Magic Quadrant pour les outils d'intégration de données

17 août 2022 - ID G00758102 - 118

Par **et 2 autres**Ehtisham Zaidi, Sharat Menon,

Le marché des outils d'intégration de données connaît un nouvel élan sous l'impulsion des exigences en matière d'intégration de données multicloud et hybrides et de modèles de conception de structures de données. Les responsables des données et de l'analyse doivent utiliser cette recherche pour évaluer les fournisseurs appropriés pour leurs cas d'utilisation d'intégration de données existants et à venir.

Hypothèses de planification stratégique

- D'ici 2024, les tâches manuelles d'intégration des données seront réduites jusqu'à 50 % grâce à l'adoption de modèles de conception de structure de données qui prennent en charge l'intégration accrue des données.

- D'ici 2024, la gestion et l'intégration des données augmentées basées sur l'IA réduiront jusqu'à 30 % le besoin de spécialistes informatiques.

- D'ici 2025, les outils d'intégration de données qui ne fournissent pas de capacités d'intégration de données hybrides multicloud via un modèle PaaS perdront 50 % de leur part de marché au profit des fournisseurs qui le font.

Définition/description du marché

Gartner définit l'intégration de données comme la discipline comprenant les modèles architecturaux, les méthodologies et les outils qui permettent aux organisations d'obtenir un accès et une livraison cohérents des données sur un large éventail de sources de données et de types de données pour répondre aux exigences de consommation de données des applications métier et des utilisateurs finaux. Les outils d'intégration de données permettent aux organisations d'accéder, d'intégrer, de transformer, de traiter et de déplacer des données couvrant divers points de terminaison et dans n'importe quelle infrastructure pour prendre en charge leurs cas d'utilisation d'intégration de données.

Le marché des outils d'intégration de données comprend des fournisseurs qui offrent un ou plusieurs produits logiciels autonomes pour permettre la construction et la mise en œuvre d'une

infrastructure d'accès et de livraison de données pour divers cas d'utilisation d'intégration de données.

Ceux-ci incluent (mais ne sont pas limités à):

- **Ingénierie** des données : utilisation des capacités des outils d'intégration de données pour concevoir des pipelines de données à l'appui de divers cas d'utilisation analytique tels que l'entrepôt de données, les lacs de données, la science des données et l'apprentissage automatique. Nous ne faisons pas ici référence à l'ingénierie des données en tant que pratique.

- **Intégration des données** cloud : migration et modernisation des charges de travail de données dans le cloud public avec une architecture qui s'étend sur site et un ou plusieurs écosystèmes cloud (hybride/multicloud) pour permettre une utilisation optimale des ressources cloud.

- Intégration des données **opérationnelles** : Prise en charge des cas d'utilisation de l'intégration des données opérationnelles/transactionnelles telles que la gestion des données de référence (MDM), l'acquisition et le partage de données interentreprises, le partage de données B2B, la synchronisation des données liées aux processus métier critiques et la prise en charge des initiatives de gouvernance des données.

- **Fabric** de données : capacités d'intégration de données fournies à l'appui des cas d'utilisation liés à la conception émergente de la structure de données. Cela inclut la possibilité de permettre un accès plus rapide aux données fiables dans des paysages distribués en utilisant des métadonnées actives, la sémantique et les capacités ML.

Les capacités de base sont des exigences fonctionnelles que chaque fournisseur d'outils d'intégration de données évalué doit prendre en charge pour être inclus dans cette recherche.

Il s'agit notamment des éléments suivants :

- **Topologie de déplacement** des données : déplacement uni/bidirectionnel/multidirectionnel des données entre les points de terminaison (par exemple, synchroniser, comparer, diffuser, consolider) via des modes physiques et virtuels, répondant aux exigences de latence par lots/microbatch/temps réel pour l'intégration des données.

- **Virtualisation des données** : exécution de requêtes distribuées sur des sources de données disparates virtuellement intégrées. Cela nécessite des adaptateurs aux sources de données, un référentiel de métadonnées et un moteur de requête distribué qui peut fournir des résultats de différentes manières (par exemple, API, JDBC) pour la consommation en aval.

- **Intégration** des données de flux : traitement des données en mouvement (p. ex., flux, événements) et provisionnement des données en cours d'eau pour la consommation, l'analyse ou le stockage en aval.

- **Services d'API :** les données en tant que service sont activées via les fonctionnalités de conception d'API pour créer et gérer des points de terminaison d'API sortants sur des

ressources de données existantes et gérer la consommation d'API entrantes pour ingérer des données internes et externes.

- **Transformation complexe** des données : fonctionnalités qui facilitent les opérations complexes de traitement des données telles que la correction des valeurs aberrantes, l'analyse sophistiquée (par exemple, exploration de texte de forme libre, journaux de télémétrie, exploration multimédia), la modélisation de données complexes (par exemple, la création automatisée de pipelines de données et la prise en charge de l'automatisation de l'entrepôt de données) et la création de transformations réutilisables.

- **Intégration de données augmentée :** Capacités qui améliorent et optimisent les opérations d'intégration de données (par exemple, les dérives de schéma auto-réparables, la récupération automatique) en utilisant largement les métadonnées (par exemple, les données d'utilisation, les journaux de transactions, les charges de travail système) et des algorithmes ML préemballés qui peuvent informer ou automatiser les tâches d'ingestion, de transformation, de combinaison et de provisionnement des données.

- **Préparation** des données : La pertinence des outils d'intégration de données pour soutenir une gamme de rôles opérationnels (p. ex. intégrateurs citoyens, analystes d'affaires) pour l'intégration de données en libre-service. L'accent est mis sur l'autonomisation du personnel non technique en utilisant diverses techniques telles que le mélange de données à faible / sans code, l'exploration visuelle et la correspondance probabiliste.

- **Portabilité de l'intégration : portabilité** de la conception des flux de données sur l'infrastructure (sur site, SaaS, fournisseur de services cloud, VPC, etc.), offrant des capacités de gestion de la charge de travail dans un environnement d'exécution propre, sûr et portable (par exemple, via la conteneurisation).

Les capacités de prise en charge se composent des fonctionnalités qui prennent en charge les cas d'utilisation prospectifs grâce à des fonctionnalités différenciées.

Il s'agit notamment des éléments suivants :

- **Prise en** charge des métadonnées : Capacités qui prennent en charge l'utilisation intensive des métadonnées (p. ex., métadonnées d'utilisation, journaux de transactions, charges de travail système) pour automatiser ou améliorer les tâches d'intégration des données.

- **Prise en charge de la** gouvernance des données : Capacités qui facilitent les mandats de gouvernance des données (par exemple, qualité des données, traçabilité des données) lors du traitement des données pour répondre à des cas d'utilisation spécifiques d'intégration de données (par exemple, MDM, partage de données)

- Prise en charge de **DataOps :** capacités de gestion des changements pour prendre en charge les données et les artefacts connexes (par exemple, intégration Git des pipelines de données, gestion des modèles de données), automatisation (par exemple, tests automatisés),

orchestration de la livraison de données (par exemple, pipelines CI/CD) avec des niveaux de sécurité appropriés pour améliorer l'utilisation et la valeur des données.

- **Prise en charge FinOps :** fonctionnalités qui permettent aux responsables des données et de l'analyse de contrôler les dépenses de manière itérative, de comprendre les performances des produits et de faire des choix concernant les compromis prix/performances, ce qui se traduit par une allocation optimale des ressources dans le cloud.

Pour qu'un fournisseur soit inclus dans ce marché, ses outils d'intégration de données doivent être en mesure de prendre en charge ces cas d'utilisation et capacités **indépendamment** des offres de produits des autres fournisseurs. Les fournisseurs qui vendent des technologies d'intégration de données dans le cadre d'autres solutions (telles que des plates-formes d'analyse, des SGBD et des applications packagées ou SaaS) ne sont pas considérés comme des fournisseurs d'outils d'intégration de données par Gartner.

Notre évaluation des outils d'intégration de données n'inclut pas les frameworks open source, les plates-formes de développement à usage général ou les interfaces de programmation. Ces frameworks ou plates-formes d'intégration de données qui nécessitent une personnalisation intensive par les développeurs pour les concevoir pour des scénarios d'intégration de données spécifiques sont exclus de ce Magic Quadrant.

Les fournisseurs évalués dans ce Magic Quadrant proposent au moins un outil commercial prêt à l'emploi spécialement conçu pour prendre en charge tous les cas d'utilisation d'intégration de données répertoriés dans ce rapport.

Magic Quadrant

Figure 1 : Magic Quadrant pour les outils d'intégration de données

CHALLENGERS | LEADERS

Informatica
Oracle
IBM
Microsoft
Qlik
SAP
Denodo
Talend
TIBCO Software
Precisely
SAS

Amazon Web Services
Fivetran
Matillion
Palantir
Safe Software
Software AG (StreamSets)
SnapLogic
Hitachi Vantara
CloverDX
K2View

ABILITY TO EXECUTE

NICHE PLAYERS | VISIONARIES

COMPLETENESS OF VISION →   As of August 2022   © Gartner, Inc

Gartner

Source : Gartner (août 2022)

Forces et mises en garde des fournisseurs

Amazon Web Services

Amazon Web Services (AWS) est un acteur de niche dans ce Magic Quadrant ; c'est un nouvel entrant cette année. Son siège social est situé à Seattle, dans l'État de Washington. Il propose AWS Glue comme outil d'intégration de données. Ses clients utilisent AWS Glue lorsque leurs magasins de données cibles principaux résident sur le cloud AWS. Gartner estime que des milliers de clients utilisent les tâches d'intégration de données et le service de catalogue AWS Glue. Ses opérations sont mondiales et couvrent divers secteurs.

Forces

- **L'intégration native réduit la complexité et améliore l'interopérabilité au sein** de l'écosystème de données AWS : l'intégration native d'AWS Glue avec d'autres services cloud AWS (tels que S3, Redshift, Athena et Lake Formation) en fait le choix d'outil d'intégration de données le plus approprié lors de la sélection d'un écosystème de données natif AWS.

- **Innovation pour répondre aux demandes de livraison** de données : AWS Glue obtient de bons résultats en matière d'innovation pour son service d'intégration de données sans serveur qui va

au-delà de la mise à l'échelle de l'infrastructure pour le traitement des données. Il prend en charge le nettoyage des données basé sur le ML et la détection des PII à grande échelle. Il étend même le traitement des données pour d'autres applications (telles que les ordinateurs portables et les IDE) via son API de sessions interactives.

- **Centralise** les métadonnées **dans son catalogue** de données : AWS Glue offre un ensemble complet de fonctionnalités de métadonnées telles que le catalogage, la recherche, le lignage, les contrôles de qualité des données, la détection des informations personnelles et la journalisation de l'accès aux données pour l'auditabilité. Par exemple, les métadonnées AWS Glue sont accessibles dans Athena pour explorer les données sous-jacentes dans S3.

Précautions
- **Connecteurs limités au-delà des services** natifs AWS : AWS Glue propose une large gamme de connecteurs natifs de la place de marché AWS pour les clouds sur site et autres. Toutefois, pour certaines sources non-AWS (telles que Mainframes, SAP, Azure ou GCP), les clients auront besoin d'options tierces, dont certaines sont disponibles sur AWS Marketplace (comme CData, tCVision et Qlik) pour compenser ce déficit.

- **Courbe d'apprentissage abrupte** : AWS Glue n'est pas convivial pour les débutants : selon les évaluations de Gartner Peer Insights, sa documentation produit et ses bibliothèques d'exemples de code ne sont pas suffisantes pour les tâches complexes d'ingénierie des données. Il existe une forte dépendance à l'expertise technique pour le codage (Python ou Scala) et le débogage, et la mise en œuvre d'un cadre d'intégration d'entreprise nécessite une expertise dans les architectures sans serveur. En plus des récentes améliorations apportées aux produits, AWS Glue a ajouté des outils sans code tels que Glue Studio et Glue Databrew pour atténuer ces défis.

