Gartner.

Magic Quadrant for Data Integration Tools

Updated 9 November 2021, Published 25 August 2021 - ID G00734633 - 107 min read

UPDATED This Magic Quadrant now includes links to relevant Corporate Transaction Notification research.

By Ehtisham Zaidi, Sharat Menon, and 3 more

The data integration tool market is seeing renewed momentum, driven by requirements for hybrid and multicloud data integration, augmented data management, and data fabric designs. This assessment of 18 vendors will help data and analytics leaders choose a best fit for their data integration needs.

Strategic Planning Assumptions

- Through 2022, manual data management tasks will be reduced by 45% through the addition of machine learning and automated service-level management.
- By 2023, AI-enabled automation in data management and integration will reduce the need for IT specialists by 20%.

Market Definition/Description

Gartner defines data integration as the discipline comprising the architectural patterns, tools and methodologies that allow organizations to access, harmonize, transform, process and move data spanning various endpoints and across any infrastructure.

The market for data integration tools includes vendors that offer a *stand-alone* software product (or products) to enable the construction and implementation of data access and data delivery infrastructure for a variety of data integration scenarios.

These include (but are not limited to):

 Data Engineering: Building, managing and operationalizing data pipelines in support of various analytics and data science demands (for example, logical data warehouse, analytics/business intelligence [BI] and machine learning, among other analytical use cases) by following defined

architectural patterns, tools and methodologies. This use case also requires the data integration tool vendors to deliver capabilities within their data integration tools that enable:

- Support for optimized delivery of analytics: This is the ability to provide access to various heterogeneous data and application sources and then connect or collect data from these sources into a target store that is optimized for delivering integrated data for analytics use cases. This includes the ability to support data warehouse deployments and to manage pushdown optimizations into these data warehouses to support the transformations needed to create data models that are most suited for an organization's evolving analytics needs.
- Data lake enablement: Data integration tools support the ingestion of data in its "native" format to data stores that support the data lake requirements of an organization (including cloud object stores or file stores). However, mere movement of data to the data lake is not enough. Organizations require their data integration tools to support the transformation and operationalization of the data (which includes data preparation, schema assignment, managing and supporting mappings, and delivering the data to supported and consuming applications).
- Self-service data preparation: The ability to provide user experiences that enable end users to develop and manage integrations independent of the tool vendors' professional services. These experiences must support multiple integration personas, particularly integration specialists and ad hoc/citizen integrators.
- Cloud Migration: This use case requires data integration tools to support the migration and modernization of data and analytics workloads to public cloud infrastructure – usually involving an architecture that spans on-premises and one or more cloud ecosystems. This use case also requires the vendors to deliver capabilities within their data integration tools that enable:
 - Data migration and consolidation: Data integration tools increasingly address the data movement and transformation needs of data migration and consolidation, such as the replacement of legacy applications or the migration to new computing environments.
 - Cloud data delivery options: The ability to deliver data integration capabilities as cloud services for hybrid, multicloud, and intercloud integration scenarios.
- Operational Data Integration: Supporting operational/transactional data integration use cases such as master data management and interenterprise data acquisition and sharing (including the ability to create data hubs for integration when needed). This includes the ability of data integration tools to integrate, consolidate and synchronize data related to critical business processes and to support data governance initiatives. Common requirements of data integration tools include:

- Sourcing and delivery of master data in support of master data management (MDM): This involves enabling the connectivity and integration of data representing critical business entities such as customers, products and employees. Data integration tools can be used to build the data access and synchronization processes to support MDM initiatives.
- Interenterprise data sharing: Organizations are increasingly required to provide data to, and receive data from, external trading partners (customers, suppliers, business partners and others).
- Data consistency between operational applications: Data integration tools provide the ability to ensure database-level consistency across applications, both on an internal and an interenterprise basis, and in a bidirectional or unidirectional manner.
- Data Fabric: A data fabric architecture enables faster access to trusted data across distributed landscapes by utilizing active metadata, knowledge graphs, semantics and ML capabilities of data integration (as well as other data management tools, including data catalogs and data governance). Data integration tools must enable the creation and delivery of data fabric design patterns that enable multiple producers and consumers of data to be brought together through better integration, collaboration and automation of data pipelines. The data fabric use case requires data integration tools to be able to deliver data in various styles (not just batch, but a combination of batch with data virtualization, streaming, messaging or API-based delivery styles). Importantly, organizations need their data integration tools to be able to both deliver integration data as data services and orchestrate these services (see below):
 - Data services orchestration: This is the ability to deploy all aspects of runtime data integration functionality as data services (for example, deployed functionality can be called via a web service interface).

Customers must be able to implement and support these use cases with independent offerings from data integration vendors; the use of third-party tools or data integration capabilities embedded in other solutions should not be required. Vendors that sell data integration technology as part of other solutions (such as analytics platforms, DBMSs, and packaged or SaaS applications) are not considered data integration tool vendors by Gartner.

Our evaluation of data integration tools does not include open-source frameworks, general-purpose development platforms or programming interfaces. Such data integration frameworks or platforms, which are "general purpose," and those that require heavy customization by developers to engineer them for specific data integration scenarios are excluded from this Magic Quadrant. Vendors evaluated in this Magic Quadrant offer at least one commercial off-the-shelf tool that is purpose-built for data integration and transformation.

Data integration tools are required to execute many of the core functions of data integration, which can be applied to any of the above scenarios. (For a detailed list and explanation of all core capabilities and use cases of tools in the data integration market, see Critical Capabilities for Data Integration Tools).

Magic Quadrant

Figure 1: Magic Quadrant for Data Integration Tools



Source: Gartner (August 2021)

Vendor Strengths and Cautions

Adeptia

Adeptia is a Niche Player in this Magic Quadrant; in the previous iteration of this research, it was also a Niche Player. Based in Chicago, Illinois, Adeptia offers Adeptia Connect as its data integration product. Its operations are primarily focused in North America and EMEA. It has a customer base of https://www.gartner.com/doc/reprints?id=1-27E7HBB5&ct=210909&st=sb 4

➡

12/11/21, 9:56 AM

Gartner Reprint

over 1,600 in diverse industries, the majority of which sit in the insurance, financial and manufacturing sectors.

Strengths

- Unified platform: Adeptia supports the core requirements of connectivity and adapters, bulk/batch data delivery, and granular data capture and propagation. The combination of data integration capability, application integration, B2B integration and trading partner management within a single product is reported as a value point. It reduces the complexity for buyers and streamlines activities for integrating flows among systems and for delivering data between systems and users.
- High productivity: The ease of use of Adeptia's data integration tooling appeals to IT teams and business roles. Efficiency at creating data pipelines is reinforced by enhancements to large-filebased data ingestion, monitoring data flow of remote access and the use of machine learning to aid data mapping, generate data schema and introspect metadata from data sources.
- Pricing and value: Customers view Adeptia's tools as attractively priced, delivering good value relative to the cost of tools. Simplified packaging for supporting its focus of self-service business data integration aims at easing expanded use for interoperating the data integration functionality in conjunction with other capabilities for prepackaged integration processes for business applications on the same platform.

Cautions

- Limited adoption coverage: While implementations of Adeptia's data integration tooling resonate well with deployments of a targeted scope – driven predominantly by interenterprise data sharing – Adeptia's highly focused approach poses a challenge in addressing the breadth of applicability in this market in some competitive situations.
- Low availability of skills: While ease of use and reduced dependence on specialist resources are emphasized in Adeptia's market focus, finding skilled implementers is an increasing need and challenge for organizations trying to implement and expand deployments as their requirements grow.
- Limited mind share: While awareness of Adeptia is growing, there remains a lack of enterprise inclination to use this vendor's data integration tools in complex implementations, which creates a barrier to enterprisewide adoption. Users of Gartner's client inquiry service rarely mention this vendor in competitive situations.

CloverDX

CloverDX is a new entrant to this Magic Quadrant and is positioned as a Niche Player. It is headquartered in Prague, Czech Republic. It offers the following data integration tool: CloverDX Data Management Platform. It has over 400 customers for this product line. Its operations are primarily

focused in North America and EMEA, with a growing presence in Asia/Pacific (APAC). Its clients are primarily in the financial services, consulting and software services, and public sectors.

Strengths

- Product focus on solving intricate data problems: CloverDX is an integrated data orchestration solution to manage complex data management processes (including error handling) within defined frameworks. For example, it can parse similar data files from multiple sources that do not conform to the same format, organize data into a data warehouse and handle complex data transformations based on templatized configurations by business users.
- Customer focus reflected on services, pricing, packaging and deployment: CloverDX simplifies onboarding customer experience through training and workshops. Its professional services and support services serve the ongoing customer needs with guaranteed SLAs. It follows a simplified pricing model by selling one complete product package that can be deployed on a customer's infrastructure (self-hosted), either on-premises or in the cloud.
- Data discovery, classification and anonymization support: CloverDX can automatically discover data domains and classify using machine learning. Its business users can change the domain in case it was misclassified. Based on user configurations, it can apply anonymization rules such as deidentifying sensitive data.

Cautions

- Limited support for diverse data integration styles: CloverDX does not support data virtualization or change data capture replication. Its streaming data integration capabilities are limited, which is achieved by listening to message streams and processed via microbatching.
- Limited metadata capabilities: CloverDX does not support data lineage and impact analysis as an out-of-the-box functionality. It does not have any active metadata capabilities like integration recommendations.
- Limited availability of quality third-party resources (integrators and service providers): CloverDX has several OEM partnerships like IBM InfoSphere MDM components and GoodData CloudConnect, and a few system integrators like EY and Accenture. However, there are not widely available CloverDX developers in the market, which forces its customers and prospects to develop the necessary competencies in-house.

Denodo

Denodo is a Leader in this Magic Quadrant; in the previous iteration of this research, it was also a Leader. It is based in Palo Alto, California. It offers the following data integration tool: Denodo Platform. Its operations are geographically diverse, with a customer base of over 900 primarily in the financial services, manufacturing and technology sectors.

Strengths

- Product strategy focused on logical and distributed architectures: Denodo enables its customers to connect distributed data through business-friendly semantic models that decouple data from its location and physical schemas. It reflects a longtime focus and strength in data virtualization that enables agile data integration and delivery. Its data preparation capabilities simplify data access for citizen roles. Its notebook capabilities enable specialist roles like data scientists to create and share their narratives by combining queries, scripts and graphics.
- R&D efforts show product improvement: Denodo says it invests 25% of its total revenue in research and development (R&D). In comparison to its previous versions, Denodo Platform 8.0 addresses limitations when handling high data volumes and infrastructure operations when managing resources like load balancing and scaling. These product improvements are evident from various practitioner reviews captured in the Gartner Peer Insights portal and from our analyst interactions.
- Sales execution with customer focus: Denodo's customers appreciate its presales support and proof-of-concept activities. As an example, it offers Denodo Express as a free product (but capped at a certain capacity) for a single user with standard features, an online community-based support, tutorials and videos. About 80% of its paying customers have tried Denodo Express at some stage in the sales cycle.

Cautions

- Limited support for diverse data integration styles: Denodo does not support change data capture within its current data integration offering. It is least suited for traditional batch-related extract, transform and load (ETL) operations given its focus on federation and distributed query processing.
- Global federated deployments require manual configuration: Denodo supports the intercloud data integration use case. However, connectivity across Denodo instances running on different geographies involve a lot of manual configuration due to data contextualization (which is expected to a certain extent) and ongoing operations support to ensure effectiveness of the collective global federated deployment.
- Data security configuration can be challenging: Denodo practitioners report challenges around the process of secure authentication configuration, SSL connections setup in cloud environments, and lack of sufficient data security implementation guidance – all leading to prolonged platform readiness for production use. Denodo has recently improved its security best practices and documentation.

Fivetran

Since the initial publication of this Magic Quadrant (25 August 2021), Fivetran was a party in the following significant corporate transaction(s). For Key Background and Considerations for Technology and Service Selection, see:

"Closed Corporate Transaction Notification: Fivetran, Data Integration Tools" (09 November 2021)

Analysis within this Magic Quadrant remains as originally published.

Fivetran is a Niche Player in this Magic Quadrant; in the previous iteration of this research, it was also a Niche Player. It is headquartered in Oakland, California. It offers the following data integration tool: Fivetran. It has over 2,400 customers for this product line. Its operations are primarily focused in North America and EMEA, with a growing presence in APAC and LATAM. The top three industries it supports are in the software services, information technology and finance sectors.

Strengths

- Product strategy focused on cloud data engineering: Fivetran is a fully managed ELT offering that integrates SaaS and on-premises data stores with cloud DBMS targets. Its ease of use, ease of support and managed response make cloud data ingestion extremely easy, with 99.9% uptime and data delivery SLAs. Business teams can ingest data with minimal IT support. Its product strategy is complemented by its marketing partnership with all cloud service providers (CSPs), major independent software vendors (ISVs) and over 200 system integrators (SIs).
- Low total cost of ownership for low data latency: Fivetran eliminates data pipeline development time and minimizes operations. Its end-to-end data automation ensures source changes are automatically updated in the target by managing schema drifts and guaranteeing reliable data replication. Its customers spend less on data operations tasks. It applies a consumption-based pricing model on the number of active rows processed.
- Fivetran managed data with prebuilt schemas: Fivetran has integrated the open-source data build tool (dbt) as its data transformation layer. dbt enables complex transformations and DataOps capabilities. Fivetran unifies orchestration of data integration, data models, tests and alerts with over 50 prebuilt modeling packages, such as digital advertising and financial packages. These packages increase data engineers' productivity by translating normalized data into analysis-ready data in minutes.

