a set of providers that may be the best fit to deliver the functionality necessary to succeed in their efforts to improve data quality.

This Critical Capabilities research assesses 15 vendors, providing 17 data quality solutions, in terms of their data quality capabilities to support the following seven use cases:

- AI and Machine Learning
- Analytics and Data Science
- Data and Analytics Governance Initiatives
- Data Integration
- Data Migration
- Master Data Management
- Operational/Transactional Data Quality

Finally, organizations must also consider each vendor's market presence, track record, financial and organizational strength, availability of skills, product support, pricing and licensing models, and depth of professional services. See the companion [Magic Quadrant for Data Quality Solutions](#) for a high-level vendor assessment that considers many of these additional characteristics.

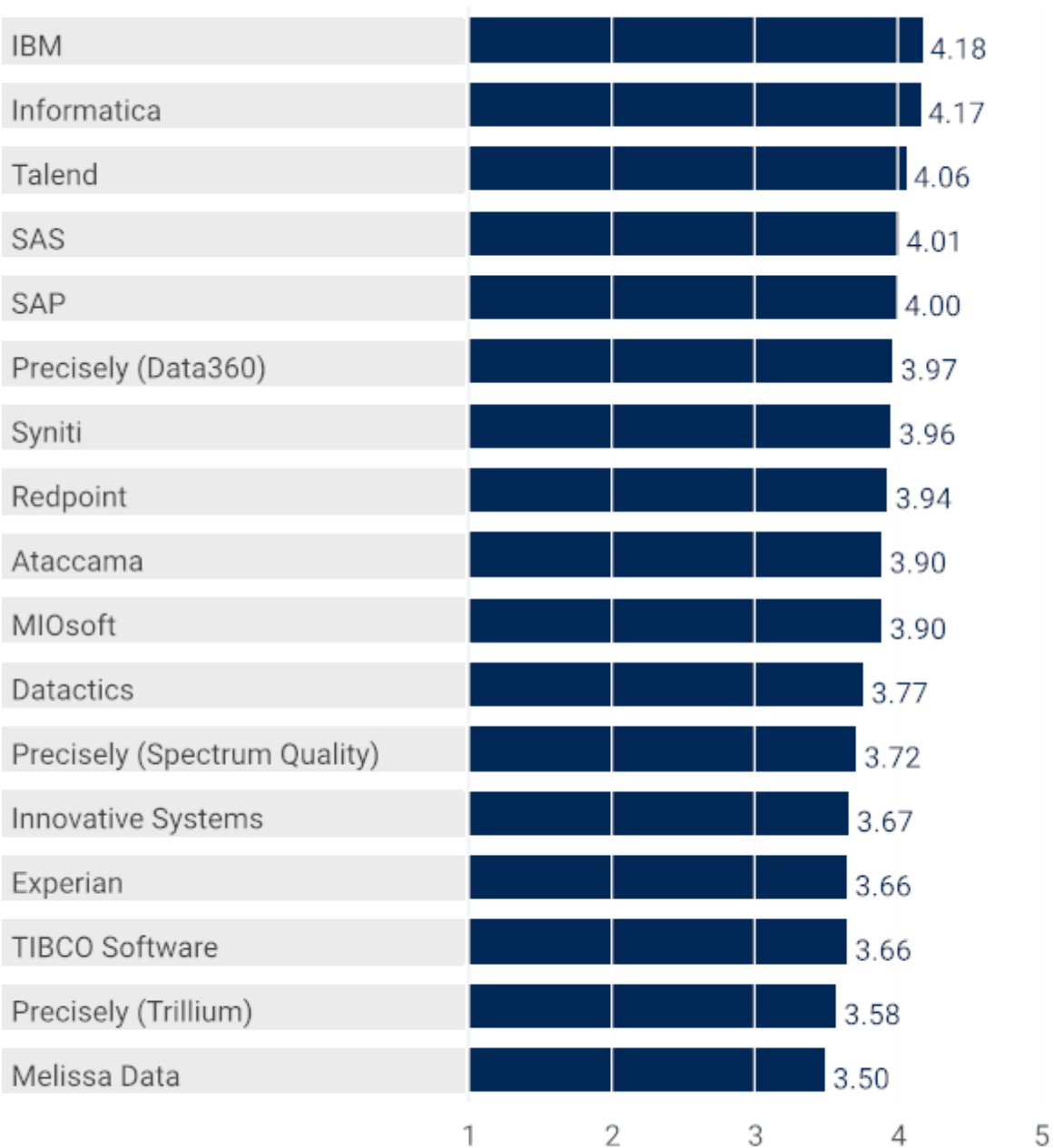
Analysis

Critical Capabilities Use-Case Graphics

Vendors' Product Scores for Analytics and Data Science Use Case



Product or Service Scores for Analytics and Data Science



As of 20 September 2021

© Gartner, Inc

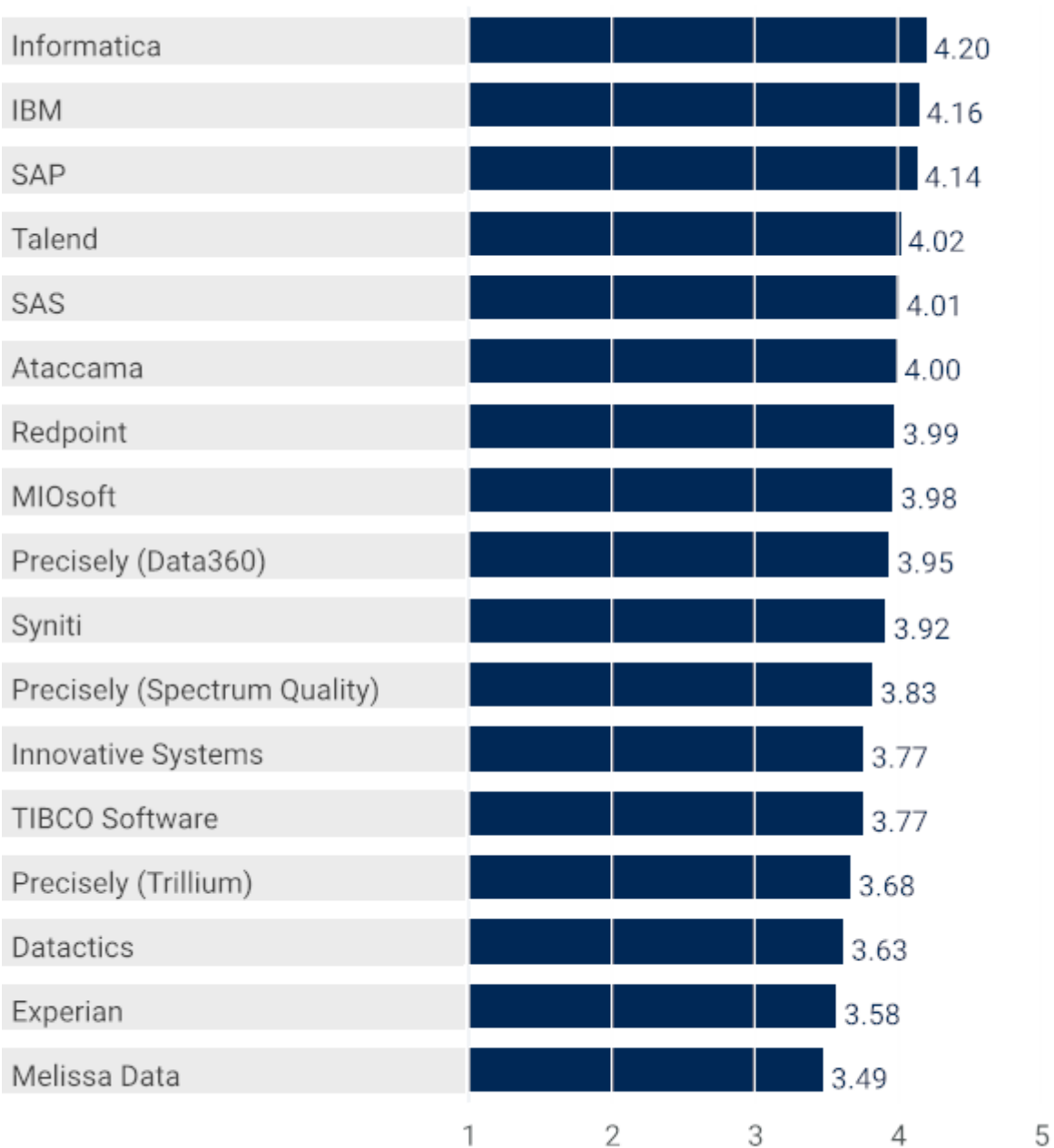
Gartner

Source: Gartner (October 2021)