- Les **praticiens signalent** plusieurs problèmes opérationnels : Les évaluations Gartner Peer Insights mettent en évidence plusieurs problèmes opérationnels tels que des temps de démarrage élevés pour la rotation du cluster et les nouvelles tâches, l'absence de paramètres de délai d'inactivité entraînant des coûts élevés et des goulots d'étranglement occasionnels des performances pendant les heures de pointe du traitement des données. Cependant, la dernière version d'AWS Glue 3.0 devrait atténuer la plupart de ces problèmes.

CloverDX

CloverDX is a Niche Player in this Magic Quadrant, the same as last year. It is headquartered in Prague, Czech Republic, and offers the CloverDX Data Management Platform, composed of CloverDX Designer and CloverDX Server. The customer base includes more than 450 organizations, mainly located in North America and Europe. Clients are commonly in the following sectors: finance and banking, software and SaaS providers, advisory and consulting, and the public sector.

Strengths
- **Broad support for complex data types and data integration scenarios**: CloverDX supports a wide variety of data types and also has robust data transformation capabilities. It supports CSV, XML, JSON and Avro data types and beyond, offering connectivity to relational datastores.

It also integrates with graph and multimodel databases such as Neo4j and CosmosDB (among several others). CloverDX has also incorporated built-in support for Kafka.

- **Simple pricing model:** Although CloverDX does not have particularly strong support for financial governance and FinOps (it does not offer ways to automatically scale deployments), the platform has a pricing model that is easy to understand. Pricing is based on user seats and server cores, so costs are transparent and predictable. This is in contrast to the consumption-based pricing models used by other vendors, which can result in unplanned or unpredictable costs.

- **Customer experience:** Clients praise CloverDX's high-touch customer support as well as its remote and self-paced training options and personalized professional services. They report that they are able to quickly find value from the product. The company has a customer advisory board that helps drive the product development roadmap, demonstrating that it is attuned to user needs.

Cautions

- **Metadata, data fabric and automation capabilities are still developing:** As a company, CloverDX understands the common complexities of the data landscape and recognizes the need to offer automation for data integration. Although the CloverDX Data Management Platform does provide automated capabilities, fully executing on this vision is a work in progress. As an example, CloverDX does not have strong support for the continuous analysis of metadata (also referred to as "active metadata"), which is important for setting the foundation for a data fabric and for additional automation.

- **Limited support for diverse data integration styles:** CloverDX does not directly support stream data integration, such as modifying data in-stream for the next consumer, although this work can usually be achieved by microbatching. CloverDX also does not support data virtualization or change data capture.

- **Limited third-party partnerships:** CloverDX is working to expand its network of integration partners (with system integrators and OEMs, for example), but currently lags behind its larger competitors in terms of number and breadth of partnerships. CloverDX needs to improve integration with adjacent data management infrastructure providers such as cloud DBMSs, external and third-party metadata management tools, data governance technology providers, and analytics and data science providers.

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 Denodo Platform on-premises and on public cloud (AWS, Azure and GCP). Its operations are geographically diverse, with more than 1,000 customers primarily in the financial services, manufacturing and technology sectors.

Strengths

- **Focus on distributed data architectures:** Denodo begins with virtualization across disparate sources and extends its capabilities through data science notebooks. It can write out its

caching tier to persistent stores. It is suitable for logical data warehouse and data fabric use cases through building business-friendly semantic models.

- **Sophisticated optimizer:** Denodo uses statistics to evaluate current operations of query patterns and then uses ML-enabled DataOps to enhance performance in terms of faster response and smaller footprint in terms of resource allocation.

- **"Try before you buy" engagement model:** Denodo's customers appreciate its presales support and proof-of-concept activities. About 80% of its paying customers have tried Denodo Express (its free product, which is capped at a certain capacity for a single user with standard features) at some stage in their sales cycle. Its pricing model is based on the number of cores driven by customer demand, varying from small departmental to enterprise deployments.

Cautions

- **Limited support for traditional bulk/batch workloads:** Denodo does not support change data capture within its current data integration offering. It is therefore not suited for traditional batch-related extract, transform and load (ETL) operations, given its focus on federation and distributed query processing.

- **Multilocation deployments require manual configuration:** Denodo supports hybrid and intercloud data integration use cases. However, connectivity across Denodo instances running in different geographies involves a lot of manual configurations (particularly in Denodo version 7.x) and ongoing operations support to ensure effectiveness of the multilocation deployment. Denodo claims that it has addressed this gap in its latest version (8.x), which requires customers to upgrade.

- **Data security configuration can be challenging:** Denodo practitioners report challenges around the process of secure authentication configuration, SSL connections setup in cloud environments and frequent timeouts. A few customers have complained of delays in the platform readiness because some configuration settings were difficult to manage when done through scripting instead of using the Denodo UI.

Fivetran

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 acquired HVR in October 2021. It offers the following data integration tools: Fivetran and HVR. It has more than 4,000 customers for these products. Its operations are primarily focused in North America and EMEA, with a growing presence in APAC. The top three industries it supports are in the software services, information technology and media sectors.

Strengths

- **Market momentum and partnerships:** Fivetran grew its customer base significantly last year, primarily targeting departmental leaders with its appeal of easy, automated extract and load data integration capabilities. It has established partnerships with major cloud service providers (CSPs), independent software vendors (ISVs) and system integrators (SIs) to further strengthen its market momentum.

- **Product strategy focused on data movement and cloud:** Fivetran focuses on physical data movement into cloud targets, making it suitable for use cases like centralizing data in cloud data stores, database replication and cloud migration. It has made several improvements to its integration infrastructure compliance and secure data movement such as secure data encryption, expanded RBAC and data residency controls, and private links (with AWS and Azure).

- **Low total cost of ownership (TCO) with quick data delivery:** Fivetran applies a consumption-based pricing model to the amount of monthly active rows processed, allowing customers to get started with their projects without the need to secure significant upfront capital expenditure. It also reduces data pipeline development times and minimizes operations. It can be set up to handle source schema changes automatically in the cloud target data store by managing schema drifts and guaranteeing reliable data replication. As a result, its customers spend less time on data operations tasks.

Cautions

- **Limited product coverage:** Fivetran offers low-latency, managed data ingestion service to the cloud. However, traditional batch-related ETL operations, data virtualization and message-oriented middleware are not supported. As a result, enterprise customers need to supplement Fivetran's data ingestion capabilities with a third-party technology if they require complex data transformation and orchestration support. Fivetran integrates with dbt Core to enable data transformation for its customers.

- **Selective metadata and governance support:** Fivetran lacks the ability for metadata analysis like discerning relationships between data assets, profiling and dynamically alerting discrepancies. It is beginning to open up its metadata via API to third-party catalogs. It also lacks essential data governance capabilities in support of data quality and enforcement of compliance rules like data masking on integrated data. Its roadmap item "governance for inflight data" is likely to introduce automatic PII detection and contextual tagging.

- **Customer support and operational issues:** The poor user interface inherited from the HVR acquisition, combined with challenges around locating suitable knowledge base documentation often forces Fivetran's HVR customers to seek product support. This challenge has created longer turnaround times for some customers. The latest release of HVR 6 with updated documentation and recent advances with Fivetran's support capabilities are expected to alleviate these challenges.

Hitachi Vantara

Hitachi Vantara has reentered the Magic Quadrant this year after its absence last year and is positioned as a Niche Player. It is based in Santa Clara, California, and offers Lumada DataOps platform, which includes Data Integration & Analytics powered by Pentaho and Data Catalog. It also offers data integration capabilities within Lumada Industrial DataOps, its industrial IoT (IIoT) platform. The vendor's operations are geographically diverse with a customer base of around 800 organizations. The top three industries it supports are financial services, software and technology, and retail and consumer products.

Strengths

- **Improved data management vision:** Lumada DataOps provides data integration capabilities with support for adjacent data management disciplines, such as data governance, metadata management and data quality, through the acquisitions of Waterline Data and Io-Tahoe. Hitachi Vantara has enabled integration across these offerings within Lumada DataOps.

- **Differentiation in IoT and edge data integration scenarios**: Due to Hitachi's balanced focus on supporting both IT and OT use cases, Hitachi can differentiate in scenarios that require integrating data from IoT devices at the edge and merging that with data from traditional sources such as databases and file systems.

- **Support for augmented data integration in specific areas:** Hitachi provides support for optimizing data persistence across multiple clouds by analyzing factors such as cost, location, governance and latency. Automation is also used for intelligent data tiering on HDFS files, dynamic selection of execution runtimes, and discovering relationships between datasets. Although there's still a lot of room for improvement here, this is a good start.

Cautions

- **Market traction reliant on Pentaho:** The low number of mentions for Hitachi Vantara among Gartner's client interactions and product reviews is indicative of slowing market momentum in data integration use cases. Customers and prospects (on Gartner inquiry calls) generally equate Hitachi Vantara to just Pentaho Data Integration, being unaware that this has been included in Lumada DataOps, providing many additional data integration capabilities, such as data virtualization via Data Services. Enterprises that do recognize capabilities beyond Pentaho believe these to be mainly relevant to IIoT use cases.

- **Nascent support for differentiating components of the data fabric:** Although Hitachi Vantara has a vision for helping customers optimize their data fabric through its data integration tools and data catalogs, support for certain differentiating components of the data fabric, such as semantic modeling and active metadata support, is at a nascent stage. Hitachi is building upon existing capabilities here, such as AI-based semantic tagging of data.

- **Limitations in cloud-native deployments:** Hitachi's data integration capabilities are not yet available as a fully managed iPaaS across public cloud infrastructure providers such as AWS and Azure. Containerization support for Pentaho was added recently, through which deployments on the public cloud of choice are now possible.

IBM

IBM is a Leader in this Magic Quadrant; in the previous iteration of this research, it was also a Leader. IBM is headquartered in Armonk, New York. IBM Cloud Pak for Data (which includes DataStage Enterprise Plus Cartridge), IBM Cloud Pak for Integration (for application integration scenarios), Cloud Pak for Data as a Service (which includes DataStage as a Service, Watson Query as a Service and Watson Knowledge Catalog as a Service) and IBM Data Replication, all target a range of data integration use-case scenarios. The vendor's customer base for this

product set is more than 10,000 organizations. Its operations are global, and its clients tend to be enterprise B2B and B2C organizations in the banking and financial services, insurance, healthcare and pharmaceuticals industries.

Strengths

- **Support for the data fabric design**: IBM software has collaborated with IBM research to embed capabilities for augmented data integration into its Cloud Pak for Data (CPD) platform and services. The incorporation of capabilities to capture and activate metadata in Watson Knowledge Catalog, the ability to support DataOps patterns for improved orchestration and agility, and the utilization of knowledge graphs to support semantic modeling and taxonomy to ontology mapping for unstructured content have further improved its support for data fabric use cases.

- **Comprehensive portfolio for operational and analytical use-case support**: IBM has a comprehensive tools portfolio within CPD that includes DataStage (for bulk/batch integration), IBM Cloud Pak for Integration (for application integration and API management), Watson Query (for data virtualization), IBM Data Replication (for data replication and synchronization) and IBM Streams (for stream data integration scenarios). Along with these capabilities, IBM CPD is well-integrated with other data management technologies including data quality, MDM and data governance.

- **Modular architecture and DataOps enablement**: IBM's data integration tools are delivered as tightly integrated and yet loosely coupled services on Red Hat OpenShift (a Kubernetes-based platform). Clients praise IBM's remote runtime capabilities, which reduce egress costs by allowing developers to build pipelines once and push down workloads to the execution environments of their choice. IBM's support for CI/CD and integration with Git (for versioning), Jenkins (for task scheduling) and other third-party task and workflow managers is highly rated.