- Limited support for diverse data integration styles: Fivetran does not support data virtualization
 or message-oriented encapsulation and movement of data with its current data integration
 offering.
- Limited metadata management and data governance support: Fivetran lacks capabilities such as discover and capture data semantics, synchronization of data models between DBMS and

modeling tools, the ability to share metadata with other tools, the ability to support enforcement of policy and compliance rules on integrated data, and data quality support.

Gaps in product coverage: Fivetran is invested in various partnerships, including Trifacta for data preparation support, Alation and Collibra for data cataloging and governance support, and dbt Labs' design build tool (dbt) for complex data transformation and orchestration tasks. This approach might work while adapting a best-of-breed approach. However, this is a caution for its prospects who prefer these capabilities to be supported natively within Fivetran.

HVR

HVR is a Niche Player in this Magic Quadrant; in the previous iteration of this research, it was also a Niche Player. Based in San Francisco, HVR offers its HVR data integration product primarily to support log-based change data capture (CDC). The vendor's customer base for this product set is over 450 organizations. Its operations are mostly in North America and EMEA, and its clients are primarily in the manufacturing, technology and finance sectors.

Strengths

- Strong focus on data replication: HVR has a singular focus on the CDC market. This enables HVR to introduce strong product capabilities in this space. These include the ability to read both online and archive logs, the capability to match source and target tables in HVR hub for data validation, the inclusion of both agent-based and agentless options for hybrid and multicloud deployments, embedded functionality to support data compression for high-volume replication, and support for multiple target models such as tables, JSON, Parquet, Avro, CSV, XML and more.
- Enhanced support for complex data transformations: HVR goes beyond simple replication of the data from source to target. It supports row-level transformations, stubbing missing values, fixing outliers and character set conversions. An SDK can be used to call external data transformation logic, if required, with orchestration done by HVR. HVR also provides a differentiated set of capabilities to extract and transform data from SAP systems, such as decoding of data from cluster and pool tables.
- Automation of data replication and synchronization tasks: HVR supports automapping of columns between source and target tables even when data type variations exist autorecovery of failed jobs, autoschema drift correction using autoexecution of DDL commands, automatic creation of "absent tables" or "soft delete" flags in the target database if elements of the source schema are suddenly removed, and more.

Cautions

Limited deployment options: HVR has very limited availability as a managed service right now.
 Customers currently manage HVR as infrastructure as a service (laaS) — in their own virtual machine (VM) on public clouds. HVR launched its Agent as a Service for Microsoft Azure in April

2021, and it has a Real-Time Data as a Service (RTDaaS) offering on the roadmap to mitigate this issue.

- Lack of support for complex ETL workloads: HVR lacks capabilities for addressing data delivery styles other than data replication. For complex bulk/batch ETL workloads, HVR has partnerships with Talend, Matillion and Idera. However, despite having the ability to enrich data, limited ability for bidirectional exchange of metadata with other tools means that customers need to stitch together HVR with other data integration tools for enabling the full breadth of data integration styles.
- Data and analytics governance support is limited: HVR does not have out-of-the-box integration
 with tools that provide data governance capabilities, such as data quality, data catalog,
 information stewardship and data privacy. HVR is gradually improving in this area, as data lineage
 for Slow Changing Dimension (SCD) Type 2 data is already provided, and capabilities such as data
 masking and data quality are on the roadmap.

IBM

IBM is a Leader in this Magic Quadrant; in the previous iteration of this research, it was also a Leader. IBM Cloud Pak for Data (which includes DataStage Enterprise Cartridge), IBM Cloud Pak for Integration, IBM Data Replication and IBM Data Virtualization Manager all target a range of data integration use-case scenarios. The vendor's customer base for this product set is more than 11,000 organizations. Its operations are global, and its clients tend to be enterprise B2B and B2C organizations.

- Modular architecture that supports intercloud and hybrid integration patterns: IBM's data integration tools are delivered as tightly integrated and yet loosely coupled services on the Red Hat OpenShift platform (which is a Kubernetes-based orchestrated platform). Clients can design integration pipelines once and then deploy them on any cloud infrastructure or on-premises environment to support multicloud hybrid integration. Customers praise IBM for its modular services and serverless execution capabilities that reduce egress costs by pushing execution closer to the data, irrespective of where it resides.
- Strong support for DataOps: IBM's customers have praised its capabilities for DataOps support. These include native Git integration (for version control and agile project management), automated test/validation using a language similar to the popular Gherkin (to specify test scenarios) and CI/CD toolchain capabilities. Customers also praise its integration portability capabilities that allow the same data pipeline to be executed in different execution environments across various CSP infrastructures.

Capabilities that support the data fabric design: IBM's customers praise its capabilities for data integration design and delivery automation. It embeds active-metadata-based insights from the IBM Knowledge Catalog into its data integration tools that assist developers with automating various repetitive tasks. IBM's support for combining different data delivery styles (batch, streaming and virtual, for example) and its recent graph support for complex data modeling have further improved its overall data fabric support capabilities.

Cautions

- Complicated licensing: IBM customers have cited challenges in understanding the changes to licensing (and how existing licensing metrics will be impacted) while upgrading to Cloud Pak for Data. Some customers report challenges in parsing licensing metrics of existing tools (DataStage, for example) and how these can be traded-in/converted to the new licensing metrics of Cloud Pak for Data. To mitigate some of these challenges (for customers upgrading to Cloud Pak for Data), IBM does offer simultaneous access to the traditional software and forward compatibility for existing data integration pipelines originally built on the traditional software.
- Low visibility and understanding of data preparation capabilities: IBM's data preparation capabilities to support the integration requirements of less technical users come embedded within the Data Refinery component of Cloud Pak for Data. However, many customers and prospects are simply unaware of these data preparation capabilities and frequently look for competing solutions for stand-alone data preparation requirements. IBM should provide a more detailed understanding of how data preparation pipelines created in Data Refinery can be operationalized for production usage in Cloud Pak for Data.
- Challenges with upgrades: Some IBM customers have cited challenges with upgrading to the latest IBM data integration offerings, including Cloud Pak for Data. A few customers state that it was costly and challenging to upgrade to the new versions of the software. IBM must provide a more detailed roadmap and support (including in-place and container-based upgrades) for moving customers to Cloud Pak for Data.

Informatica

Informatica is a Leader in this Magic Quadrant; in the previous iteration of this research, it was also a Leader. Informatica is headquartered in Redwood City, California. It offers the following data integration tools as part of its Intelligent Data Management Cloud: Informatica Intelligent Cloud Services (IICS) (which includes the following services related to data integration: Cloud Data Integration Elastic, Cloud Mass Ingestion, Cloud Integration Hub and Cloud B2B), Data Engineering Integration, Enterprise Data Preparation, Enterprise Data Catalog, Data Engineering Streaming, PowerCenter, and PowerExchange. Informatica has over 10,000 customers for these product lines. Its operations are geographically diversified and its clients are primarily in the financial services, healthcare and public sectors.

Strengths

- Product investments aligned to the data fabric vision: Informatica has continued to invest in tools that are aligned to the data fabric design vision. Informatica CLAIRE, which is an active-metadata-based AI and ML engine, continues to be highlighted by customers for enabling automation in data integration design and delivery. Informatica has also invested in enabling knowledge graphs (through recent acquisitions of Compact Solutions and GreenBay Technologies) as part of its data integration offerings to support complex modeling tasks involving multirelationship data. Finally, customers stitching data fabric designs benefit from the deep integration of Informatica Enterprise Data Catalog with data integration pipelines, which enables active metadata sharing for insights and automation of tasks.
- Strength in data engineering use cases: Informatica's customers praise its scalable data engineering tools for their ability to handle all data movement topologies. Customers also call out the scalability and performance optimization of its tools for complex transformations requiring push-down optimization and native Spark elastic and serverless capabilities for massively parallel processing (MPP) support. Informatica Cloud Mass Ingestion has been praised by data engineers who are looking to support real-time replication from file, database, applications and event streams through one common platform.
- Strong execution for operational data integration use cases: Beyond just analytics and data science use cases, customers looking to implement a data hub for integration, governance and sharing (among other operational use cases) call out Informatica's solutions as relatively mature. The Cloud Integration Hub offering is frequently selected by customers in competitive situations for its ability to support all data modalities (including batch, virtual, streaming and API-based integration) and for its ability to integrate and deliver data independently to a multicloud hybrid environment.

- Challenges with manual migration from PowerCenter to IICS: Although Informatica continues to support PowerCenter on-premises deployments, a growing number of customers are looking to migrate existing workloads from PowerCenter to IICS. Customers looking to manually repoint PowerCenter mappings to IICS mappings have faced challenges ranging from cost to performance and even productivity of data teams. Reference customers have stated that they needed significant upfront planning and Informatica (or partner) professional services support to successfully make this migration. To accelerate these migration efforts, Informatica has launched a modernization program that includes automated migration utilities.
- Less visibility and understanding of Informatica's new pricing model: Informatica has launched a new consumption-based pricing model based on Informatica Processing Units (IPUs), which is a unit of software licensing capacity measured by service usage. While this model does enable better scaling in cloud environments, existing customers cite challenges in upgrading to this

model. Customers find it difficult to forecast usage upfront to understand how many IPUs should be licensed, and they need a better understanding of how existing licensing can be synchronized and mapped to the new model. In order to mitigate some of those challenges, Informatica has launched a new tool that can assist customers with forecasting their IPU needs and providing a degree of overdraft protection on unexpected consumption.

Some DataOps-related challenges: A few Informatica customers requested better DataOps capabilities, including improved scheduling capabilities for pipelines. Informatica has invested in CI/CD support, better regression testing and improved integration with Git, but existing customers are mostly unaware or report challenges with such capabilities.

Matillion

Matillion is a Niche Player in this Magic Quadrant, the same as last year. It has dual headquarters in Manchester, U.K., and Denver, Colorado. It offers the following data integration tools: Matillion ETL and Matillion Data Loader. It has over 1,050 customers for these product lines. Its operations are geographically diverse in North America, EMEA and APAC. Its clients are primarily in the business services, software and manufacturing sectors.

Strengths

- Product strategy focused on cloud data engineering: Matillion is primarily an ELT (extract, transform and load) engine that exploits the separation of storage and compute within cloud DBMSs by pushing down complex transformations as SQL capability. It offers Matillion ETL to data engineers for integrating data in the cloud DBMS with full developer capabilities such as version control, compute clustering and code sharing via Matillion Exchange. It offers Matillion Data Loader to analysts and citizen integrators for easy ingestion of data into cloud data stores.
- Deployment options solve common data sovereignty issues: Matillion ETL is deployed as machine images within customer's preferred cloud region. It complies with data sovereignty regulations (including Gov Cloud) because customers' data never leaves the sovereign region. This is a unique advantage over SaaS-based integration vendors who must deploy data centers in every regulated region based on customer demands.
- High level of customer satisfaction: Matillion's customers are pleased with its presales support, product functionality, pricing and customer services. For example, it offers a predictable spend using Matillion Hub, which gives a unified view of usage across all cloud service providers and cloud data warehouses. Customers can drill into consumption of credits (based on the amount of time they use Matillion) weekly/monthly to track the level of usage they are expecting. Customers can also add or remove users as needed.

- Limited support for diverse data integration styles: Matillion does not support stream data integration or message-oriented encapsulation and movement of data. It does not have built-in data virtualization capabilities. However, it simplifies the customer experience of the underlying cloud DBMS for data federation via external table objects.
- Limited data governance support: Matillion lacks capabilities such as the ability to support enforcement of policy and compliance rules on integrated data and overall data quality support. It offers basic lineage, and it partners with Collibra for end-to-end data lineage tracking and impact analysis reporting.
- Inadequate product documentation: While Matillion has improved its documentation in the last year, a few of its customers seem displeased in this area: It is sometimes hard to understand the use of a component by reading the documentation, and some customers have sought help in demonstrating the use of various components.

Microsoft

Microsoft is a Leader in this Magic Quadrant; in the last iteration of this research, it was a Challenger. Its SQL Server Integration Services (SSIS) tool targets on-premises data integration tasks, while Azure Data Factory (ADF) targets Azure-based data integration tasks. The number of paid customers for its data integration tools is not disclosed by Microsoft. Its operations are geographically diverse, and its clients range from small and midsize businesses to large enterprises.

Strengths

- Relevant capabilities and TCO: Overall low total cost of ownership (TCO), speed of implementation, ease of use, and the ability to integrate with other Microsoft technologies are frequently the main reasons this data integration tooling is chosen over alternatives. The Azure Data Factory functionality of a low-/no-code environment increasingly appeals to the incumbent SSIS customer base as it migrates to the Azure environment.
- Productivity in data delivery-, process- and business-oriented use: Ability to redeploy SSIS
 artifacts to Azure Data Factory extends hybrid deployment capabilities and supports combined
 use of ETL, enterprise business workflows and data preparation. Improved cohesive access of
 metadata and multicatalogs sets out for supporting active data profiling requirements of ad hoc
 and citizen integrators.
- Widespread tool presence and usage experience: Broad familiarity with the implementation of Microsoft technologies spurs increasing usage of Azure Data Factory. Wide choices in terms of community collaboration, training, and third-party documentation and guidance for deployment practices are key points of value.

- Budget prediction: While the pricing model of Azure's data integration tooling is transparent and flexible (supported by the Azure price calculator and cost management tooling), some customers not well-acquainted with the Azure ecosystem find cost planning difficult in determining scenarios and budget projections.
- Perception of Microsoft-centric capabilities: Although Microsoft has a large customer base and a strong focus for developers to leverage its data integration tools to connect to non-Microsoft data sources, the migration to the Azure environment has generated market perceptions that the vendor's data integration tooling has limited appeal to non-Microsoft-centric environments.
- Deployment and usage experience: Although users of Microsoft deploy a mix of cloud and onpremises runtime models, easier implementation is desired by organizations lacking experience in artifacts migration between deployment choices, including multicloud and intercloud. Microsoft's focus on serverless runtime and continuous integration/continuous delivery (CI/CD) efficacy aims at simplifying data integration deployment and usage experience of customers.