Vendors' Product Scores for Data Integration Use Case



Product or Service Scores for Data Integration



As of 20 September 2021

© Gartner, Inc

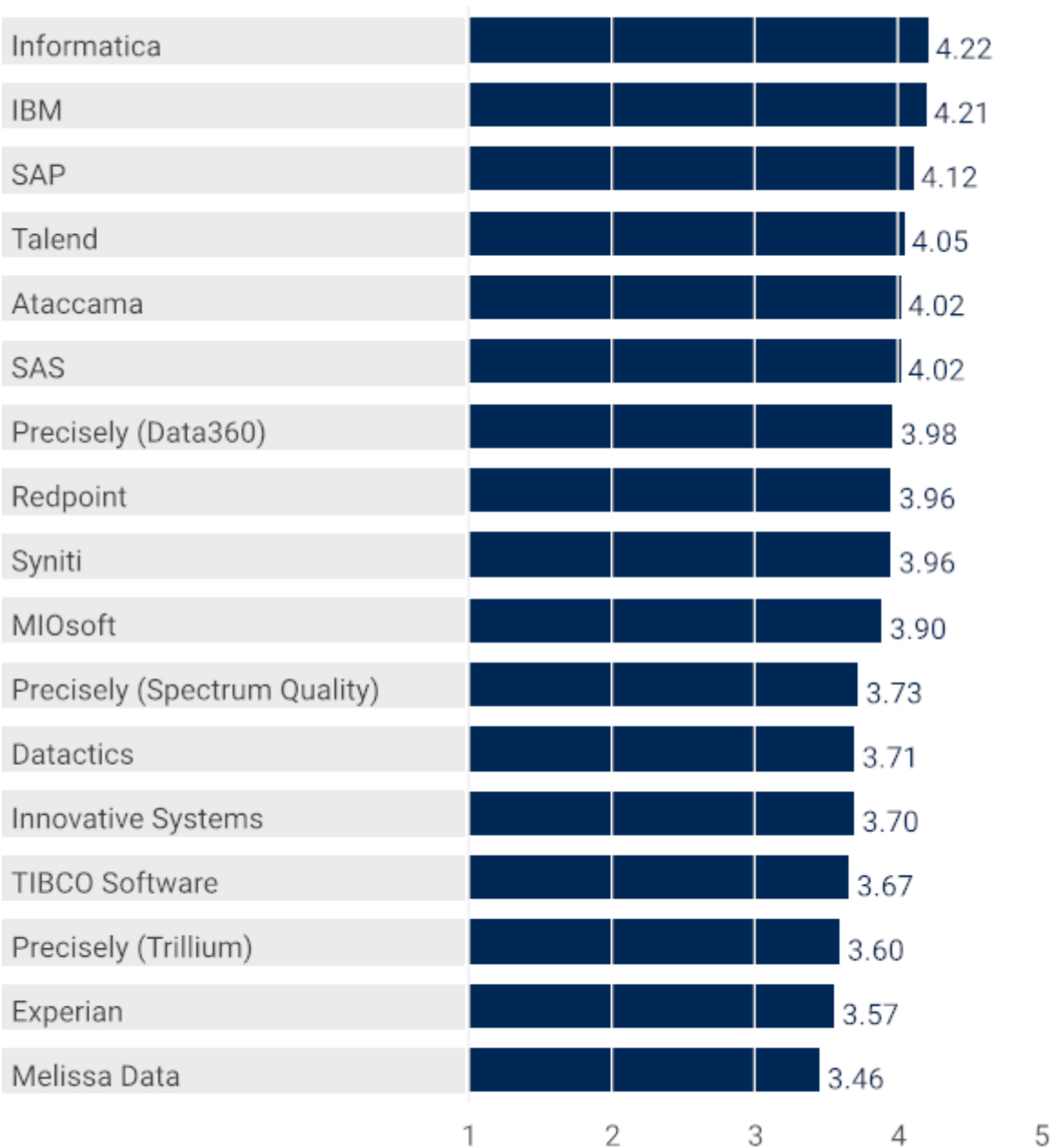
Gartner

Source: Gartner (October 2021)

Vendors' Product Scores for Data Migration Use Case



Product or Service Scores for Data Migration



As of 20 September 2021

© Gartner, Inc

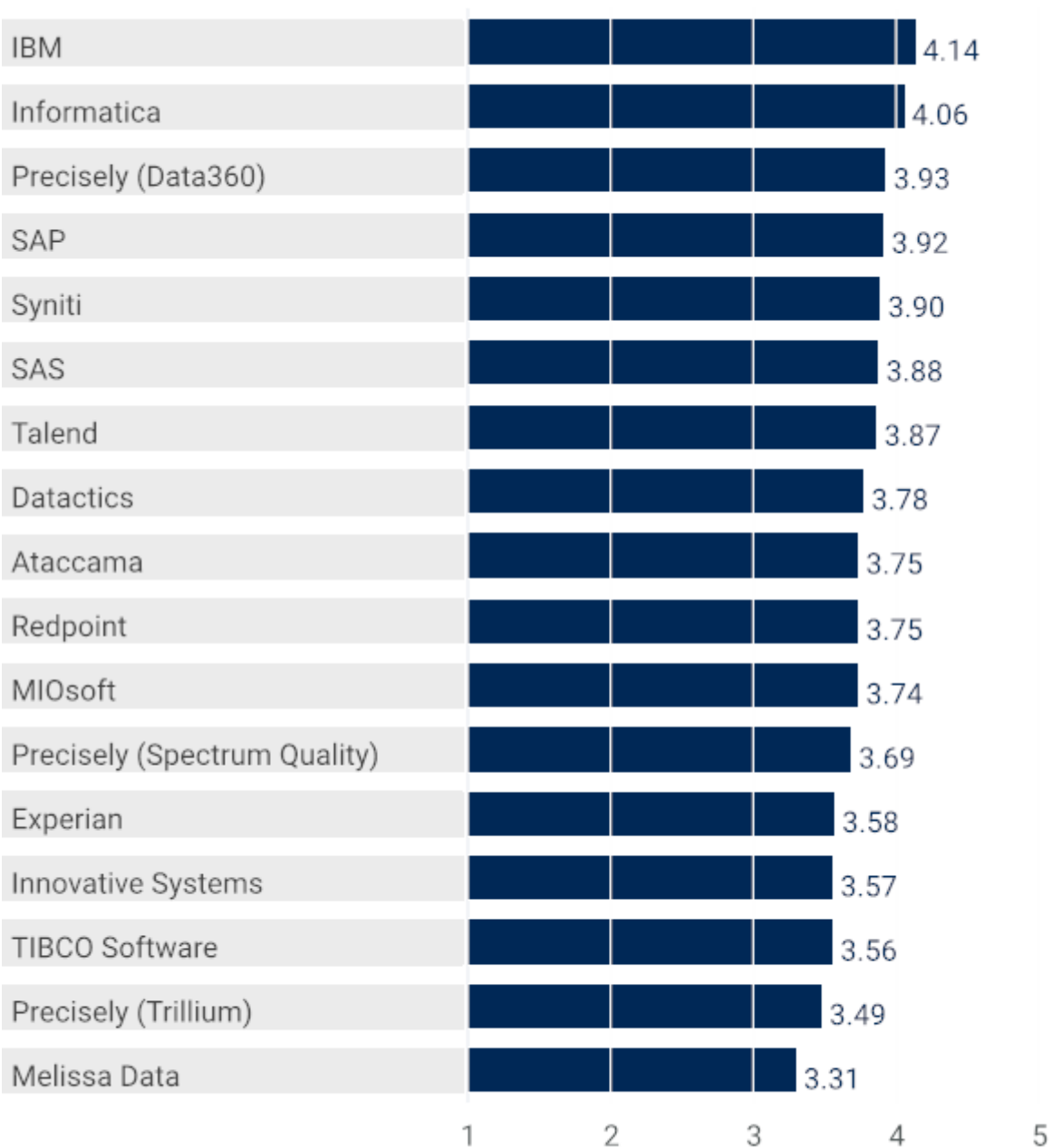
Gartner

Source: Gartner (October 2021)