Cautions

- **Data replication product maturity**: IBM delivers change data capture (CDC) technology through currently supported IBM Data Replication packages. However, some of its customers are still on its legacy CDC tooling. Reference customers of IBM have reported a below-par UI for this product set. Some customers have highlighted challenges with workload monitoring, performance optimization and high availability. IBM should ramp up its migration to the IBM Data Replication services on IBM Cloud Pak for Data (from legacy IBM CDC offerings) and improve its capabilities for stream data integration support to ensure that existing clients don't have to evaluate third-party data replication solutions.

- **Lack of clarity on migration and upgrades**: Although IBM is aggressively trying to migrate legacy IBM Infosphere Information Server customers to IBM Cloud Pak for Data, some customers report lack of clarity around best practices and license portability, and cite the lack of a well-structured migration and upgrade roadmap. Although IBM does provide embedded migration tooling to help assist with migration to CPD, customers have little knowledge of this capability.

- **Market perception and revenue growth deceleration**: Some IBM prospects (on Gartner's client inquiry service) continue to have a perception that IBM's tools are expensive, complex to implement and targeted toward large organizations that possess strong data engineering skill sets. Gartner's analysis of our proposal review data suggests that IBM's tools were rarely evaluated or considered by SMEs unless they were existing IBM customers. IBM's positioning of Cloud Pak for Data as a modular set of services (rather than an end-to-end platform), and pivoting to a serverless, metered licensing model, should alleviate some of these concerns.

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 several on-premises data integration products including PowerCenter, PowerExchange, Data Engineering Integration, Enterprise Data Preparation and Data Engineering Streaming. It also provides various data integration services as part of its Intelligent Data Management Cloud (IDMC) platform including Cloud Data Integration, Cloud Data Integration Elastic, Cloud Mass Ingestion, Cloud Integration Hub and Cloud B2B Gateway. Informatica has more than 5,700 customers for these product lines. Its operations are geographically diversified, and its clients are distributed across multiple industries, with the top being in the financial services, healthcare and public sectors.

Strengths
- **Augmented data integration support:** One of the main reasons highlighted by customers to evaluate and select Informatica's data integration tools is its automation support for complex and repetitive data integration tasks. Informatica has significantly invested in CLAIRE, its active metadata-driven ML engine, which performs continuous analysis over all collected metadata to significantly automate schema drift, data pipeline orchestration, performance monitoring, and optimization and data modeling.

- **Improved pricing and licensing aligned to cloud adoption:** Informatica has pivoted to a less-complex, consumption-driven licensing model based on Informatica Processing Units (IPUs). This common unit of capacity can be used across its entire set of cloud services offered under the IDMC umbrella. Users can subscribe to a certain number of IPUs (based on forecast usage), which can then be used interchangeably across all its major data integration product lines. Early adopters of this new pricing model report ease of understanding, adoption and scaling.

- **Operational data integration use-case delivery:** Customers looking for a technical implementation of the data hub architecture to support operational data integration use cases report Informatica's solutions as mature. The Cloud Integration Hub offering is frequently evaluated and selected by customers for its ability to support all data modalities (including batch, virtual, streaming and API-based integration) and for its ability to integrate, govern and share across a multicloud/hybrid environment for data and application integration scenarios.

Cautions

- **Reported challenges with PowerCenter to Informatica Cloud migration**: Some customers report challenges while migrating from PowerCenter to Informatica Cloud. Informatica provides a migration utility tool (that automated a certain percentage of manual mapping conversion tasks) at an additional cost. Some customers reported that their developers had to invest time so that they could efficiently utilize the upgrades (and certain new features) of Informatica Cloud to optimize existing integration workloads in the new cloud environment. Some early migration customers reported the need for manual workarounds and postmigration testing. Informatica has made investments in this area by training and certifying various global system integration partners including Accenture, KPMG and Deloitte that can assist Informatica customers in their migration projects.

- **Go-to-market overly focused on end-to-end data management scenarios**: Informatica delivers a broad portfolio of tools to support a broad range of data integration scenarios. Although this is well-appreciated by large enterprise clients, SMEs and business departments looking to start their journeys with basic data ingestion pipelines often find Informatica's product set and messaging (focused as it is on augmented integration, CLAIRE and data fabric) as overwhelming and complex to navigate. Informatica should balance end-to-end best-of-breed functionality with targeted best-fit engineering tools to support "land and expand" strategies. To address some of these challenges, Informatica recently launched a free Data Loader service to support simplified workflows for building data pipelines.

- **Call for DataOps-related enhancements**: Some customers stated that they were unaware of how Informatica's data integration tools integrate and interoperate with popular third-party or open-source orchestration and task workflow management tools like dbt, Apache Airflow, Luigi, Prefect and Dagster. Data engineers appreciate the low-code integration support of Informatica, but stated that they were not aware of its extensibility features to high-code repositories for certain use cases that require coding. Some customers also called out the need for improvements in change management, versioning and CI/CD capabilities within Informatica's tool portfolio.

K2View

K2View is a new entry to this Magic Quadrant and is positioned as a Niche Player. It is headquartered in Yokneam, Israel, and offers the K2View Data Product Platform as its data integration tool. The customer base includes more than 50 organizations, mainly in North America and Europe. Clients are commonly in large enterprises in the communications and media, financial services and healthcare verticals.

Strengths

- **Business focus and usability:** K2View is clearly focused on addressing business challenges through its data integration technology. As opposed to other products, which may be targeted toward technical personas, K2View's strategy and go-to-market aligns with common business outcomes and use cases, such Customer 360, legacy data migration, test data management, and operational intelligence for real-time insights (such as churn prediction, credit scoring or fraud detection).

- **Unique "Micro-Database" architecture:** K2View supports a distributed and scalable architecture where the dataset for each business entity is managed in its own Micro-Database. This allows K2View to concurrently manage billions of Micro-Databases with near-real-time data movement, including the ability to apply in-flight data transformations.

- **Flexible data virtualization capabilities:** K2View offers a capability called "dynamic data virtualization," in which some data can be accessed from the source via virtualization and other data can be accessed from a Micro-Database, which is asynchronously synced with the sources. This ability to decide which data will be stored physically, and which will be virtualized, provides advantages in terms of performance optimization and load management on source systems.

Cautions
- **Low market visibility and mind share:** K2View is not as well-known as some incumbent vendors in this market. While not necessarily a drawback, some Gartner clients evaluating data integration tool vendors were slightly hesitant to evaluate K2View due to its relatively low market visibility and mind share. K2View was mentioned infrequently by Gartner clients (on our client inquiry service) especially when those clients were evaluating data integration tools for analytical use cases.

- **Augmented capabilities are limited:** K2View has some augmented capabilities, but these are still developing. For example, K2View is not currently able to use AI/ML algorithms to automate or augment sophisticated data transformation, and it cannot adapt existing data transformations in a proactive, nonhuman-initiated fashion.

- **Pricing is not fully transparent:** K2View uses multiple pricing models, including traditional pricing based on the number of data sources, contract length and level of support. The company also has a consumption-based subscription pricing model, which is based on data reads/writes and data storage. Although these pricing models are also used by other vendors, a lack of detail on the vendor's website means that it is not easy for potential customers to evaluate the pricing structure at a glance and forecast ongoing spend and TCO with K2View as they scale their usage.

Matillion

Matillion is a Niche Player in this Magic Quadrant, the same as last year. It has dual headquarters in Denver, Colorado, and Manchester, U.K., and offers Matillion ETL and Matillion Data Loader as its data integration tools. The customer base includes more than 1,150 organizations, mainly in North America and Europe. Clients are commonly in the business services, software and manufacturing sectors.

Strengths
- **Focus on cloud data engineering:** Matillion offers two products: Matillion ETL is used for building data pipelines, transforming data and preparing data for consumption by analytics tools. Matillion Data Loader focuses on extracting data from source systems and loading it into

cloud data platforms. Matillion uses an ELT approach, leveraging SQL-based push down transformations and native optimizations in cloud data platforms.

- **Strong market momentum, recognition and third-party partnerships**: Matillion has deep partnerships with popular data management vendors and system integrators, and was named Snowflake Technology Partner of the Year for Data Integration in 2021. Matillion has a strong integration with Snowflake for out-of-the-box modeling into Snowflake's data warehouse environment. Matillion also partners with Collibra for data governance, and it conducts co-selling and provides integrations with multiple other data and analytics companies, such as Databricks and ThoughtSpot.

- **Deployment options address common data sovereignty issues**: Matillion is deployed as a virtual machine image in the customer's cloud data warehouse, so the client company has full control over the data, which never leaves the customer's environment. This allows Matillion to address data sovereignty issues and privacy concerns, and means that the tool can be used by government agencies, such as in the AWS GovCloud environment.

Cautions

- **Limited support for metadata and data governance**: Matillion does not have full support for all metadata management capabilities, such as semantic modeling, taxonomy and ontology mapping, knowledge graphs, impact analysis, and active metadata support for automation. As another example, Matillion partners with other metadata management and active metadata vendors, such as Collibra, for data lineage, and does not currently offer these features natively.

- **Limited data observability and data virtualization support**: Matillion currently does not offer mature capabilities related to data observability, such as data pipeline monitoring and ML-based recommendations for optimizing data integration workload performance. These features are on the company's product development roadmap. For data virtualization, Matillion supports basic virtual data access via external tables, but lacks advanced data virtualization capabilities. For example, support for external query acceleration engines like Impala or Presto and dynamic query optimization is currently lacking.

- **Customers report difficulty with upgrades**: Although Matillion does a good job of introducing new features, and has regular product releases, some customers feel that the company's testing of new features and communication about product changes could be improved.

Microsoft

Microsoft is a Leader in this Magic Quadrant. In theprevious iteration of this research, it was also a Leader. It is headquartered in Redmond, Washington. It offers SQL Server Integration Services (SSIS) for on-premises data integration tasks, Azure Data Factory (ADF) for hybrid data integration tasks, and Power Query for data preparation tasks. Its operations are geographically diverse, and its clients range from small and midsize businesses to large enterprises. Gartner estimates a customer base of about 20,000 companies.

Strengths

- **Native integration reduces complexity and improves interoperability within the Azure data ecosystem:** ADF's native integration with other Azure cloud services (like ADLS, Microsoft Purview and Azure Synapse Analytics) makes it the most suitable data integration tool choice when selecting an Azure-native data ecosystem. For example, lineage metadata from ADF can be accessed within Purview along with business metadata.

- **Product strategy focused on both enterprise and departmental use cases:** Microsoft positions ADF and SSIS for technical roles with its data engineering capabilities, which include curating trusted data assets and managing the full life cycle of data pipelines. It positions Power Query for citizen roles with its core data preparation capabilities and ease of use to simplify data access.

- **Go-to-market partnerships:** Microsoft leverages several partnership channels with independent software vendors (ISVs) operating on Azure cloud (such as Snowflake and Databricks for dbPaaS), and also with global system integrators (SIs). It offers several jump-start data integration projects with demo code libraries and tutorials. It also operates a lift-and-shift migration program of SSIS workloads to ADF toward retaining legacy customers during their data modernization efforts.

Cautions

- **Integration complexities in hybrid architectures:** Microsoft's portfolio of cloud and on-premises data integration tools is good enough to support various deployment options such as public cloud, private cloud and on-premises. However, it leaves users to address the decisions around data placement (where they will manage specific datasets) and workload placement (where they will run diverse data processing workloads).

- **Cloud workload performance and spend prediction:** Although 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 that are not well-acquainted with the Azure ecosystem can find cost planning difficult when determining scenarios and budget projections. Some report the need to add multiple nodes to speed up data processing, which is the least economical route.

- **Error reporting and debugging is sometimes inconsistent:** Several Gartner Peer Insights reviews highlight the lack of clarity on error messages and how troubleshooting becomes a challenging task. In addition to product improvements, Microsoft should also provide DataOps best practice guidance such as code templates and libraries, video tutorials and partner accelerators library.