Oracle

Oracle is a Leader in this Magic Quadrant, the same as last year. Based in Redwood Shores, California, Oracle offers the Oracle GoldenGate (OGG) platform, Oracle Data Integrator, Oracle Big Data SQL (BDSQL), and data integration services within Oracle Integration Cloud (i.e., iPaaS) and Oracle Cloud Infrastructure (OCI) Data Integration. Oracle's customer base for these products is over 15,000 organizations. Its operations are geographically diversified, and its clients are primarily in the financial services, manufacturing and public sectors.

- Aligned to the data fabric use case: Oracle has a strong product portfolio for supporting a data fabric design. OCI Data Catalog can extract metadata from all Oracle services and use that to prescribe improvement in the way these services are run and orchestrated. Oracle can automate many aspects of data integration, such as self-healing of jobs, automatic schema drift correction, and autoassigning of clusters to run serverless OCI Data Integration or OCI GoldenGate jobs.
- Balanced support for all data delivery styles: Oracle provides Oracle Data Integrator for batch workloads, Oracle GoldenGate for replication, Oracle GoldenGate Stream Analytics for streaming workloads, and Big Data SQL for data virtualization. These products are available as cloud services as well. Oracle can therefore support customers looking for an "all in one" data integration platform.
- Enterprise-grade products for mission-critical workloads: Through GoldenGate, Oracle differentiates itself most when organizations need to run critical applications with extremely strict SLAs for uptime, such as payment systems and ticketing systems. Large enterprises continue

relying on it for replication across complex data ecosystems. All Oracle data integration services provide excellent uptime guarantee, with credit refunds when failing to meet your SLA.

Cautions

- Perception of data integration being OCI-centric: Oracle needs to better educate the market that stand-alone tools (ODI, OGG and BDSQL) continue to have general availability (GA), and that it has its own long-term product roadmaps outside of the OCI ecosystem. Oracle has the unique opportunity to compete as both a cloud service provider through OCI and an independent data integration provider through its stand-alone products. But for this, its multicloud strategy around integration with rival CSP services needs more clarity.
- Perception of higher price compared with competitors: While Oracle has introduced flexible, metered solutions, market perception of pricing is still an issue. For instance, Oracle is frequently at risk of losing out on Oracle GoldenGate deals against other data replication vendors, especially when evaluated by small and midsize businesses (SMBs), which have limited budgets.
- Limited traction for stand-alone data virtualization: While Oracle does provide Big Data SQL for data virtualization, Gartner rarely sees this product being evaluated when the customer requires only a point solution to support this data delivery style.

Precisely

Precisely is a Challenger in this Magic Quadrant, the same as last year. Based in Burlington, Massachusetts, Precisely offers Connect, Ironstream and Spectrum. The vendor's customer base for this product set is over 2,300 organizations. Its operations are mostly in North America and EMEA, and its clients are primarily in the financial services, insurance and healthcare sectors.

- Data migration into the cloud is a strong use case: Precisely continues to have technical expertise in reading data from IBM iSeries, IBM zSeries and other mainframe systems through Connect and in moving that data into modern cloud data stores such as Snowflake as part of cloud migration projects. Precisely also excels at migrating jobs built on on-premises data lakes into solutions for managed Hadoop on the public cloud, such as those provided by Databricks and Cloudera.
- Strong potential for data management platform play: Precisely has a strong focus on the data integrity practice through its products for data integration, data quality and enrichment, data governance, and location intelligence. The acquisition of Infogix now adds potential data governance capabilities as well. Precisely has many of the building blocks in place for bringing a complete data management platform to the market.
- Praised for ease of use: The peer user community has praised Precisely products for their graphical user interface that enables easy design and configuration of data flows and replication

jobs. The ease of monitoring these data flows through data lineage and the ability to create custom dashboards for metadata observability were also commended.

Cautions

- Lack of clarity behind acquisitions: Precisely has acquired many data management vendors in the past few years. To support customers looking for a holistic data management strategy from Precisely, it needs to work on aspects such as mitigating capability overlaps across the portfolio, unifying the user experience across all products and streamlining pricing models.
- Does not support augmented data integration: Precisely does not currently provide in-tool recommendations to the users to select or optimize data transformations, execution engines and so on. This is an essential capability to support the data fabric use case and is one of the few true differentiators in the largely mature data integration tool market.
- Not suitable for hybrid deployment: Precisely lacks deployment options to support its customers' hybrid and multicloud ecosystems. There are no managed services for cloud deployment for the Connect and Ironstream products as of now. Precisely is working on this: Kubernetes-based deployment and a beta version for Precisely Connect SaaS are both 2021 product roadmap items.

Qlik

Qlik is a Challenger in this Magic Quadrant; in the last iteration of this research, it was also a Challenger. Based in King of Prussia, Pennsylvania, Qlik targets a range of data replication and metadata management tasks through its Qlik Replicate, Qlik Compose, Qlik Enterprise Manager and Qlik Catalog products. Capabilities from its recent acquisition of Blendr.io adds support for integrating cloud data and API sources. Its customer base for this product set is more than 3,000 organizations globally. Qlik's operations are predominantly based in North America and EMEA, and its clients tend to be enterprises in the healthcare, financial services, retail and manufacturing sectors.

- Traction of core functionality: Qlik offers strong data integration functionality in the key areas of data replication and synchronization technology applied to heterogeneous data types, with a historical strength in addressing complex use cases like mainframe and SAP data. With revenue growth that is again well above market average in 2020, Qlik progresses its track record of supporting data consistency and expands support across data ingestion/integration to data cataloging and event and batch analytics.
- Productivity for multipersonas: Qlik's tooling supports data integration specialists as well as, increasingly, less technically skilled personnel in order to make data available through real-time data streaming, data warehouse automation and data lake creation. The role-based interface allows citizen integrators to integrate data quickly for their purposes. Design and administrative

tools are used by enterprise architects and designers when more robust solutions are required. The result is a mix of time to value of delivery that matches many different business needs.

Ease of operating with incumbent infrastructure: Users of Qlik's data integration tooling often like its ease of use – including ease of installation and configuration, automated code generation, and nonintrusive implementation – as main reasons for selecting the vendor in competitive situations.

Cautions

- Breadth of market coverage: While Qlik's traction for data replication, data warehouse automation and data lake population is well-established, prospective customers are predominantly drawn to its data replication/synchronization. This presents disadvantages in competitive bids when customers look to combine and interoperate between diverse data delivery styles beyond CDC/replication.
- Guidance for implementation practices: Some implementations cited a need for more-extensive guidance for architecture and delivery from the vendor, and for tightening links to governance support of data stewards, in relation to the breadth of data management activities.
- Skills availability: Finding skilled resources was reported as a challenge for some customers and prospects when trying to deploy, interoperate and expand the use of Qlik's portfolio of tooling in complex integration environments.

Safe Software

Safe Software is a Niche Player in this Magic Quadrant, the same as last year. It is headquartered in Surrey, British Columbia, and offers a data integration platform called Feature Manipulation Engine (FME), which includes FME Desktop and FME Server. The customer base for this product set is more than 6,800 organizations (excluding OEMs), mainly in EMEA and North America. Its clients are primarily in the government; utilities; energy (oil and gas); and architecture, engineering and construction (AEC) sectors.

- Spatial data integration: Safe Software's strength is in geospatial data ingestion, integration and sharing. The vendor offers an extensive collection of connectors for spatial data sources (such as Esri's GIS) and data management platforms. Customers can use these connectors to connect spatial data to business systems, such as CRM systems, and can take advantage of prebuilt transformations that provide relevance for location intelligence and interenterprise data sharing.
- No-code/low-code data preparation capabilities: Safe Software offers a low-code UI that leverages 500-plus prebuilt transformers to enable users to share and model data without requiring substantial technical knowledge. Using FME Server Apps, customers can build interactive

web apps via a no-code interface and can also create branded landing pages to present a collection of multiple apps.

 Execution, support and customer relationships: This vendor offers a variety of deployment options, including on-premises, cloud-hosted, via cloud marketplaces and containerization. Customers praise Safe Software for ease of installation and setup, deployment, ongoing support, and quality of training and support documentation. Customers also report that the company has a particularly supportive user community.

Cautions

- Limited traction beyond spatial data: Safe Software does support all data integration use cases but is primarily used for spatial data and data intelligence use cases, and its strengths play heavily to this focus. Based on data from Gartner's client inquiry and proposal review services, clients rarely evaluate Safe Software for situations that do not involve spatial data.
- Limited product maturity for metadata management: Safe Software is working toward providing a more comprehensive ability to discover and capture metadata to inform and automate integration, design and delivery, but some of these active metadata features are new, and the product offerings may not be fully comprehensive. Capabilities such as active-metadata-based recommendations for developers, metadata exchange with third-party information governance tools, and data lineage and impact analysis reports are not yet fully supported.
- Limited DataOps support: FME offers many change management capabilities, such as Git integration for version control, automated deployment and workflow updates via FME Server Automations, as well as the ability for users to update workflows via an API, but it lacks full CI/CD support. Workflow management for the development process, such as authorizations and approvals, is not offered.

SAP

SAP is a Leader in this Magic Quadrant, the same as last year. It is based in Walldorf, Germany. It offers the following data integration tools: SAP Data Intelligence, SAP Data Services, SAP Landscape Transformation (SLT) Replication Server and SAP Integration Suite, along with integration capabilities within the SAP HANA platform. Its operations are geographically diverse, with a customer base of over 74,000, and the top three industries it supports are consumer products, oil and gas, and retail sectors.

Strengths

Product strategy focused on unification, hybrid and scale: SAP leverages its flagship on-premises offerings (SAP Data Services and SLT Replication Server) into a hybrid data management strategy (with SAP Data Intelligence Cloud) to meet the needs of all customers. It provides an end-to-end integration strategy for its customers by enabling them to combine and switch between multiple

integration styles across on-premises and cloud. SAP Integration Suite offers a unified integration platform for end-to-end process and data integration, as well as API management, among other capabilities.

- Caters to the data processing needs of enterprisewide users: SAP Data Intelligence is one solution for data cataloging and data integration. It orchestrates heterogeneous data processing engines, including SAP HANA, and several open-source engines across distributed infrastructures, cloud and on-premises. Citizen users can visually explore and discover data in a no-code setup, data engineers can build full-blown data integration pipelines with support for complex transformations and data quality, and data scientists can use their notebook applications for combining data exploration and integration.
- Industry-specific business content packages: SAP, along with its partners, offers various industryspecific domain expertise and integration solutions with predefined data flows, metadata and ML models for a wide variety of use cases. For example, in partnering with Wipro, it offers "Cognitive Customer Service" to improve customer service effectiveness by predicting the reason for a customer call.

Cautions

- New customers struggle with initial setup: Several new SAP data integration customers express concerns around initial setup complexities and delays. There is a shortage of skilled expertise in the market for its newer products like SAP Data Intelligence. Many customers recommend using SAP service partners to address this challenge.
- Limited integration with non-SAP products: Some practitioners face compatibility challenges between SAP's data integration product set and non-SAP products. For example, data scientists face several integration challenges (like ease of data/metadata sharing) between SAP Data Intelligence and their machine learning tool of choice. However, its 2021 roadmap capabilities to support several open standards and languages (like Kafka, MQTT, Python and R) are expected to address some of these limitations.
- High pricing: Several SAP data integration customers report high licensing costs and a longer time to achieve ROI objectives. On the balance, it's cloud pricing strategy provides specific customer solutions, such as bundling multiple products into single licenses and sharing resources, which might reduce the overall cost.

SAS

SAS is a Challenger in this Magic Quadrant; in the last iteration of this research, it was a Leader. Based in Cary, North Carolina, SAS offers the following data integration products: SAS Data Management, SAS Data Integration Studio, SAS Federation Server, SAS/ACCESS in SAS Viya, SAS Data Loader for Hadoop, SAS Event Stream Processing on SAS Viya and SAS Data Preparation on

SAS Viya. The vendor's customer base for these products is around 16,000 organizations. Its operations are geographically diversified, and its clients are primarily in the banking, government and services sectors.

Strengths

- Enhanced support for diverse user persona: SAS provides SAS Data Integration for ETL developers and data engineers, SAS Data Preparation for business users, SAS Studio Analyst for business analysts, SAS Information Governance for data stewards, and more. These tools place SAS in a good position to leverage the recent trends toward data democratization.
- Flexibility in the execution engine: SAS enables execution of complex transformations and advanced data quality functions within Spark on Databricks and Hadoop, or within databases such as Teradata. SAS Data Integration Server leverages a "design once, deploy repeatedly" architecture, enabling users to deploy the same job to be scheduled to run in batch, in real time or as a service.
- Positive customer experience for professional services: Customers commend SAS for its services being very professional and solution-oriented. SAS scored over 4 out of 5 by the peer user community for "services and support."

