Magic Quadrant pour les solutions de qualité de données augmentée

10 mars 2025 - ID G00813007 - 58 min de lecture

Par Melody Chien , Divya Radhakrishnan et 1 autre

Les solutions de qualité de données augmentée transforment les processus traditionnels de qualité de données et le marché, grâce à l'IA/GenAI qui fournit des données fiables et prêtes à l'emploi. Cette étude aide les responsables des données et de l'analyse à comprendre les technologies émergentes et le paysage des fournisseurs afin de prendre de meilleures décisions d'achat.

Hypothèses de planification stratégique

- D'ici 2027, 70 % des organisations adopteront des solutions modernes de qualité des données pour mieux soutenir leur adoption de l'IA et leurs initiatives commerciales numériques.
- D'ici 2027, 80 % des principaux fournisseurs de qualité de données exploiteront les grands modèles linguistiques (LLM) et le traitement du langage naturel (NLP) pour permettre l'inférence interactive des utilisateurs afin d'améliorer la productivité des utilisateurs et l'efficacité des outils.
- D'ici 2027, l'application de l'IA générative (GenAI) accélérera de 40 % le délai de rentabilisation des programmes de gouvernance des données et des analyses et de gestion des données de référence (MDM).

Définition/Description du marché

Gartner définit les solutions de qualité augmentée des données (ADQ) comme un ensemble de fonctionnalités permettant d'optimiser l'expérience de qualité des données. Elles visent à améliorer la découverte d'informations, les suggestions d'actions optimales et l'automatisation des processus en exploitant les fonctionnalités d'IA/machine learning (ML), l'analyse de graphes et l'analyse des métadonnées. Chacune de ces technologies peut fonctionner indépendamment ou en coopération pour créer des effets de réseau permettant d'accroître l'automatisation et l'efficacité dans un large éventail de cas d'utilisation de la qualité des données. Ces solutions dédiées incluent diverses fonctionnalités telles que le profilage et la surveillance ; la transformation des données ; la découverte et la création de règles ; la mise en correspondance, la liaison et la fusion ; la prise en charge active des métadonnées ; la correction des données et l'ergonomie basée sur les rôles.

Ces solutions packagées facilitent la mise en œuvre et le soutien de l'assurance qualité des données, généralement intégrées à une stratégie plus large de données et d'analyse (D&A). Parmi les cas d'utilisation existants et à venir, on peut citer :

- Développement de l'analyse, de l'intelligence artificielle et de l'apprentissage automatique
- Ingénierie des données
- Gouvernance D&A
- Gestion des données de base
- Qualité des données opérationnelles/transactionnelles

Capacités indispensables

Les capacités indispensables pour ce marché incluent :

- Il s’agit de la capacité d’accéder et d’appliquer la qualité des données à un large éventail de sources de données, notamment les sources de données internes/externes, au repos/en streaming, sur site/dans le cloud et relationnelles/non relationnelles.
- **Profiling and monitoring/detection:** This involves 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 to enable them to identify data quality issues. Results from profiling should be able to drive the ongoing monitoring for data quality issues based on preconfigured, custom-built monitoring rules (or adaptive

rules) and alert violations. This automatically detects outliers, anomalies, patterns and drifts.

It also includes a monitoring dashboard, log files or audit trail for compliance requirements. Augmented profiling capabilities include automatic semantic discovery of data elements through analysis or integration to metadata management tools, and selection or recommendation of algorithms based on semantic values to enhance analysis.

Augmented solutions should provide recommendations to users about new alerts to add based on anomalies detected. Solutions should actively learn from user behavior and feedback about which issues are not relevant and then manage notifications accordingly.

- **Matching, linking and merging:** This involves 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. Augmented solutions utilize AI and ML models to automatically suggest potential matches and can tune the results based on user feedback. For merging tasks, consolidation rules for merging data are automatically suggested and refined based on user feedback. User involvement in terms of selecting algorithms, constructing specific match and consolidation rules, and configuration and tuning or matching parameters should be minimal.

- **Business-driven workflow and issue resolution:** These are processes and a 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 as facilitated by collaboration, pervasive monitoring and case management. Augmented solutions can initiate and assign data quality issues leveraging and activating business, technical and operational metadata.

- **Rule discovery, creation and management:** This is 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. Augmented solutions will support the creation of ML-supported data quality rules through a training interface that uses:

    - Unsupervised algorithms that can infer and create data quality rules automatically

- Natural language statements that describe and can execute a data quality rule

Standard Capabilities

The standard data quality capabilities for the stand-alone or unified data management platform market include:

- **Data curation and enrichment:** Augmented solutions support integration with third-party AI models, such as large language models (LLMs), to perform data quality functions. They also can integrate externally sourced data to improve completeness and add value.
- **Active metadata support:** This is the ability to collect, discover or import metadata from partners as well as to build or import lineage to perform rapid root cause analysis of data quality issues and impact analysis of remediation. This also involves applying passive and active metadata findings and making use of metadata-based rule recommendations and associations, as well as data discovery and cataloging. It includes a metrics view based on critical data elements.
- **Deployment environment, architecture and integration with other applications:** This covers styles of deployment, hardware and operating system options; configuration of data quality operations and processes; and interoperability with third-party tools. Augmented solutions can integrate observability and monitoring metrics with DataOps tools to improve notification, analysis and orchestration of issues as they pertain to data flows.
- **Multidomain support and address validation/geocoding:** This is the ability to address both multiple data subject areas (such as various master data domains and vertical industry domains) and depth of packaged support (such as prebuilt data quality rules) for these subject areas. This also includes capabilities that support location-related data standardization and cleansing, as well as completion for partial data in real-time or batch process. For augmented solutions, multidomain support should be coupled with the ability to automatically recommend or deploy any prepackaged content based on the explicit or inferred semantics of the data being profiled.
- **Role-based usability:** This is the suitability of the solution to engage and support the various roles (especially nontechnical business roles) required in a data quality initiative. Roles include data engineers, stewards, data quality analysts, data architects, data integration analysts, business analysts, data preparation experts and so on.

- **Data transformations (parsing, cleansing and standardizing):** This involves the decomposition and formatting of diverse datasets based on government, industry or local standards; business rules; knowledge bases; metadata; and machine learning. This also involves the modification of data values to comply with domain restrictions, integrity constraints or other business rules. Augmented solutions may use a combination of supervised and semisupervised AI and ML models to parse, standardize and cleanse data.

Magic Quadrant

Figure 1: Magic Quadrant for Augmented Data Quality Solutions

Gartner

Vendor Strengths and Cautions

Ab Initio

Ab Initio is a Challenger in this Magic Quadrant. This is its first appearance in the Magic Quadrant. It is headquartered in Lexington, Massachusetts. Its data quality product is Ab Initio Data Quality Environment (DQE), which can be sold as a stand-alone product or included as data quality components in Ab Initio Data Platform. Ab Initio Data Platform is a unified data management platform that contains data integration, data quality, metadata management and data governance features. Ab Initio supports deployment on-premises or in the cloud via Kubernetes or virtual machines. It does not have a SaaS offering, while its

platform as a service (PaaS) offering is a standard single-tenant deployment using Kubernetes/containers. Ab Initio currently has around 1,900 customers in the data quality products. Its operations are geographically diverse and primarily targets mid- or large-sized enterprises, with primary customers across sectors including financial services, telecom and healthcare.