Oracle

Oracle is a Leader in this Magic Quadrant, the same as last year. Based in Austin, Texas, Oracle offers the Oracle GoldenGate (GG) platform, Oracle Data Integration Suite (ODI), Oracle Big Data SQL (BDSQL), data integration services within Oracle Integration Cloud and Oracle Cloud Infrastructure (OCI) Data Integration Services. Oracle's customer base for these products is more

than 16,000 organizations. Its operations are geographically diversified, and its clients are primarily in the financial services, telecommunications and retail/e-commerce sectors.

Strengths

- **Technical superiority in data replication:** Oracle GoldenGate stands out in the data replication/change data capture space through its multithreaded log capture API, edge deployments and microservices-based architecture. GoldenGate is often deployed to support mission-critical applications that need to always be highly available.

- **Balanced focus across OCI and multicloud deployments:** Although OCI Data Integration is a strong part of the OCI ecosystem, Oracle continues to deliver enhancements for its stand-alone data integration products for multicloud deployments. Native integration for Snowflake with Oracle GoldenGate (with an option for external staging in other public cloud object stores) is an example of this multicloud focus.

- **Clear vision for integration convergence:** Oracle's vision of combining data integration with application integration by designing reference architectures utilizing Oracle data integration products alongside Oracle Integration Cloud sets it up for long-term differentiation.

Cautions

- **Relatively limited adoption for data virtualization:** Oracle isn't evaluated for data virtualization as frequently as other stand-alone data virtualization product vendors. Due to relatively limited adoption for Big Data SQL as compared to other Oracle products for data integration, some customers find the current user community around the product to be too small. Oracle's recent push to enable virtual data access across multiple clouds should accelerate adoption going forward.

- **Perception of alignment to complex use cases only:** Due to Oracle's track record of supporting highly complex transactional workloads, customers often do not consider Oracle for integration scenarios requiring a "good enough" solution, such as low-code data ingestion into cloud data warehouses. As of now, end-user organizations remain largely unaware of the low-code capability and augmented user experiences within Oracle's products that could support Oracle's penetration into the SMB market.

- **Improvements required in troubleshooting and monitoring:** Some customers have expressed dissatisfaction with the level of in-product assistance provided for troubleshooting errors. Some have also asked for improvements in monitoring capabilities to detect errors more quickly.

Palantir

Palantir is a new entrant to this Magic Quadrant and is positioned as a Visionary. It is headquartered in Denver, Colorado and offers the Palantir Foundry Data Integration Suite as its data integration platform. The customer base includes 237 organizations, mainly in North America and Europe. Clients are commonly in the following sectors: government, healthcare and pharmaceuticals, law enforcement, aerospace, and oil and gas.

- **Highly skilled embedded engineers that can build and maintain complex data applications:** Palantir's clients are mainly large organizations that use the company's forward deployed engineers to build, deploy and maintain end-to-end analytic applications. Reference customers praise the deep support provided when building ontologically aligned and taxonomically defined data services, which often require complex data transformation, data orchestration and semantic modeling.

- **Data fabric vision:** One of Foundry's biggest differentiators is its integration of a fully featured semantic layer — the Foundry Ontology — within the same stack of capabilities as connectivity, data transformation, metadata management and data security. Integrated data can be mapped to a graph-oriented semantic layer, which can synthesize multimodal attributes. Foundry Ontology, once developed, supports the representation of future integrated data as knowledge graphs. Additionally, the Foundry semantic engine can extend the Foundry Ontology with its Ontology Gateway — an API layer supporting access to all elements in ontologies and associated pipelines.

- **Extensive support for data security, privacy and governance:** Palantir's clientele includes some of the world's largest defense agencies, state and central governments, and law enforcement agencies. The platform preserves the security states (including role-based, classification-based, and purpose-based models) across all data and model versions. It enables branching of all data and code in a secure sandbox; provides security-aware search, indexing, and cataloging; and extends through security APIs that can be used to integrate with existing workflows and third-party security systems.

Cautions

- **Lack of knowledge transfer and skills availability:** Palantir customers in Gartner client interactions emphasize reliance on forward-deployed engineers to maintain their data applications. Although most customers are satisfied with the quality of this arrangement, they state challenges with knowledge transfer that would reduce their reliance on Palantir support and specify that sourcing Palantir skills from the market has been challenging. Palantir's incorporation of a knowledge transfer program and a first wave of joint forays into the market with consulting partners should alleviate some of these concerns.

- **High-cost, high-value perception:** Palantir targets large organizations that require support for complex data management and analytics use cases. Also, Palantir's pricing model — in which it charges organizations on "configuration of and access to a scoped use case" — reflects the development of end-to-end data integration use cases. As a result, small to midsize organizations have a perception of Palantir as being too comprehensive (and expensive) for their basic data integration requirements. Palantir must demonstrate a more modular approach to better position itself for SMEs and midsize organizations.

- **Extensibility and integration with other data management ISVs:** Some Palantir customers report that the vendor has a tendency to engage in scope expansion that targets the most

difficult integration use cases by including mundane data integration tasks that could be managed with lower-cost tools. Even though Palantir has recently entered into partnerships with IBM, Microsoft and Google (among others), it must expand on its integration with other incumbent data management vendors to showcase better interoperability with the organizations' existing data management infrastructure.

Precisely

Precisely is a Challenger in this Magic Quadrant, the same as last year. Based in Burlington, Massachusetts, Precisely offers Connect, Ironstream and Spectrum as its data integration products. The vendor's customer base for this product set is more than 2,500 organizations. Its operations are geographically diversified, and its clients are primarily in the financial services, insurance and healthcare sectors.

Strengths

- **Differentiating capabilities for customers' data modernization initiatives**: Many customers use Precisely for moving data from legacy systems (such as IBM iSeries, IBM zSeries and Hadoop-based data lakes) into cloud data warehouses, lakehouses and message queues as part of their data modernization initiatives. Additionally, Precisely's ability to observe, mask and format different mainframe data types is sought after by its customer base, as most other vendors no longer focus on mainframes.

- **Ease of configuration and development**: Precisely's intuitive GUI, which facilitates easy and quick development, configuration and bug fixing, is highlighted as praiseworthy by customers.

- **Strong vision for enabling an open data ecosystem**: Precisely provides products for data integration, data quality and enrichment, master data management, and data governance, as part of its Data Integrity Suite. This is a modular data management platform that can run on open frameworks like Spark and interoperate with third-party data management solutions such as CI/CD tools and data lineage solutions.

Cautions

- **Deployment options are limited**: Precisely is neither available as an iPaaS for cloud-native deployments nor deployable on edge devices to support a distributed data and analytics architecture. Deployment on cloud infrastructure (and the availability of Connect ETL's batch capabilities) as Docker containers mitigates this challenge to a degree. Also, a SaaS-based data integration module within the Data Integrity Suite is on the roadmap.

- **Diminished value for specific data delivery styles**: Although Precisely differentiates itself with a comprehensive data management portfolio, its value diminishes when customers are looking for leading products for specific data delivery styles. For instance, automatic schema drift handling and agent-based architecture for data replication are roadmap items. Precisely is evaluated a lot more frequently when the source for replication is a mainframe system than when it's not. Precisely Spectrum Quality is rarely evaluated as a stand-alone data virtualization product.

- **Lack of alignment to the data fabric vision:** Precisely's limited augmented data integration and active metadata management capabilities continue to hamper its support for data fabric use cases.

Qlik

Qlik is a Challenger in this Magic Quadrant, the same as last year. Based in King of Prussia, Pennsylvania, Qlik targets a range of data integration use cases through its Qlik Replicate, Qlik Compose, Qlik Enterprise Manager, Qlik Cloud Services, Qlik Application Automation and Qlik Catalog products. 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.

Note: In October 2020, Qlik acquired Blendr.io, a provider of iPaaS technology for data and application integration. In August 2021, it also acquired Nodegraph, a metadata management startup focused on interactive data lineage, impact analysis and data governance.

Strengths
- **Data engineering use cases for data warehouse automation:** Reference customers of Qlik praise its real-time data replication and complex data transformation in support of data warehouse automation and data lake enablement scenarios. Native connections to legacy databases, file systems, object stores and Kafka (for stream data integration scenarios) were highly rated by Qlik's clients. The maturity of Qlik Compose for rapid development of data models (through automated modeling, DDL and ETL generations) was also cited as a key factor by Qlik's clients for evaluating and selecting Qlik.

- **Key acquisitions improve key capabilities:** Qlik's recent acquisition of Nodegraph has significantly improved the data lineage capabilities of its data catalog service. The acquisition of Blendr.io (which was renamed Qlik Application Automation) extends its application integration, reverse-ETL and API integration. Finally, the acquisition of Big Squid (now called Qlik AutoML), has enhanced its augmented data integration capabilities. Qlik AutoML also augments data preparation tasks within Qlik Data Catalog and provides self-healing schema evolution within Qlik Compose.

- **Self-service data integration support:** Qlik's tooling supports data integration specialists, as well as less-technical personnel in order to make data available through real-time data streaming. The role-based interface of Qlik Compose allows citizen integrators to develop data marts in their data warehouses. On the other hand, the design and administrative tools within Qlik Compose are used by enterprise architects and data engineers when more-robust data models are required.

Cautions
- **Lack of focus on operational data integration:** Although Qlik's focus on data warehouse automation and data engineering continues to gather momentum, its support for operational data integration use cases needs improvement. Several Gartner clients have stated that Qlik

should focus on enabling data-hub-based architectures for efficient integration, governance and sharing for complex application integration scenarios.

- **Challenges with performance optimization and synchronization of replication jobs**: Although Qlik Replicate's heterogeneous compare and repair has improved, some customers continued to cite challenges with setting up and configuring replication jobs. They have stated that the setup is cumbersome and that it took them significant time, effort and intervention from Qlik support agents to identify, fix and resynchronize replication workloads, which lead to downtime. Qlik should improve its monitoring, task observability and performance tuning.

- **Limited support for data virtualization use cases**: Qlik's ability to support data federation and virtualization is currently limited to registering datasets in Qlik Catalog and delivering derived datasets via APIs solely for analytics consumption. Customers looking for advanced data virtualization — including creating virtual data hubs and marts, dynamic query optimization, support for external analytics query acceleration tools and MPP engines — will need to evaluate third-party data virtualization tools.

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 the FME Enterprise Integration Platform, which is composed of FME Desktop, FME Server, FME Cloud and FME Mobile. The customer base includes more than 9,300 organizations, mainly in Europe and North America. Clients are commonly in the government, utilities, energy, architectural engineering and construction, and telecommunications sectors.

Strengths
- **Spatial data integration**: Safe Software provides strong support for spatial data integration, spatial data formats, spatial data transformations and processing data from IoT "things" such as internet-connected temperature sensors. Although spatial data is clearly Safe Software's specialty, the company's support for nonspatial data types and use cases is also robust. As such, this platform will work for a variety of organizations and data integration scenarios. Safe Software supports geospatial data integration with a range of sources and targets, including DbPaaS vendors such as Snowflake and Amazon Redshift.

- **Execution and support**: Safe Software has clearly packaged deployment options and pricing/licensing models that are targeted toward certain industry verticals. As an example, Safe Software has offerings that are specifically for local governments or utility companies, with pricing based on the population served. Deployment options include on-premises, in the customer's cloud, and as a vendor-managed cloud offering.

- **Support for nontechnical users via unique, user-centric interfaces**: Safe Software supports business-level users by provisioning a low-code/no-code graphical interface that is part of the data integration platform, but the company is innovating by also offering mobile apps and augmented reality. FME AR allows users to interact with data via their smartphones, such as visualizing water pipe locations underground while the user walks along the street. These

innovative features can help nontechnical users to leverage data in their day-to-day work environments.

