

Critical Capabilities for Data Integration Tools

Published: 8 October 2019 **ID:** G00377592

Analyst(s): Nick Heudecker, Eric Thoo, Ehtisham Zaidi

Data integration tools address a wide range of use cases that rely on key data delivery capabilities. This research will help data and analytics leaders identify vendors' relative strengths across these capabilities and select the right tool in support of their data management solutions.

Key Findings

- All data integration tool vendors were rated as “meeting/exceeding expectations” for their support for bulk/batch data movement. However, support for other data delivery styles (data virtualization and data replication, for example) is less consistently delivered.
- Metadata support (both passive and active metadata) is now critical as organizations continue to focus on utilizing metadata to optimize (or even automate) integration flows. The cohort of products in this evaluation averaged 3.6 and 3.7 for active and passive metadata capabilities, respectively. While adequate, these capabilities must improve.
- The desire for advanced data virtualization has been met by a larger group of products and with more success than last year. Roughly one-third of products scored below 3.0, or “meets requirements.” In the previous version of this analysis, nearly half the products scored below 3.0.

Recommendations

For data and analytics leaders looking for data integration tools to support data management solutions:

- Evaluate data integration tool vendors on their ability to **support both passive and active metadata**. Active metadata support and analysis provide the foundation for improved data integration optimization and augmented data management capabilities.
- Complement existing data integration architectures by incorporating tools that support (or will support in the near term) mature data virtualization capabilities.
- Focus on data integration tools that **support role-based data integration** through **mature data preparation capabilities**. Prioritize those tools that include the option to promote and operationalize individually developed integration processes into enterprise-capable ones.

What You Need to Know

The 10 critical capabilities defined in this research, and used to assess the featured vendors, are a subset of the evaluation criteria used in the companion “Magic Quadrant for Data Integration Tools.” They represent the prevalent functionality (of one or multiple products) required to support the six use cases most commonly encountered by Gartner clients.

The ratings allocated to each vendor are driven by the results of a survey of reference customers’ usage and satisfaction, as well as Gartner analysts’ insights obtained through client interactions, vendor briefings and other relevant sources. The degree to which organizations believe a given capability meets their needs, along with the frequency of usage across deployments, also influences the rating.

Enterprises are increasingly reporting a wide range of data integration challenges and use-case scenarios. They are extending their plans for capabilities that serve comprehensive data and analytics requirements. A narrow portfolio or a single type of data delivery — such as bulk/batch extraction, transformation and loading (ETL) alone — is no longer sufficient. Data virtualization, data replication/synchronization, message-oriented data movement and stream data integration are increasingly essential to an overall data integration tool portfolio.

Digital business transformation is driving data integration tool vendors to meet demands for:

- **Data delivery styles** — A full range of data delivery styles based on common metadata allows organizations to intertwine and switch between them to execute modern data and analytics use cases (see “Modernize Your Data Integration Architecture for Digital Business by Combining Data Delivery Styles”).
- **A mix of real-time data delivery** — For example, a combination of change data capture (CDC)/replication and message-oriented data movement can interoperate with event streams, message queues and the enterprise service bus (ESB) (see “Market Guide for Event Stream Processing”).
- **A wide range of integration use cases** — These might include data integration for optimized analytics, data consistency between operational applications, data migration and consolidation, master data management (MDM), interenterprise data sharing, and support for data services orchestration.
- **Integration platform as a service (iPaaS) delivery options** — Organizations are looking for iPaaS delivery options to ease integration, introduce flexibility and agility as they continue to converge application and data integration, and reduce the time to integration (see “Use iPaaS to Extend Your Data Integration Strategy to the Cloud in Hybrid Ways”).
- **Data integration capabilities** — These support the data preparation activities of business-facing integration personas, including citizen integrators and business analysts. Data integration tools must provide support for new data management roles such as data engineers. They must help them create, manage and even automate data pipelines for analytics and data science, and operational data integration requirements (see “Market Guide for Data Preparation” and “Toolkit: Job Description for the Role of a Data Engineer”).

- **Data management approaches** — For example, logical data warehouses and data lake architectures are driving optimized repository architectures and employ a combination of data delivery capabilities to assimilate data involving a variety of datasets (see “The Practical Logical Data Warehouse: A Strategic Plan for a Modern Data Management Solution for Analytics”).
- **The distribution of required computing workloads into parallelized processes and nonrelational structures** — This will continue to advance the ability of data integration tools to interact with stream/event data sources, and to deliver data to, and execute integration tasks in, platforms associated with distributed environments (such as data lakes).
- **The expansion of vendors’ capabilities into application integration** — Expansion provides opportunities to use tools that exploit common areas of both technologies to deliver shared benefits. Organizations have begun to pursue data integration and application integration in a synergistic way to exploit the intersection of the two disciplines (see “Innovation Insight for Hybrid Integration Platforms”).

Additionally, increasing focus on digital business transformation is creating the following trends in the data integration tool market:

- While there is an increased focus on data integration tools that provide a portfolio of data delivery styles, there is still room in the market for specialists. These specialist data integration tool vendors provide small footprint solutions and are focused specifically on becoming the best at “doing one thing.”
- In the market today, some providers are exploring alternatives to data virtualization by deploying distributed data integration services processing. Data virtualization — in a classic sense — consists of read-only logical views of data and various levels of cache management processing optimization in a balanced model. It can be deployed as a tier or layer in a services architecture, or even during API development. Distributed data integration services processing, on the other hand, introduces coordinated remote processes that both read and write data. Classic data virtualization vendors are now introducing this capability, and some data integration vendors have opted to eschew classic virtualization due to their perceived understanding of how remote services will enable multicloud integration (see “Market Guide for Data Virtualization”).

Buyers must avoid adopting data integration tools that have isolated capabilities and do not provision bidirectional metadata sharing with other data management tools. These include data quality tools, data governance, metadata management solutions and MDM tools.

How to Use This Research

Data and analytics leaders should use this research to identify which of the evaluated data integration tools from vendors best support the 10 critical capabilities selected by Gartner, and how these functional capabilities match the six key use cases. They should also evaluate products based on how they support their enterprise’s current needs, and how compatible they are with the vendor’s current product set and future vision (see “The State and Future of Data Integration: Optimizing Your Portfolio of Tools to Harness Market Shifts”).

The unevenness in what some offerings can and cannot do means that buyers might decide to support isolated capabilities without a vision for a broader data integration architecture. This market fragmentation also poses a challenge for organizations that want to address data integration holistically as an ongoing discipline.

While vendors' product scores reflect customer feedback that has been integrated with analysts' input about the critical capabilities, product capabilities do not provide a complete evaluation of either vendor or tool on their own. It is essential to also consider each vendor's market presence, track record, financial and organizational strength, availability of skills, product support and outlook, such as its vision and ability to adapt to market changes and disruptions (see "The State and Future of Data Integration: Optimizing Your Portfolio of Tools to Harness Market Shifts").

With this in mind, Gartner recommends the use of this Critical Capabilities research alongside its companion "Magic Quadrant for Data Integration Tools" to understand the vendor landscape beyond product capabilities. We also recommend using Gartner's "Toolkit: RFP Template for Data Integration Tools" to ensure you gain an overview of the breadth of product capabilities relevant to this market.