Cautions

- Limited vision as an independent, stand-alone data integration tools provider. As part of SAS's plans to converge capabilities across data management and analytics, SAS wants SAS Data Management to be a part of SAS Viya 4 in the future, with SAS/ACCESS, SAS Data Preparation, SAS Information Governance and SAS Studio Analyst already migrated. SAS needs to proactively educate the market that data integration capabilities will continue to be available as stand-alone products. Gartner rarely sees examples of SAS competing for data integration deals unless other SAS systems are already a major part of the client's data and analytics ecosystem.
- Lacks vision for supporting data fabric designs: SAS does not provide graph-based semantic modeling capabilities. SAS does not yet have a clear vision of how its different data management products (data catalog, data integration and so on) can be used to support its customers' data fabric designs. This is especially critical because some of SAS's closest competitors have advanced their vision to support data fabric designs in the past year.
- Configuration and installation challenges: Multiple customers in the peer user community have stated that the installation and configuration of SAS products can be challenging and timeconsuming and can sometimes require a SAS expert. SAS Viya 4 is cloud-native and runs on containers, which is expected to mitigate these challenges going forward.

SnapLogic

SnapLogic is a Visionary in this Magic Quadrant; in the previous iteration of this research, it was also a Visionary. Headquartered in San Mateo, California, SnapLogic offers the SnapLogic Intelligent Integration Platform as its data integration offering. It operates mostly in North America and EMEA, with a customer base estimated at 1,100 organizations from diverse sectors including technology, retail, manufacturing, healthcare, financial services and transportation.

Strengths

- Productivity driven by augmented guidance and automation: SnapLogic's ability to empower less technical data integration roles reduces reliance solely on highly skilled specialist integrators. Embedded ML for automated guidance helps accelerate the development of integration flows, which can reduce the reliance on detailed documentation. Its focus on ease of use and on augmented data integration provides automated guidance embedded to simplify the development of integration flows and reduce time to value. These are key points of value for adoption.
- Pricing model simplicity and trial version: Pricing simplicity encourages self-service license options to facilitate adoption and onboard prospects faster. A free trial version and a newly added user interface via its Fast Data Loader are designed to ease exploration of product features, including data loading to cloud data warehouses.
- Sales and marketing strategies: SnapLogic continues to go to market through cloud partners to strengthen channels for data integration deal flows by leveraging co-marketing activities with ISV providers like AWS, Snowflake and SAP.

Cautions

- Market perception of cloud-only relevance: While SnapLogic has been educating in its market reach in the data integration tools market, perception of SnapLogic's credentials relate largely to its cloud-enablement scenarios, which adversely affect some competitive evaluations.
- Limited guidance and support for implementations: Deployments in increasingly complex scenarios are raising customers' expectations of SnapLogic's implementation support and guidance. They want easier ways to familiarize with implementation best practices and proven solutions, including those from third-party providers and peer user communities, while SnapLogic builds out its ISV and SI partner ecosystems.
- Lack of data management versatility: Although SnapLogic supports data transformation for data profiling, the need for data governance support was expressed in some data integration activities as a challenge. Some customers state that they need improved coverage of data management support, including comprehensive data quality and collaborative operationalization in data fabric scenarios.

Talend

Talend is a Leader in this Magic Quadrant; in the previous iteration of this research, it was also a Leader. Talend is headquartered in Redwood City, California. Its data integration tools include the Talend Data Fabric and the Talend Data Catalog. Talend has over 6,000 licensed customers for this product line. Its operations are geographically diversified, and its clients represent companies in a variety of sectors, such as media and services, financial services, and manufacturing.

Strengths

- Governed approach to operationalization of self-service data preparation: Talend provides Stitch Data Loader as a low-code UI for nontechnical users to ingest data into sandbox environments. They can then use Talend Data Inventory and Talend Data Catalog to find, inventory, profile and tag datasets of relevance. Talend then makes it easy for them to prepare the data of choice in Talend Data Preparation by using its embedded ML capabilities to prepare the datasets without any coding. Finally, once business users are satisfied with their data model, they can invite data engineers to review and operationalize their data models to Talend Studio (where the data can be taken to the data warehouse for production). At the same time, information stewards can review the data for conformance to policy and governance rules in the Talend Data Stewardship application.
- Strong capabilities for the data engineering use case: Talend has improved the Talend Pipeline Designer tool to enable better integration code portability. Data engineers can now build data pipelines once and redeploy across multiple infrastructure and execution environments (on-premises, cloud or even hybrid using Talend remote engines) through integration with popular containerization services. Talend has also invested heavily in native integration with Spark (Databricks) and Apache Beam for more optimal pushdowns and complex job execution.
- Improved capabilities to support DataOps: Talend has significantly upgraded its support for DataOps in the Talend Data Fabric. Enhancements include support for CI/CD and native integration with popular version control and project management tools like Git, Jenkins and Maven, as well as support for automated testing and validation and visual impact analysis.

- After-sales support challenges: Several Talend customers call out the need for a more comprehensive online product documentation guide, particularly for supporting the growing Talend data integration tools portfolio. Some Talend customers also cited slow turnaround times for issues and bug fixes and requested faster planned upgrades to Talend Cloud.
- Limited traction for data replication: Talend Data Fabric supports log- and trigger-based change data capture (CDC) functionality natively and even contributes to the Apache Falcon project (to support CDC-like capabilities to Hadoop). But Gartner clients that use Talend exhibit limited understanding of Talend's capabilities in data replication and data synchronization. Due to this

lack of clarity, Gartner regularly fields calls with Talend prospects (and even existing customers) who are evaluating other vendors for their data replication use cases.

Challenges with licensing and packaging clarity: A small but significant number of Talend customers cited challenges with understanding its cloud pricing models, which include a combination of named user licenses and new metrics like the "remote engine" and tokens. Talend customers also need additional clarity on the packaging of different data integration tools within the Talend Data Fabric and how they can start small and scale with different services. Talend has recently announced new packages, which include product trials, entry level tools, platforms and add-ons to alleviate some of those packaging challenges.

TIBCO Software

TIBCO Software is a Challenger in this Magic Quadrant; in the previous iteration of this research, it was also a Challenger. TIBCO is based in Palo Alto, California. Its main data integration tools are TIBCO Data Virtualization (TDV), TIBCO Cloud Integration, TIBCO Cloud Metadata, TIBCO Messaging, TIBCO Streaming and TIBCO Omni-Gen (which includes the iWay suite of products from Information Builders). The vendor's customer base for this product set is more than 6,000 organizations. Its operations are geographically diversified, and its clients include companies in the financial services, telecommunications and manufacturing sectors.

Note: On 5 January 2021, TIBCO announced that it had closed the acquisition of Information Builders. For a more in-depth understanding of this acquisition and its potential implications, see Corporate Transaction Notification: TIBCO Software.

- Modular tools that support a "connect and collect" vision for data integration: TIBCO's data integration tools allow customers to maintain a balance between connecting to data (using TDV) and collecting/moving data (using TIBCO's other integration tools). TIBCO's data integration tools support modular use (that is, they are loosely coupled), allowing customers to start small and scale in any environment based on their use-case demand.
- Leadership in Internet of Things (IoT) and streaming scenarios: TIBCO's customers praise its data integration services for their scalability, performance optimization and maturity for supporting their stream data integration scenarios. TIBCO Streaming has consistently exhibited strength in competitive situations versus other stream/event stream processing tools, based on Gartner client inquiry feedback.
- Improved bulk/batch data integration capability: TIBCO completed its acquisition of Information Builders (ibi) in January 2021, which also gave it access to ibi's Omni-Gen platform (which TIBCO now markets as TIBCO Omni-Gen). Omni-Gen has been historically highly rated for its mature ETL

capabilities for complex bulk/batch data integration scenarios, and some implementations indicated significant improvements in its ETL support through the use of Omni-Gen.

Cautions

- Packaging challenges: TIBCO has grown over the years with several acquisitions, and with its latest acquisition of ibi, the company's data integration portfolio has become quite complex. Gartner receives a significant number of inquiries from TIBCO prospects and customers who are unclear on which tools and services need to be combined or used separately to accomplish their current and upcoming data integration requirements. TIBCO's efforts in packaging its solutions under the "Connect, Unify & Predict" segments have alleviated some of these challenges to a certain extent.
- Limited active-metadata-based insights to support augmented data integration: TIBCO has continued to invest in active metadata capabilities through its investments in the TIBCO Cloud Metadata product. However, data engineering teams using TIBCO request more active-metadata-based decision insights to support the automation of complex integration tasks. Customers request more seamless sharing of metadata insights from TIBCO Cloud Integration in TIBCO's other data integration tools.
- Challenges related to DataOps: A few TIBCO customers request improved DataOps capabilities, including better support for CI/CD (for example, through integration with Git or Jenkins) and efficient debugging/error-handling workflows through automated test and validation functionality improvements. Some of these challenges are mitigated through the inclusion of the Deployment Manager module with TDV.

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

 CloverDX: CloverDX has been included as a first-time entrant in this Magic Quadrant because it met the inclusion criteria on number of customers and has also included multiple data integration styles (including batch, real-time and event-based data integration).

Dropped

- Actian: Actian was not included in this year's Magic Quadrant because it did not meet the updated inclusion criteria, which requires demonstrable market presence and execution as a stand-alone (and independent) data integration tool.
- Information Builders: Information Builders was acquired by TIBCO Software.
- Hitachi Vantara: Hitachi Vantara was not included in this year's Magic Quadrant because it did not meet the updated inclusion criteria. The required information was not made available by the vendor to substantiate its ability to meet the latest inclusion criteria for this research.

Inclusion and Exclusion Criteria

The inclusion criteria represent the specific attributes that Gartner analysts believe are necessary for inclusion in this research.

To qualify for inclusion, the vendor's data integration tool (or tools) must be a "stand-alone" product directly usable by the buyer. To use the product, customers should be able to procure only the data integration tool as an independent offering and not as a part of some other offerings — such as another form of tool suite, an application or other technology solution — of which the data integration capabilities are an "embedded" subset.

The data integration tool (or tools) must demonstrate various data delivery styles (from the list below) and be flexible enough to combine these styles for delivering various customer use cases:

- Bulk/batch data movement: Bulk and/or batch data extraction and delivery approaches (such as support for ETL/ELT/ETLT) are used to consolidate data from distributed databases and formats. This capability draws on data from across systems and organizational boundaries and can play a role in all use cases in this research.
- Data virtualization: Data virtualization executes queries against distributed data sources to create virtual, integrated views of data "in memory." Virtual views require adapters to data sources, a metadata repository and a distributed query engine that can provide results in various ways for downstream consumption.
- Data replication and data synchronization:
 - Data replication implies a simple copy of data and schema from one location to another, always in a physical repository. Replication does not change the form, structure or content of the data it moves.
 - Data synchronization focuses on establishing and maintaining consistency between two separate and independently managed create, read, update, delete (CRUD) instances of a shared, logically consistent data model for an operational data consistency use case. Synchronization also maintains and resolves instances of data collision, with the capability to establish

embedded decision rules for resolving such collisions (sometimes through schema-drifthandling mechanisms and sometimes through other options).

- Stream data integration: This is the ability to address data integration requirements through interoperability with streams/events, including provisioning of data in-stream for enabling downstream consumption, analysis or storage.
- **Data services orchestration** This covers both message-oriented middleware and API services:
 - These capabilities allow data integration tools to encapsulate data in messages that various applications can read. Data is exchanged in real time, often via message queues like Kafka or by using message-oriented middleware such as Java Message Service (JMS), IBM MQ and RabbitMQ.
 - This capability allows data as a service, enabled through API design and delivery capabilities, to create and manage outbound API endpoints over existing data assets, and to handle inbound API consumption to ingest internal and external data.

Beyond the data delivery styles called out in the list above, the data integration tools must exhibit the following capabilities to be included in this Magic Quadrant research:

- Data movement topology: This involves uni-/bi-/multidirectional movement of data across endpoints (such as synchronize, compare, broadcast and consolidate), via physical and virtual modes, while meeting batch/microbatch/real-time latency requirements.
- Range of connectors/adapters (sources and targets): This is native access to relational and nonrelational DBMS products, plus access to nonrelational legacy data structures, flat files, XML and message queues, cloud-based data asset types (including data of SaaS applications and cloud data stores), and streaming data.
- Complex data transformation: These include capabilities that ease complex data processing operations, such as fixing outliers, stubbing missing values, sophisticated parsing, automating data warehouses and creating reusable transformations.
- Metadata and data modelling support: This includes automated metadata discovery (such as profiling new data sources for consistency with existing sources), lineage and impact analysis reporting, and the ability to synchronize metadata across multiple instances of the tool. It also involves an open metadata repository, including mechanisms for bidirectional sharing of metadata with other tools:
 - Active metadata capabilities support the extensive use of metadata (for example, usage data, transaction logs and system workloads) that can automate the broader data management and

operations tasks through ML and AI.

- Augmented data integration support: These capabilities improve and optimize data integration operations (such as self-healing schema drifts and autorecovery) using extensive use of metadata (for example, usage data, transaction logs and system workloads) and prepackaged ML algorithms that can inform or automate the tasks to ingest, transform, combine and provision data.
- Support for data preparation capabilities: This is the usability of data integration tools both for data engineers and citizen integrators and their suitability to support a range of business roles (for example, citizen integrators and business analysts) for self-service. The emphasis is on empowering nontechnical staff using various techniques such as low-/no-code data blending and visual exploration.
- Design and development support: This includes a graphical design/development environment and team development capabilities, such as version control and collaboration. It also includes multiple versions running in disparate platforms and multiple instances of service deployments in production environments, as well as alternative or collaborating development environments.
- API services: This is data as a service enabled through API design capabilities to create and manage outbound API endpoints over existing data assets, and to handle inbound API consumption to ingest internal and external data.
- DataOps support: Increasingly, organizations also expect DataOps support that is, the ability to create a collaborative environment for IT/data engineers to iteratively work with data consumers through continuous collaboration and through better communication for more repeatable and optimized data delivery. Other requirements include change management capabilities to data and related artifacts (such as Git integration of data pipelines and data model management), automation (for example, automated testing), and orchestration of data delivery (such as CI/CD pipelines) with appropriate levels of security to improve the use and value of data.
- Integration portability: This is data flow design portability across the infrastructure (at onpremises, SaaS, CSP and virtual private cloud [VPC]), providing workload management capabilities in a clean, safe and portable runtime environment (like containerization).
- Data quality/governance support: This is the ability to import, export and directly access metadata with/and from data profiling, data quality tools and/or other data-governance-enabling technologies (such as MDM, information stewardship, metadata management and data catalog tooling). Accepting business and data management rule updates from data stewardship workflows and sharing data profiling information with such tools is highly desired. Capabilities that assist data governance mandates (such as data quality and data lineage) while handling data for meeting specific use cases (master data management, data sharing and so on) are also needed.