- **Limited traction outside of spatial data integration:** Although Safe Software's platform is capable of handling both spatial and nonspatial data, customers without spatial data use cases are less likely to choose Safe Software. This is driven by the company's product marketing — it is clearly pitched as a spatial data integration solution — which limits the number of customers that consider purchasing this product.

- **Metadata capabilities are still developing:** FME includes a "SchemaScanner" transformer to harvest the schema from the source data, which can then be compared to what users expect, but this process is not fully automated. The current workflow is a step in the right direction toward using metadata to detect patterns and make recommendations, but the company's support for metadata activation is not yet fully developed.

- **Limited augmented capabilities:** The product's breadth of augmented capabilities is limited. This includes a lack of metadata support as mentioned above, but also the lack of concepts such as knowledge graph support, FinOps and out-of-the-box techniques to augment or automate sophisticated transformation logic (e.g., using ML models).

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. It also offers integration capabilities within the SAP HANA platform. Its operations are geographically diverse, with a customer base of more than 18,000 companies. The top-three industries it supports are automotive, consumer products and public sector.

Strengths

- **Native integration with SAP applications:** SAP data integration solutions are tightly integrated with various SAP applications like ECC, B/W and HANA. Native integration reduces complexity and improves interoperability within the data ecosystem. In addition, its data integration solutions are preintegrated with data quality, data governance, data preparation and data stewardship capabilities.

- **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 delivers 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.

- **Centralizes metadata in its data catalog:** SAP Data Intelligence offers a rich set of metadata capabilities such as cataloging, search, lineage, data quality checks, PII detection and data access logging for auditability. In addition, its active metadata capabilities enable data and metadata discovery to drive recommendation engines, warnings, rules, and semantic discovery, including a business glossary in the data catalog.

Cautions

- **Perception of high license costs:** Several SAP data integration customers report high licensing costs. Also, starter and small organizations rarely consider SAP data integration products due to its high price perception. However, SAP'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.

- **New customers struggle with initial setup:** Several new SAP data integration customers known to Gartner have expressed concerns around initial setup complexities and delays. Adding to this, there is a shortage of skilled expertise in the market for its newer products like SAP Data Intelligence. Many customers are depending on SAP service partners to address this challenge. However, SAP is moving toward containerized deployments, which should reduce additional consulting services required for the initial setup.

- **Limited integration with non-SAP products:** A significant number of SAP customers report challenges with the lack of functionality of SAP data integration tools to replicate data from SAP tables to non-SAP targets (like AWS, GCP or Azure). Customers report that SAP's data integration tools (like SAP Data Services and SLT Replication Server) are not suited for these tasks (and sometimes have licensing concerns for replicating data to non-SAP targets). Therefore, they procure third-party tools like Theobald, Aecorsoft and SNP Glue that are able to query the SAP application layer and replicate data to non-SAP targets. SAP Data Intelligence Cloud is expected to address this gap.

SAS

SAS is a Challenger in this Magic Quadrant, the same as last year. Based in Cary, North Carolina, SAS offers the following data integration products: SAS Studio Engineer on SAS Viya, SAS Event Stream Processing on SAS Viya, SAS Intelligent Decisioning, SAS In-Database Technologies, SAS Data Management and SAS Data Integration Server. The vendor's customer base for these products is around 16,000 organizations. Its operations are geographically diversified, and its clients span many industries with the top sectors being financial services and government.

Strengths

- **Flexibility of design options for data engineers:** SAS Studio Engineer on SAS Viya enables data integration jobs to be designed either in a GUI; by writing custom code in SAS, SQL and Python; or by invoking code within databases such as Hadoop, Databricks, Azure Synapse or Teradata using SAS in-database technologies. Once designed, these jobs can run in batch mode, streaming mode or "as a service."

- **Continued support for collaborative data engineering:** SAS Studio Engineer provides persona-based interfaces, including specialized interfaces for both nontechnical business users and analysts. SAS Intelligent Decisioning is a business-user-focused tool to develop business rules to monitor and validate data. SAS's AI-driven data transformation suggestions aid in accelerating and streamlining data preparation processes.

- **Enhanced integrations for DataOps support:** SAS integrates with Git, Bitbucket, Jenkins and Azure DevOps among others for building DataOps pipelines. SAS artifacts can work with CI/CD pipelines using APIs and CLIs. In addition to SAS's native scheduler, SAS Job Flow Scheduler, integration with third-party process orchestration solutions like Apache Airflow, Dagster and Prefect is provided.

Cautions

- **Challenges with SAS's vision as an independent data integration vendor:** Customers tend to evaluate SAS for its data integration capabilities primarily when they've already invested in its analytics and data science solutions. SAS is playing into this tendency by positioning SAS Studio Engineer as a part of SAS Viya, which is a unified AI, analytics and data management platform. This further exacerbates the challenge of positioning itself as an independent data integration vendor.

- **Issues with pricing:** Customers point out high license costs as one of the main areas of concern for SAS Data Management. With SAS Viya being a cloud-native platform, customers expect SAS Studio Engineer to be available at more competitive price points and to complement its authorized user-based model by embracing modern consumption-based models.

- **Lacks vision for data fabric support:** SAS does not provide mature capabilities to support active metadata management or semantic data modeling. Strategic partnerships with vendors across technologies like metadata management and data governance are few, thereby limiting opportunities to co-sell unified solutions for complete data fabric support. Partnerships with vendors through the Egeria project are a step in the right direction.

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 1,400 organizations from diverse sectors including technology, consumer goods, retail and manufacturing.

Strengths

- **Product strategy focused on prebuilt connectors called Snaps:** SnapLogic offers more than 600 dedicated Snaps that enable code-free integrations such as flattening nested data structures, bulk load and streaming data. Snaps help developers to connect multiple endpoints together, eliminate data silos, obtain complete insights, and yield high-performing business results.

- **Productivity driven by augmented guidance:** SnapLogic empowers less-technical citizen roles to reduce their reliance on highly skilled data engineers through its ease of use and embedded ML guidance, called "Iris." A single platform with multiple interfaces for all personas makes it more compelling for both data preparation and data engineering use cases.

- **Established role-based go-to-market strategy:** SnapLogic facilitates quicker adoption and onboarding of prospects through its simple pricing and trial versions. It targets departmental leaders with its Fast Data Loader managed service, which can load data into cloud data stores with prepackaged, ready-to-run pipelines.

Cautions
- **Market perception of cloud-only relevance:** SnapLogic suffers the market perception that its integration capabilities are limited to cloud applications despite it offering on-premises solutions. This adversely affects some competitive evaluations known to Gartner. SnapLogic is educating its prospects and customers for its broader reach in the data integration tools market.

- **Limited guidance and support for implementations:** Many SnapLogic customers report difficulties around troubleshooting issues, and noted poor product documentation and training manuals. Deployments in complex scenarios have raised customer expectations about SnapLogic's implementation support and guidance.

- **Limited governance support:** SnapLogic lacks some capabilities such as scoring/certifying data assets for ownership and tagging/masking sensitive data in support of data governance use cases. Some customers state that they need comprehensive data quality assistance.

Software AG (StreamSets)

Software AG (StreamSets) is a new entrant to this Magic Quadrant and is positioned as a Niche Player. It is headquartered in San Francisco, California, and offers the StreamSets DataOps Platform as its data integration tool. The customer base includes more than 150 organizations, mainly in North America and Europe. Clients are commonly in the financial services, technology, telecommunications, insurance and healthcare sectors.

Note: On 19 April 2022, StreamSets was acquired by Software AG, a provider of API management, iPaaS and business process management software.

Strengths
- **Support for streaming, message queues and event-broker platforms:** StreamSets takes a streaming-first approach to ingestion, enabling continuous data delivery. Supported streaming and event brokers include: Apache Kafka, Amazon Kinesis, Amazon SQS, Azure Event Hub, Azure IoT Hub, Google Pub/Sub and RabbitMQ. Because many vendors don't have mature stand-alone DBMS independent stream data integration and data replication, this is a differentiator; StreamSets displays particular strengths in these areas.

- **Engineering-focused DataOps, observability and orchestration:** The StreamSets platform supports orchestration for combining multiple data delivery styles, such as streaming ingestion

into a data lake, followed by batch ETL from a data lake to a data warehouse. StreamSets has mature automatic schema drift detection and resolution, and can send alerts to users. The platform has built-in version control and also supports third-party CI/CD tools like GitLab and Jenkins. This combination of capabilities yields the kind of stable data integration pipeline delivery that appeals to data engineers for analytics and data science use cases.

- **Future-oriented market understanding and product alignment**: StreamSets has a strong vision for the future of data integration, especially as it relates to augmentation, automation and monitoring. As an example, StreamSets has a single-pane-of-glass control plane for management and monitoring, which provides real-time operational transparency. Features like this improve operational intelligence and pave a clear roadmap for future enhancements in augmented data integration and FinOps support.

Cautions

- **Postacquisition uncertainty**: StreamSets' recent acquisition by Software AG has raised some postacquisition concerns related to its product roadmap, packaging, positioning, licensing and postimplementation maintenance and support. As a parent company, Software AG is more relevant in the application integration space, with the acquisition of StreamSets giving it an entry into the data integration market. Software AG has committed to its investments in StreamSets as a stand-alone product along with a detailed roadmap regarding upcoming enhancements, which should alleviate some of these concerns.

- **Limited native support for data cataloging and data lineage**: StreamSets does not currently support data cataloging or lineage from within the DataOps platform, but does integrate with third-party tools like Collibra and Apache Atlas for metadata management support.

- **Market traction currently limited to stream data integration**: StreamSets does not have strong name recognition beyond stream data integration use cases. Although StreamSets has invested effort into its CDC capabilities, organizations who are looking for other data delivery styles, such as bulk/batch data movement or data virtualization, are more likely to consider other vendors.

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 San Mateo, California. Its data integration tools include the Talend Data Fabric (which includes the Talend Studio, Stitch, Talend Big Data, Management Console, API Services, Data Inventory, Pipeline Designer, Data Preparation and Data Stewardship) and the Talend Data Catalog. Talend has more than 7,200 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.

Note: Thoma Bravo, a private equity investment firm, announced the acquisition of Talend on 2 September 2021. Talend acquired Gamma Soft, a provider of log-based change data capture technology for data replication and synchronization support, in April 2022.

Strengths

- **Focus on improved data quality while creating data pipelines**: The Talend Data Fabric platform leverages metadata analysis to provide an initial "trust score" to data pipeline designers. This is an aggregate view of the quality of metadata that measures the validity, completeness, discoverability and usage/popularity of the discovered datasets in the Talend Data Inventory Module. This can be utilized by data engineers to build more optimal data pipelines in the Talend Studio and Pipeline Designer, and to reduce data quality and technical debt caused by schema drift.

- **Improved capabilities for self-service data preparation and data engineering**: Talend's reference customers praise its governed approach to self-service data preparation. Power users can inventory distributed data assets in the Talend Data Catalog. They can then view and improve the quality of the data in the Talend Data Inventory module and allow data stewards to enforce policies in Talend Data Stewardship. Once the stewards are content, the power users can perform their own transformations using the low-code environment of the Talend Pipeline Designer, or invite data engineers to perform complex transformations and operationalize the self-service findings in Talend Studio.

- **Improved support for data replication scenarios**: Talend recently concluded the acquisition of Gamma Soft, a provider of log-based change data capture technology (CDC). This acquisition will enable Talend to augment its existing data replication capabilities (within the Talend Studio) with the native CDC adapters and APIs from Gamma Soft for real-time interenterprise data sharing and database synchronization scenarios.

Cautions

- **Data observability challenges**: Several Talend customers have called out the need for improved data observability within Stitch and other data integration modules of the Talend Data Fabric. Customers have also requested improved diagnostics, root cause analysis and more-efficient performance monitoring of jobs. Others have asked for deeper integration between the Talend Remote Engines (that allow customers to run jobs in different environments) and the Talend Management console for better tracking and management of the remote agents.