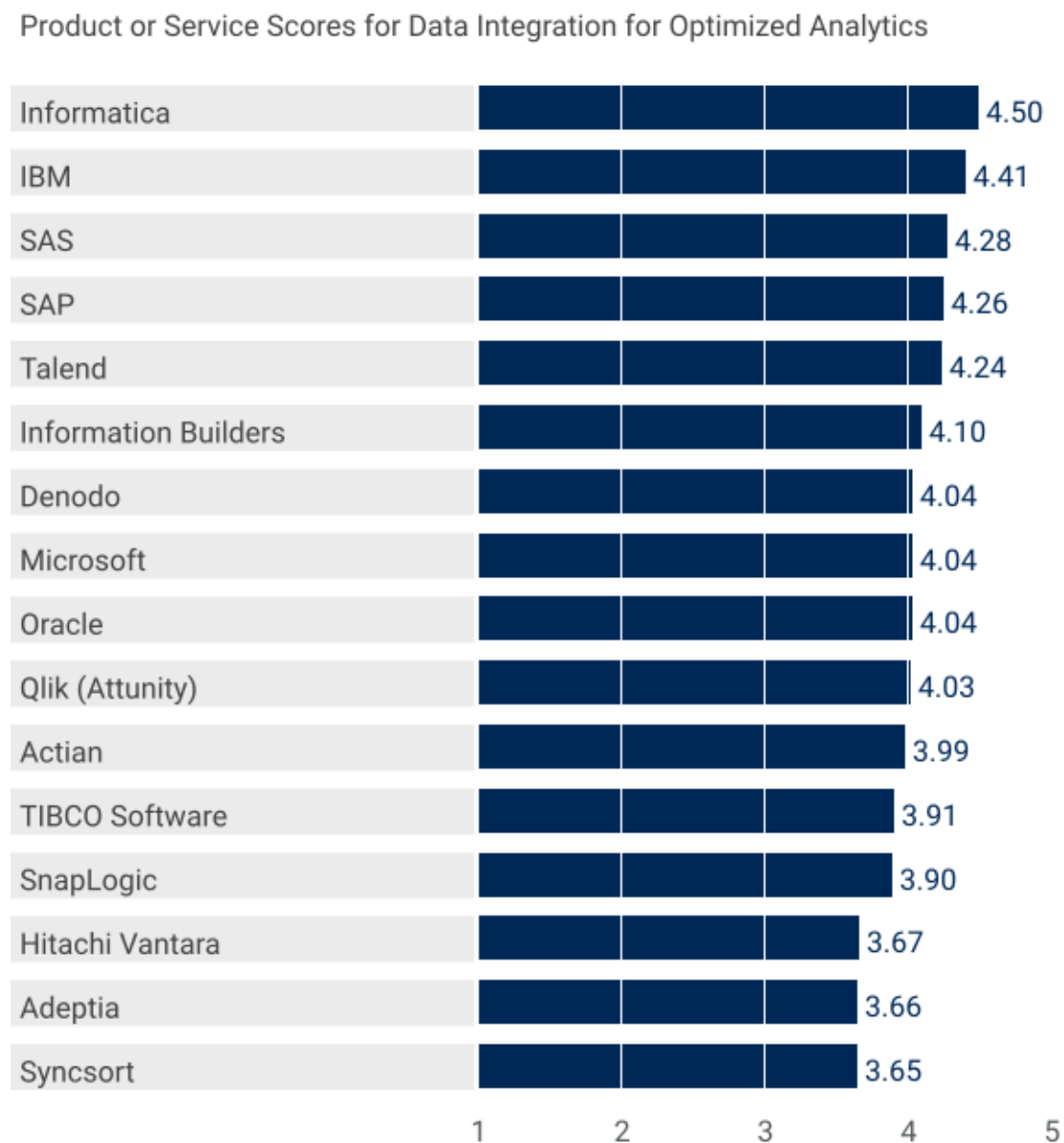
With the successful deployment and standardization of data integration tools, and proper alignment of data integration capabilities to use cases, it is possible to achieve a data integration environment that is adaptable. In other words, the data integration environment changes dynamically to optimize the delivery of data for the business services and applications that will consume it (see "Top Three Acts to Renew Your Data Integration Strategy for Digital Business").

As mentioned, the critical capabilities assessed in this document represent a subset of the evaluation criteria that Gartner recommends when selecting vendors and tools. Therefore, the positioning of vendors in the graphics and tables do not represent overall vendor positioning in the market, and do not always coincide with positioning of vendors in the corresponding Magic Quadrant.

Analysis

Critical Capabilities Use-Case Graphics

Figure 1. Vendors' Product Scores for the Data Integration for Optimized Analytics Use Case

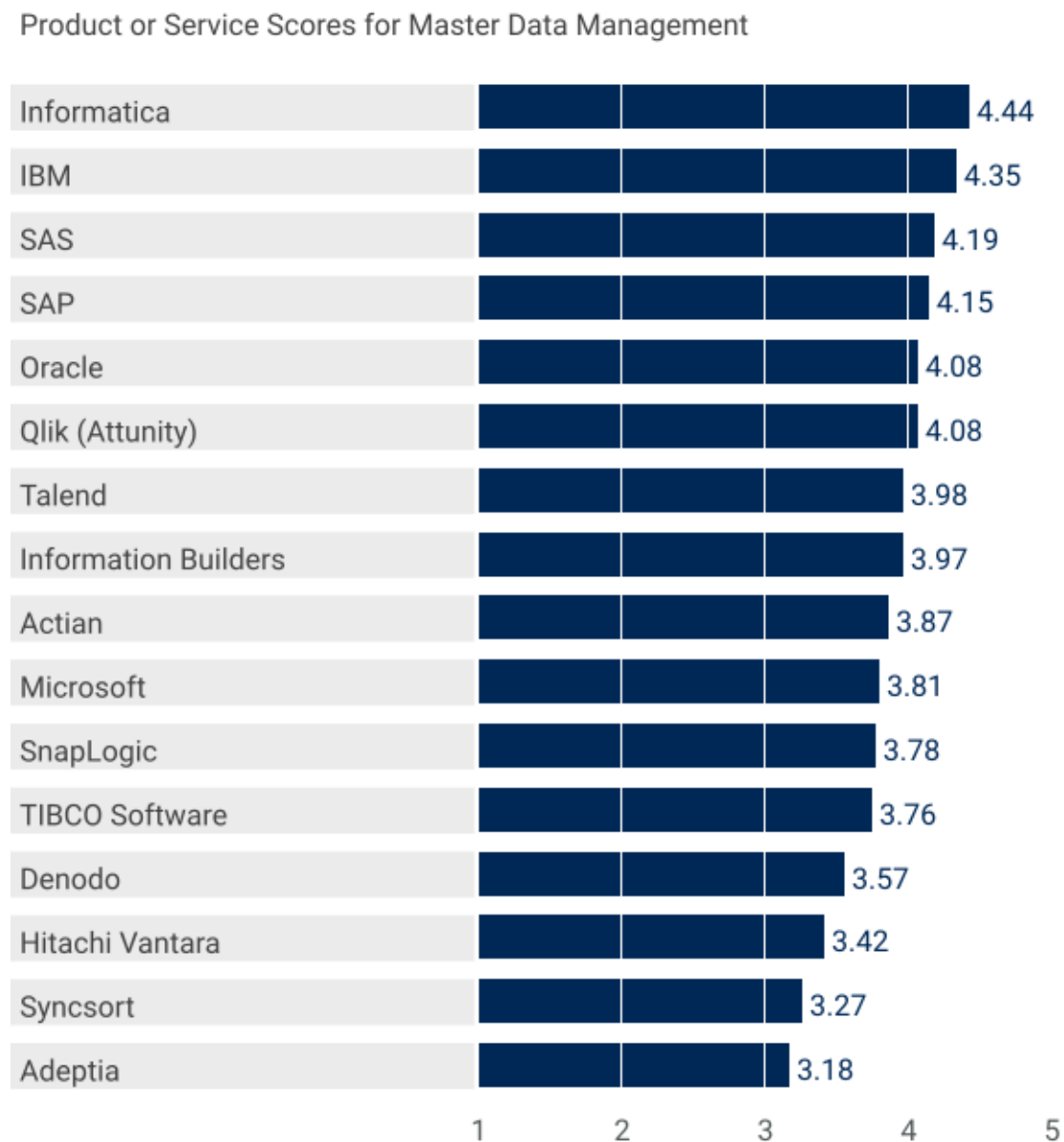


As of 13 September 2019

© Gartner, Inc

Source: Gartner (October 2019)