12/11/21, 9:56 AM

Gartner Reprint

- Runtime platform support: This is the ability to deploy and run data integration tools on multiple platforms including Windows, UNIX and/or Linux operating systems.
- Service enablement support: This is the ability to deploy functionality as services, including manners in which functionality can be called via a web service interface.
- Support for the delivery of data integration functionality as cloud services: This could be done through a hosted, containerized, PaaS, IaaS or SaaS delivery mechanism. The ability to perform integration across a hybrid and possibly a multicloud and intercloud ecosystem is highly desired.

In addition, vendors must satisfy the following quantitative requirements regarding their market penetration and customer base:

- Revenue or customer count:
 - Either generate at least \$30 million in software revenue from data integration tools in the calendar year 2020 that is, from perpetual license with maintenance, subscription with support (which would include payment only for data integration software), or through a consumption-based licensing model where the consumption metrics are being used *only* for the data integration software (on an annual basis).
 - Or maintain at least 400 paying customers for its data integration tools in production. (The number of downloads without license or maintenance revenue is informative, but not a qualifying piece of information.)
 - Note: Gartner will use as many independent resources for validating this information as possible, specifically to validate provided information. Gartner may request that you provide a written confirmation of achievement of the following requirements. The confirmation must be from an appropriate finance executive within your organization.
- Geography: The vendor must support data integration tool customers in at least two of the following geographic regions or specific national markets: North America, South America, EMEA and Asia/Pacific. Specific exceptions may be possible relative to also meeting the revenue or verifiable supported customer counts – understanding that Gartner analyst decisions are final.
- Market Presence: Demonstrated market presence will also be reviewed and can be assessed through internal Gartner search, external search engines, Gartner inquiry interest, technical press presence and activity in user groups or posts. A relative lack of market presence could be determined as a reason to exclude a product/service offering.

Vendors that focus on narrow use cases that are too specific for the broader data integration market were excluded. In the past, some vendor/supplier tools were excluded because:

- They focused on only one horizontal data subject area for example, the integration of customeridentifying data.
- They focused on only a single vertical industry.
- They served only their own internally managed data models and/or architectures (including tools that only ingest data to a single proprietary data repository). Or, they were used by a single data discovery/visualization, analytics/BI tool, data science/ML platform, or DBMS/data management solution for analytics, data lake management, data warehouse automation or cloud ecosystem vendor. These vendors use their data integration tools only to ingest/integrate data into their own repository or within the confinement of their own broader tool/platform or ecosystem.
- They provided data integration as a capability embedded within their broader data management/analytics/data science platform but did not provide a stand-alone/independent or commercially off-the shelf available data integration tool product.
- Vendors that only provide support for open-source platforms/frameworks or development platforms, which need to be heavily engineered/customized for specific data integration tasks/use cases and/or are specific to a single data integration/data delivery style (such as stream data integration only).
- Vendors that provide adapters or drivers to various data and analytics sources and targets, thereby indirectly supporting data integration, but these vendors do not market a stand-alone data integration tool.
- Vendors that only provide self-service data preparation tools for citizen integrators, power users, analysts and line-of-business (LOB) users, but these tools do not have the ability to physically move data or operationalize these self-service data flows and models into production through data movement, governance and sharing, if and when needed.

Honorable Mentions

Actian, owned by HCL and headquartered in Palo Alto, California, offers the DataConnect product as its primary data integration tool. Its operations are geographically diverse, and it has over 8,000 customers. Its customers tend to be enterprise B2B and OEM partners. Its data integration tooling aligns with its cloud-based data warehouse, Avalanche, for meeting targeted data needs. The ability to handle complex workloads, wide range of connectivity, and bulk/batch data movement are among major factors for buyers adopting Actian's data integration capabilities. Actian was not included in this year's Magic Quadrant because it did not meet the updated inclusion criteria, which requires demonstrable market presence (and execution) as a stand-alone data integration tool. Gartner inquiry analysis (which includes vendor proposal reviews) shows relative lack of inclusion of Actian DataConnect tool by end-user clients, in competitive situations, for data integration scenarios that don't involve Actian's data warehouse as a target data store.

- Amazon Web Services (AWS), headquartered in Seattle, Washington, is the largest cloud service provider in the world by revenue, with an international presence and a global client base across all major industries. It offers AWS Glue as its serverless data integration service that enables data access and discovery. Glue has a data catalog and schema registry, as well as a connector marketplace with a growing number of integrations with third-party and on-premises data stores. Glue has multiple persona-specific job authoring tools and automated ETL code generation for its Python and Apache Spark-based serverless ETL engines. Its job-authoring experiences include IDEs and notebooks for data engineers, AWS Glue Studio for ETL developers, AWS Glue DataBrew for citizen data scientists, and a task orchestration system to schedule and run end-to-end batch and real-time workloads. It offers data replication via its AWS Database Migration Service (DMS) to ingest data into AWS cloud, ingestion from SaaS systems with Amazon AppFlow. It offers data virtualization via its Amazon Athena query service. At this time, we do not see enough evidence to support Glue as an independent/stand-alone offering that enterprises procure to integrate and share data in non-AWS environments. However, AWS continues to invest in new features and its adoption is growing at a fast pace, making it a good choice for AWS customers.
- Confluent, headquartered in Mountain View, California, offers Confluent Platform, which is a selfmanaged Kafka distribution for enabling stream data integration and stream analytics use cases within organizations. Confluent also provides Confluent Cloud, which is a fully managed, serverlerless and cloud-native service for Apache Kafka. Over the years, Confluent has become significant in the data integration tools market due its popularity among organizations that need Kafka distribution support for their streaming, messaging and event-based data integration scenarios. Organizations looking to scale with Kafka prefer Confluent for its additional capabilities for data connectivity, security, governance and orchestration. Based on its capabilities, Confluent is fairly often used for stream data ingestion (a functional subset of stream data integration). Streaming data that enters Confluent Platform or Confluent Cloud can be landed in a data store (such as a Hive table or Snowflake) using off-the-shelf connectors from Confluent without custom user programming. This common pattern is essentially an ELT usage scenario (but Confluent also supports ETL for real-time data streams). The other three usage scenarios supported by Confluent are (1) messaging support for continuous intelligence [its biggest usage scenario], (2) messaging for transactions [message queuing] and (3) event stream processing (ESP)/stream analytics. Even though Confluent has become popular in the stream data integration subset of the broader data integration tools market, it was excluded from this years' MQ evaluation because it does not provide data virtualization, data services orchestration, or automated discovery and acquisition of metadata from data sources.
- DataStreams, headquartered in South Korea, provides the TeraONE Data Fabric platform, which provides capabilities for data integration, data governance and big data management. It also provides stand-alone products for specific data integration patterns, such as TeraStream for batch data processing, TeraStream BASS for streaming ETL, DeltaStream for CDC, TeraONE for Data Lake and TeraONE Super Query for data virtualization. While data governance is the biggest use

case with its IRUDA platform, it supports analytical use cases as well through its own visualization tools, SuperVisual and TeraOne Idea, and through partnerships with Tableau, Qlik and others. Most of DataStreams' revenue comes from South Korea, and this is the reason it could not meet the inclusion criteria for this Magic Quadrant.

- eQ Technologic, headquartered in Costa Mesa, California, provides the eQube Data as a Service (eQube-DaaS) platform. eQube-DaaS is closely aligned to the data fabric design and offers a lowcode/no-code integration platform that supports multiple data integration techniques (through embedded support for batch, streaming, messaging, APIs, data virtualization and application integration). The eQube-DaaS platform is made up of three loosely coupled (yet tightly integrated) offerings: eQube-MI (for data integration and data migration use cases), eQube-AG (for application integration and API gateway) and eQube-TM (for data model management and data transformation maps). eQ Technologic also offers an independent data virtualization service as part of the eQube-DaaS platform. This service can create virtual views of data from heterogeneous sources and then expose these integrated views as REST, API or Open Data Protocol (OData) services. Beyond data virtualization, eQ Technologic also provides support for stream data integration and messaging use cases within the eQube-DaaS platform. Asset-heavy industries (including aerospace and defense, ship building, high tech, auto and machinery, energy, consumer packaged goods [CPG] and so on) looking to establish enterprise search, API creation, synchronization of multi product life cycle management (multi-PLM) systems and synchronization of PLM data with popular ERP systems (like SAP) will also benefit from eQ Technologic's native connectors to industrial applications and IoT systems. While eQube-DaaS did offer all the data integration capabilities for this market, it did not make the inclusion criteria related to market execution for this year's Magic Quadrant iteration.
- Google Cloud Platform (GCP), headquartered in Mountain View, California, provides a comprehensive set of data integration capabilities as part of its data management tools portfolio. GCP offers Cloud Data Fusion, which is a fully managed and cloud-native enterprise data integration service that delivers ETL/ELT capabilities and comes with a significant number of preconfigured, native connectors to various popular heterogeneous data sources, including databases, enterprise applications, SaaS applications, IoT data sources and other third-party cloud services. Cloud Data Fusion is built with an open-source core (on Cask Data Application Platform [CDAP]), which enables pipeline portability and orchestration to various execution engines and environments. GCP also offers a new data replication service, Datastream, that delivers log-based, change-data-capture-based data delivery to Google's serverless data warehouse, BigQuery, and Google databases like CloudSQL, Spanner, Bigtable and Firestore. For customers looking to migrate data to GCP's data stores, GCP offers the Database Migration Service to support bulk-/batch-based data migration. For data preparation requirements, Google embeds Trifacta's data preparation capabilities and offers the service as "Dataprep by Trifacta." For messageoriented data movement, Google offers the Pub/Sub integration service, and for customers looking to support IoT data ingestion, operational intelligence and streaming analytics use cases, Google

offers the Cloud Dataflow service. Although GCP offers comprehensive data integration services to support many customer use cases, Gartner would like to see more evidence that showcases the use of GCP's data integration tools by customers in multicloud/hybrid and intercloud scenarios, where GCP's data integration services are being used to integrate the data with non-Google data sources and targets.

- Hitachi Vantara, based in Santa Clara, California, offers Pentaho Data Integration (PDI) and Lumada Data Catalog within the Lumada DataOps Suite, which comprises analytics and data management technologies. The vendor's operations are geographically diverse with a customer base of over 880 organizations. PDI has a strong following and has momentum for its mature ETL/ELT capabilities (given its roots as an open-source service). Beyond PDI, Hitachi Vantara has expanded its portfolio to deliver streaming, messaging and API-based data delivery services. The recent acquisitions of Waterline Data and Io-Tahoe have improved the support for smart metadata discovery, data cataloging, metadata management and data governance capabilities within the Lumada DataOps Suite. Companies favor Hitachi Vantara's data integration tools for their support for IoT data integration and edge data access use cases. The vendor has also aligned its data integration services to the DataOps discipline by improving the modularity of its various tools within the Lumada DataOps platform. Hitachi Vantara has improved DataOps-enabling capabilities - such as data pipeline automation, CI/CD and Git integration, and DataOps support for accessing and distributing data – by enabling developed artifacts to be made embeddable and available as microservices. Hitachi Vantara was not included in this year's Magic Quadrant because it did not meet the updated inclusion criteria and the required information was not made available by the vendor to substantiate its ability to meet inclusion criteria.
- Irion, headquartered in Torino, Italy, provides Irion EDM, a metadata-driven data management platform that offers various data integration capabilities as modular services. Irion EDM provides Declarative Extract Load & Transform (DELT) data integration technology, which enables its users to abstract the underlying complexity of data objects (and models) and declare the required output and mappings ahead of time. Beyond a declarative way to develop and deliver ELT pipelines, Irion EDM also provides the ability to analyze all forms of passive metadata (to convert this into active metadata) and to use the output of this analysis to support automating various data integration and data transformation tasks. Irion EDM customers can start with data integration and then expand to cover other use cases requiring data quality, master data management, metadata management and data governance capabilities using Irion's embedded support for these capabilities within the Irion EDM platform. Founded in 2004, Irion has successfully ramped up data management and integration capabilities within its platform, but did not qualify for this Magic Quadrant due to its limited geographical presence (beyond Italy) and due to not meeting the revenue and customer count thresholds for this market.

- K2View, headquartered in Israel, offers the K2View Data Fabric platform. It approaches data integration with autodiscovery and modeling of business entities such as customer, location and device. Its patented "micro-database" technology integrates, cleanses, masks and manages data for a business entity in its own microdatabase on Cassandra. Every microdatabase is compressed and encrypted, and enables holistic high-speed data access at scale. It fits the data fabric use case very well. Its primary customer base is from telco, financial services and healthcare sectors. Its common use cases include customer 360, data privacy and governance, cloud migration, and data preparation and delivery into data lakes and warehouses.
- StreamSets, headquartered in San Francisco, California, provides the StreamSets DataOps Platform. Composed of StreamSets Data Collector, StreamSets Transformer and StreamSets Control Hub, the platform enables organizations to use streaming, batch and change data capture integration styles to build modern data pipelines. The company is a proponent of the DataOps methodology, which includes a substantial focus on providing speedy and flexible access to data. Founded in 2014, StreamSets has made good headway into the enterprise market, particularly for stream data integration, but the company's total revenue/number of clients did not meet the inclusion criteria for this Magic Quadrant.