- **Support challenges**: A few Talend customers have reported concerns with after-sales support, training and documentation. Customers stated the need for improved documentation with better examples, more descriptive error messages (so that errors can be more easily replicated for seeking Talend support) and, in some cases, faster response and turnaround time from Talend when the customer is on the basic support level.

- **Concerns with pricing and scaling**: A small but significant number of Talend clients have reported challenges with unpredictable pricing as they scale data volumes with Stitch. Some customers also reported concerns with price increments during license renewals with other Talend data integration tools. Talend should work with prospects to better gauge their requirements and be more proactive in suggesting pricing options that balance a land-and-expand strategy with options to scale. Talend introduced pricing and packaging changes in 2021 to address some of these 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 data integration tools portfolio includes TIBCO Data Virtualization (TDV), TIBCO Cloud Integration, TIBCO Messaging, TIBCO Streaming and TIBCO OmniGen. The vendor's customer base for this product set is more than 7,500 organizations. Its operations are geographically diversified, and its clients include companies in the financial services, telecommunications and manufacturing sectors.

Strengths

- **Improved interoperability across products**: TIBCO Cloud Metadata is capable of sharing metadata across all of TIBCO's data and analytics products. TIBCO Cloud Passport provides customers the flexibility to utilize various TIBCO Cloud services (including TIBCO Cloud Integration) through a single consumption model. This is part of TIBCO's efforts to position its capabilities as a loosely coupled, highly cohesive offering.

- **Specialization for most data delivery styles**: TIBCO Software has always been rated well in stream data integration through TIBCO Streaming and TIBCO Cloud Events, in messaging through TIBCO Messaging, and in data virtualization through TDV. A recent acquisition extends this portfolio to ETL and ELT as well. TIBCO is now focused on capabilities that can combine these styles, such as streaming data virtualization within TDV.

- **Strong data fabric vision**: TIBCO Agile Data Fabric embeds its AI/ML capabilities into all products through plug-in algorithms for augmented data management. Newly patented technology around data classification and tagging supports creation of business-enriched logical models over data. The user experience across multiple products has also been improved to bolster the data preparation component of the data fabric.

Cautions

- **Lacking in DataOps support**: TIBCO has been relatively slow to react to the rapidly escalating customer requirements around DataOps, such as end-to-end orchestration of data delivery with automated steps for data management code versioning, code testing and code deployment into production. Although TIBCO does support integrations with some CI/CD and version control tools like Git, and TDV Deployment Manager can be used to move models across dev/prod/test environments, a holistic vision for DataOps is missing.

- **Limited traction for the cloud migration use case**: Customers rarely evaluate TIBCO when migrating analytical workloads to public cloud and maintaining a hybrid cloud environment. TIBCO currently provides containerization support for its data integration products, with full SaaS availability already on its 2022 roadmap. Also, although TDV does play a role in this use case, log-based bidirectional data replication is often a requirement here, which is an area of improvement for TIBCO.

- **Need for improved documentation**: Customers have requested more content to be added to the assistance wiki for TDV, better learning materials for users new to TDV and more updated

documentation on system integration to third-party platforms. TIBCO is on the path toward addressing this by launching a new digital community for customers and partners that will include customer support and product documentation, among other things.

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

The following vendors are included in this Magic Quadrant as they met the inclusion criteria:

- AWS

- Hitachi Vantara

- K2View

- Palantir

- Software AG (StreamSets)

Dropped

- **Adeptia:** Adeptia is dropped from this year's Magic Quadrant because the vendor positions itself as a citizen integration solution focused on self-service data access, integration and sharing. Adeptia did not share the required information to substantiate its ability to meet the latest inclusion criteria for this research (which includes and requires use cases beyond self-service data preparation and integration for citizen integrators).

- **HVR:** Fivetran (already on this Magic Quadrant) acquired HVR in October 2021.

Inclusion and Exclusion Criteria

The inclusion criteria represent the specific attributes that analysts believe are necessary. To qualify for inclusion, the vendor's data integration tool (or tools) must offer at least one "stand-alone" product directly usable by the buyer. The vendors must offer a generally available software product that meets Gartner's definition of a data integration tool. To use the product, customers should be able to procure 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) 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:** 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:** 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 (using schema-drift-handling mechanisms and other means).

- **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:

  - This capability allows data integration tools to encapsulate data in messages that various applications can read. Data is exchanged in real time, often via message queues like Apache 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:

- **Range of connectors/adapters (sources and targets):** This includes native connectors to seamlessly access 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. Preference is given to connectors that work out-of-the-box, as opposed to connectors that must be customized by end users.

- **Data movement topology:** Uni-/bi-/multi-directional movement of data across endpoints (e.g., synchronize, compare, broadcast, consolidate) via physical and virtual modes, meeting batch/microbatch/real-time latency requirements.

- **Complex data transformation support:** Capabilities that support complex data processing operations such as fixing outliers, sophisticated parsing (e.g., free-form text mining, telemetry logs, media mining), complex data modeling (e.g., automated data pipeline creation for Data Vault modeling) and creating reusable transformations.

- **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 (e.g., 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.

- **Integration portability:** This is data flow design portability across the infrastructure (at on-premises, SaaS, CSP and virtual private cloud [VPC]), providing workload management capabilities in a clean, safe and portable runtime environment (like containerization).

- **Metadata and data modeling 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. Capabilities must be provided for extensive use of metadata (e.g., usage data, transaction logs, system workloads) to automate/improve data integration and operations tasks.

- **Data governance and information stewardship support:** Capabilities that assist data governance mandates (e.g., data quality, data lineage) while handling data for meeting specific use cases (e.g., master data management, data sharing). 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.

- **DataOps support:** Change management capabilities to data and related artifacts (e.g., Git integration of data pipelines, data model management), automation (e.g., automated testing), orchestration of data delivery (e.g., CI/CD pipelines) with appropriate levels of security to improve the use and value of data.

- **FinOps support:** Capabilities that enable D&A leaders to iteratively control spending, understand product performance and make choices regarding price-to-performance trade-offs, resulting in optimal allocation of resources in the cloud.

- **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. The ability to deliver integration services via a PaaS delivery model is highly desired by customers.

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 2021 — that is, from perpetual license with maintenance, or 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 **450 subscription or maintenance paying customers** (where "customers" does not mean individual "user" license seats), 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.)

- **Geography:** Have market presence in **at least three of the following regions** (regional market presence is defined as a minimum of 5% of the revenue of the verified production customer base, as well as the existence of dedicated sales offices or distribution partnerships in a specific region):

  - North America (Canada, Mexico and the U.S.)

  - Central and South America

- Europe (including Western Europe and Eastern Europe)

- Middle East and Africa (including North Africa)

- Asia/Pacific region (Including Japan)
- **Market presence:** Demonstrated market presence will be assessed through internal Gartner search, external search engines, Gartner inquiry interest, technical press presence and activity in user groups. A relative lack of market presence could be determined as a reason to exclude a product/service offering.

- Have a data integration tool service **generally available** as of midnight, U.S. Eastern Daylight Time on 2 April 2022. This includes any new functionality added to the service by the specified date. We do not consider beta, "early access," "technology preview" or other not generally available functionality or services. Additionally:

  - Any acquired product or service must have been acquired and offered by the acquiring vendor as of 2 April 2022. Acquisitions after this date were considered under their preacquisition identities, if appropriate, and are represented separately until the publication of the following year's Magic Quadrant.

Vendors that focus on narrow use cases that are too specific for the broader data integration market were excluded. Certain vendor tools were excluded because:

- They focused on only one horizontal data subject area — for example, the integration of customer-identifying data.

- They focused on only a single vertical industry.

- They served only their own internally managed data models and/or architectures (such as providing data integration 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 generally 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

- **Cambridge Semantics** is headquartered in Boston, Massachusetts, and offers Anzo Knowledge Graph Platform. Anzo simplifies the integration, modeling and blending of multirelationship data across silos into insight-rich knowledge graphs at enterprise scale. Users can manage their metadata in the Anzo platform by connecting to data sources and then mapping their relationships. Its ontology management enables users to describe their business concepts, map their data assets, deploy a knowledge graph and drive business decisions. It also provides deep data integration and unification for AI and ML models. There is a good adoption among enterprises toward data fabric use cases. At this time, we do not see sufficient demonstrable market presence from Cambridge Semantics as a stand-alone data integration product vendor, which is why it was not included on this year's Magic Quadrant.

- **CData Software** is headquartered in Chapel Hill, North Carolina, and offers products under three segments. Under the first segment of data connectivity, it provides CData Drivers and CData Connect Cloud, which have more than 3,000 product SKUs. A product SKU is a unique combination of a specific connector and the technology on which it is supported, such as a CData Salesforce Connector available for JDBC, ODBC, Python, Tableau or PowerBI. CData Connect Cloud provides a library of connectivity via a "Data Connectivity as a Service" cloud offering built for business users. Under the second segment of data integration, it provides CData Sync and CData DBAmp. The former is a log-based change data capture product supporting more than 100 data sources, while the latter is a specialist product supporting Salesforce to SQL Server integration. Under the third segment of application integration, it provides CData Arc, which supports multiple communication protocols for B2B integration for secure Electronic Data Interexchange (EDI), Managed File Transfer (MFT) and APIs. At this time, we do not see sufficient demonstrable market presence from CData Software as a stand-alone data integration product vendor, which is why it was not included on this year's Magic Quadrant.

- **Confluent** is headquartered in Mountain View, California, and offers Confluent Cloud and Confluent Platform as its data integration tools. Confluent Cloud is a fully managed, serverless, and cloud-native service for Apache Kafka. Confluent Platform is a self-managed Kafka distribution. Both products are used for enabling stream data integration and stream analytics, both through Kafka and through a portfolio of tools designed to work with Kafka. Many organizations are using Kafka for streaming, messaging and event-based data integration

scenarios. Despite its popularity, organizations struggle to maintain Apache Kafka deployments on their own, which can be operationally burdensome. Kafka does not provide full support for all data integration capabilities (such as orchestration, security and governance) and does not have a schema registry for metadata. Confluent helps address these gaps. Confluent is currently a very popular choice for customers looking to support stream data integration, but was excluded from this year's Magic Quadrant evaluation because it does not provide some essential functions including data virtualization, augmented data integration or data preparation.

- **Data Virtuality** is headquartered in Leipzig, Germany, and offers the Data Virtuality Platform and Pipes as its data integration tools. Data Virtuality uses data virtualization to create a unified semantic layer that helps users access data, even when the data is stored in different locations. The Data Virtuality Platform combines both data virtualization and ETL/ELT, providing flexibility depending on the requirements. Customers connect their data sources to the Data Virtuality Platform, create a data logic to map the connections, and then deliver data to the data targets. Common use cases include real-time data access, creating a logical data warehouse, data federation, data governance, GDPR compliance and creating a self-service semantic layer. Data Virtuality also provides a simplified version of its platform, which is called "Pipes," that is a "data loader" tool that is used for data ingestion and replication. Organizations can use Pipes to load data into a cloud data warehouse. Data Virtuality was excluded from this year's Magic Quadrant because it did not exhibit enough market presence (see our Inclusion Criteria section). Also, Gartner did not see Data Virtuality in a significant number of competitive situations for use-case scenarios not involving data virtualization.

- **DataStreams** is headquartered in South Korea and provides the TeraONE Data Fabric platform, which provides capabilities for data integration, data governance and database 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. Although 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. At this time, DataStreams does not meet our inclusion criteria for market presence, and therefore was not included in the Magic Quadrant this year.

- **dbtLabs** is headquartered in Philadelphia, Pennsylvania, and offers dbt (data build tool) to transform data within data warehouses. Started as an internal productivity tool at Fishtown Analytics (a service company then), it has become dbtLabs (a product company now). It focuses on simplifying the transformation tasks within ELT use cases. It uses DataOps principles to drive velocity including CI/CD deployment practices and test-driven-development (TDD) frameworks to cut down on unmanaged code. Developers use dbt to express a piece of business logic and manage the end-to-end data pipeline process from development to production. There is a wide adoption of the tool among enterprises and a rapidly growing community of users. However, it did not meet the inclusion criteria on support for core capabilities including data movement topology (does not provide data ingestion) and streaming data integration capabilities.