Figure 2. Vendors' Product Scores for the Master Data Management Use Case



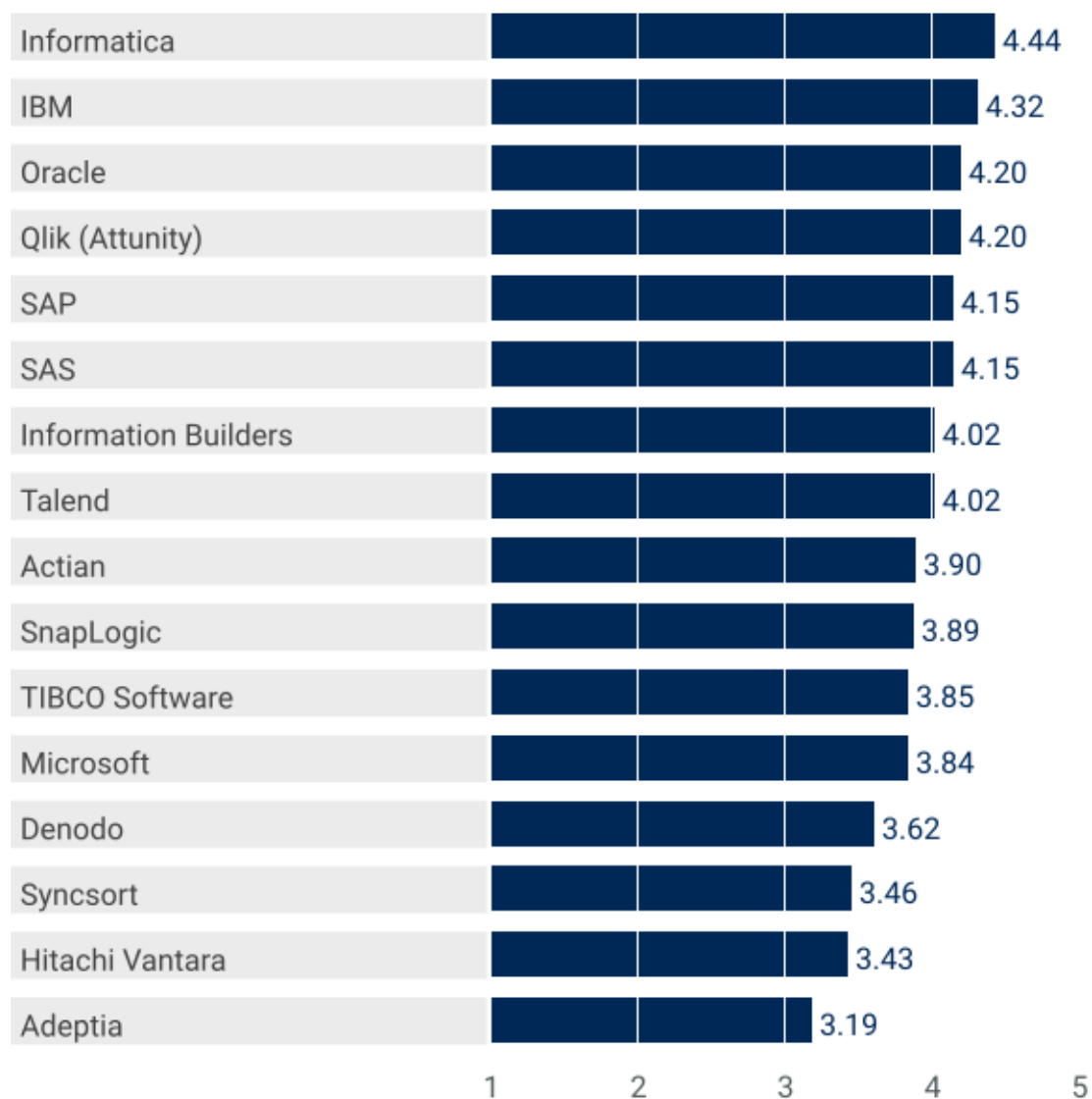
As of 13 September 2019

© Gartner, Inc

Source: Gartner (October 2019)

Figure 3. Vendors' Product Scores for the Data Consistency Between Operational Applications Use Case

Product or Service Scores for Data Consistency Between Operational Applications



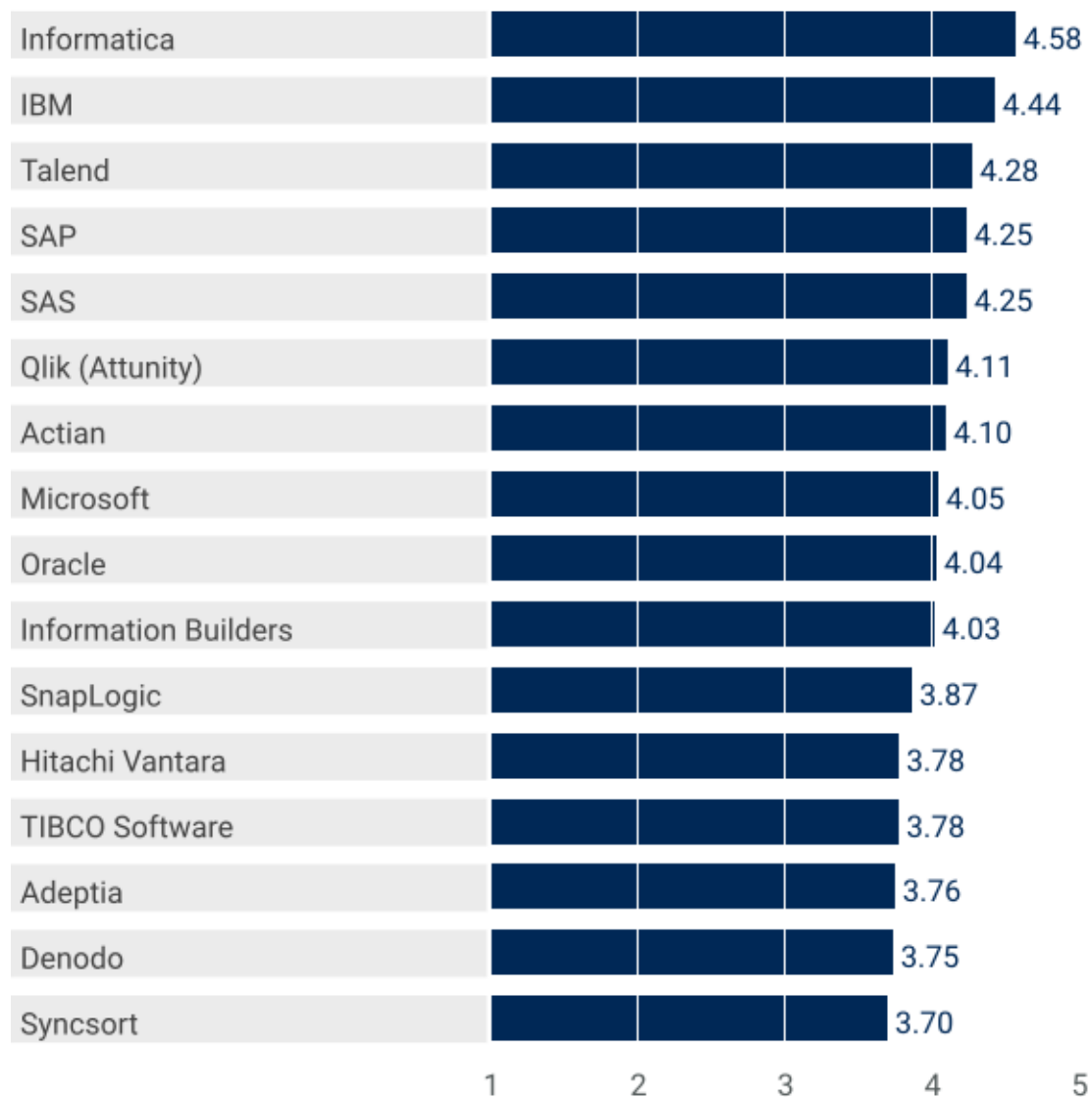
As of 13 September 2019

© Gartner, Inc

Source: Gartner (October 2019)

Figure 4. Vendors' Product Scores for the Interenterprise Data Acquisition and Sharing Use Case