Strengths

- **Product offerings:** Ab Initio offers a stand-alone data quality product as well as a unified data management and governance platform that embeds data quality (DQ) features. Unlike most software companies, Ab Initio supports all versions of its products with long-term backward compatibility. It also supports distributed architecture, which allows its customers from on-premises or Kubernetes in the cloud environments to use the same core products and the same licenses.

- **Technical support and customer services:** Gartner Peer Insights reviews indicate technical support for implementation and speedy responses from customer services to solve their ongoing issues. The collaboration between the account management teams and support teams also results in high customer satisfaction.

- **Performance and scalability:** Ab Initio's product delivers high performance and throughput when processing large volumes of data, and is easily scaled based on the business's requirements. Peer Insights reviews also indicate positive feedback on gaining business efficiency with Ab Initio's faster data processing.

Cautions

- **Industry and enterprise focus:** Although Ab Initio's general-purpose product is applicable across all industries, the company focuses its sales and marketing efforts on highly regulated industries (e.g., financial services and healthcare) with strong compliance requirements. In addition, it primarily targets midsize or large enterprises, with less focus on smaller organizations.

- **Multidomain support:** Ab Initio provides limited support in data domains such as location and geospatial data, IoT data and physical assets. There is also no prebuilt packaged functionality to support this data.

- **Training and documentation:** Based on feedback from Gartner Peer Insights, Ab Initio's customers express challenges accessing training material such as online tutorials, webinars and videos. Documentation, such as error messages, can sometimes be

confusing. Search function performance in the document is limited as well. The vendor has addressed this issue with the latest version.

Anomalo

Anomalo is a Niche Player in this Magic Quadrant. This is its first appearance in the Magic Quadrant. Founded in 2018, it is headquartered in Palo Alto, California. Its data quality product is named Anomalo. The vendor has nearly 60 customers located mostly in North America, with a few in EMEA. Its clients are primarily in the financial services, insurance and consumer goods sectors.

Strengths

- **Data profiling, and monitoring features:** Anomalo provides solid capabilities in monitoring driven by AI/machine learning (ML) to empower users to detect and resolve DQ issues. The platform can perform sophisticated profiling on any tabular data to detect anomalies through unsupervised ML. The profiling also comes with over 40 custom checks to detect deviations from expected patterns or violations of monitoring rules at the record-level inspection.

- **GenAI-enabled rule creation:** Anomalo leverages NLP and LLMs to support rule creation. The tool converts rules into SQL when users provide desired logic in NLP. Anomalo's LLM can also identify and configure the most relevant data quality rule, populating custom options as needed, if desired outcomes are provided. These features help users at various skill levels create appropriate data quality rules on their own.

- **Unstructured data support:** Anomalo provides support for unstructured data analysis that is catered by analyzing text data stored within SQL table columns, enabling users to detect context-specific patterns, anomalies and themes within unstructured text values. The vendor has added unstructured data monitoring, which provides unsupervised analysis of text-based documents targeted for GenAI applications.

Cautions

- **Market presence and mind share:** Anomalo is a startup company considered relatively small compared with its competitors. It has a limited market presence with a small customer base and workforce. Anomalo currently has 100% direct sales in five countries and does not have support for VAR or distributed partnership. It is rarely mentioned in

Gartner client inquiry, and it is less likely than other vendors in this report to be shortlisted in competitive situations seen by Gartner.

- **Innovation focus and integration:** Anomalo invests AI and GenAI technologies in specific areas such as profiling, detection, monitoring and rule creation, but has not yet widely applied the same technologies to other DQ functionalities. It provides limited lineage support, which is currently only at the table level for data in Databricks, Snowflake and BigQuery. Integration with broader data management and data governance tools is expected. The vendor currently works primarily with Alation and Atlan.

- **Feature limitations:** Some features are not fully built out by Anomalo. For example, Anomalo is only available in the English language, and the tool lacks a built-in user interface to parse or standardize data, instead performing such functions through the Anomalo Data Transformation API. Its data resolution workflow relies on integrations with tools like Jira and ServiceNow for certain capabilities, and the overall product may not be intuitive to business users that lack foundational knowledge of the data in which they are interested.

Ataccama

Ataccama is a Leader in this Magic Quadrant. It is headquartered in Toronto, Canada. Its data quality product is Ataccama ONE Data Quality Suite, which is part of the Ataccama ONE platform. The platform provides integrated data management and governance capabilities, blending metadata management and master data management capabilities with its core data quality offerings. Ataccama has around 570 customers for this product line. Its operations are mostly in North America and EMEA, with clients primarily in the financial services, manufacturing and insurance sectors.

Strengths

- **Production innovation:** Ataccama introduced its ONE AI agent in the Ataccama ONE platform. ONE AI provides AI assistant capabilities to minimize lengthy and manual processes such as asset description generation, and generating or suggesting DQ rules. It also has built-in ML algorithms to automatically perform data transformation, detect outlier and anomaly at the record level, and conduct time series DQ analysis and learnable ML suggestions for data-matching proposals

- **Data quality support in Snowflake:** Ataccama introduced a free DQ app in the Snowflake AI Data Cloud. The tool is available directly in Snowflake UI, and provides more than 50

predefined DQ rules for data validation. This app is free to all Snowflake customers and does not require purchasing the Ataccama ONE unified platform. Data is processed right within the Snowflake platform to minimize the data movement. Gartner sees Snowflake customers expanding into Ataccama's broader platform features due to this offering and adding to its revenue.

- **AI-enabled documentation portal:** Ataccama launched an online documentation portal based on AsciiDoc architecture and powered by ONE AI. With that, users can run a full-text search and simply ask what they want to know instead of reading lengthy documentation. The new portal also provides a feedback loop mechanism to improve the content. The document portal is available for the public as well.

Cautions

- **SaaS multitenant deployment option:** Ataccama currently does not provide SaaS with a multitenant deployment option. It does plan to have that available in early 2025. PaaS with a single tenant deployment is available through Ataccama Cloud, a vendor-managed cloud environment.

- **Unstructured data support:** Ataccama has not yet demonstrated more development in providing unstructured data support. Most capabilities for unstructured data quality are on Ataccama's roadmap. The vendor currently provides support for text-based data type natively. Ataccama ONE allows integration with external services focused on specialized categorization of unstructured files, enabling the metadata results to then be used in all data quality processes, such as data classification and DQ evaluation, which are not locally available in Atacama's products.

- **Consulting support and training:** Gartner Peer Insights reviews indicate that Ataccama customers heavily rely on Ataccama consultants because the tool is complex to use for beginners. Training material for various levels of skills is limited.

CluedIn

CluedIn is a Niche Player in this Magic Quadrant. It is headquartered in Denmark. CluedIn's core product, CluedIn, is distributed under two product deployments — master data management and data quality — each of which can be deployed stand-alone. CluedIn offers self-hosted, PaaS, managed service and SaaS deployment options. Its data quality product has nearly 90 customers, which are mostly in EMEA, with some in North America. Its clients are primarily in the financial services, insurance and consumer goods sectors.