- **eQ Technologic** is headquartered in Costa Mesa, California, and provides the eQube Data as a Service (eQube-DaaS) platform. eQube-DaaS is closely aligned to the data fabric design and offers a low-code/no-code integration platform that supports multiple data integration techniques (through embedded support for batch, streaming, messaging, API-based delivery, data virtualization and application integration). The eQube-DaaS platform is made up of three loosely coupled 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). eQube's data virtualization service is integrated as part of the eQube-DaaS platform and plays a critical role in offering data as a service. Asset-heavy industries (including aerospace and defense, auto and machinery, energy, shipbuilding and high tech) looking to establish a data fabric design for use cases including enterprise search, common data model, API creation, synchronization of multiproduct life cycle management (multi-PLM) systems and their data with popular ERP systems (like SAP) will benefit from eQube's native integration support for industrial applications and IoT systems. Although 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.

- **Google** is headquartered in Mountain View, California, and provides a comprehensive set of data integration capabilities as part of its data management portfolio. Its Google Cloud Platform (GCP) offers Cloud Data Fusion, which is a fully managed and cloud-native data integration service that delivers ETL/ELT capabilities. Additionally, Cloud Data Fusion also supports a low-code data preparation environment for self-service data transformation by citizen integrators. GCP also offers a data replication service, Datastream, that delivers a log-based CDC capability to support real-time data movement to BigQuery (GCPs data warehouse), and other GCP databases including CloudSQL, Spanner, Bigtable and Firestore. For message-oriented data movement, Google offers the Pub/Sub integration service, and for customers looking to support IoT data ingestion and streaming analytics use cases, Google offers the Cloud Dataflow service. GCP has also added an orchestration service called Cloud Composer that supports data orchestration and life cycle management. Finally, GCP offers Dataplex, which centrally manages and governs data across all these GCP offerings. Although GCP offers comprehensive data integration services to support all use cases, it did not meet our inclusion criteria on market presence (specifically due to the low number of Gartner client inquiries mentioning GCP in support of stand-alone data integration use cases).

- **Nexla** is headquartered in San Mateo, California, and offers the Unified Data Operations platform. This platform provides a low-code/-no-code, automation-based approach for diverse roles (such as data analysts, data engineers, business ops teams and data scientists) to achieve self-service data integration, preparation, monitoring and delivery. At the heart of the platform are Data Connectors and Data Products (Nexsets), which can be autogenerated or manually created for user collaboration. Nexsets are logical entities that contain metadata related to datasets, data transformations, data access controls, data errors and alert configurations, to help accelerate and standardize data sharing and collaboration within an organization. Connecting Nexsets with different data systems generates ETL/ELT integration, streaming data integration, API integration, data APIs and more. At this time, Nexla does not

meet our inclusion criteria for market presence, despite having a technically sound offering, and was therefore not included in the Magic Quadrant this year.

- **Striim** is headquartered in Palo Alto, California, and offers Striim Platform and Striim Cloud as its data integration tools. Striim Cloud is a unified stream data integration and stream analytics solution that operates as a fully managed service. Striim Platform offers a self-hosted version of the same core software that can be deployed either in the cloud or on-premises. While other stream data integration tools mainly focus on data ingestion, with limited support for transformation, Striim is able to provide full support for stream data ingestion and integration, so it can be used for in-flight transformation, real-time enrichment and stream analytics. Customers praise Striim's log-based change data capture (CDC) capabilities, which are commonly used to stream data from on-premises and disparate cloud environments to cloud targets. In addition, Striim provides data delivery validation, metadata management, pipeline monitoring, event preview, and alerts, for continuous monitoring to ensure business SLAs and SLOs for real-time data delivery. Striim was excluded from this year's Magic Quadrant because it did not meet the inclusion criteria for revenue/customer count threshold or market presence.

Evaluation Criteria

Ability to Execute

Gartner analysts evaluate providers on the quality and efficacy of the processes, 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.

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 and skills. 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.

- The ability of offerings to allow interoperability to open-source solutions and third-party offerings. Some consumers are prepared to accept products from many different suppliers and assemble them together on their own. Interoperability is therefore 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).

- The ability to support DataOps and orchestration capabilities to enable agile pipeline delivery, management, operationalization and maintenance. CI/CD integration, support for Git and Jenkins, support for regression test/validation, support for data observability, support for schema drift handling and so forth are all expected.

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. This criterion views 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 the vendor's product lines, ARRs, profitability and growth in new geographies/use cases.

- Product/services growth in revenue to determine the vendor growth in the data integration software market.

- Growth in high-momentum use cases such as cloud data integration. Revenue growth through cloud integration tools (iPaaS, SaaS, etc.).

- Other metrics to determine financial viability and spend on R&D efforts to continue differentiation and growth in product lines.

- Investment in skills, people and persona for product development, delivery and support. Retention and growth metrics — both are necessary.

Sales Execution/Pricing

This measures 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.

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. Buyers 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 consider 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 vendors 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.

- Finally, data management as a discipline needs to track, predict and preempt overall cost associated with cloud integration workloads. This is especially important as data and analytics teams are becoming distributed across various domains and are increasingly being placed in various lines of businesses. This makes it important for data management leaders to have the ability to associate the cost of running data integration workloads to the business value they provide. This can be done by analyzing performance and system metadata to best allocate processing capacity. Vendors will need to provide tools that can automate aspects of this through FinOps approaches.

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 looking for vendors that are adding features 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 support 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 can assist with moving away from environments that seem to have lost traction — for example, Hadoop — and toward modern sources and targets, such as cloud database platforms as a service and cloud applications.

- Vendors that support using AI/ML to automate complex and repetitive data integration tasks such as data transformation, orchestration, parts of data modeling and data delivery are preferred.

- Vendors that provide open ecosystems to support independent data integration (which does not depend on any cloud infrastructure specifically or a DBMS or a proprietary data exchange format or data storage format) are preferred.

Marketing Execution

This is the clarity, quality, creativity and efficacy of delivering 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.

We rated the vendors on:

- 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 whether they and their customers can specifically benefit from specialized horizontal or vertical capabilities, geographically targeted approaches, or partner-supported implementation practices.

- Ability of the vendor to support and become a part of open community channels for best practices sharing, sharing connectors/code/mappings/other assets or support open metadata sharing standards.

- Finally, how the vendor is rated and perceived on community review and evaluation channels like the Gartner Peer Insights.

Customer Experience

These services should enable customers to achieve anticipated results. Specifically, this includes quality supplier/buyer interactions, technical support and 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 support.

- 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 evaluate the ability of the vendor to maintain parity between cloud and on-premises software and measure after-sales support, migration ease, ease of deployment and overall maintenance and support.

- Customer feedback on the ability of their vendors to meet roadmap deliverables, technical knowledge sharing, skills enablement and augmentation, and training will all be evaluated.

- 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.

We rated the vendors on:

- 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 active metadata, metadata management overall and semantic tiers.

- Also, they must demonstrate that they can provide ongoing support for the massive bulk/batch data movement market and for other data delivery styles including replication, streaming, API, virtualization and messaging.

- 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

Evaluation Criteria ↓	Weighting ↓
Marketing Execution	Medium
Customer Experience	High
Operations	Low

Source: Gartner (August 2022)

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 listen, understand customer demands, and shape or enhance market changes with their added vision.

We rated the vendors on:

- The ability to formulate product vision around multicloud/intercloud and hybrid data integration capabilities. Also, how the vendors are participating in a cloud ecosystem

- The ability to provide core data integration services including all kinds of data movement topologies, support for bulk/batch, streaming, replication, API, messaging and support for the interoperating and combining of these data delivery styles if needed.

- The ability to provide "advisors" and other insights in design, development, deployment and management of integration services to support decision insights and decision automation.

- 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. Existing support for delivery capabilities through microservices is a plus.

- The ability to contribute to data fabrics/mesh designs. We are looking for support for knowledge graphs, graph modeling, semantic modeling and enrichment, taxonomy to ontology

mappings support, graph analysis, and graph query to support data fabrics. It would be interesting to view support for data product delivery through governance to distributed data product teams in domains.

- Ability to showcase capabilities (and upcoming ones on the roadmap) to support DataOps for agile and automated delivery of data integration pipelines. We will look for capabilities that support agile data engineering practices.

- Finally, the vision to support financial governance through plans for FinOps support and current capabilities to validate this.

Marketing Strategy

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

We rated the vendors on:

- Ability 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.

- Ability to showcase a complex portfolio through clear differentiated messaging justifying purchase and clarifying use of each product/SKU.

- 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

- Vendors' ability to become a part of different cloud ecosystems as an independent data integration partner for the customer

- Strategy to grow beyond existing markets, use cases, geographies, and core capabilities

- Demonstrated evidence of improvement in communication of existing and upcoming tools/services

- Affiliate partnerships

- Growth through varying channels (OEMs, VAR, Sis, consulting companies, join go-to-market, partnerships with vendors in the data and analytics space, etc.)

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.

We rated the vendors on:

- 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 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.

- We will assess vendors' roadmap and existing capabilities to support interoperability (through open metadata exchange) with other tools (their own or third party)

- We will also assess roadmaps to support DataOps and FinOps to support agile data engineering and cost optimization measures.

- Finally, we will be looking out for a roadmap (through demonstrable evidence) that supports seamless on-premises to cloud migration of tools or version/system migration in general. This will include changes to tool pricing as well.

Business Model

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

We rated the vendors on:

- 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.

- It is 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 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 able to interoperate with existing tools within the customer base rather than having to replace all current software.

- We will be looking out for vendors to create a niche for themselves in this complex market. How vendors carve out differentiation, land-expand, grow and target specific differentiated use cases, persona, delivery models and even operating models is evaluated.

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.

- The vendors' ability to provide vertical knowledge and expertise in presales, sales, implementation and support for targeted verticals is evaluated in this section.

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 and deliver workloads to the most optimal execution engines through containerized services and microservices architecture.

- 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.

- Provide support for financial governance to enable business units, CDOs and CFOs to support cost optimization as workloads move to the cloud.

- 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.

- Level and performance of support in different geographies and skills availability to support after sales maintenance, etc.

Table 2: Completeness of Vision Evaluation Criteria

Evaluation Criteria ↓	Weighting ↓
Market Understanding	High
Marketing Strategy	High
Sales Strategy	Medium

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

Source: Gartner (August 2022)

Quadrant Descriptions

Leaders

Leaders are front-runners in their capability to support the combination of different data delivery styles (for example, the ability to combine and switch between ETL/ELT, replication, messaging, API integration and data virtualization based on their use-case demands).

Leaders exhibit significant market mind share and 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. Leaders have an established market presence, significant size and a multinational presence.

In 2022, 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 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 vendors embed AI/ML toolkits, which utilize active metadata (as input) to start automating various aspects of data integration design and delivery. Most vendors in the Leaders quadrant provide capabilities to deliver the data fabric, although some require customization.

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 efficient data engineering through self-service data preparation capabilities and integration portability. They also provide the ability to deliver pipelines and code through containerized services. Leaders identify the need for financial governance, especially for integration workloads running in the cloud and support needs to balance flexibility with cost optimization.

Leaders are strong in establishing their data integration tools to support both traditional and new data integration patterns to capitalize on market demand.

Challengers

Bulk/batch delivery styles (such as replication, streaming or data virtualization) are no longer differentiating but a "must have" feature. As such, this capability is now considered under 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 2022, the Challengers listed in this research exhibit a strong understanding of the current data integration market demand and exhibit both the credibility and viability to deliver. Some Challengers are 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 can productize and market (1) data replication as a key strength for targeted use cases such as cloud data migration, data warehouse automation and data lake enablement or (2) data virtualization for faster turnaround time to analytics or (3) extensive support for moving data and workloads from nonrelational DBMSs such as Hadoop to cloud data warehouses or cloud data lakes.