Vendors' Product Scores for D&A Governance Initiatives Use Case



Product or Service Scores for D&A Governance Initiatives



As of 20 September 2021

© Gartner, Inc

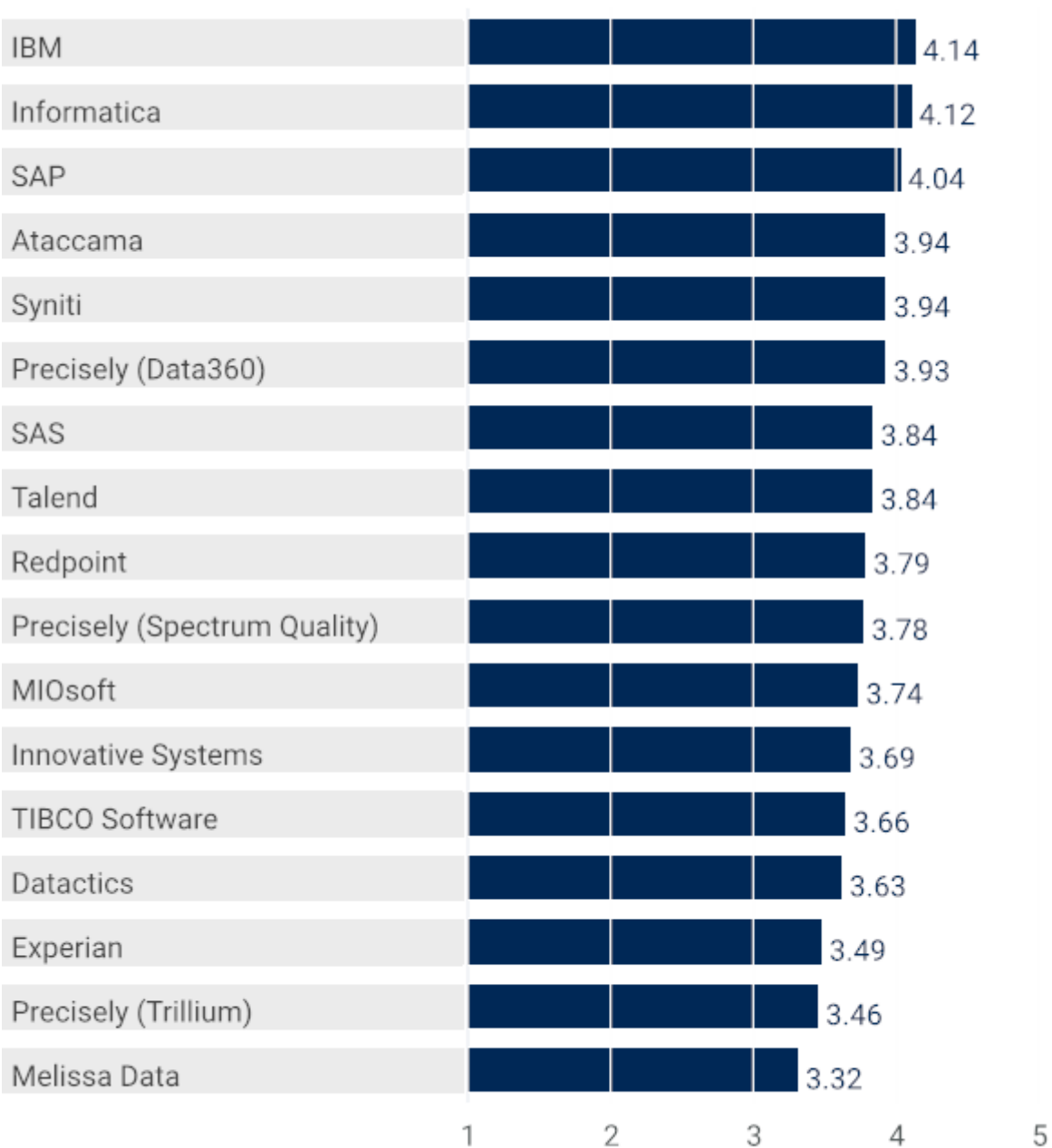
Gartner

Source: Gartner (October 2021)

Vendors' Product Scores for Master Data Management Use Case



Product or Service Scores for Master Data Management



As of 20 September 2021

© Gartner, Inc

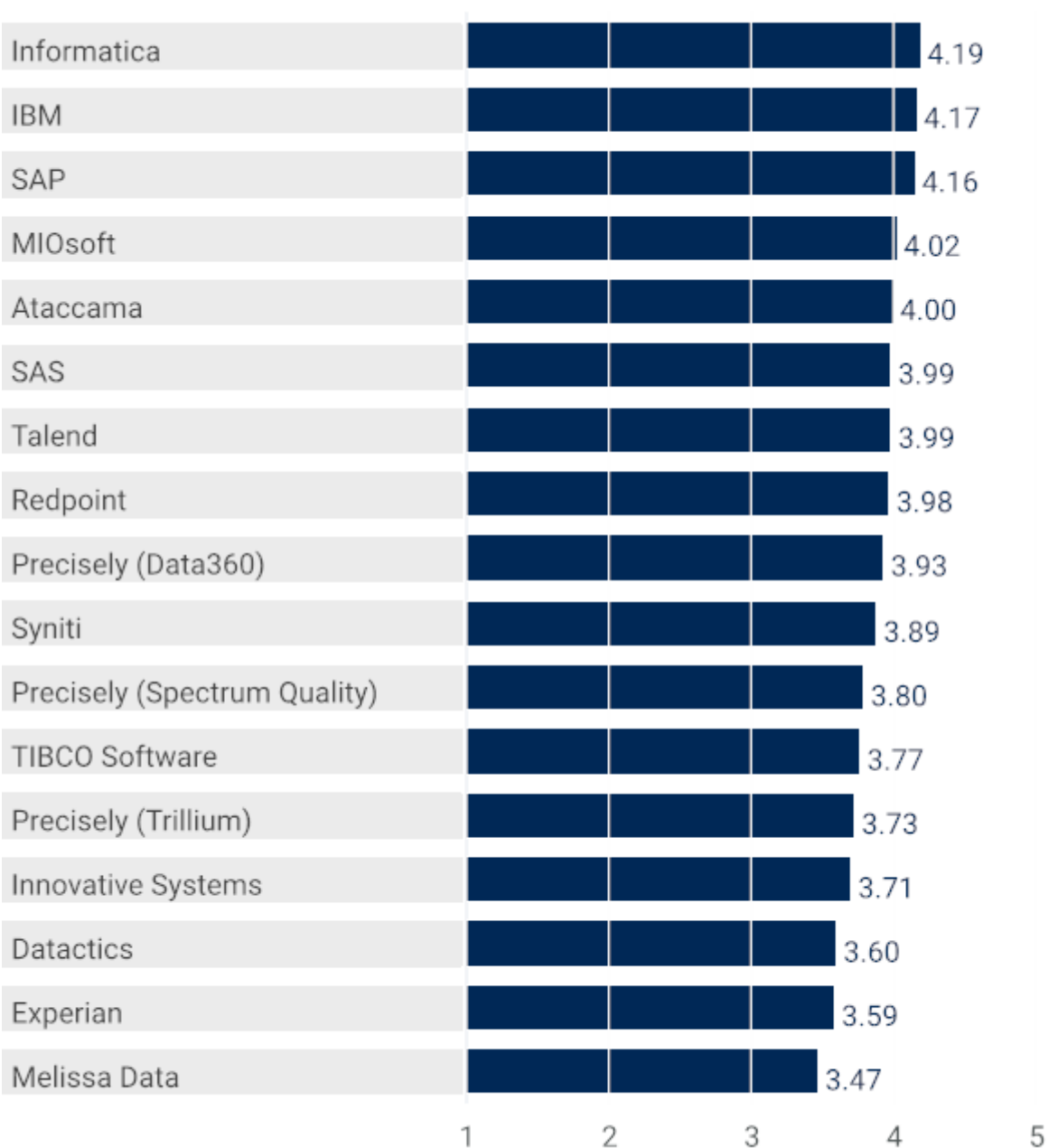
Gartner

Source: Gartner (October 2021)