Product or Service Scores for Interenterprise Data Acquisition and Sharing



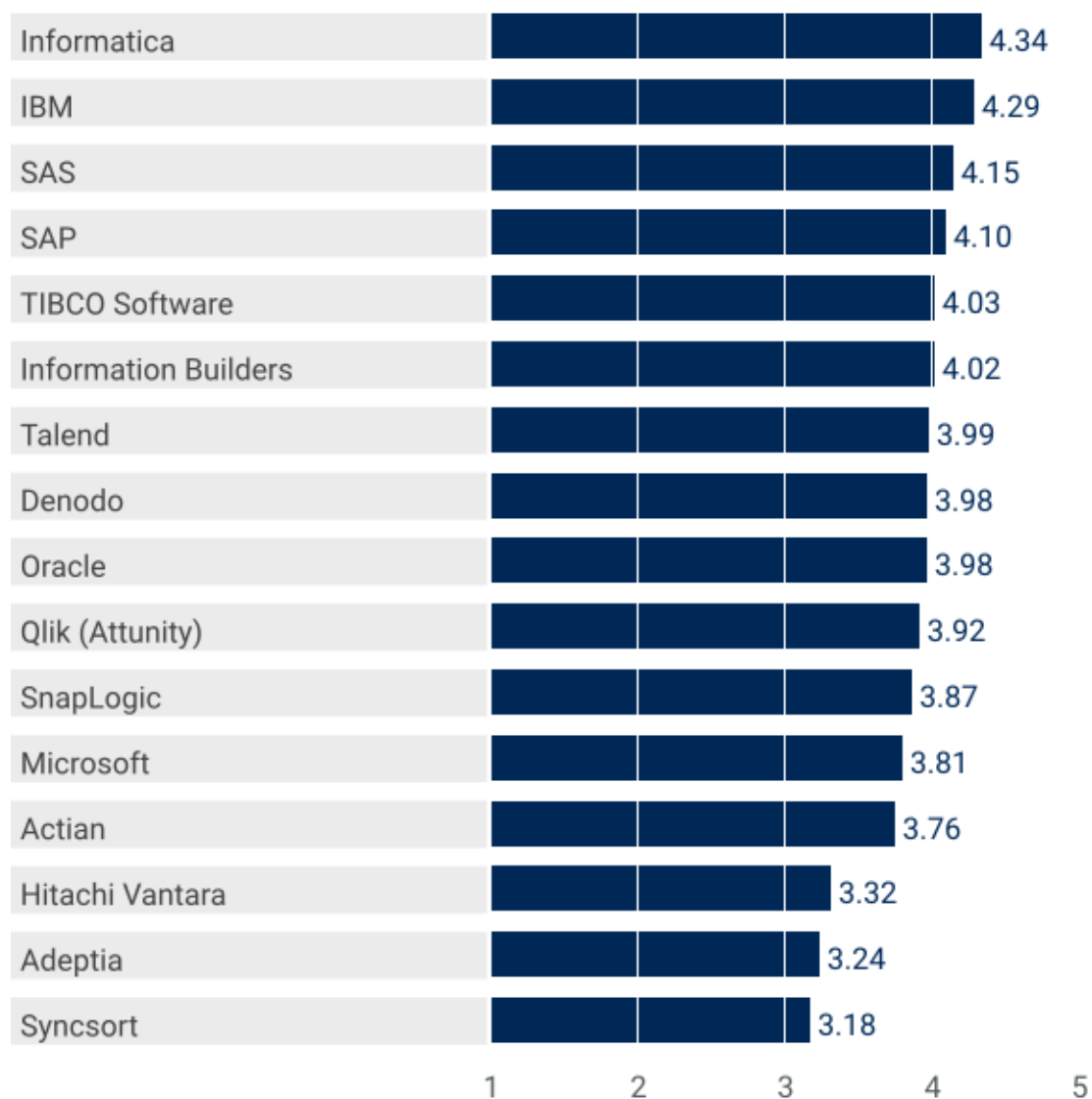
As of 13 September 2019

© Gartner, Inc

Source: Gartner (October 2019)

Figure 5. Vendors' Product Scores for the Data Services Orchestration Use Case

Product or Service Scores for Data Services Orchestration

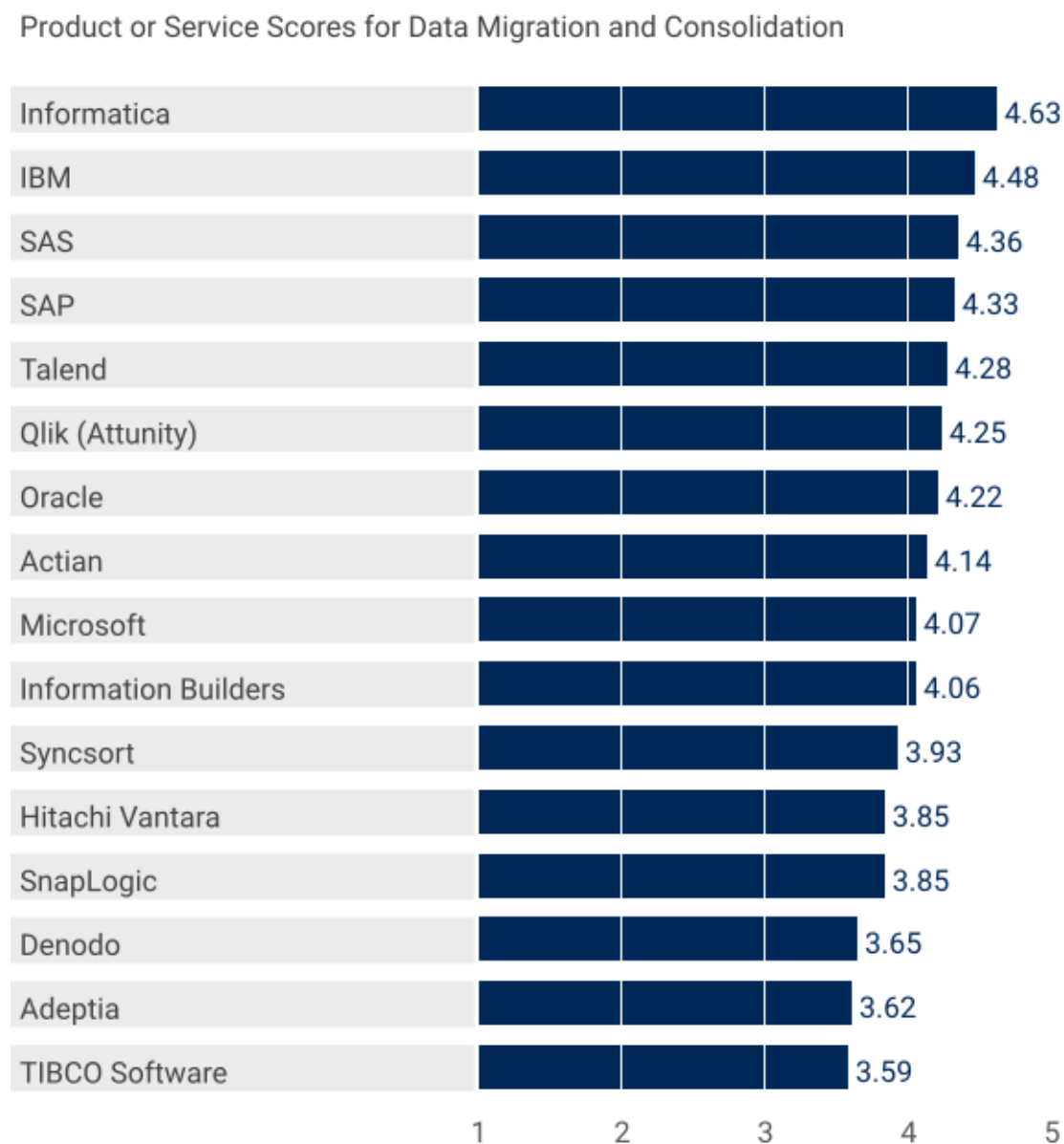


As of 13 September 2019

© Gartner, Inc

Source: Gartner (October 2019)

Figure 6. Vendors' Product Scores for the Data Migration and Consolidation Use Case



As of 13 September 2019

© Gartner, Inc

Source: Gartner (October 2019)

Vendors

Action

Action's DataConnect product supports the data integration capabilities required by this market — spanning real-time, messaging-style and bulk/batch-oriented data delivery. Action was acquired by HCL Technologies and Sumeru Equity Partners on 17 July 2018 but will continue to operate as a

separate company. More than 8,000 customers (including OEMs/embedded) are using DataConnect.

Reference customers for Actian reported that its bulk/batch data movement is a key strength (thus rating it very highly), and that they take advantage of this capability in synergy with the analytics offerings of Actian's broader portfolio. Actian's data integration tools are capable of supporting a range of use cases. However, it is Actian's strength in bulk/batch data movement — combined with support for real-time, message-oriented data movement for B2B integration — that garnered it strong scores for the data migration and consolidation and interenterprise data acquisition and sharing use cases.