Evaluation Criteria

Ability to Execute

Gartner analysts evaluate providers on the quality and efficacy of the processes, systems, methods or procedures that enable IT provider performance to be competitive, efficient and effective, and to positively impact revenue, retention and reputation within Gartner's view of the market. Ultimately, technology providers are judged on their ability to capitalize on their vision.

In this research, we evaluate the vendors' Ability to Execute in the data integration tool market by using the following criteria.

Product/Service

These are core goods and services that compete in and or serve the defined market. This includes current product and service capabilities, quality, feature sets, skills and so on. This can be offered natively or through OEM agreements/partnerships as defined in the market definition and detailed in the subcriteria.

We rated the vendors on:

The vendors' capabilities that address current market requirements. These include but are not limited to bulk/batch data movement, CDC-based data replication and synchronization, data services orchestration, stream data integration for real-time use cases, data migration support,

support for data engineering for analytics and data science, and other integration efforts for operational use cases (like MDM).

- The degree of openness of the vendor technology and product strategy that is, the ability to exchange metadata with third-party offerings.
- Some consumers are prepared to accept products from many different suppliers and assemble them together on their own. Therefore, the ability of offerings to allow interoperability to opensource solutions and third-party offerings is appreciated by end users.
- Connectivity options to not only nonrelational databases, cloud applications and cloud data stores (such as cloud object stores and cloud data warehouses), but also traditional stores (including relational databases and enterprise applications).
- Connecting data integration activities to data quality and governance becomes integral in supporting those operational data integration use cases that require sharing high-quality data along with its lineage, such as master data management and B2B data sharing.
- The ability to offer both serverless metered pricing options (for net new use cases), and traditional pricing models such as node-/core-based models (when the use cases do not require flexibility of compute).

Overall Viability

Viability includes an assessment of the organization's overall financial health, as well as the financial and practical success of the business unit. These criteria view the likelihood of the organization to continue to offer and invest in the product as well as the product position in the current portfolio.

We rated the vendors on:

- The appropriateness of the vendor's financial resources, the continuity of its people and its technological consistency and how that affects the practical success of the business unit or organization in generating business results.
- The growth of their product lines, their ARRs, profitability and also growth in new geographies/use cases.
- Product/services growth in revenue to determine the vendor growth in the data integration software market.
- Other metrics to determine financial viability and spend on R&D efforts to continue differentiation and growth in product lines.

Sales Execution/Pricing

This is the organization's capabilities in all presales activities and the structure that supports them. This includes deal management, pricing and negotiation, presales support, and the overall effectiveness of the sales channel.

We rated the vendors on:

- The ability of vendors to offer modular solutions. Organizations increasingly seek "modularity" or the capability to isolate specific required functions in data integration that are then reflected in their implementation approach and cost allocation.
- Ability to provide tools and capabilities through different pricing models appropriate by use cases, persona and environment is rated highly.
- The ability of vendors to support buyers that are looking for new pricing metrics that abstracts them from the underlying metrics of cloud pricing. They are looking for vendor options that support serverless metered pricing metrics that are a true reflection of the actual work done and that can separate compute from storage/infrastructure.
 - Having said that, organizations are also wary that serverless metered pricing options that don't take into account good financial governance can soon get out of control. Gartner will be closely evaluating vendors on their ability to enforce financial governance metrics into their pricing models and licensing metrics.
- On the ease with which customers can hold them accountable for agreed upon SLAs. Buyers are evaluating ways through which they can hold vendors accountable for the promised SLAs (in terms of uptime, turnaround times to issues, bug fixes, migrations and so on). Providers must demonstrate ways through which customers can escalate and attain credits/discounts when SLAs are not met.

Market Responsiveness/Track Record

This measures vendors' ability to respond, change direction, be flexible and achieve competitive success as opportunities develop, competitors act, customer needs evolve and market dynamics change. This criterion also considers the provider's history of responsiveness to changing market demands.

We rated the vendors on:

We are looking for evidence on how the vendors "course-corrected" to changing buyer requirements in terms of use cases, upcoming capabilities, pricing models and support requirements.

- As an example, managed services options for maintaining data pipelines and handling schema drifts are in demand, particularly by business teams and citizen integrators. Providers that can enable these requests are therefore selected over others that are still focused on IT teams alone.
- The requirements to enable data fabric designs are also increasing. We are therefore looking for those vendors that are adding features (including those on the roadmap), formulating partnerships (such as the ones with graph providers, semantics technologies and so on) to enable more comprehensive data fabrics.
- Even though solutions that provide low-code/no-code UIs are preferred, we are also getting requests from data engineering teams for tools that can open up "portals" for custom coding and importing scripts created in languages such as R and Python for highly advanced transformations.
- The market is also looking for vendors that have the know-how of moving away from environments that seem to have lost traction — for example, Hadoop — and moving toward other popular data/sources and targets, such as cloud database platforms as a service and cloud applications.

Marketing Execution

This is the clarity, quality, creativity and efficacy of programs designed to deliver the organization's message in order to influence the market, promote the brand, increase awareness of products and establish a positive identification in the minds of customers. This "mind share" can be driven by a combination of publicity, promotional activity, thought leadership, social media, referrals and sales activities.

- Brand recall value has a high premium in a mature market like data integration.
- Providers must develop a means of converting community "chatter" and excitement to support delivery and go-to-market campaigns.
- The overall effectiveness of the vendor's marketing efforts which impact its mind share, market share and account penetration – is important.
- The ability of the vendor to adapt to changing demands in the market by aligning its product message with new trends and end-user interests was part of the evaluation.
- Suppliers need to be aware of emerging best practices for data management infrastructure, and if they and their customers can specifically benefit from specialized horizontal or vertical capabilities, geographically targeted approaches or partner-supported implementation practices.

These products and services and/or programs should enable customers to achieve anticipated results with the products evaluated. Specifically, this includes quality supplier/buyer interactions, technical support or account support. This may also include ancillary tools, customer support programs, availability of user groups and service-level agreements.

We rated the vendors on:

- We will evaluate the level of satisfaction expressed by customers with the vendor's product support and professional services.
- We will also look at customers' overall relationship with the vendor, the experience while upgrading software versions, the learning curve for new users given the training resources available, and customer perceptions of the value of the vendor's data integration tools relative to cost and expectations.
- We will look at various platforms for such data points including but not limited to our interactions with end users in inquiries, Peer Insights data, surveys, customer reference calls, touchpoints across various Gartner and external events, community chatter, and vendor briefing data.
- The distinction between advanced use cases and "pedestrian" applications is becoming more pronounced. The evaluation this year is focused on separating success in "traditional" market delivery from success in "innovative" market delivery in reviewing the customer experience.

Operations

This is the ability of the organization to meet goals and commitments. Factors include quality of the organizational structure, skills, experiences, programs, systems and other vehicles that enable the organization to operate effectively and efficiently.

- Operations are not specifically differentiating to end-user markets, but product management consistency and support/maintenance practices add to the overall customer experience and to the stability of senior staff.
- Suppliers need to demonstrate a new balance in their R&D allocation to ensure they are positioned for deployment with greater focus on data services, metadata management and semantic tiers.
- Also, they must demonstrate that they are well-positioned to provide ongoing support for the massive bulk/batch data movement market.
- Partner programs, skills augmentation, improvements in support and services, training materials and programs, and delivery with external service providers are all important in this evaluation criteria.

Table 1: Ability to Execute Evaluation Criteria

Evaluation Criteria ↓	Weighting ↓
Product or Service	High
Overall Viability	High
Sales Execution/Pricing	High
Market Responsiveness/Record	Medium
Marketing Execution	Medium
Customer Experience	High
Operations	Low

Gartner (August 2021)

Completeness of Vision

Market Understanding

This is the ability to understand customer needs and translate them into products and services. Vendors that show a clear vision of their market can listen, understand customer demands, and shape or enhance market changes with their added vision.

- The ability to formulate product vision around multicloud/intercloud and hybrid data integration capabilities
- The ability to provide "advisors" and other insights in design, development, deployment and management of integration services

- The ability to automate and augment data integration design and delivery through active metadata analysis and recommendation engines
- The ability to provide business-user-friendly UIs through self-service data preparation modules and to allow skilled users to operationalize self-service findings and flows
- The ability to work with data services through API management and integration and to deliver application and data integration flows together if the use case demands it.
- The ability to contribute to data fabrics/mesh network designs

Marketing Strategy

This is the clear, differentiated messaging consistently communicated internally and externalized through social media, advertising, customer programs and positioning statements.

We rated the vendors on:

- Provider's need to improve community, internal and external messaging in order to differentiate their offerings from various other categories in the market — for example, basic ingestion tools as well as integration tools supporting only open-source frameworks.
- Clear messaging on trial/freemium to full enterprise offerings (with differentiators across each support/SLA level)
- Good visibility into the product portfolio along with features across each separate tool including
 possible overlaps, ways to buy, means to procure, support tiers and licensing is now a must.
- Efforts and investments, with demonstrations, into partner programs, training programs, OEM/value-added reseller (VAR), other partnerships, cloud provider partnerships, SI partnerships and so on.
- Demonstrated proof of expansion in training, certifications and availability of talent in the market (through partner programs, training and so on).

Sales Strategy

This involves the strategy for selling and using the appropriate networks, including direct and indirect sales, marketing, service, and communication. This includes partners that extend the scope and depth of market reach, expertise, technologies, services and customer base.

We rated the vendors on:

Expansion in sales partner networks

- Strategy to grow beyond existing markets, use cases, geographies and critical capabilities
- Demonstrated evidence of improvement in communication of existing and upcoming tools/services
- Affiliate partnerships
- Evidence of becoming part of cloud ecosystems that enhance growth and sales

Offering (Product) Strategy

This is an approach to product development and delivery that emphasizes market differentiation, functionality, methodology and features as they map to current and future requirements.

- This section is about aligning existing tools and roadmaps with future market direction.
- We are looking for tools that can deliver distributed data integration across on-premises, cloud, intercloud and edge ecosystems.
- Tools must exhibit improvement in automation-oriented capabilities.
- While advanced capabilities are needed, the tools must not drop existing and "traditional" requirements of data integration, including bulk/batch capabilities, supporting hybrid/on-premises sources and targets, supporting developers and so on.
- There is now significant increased expectation on "active" metadata understanding, conversion, utilization and analysis:
 - This active metadata is used in profiling, machine learning, evaluation of assets and comparison with existing integration upon connection.
 - Self-correcting optimization in processes is now important and expected.
 - Utilizing metadata to assist in user "push" recommendations for new data assets and to create semantic knowledge graphs to assist with data fabric design that enables a more consistent (and application-neutral) semantic model for integration — is considered a differentiator.
- Given the requirement for data integration tools to support diverse environments for data, delivery models and platform-mix perspective, we assess vendors on the degree of openness of their technology and interoperability with other data and analytics tools.

Business Model

This is the design, logic and execution of the organization's business proposition to achieve continued success.

We rated the vendors on:

- Although broad, all-inclusive delivery models represent one "best-of-breed" solution approach, it is also both expected and reasonable to assume that tightly targeted models for traditional delivery needs can cut delivery cost, increase adoption and deliver specific integration needs to end-user organizations.
- The overall approach the vendor takes to execute on its strategy for the data integration tool market — including diversity of delivery models, packaging and pricing options, and partnerships is important.
- The ability of the vendors to provide both "current" requirements through best-fit engineering tools versus future requirements through end-to-end platforms or best-of-breed options is a good measure of this category.
- The business proposition must include the ability for end-user organizations to try before they buy, and the tools must be in a position to interoperate with existing tools within the customer base rather than having to replace all current software.

Vertical/Industry Strategy

The strategy to direct resources (sales, product and development), skills and products to meet the specific needs of individual market segments, including verticals.

We rated the vendors on:

- We look at the degree of emphasis the vendor places on vertical solutions, and the vendor's depth of vertical market expertise.
- Although most prospects are not looking for data integration tools focused on a specific vertical/domain because they rightly treat data integration as industry agnostic, some organizations might favor tools that are able to create a specific industry model, ontology or knowledge graph based on an industry-specific taxonomy.
- Vertical/domain specific solution accelerators, KPIs, best practices and other industry starter templates might be favored by some buyers as well, in addition to industry/domain experts being a part of the professional services provided.

Innovation

This is direct, related, complementary and synergistic layouts of resources, expertise or capital for investment, consolidation, defensive or preemptive purposes.

We rated the vendors on:

- The current innovation demands in the market are centered on managing location-agnostic capability in data integration that is, the ability to not have to move or replicate data necessarily but to connect to data in-place when feasible and take the processing to the data (rather than vice versa) to execute integration.
- Integration should run on-premises and in the cloud, and switch between them.
- As data becomes highly distributed, data integration activities are also required to become easily distributable to any data location, and vendors should be able to recommend/determine when data needs to be moved for optimal processing.
- Converging data and application integration approaches is now expected.
- ML-based automation using internal analytics on all kinds of collected metadata to support integration activities is another area of improvement that the market currently demands.
- The growing diversity of users indicates a much higher demand for administrative, auditing, monitoring and even governance controls that utilize job audit statistics.
- Graph analysis to determine user classification and optimization "hints" are also increasingly demanded.
- Finally, because the increase in the number of data pipelines is inevitable, organizations are expecting DataOps-oriented capabilities that can support CI/CD; project management capabilities such as Git and Jenkins; automated testing/validation; the handling of various environments in an agile manner; sandboxes on demand and management of them; and agile pipelines creation, reuse, execution, management and so on.