Vendors' Product Scores for Operational/Transactional Data Quality Use Case



Product or Service Scores for Operational/Transactional Data Quality



As of 20 September 2021

© Gartner, Inc

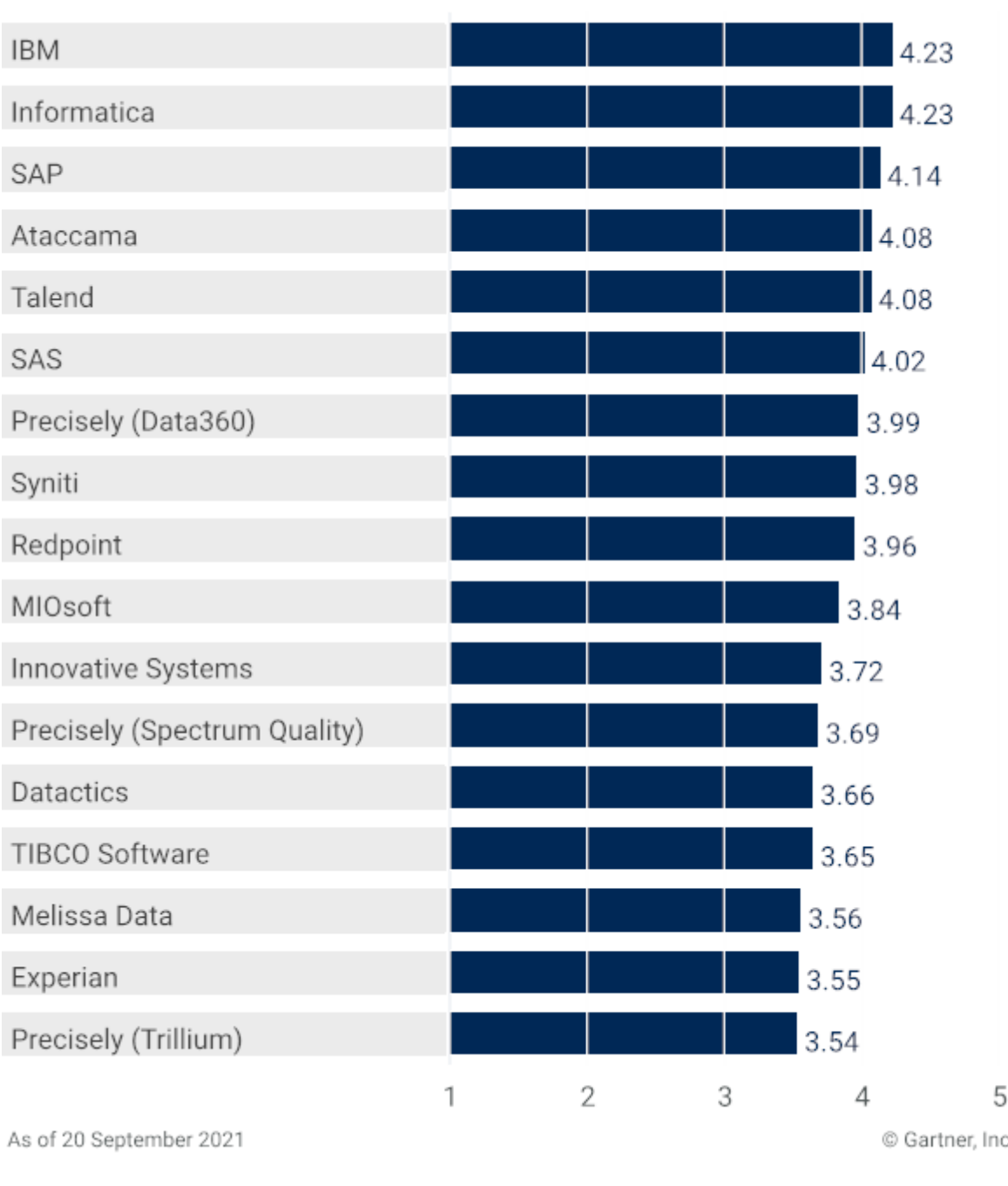
Gartner

Source: Gartner (October 2021)

Vendors' Product Scores for AI and Machine Learning Use Case



Product or Service Scores for AI and Machine Learning



Source: Gartner (October 2021)

Vendors

Ataccama

Ataccama has headquarters in Toronto, Canada. Its data quality product is Ataccama ONE Data Quality Suite. Ataccama has an estimated over 400 customers for this product line. It also has two freemium data quality products: Ataccama Data Quality Analyzer and Ataccama ONE Profiler. As reported by the vendor, its solution is primarily used in the banking and securities, healthcare, and public sectors.

Ataccama's profiling and scalability/performance are often rated as the key strengths of the solution. Customers also praise its core data quality functionality like matching/linking/merging. These critical capabilities contributed to Ataccama top-five score in the AI/machine learning, data migration and data integration data quality use cases. Ataccama's weaknesses relate specifically to interactive visualization and business-driven workflow, which contribute to its comparatively lower score for the data and analytics

governance, as well as analytics and data science use case. Ataccama's acquisition of Tellstory should bring further enhancements to analytics and visualization capabilities to strengthen this area.

Also note that Ataccama ONE platform extends support for other data management tasks around metadata and master data management. The solution is metadata-driven and augmented by artificial AI/ML capabilities to provide holistic data quality capabilities.

Datactics

Datactics has headquarters in Belfast, U.K. Its data quality product is Self-Service Data Quality platform, which contains four components: DQ Manager, DQ Clinic, FlowDesigner and AI Server. Datactics has an estimated 110 customers for these product lines. As reported by the vendor, its solution is primarily used in the financial services, banking and government sectors.

Datactics achieves favorable ratings for its profiling, interactive visualization/analytics and role-based usability, which makes it a suitable candidate for analytics and data science as well as data governance use cases. However, with lower-than-average ratings for multidomain support as well as parsing, standardization and cleansing for verticals outside of its strength, it records lower scores for master data management and operational/transactional type data quality use cases.

Also note that Datactics provides a range of AI functionalities, including prebuilt models for entity resolution, ML-based matching, outlier detection, and knowledge graph-based modeling and analysis. The focus is to offer a low-code platform to nontechnical users for rapid implementation across data quality use cases.

Experian

Experian has its corporate headquarters in Dublin, Ireland. Its data quality products include Experian Aperture Data Studio, Quick Address Systems (QAS) Pro (a web-based address validation product) and other data validation services/products. Experian has an estimated 6,400 customers for these product lines, with the majority of them using data validation services. As reported by the vendor, its solution is primarily used in the financial services, health and retail sectors.

Customers highlight Experian's strength in profiling and role-based usability, making it a favorable choice for business-user-heavy use cases like D&A governance and analytics. Experian derives this strength from years of experience working in the customer data domain. However, due to a lack of more innovation in metadata or AI across a few capabilities — such as cleansing, visualization/analytics and business-driven workflow — it scores lower in other use cases.

Also note that Experian differentiates with its ample data enrichment options, which augment its data quality capabilities by applying the experience in the form of supervised learning. The vendor provides a rich reference dataset to enrich customer data across multiple verticals.

IBM

IBM has headquarters in Armonk, New York, U.S. Its data quality products are IBM InfoSphere Information Server for Data Quality, InfoSphere QualityStage and Watson Knowledge Catalog (WKC), powered by IBM Cloud Pak for Data. IBM has an estimated 2,500 customers for these product lines with substantial usage across most sectors.

IBM offers metadata-driven and machine-learning-based automation as well as comprehensive functionality with both depth and breadth across all data quality use cases. IBM's functionality for profiling, multidomain support, business-driven workflow, rule management, data validations and role-based usability is its key

strength. As a result, IBM scores well in most use cases, and tops the chart for the data and analytics governance initiatives. However, it could benefit by further improvements to interactive visualization and matching/linking/merging capabilities, especially for the on-premises versions of its solution.

Note here that Watson Knowledge Catalog aims to provide most data management functionalities from the same platform as a single experience and to support end-to-end DataOps and AI life cycle scenarios. IBM also offers MDM, which is integrated into its Watson Knowledge Catalog.

Informatica

Informatica has headquarters in Redwood City, California. Its data quality products are Informatica Data Quality (IDQ), Informatica Cloud Data Quality (CDQ), Informatica Axon Data Governance and Enterprise Data Catalog, Informatica Data Engineering Quality and Informatica Data as a Service. Informatica has an estimated over 6,000-plus customers for these product lines and, as reported by the vendor, its solutions are used across most sectors.