Strengths

- **Strategic partnership with Microsoft:** CluedIn has a strong partnership with Microsoft. The CluedIn product is a native solution for master data management and data quality in Microsoft Azure. It also integrates natively with 27 Azure services, such as Microsoft Purview, Azure OpenAI and Power BI. It can be sold through Azure Marketplace. This partnership provides extensive resources in marketing and sales opportunities to CluedIn.

- **Data management platform based on Azure OpenAI:** CluedIn uses Azure OpenAI as part of its AI/ML layer for supporting data management capabilities such as MDM, data governance and data quality. The vendor provides AI-driven mapping, profiling, validation, standardization, rule generation and lineage. The product also enriches, corrects, validates and standardizes data with conversational prompts in different languages.

- **Pricing and licensing strategy:** CluedIn provides competitive pricing options, including consumption-based, record-based options and a free SaaS option. Its community version offering is free of charge up to 10,000 records and 4 CPUs, with additional fees for processing/CPU cores. The community version has all of the standard commercial version features and can be deployed on any hyperscaler supporting Kubernetes or Docker.

Cautions

- **Exclusive partnership with Microsoft:** CluedIn currently has an exclusive partnership with Microsoft. The vendor's product strategy is primarily targeted at Microsoft Azure's environment as a master data management tool. Although CluedIn's close partnership with Microsoft offers advantages such as access to Microsoft's large partner network, availability of the full feature set of CluedIn outside of Azure is limited. CluedIn's product can be deployed to non-Azure clouds, but not all features are fully supported by the vendor outside of the DQ integration environment.

- **Product marketing strategy:** The CluedIn product is currently marketed as an MDM tool instead of a data quality tool. Data quality features are under the umbrella of MDM tools, which causes confusion to its potential customers.

- **Market growth and traction:** Although CluedIn showed revenue growth in 2024, the new customer numbers remained flat. Its geolocation presence is limited to Denmark, Australia, the U.K., Belgium and Germany, and it has no presence in North America. CluedIn is rarely mentioned by users of Gartner's client inquiry service.

DQLabs

DQLabs is a Visionary in this Magic Quadrant. It is headquartered in Pasadena, California. Its data quality product is DQLabs Modern Data Quality Platform. DQLabs currently has 125 direct customers for its data quality product and a larger number of users via its OEMs with Hitachi Pentaho DQ and Quest erwin Data Intelligence. Its customers are mostly in North America, with some in EMEA, and they are primarily in the banking, technology and healthcare sectors.

Strengths

- **Sales growth:** DQLabs has demonstrated solid growth in revenue and its customer base for its data quality and data observability products. The vendor claims an annual revenue growth rate from 70% to 80% and a double increase in its number of customers from the previous year.
- **DQ and data observability integration:** DQLabs offers a unified platform that integrates data quality and data observability. By integrating these two components, the vendor provides a streamlined process to detect, monitor and resolve data quality issues across the data systems. The two components can also be purchased separately as two different modules.
- **Innovation:** DQLabs invests in AI innovation to improve automation of DQ processes. It introduced automated anomaly detection, root cause analysis and compliance risk monitoring. DQLabs also built its AI Center of Excellence with more than 50 dedicated experts, driving innovations in AI and GenAI.

Cautions

- **Market presence and mind share**: Despite the growth of sales, DQLabs is still lacking market presence, with rare mentions by users of Gartner's client inquiry service. There is limited market execution via social media channels to showcase its thought leadership or technology innovation.
- **Unstructured data support**: DQLabs currently provides limited capabilities in unstructured data, compared with its competitors. For example, applying validation logic or highlighting outliers of data sources from one another, or parsing and classifying data from unstructured data sources into specific categories based on user defined rules are

only partially supported. The platform's basic features need more development effort to fully support the requirements of unstructured data.

- **Cost for diverse environments:** Although DQLabs charges a fair price for basic usage, it charges extra for connecting to additional data sources (e.g., Snowflake, Databricks); data pipeline, data catalog or analytics applications (e.g., Kafka, Collibra, Tableau); and messaging tools (e.g., Teams, Slack). Enterprises with complicated and diverse environments are likely to see a high cost for their implementation. DQLabs does offer enterprise license agreements that may make diverse deployments more cost-effective.

Experian

Experian is a Challenger in this Magic Quadrant. It is headquartered in Dublin, Ireland. Its data quality products include Experian Aperture Data Studio, Experian Governance Studio, Experian Batch, and several data validation and enrichment services such as Experian Address Validation, Experian Email Validation and Experian Phone Validation. Aperture Data Studio supports on-premises, private or public cloud, or hosted infrastructure. Experian has an estimated 6,000 customers for these products, with most using data validation and enrichment services. Its operations are geographically diversified, and its clients are primarily in financial services, retail, public sector and insurance.

Strengths

- **Augmented features:** Experian continued building out product features by delivering ML-powered profiling, smart rules suggestions, auto tagging, standardization, matching and clean-up transformations. Natural language processing is currently supported to create complex transformation functions. GenAI and natural language-based interactions to generate recommendations for new rules and for new transformations are in pilot now with a limited customer base.

- **Identity resolution:** Experian's data enrichment service offerings and prebuilt packages with industry-specific rules (e.g., financial credit, housing, education, healthcare) enable comprehensive identity resolution activities for a person or business entity.

- **Integration:** Gartner Peer Insights reviews indicate that Experian customers like Aperture Data Studio's intuitive approach to developing profiling and transformation functions. The tool also provides expanded support for automation, with over 100 APIs now available to automate processes and integrate with third-party applications. Pushdown processing

has also been added for common database environments, minimizing data movement requirements or the need to write complex SQL code for profiling or validations.

Cautions

- **Focus:** Experian's primary focus is party data content, servicing business and consumer-related information. Experian also states coverage for other data types, e.g., financial, products and materials. Organizations with strong data quality requirements across diverse data domains should carefully evaluate Experian's capabilities against their data domain priorities.
- **GenAI innovations:** Experian's current solutions have limited features that leverage GenAI, compared with other vendors in this research. Although the release is not yet generally available, Experian does report a roadmap of anticipated innovations planned for the v3.0 release, targeted for availability in 1H25.
- **Unstructured data support:** Experian solutions do not provide built-in support to parse and classify content from unstructured data sources, nor do they analyze for outliers, generate context-specific metadata about the information, or perform sentiment analysis. These entity extraction requirements are currently met through the use of custom parsers for semistructured, JSON-formatted content.

IBM

IBM is a Challenger in this Magic Quadrant. It is headquartered in Armonk, New York. Its data quality products are Cloud Pak for Data (specifically, IBM Knowledge Catalog [IKC] and IBM Match 360), IBM Databand, IBM DataStage, IBM Manta Unified Data Lineage, and legacy products InfoSphere Information Analyzer and QualityStage. IKC utilizes container-based and microservices-based architecture based on the Red Hat OpenShift layer for on-premises and/or cloud. Gartner estimated around 2,800 customers for these product lines. Its operations are geographically diversified, with clients in various sectors.