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 mainframe data, analytics/BI or data science. Gartner realizes that many customers have specialized demands for their most urgent or upcoming projects. We also recognize vendors in the Challengers quadrant that (if needed) can scale to support broader data integration use cases but can also customize their offerings for specific/traditional use cases, data types, data sources/targets, execution environments or specific CSPs.

Overall, the market is pushing Challengers to embrace the market vision of a data fabric, automated data integration and multicloud/hybrid data integration 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 self-service data preparation), performance optimization, query optimization, and movement of workloads to data stores and engines that are best suited for processing. Overall, this move toward enabling data fabric architectures with support for financial governance automation through FinOps is a key driver 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 2022, the Visionaries in this Magic Quadrant have focused early on futuristic capabilities and go-to-market strategies to capitalize on their capacity to leverage:

• Augmented data integration through the data fabric design

• Representing data and metadata in business taxonomies and ontologies with support for semantic modeling

• The ability to provision a knowledge graph of data entities that represent multirelationship data in business context

• Serverless integration that supports a multicloud and hybrid cloud integration architecture

• The delivery of various data integration functionalities as loosely coupled API/microservices

• Seamless orchestration through DataOps techniques

• iPaaS delivery models led by the convergence of data integration, application integration and API integration/API management use cases

• Deliver data applications as data products and support federated governance.

In addition, a significant driver of vision is the ability of tools to connect to and analyze all forms of metadata — both passive and, increasingly, active metadata. With this, tools provide key statistics to developers and citizen integrators that aid with integration design and automation. Visionaries should lead the push toward the utilization of graphs, semantics, knowledge graphs and AI/ML for significant automation in data integration design, delivery and maintenance. Visionaries sometimes lack market mind share or credibility beyond their established customer base, their main use cases or very specific application domains/verticals. They may lack partnerships and (or) tight integrations with other incumbent data management vendors including third-party metadata management, data governance and data quality solutions. Visionaries could still be ramping up partnerships with system integrators, consulting companies and other partners that can ensure consistent support, implementation, training or after-sales support to their clients worldwide and beyond their main established markets. 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

Because this is a mature market, Niche Players generally do not generally exhibit gaps in primary 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. Many have very strong offerings for a specific range of data integration problems — for example, a set of specific 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 often exhibit advantages in pricing in their small footprint and in vertical or horizontal solutions and can't complement an organization's other data management technologies. Niche Players are known for solving one part of the data integration problem well through a targeted solution and may be a "good enough" solution for organizations with less complex needs.

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, but may be lacking maturity on certain features that display market vision.

Context

The market for data integration tools continues to evolve and is supported by strong levels of consolidation, strong revenue growth of 11.8% in 2021 (to reach $3.8 billion at the end of 2021) and rapid 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 Kafka), 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 and are seeking tools with targeted capabilities. As a result, they are interested in tools that are specialists in one data delivery style (for example, data replication, data ingestion, API integration, self-service data preparation or data virtualization). 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.

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

In addition, vendors in the Leaders quadrant are focused on automating various aspects of data integration. These include design, ingestion, schema mapping, schema drift detection and corrections, dataOps automation, orchestration automation, next-best transforms, automated lineage and impact analysis, and infrastructure management. These capabilities for augmented data integration demand a new data integration design — one that supports a balance of connect and collect data integration strategies.

Active-metadata-enabled data integration augmentation 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 2022. 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, will provide the necessary 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 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. Vendors that support just one cloud provider (CSP) or only their own databases or applications will fall behind due to valid lock-in and CSP dependence 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 ingest or replicate the data from their operational DBMSs to cloud data warehouses. This has been a significant driver of growth for many data integration providers (such as Fivetran, Matillion, Qlik and Software AG [StreamSets]). These providers have formed significant partnerships with CSPs like AWS, GCP and Microsoft Azure, along with popular cloud data warehouse vendors such as Snowflake and Teradata, to deliver integrated data to cloud data warehouses and data lakes for analytics. This task is often termed as data warehouse automation and is seen as a differentiator by organizations evaluating tools.

This year, organizations have also started evaluating 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. 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 now 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 serverless execution capabilities. Support for open-source data transformation tools (like DBT), orchestration solutions (like Apache Airflow), and open-source ingestion capabilities using crowdsourced connectors (from vendors like Airbyte, for example) are a sign of vision for this edition of the Magic Quadrant.

For the first time, we have also started evaluating vendors on their ability to iteratively control spending, understand product performance, and help make choices regarding price-to-performance trade-offs, resulting in optimal allocation of resources in the cloud. This is what we are calling FinOps, and it's vital to address the growing concerns of overspending, overbudgeting and cost optimization in the cloud.

As more and more subject matter experts become a part of data integration tasks, their needs should be better supported through native integration with self-service data preparation tools and modules. These modules should support less technically skilled personnel with low-code/no-code integration environments. This capability is now a must-have. Moreover, vendors 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 Assessing the Relevance of Data Virtualization in Modern Data Architectures). In 2022, data virtualization has again been a key criterion for evaluating vendors. In addition, we see a surge in demand for tools that can integrate "data in motion" through stream data integration technologies that provide native connectors to event data sources (like clickstreams, logs and IoT) and the ability to capture, enrich and deliver event data to data stores/applications for real-time analytics and other use cases (see Market Guide for Event Stream Processing).

Market Overview

The data integration tools market continues to push toward distributed and augmented data management. This push is inherent in the modern data fabric architecture (see Data and Analytics Essentials: Data Fabric). *The market has realized that those data integration tools that do not balance "collect"- with "connect"-based data management architectures will always result in data silos and/or poorly integrated infrastructures.* Moving forward, organizations will need to monitor trends that are affecting enterprise requirements and vendor offerings. We highlight some of these below:

- **Data integration market experiences double-digit growth**: The market grew at 11.8% in 2021 as compared with 6.8% in 2020 (see Market Share: Data Integration Software, Worldwide, 2021). The increased growth in this market is part of the broader trend of postpandemic recovery reflected in higher growth of software markets overall, as software growth increased from 9.0% in 2020 to 16.0% in 2021. This increased growth in 2021 for the data integration market is reflected in our updated forecasts for growth as well. Five-year compound annual growth rate (CAGR) for the 2021 to 2026 time frame forecast is now 8.5% (see Forecast: Enterprise Infrastructure Software, Worldwide, 2020-2026, 2Q22 Update), updated from 7.0% same time last year. Cloud adoption continues to be significant, with the iPaaS market growing by 40.2% in 2021.

- **Market leaders continue to lose ground to smaller vendors**: The top five vendors in this market (based on their market share) had a collective market share of 71% in 2017. This number has been steadily dropping over the years, and in 2021, the collective market share was only 52%. A similar trend can be seen when analyzing the top three or even top 10 vendors. One of the main reasons for this is that market share leaders such as Informatica and Talend are ceding market share as they focus their growth efforts primarily on their iPaaS products. Another reason is that smaller vendors with more focused offerings (such as those targeted toward a specific data integration capability like data ingestion or data virtualization etc) continue to disrupt larger vendors with their land-and-expand strategies.

- **Market growth is being driven by support for modern data delivery styles and cloud data ecosystems**: Vendors gaining market share have a common theme: They focus on leadership in specific data integration styles such as data virtualization or data replication, and/or they focus on data integration delivered as a native and managed cloud service. Some high-growth vendors in the first category are Denodo, Qlik, Software AG (StreamSets), Striim and Confluent. These vendors collectively grew their revenue by 32% in 2021. Some high-growth vendors in the second category are AWS, Fivetran, Google, Matillion and Microsoft. These vendors collectively grew their revenue by 94% in 2021. Although these growth numbers are off a much smaller revenue base, the gradual decline in market share for the larger and established vendors shows that they will need to find the right balance between all-encompassing platform solutions and easily accessible point solutions to keep pace (see Market Share Analysis: Data Integration Software, Worldwide, 2021). Some providers have focused on vendor acquisitions to gain maturity in specific data delivery styles in the last couple of years. The most prominent examples include Fivetran acquiring HVR Software (a log-based data replication vendor), Qlik acquiring Blendr.io (for application integration), Talend acquiring Gamma Soft (a log-based change data capture vendor) and TIBCO Software acquiring Information Builders (a batch ETL/ELT technology provider).

- **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 10 (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. The data fabric is an architecture pattern that informs and automates the design, integration and deployment of data objects. This approach results in faster, informed and, in some cases, completely automated data access and sharing. In 2021, some vendors have been able to combine most (if not all) of these capabilities needed to deliver the data fabric 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 (see Quick Answer: What Is Data Fabric Design?).

- **Data mesh starts gaining traction**: One of the key reasons why organizations investigate the data mesh design is that they require decentralized and domain-oriented data delivery of "data products." This push is gathering pace because business teams believe that they are often left waiting for data engineering resource attention and that the data pipelines delivered by engineering teams do not do justice to their time-to-market SLAs. Business teams also request more data integration and modeling ownership to impart subject matter expertise into their data products through semantic modeling, ontology creation and knowledge graph support. As a result, organizations are actively evaluating data integration tools that can balance centralized data pipeline delivery with technical capabilities to provision decentralized data product delivery through capabilities such as data-as-a-service, DataOps-based CI/CD, infrastructure-as-code, GIT integration, scheduling and testing automation, and other agile capabilities to deliver data as a product to domains and business teams. See Data Fabric or Data Mesh: How to Decide Your Future Data Management Architecture.

- **FinOps and financial governance are to be taken seriously for cost optimization in cloud data integration efforts**: This is the first time that Gartner is introducing FinOps as a key capability for evaluating vendors in this Magic Quadrant research. Data integration tools need to track, predict and preempt overall cost associated with cloud integration workloads as data and analytics teams get distributed across various domains and are being placed increasingly in various lines of businesses. This makes it important for data and analytics leaders to have the ability to associate cost of running data integration workloads to the value associated with them and have control over allocating processing capacity to workloads they deem to be important through optimal analysis of performance and system metadata and the ability to associate value to cost. Vendors will therefore need to provide tools that can enable financial governance and automate aspects of this through FinOps approaches that also reflect in their pricing models and governance models. FinOps for data integration comprises the capabilities that enable data and analytics leaders to iteratively control spending, understand product performance and make choices regarding price-to-performance trade-offs, resulting in optimal allocation of resources in the cloud for efficient cost optimization.

- **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 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:** 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. For additional information, see Data and Analytics Essentials: DataOps. Those data integration tools that can support DataOps are naturally being favored over those that do not. Based on our inquiries with clients, key aspects being requested include the ability to deliver data pipelines through CI/CD. 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 tooling like GIT, Jenkins and Maven. Data engineering departments in end-user organizations prefer data integration tools that can help them balance the low-code capabilities (offered as part of the data integration tool) with options to invoke code developed in external Python libraries; in dbt (for complex transformation where manual coding is preferred); and in Apache Airflow, Luigi, Prefect, Dagster and so on for task workflow scheduling, 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.

- **Support for hybrid and intercloud data management is now critical:** 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 implementations use both on-premises and cloud environments. Hybrid data management and integration support has therefore become 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 revision.

- **Independent data integration tools are needed to prevent application/cloud service provider (CSP)/database lock-in:** There needs to be a clear focus on independent data integration tools that do not 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 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, high egress costs and data silos, resulting in the inability of organizations to reuse integrated data for general-purpose use cases. Although 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 must favor independent data integration tools to partner with those provided by CSPs, DbPaaS or even their analytics/BI or data science vendors.

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 provider team. Peer feedback from Gartner Peer Insights, comprising peer-driven ratings and reviews for enterprise IT solutions and services covering more than 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.