Informatica continues to deliver comprehensively across all data quality use cases with one of the most extensive sets of data quality capabilities. Customers continue to acknowledge its capabilities as one of the highest-rated, especially for profiling, core data quality functionality (parsing, standardizing and cleansing), matching/linking/merging, business-driven workflow and rule engine/data validation. Informatica makes extensive use of a metadata-based AI/ML engine called CLAIRE to augment the understanding of nontechnical and technical users for issue identification and remediation. Informatica can further improve its utilization of more AI techniques to assist with other data quality capabilities such as automated issue resolution in future releases. Customers also expect Informatica to offer the same level of advanced capabilities for its Cloud Data Quality offering that is currently part of the on-premises versions. Customers also comment that the solution's usability at times is impacted because customers have to work across different components to meet all the requirements.

Note here that Informatica offers depth and expertise across solutions for all data management needs, such as data integration, metadata and master data management, which are all augmented by efficient utilization of CLAIRE-driven insights.

Innovative Systems

Innovative Systems has headquarters in Pittsburgh, Pennsylvania, U.S. Its data quality products include the Enlighten Data Quality Suite and FinScan, both of which reside on the Synchronos Enterprise Customer Platform. Innovative Systems has an estimated 1,046 customers for these products. As reported by the vendor, its solution is primarily used in the banking and securities, insurance, and media sectors.

Innovative Systems offers stable core data quality functionality coupled with rich knowledge bases augmenting its ratings across matching/linking/merging, core functionality (parsing, standardizing and cleansing) and role-based usability. As a result, Innovative Systems' score is higher for the data integration, data migration and master data management use cases than its other use cases. However, the solution can improve its profiling, visualization/analytics and rule recommendations by incorporating usage of metadata and AI. Innovative Systems received lower scores for other use cases.

Also worth noting here is that the same platform extends offerings in data integration and for other data management needs. This allows Innovative Systems to target most of the data management functions and address data quality needs across these use cases.

Melissa Data

Melissa Data has headquarters in Rancho Santa Margarita, California. Its data quality products include Contact Zone, Data Quality Suite, Data Quality Components for SQL Server, Sentient Suite and Unison. Melissa Data has an estimated 1,000 customers for these products. As reported by the vendor, its solution is primarily used in the communications, financial service and healthcare industries.

Melissa Data's data validation and role-based usability capabilities are often rated as key strengths of the solution, owing to very detailed and specialized knowledge bases. This contributes to higher scores for AI/machine learning data quality use cases than its other use cases. However, the solution offers other basic data quality capabilities for profiling, interactive visualization/analytics, matching/linking/merging and multidomain support and not enough support for business-driven workflow, resulting in lower scores for other use cases in this survey. The solution has focused on contact data validation and verification with some usage outside of that domain — for example, in the life sciences and healthcare domains.

Note here that Melissa Data offers robust verification and validation data quality solutions for various data objects such as business entities, names, addresses, geocodes, emails and phone numbers using SaaS or API calls.

MIOsoft

MIOsoft has headquarters in Madison, Wisconsin. Its data quality products are MIOvantage Platform and MIOvantage DQ Explorer. MIOsoft has an estimated 465 customers for these product lines. As reported by the vendor, its solution is primarily used in the telecommunications, insurance and public sectors.

MIOsoft has above average ratings especially for scalability/performance, multidomain support, matching/linking/merging and role-based usability, and it is reflected as a suitable solution for AI/machine learning, operational/transactional, data integration and data migration use cases. The solution is also praised by end users for its powerful entity resolution engine for larger volumes of data. However, MIOsoft's business-driven workflow, rule engine/data validation and visualization/analytics capabilities require more work, reflecting possible scope of improvement with more augmented capabilities driven by metadata and AI in the next releases. With that, MIOsoft will be able to better cater growing needs of data and analytics governance-like use cases in the future.

Note that MIOsoft continues to receive consistent positive feedback for its scalable data platform for the past several years. To complement the positive feedback, the vendor continues to invest and improve scalability/performance as a part of its roadmap.

Precisely (Data360)

Precisely has headquarters in Burlington, Massachusetts. Precisely offers Data360, which comes from its recent acquisition of Infogix and was an existing participant in the last version of the Critical Capabilities for Data Quality Solutions (published in 2020). As reported by the vendor, Data360 has close to 200 customers and is primarily used in the banking and securities, insurance, and healthcare sectors.

Data360 is a metadata-driven solution at the core and offers above average capabilities for profiling, interactive visualization/analytics, business-driven workflow/usability and rule engine/data validation. This makes this solution a favorable option for business-user-dependent use cases like data/analytics governance and analytics/data science. The solution continues to augment capabilities by focusing on metadata-driven insights and simultaneously using supervised learning to enhance its capabilities for issue identification and remediation. However, customers expect to have better capabilities for content across

multiple data domains, improvement in core data quality functions (parsing/standardization/cleansing) and scalability/performance. This will allow Precisely to cater to other data quality use cases more efficiently.

Note that the same Data360 platform caters to extensive metadata management, including a data catalog, and it caters to analytics needs using Data360 Analyze.

Precisely (Spectrum Quality)

Precisely also offers Spectrum Quality, which comes from its earlier acquisition of Pitney Bowes Software. As reported by the vendor, Spectrum Quality has more than 3,000 customers and is primarily used in the banking and securities, insurance and government sectors.

Spectrum Quality offers a rich set of data quality capabilities that are especially suitable for contact data. It has specialized its functionalities based on years of experience for the customer data domain. Spectrum Quality offers above average capabilities for matching/linking/merging, business-driven workflow and role-based usability of the solution. The solution has focused on improving its matching capabilities extensively in past years. This is the stronghold of Spectrum Quality and where most of the innovation has happened. This allows the solution to be a favorable option for master-data-management-like use cases where there is an extensive need for entity resolution. However, the solution lags others in bringing innovation driven by integrated use of metadata and AI assistance across several other capabilities such as profiling, analytics, cleansing, rule inference/recommendation and automated data validation. Augmentation across these capabilities in next releases will allow Spectrum to achieve better scores for other use cases.

Note that Spectrum offers rich functionalities to support global addressing requirements, ranging from medium to large volumes of address data, that derives its strength from years of experience of working in the mailing solutions business.

Precisely (Trillium)

Precisely also offers Trillium Quality, which comes from Syncsort product line (Syncsort acquired Trillium in December 2016). As reported by the vendor, Trillium Quality has more than 1,000 customers and is primarily used in the banking and securities, insurance, and communications sectors.

Trillium Quality has been part of the data quality market for a long time and continues to derive strength from its experience across key capabilities of profiling and core data quality functions (parsing/standardization/cleansing). The solution also performs well in data quality scenarios testing for better scalability and performance. This allows Trillium Quality to achieve better scores across data integration and migration use cases compared with others. However, the solution exhibits a clear lack of innovation to make integrated use of metadata- and AI-driven augmentation across most data quality capabilities. Customers expect the solution to offer better capabilities for visualization/analytics over profiling and monitoring results, metadata-driven rule recommendations, automated data validation, and business-driven workflow/usability. These kinds of improvements will allow Trillium Quality to match up to the demand side and perform better across several other data quality use cases.

Note that Precisely also offers robust data integration capabilities from its original Syncsort product line that is integrated with Trillium solutions.

Redpoint

Redpoint has headquarters in Wellesley Hills, Massachusetts. Its data quality product is Redpoint Data Management. The vendor has an estimated 300 customers for this product. As reported by the vendor, its

solution is primarily used in the retail, financial services and healthcare sectors.

Redpoint Data Management provides a broad set of functionalities to support all data quality use cases and continues to have a strong focus on customer data domain. Its perceived strengths continue to be around core data quality functions (parsing, standardizing and cleansing), matching/linking/merging, scalability/performance and role-based usability. Consequently, Redpoint scores well in data integration, data migration, and AI and machine learning use cases in comparison with other use cases. However, the solution can improve its capabilities across profiling, interactive visualization/analytics and business-driven workflow, as well as by adding more augmented capabilities driven by better use of its Automated Machine Learning in next releases, especially outside the customer domain.