The vendor's rating for data virtualization, however, was slightly below the mean survey rating. Its lack of focus in this area contributed to relatively low scores in the data services orchestration and master data management use cases.

Reference customers continue to cite its range of connectivity options as a relative strength, thus rating the vendor highly for critical capabilities including access to multiple data sources and scalability/performance. This extends the vendor's applicability to the data migration and consolidation, and interenterprise data acquisition and sharing use cases.

Adeptia

Adeptia provides Adeptia Connect as its data integration product. An estimated 1,400 organizations currently deploy this offering. Adeptia's data integration tools focus on enabling business users to build integration tasks that facilitate data sharing and flows between applications and users, with an emphasis on simplifying B2B integration. Reference customers provided positive feedback and ratings for the following critical capabilities, which link data movements to business processes: bulk/batch data movement, data preparation and usability, and message-oriented data movement.

While Adeptia's data integration tools support a range of use cases, the vendor's strength is in bulk/batch data movement, message-oriented data movement and B2B integration. As a result, it is most suited to the interenterprise data acquisition and sharing, data integration and optimized analytics, and data migration and consolidation use cases.

The vendor was rated below average for the stream data integration critical capability, which contributed to relatively low scores for the data services orchestration and data consistency between operational applications use cases. Adeptia was also not rated highly for data virtualization and data replication/synchronization relative to most other capabilities. These factors contributed to Adeptia scoring lower for other use cases in this research.

Denodo

Denodo offers the Denodo Platform product, which reflects this vendor's longtime focus on, and strength in, data virtualization. Denodo's customer base for this product set is estimated to be around 700 organizations. It offers data virtualization that abstracts access to a diverse range of underlying data structures, types and sources in the ground and cloud. Reference customers favor

Denodo for its ability to connect to diverse sources and virtualize data access, enable logical abstraction, and support data preparation and cataloging. The Scheduler component, along with the Remote Tables feature, supports materializing views in Denodo to an external data store. These capabilities enable physical bulk/batch movement of data triggered by time, events or conditions.

Reference customers liked Denodo's capabilities for data connectivity and data delivery supported by metadata management, which contributed to strong scores for the data integration for optimized analytics and data services orchestration use cases. A focus on data catalog functionality — i.e., allowing users to inventory distributed data assets connected to Denodo, and to collect, access and use metadata to inform data integration activities — increases the tool's relevance for the data consistency use case.

Denodo was rated below the survey average for its bulk/batch data movement and data replication/synchronization capabilities this year. Therefore, its consolidated scores for the data migration and consolidation and master data management use cases were lower than for other supported use cases.

Hitachi Vantara

Hitachi Vantara offers Pentaho Data Integration, Hitachi Streaming Data Platform (HSDP) and Hitachi Data Instance Director (HDID). The vendor's customer base for this product set is more than 2,000 organizations globally. Hitachi Vantara continues to focus and evolve its data integration offerings by supporting the data needs arising from IT/OT, edge computing, and integration of IoT.

Reference customers provided positive feedback regarding the vendor's connectivity for data access and bulk/batch data movement capability, and its usage in conjunction with stream data integration functionality to link data movements to analytics and operations processes. However, the Hitachi Vantara product set provides limited data virtualization and message-oriented data movement capabilities. It also exhibits usability concerns as integrated deployment and data management requirements grow in complexity. The results are reduced relevance for the master data management and operational data consistency use cases.

Nevertheless, its data integration tools continue to support a broad range of data integration use cases. The vendor's strength in aligning bulk/batch data movement with its analytics platform garnered it relatively strong scores for the data integration for optimized analytics, data migration and consolidation, and interenterprise data acquisition and sharing use cases.

IBM

Key IBM products supporting data integration include IBM InfoSphere Information Server, IBM InfoSphere Classic Federation Server for z/OS, IBM InfoSphere Data Replication, IBM App Connect, IBM Streams, and IBM Data Refinery. The vendor's customer base for this product set is more than 11,000 organizations.

This comprehensive product portfolio offers extensive data integration use-case support and addresses all of the critical capabilities for this market. Reference customers indicated very positive experiences with the company's comprehensive coverage of data integration styles, along with

strong metadata support. IBM received strong scores from reference customers for the following critical capabilities: access to multiple data sources, data replication/synchronization, message-oriented data movement, and metadata support (both active and passive). These areas of strength are reflected in IBM's continued strong performance in the data integration for optimized analytics, data migration and consolidation, interenterprise data acquisition and sharing, and master data management use cases.

IBM's data integration offerings align with its broader portfolio — including data preparation and the use of active metadata supported by machine learning — to aid schema recognition, processing automation and integrating multicloud environments. This makes IBM's data integration portfolio also relevant to the data consistency between operational applications and data services orchestration use cases (for which IBM received strong use-case scores).

IBM received an “exceeded expectations” rating for all critical capabilities except for data preparation and usability, for which surveyed reference customers rated it lower than the mean survey rating (although still ahead of the basic requirements of the market).

Informatica

Informatica's data integration products are built on Informatica's metadata-driven AI engine, CLAIRE, and are offered as part of the Informatica Intelligent Data Platform. The platform comprises: PowerCenter, PowerExchange, Data Replication, B2B Data Transformation, B2B Data Exchange, Data Integration Hub, Data Services, Intelligent Cloud Services, Big Data Management, Big Data Integration Hub, Big Data Streaming, Enterprise Data Catalog, Enterprise Data Preparation, and Edge Data Streaming. The vendor's customer base for this product set is more than 10,000 organizations.

Informatica continues to be rated highly by its reference customers, and it was positioned in the top two vendors for its product set in eight out of the 10 critical capabilities identified by Gartner this year. In synergy with strong metadata support, multicloud integration and hybrid deployment options, Informatica's data integration tools address a broad range of data and analytics requirements. Reference customers regarded the vendor's bulk/batch data movement and breadth of data access capabilities as particularly strong, followed by data replication/synchronization and active metadata support.

Reference customers, however, continue to score Informatica lower for data virtualization than other capabilities. This is in part due to the lack of a dedicated data virtualization product. But it is also due to the lack of marketing around viable alternatives to data virtualization such as microservices and APIs, which Informatica does provide. Still, this score only marginally impacted its overall score for the data services orchestration use case.

Informatica supports a full range of data integration use cases and scored particularly strong for the data integration for optimized analytics, data migration and consolidation, and interenterprise data acquisition and sharing use cases. The vendor also scored well above average for the data consistency between operational applications, and master data management use cases.

Information Builders

Information Builders offers the Omni-Gen data integration platform, which is composed of a central platform plus additional tools, including iWay Service Manager, iWay DataMigrator, and iWay Universal Adapter Suite. Components may be bought separately. Information Builders' customer base for this product set is estimated to be more than 950 organizations.

In order to fulfill a broad range of data integration functionality capable of supporting diverse use cases, Information Builders supports an extensive range of connectors, which are also used by a wide range of competitors. Reference customers reported favorable experiences with bulk/batch data movement and scalability/performance when optimizing the deployment of the selected critical capabilities.

Deployments in conjunction with other technologies from the Information Builders' portfolio can address a diverse range of data management problems. However, reference customers scored Information Builders lower for data preparation and usability, and support for active metadata than for other capabilities.

A mix of strengths in combining data access and movement with the encapsulation of data into messaging and stream-oriented flows contributed to fairly consistent scores for Information Builders across each of the key use cases. Its product set showed strongest relevance for the data integration for optimized analytics, data migration and consolidation, and interenterprise data acquisition and sharing use-case scenarios.

In addition, the vendor's focus on evolving a cohesive data integration environment that can operate with ESB technologies and data preparation aligns well with the master data management, data consistency between operational applications, and data services orchestration use cases.

Microsoft

Microsoft offers data integration capabilities via SQL Server Integration Services (SSIS), which is included in the SQL Server DBMS license. The vendor also includes data integration as part of its Azure Data Factory (ADF) cloud data integration service. Microsoft SQL Server deployments include SSIS for data integration. Microsoft does not report a specific customer count for its data integration tools.