Strengths

- **Integrated data management platform**: IBM offers a comprehensive and integrated platform for managing enterprise data governance, including data quality, metadata and policy management. Capabilities from IBM reflect the vendor's market understanding and, therefore, its investments in key technologies to build an integrated data

management platform that embeds active-metadata-based insights and provides persona-based user experiences and container-based deployment.

- **Acquisitions:** With its most recent acquisitions of Manta Software (data lineage vendor in October 2023) and StreamSets (real-time data integration vendor in July 2024), IBM expands its data management technologies by providing visibility into data flows. This visibility allows users to pinpoint the source of issues and address real-time data ingestion of structured and unstructured data, as well as process and deliver them for complex data engineering pipelines in hybrid, multicloud environments.
- **Offering focus:** IBM provides Knowledge Accelerators that can be loaded into the IBM Knowledge Catalog to provide predefined, extensive and curated glossaries to improve data classification, regulatory compliance and self-service analytics. The Accelerators can deliver industry-focused glossaries using LLM technology to align data quality programs with specific industry focus.

Cautions

- **Innovation:** For the past year, IBM's innovation related to data quality process in IBM Knowledge Catalog has lagged behind its competitors. For example, its customers want more sophisticated data profiling and interactive visualization for analyzing the results. IBM has much in the roadmap planning stage for this but has yet to deliver it.
- **Unstructured data support:** IBM currently provides very limited functionality in its current releases, and most features to support unstructured data are on its roadmap. The pace of IBM's development in this area is slow and lagging other vendors in this research.
- **GenAI-enabled DQ features:** Most of IBM's GenAI-enabled DQ features are partially supported or on the future roadmap — for example, the ability to leverage LLMs to standardize structured data, or enrich or obtain missing value, and the ability to generate DQ rules. IBM has an AI assistant in preview that provides answers to natural language questions and guides users through certain activities in the product.

Informatica

Informatica is a Leader in this Magic Quadrant. It is headquartered in Redwood City, California. Its data quality products are Intelligent Data Management Cloud (IDMC) and Informatica Data as a Service to verify addresses, phone numbers and emails, and Informatica Data Quality (IDQ, on-premises solution). IDMC supports multitenant SaaS or

PaaS, and its runtime component may be installed on-premises or private cloud. Data as a Service offerings are SaaS only. Informatica has 5,000 customers for all product offerings. Its operations are geographically diverse, with clients across all major sectors.

Strengths

- **Revenue:** Informatica reported a 6% growth in total revenue for fiscal year-end 2023. As of 3Q24, Informatica reported a 60% year-over-year (YoY) growth in its Cloud Data Quality business. Informatica reported a 35.8% non-GAAP operating margin as of 3Q24, representing a 450 basis point improvement YoY.

- I**nnovation:** CLAIRE GPT and AI copilot are generally available and contribute to cloud business growth. CLAIRE GPT provides a natural language interface to users. CLAIRE AI copilot automates routine data management tasks. Informatica launched Cloud Data Access Management (CDAM) in 1Q24, providing AI-powered solutions to automate data access policy enforcement, following its acquisition of Privitar in August 2023. CDAM is integrated into IDMC to support data access governance.

- **Partnerships:** Informatica has a strong global partner ecosystem with over 500 global partners and 7,500 certified implementation partners. Informatica announced general availability (3Q24) of IDMC platform services for Oracle Cloud Infrastructure (OCI). Data quality integrations with Microsoft Fabric continue to be enhanced, with plans for remediation features in 2025. In June 2024, Informatica announced expanded IDMC support for Databricks and Snowflake. In April 2024, INFA announced expanded IDMC support for Google BigQuery and AI use cases.

Cautions

- **Complex use cases:** Clients with requirements for complex data transformations should evaluate differences between IDMC and Informatica Data Quality. IDMC has appealing modern features and a strong future roadmap for AI-powered data transformations and guidance, but some clients report that some complex scenarios are not yet supported — for example, real-time data consistency monitoring for system-to-system comparisons.

- **Support:** Informatica has a large number of on-premises clients that use IDQ along with Axon and EDC. Informatica now executes with a cloud-only (IDMC) product strategy. While Informatica continues to support clients on these solutions, the focus is to provide essential maintenance updates moving forward. INFA does offer migration programs to enable customers to modernize these implementations to IDMC.

- **Pricing and licensing:** Gartner inquiry calls and Peer Insights reviews highlight concerns with IDMC's pricing model due to the requirement to forecast annual feature usage, as unused prepurchased credits do not carry over to subsequent years.

Irion

Irion is a Niche Player in this Magic Quadrant. This is its first appearance in the Magic Quadrant. It is headquartered in Turin, Italy. Its data quality product is part of Irion EDM, an enterprise data management platform. Irion EDM can be deployed on-premises, hosted off-premises and deployed in fully managed IaaS environments. The vendor has around 70 customers that are located mostly in EMEA, with a small number located in North America. Its clients are primarily in the banking, securities, and insurance industries, with a small percentage in the utility sector.

Strengths

- **Support in distributed data quality processes:** Irion EDM is an enterprise data management platform with focus on data quality and governance capabilities. Using a hub and spoke architecture, the vendor provides a centralized hub that can connect to Irion or third parties' spokes, which aggregate and transform data from diverse sources. The platform also creates and executes data quality checks and business rules to where they are needed. The vendor currently provides connectors to more than 200 data sources.

- **Technology innovation:** Irion EDM is powered by Data Artificial Intelligence SYstem (DAISY), which provides a copilot wizard to quickly create data transformation or DQ rules. It also automates manual DQ processes by offering data profiling to automatically identify the characteristics of a new dataset, and recommends or enables the generation of rules to validate formats, key fields, domain values and business logic.

- **Customer experiences:** Based on Gartner Peer Insights feedback, Irion customers are typically happy with the vendor's customer support, especially its strong consultancy skills that help customers smoothly manage projects.

Cautions

- **Sales and geographic strategies:** Irion is currently operating primarily in Italy and Spain, and its clients are predominantly located in EMEA (96% of sales) and North America (4% of sales). Irion is primarily doing direct sales on its own (92% of revenue is direct sales).

Irion will need to expand its outreach through partnerships to cover other geographic locations.

- **Performance and scalability:** Gartner Peer Insights reviewers report that loading and replicating data within the Irion environment to perform DQ checks can lead to inefficiencies. Push-down process to some data environments, such as cloud data warehouse, is not available at this time.

- **Matching, linking and merging:** Irion's support in these areas is considered basic and traditional, primarily based on exact-value fuzzy matching. Packaged learnable AI algorithms and techniques to automate or augment matching based on user's actions and linguistic and cultural nuances are not available as out-of-box offerings.

Precisely

Precisely is a Challenger in this Magic Quadrant. It is headquartered in Burlington, Massachusetts. Its data quality products include the Precisely Data Integrity Suite, Trillium Quality, Spectrum Quality and Data360. Data Integrity Suite is a SaaS offering or it can be deployed on-premises with agents. Precisely's other data quality solutions support on-premises, private, hosted or public cloud deployments. Gartner estimates approximately 4,900 data quality solution customers. Its operations are geographically diverse, with clients in all sectors, including financial services, insurance and telecommunications sectors.