Note that Redpoint Global extends the same platform beyond data quality; it also provides built-in capabilities for larger data integration and application data management needs of enterprises and is mostly centered around customer domain and customer journey support.

SAP

SAP has headquarters in Walldorf, Germany. Its data quality products include SAP Information Steward, SAP Data Services, and SAP Data Intelligence (formerly Data Hub). SAP has an estimated 24,700 customers for these product lines, with substantial usage across most sectors.

SAP's strength in data quality solutions is spread across core data quality capabilities of parsing, standardizing and cleansing, as well as a bunch of other critical capabilities of profiling, visualization/analytics, robust multi-data domain support and scalability of the platform. SAP derives its strength from years of experience, market share in the data management space and working with a large variety of clients from different sectors, thus contributing to enriched prepackaged functionalities and supervised learning. However, there is a need to improve some capabilities such as matching/linking/merging and role-based usability making use of metadata- and AI-driven augmentation, especially for business users in a more native format. SAP scores well across all use cases and is top three in multiple use cases, but it could improve in the business-user-heavy data governance use case.

It is worth mentioning here that the SAP combined portfolio offers broad and deep data management functionality in both SAP and non-SAP environments, and it is positioned well to make better use of metadata- and AI-driven insights.

SAS

SAS has headquarters in Cary, North Carolina. Its data quality products are SAS Data Management, SAS Data Quality, SAS Data Loader for Hadoop, SAS Data Governance, SAS Data Preparation on SAS Viya and SAS Event Stream Processing on SAS Viya. SAS has an estimated 2,700 customers for these product lines. As reported by the vendor, its solution is primarily used in the banking, government and services sectors.

SAS's combined data quality product portfolio can support most traditional and emerging data quality needs across a variety of business scenarios. SAS offers robust functionality through its highly scalable stand-alone data quality and embedded data quality offerings for profiling and visualization/analytics. SAS also derives strength across core data quality functions (parsing, standardizing and cleansing) from specialized focus on its knowledge bases, which are prepackaged and can be customized as well. This allows SAS to score above average for most use cases, and it has one of the leading scores for the analytics and data science use case. However, customers tend to expect more from business-driven workflow, rule and data validation, and business-driven usability of the different offerings. SAS could make better use of metadata-

and AI-driven augmentation, especially by using the depth of its platform (it requires an augmented data catalog and native capabilities for AI-assisted data quality).

Note that SAS Data Quality and Data Governance components are delivered by default for all SAS Viya customers without additional cost. And with the SAS event stream engine, data quality functions can also be pushed out to the edge.

Syniti

Syniti has headquarters in Needham, Massachusetts. Its data quality product is Syniti Knowledge Platform. Syniti has an estimated 450 customers for this product. As reported by the vendor, its solution is primarily used in the pharmaceuticals and life sciences, manufacturing, and food sectors.

Syniti has worked extensively to consolidate its portfolio of data quality capabilities as part of a single data management platform called Syniti Knowledge Platform. This has allowed Syniti to significantly augment its capabilities of core functionality (parsing, standardizing and cleansing), profiling, analytics, multidomain support, business-driven workflow, usability and rule recommendation. Syniti derives its strength from better utilization of metadata and knowledge graphs at its core to build the knowledge tier, which it calls Deep Guidance. However, customers expect further improvements from the platform for profiling, matching/linking/merging and scalability/performance by making use of AI/ML techniques. This will allow the platform to have better acceptance for data science and AI-/machine-learning-like data quality use cases.

Also, note that Syniti Knowledge Platform is an integrated suite of data management solutions. It supports primary data management functions of data transformation, data migration, data governance and master data management where data quality capabilities are applied to improve the outcomes.

Talend

Talend has headquarters in Redwood City, California. Its data quality products include Talend Data Fabric and Talend Data Catalog. Talend has an estimated over 1,854 licensed customers for this product line. It also has two freemium data quality products: Talend Open Studio for Data Quality and Talend Data Preparation Free Desktop. As reported by the vendor, its solution is primarily used in the media services, financial services and manufacturing sectors.

Talend is often rated highly for its solutions capability to serve both nontechnical and technical users across multidomains. The solution also packs above average capabilities for profiling, analytics, core data quality functions (parsing, standardizing and cleansing) and rule engine/data validation. This makes Talend's offerings suitable for the majority of the data quality use cases, scoring close to the top in analytics/data science data quality use cases. However, as expected by customers, Talend is positioned well to make broader use of metadata and AI-driven insights to augment these capabilities further for general use beyond just data preparation.

Note that Talend, along with its extensive focus on a wide range of data management needs, offers robust data quality, data integration and metadata management functionalities from the same platform. This places Talend in a favorable position to further its augmented data management capabilities.

TIBCO Software

TIBCO Software is based in Palo Alto, California. It offers TIBCO Omni-Gen Data Quality (acquired from Information Builders). TIBCO has an estimated 320-plus customers for this product. As reported by the

vendor, its solution is primarily used in the financial services, healthcare and public sectors.

TIBCO offers a complete data quality suite that has its perceived strength across interactive visualization and analytics, as well as good multidomain support, which allows the solution to achieve comparatively better scores for MDM use cases. However, the solution lacks innovation with little AI-/ML-driven automation natively available as part of the platform and not much focus on making use of metadata-driven insights. The solution can potentially make gains by possible integration to TIBCO's other product portfolio – such as Spotfire's data preparation, which offers better integration to AI/ML capabilities – or from TIBCO's investment in its metadata product. TIBCO commented that it is building an integrated TIBCO DQ platform, and will soon make an announcement of it.

Also, note that TIBCO's Omni-Gen Data Quality is part of the broader data management platform that provides robust functionality and also caters to data integration and master data management needs. TIBCO also has additional existing market-leading offerings across these data management requirements.

Context

This Critical Capabilities research note is designed to help data and analytics leaders select a suitable data quality vendor solution. However, product capabilities alone should not form the basis of a purchasing decision. This Critical Capabilities report must be used in conjunction with its corresponding [Magic Quadrant for Data Quality Solutions](#) report for guidance on the market positions of data quality solution vendors.

The Magic Quadrant and Critical Capabilities research complement each other in terms of focus. The Magic Quadrant's analysis covers several dimensions across two axes: Ability to Execute and Completeness of Vision. This Critical Capabilities report focuses on a vendor's product or service – on the Ability to Execute axis. It equates to a “double-click” drill down, with a focus (in this instance) on the nine functional capabilities that, in Gartner's view, are critical for supporting the major data quality use cases.

Complete the selection process by performing a thorough RFP ([Toolkit: RFP Template for Data Quality Solutions](#)) and proof of concept to ensure that any prospective solution fits the requirements, the practice and the variety of roles and skills involved across your organization. Also, it is equally important to analyze cost-saving opportunities by looking at nontechnological characteristics of vendors, such as acquisition processes, pricing models, speed of deployment, total cost of ownership, availability of skills, and support and service capabilities.

Product/Service Class Definition

The data quality solutions market consists of vendors offering software products that provide the following data quality functionality:

- **Connectivity:** The ability to access and apply data quality rules to a wide range of data sources, including internal and external, on-premises and cloud, and relational and nonrelational data sources.
- **Data profiling, measurement and visualization:** Data analysis capabilities that give business and IT functions (especially those supporting business users) insight into the quality of data and that help them identify and understand data quality issues.
- **Monitoring:** Capabilities that assist with the ongoing understanding and assurance of data quality through the monitoring of, and alerting to, possible data quality issues or events.

- **Parsing:** Built-in capabilities that decompose data into its component parts.
- **Standardization and cleaning:** Built-in capabilities that apply government, industry or local standards, business rules or knowledge bases to modify data for specific formats, values and layouts.
- **Matching, linking and merging:** Built-in capabilities that match, link and merge related data entries within or across datasets, using a variety of techniques such as rules, algorithms, metadata and machine learning.
- **Multidomain support:** Packaged capabilities aimed at specific data subject areas, such as customer, product, asset or location.
- **Address validation/geocoding:** Capabilities that support location-related data standardization and cleansing, and completion for partial data in real-time or batch processes.
- **Data curation and enrichment:** Capabilities that integrate externally sourced data to improve completeness and add value.
- **Business rule development and implementation:** Capabilities that create, deploy and manage business rules that can then be called within the solution or by third-party applications for data validation purposes.
- **Issue resolution and workflow:** Process workflows and user interfaces that enable nontechnical business users to identify, quarantine, assign, escalate, resolve and monitor data quality issues.
- **Metadata management:** Capabilities that capture, reconcile and interoperate metadata relating to the data quality process.
- **DataOps environment:** Collaboration of the data management practice focused on improving the communication, integration and automation of data flows between data managers and data consumers across an organization.
- **Deployment environment:** Styles of deployment, hardware, operating system and maintenance options for deploying data quality operations.
- **Architecture and integration:** Commonality, consistency and interoperability among various components of the data quality toolsets (including third-party tools).
- **Usability:** Suitability of the solution to engage and support the various roles (especially nontechnical business roles) required in a data quality initiative.