SSIS is most often deployed with SQL Server for bulk/batch data movement, and to integrate with Microsoft-centric environments that include supporting analytics, data management and end-user data manipulation using Microsoft Office tools. Its data preparation and usability capability (which has been rated well above the survey mean by reference customers) results in favorable time to value for data integration tool deployments. Microsoft received one of the best scores for bulk/batch data movement, reflected in its strong support for the interenterprise data acquisition and sharing use case.

Microsoft also offers data integration with ADF to support iPaaS. The ability to redeploy SSIS artifacts to ADF, plus an enhanced deployment option for SSIS on Linux, extends Microsoft's deployment capabilities and supports the combined use of ETL, message-oriented data movement and data preparation. While Microsoft's data integration tools apply to the full range of our use

cases, strengths in synergy with business workflows contributed to its alignment with the data integration for optimized analytics and data migration and consolidation use cases.

Reference customers scored Microsoft below average for its data virtualization and data replication/synchronization capabilities. This is reflected in comparatively low scores for the master data management and data consistency between operational applications use cases.

Oracle

Oracle's data integration tools include: Oracle Data Integration Platform Cloud Service (which includes Oracle GoldenGate Cloud Service and Oracle Data Integrator Cloud Service); Oracle GoldenGate (for data replication, synchronization and big data streaming); Oracle Data Integrator (ODI — for bulk/batch integration); Oracle Big Data SQL (for data virtualization scenarios); Oracle Service Bus (for messaging); and Oracle Integration Cloud Service (iPaaS). The vendor's customer base for this product set is more than 12,000 organizations.

Reference customers highlighted Oracle's strengths in data replication/synchronization, access to multiple data sources, and bulk/batch data movement, with Oracle receiving strong scores across these three capabilities. Reference customers also gave positive feedback on Oracle's scalability/performance optimization and stream data integration delivery styles, as well as on the product set's synergy with Oracle's broader portfolio for application integration support. The resulting scores reflect Oracle's relevance for all key use cases, including a particular affinity for the data migration and consolidation, data consistency between operational applications and master data management use-case scenarios. Also, support for stream data integration, combined with strong ELT, real-time-oriented CDC, diverse data access, and metadata support are particularly relevant to the evolving integration requirements for the data integration for optimized analytics use case.

Some reference customers, however, did cite a need for improvement across the data virtualization, message-oriented data movement, data preparation and usability, and metadata support capabilities. This resulted in a lower use-case score for data services orchestration than for other use cases (although Oracle's score was still above the market average).

Qlik (Attunity)

Note: Qlik completed its acquisition of Attunity in May 2019.

Attunity has a reputation for delivering data replication and synchronization technologies that can be applied to heterogeneous data sources and types. It also has a historical strength in addressing mainframe data, and a current focus on real-time data replication to cloud and nonrelational data stores. Attunity's data integration tools — Attunity Replicate, Attunity Compose, Attunity Visibility and Attunity Enterprise Manager — cover core requirements for data integration functionality. The offerings place emphasis on replicating data to and from the cloud, accelerating data warehouse deployments and providing Hadoop workload optimization support. Attunity's customer base for this product set is estimated to be more than 2,500 organizations globally.

Reference customers cited the strength of its CDC-based data replication/synchronization, stream data integration, and bulk/batch data movement capabilities as key reasons for using Attunity. Integration activities involving Apache Kafka, Spark and Hadoop in cloud environments use real-time intertwining of data movement and event streams. Attunity does not provide data virtualization and so does not resonate well with the data services orchestration use case. With an expanding range of connectors for relational, nonrelational, cloud and mainframe data sources and targets, Attunity supports a diverse range of data integration scenarios. Its scores reflect Attunity's particular relevance for the data consistency between operational applications, data migration and consolidation, data integration for optimized analytics, and interenterprise data acquisition and sharing use cases.

SAP

SAP's key data integration offerings include SAP Data Services, SAP Replication Server, SAP Landscape Transformation Replication Server, SAP Data Hub, SAP HANA, and SAP Cloud Platform Integration Suite. The SAP HANA platform includes SAP HANA Smart Data Integration for bulk/batch and low-latency data movement, SAP HANA Smart Data Access for data virtualization, and SAP HANA Streaming Analytics for stream data integration. SAP's customer base for this product set is estimated at more than 60,000 organizations.

Reference customers reported that the vendor's bulk/batch data movement, access to multiple data sources, scalability/performance optimization, and data replication/synchronization capabilities are generally very strong. Conversely, they scored the vendor's data preparation and usability relatively low, while desiring easier deployment in an integrated manner required across data integration use cases. SAP Data Hub 2.5, released in May 2019, adds self-service data preparation capabilities, but this feature is likely too new to be reflected in reference customer feedback. SAP continues its efforts to simplify seamless administration, monitoring and orchestration of all integration technologies, especially with the SAP Data Hub offering that focuses on integration needs of SAP and non-SAP environments.

The SAP product set has comprehensive data integration functionality that supports all key use cases. Its positive scores for all critical capabilities reflect SAP's relevance and strengths across the data integration for optimized analytics, interenterprise data acquisition and sharing, and data migration and consolidation use-case scenarios. In addition, SAP enables integration flows of variable latency through its support for stream data integration and message-oriented movement data delivery styles. This positions the vendor well for the master data management and data consistency between operational applications scenarios.

SAS

SAS's data integration tool portfolio comprises SAS Data Management, SAS Data Integration Studio, SAS Federation Server, SAS/ACCESS, SAS Data Loader for Hadoop, SAS Data Preparation and SAS Event Stream Processing. SAS's customer base for this product set is estimated to be 14,000 organizations.

SAS covers the full range of data integration functionality and is capable of supporting all key use cases in this report. Use of this vendor's data integration tool portfolio is mostly in analytics

scenarios, but this year the company's data management platform has a strong showing in operational real-time data processing and metadata-driven integration of sources.

Reference customers scored SAS particularly high for its access to multiple data sources, bulk/batch data movement, data virtualization, and scalability/performance optimization capabilities, reflecting its historical depth in analytics/business intelligence (BI) and statistical reporting. The synergistic approach across its portfolio — combining data movement, virtualized data delivery, and advances in related SAS data and analytics portfolios — aligns SAS specifically to the data integration for optimized analytics, data migration and consolidation, and interenterprise data acquisition and sharing use-case scenarios.

Reference customers did, however, highlight the need for improvement in its support for stream data integration, data replication/synchronization and message-oriented data movement, capabilities for which SAS received scores that were below the mean. As a result, SAS's score for the data consistency between operational applications use case was slightly lower than for other use cases (although still deemed adept and above average by reference customers).

SnapLogic

SnapLogic offers the SnapLogic Intelligent Integration Platform. The vendor's customer base is around 2,500 organizations.

As an enterprise integration platform, SnapLogic is a relatively new entrant to the more specific data integration market. The flexibility of its platform, encapsulated in functionality called "Snaps," has been generally well received among its customers and the overall market. However, because its Intelligent Integration Platform targets general enterprise integration use cases, it isn't a specialized tool for data integration, which impacted its scoring for this evaluation.

SnapLogic scored well above 3.0, or "meets requirements," for each use case presented in this research. Reference customers cited its capabilities around bulk/batch data movement and active metadata as particularly strong. However, it scored below average for passive metadata, and its data preparation capabilities were viewed as less sophisticated than the majority of its competitors.

Syncsort

Syncsort's data integration tools — offered via the Syncsort Connect family of products — focus on data integration capabilities that work well for relational, nonrelational and legacy mainframe data integration needs. Syncsort offers Connect CDC, Connect for Big Data, Connect ETL, Connect AppMod, Connect Sort, Ironstream (which integrates security and operational data from z/OS mainframe and IBM i systems for use in technologies such as Splunk and ServiceNow) and Elevate MFSort (for mainframe data integration). Syncsort has a customer base of around 2,250 organizations.

A focus on offloading mainframe data processing workload to Hadoop or cloud data warehouses, along with containerization, cloud and data quality support, extends Syncsort's relevance in the data integration tools market.

Reference customers cited functionality for bulk/batch data movement and stream data integration as very strong capabilities for this vendor. They rated the Syncsort tools' data preparation and usability capability as meeting their basic needs but expressed some concerns regarding dependency on highly skilled developers. Use-case scores reflect relevance across the full range of scenarios, with the data migration and consolidation, and interenterprise data acquisition and sharing use-case scenarios showing relatively strong applicability.