Geographic Strategy

The provider's strategy is to direct resources, skills and offerings to meet the specific needs of geographies outside the "home" or native geography — either directly or through partners, channels and subsidiaries — as appropriate for that geography and market.

We rated the vendors on:

 The ability for vendors to provide their customers with local support with differing levels of confidence in the various approaches possible (that is, VARs, resellers, channel partners, OEM offerings and distributors)

- The ability to provide continuity of support across regions
- The ability for development platforms to monitor where data originates with jurisdictional cognizance, and where it is eventually delivered.
- Their ability to address the possible violation of national laws due to data movement
- The vendor's strategy for expanding into markets beyond its home region/country and its approach to achieving global presence (for example, direct local presence and use of resellers/distributors) are crucial for capitalizing on global demands for data integration capabilities and expertise

Table 2: Completeness of Vision Evaluation Criteria

Evaluation Criteria ↓	Weighting ↓
Market Understanding	High
Marketing Strategy	Medium
Sales Strategy	Medium
Offering (Product) Strategy	High
Business Model	Low
Vertical/Industry Strategy	Low
Innovation	High
Geographic Strategy	Low

Gartner (August 2021)

Quadrant Descriptions

Leaders

Leaders in the data integration tool market are front-runners in their capability to support the combination of these different data delivery styles (for example, the ability to combine and switch between ETL, replication and virtualization based on their use-case demands). Leaders exhibit significant market mind share, and resources skilled in their tools are readily available. These vendors recognize the need for new and emerging market demands — often providing new functional capabilities in their products ahead of demand — by identifying new types of business problems to which data integration tools can bring significant value. Examples of deployments that span multiple projects and types of use cases are common among Leaders' customers. Leaders have an established market presence, significant size and a multinational presence — either directly or through a parent company.

In 2021, Leaders in this market have started delivering on the data fabric promise — that is, their ability to balance collecting data with connecting to data. They automate the process of collecting all types of metadata (not just passive) and then represent the metadata (and sometimes data) in a graph (to preserve context). This is then followed by improving the data modeling process by enriching the models with agreed-upon semantics. Finally, some of these vendors embed AI/ML toolkits, which utilize active metadata (as input) to start automating (or at the very least, informing) various aspects of data integration design and delivery. Most vendors in the Leaders quadrant provide capabilities to deliver the data fabric, although some might require significant customization.

Leaders have been advancing their metadata capabilities, with some introducing highly dynamic optimization and advanced design assistance functions. They have been extending their capabilities to allow for ML over this active metadata to assist developers with various degrees of support and automation in integration design and implementation. Leaders are adept at providing tools that can support both hybrid integration and multicloud integration options, bridging the data silos that exist across on-premises and multicloud ecosystems. Leaders allow organizations to remain independent in data integration as they look to deploy workloads across multiple CSPs, and they allow organizations to effectively provision cloud ecosystems.

Leaders provide the ability to enable efficient data engineering through self-service data preparation capabilities that can be operationalized to enable integration portability through serverless execution capabilities. They also provide the ability to deliver pipelines and code through containerized services. Finally, some Leaders have also added DataOps-enabling capabilities to support CI/CD and integration with version control and project management tools like Git.

Leaders are strong in establishing their data integration tools as enterprise standard in at least one primary use case, with the capabilities to deliver in multiple use cases. Their data integration tools become a critical component of the modern data management infrastructure. Their tools support both traditional and new data integration patterns to capitalize on market demand.

Challengers

In 2021, there was a shift in market realization that data delivery styles other than bulk/batch (such as replication, streaming or data virtualization) are no longer differentiating but more a "must have" — attributing this capability to execution rather than vision. In line with this market shift, Challengers have been making significant strides in delivering these capabilities within a broader metadata-driven data integration toolset.

In 2021, the Challengers listed in this research constitute vendors that exhibit a strong understanding of the current data integration market demand and exhibit both the credibility and viability to deliver on that demand. Some Challengers are extremely mature in specific core capabilities, which enables them to deliver targeted use cases faster and with a better overall TCO than other vendors. These vendors have developed best practices for leveraging their strongest product capability in new delivery models. For example, they have the ability to productize and market (1) data replication as a key strength for targeted use cases such as cloud data migration or (2) data virtualization for faster turnaround time to analytics.

Challengers generally have substantial customer bases. They exhibit strong market presence, although implementations may be of a single-project nature or reflect multiple projects of a single type — for example, predominantly data replication or application integration, or use cases specific to IoT or mainframe data or geospatial data type. Gartner realizes that, while the market vision is for the vendors to deliver all data delivery styles across a multicloud/hybrid environment and support all popular data sources/targets, many customers could have specialized and targeted demands for their most urgent or upcoming projects. Therefore, we also recognize vendors in the Challengers quadrant that (if needed) can scale to support most data integration use cases, but can also customize their offerings for specific use cases, data types, data sources/targets, execution environments or specific CSPs in demand.

Overall, the market is pushing Challengers to embrace the market vision of a data fabric and automated data integration, which ensures that they utilize metadata and perform ML over this metadata to deliver solutions that can automate various data integration tasks. These tasks include automated profiling, automation of repetitive transformations, data preparation (and the ability to use ML to automate the operationalization of self-service data preparation models), performance optimization, query optimization, and movement of workloads to data stores and engines that are best-suited for processing, among others. Overall, this move toward enabling data fabric architectures is a key area in 2021 that will determine which Challengers can move into the Leaders quadrant next year.

Visionaries

Visionaries demonstrate a strong understanding of emerging technology and business trends or focus on a specific market need that is far outside of common practices, while also possessing capabilities that are expected to grow in demand. In 2021, the Visionary in this Magic Quadrant has focused early on alternative go-to-market strategies or specific capabilities to capitalize on their capacity to leverage:

- Augmented data integration through the data fabric design
- Serverless integration tooling that supports a multicloud and hybrid cloud integration architecture
- The growth in demand for connectors and delivery of various data integration functionalities as loosely coupled API/microservices
- Seamless orchestration of various data integration components and delivery styles through DataOps techniques
- iPaaS delivery models to capture the trend of cloud data integration and the convergence of the application integration and data integration tools market, led by the convergence of data integration, application integration and API integration/API management use cases

In addition, a significant driver of vision for the market this year has been the ability of tools to connect to and analyze all forms of metadata — both passive and, increasingly, active metadata. With this, tools can provide key statistics to developers and citizen integrators that aid with integration design and, increasingly, integration automation. The Visionary listed in this quadrant this year is doing all these things. Visionaries should lead the push toward the utilization of graphs, semantics, knowledge graphs and Al/ML for significant automation in data integration design, delivery and maintenance. Visionaries sometimes lack market mind share or credibility beyond their customer base, their main use case or single application domain. They may also fail to provide a comprehensive set of product capabilities — including those that focus on a single data integration style — and simply import, export or leverage that primary data integration style to deliver customer use cases. They may lack the installed base and global presence of larger vendors. Finally, Visionaries may be established players in related markets that have only recently placed an emphasis on data integration.

Niche Players

With the market now matured, Niche Players generally don't exhibit gaps in primary market functionality or features. Instead, they are simply challenged to improve their execution or have not identified a specific market approach that expands use cases for their technology. This means that almost every Niche Player will be able to deliver against standard market expectations, both in functionality and cost/price options.

Niche Players may not appear very frequently in competitive situations for comprehensive and/or enterprise-class data integration deployments (where the customers are purchasing their tools as the only tool to deliver all data integration use cases). Many have very strong offerings for a specific range of data integration problems — for example, a set of technical environments, data delivery styles (batch, replication, streaming or virtual), application domains or use-case scenarios — and deliver substantial value for their customers in the associated segment. Niche Players now exhibit particular advantages in pricing in their small footprint and even in vertical or horizontal solutions. This makes them ideal candidates to be a best-fit solution that complements other technology in the data management infrastructure of an organization.

Niche Players are known for solving one part of the data integration problem well through a targeted solution. This could be data ingestion to the cloud, support for complex ETL workloads, B2B integration, real-time data replication, location intelligence through geospatial data integration, or data integration support for business teams through managed services delivery of their software. Importantly, Niche Players in this market have demonstrated their capability to outperform dozens of tool and solution offerings that were considered and eventually excluded from this Magic Quadrant. Finally, Niche Players may be lacking maturity on certain features that display market vision. These features might include multicloud/hybrid support or support for data fabric architectures, or even the ability to inform and automate data integration design and delivery through active metadata support and analysis.

Context

The market for data integration tools continues to evolve and is supported by strong levels of market growth and adoption. More data and analytics leaders are realizing that data integration is a critical component of their data management infrastructure. They understand that they need to employ data integration functions to share data across all organizational and systemic boundaries. Therefore, organizations are increasingly seeking a comprehensive range of improved data integration capabilities to modernize their data, analytics and application infrastructures.

Data and analytics leaders must navigate a market brimming with products that claim to solve a range of data integration problem types. However, not all vendor solutions have experience in, or evenly provide, all the relevant capabilities needed across our key data integration use cases (see the companion Critical Capabilities for Data Integration Tools). Some vendors focus heavily on providing solutions focused on just one data integration style such as bulk/batch (through ETL or ELT), data replication (through CDC), messaging through (APIs), or virtual (through data virtualization). But they may place less emphasis on the important capability of interoperating, orchestrating or even combining these different data delivery styles (ETL with data virtualization or data replication or messaging through API integration, for example) for accomplishing key use cases.

Some organizations have determined that basic functions are adequate for them and, therefore, are seeking tools with focused and targeted capabilities. As a result, they are interested in evaluating and

procuring tools that are specialists in one data delivery style (for example, data replication, data ingestion, API integration, self-service data preparation or data virtualization). Also, some organizations prefer tools that can support one use case (such as cloud data ingestion and migration), one data type (such as IoT data integration) or one scenario (such as location intelligence through geospatial data integration focus). Such organizations can confidently start with the vendors in the Niche Players quadrant and then ensure that the shortlisted vendors can also support their upcoming use cases for the medium and long term (using the companion Critical Capabilities for Data Integration Tools).

Organizations that seek tools that are generalists (or best-of-breed in data integration) and can support multiple use cases through a combination of different data integration styles can evaluate the vendors mentioned in the Challengers and Leaders quadrants.

In addition, vendors in the Leaders quadrant are focused on new demands for automation in various aspects of data integration. These include design, ingestion, schema mapping, schema drift detection and corrections, next-best transforms, automated lineage and impact analysis, and infrastructure management and orchestration. These capabilities for augmented data integration demand a new data integration design — one that supports a balance of connect and collect data integration strategies. This design utilizes the collection and analysis of all forms of metadata (not just technical metadata) to provide insights that enable automated data integration delivery (see Data Fabrics Add Augmented Intelligence to Modernize Your Data Integration). Data fabrics are defining the direction of this market, and while Visionaries have a strong product to meet the demand, Leaders have been able to productize this architecture and deliver it at scale due to their superior reach and presence.

Active-metadata-enabled data integration augmentation and automation is a significant driver of market vision this year. Metadata as a byproduct of the design and operations management of a data integration platform is a minimum requirement of data integration tools in 2021. Platforms and solutions are now expected to provide continuous feedback regarding the profiles, quality, location, performance optimization, lineage, use cases, access points, context, frequency of access and content analysis of integrated data assets. As far as architects and solution designers are concerned, this feedback was long overdue. It is expected that graph analytics, powered by every conceivable type of metadata (both passive and active), will provide the necessary dynamic data fabric designs for introducing ML capabilities into data integration platforms (see How to Activate Metadata to Enable a Composable Data Fabric). This capability for active-metadata-based integration has been weighted very highly to define the vision of the market this year by Gartner analysts.

Gartner sees that the urgent need to acquire and integrate data across multiple CSPs, typically for hybrid cloud and intercloud integration, is becoming crucial to many data integration use cases. The COVID-19 pandemic has accelerated the movement to the cloud, and data integration tools are leading the way to support this shift. However, integration tool vendors that support either the

integration of data for just one cloud provider (typically their own cloud) or integration scenarios utilizing their own databases or applications will fall behind on customer demand due to valid lock-in and CSP independence concerns.

An interesting trend from our inquiries revealed that an increasingly high number of data and analytics leaders are investigating and adopting tools that can support data ingestion and replication. This increase is because organizations are looking to utilize their data integration tools to ingest or replicate the data from their operational DBMSs to cloud data warehouses supported by dbPaaS. This has been a significant driver of growth for many data integration providers (such as Fivetran, HVR, Matillion and Qlik). These providers have formed significant partnerships with CSPs like AWS and Microsoft Azure, along with popular cloud data warehouse vendors such as Snowflake and Teradata, to deliver integrated data from on-premises data stores and applications to cloud data warehouses and lakes for analytics. They are doing this through forward engineering, often as ready for consumption as integrated data with schema assigned for analytics and data science use cases.

This year, organizations have also started evaluating their data integration tool vendors for their ability to support integration portability and improve the ability to deliver integration flows, mappings, assets and pipelines in an agile and orchestrated manner. Support for DataOps has been rated as a significant driver for vision by our analyst team in evaluating vendors. This is in line with market demand to support infrastructure as code and to enable the ability to port integration pipelines for optimal executions to environments that best support the required SLAs. Organizations were asking their data engineers to support DevOps capabilities for their data pipelines, but now they expect vendors to deliver an increasing array of such capabilities embedded in solutions ranging from support for CI/CD and integration with Git and Jenkins to providing automated scheduling and serverless execution capabilities. The ability to embed such capabilities was a sign of vision for this edition of the Magic Quadrant.