Critical Capabilities Definition

The functional capabilities that Gartner assesses to be critical in support of various data quality use-case requirements are listed in this section.

Profiling

The statistical analysis of diverse datasets (ranging from structured to unstructured data and from on-premises to cloud) to give business users insight into the quality of data and enable them to identify data quality issues.

Parsing, Standardizing, Cleansing

The decomposition and formatting of diverse datasets based on government, industry or local standards, business rules, knowledge bases, metadata and machine learning; modification of data values to comply with domain+ restrictions, integrity constraints or other business rules.

Interactive Visualization/Analytics

The interactive analytical workflow and visual output of statistical analysis to help business and IT users identify, understand and monitor data quality issues and discover patterns and trends over time through, for example, reports, scorecards, dashboards and mobile devices.

Matching, Linking and Merging

Matching, linking and merging of related data entries within or across diverse datasets using a variety of traditional and new approaches, such as rules, algorithms, metadata, AI and machine learning.

Multidomain Support

The ability to address multiple data subject areas (such as various master data domains and vertical industry domains) and depth of packaged support for these subject areas.

Business-Driven Workflow

The processes and user interface to manage the data quality issue resolution through the stewardship workflow, and to enable business users to easily identify, quarantine, assign, escalate and resolve data quality issues facilitated by collaboration, pervasive monitoring and case management.

Scalability and Performance

The ability to deliver suitable throughput and response times to satisfy performance SLAs in both batch and real-time modes, given increasing data volumes for diverse datasets (ranging from structured to unstructured data, on-premises to cloud, and traditional to open-source databases).

Role-Based Usability

The suitability of the solutions to support and empower various roles — such as data steward, data engineer or data scientist — using self-service techniques such as collaboration, business-friendly interfaces and data preparation.

Rule Management and Data Validation

The ability to design, create and deploy business rules for specific data values. The rules can be called within the solution or by third-party applications for data validation purposes, which can be done in batch or real-time mode.

Use Cases

The common use cases that Gartner identifies to be supported by various enterprise-level data quality solutions are listed in this section.

Analytics and Data Science

DQ capabilities supporting interpretation of structured/unstructured sources for operational analytics, performance management, sentiment analysis and other analytic scenarios.

This use case increasingly involves combining diverse datasets — including unstructured data, Internet of Things (IoT) data and streaming data — of unknown quality in data warehouse or data lake environments. Therefore, this use case has a heavier emphasis on data profiling, matching and interactive visualization and

analytics capabilities. In addition, usability, performance and scalability are key. Data quality capabilities are useful for supporting components of analytics and data science solutions.

Note: This use case focuses on the degree to which data quality solutions support analytics in general. Vendors appearing here may or may not sell analytics solutions.

Data Integration

Data quality capabilities applied within data integration processes in support of both data consolidation for analytics and operational data/application integration.

Data integration initiatives cannot be successful without mechanisms to assure the quality of the data being integrated and delivered. With increasing complexity and diversity of data sources, the key critical capabilities include matching, core data quality (parsing, standardizing and cleansing) function, and performance/scalability. Data quality capabilities are one component of a comprehensive data integration solution.

Note: This use case focuses on the degree to which data quality solutions support data integration. Vendors appearing here may or may not sell data integration solutions, and their assessment is independent of such offerings.

Data Migration

Data quality capabilities used in the context of a data conversion, migration or modernization initiative (such as conversion from legacy to modern applications).

These initiatives require a strong focus on identifying data quality issues upfront. Therefore, this use case emphasizes data profiling and core data quality function while also demanding strong scalability and performance to support large-scale migration efforts. Data quality capabilities are one component of a comprehensive data migration solution.

Note: This use case is focused on the degree to which data quality solutions support data migration. Vendors appearing here may or may not sell data migration solutions or services, and their assessment is independent of such offerings.

D&A Governance Initiatives

Data quality capabilities supporting the data governance initiative and its associated key roles (data stewards and data owners, for example).

Information leadership roles (such as the chief data officer) and initiatives focus on increasing the value of data assets while managing risks and compliance. The data governance initiative requires superior capabilities for data profiling, visualization, workflow and usability to support data governance roles at all levels. This includes data stewards, members of data governance boards/councils, and other business-side stakeholders. These roles are increasingly supported by nontechnical individuals.

Note: This use case is focused on the degree to which data quality solutions support data governance. Vendors appearing here may or may not sell data governance solutions or services, and their assessment is independent of such offerings.

Master Data Management

Data quality capabilities applied to various key master data domains in the context of MDM initiatives and the deployment of custom or packaged MDM solutions.

This use case emphasizes the matching, workflow and multidomain capabilities of the solutions due to the common requirements to resolve master data that is authored in disparate sources. Data quality capabilities are one component, among many, that make up a comprehensive master data management solution.

Note: This use case focuses on the degree to which data quality solutions support master data management. Vendors appearing here may or may not sell master data management solutions, and their assessment is independent of such offerings.

Operational/Transactional Data Quality

Data quality capabilities applied to controlling the quality of data created by, maintained by and housed in operational/transactional applications, including IoT systems.

As data quality controls are increasingly applied upstream, closer to the source of data, the ability to embed data quality capabilities in operational applications is key. This use case emphasizes the core data quality operations (including parsing, standardizing and cleansing), workflow and multidomain, as well as strong scalability and performance in the face of ever-increasing transaction volumes and velocity.

AI and Machine Learning

Data quality capabilities applied to controlling the quality of data used for training models or algorithms and actual data feed.

The data quality solutions can help data scientists prepare their datasets by fixing data problems, transforming data value and detecting any data bias. With increasing complexity and diversity of data sources and AI use cases, the key critical capabilities include profiling, standardizing, cleansing, matching, scalability and performance.

Note: This use case focuses on the degree to which data quality solutions support AI/ML algorithm development. Vendors appearing here may or may not sell AI/ML or data science platforms, and their assessment is independent of such offerings.

Vendors Added and Dropped

Added

TIBCO Software: TIBCO has been included as a first-time entrant because it meets the inclusion criteria after its acquisition of Information Builders.

Datactics: Datactics has been included as a first-time entrant because it meets the inclusion criteria.

Dropped

Oracle: Due to lack of vendor’s responses, Gartner could not confirm that Oracle meets this year’s inclusion criteria.

Infogix: Infogix was acquired by Precisely in June 2021. However, the solution is represented in the assessment as Precisely Data360.

Information Builders: Information Builders was acquired by TIBCO Software in January 2021.

Inclusion Criteria

In the context of this Critical Capabilities analysis, we are using the same inclusion criteria as our [Magic Quadrant for Data Quality Solutions](#).

The inclusion criteria represent the specific attributes that vendors must have in order to be included in this research.

To be *included*, vendors had to fulfill all the following criteria:

- Offer stand-alone software solutions that are positioned, marketed and sold specifically for general-purpose data quality applications. Vendors that provide several data quality product components must demonstrate that these are integrated and collectively meet the full inclusion criteria for this Critical Capabilities research.
- Deliver core data quality functions for, at minimum, profiling, interactive visualization, business rule creation, rule-based data validation, parsing, standardization, cleansing, matching, multidomain data support and business-driven workflow.
- Support the above functions in both scheduled (batch) and interactive (real-time) modes.
- Enable large-scale deployment via server-based and cloud-based runtime architectures that can support concurrent users and applications.
- Support multiple data domains and diverse use cases across different industries.
- Maintain an installed base of at least 100 production customers (different companies/organizational entities) for their flagship data quality products and not individual smaller modules or capabilities.
- Achieve at least \$5 million in total recognized revenue (per generally accepted accounting principles [GAAP] definition) for software (license, maintenance and subscription) relating to data quality solutions in the calendar year 2020.
- Have a customer base for production deployment that includes customers in multiple countries and in more than one region (i.e., North America, South America, EMEA and Asia/Pacific).
- Support data quality functionality with packaged capabilities to process data in at least two different languages: The user interface can support multiple languages as well.
- Include a complete solution addressing administration and management, as well as end-user-facing functionality, for four or more of the following types of users: data steward, data architect, data quality analyst, data engineer, data scientist and casual user.
- Provide out-of-box, and prebuilt data quality rules for the purpose of data cleansing, standardization and transformation, based on common industrial practices.
- Support integrability or interoperability with other data management solutions such as metadata management, master data management or data integration solutions from third-party tools.
- Provide direct sales and support operations, or have a partner providing sales and support operations, in at least two of the following regions: North America, South America, EMEA and Asia/Pacific.