Strengths

- **Partnerships:** Precisely demonstrated 8% YoY growth in data quality revenue. It has co-sell partnerships with technology vendors such as Amazon Web Services (AWS), Databricks, Microsoft and Snowflake, including supported integrations. Precisely also provides matching, address, email and phone validation features via an OEM relationship with Salesforce. Its partnerships further extend Precisely's sales opportunities and industry coverage.

- **Portfolio:** Precisely's Data Integrity Suite leverages AI to generate business descriptions of cataloged data, and GraphQL is utilized in the address data enrichment process. ML-based analysis learns from past data cleansing decisions to refine future suggestions. Specialized features of Spectrum, Trillium and Data360 are exposed within the Data Integrity Suite, which provides a consistent user experience overall.

- **Implementation:** Gartner Peer Insights reviews indicate positive feedback regarding Precisely's services, helping clients determine where to begin, with alignment to business

goals. Implementation time frames range from a week to a few months, depending on engagement scope.

Cautions

- **Coverage:** Precisely's solutions provide their deepest functionality for transformation, enrichment, matching, and resolution of address- and location-related information. Precisely states support for a wide variety of data domains beyond location. However, organizations that require broad coverage of diverse data domains should carefully evaluate Precisely's solutions for a comprehensive fit across all data scenarios.
- **LLM innovations:** Precisely's solutions support interactions with external LLMs, but most enhancements in this area remain on the roadmap. For example, the use of LLMs/NLP is planned for rule discovery and generation, and to support issue resolution workflows. Companies should evaluate these feature areas where they are relevant to their requirements and monitor them for future general availability.
- **Strategy and support:** Precisely indicates that most significant enhancements are planned for the Data Integrity Suite. Historical products continue to support existing use cases and receive minor updates, with bridges to the Data Integrity Suite where appropriate. Organizations using historical solutions should contact Precisely about long-term support changes against their business requirements and consider migration implications or solution alternatives where relevant.

Qlik

Qlik is a Leader in this Magic Quadrant. It is headquartered in King of Prussia, Pennsylvania. Qlik consolidated all data governance and data quality capabilities under Qlik Talend to facilitate data foundations for AI, analytics and operations, following its May 2023 Talend acquisition. Its data quality products include Qlik Talend Cloud, Talend Data Fabric, Talend Data Catalog and Qlik Answers. Talend Open Studio (a free, open-source product) has been discontinued. Qlik supports hybrid cloud and on-premises offerings, and a SaaS offering — Qlik Talend Cloud. Qlik has around 3,000 active customers for its data quality offering. Its operations are geographically diversified, and clients are primarily in financial services, manufacturing and retail/wholesale industries.

Strengths

- **Revenue growth:** Qlik continues to grow strongly, with a revenue increase of more than 100% in its overall data management software for 2023 based on Gartner's market share report. Qlik also claimed 37% new license growth. The cross-selling opportunities within the existing Qlik customer base (more than 40,000) have great potential for boosting its sales. Its strong partnership network (more than 1,850 partners) also expands its outreach opportunities.

- **Technology innovation:** Qlik has demonstrated its technology innovation in the DQ area to support AI use cases. For example, it provides a Trust Score for AI to evaluate the DQ for AI use cases, and new OpenAI and Pinecone integrations supporting retrieval-augmented generation (RAG) pipelines for unstructured and structured data. Qlik also now supports running data preparations through API calls, enabling easy automation and enhancing data transformation capabilities.

- **Unstructured data support**: Qlik announced its acquisition of Kyndi in January 2024. This acquisition was aimed to add support for interpreting and processing unstructured data. By leveraging the combination of Qlik Cloud's structured data and Kyndi's unstructured data technologies, Qlik is well-positioned to enhance its capabilities in processing and utilizing data with trusted quality.

Cautions

- **AI/ML-driven rule management:** Qlik currently does not support using ML techniques to infer and generate DQ validation rules across multiple attributes. It does not support the identification of duplicate or similar rules through the analysis of metadata, and then the notification of developers of duplicate rule creation.

- **Reporting and dashboard:** Qlik provides prebuilt reports and dashboards for data quality assessment, but provides a limited customization ability to those reports.

- **Learning curve and documentation:** Gartner Peer Insights reviewers commented on the complexity of the tool, and that the learning curve for new users can be steep. Its customers also expressed a desire for more comprehensive documentation and training resources to reduce the learning curve.

SAS

SAS is a Niche Player in this Magic Quadrant. It is headquartered in Cary, North Carolina. Its data quality products are part of components in the SAS Viya platform and SAS Event

Stream Processing, as well as SAS Data Management Advanced and SAS Data Quality Desktop from SAS 9 Platform. SAS Viya is a containerized cloud-agnostic platform that can be deployed on-premises or on cloud platforms. The SaaS option is currently not available. Gartner estimated around 2,670 customers for these product streams. Its operations are geographically diversified, with clients in various sectors.

Strengths

- **Unstructured data support:** SAS demonstrates strong unstructured data support by offering capabilities to analyze unstructured data sources and generate context-specific metadata. SAS Viya can also integrate Visual Text Analytics features to facilitate the extraction of entities and topics. It can also apply validation logic and identify outliers across data sources.

- **Data enrichment via synthetic data generation:** SAS Data Maker (based on technology from Hazy's software through the acquisition in November 2024) is a new tool to generate synthetic data. It includes a GUI interface and an API for synthetic data generation and quality assessment. The tool is currently in review in Azure marketplace and is expected to be in GA in 1H25.

- **Integration within the SAS ecosystem and with open-source tools:** SAS provides integration with other SAS products like SAS Data Integration Studio and SAS Visual Analytics that allow a complete analytics ecosystem. The vendor has also strengthened integration with open-source tools such as Great Expectations. SAS customers can use Great Expectations as a business rule engine for rule creation, and apply them in SAS environments.

Cautions

- **Marketing strategy and execution:** SAS positions and markets the SAS Viya primarily as a DSML and analytics platform, even though it also has data quality and data integration features. The vendor's marketing strategy has been skewed more toward the data science and analytics area than data quality and data management. Gartner client inquiry data reveals a reduction in mind share for SAS data quality solutions. It is rarely mentioned by users of Gartner's client inquiry service and is not often shortlisted during the vendor selection process.

- **GenAI functionality:** SAS has not been an early adopter of GenAI and lags in its adoption compared to other vendors in this research. Many of its GenAI functions embedded in the

platform are currently in preview form. SAS Viya Copilot is in private review. Some features, such as autogenerating technical data quality rules using natural language processing, are on the roadmap.

- **Cost and training:** Gartner Peer Insights reviews indicate that SAS products come with higher-than-market-average license and implementation costs, which can be limiting for budget-conscious small and midsize businesses. Its sophisticated features can be complex for beginners, requiring training and expertise to utilize tools effectively. SAS customers should consider leveraging the vendor's offering of online/in-person training and consulting services.