Syncsort has a continuing focus on enabling Hadoop distributions in conjunction with data quality/governance, CDC, and enabling mainframe file and streaming data ingestion. This broadens the relevance of the vendor's tools for the data integration for optimized analytics and data consistency between operational applications use-case scenarios.

Reference customers did, however, regard data virtualization, metadata support and overall usability as weaker than other capabilities. This contributed to lower scores in the other use cases, particularly data services orchestration and support for master data management.

Talend

Talend's data integration products include: Talend Open Studio, Talend Data Fabric, Talend Data Management Platform, Talend Big Data Platform, Talend Data Services Platform, Talend Integration Cloud, and Talend Stitch Data Loader. The vendor's paying customer base for this product portfolio is more than 3,200 organizations.

According to reference customers, the vendor's established strengths are in access to multiple data sources, bulk/batch data movement, and scalability/performance optimization. However, Talend's data virtualization and message-oriented data movement capabilities are less-frequently used (and were rated slightly lower than other capabilities in our reference customer survey). Talend's support for the stream data integration capability was rated among the highest of vendor products evaluated, which suggests strong support for event-stream-processing-related scenarios.

As part of the vendor's basic support for data services, datasets can be exposed as web services, APIs or Spark queries. Overall, Talend is well positioned for broadening usage across data management and integration scenarios. Its synergy with bulk/batch data movement and stream data integration is reflected in a strong affinity with the data integration for optimized analytics, data migration and consolidation, and interenterprise data acquisition and sharing use cases. In addition, the company's commitment to Apache Hadoop, Beam and Spark — along with support for multicloud environments and strong links between its data integration and data preparation capabilities — positions it well to provide integration support for nonrelational data stores.

TIBCO Software

TIBCO's data integration tool offerings include: TIBCO Data Virtualization (TDV), TIBCO Cloud Integration, TIBCO EBX, TIBCO Streaming, TIBCO Messaging (which includes TIBCO FTL, TIBCO eFTL, TIBCO Flogo Enterprise, and TIBCO Enterprise Message Service) and TIBCO Spotfire. TIBCO Spotfire is included for its embedded data preparation and data catalog capabilities. The vendor's customer base for this product set is more than 5,000 organizations.

Reference customers gave extremely positive feedback on TIBCO's message-oriented data movement, with the vendor receiving the highest rating out of all vendors evaluated for this capability. They also called out the vendor's strong performance for advanced data virtualization, which includes pushing processing down to data sources to minimize data movement and reducing time to value.

TIBCO's reference customers also scored it relatively high for access to multiple data sources and stream data integration. This is in line with healthy scores for the data integration for optimized analytics and data services orchestration use-case scenarios. TIBCO received below-average scores for its bulk/batch data movement and data replication/synchronization capabilities. As a result, it was not rated so highly for the data migration and consolidation, and interenterprise data acquisition and sharing use cases.

Context

The market for data integration tools continues to evolve and is supported by strong levels of adoption. Data and analytics leaders must navigate a market brimming with products claiming to solve a range of data integration challenges. However, not all have experience in — or evenly provide — every relevant capability needed across our key use cases. Some vendors focus heavily on moving data in bulk or batch but may place less emphasis on capabilities such as data virtualization, real-time, message-oriented integration, stream data integration, or data replication/synchronization.

Some organizations have determined that basic functions are adequate and are therefore seeking tools with focused and targeted capabilities. As a result, they are interested in evaluating and procuring tools that are specialists in one data delivery style. Others have decided to initiate an enterprisewide effort to optimize the breadth of approaches and are seeking tools that are generalists in data integration and have varied data delivery styles.

More and 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. Organizations are therefore increasingly seeking a comprehensive range of improved data delivery capabilities to modernize their data, analytics and application infrastructures.

In 2019, traditional integration has begun to shift from the bulk/batch dominance for delivery in the market, but only slightly. More than 80% of all organizations in the companion Magic Quadrant survey are making significant use of bulk/batch, but as much as 40% are also utilizing data virtualization, message queues or simple replication with layers of data processing afterward. While many organizations have traditional solutions in place, modern demands have increased the utilization of message-oriented data movement, data virtualization and data replication/synchronization. Somewhere between 35% and 45% of all organizations surveyed are using at least two of these alternative approaches — and an even higher percentage of leading or large organizations are doing so (see “Market Guide for Data Virtualization”).

In the context of digital business, “business moments” — that is, opportunities of short duration or a point in time that sets in motion a series of events involving people, businesses and things — are increasingly attracting the attention of enterprises. They want to harness data to seize these moments, which will require data integration support that includes a focus on stream/event-oriented capabilities. This is where stream data integration is now becoming a strong capability and a relevant data delivery style for modern data integration requirements (see “Market Guide for Event Stream Processing”).

In continuation from 2018, organizations are seeking solutions that facilitate role-based data integration, which includes capability to promote or manage the workflow of converting individually developed processes into enterprise-capable ones (see “Market Guide for Data Preparation Tools”).

A mix of data integration approaches thus becomes crucial, spanning physical delivery to virtualized delivery, and bulk/batch movements to event-driven granular data propagation. In particular, 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 of this data are potentially neither practical nor viable. This is driving an increase in demand for the connection to (and not just collection of) data.

The distribution of required computing workloads to parallelized processes in Hadoop, alternative nonrelational repositories and data lakes will continue to advance the ability of data integration tools to interact with, deliver data to, and execute integration tasks in, platforms associated with emergent data and analytics environments.

The era of metadata has begun — metadata as a byproduct of the design and operations management of a data integration platform is a minimum requirement of data integration tools in 2018. Platforms and solutions are now expected to provide continuous feedback regarding the profiles, quality, use cases, access points, context, performance of integration flows, frequency of access and content analysis of integrated data assets. As far as architects and solution designers are concerned, this feedback is long overdue. It is expected that graph analytics powered by every conceivable type of metadata (both passive and active) will provide the necessary information for introducing machine learning capabilities into data integration platforms.

Finally, the need to acquire and integrate data from cloud environments is becoming critical to many data integration use cases. The expansion of vendors’ capabilities into application integration provides opportunities to use tools that exploit common areas of both technologies to deliver shared benefits. Organizations have begun to pursue data integration and application integration in a synergistic way in order to exploit the intersection of the two disciplines. This combined capability of integration patterns is a key component in enabling a HIP-inspired infrastructure.

Product/Service Class Definition

The data integration tool market comprises products that enable the construction and implementation of data access and data delivery infrastructure for a variety of usage scenarios. Traditionally, separate market segments and vendors existed for the various classes of data integration technology (also known as data delivery styles), such as ETL, stream data integration, data virtualization and replication. As vendors extend their capabilities to broaden their coverage of

data integration styles, the merging of discrete, single-purpose data delivery styles into a market of vendor tools supporting multiple data delivery styles has become common.

These multifunction data integration tools will include the core elements of data integration and be deployable in a variety of use cases. Ideally, users want to model integrated views and data flows once and deploy them in various runtime modes — from batch to real time, bulk to granular, and physical to virtualized.

Critical Capabilities Definition

Access to Multiple Data Sources

Data integration tools are now expected to provide native access and connectivity options to relational DBMS products, plus access to nonrelational data stores (such as NoSQL), Hadoop distributions, flat files, XML and message queues, and data asset types (such as JSON or XML).

Buyers of data integration tools are particularly severe on data integration tools that do not provide access to traditional and modern data sources, data types, applications, cloud object stores and file systems (among others), including:

- Relational databases
- Nonrelational databases
- Various file formats
- XML, JSON and other popular data assets and formats
- Packaged applications, such as those for customer relationship management (CRM), enterprise resource planning (ERP) and supply chain management (SCM)
- SaaS and cloud-based applications and sources
- Industry-standard message formats, such as electronic data interchange (EDI), Health Level Seven International (HL7) and the Society for Worldwide Interbank Financial Telecommunication (SWIFT)
- Parallel distributed processing environments, such as Hadoop Distributed File System (HDFS) and other nonrelational-type repositories (e.g., graph, table-style, document store and key-value DBMSs)
- Message queues, including those provided by application integration middleware products and standards-based products (such as Java Message Service)
- Access to streaming data via event stream processing (IoT, log, sensor data, etc.)
- Data types of a less-structured nature, such as those associated with social media, web clickstreams, email, websites, office productivity tools and content

- Emergent sources, such as data on in-memory repositories, mobile platforms and spatial applications
- Screen-scraping and/or user interaction simulations (e.g., scripts to interact with the web, 3270 or VT100 terminals)

Scalability/Performance

The ability of data integration tools to deliver suitable throughput and response times to satisfy performance SLAs for all data integration use cases, and all data granularity and latency requirements, given the increasingly large data volumes and data diversity needs.