The following types of vendor were excluded from this Magic Quadrant, even if their products met the above criteria:

- Vendors that meet the above criteria but are limited to deployments in a single, specific application environment, industry or data domain are excluded.
- Vendors that support limited data quality functionalities or that address specific data quality problems (for example, address cleansing and validation) are excluded because they do not provide the complete suite of functionality expected of today’s data quality solutions.
- Vendors that support only on-premises deployment and have no option for cloud-based deployment on any public cloud environment (for example, AWS, Azure or Google Cloud) are excluded.
- Vendors that operate in only a single country and support only one language are excluded.
- Vendors that lack the integrability or interoperability with other data management solutions such as metadata, MDM or data integration solutions are excluded.

Table 1: Weighting for Critical Capabilities in Use Cases

<i>Critical Capabilities</i> ↓	<i>Analytics and Data Science</i> ↓	<i>Data Integration</i> ↓	<i>Data Migration</i> ↓	<i>D&A Governance Initiatives</i> ↓
Profiling	25%	5%	20%	20%
Parsing, Standardizing, Cleansing	5%	20%	20%	5%
Interactive Visualization/Analytics	20%	5%	10%	30%
Matching, Linking and Merging	15%	20%	10%	0%
Multidomain Support	0%	15%	5%	0%
Business-Driven Workflow	0%	5%	5%	20%

<i>Critical Capabilities</i> ↓	<i>Analytics and Data Science</i> ↓	<i>Data Integration</i> ↓	<i>Data Migration</i> ↓	<i>D&A Governance Initiatives</i> ↓
Scalability and Performance	10%	20%	15%	0%
Role-Based Usability	20%	5%	5%	25%
Rule Management and Data Validation	5%	5%	10%	0%

Source: Gartner (October 2021)

This methodology requires analysts to identify the critical capabilities for a class of products/services. Each capability is then weighted in terms of its relative importance for specific product/service use cases.

Critical Capabilities Rating

Each of the products/services that meet our inclusion criteria has been evaluated on the critical capabilities on a scale from 1.0 to 5.0 (see Note 1).

Table 2: Product/Service Rating on Critical Capabilities

<i>Critical Capabilities</i> ↓	<i>Ataccama</i> ↓	<i>Datactics</i> ↓	<i>Experian</i> ↓	<i>IBM</i> ↓	<i>Informatica</i> ↓
Profiling	4.3	4.0	3.7	4.4	4.5
Parsing, Standardizing, Cleansing	4.1	3.8	3.5	4.2	4.3
Interactive Visualization/Analytics	3.5	3.5	3.0	4.0	3.8
Matching, Linking and Merging	4.0	3.7	3.5	4.1	4.2

<i>Critical Capabilities</i> ↓	<i>Ataccama</i> ↓	<i>Datactics</i> ↓	<i>Experian</i> ↓	<i>IBM</i> ↓	<i>Informatica</i> ↓
Multidomain Support	4.0	3.3	3.5	4.1	4.1
Business-Driven Workflow	3.8	3.7	3.3	4.1	4.0
Scalability and Performance	4.1	3.5	3.8	4.2	4.3
Role-Based Usability	3.5	4.0	4.4	4.1	4.0
Rule Management and Data Validation	4.1	3.5	3.5	4.3	4.1

Source: Gartner (October 2021)

Table 3 shows the product/service scores for each use case. The scores, which are generated by multiplying the use-case weightings by the product/service ratings, summarize how well the critical capabilities are m

Table 3: Product Score in Use Cases

<i>Use Cases</i> ↓	<i>Ataccama</i> ↓	<i>Datactics</i> ↓	<i>Experian</i> ↓	<i>IBM</i> ↓	<i>Informatica</i> ↓
Analytics and Data Science	3.90	3.77	3.66	4.18	4.17
Data Integration	4.00	3.63	3.58	4.16	4.20
Data Migration	4.02	3.71	3.57	4.21	4.22
D&A Governance Initiatives	3.75	3.78	3.58	4.14	4.06

<i>Use Cases</i> ↓	<i>Ataccama</i> ↓	<i>Datactics</i> ↓	<i>Experian</i> ↓	<i>IBM</i> ↓	<i>Informatica</i> ↓
Master Data Management	3.94	3.63	3.49	4.14	4.12
Operational/Transactional Data Quality	4.00	3.60	3.59	4.17	4.19
AI and Machine Learning	4.08	3.66	3.55	4.23	4.23

Source: Gartner (October 2021)

To determine an overall score for each product/service in the use cases, multiply the ratings in Table 2 by the weightings shown in Table 1.

Evidence

The analysis in this research is based on information from several sources, including:

- An RFI process that engaged vendors in this market. It elicited extensive data on functional capabilities, customer base demographics, financial status, pricing and other quantitative attributes.
- Interactive briefings and product demonstrations in which vendors provided Gartner with updates on their strategy, market positioning, recent key developments and product roadmap.
- Feedback about solutions and vendors captured during conversations with users of Gartner's client inquiry service.
- Market share and revenue growth estimates developed by Gartner's technology and service provider research unit.
- Peer feedback from Gartner Peer Insights, comprising peer-driven ratings and reviews for enterprise IT solutions and services covering over 300 technology markets and 3,000 vendors.

Critical Capabilities Methodology

This methodology requires analysts to identify the critical capabilities for a class of products or services. Each capability is then weighted in terms of its relative importance for specific product or service use cases. Next, products/services are rated in terms of how well they achieve each of the critical capabilities. A score that summarizes how well they meet the critical capabilities for each use case is then calculated for each product/service.

"Critical capabilities" are attributes that differentiate products/services in a class in terms of their quality and performance. Gartner recommends that users consider the set of critical capabilities as some of the most important criteria for acquisition decisions.

In defining the product/service category for evaluation, the analyst first identifies the leading uses for the products/services in this market. What needs are end-users looking to fulfill, when considering products/services in this market? Use cases should match common client deployment scenarios. These distinct client scenarios define the Use Cases.

The analyst then identifies the critical capabilities. These capabilities are generalized groups of features commonly required by this class of products/services. Each capability is assigned a level of importance in fulfilling that particular need; some sets of features are more important than others, depending on the use case being evaluated.

Each vendor's product or service is evaluated in terms of how well it delivers each capability, on a five-point scale. These ratings are displayed side-by-side for all vendors, allowing easy comparisons between the different sets of features.

Ratings and summary scores range from 1.0 to 5.0:

1 = Poor or Absent: most or all defined requirements for a capability are not achieved

2 = Fair: some requirements are not achieved

3 = Good: meets requirements

4 = Excellent: meets or exceeds some requirements

5 = Exceptionnel : dépasse largement les exigences

Pour déterminer un score global pour chaque produit dans les cas d'utilisation, les notes des produits sont multipliées par les pondérations pour obtenir le score du produit dans les cas d'utilisation.

Les capacités critiques sélectionnées par Gartner ne représentent pas toutes les capacités d'un produit ; par conséquent, peuvent ne pas représenter les plus importants pour une situation d'utilisation ou un objectif commercial spécifique. Les clients doivent utiliser une analyse des capacités critiques comme l'une des nombreuses sources d'informations sur un produit avant de prendre une décision de produit/service.

**Learn how Gartner
can help you succeed**

Become a Client

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](#)."

[About](#) [Careers](#) [Newsroom](#) [Policies](#) [Site Index](#) [IT Glossary](#) [Gartner Blog Network](#) [Contact](#) [Send Feedback](#)

Gartner

© 2021 Gartner, Inc. and/or its Affiliates. All Rights Reserved.