Vendors Added and Dropped

We review and adjust our inclusion criteria for Magic Quadrants as markets change. As a result of these adjustments, the mix of vendors in any Magic Quadrant may change over time. A vendor's appearance in a Magic Quadrant one year and not the next does not necessarily indicate that we have changed our opinion of that vendor. It may be a reflection of a change in the market and, therefore, changed evaluation criteria, or of a change of focus by that vendor.

Added

- **Ab Initio:** Ab Initio is an established vendor with wider data management capabilities as part of the larger Ab Initio Data Platform. Its stand-alone offering for data quality is called Ab Initio DQE, which meets the current inclusion criteria.
- **Anomalo:** Anomalo is a new entrant into the augmented data quality solutions market and meets the inclusion criteria.
- **Irion:** Irion's platform, Irion EDM, is an enterprise data management platform and includes data quality as a capability. Irion meets the inclusion criteria.

Dropped

- **Collibra:** The vendor was dropped because it does not provide cloud-based/SaaS deployment options.
- **Datactics:** The vendor was dropped due to lack cloud-native data quality capabilities and unstructured data support.

- **MIOsoft:** The vendor was dropped due to lack of support for augmentation of critical data quality functions that leverage AI/ML features, graph analysis and metadata analytics in its product offerings at the time of evaluation.
- **SAP:** The vendor was dropped because the product SAP Datasphere was not positioned for general-purpose data quality use cases and required additional SAP components to fully address data quality scenarios at the time of evaluation.

Inclusion and Exclusion Criteria

To qualify for inclusion, vendors must meet all following inclusion 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 or unified data management platforms must demonstrate that these are integrated and collectively meet the full inclusion criteria for this Magic Quadrant.
- Deliver critical augmented data quality functions at a minimum (descriptions are the same as given in the Market Definition):
  - Profiling and monitoring/detection
  - Data transformations
  - Rule discovery, creation and management
  - Matching, linking and merging
  - Active metadata support
  - Usability, workflow and issue resolution
  - Alerts, notifications and visualization
  - Unstructured data support
- Support augmentation of the critical data quality functions listed above by leveraging AI/ML features (supervised, semisupervised or unsupervised methods, NLP-based or LLM-supported), graph analysis and metadata analytics (active metadata).
- Support the above functions in both scheduled (batch) and interactive (real-time) modes.

- Enable large-scale deployment via server-based or cloud-based runtime architectures that can support concurrent users and applications. Cloud-based/SaaS version should support all critical data quality functions independently as mentioned in the above criteria.
- Maintain an installed base of at least 50 production paying customers (different companies/organizational entities) for their flagship data quality product (not individual smaller modules or capabilities). The customers must be running in production for at least six months.
- 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, database administrator, data integration analyst, data scientist, data analyst, business intelligence analyst and a citizen user.
- Provide out-of-the-box and prebuilt data quality rules for the purpose of data profiling and monitoring, cleansing, standardization, and transformation, based on common industrial practices.
- Support integrability and interoperability with other data management solutions such as metadata management, master data management and data integration solutions from third-party tools.
- Provide direct sales and support operations, or a partner providing sales and support operations in at least two of the following regions: North America, South America, EMEA and Asia/Pacific.
- The customer base for production deployment must include customers in multiple countries and in more than one region (North America, South America, EMEA and Asia/Pacific), and be representative of at least three or more industry sectors.
- The solution must demonstrate the capabilities as available in general availability from 15 October 2024.

The following types of vendors 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.

- Vendors that support limited data quality functionalities — no augmentation and automation or addressing of very specific data quality problems (for example, focusing on only address cleansing and validation). They do not provide the complete suite of data quality functionality expected from today's augmented data quality solutions.
- Vendors that support only on-premises deployment and have no option in cloud-based deployment on any public cloud environment (for example, AWS, Azure or Google Cloud).

Evaluation Criteria

Ability to Execute

Gartner analysts evaluate technology vendors on the quality and efficacy of the processes, systems, methods and procedures that enable their performance to be competitive, efficient and effective, and to positively impact their revenue, retention and reputation within Gartner's view of the market.

Gartner evaluates vendors' Ability to Execute in the augmented data quality solutions market by using the following criteria:

- **Product or service:** The ability to address current market needs, which also includes serving the much-needed augmented data quality capabilities. These include using AI and ML to measure data quality, recommending data quality rules, and possible data remediation solutions. We also consider the ability of its product to allow interoperability to open-source solutions and third-party offerings
- **Overall viability:** The vendor's financial strength (as assessed by revenue growth, profitability and cash flow) and the strength and stability of its people and organizational structure. This criterion reflects buyers' increased openness to considering newer, less-established and smaller providers with differentiated offerings. We also consider growth in high-momentum use cases such as AI-ready data support.
- **Sales execution/pricing:** The flexibility, effectiveness, ease of adoption and market suitability of the vendor's pricing and licensing model in light of current and future customer demand trends and spending patterns (for example, operating expenditure and flexible pricing). We consider the ability to provide tools and capabilities through different pricing models appropriate by use cases, persona and environment.

- **Market responsiveness/record:** The degree to which the vendor has demonstrated the ability to respond successfully to market demand for data quality capabilities over an extended period. We look at evidence of how the vendor has "course corrected" to changing buyer requirements in terms of use cases, upcoming capabilities, pricing models and support requirements. For example, we consider the vendor's ability to respond to current market demands for solutions that support AI/ML to automate complex and repetitive data quality tasks such as data transformation and rule creation.
- **Marketing execution:** The overall effectiveness of a vendor's marketing efforts, the degree to which it has generated mind share and the magnitude of the market share achieved as a result. We look at the ability of the vendor to adapt to changing demands in the data quality market by aligning its product message with new trends and end-user interests.
- **Customer experience:** The level of satisfaction expressed by customers with a vendor's product support and professional services. We also assess their overall relationship with the vendor, as well as customer perceptions of the value of the vendor's data quality solution relative to costs and expectations. We evaluate customer feedback on a vendor's ability to meet roadmap deliverables, technical knowledge sharing, skills enablement, augmentation and training.
- **Operations:** The stability of key staff and other means that enable the vendor to operate effectively and efficiently. The areas that we consider include partner programs, skills augmentation, improvements in support and services, training materials and programs, and delivery with external service providers.

Table 1: Ability to Execute Evaluation Criteria

Evaluation Criteria	*Weighting*
Product or Service	High
Overall Viability	Medium
Sales Execution/Pricing	High
Marketing Execution	Medium

Evaluation Criteria	*Weighting*
Customer Experience	High
Market Responsiveness/Record	Medium
Operations	Low

Source: Gartner (March 2025)

Completeness of Vision

The evaluation covers current and future market direction, innovation, customer needs and competitive forces, as well as how well they correspond to Gartner's view of the market.