As more and more subject matter experts and citizen integrators cite willingness to become a part of data integration, their needs should be better supported by data integration tool vendors through native integration with self-service data preparation tools and modules. These modules should enable integration support for less technically skilled persona with low-code/no-code integration environments. This capability is now a must-have and signifies how well a vendor is executing on the market need. Moreover, those vendors that support this need must also support data engineers who are tasked with operationalizing self-service models to production environments after guaranteeing governance and compliance.

Finally, a mix of data integration approaches has remained crucial, spanning from physical delivery to virtualized delivery, and from bulk/batch movements to event-driven granular data propagation. When data is being constantly produced in massive quantities and is always in motion and constantly changing (for example, IoT platforms and data lakes), attempts to collect all this data are neither practical nor viable. This is driving an increase in demand for connection to data, not just the collection of it (see Market Guide for Data Virtualization). In 2021, data virtualization has again been

a key criterion for measuring the vision score of evaluated vendors. But the ability of the tool to deliver data virtualization in combination with other data integration styles — including batch, streaming and messaging (which is necessary for optimal performance optimization) — has been weighted as an execution criterion.

Market Overview

The data integration tools market continues to push toward distributed and dynamic datamanagement-based designs. This push is inherent in the modern data fabric architecture (see Data Fabrics Add Augmented Intelligence to Modernize Your Data Integration). The market has realized that those data integration tools that do not balance "collect"- with "connect"-based data management architecture strategies will always result in data silos and/or poorly integrated infrastructures. In 2021, we have witnessed a variety of trends dictating the data integration tools market. Some of these trends and directions were inevitable, while some have emerged as urgent due to the disruptions (and possible business opportunities) resulting from the COVID-19 pandemic. Moving forward, organizations will need to monitor and exploit the major trends that are affecting enterprise requirements and vendor offerings in the data integration tool market.

We highlight some of these below:

- Data integration remains a stable growth market: The data integration tools market grew at 6.8% in 2020 as compared with 7.0% in 2019. The market was largely not impacted by the pandemic, and this is because data integration is key to running mission-critical systems not a luxury/discretionary tool purchase. This market is therefore expected to grow at a similar trajectory in the near future, with five-year compound annual growth rate (CAGR) for the 2020-2025 time frame forecast to be 6.5% (see Forecast: Enterprise Infrastructure Software, Worldwide, 2019-2025, 2021 Update). Cloud adoption continues to be significant, with the iPaaS market growing by 39% in 2020.
- Market leaders continue to lose ground to smaller vendors: The top five vendors in this market (based on overall market share) had a collective market share of 71% in 2017. This dropped to 56% in 2020. A similar trend can be seen when analyzing the top three or even top 10 vendors. There are multiple reasons for this shift. Market leaders such as Informatica and Talend are ceding market share as they focus their growth efforts primarily on their iPaaS products, which is detrimental to the growth of their traditional data integration products. Vendors gaining market share are sharing a common theme: They focus on leadership in a specific data integration style such as data virtualization or data replication (for example, Qlik, HVR and Denodo), and/or they focus on data integration delivered as a native cloud service (for example, Fivetran, Matillion and Microsoft Azure). Larger and established vendors will need to find the balance between allencompassing platform solutions (through concepts such as the data fabric) and easily accessible point solutions to keep pace (see Market Share Analysis: Data Integration Software, Worldwide, 2020).

- Data fabric is critical and driven by the end-user push toward augmented data integration: Another huge Completeness of Vision criterion that the market is demanding is augmented data integration design and delivery. The COVID-19 pandemic has only fast-tracked this strategic direction of the market. Data and analytics leaders are realizing that they cannot keep investing in manual data integration; they need automation support. Data integration teams (in terms of individual members) are constantly contracting – the median number of individuals in teams is less than 15 (based on anecdotal evidence from our inquiries). And while team sizes are reducing, the amount of data and, hence, the number of data integration requirements are growing exponentially. This gap between demand and supply is pointing toward an urgent focus on automation and augmentation. Augmented data integration demands a renewed focus on the data fabric architecture design, which is a key use case for this year. A data fabric is an architecture pattern that informs and automates the design, integration and deployment of data objects regardless of deployment platforms or architectural approaches. This approach results in faster, informed and, in some cases, completely automated data access and sharing. The data fabric requires various data management capabilities to be combined and used together. These include augmented data catalogs, semantic data enrichment, utilization of graphs for modeling and integration, and finally, an insights layer that uses AI/ML toolkits over metadata graphs to provide actionable recommendations and automated data delivery. In 2021, some vendors have been able to combine most (if not all) of these capabilities into productized solutions, which signifies leadership. Others are going in this direction through partnerships, merger and acquisition (M&A) activity, product enhancements and, more frequently, a combination of all these.
- Data engineering requirements are fueling the next round of investments: Data engineering is the discipline of translating data into usable forms by building and operationalizing data pipelines across various data and analytics platforms. It goes beyond the traditional data management practices to include software engineering and infrastructure operations practices (see How to Build a Data Engineering Practice That Delivers Great Consumer Experiences). An example is using coding languages like Python and Scala to automate data pipeline builds, regression tests, deployments and operations monitoring. With more and more data infrastructure running on the cloud, platform operations are becoming a core part of data teams' responsibilities. The market around data engineering is still emerging, and there are no set industrywide standards, which limits the use-case application. Data Integration tools are stepping up to this need and providing various built-in capabilities to assist end-user customers. Therefore, organizations prefer those data integration tools that embed capabilities that assist with allowing data engineers to build, manage, operationalize and even retire data pipelines in an agile and trusted manner, as well as run their pipelines in various execution environments. Data integration tools that allow optimizing code and pipeline execution through pushdowns, containerizations and serverless execution are being preferred in competitive RFPs.
- DataOps needs to be supported as an emerging discipline: Even though data integration tools don't by themselves provide all capabilities necessary to deliver DataOps, they certainly support

DataOps enablement. DataOps is not a single tool or process but is instead a focus on building collaborative workflows to make data delivery more predictable. DataOps focuses on connecting people, processes and technology to integrate data across an organization. By placing particular attention on key performance indicators (KPIs) such as time to delivery, cost savings, and code quality, using DataOps techniques helps organizations create data pipelines that are rapid, flexible and reliable. For additional information, see Introducing DataOps Into Your Data Management Discipline. Those data integration tools that can support DataOps are naturally being favored over those that don't. Based on our inquiries with clients, key aspects being requested include the ability to deliver data pipelines through CI/CD for agility. Organizations also request capabilities that support automated testing and validation of code. Leading tools also provide the ability to integrate their tools with project management and version-control popular tooling like GIT, Jenkins and Maven, for example. Some data integration tools also enable organizations to manage different nonproduction environments (such as sandbox, development and test/quality assurance [QA]) in an agile manner.

- Hybrid and intercloud data management must be supported: Cloud architectures for data management span hybrid, multicloud and intercloud deployments. There are both risks and benefits in managing data across diverse and distributed deployment environments. Data location impacts performance, data sovereignty, application latency SLAs, high-availability and disaster recovery strategies, and financial models. Gartner estimates nearly half of data management and integration support has been pivotal in the market. Data integration tools are expected to dynamically construct or reconstruct integration infrastructure across a hybrid data management environment. Those tools that can support integrating data across different cloud infrastructures and synchronize it with on-premises data sources and targets have been given more vision scores in this year's Magic Quadrant.
- Independent data integration tools are needed to prevent application/CSP/database lock-in: Another major development this year has been the focus on independent data integration tools that don't necessitate the movement and persistence of data into a specific vendor repository or cloud ecosystem. This is more important than ever because embedded data integration capabilities delivered by vendors as part of a broader application (such as an analytics and BI or CRM tool) or database, or even CSP-specific data integration solutions, might make it easy for organizations to ingest data into one database, application or CSP ecosystem. However, these same embedded integration capabilities do very little to allow organizations to integrate data across different data stores, applications or multicloud/hybrid environments. This could lead to potential vendor lock-in challenges and data silos, resulting in the inability of organizations to reuse integrated data for general-purpose use cases. This is something that organizations can illafford due to their cost optimization initiatives, especially during the COVID-19 era. While the CSP native data integration tools have started becoming more open in terms of allowing for two-way integration (to and from their own cloud data stores), those organizations that are looking to invest

in a general-purpose (and independent) data integration tool for use cases involving more than one cloud service provider generally favor independent data integration tools to partner with those provided by CSPs.

- Convergence of application and data integration is prevalent in the current data integration tools market: The increasing synergy between data integration tooling capabilities and application integration provides opportunities to exploit common areas of both technologies to deliver shared benefits, such as use of CDC tooling that publishes captured changes into message queues. By combining data integration and application integration, organizations are exploiting the intersection of the two disciplines. They align both infrastructure for composing applications and data pipelines for supporting a broad range of operational flows, while simultaneously optimizing costs and competencies. This convergence trend is driving demands for data integration patterns, use cases, endpoints and deployment models to support a hybrid integration platform strategy (see How to Deliver a Truly Hybrid Integration Platform in Steps). As a result, we see organizations preferring those data integration tools that can be delivered via an iPaaS model, particularly when they are looking to drive common projects that need application and data integration convergence.
- Cost optimization through modularity and tightly integrated (but loosely coupled) data integration tool options is critical: COVID-19 has brought back focus on cost optimization in data integration. The market is now expected to deliver modular solutions through loosely coupled architectures, where specific capabilities that are urgently needed by organizations (such as data catalogs or data virtualization) can be delivered and consumed as needed (without having to buy the entire platform upfront). Vendors are therefore now expected to break out key capabilities from broader platforms and deliver them through loosely coupled (but highly integrated) solutions and deployment platforms (such as delivering key capabilities as microservices or through containers). The market is also favoring data integration tools that can be procured and accessed through favorable pricing models based on pay-as-you-go, freemium and serverless metered pricing models. (These new pricing models allow customers to try functionality before buying the full platform and to pay for actual usage in popular cloud environments). These new pricing options should be in addition to, and not a replacement for, traditional performance-based pricing metrics because a part of the market audience continues to value price predictability over pay-as-you-go pricing models.

Evidence

The analysis in this Magic Quadrant 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 in which vendors provided Gartner with updates on their strategy, market positioning, recent key developments and product roadmap.
- Feedback about tools 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 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.

Evaluation Criteria Definitions

Ability to Execute

Product/Service: Core goods and services offered by the vendor for the defined market. This includes current product/service capabilities, quality, feature sets, skills and so on, whether offered natively or through OEM agreements/partnerships as defined in the market definition and detailed in the subcriteria.

Overall Viability: Viability includes an assessment of the overall organization's financial health, the financial and practical success of the business unit, and the likelihood that the individual business unit will continue investing in the product, will continue offering the product and will advance the state of the art within the organization's portfolio of products.

Sales Execution/Pricing: The vendor's capabilities in all presales activities and the structure that supports them. This includes deal management, pricing and negotiation, presales support, and the overall effectiveness of the sales channel.

Market Responsiveness/Record: Ability to respond, change direction, be flexible and achieve competitive success as opportunities develop, competitors act, customer needs evolve and market dynamics change. This criterion also considers the vendor's history of responsiveness.

Marketing Execution: The clarity, quality, creativity and efficacy of programs designed to deliver the organization's message to influence the market, promote the brand and business, increase awareness of the products, and establish a positive identification with the product/brand and organization in the minds of buyers. This "mind share" can be driven by a combination of publicity, promotional initiatives, thought leadership, word of mouth and sales activities.

Customer Experience: Relationships, products and services/programs that enable clients to be successful with the products evaluated. Specifically, this includes the ways customers receive technical support or account support. This can also include ancillary tools, customer support programs (and the quality thereof), availability of user groups, service-level agreements and so on.

Operations: The ability of the organization to meet its goals and commitments. Factors include the quality of the organizational structure, including skills, experiences, programs, systems and other vehicles that enable the organization to operate effectively and efficiently on an ongoing basis.

Completeness of Vision

Market Understanding: Ability of the vendor to understand buyers' wants and needs and to translate those into products and services. Vendors that show the highest degree of vision listen to and understand buyers' wants and needs, and can shape or enhance those with their added vision.

Marketing Strategy: A clear, differentiated set of messages consistently communicated throughout the organization and externalized through the website, advertising, customer programs and positioning statements.

Sales Strategy: The strategy for selling products that uses the appropriate network of direct and indirect sales, marketing, service, and communication affiliates that extend the scope and depth of market reach, skills, expertise, technologies, services and the customer base.

Offering (Product) Strategy: The vendor's approach to product development and delivery that emphasizes differentiation, functionality, methodology and feature sets as they map to current and future requirements.

Business Model: The soundness and logic of the vendor's underlying business proposition.

Vertical/Industry Strategy: The vendor's strategy to direct resources, skills and offerings to meet the specific needs of individual market segments, including vertical markets.

Innovation: Direct, related, complementary and synergistic layouts of resources, expertise or capital for investment, consolidation, defensive or pre-emptive purposes.

Geographic Strategy: The vendor's strategy to direct resources, skills and offerings to meet the specific needs of geographies outside the "home" or native geography, either directly or through partners, channels and subsidiaries as appropriate for that geography and market.

Learn how Gartner can help you succeed

Become a Client

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

About Careers Newsroom Policies Site Index IT Glossary Gartner Blog Network Contact Send Feedback

Gartner.

 $\ensuremath{\textcircled{\sc 0}}$ 2021 Gartner, Inc. and/or its Affiliates. All Rights Reserved.