Gartner assesses vendors' Completeness of Vision in the augmented data quality solutions market by using the following criteria:

- **Market understanding:** The degree to which the vendor leads the market in new directions (in terms of technologies, products, services or use cases). Gartner evaluates the vendor's ability to adapt to significant market changes and disruptions, such as by providing data quality process automation with metadata and ML-driven rule recommendation, or by supporting unstructured data quality for preparing data for AI initiatives. Also considered is the degree to which vendors are aligned with the significant trend of convergence with other data-management-related markets — specifically, the markets for data integration tools, metadata and MDM solutions.
- **Marketing strategy:** Messages, consistently communicated internally and externally through channels, social media, advertising, customer programs and positioning statements. Also considered are the degree to which the vendor's marketing approach aligns with and/or exploits emerging trends (such as adaptive data and analytics governance and business-centric data quality programs) and the overall direction of the market.
- **Sales strategy:** The use of partnerships and also aligning sales models with customers' preferred buying approaches, such as freemium programs and subscription-based

pricing. We consider vendors' growth through varying channels (e.g., OEMs, VARs, SIs, hyperscaler marketplaces, consulting companies, joint go to market, partnerships with vendors in the D&A space).

- **Offering (product) strategy:** The degree to which the vendor's product portfolio and roadmap reflects demand trends, fills current gaps or weaknesses, and emphasizes competitive differentiation. Also considered is the breadth of the vendor's strategy regarding a range of product and service delivery models, from traditional on-premises deployment to SaaS and cloud-based models.
- **Business model:** The vendor's overall design and approach to executing its strategy for the augmented data quality solutions market. This approach includes delivery models, funding models (public or private), development strategies, packaging and pricing options, and partnership types (such as joint marketing, reselling, OEM and system integration/implementation).
- **Vertical/industry strategy:** The degree of emphasis that the vendor places on vertical market solutions. The depth of its vertical market expertise and provision of prebuilt data quality rules or libraries are also considered.
- **Innovation:** The extent to which the vendor demonstrates creative energy in thought leadership and in differentiating ideas and product roadmaps that could significantly extend or even reshape the market in a way that adds value for customers. Particularly, we examine how well vendors support— or plan to support — key trends with regard to personas, data diversity, latency, data quality analytics, intelligent capabilities and deployment, for example.
- **Geographic strategy:** The vendor's strategy in light of global demand for data quality capabilities and expertise. We evaluate the vendor's ability to provide customers with local support, via VARs, other resellers, and channel, vertical/industry and SI partnerships.

Table 2: Completeness of Vision Evaluation Criteria

Evaluation Criteria	*Weighting*
Innovation	High
Market Understanding	High

Evaluation Criteria	*Weighting*
Marketing Strategy	Medium
Sales Strategy	Medium
Offering (Product) Strategy	High
Business Model	Low
Vertical/Industry Strategy	Medium
Geographic Strategy	Low

Source: Gartner (March 2025)

Quadrant Descriptions

Leaders

Leaders demonstrate strength in depth across the full range of data quality functions, including the core data quality capabilities that have existed for years, and bring more automation and augmentation to data quality solutions.

Leaders exhibit a clear understanding of dynamic trends in the data quality market. They explore and execute thought-leading and differentiating ideas, and they deliver product innovation based on the market's demands. Leaders also provide additional insights of data by leveraging advanced technologies such as AI/ML, knowledge graph, active metadata or NLP to minimize manual effort.

Leaders align their product strategies with the latest market trends. These trends include focusing on a nontechnical audience, trust-based governance, growth in data diversity, low data latency, in-depth data quality analytics (not just reporting) and intelligent capabilities. Other trends are new delivery options (such as cloud, hybrid cloud and IoT edge deployment), and alternative pricing and licensing models (such as meter-based, consumption-based or pay as you go, or persona/use-case-based).

Leaders address all industries, geographies, data domains and use cases. Their products support multidomain and alternative deployment options such as SaaS or microservices. They offer excellent support for business roles and easy-to-use visualization, and they include out-of-the-box machine learning capabilities and predictive analytics.

Leaders offer extensive support for a variety of traditional and new data sources (including cloud platforms, data lake, IoT platforms, Apache Hadoop and mobile devices), a trust-based governance model, and delivery of enterprise-level data quality implementations.

Leaders have significant size, an established market presence and a multinational presence (either directly or through a parent company). Leaders also undertake clear, creative and effective marketing, which influences the market, promotes their brand and increases their mind share.

Challengers

Challengers in this research exhibit a strong understanding of the current demands of the augmented data quality market as well as both the credibility and viability to deliver. They also have solid sales and marketing execution. Therefore, they generally have substantial customer bases. Some Challengers are mature in specific capabilities and use cases, which enables them to deliver targeted use cases faster and with a better overall total cost of ownership than other vendors (sometimes even Leaders). Some may also focus on certain ecosystems. These vendors have developed best practices for leveraging their strongest product capability in new delivery models.

Challengers may not have the same breadth of offering as Leaders and/or, in some areas, may not demonstrate as much thought leadership and innovation. For example, they may focus on a limited number of data domains (e.g., customer, product or location data). They may not exhibit full understanding of concepts like augmented data quality and may not have features related to these as part of their vision. They may not have built-in metadata management capabilities but can provide integration with metadata management platforms. The data observability features are basic and not comprehensive.

Challengers may lack capabilities in areas such as real-time profiling on streaming data, ML-based anomaly detection, predictive analysis or support of complicated data landscapes. Leveraging LLMs is primarily in the vision or prototyping stage.

Compared with Leaders, Challengers often exhibit less understanding of some areas of the market, and their product strategies may suffer from a lack of differentiation.

Visionaries

Visionaries are innovators and demonstrate a strong understanding of emerging technology and business trends, or they focus on a specific market need that is far outside of common practices, while also possessing capabilities that are expected to grow in demand. They are aligned with the market in adding features related to automation and augmentation as a part of their roadmaps. They also have a vision toward more unified platforms that allows convergence of capabilities or integration plans with adjacent markets for automation of data quality processes.

Visionaries should lead the push toward the utilization of knowledge graphs, semantics, active metadata and AI/ML for significant automation in data quality design, delivery and monitoring. Visionaries also focus on a nontechnical audience, trust-based governance, growth in data diversity, low data latency, data quality analytics and intelligent capabilities. Also included are new delivery options (such as container-based or IoT edge deployment) and alternative pricing models (such as open source and subscriptions). Visionaries' product capabilities are mostly aligned with these trends, but not as completely as Leaders.

Although Visionaries can deliver good customer experiences, they may lack the scale, market presence, brand recognition, customer base and resources of Leaders. They have good vision but are relatively slow in execution of these great vision elements.

Niche Players

Niche Players often specialize in a limited number of industries, geographic areas, market segments (such as small and midsize businesses) or data domains (such as customer data or product data). They often have strong offerings for their chosen areas of focus and deliver substantial value for customers in those areas. Niche Players may not appear frequently in competitive situations for comprehensive and/or enterprise-class data quality deployments. Many have strong offerings for a specific range of data quality challenges.

Niche Players typically have limited market share and presence, and have limited functionalities or lack financial strength. Niche Players often have to catch up with the latest innovations, such as active metadata support, machine learning and observability.

Niche Players often exhibit advantages in pricing within their established footprint and in vertical or horizontal solutions, but sometimes cannot complement an organization's other data management technologies. Niche Players are known for solving one part of the data quality problems well through a targeted solution and may be the right solution for organizations with less complex needs.

Context

The acceleration of AI and GenAI adoption and the dynamics of business change are creating demand for greater scale and shorter time to value in relation to data management. Data quality is a particular concern, because trusted, high-quality data is vital to the success of AI and business initiatives. According to the 2024 Gartner AI Mandates for the Enterprise Survey, [1] on average, about 40% of AI prototypes make it into production, and participants reported data availability and quality as a top barrier to AI adoption. In addition, more than 64% of organizations stated that data quality and data governance remain one of their top five investment areas in the next two to three years, according to Gartner's 2024 Evolution of Data Management as a Dedicated Function Survey. [2]

All these challenges drive the adoption of augmented data quality solutions. Data is useful only if its quality, content and structure is documented and well-understood. The cost of dirty, insufficient and/or inaccurate data remains a substantial threat. Delivering reliable, trusted and timely data for business consumption is a continuous effort and process that can be supported with modern technologies in augmented data quality solutions.

In this report, we assess 12 vendors. Some are more advanced in augmentation and automation, and some are picking up speed toward the same goal. Use this Magic Quadrant to help find the right vendor and product for your organization's needs. Gartner strongly advises against selecting a vendor solely because it is in the Leaders quadrant. A Challenger, Niche Player or Visionary could be the best match for your requirements. Use this Magic Quadrant in combination with the companion **Critical Capabilities for Augmented Data Quality Solutions**, as well as Gartner's client inquiry service and Peer Insights portal.

Market Overview

The augmented data quality solutions market continually shows steady growth and remains vibrant due to emerging and increasing demands from AI and GenAI adoption and ongoing digital transformation initiatives. AI-related technologies continually disrupt the augmented data quality solutions market. The disruption brings an evolution in how we handle data quality processes and retrieve insights into data. The augmented data quality solutions, driven by AI and GenAI and metadata, brings automation to data quality processes to shorten the time to value, and augmentations to broaden understanding of data. Leading vendors' augmented capabilities provide greater insights and automation. There are two important market phenomena — AI for DQ and DQ for AI.

AI for Data Quality

AI technologies augment the data quality life cycle of "discover, assess, associate, validate, correct and monitor" and are focused on working with individual datasets more efficiently. AI and ML techniques analyze datasets and detect anomalies, or infer rules using functional dependency analysis. Dataset cleansing can also be automated using clustering algorithms to spot outliers and recommend suggested corrections. This approach allows for a more in-depth and rapid execution of data quality actions on individual datasets.

A combination of supervised, semisupervised and unsupervised methods are being deployed to enhance data quality capabilities and automate processes. The offerings from mainstream vendors are primarily supervised learning techniques in data quality use cases where the entities and relationships in the data are well-understood. Unsupervised techniques actively learn from data to detect patterns and identify outliers within datasets. Matching customer entities is an example of leveraging unsupervised matching algorithms that learn from data content and users' responses.

NLP and LLM is effective in understanding, parsing and processing human languages. Within the data quality area, NLP is especially helpful to profile, parse, match, standardize and cleanse data using natural language. Business users can also create new data quality requirements through conversational approaches that allow them to specify their data quality rule in conversational English — for example, "The height of product A must be more than 10 inches."

Data quality vendors are actively carrying out their GenAI initiatives by bringing ChatGPT-like functions to their data quality solution. Most are using commercial OpenAI (such as MS

Azure OpenAI Service, Google Vertex AI) for security. They may build their own proprietary interface to LLMs for translating the input and output.

Data Quality for AI

High-quality data is essential for building accurate and reliable AI systems. Augmented data quality ensures the integrity of data used for AI by providing profiling, monitoring and detection features to ensure the quality of data in the entire data pipeline. The integration with third-party data also provides data enrichment opportunities for improving data quality. There is also a new focus on unstructured data quality that includes the assessment of sensitive or personal content and identifies the uniqueness of content, building a deep connection between them to prepare data pipelines for retrieval augmented generation (RAG) and fine tuning.

Expansion of Data Quality Use Cases

- **AI-ready data:** Augmented data quality solutions help provide technologies to deliver data for AI use cases. It helps assess the data quality, monitor changes in schema and DQ, validate the accuracy of data, and fix errors or prepare dataset for AI use cases (see **Follow These Five Steps to Make Sure Your Data Is AI-Ready**).
- **Data products:** Augmented data quality solutions help ensure the quality of datasets used to create a data product that accurately reflects real-world information, can be reused and reshared, and is compliant with related regulations or policies.

New Critical Feature: Unstructured Data Support

Due to the rise of AI adoption, the technology innovation toward unstructured data is evolving. D&A leaders are interested in using unstructured data for RAG. Data quality is part of the process and requirement.

Augmented data quality solutions provide an ability to analyze unstructured or semistructured data to highlight data quality issues based on semantic analysis and business validation logic, and generate context-specific metadata. AI/ML capabilities are leveraged to validate the accuracy, completeness, and consistency of unstructured data by assessing data quality based on the availability and completeness of metadata through the application of specific validation logic. Vendors are typically leveraging one or some of the following approaches:

- Codevelop with hyperscaler
- Use open-source LLMs
- Customize/fine-tune commercial LLMs
- Build their own LLMs

Market Share and Growth

From a revenue perspective, the data quality solutions market has shown accelerated growth. The data quality solutions market reached $2 billion in 2023 (see **Market Share: Data and Analytics Software, Worldwide, 2023**). The top three vendors by revenue in 2023 (SAP, Experian and Precisely) had 47% market share, an increase of 1% over 2022. The share held by the top three vendors has hovered around 45% for the last three years, signaling a market driven by its top vendors.

The overall market is split into two segments: traditional data quality solutions and augmented data quality solutions. SAP, representing a traditional data quality vendor is rated as top market share vendor, due to its large customer base. The market has started to recognize vendors that have a strong roadmap for augmented capabilities — vendors that did not demonstrate this were seen as lacking vision for this evaluation. Vendors do see more competitive advantage if they invest more in augmented data quality in their product offering, and those that fail to pivot will be irrelevant after the market consolidates.

Market Trends: Convergence

Converge With Tools (Metadata, Data Integration and D&A Governance Tools)

Data quality is increasingly converging with other data management and data governance segments. An overwhelming majority of augmented data quality solutions that Gartner tracks have products in other segments, especially metadata, data integration and D&A governance tools. (See , and .) This convergence indicates the nature of data quality demand across the entire data pipeline and life cycle from various aspects.

Convergence avec les outils d'observabilité des données

The rise of the concept of data observability also adds a new technology stream to augmented data quality solutions. Data observability is a technical capability to understand the health of an organizations' overall data and data landscape (see). This will be an

extension of augmented data management, combining features from ADQ, active metadata and DataOps.

In their approach to finding data issues, data observability solutions target automated anomaly and outlier detection with the ability to reuse these algorithms for rule development. Many vendors in the DQ market have developed more comprehensive observability features in their tools. The stand-alone data observability tools do not provide data remediation features. Therefore, the partnership between the two forms a strong ecosystem to address data quality issues end to end.

⊕ Evidence

⊕ Evaluation Criteria Definitions