Bulk/Batch Data Movement

Bulk and/or batch data extraction and delivery approaches (such as support for ETL/ELT/ETLT) 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 report.

Bulk/batch data movement continues to be the most popular data delivery style in the data integration tools market, with more than 80% of surveyed organizations reportedly using it for their use-case requirements. (This is based on Gartner's 2018 Data Integration Customer Reference Survey.)

Message-Oriented Data Movement

This capability allows data integration tools to encapsulate data in messages that various applications can read. This allows them to exchange data in real time, often via message queues, but now also via data as a service or even data services orchestration mechanisms.

Data integration tools may need to interoperate with application integration technology, such as when exposing extracts of data from sources as a service to be provisioned via an ESB, and enabling interactions with message request/reply, publish-subscribe and routing. Message-oriented data movement is a data delivery style that utilizes a single record in an encapsulated object. This may or may not include internally defined structures (XML), externally defined structures (EDI), a single record, or other source that delivers its data for action to the data integration process. Apache Kafka is now becoming a very popular and important message queue as both a data source and data target. Organizations evaluating data integration tools are advised to look for Kafka support (if needed) either as an included capability or on the near-term roadmap of their data integration vendor.

Passive Metadata Support

Passive metadata includes an automated acquisition of metadata, data model creation, documentation and maintenance. It also includes lineage and impact analysis reporting, an open metadata repository, and synchronization of metadata with an end-user interface to view and work with metadata.

Passive metadata is static metadata that is either updated manually or captured periodically for design versions. The primary distinction from active metadata is that passive metadata consists primarily of documentation, ranging from fixed schema of sources and/or targets all the way through to business definitions acquired in a glossary and maintained as a formal data dictionary.

As the increasingly important heart of data integration capabilities, passive metadata and data modeling requirements include:

- Automated discovery and acquisition of metadata from data sources, applications and other tools
- Discernment of relationships between data models and business process models
- Data model creation and maintenance
- Physical-to-logical model mapping and rationalization
- Ability to define model-to-model relationships via graphical attribute-level mapping
- Lineage and impact analysis reporting, in graphical and tabular formats
- An open metadata repository, with the ability to share metadata bidirectionally with other tools
- Synchronization of metadata across multiple instances of the tools
- Ability to extend the metadata repository with customer-defined metadata attributes and relationships
- Documentation of project/program delivery definitions and design principles in support of requirements' definition activities
- A business analyst/end-user interface to view and work with metadata

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 virtualization continues to be the fastest-growing delivery style in the market, based on Gartner's 2018 Data Integration Customer Reference Survey. As many as 40% of the organizations surveyed utilize data virtualization, message queues or simple replication, with layers of data processing afterward. Somewhere between 35% and 45% of all organizations surveyed are using at least two of these alternative data delivery styles — and an even higher percentage of leading or large organizations are doing so. Buyers are expressing a desire to use data virtualization particularly in use cases requiring a common semantic layer across several distributed (and heterogeneous) data sources.

Active Metadata Support

The ability of the data integration tool to deliver machine-learning-enhanced metadata discovery, and internal analytics to support, optimize and even automate human data management and integration tasks.

Due to the explosion of data in today's highly connected and digital business environments, the growth in data volume and diversity is fast exceeding the ability to process and integrate this data. Organizations, therefore, expect their data integration tools to provide abilities to autoexecute transformations through machine learning capabilities. In order to support any machine-learning-based automation, enterprises pursuing frictionless sharing of data must have metadata capabilities that far exceed passive metadata practices. Passive metadata is metadata that is static in nature, usually emerges at design time and often requires human or manual updates. Passive metadata most often consists of simple documentation or design-time technical metadata.

Organizations now need their data integration tools to provide continuous access, analysis and feedback on metadata parameters such as frequency of access, data lineage, performance optimization, data quality (based on feedback from supporting data quality/data governance/information stewardship solutions) and context. As far as architects and solution designers are concerned, this feedback is long overdue. It is expected that graph analytics powered by every conceivable type of metadata will provide the necessary information for introducing machine learning capabilities into data integration platforms. The result will be systems that utilize both cost-based and priority-based optimization in a policy-driven solution that will eventually consider combinations of data across on-premises and multicloud deployments. These systems will be able to dynamically relocate data, provide data processing services, and coordinate with governance policies that consider issues such as rights to privacy and legal jurisdictions and provenance.

Data Replication/Synchronization

Allows data integration tools to provision a simple copy of data by physically moving it in near real time from one location to another, always in a physical data store. It can be a basis for all data integration styles, but does not change the form, structure or content of the data it moves.

Data replication supports high-volume and mission-critical scenarios and keeps operational data consistent across multiple systems. It mirrors data, such as enabling CDC between two or more DBMSs, schemas and other data structures, whether of the same type or different types.

Some data replication tools also provide data synchronization capabilities. Data synchronization specifically 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.

Data replication/synchronization is now extremely popular with organizations trying to support near-real-time data delivery between traditional and modern data stores.

Stream Data Integration

The ability to address data integration requirements through interoperability with streams/events, including provisioning of data in-stream for enabling downstream consumption or analysis.

Stream data integration primarily focuses on the ingestion and processing of data sources targeting real-time data integration requirements. This capability allows the data integration tool to filter and enrich the data, and optionally calculate time-windowed aggregations before storing the results in a database, file system or some other store, such as an event broker.

Stream/event data consists of datasets that follow a consistent content and structure over long periods of time. The datasets have large numbers of records that report status changes for the connected device or application, or continuously update records with new values.

This capability includes the ability to incorporate event models, inferred row-to-row integrity, and variations of either of those models or the inferred integrity with alternative outcomes. These outcomes may or may not be aggregated or parsed into separate event streams from the same continuous stream. The logic for this approach is embedded in the data stream processing code. This is particularly popular for integration needs concerning event/stream data from IoT data sources, log data, sensor data, clickstream data and other such real-time event streams.

Data Preparation and Usability

The suitability of data integration tools to support and empower a range of business roles. Emphasis is on engaging nontechnical roles (such as citizen integrators, analysts and data engineers) outside and in support of the IT organization, using techniques and tools such as data preparation.

Data integration tool vendors provide data preparation either through a stand-alone data preparation offering or by including data preparation as a capability within a broader data integration platform.

Data integration tool vendors must address users' data preparations needs, but also ensure that they provide ways for self-service data preparation flows to be operationalized and converted to trusted integration flows through bidirectional metadata sharing and analysis support.

Use Cases

Data Integration for Optimized Analytics

Data is extracted from operational systems and data sources, and transformed, merged and delivered to provision integrated datasets for optimized and repeatable analytics.

Analytics/BI and data science/machine learning involves presenting data in a way that gives insight to the business in support of decision making. The data usually comes from several different (and often operational) sources, such as transactional applications and data repositories that may reside in diverse environments, including on-premises (and now more and more in the cloud). Data

integration tools are foundational in provisioning integrated, transformed and contextualized data-to-data stores that are optimized for analytics/BI and data science use cases. They include data warehouses, data marts, data hubs and cloud data stores.

This use case continues to be a major driver for data integration adoption.

Master Data Management

Data integration tools can be used to integrate, consolidate and synchronize master data related to critical business processes.

This might include data representing a wide range of master data domains — such as customers, products or employees — that are of key importance for master data management (MDM) success.

MDM is the discipline for ensuring the consistency of an enterprise's master data. Different types of MDM approaches will require different data delivery styles (see “Select the Best Master Data Management Implementation Styles for Your Needs”). These might include creating a single view of the master data entity, or provisioning the data for various applications that need to consume it.

Data Consistency Between Operational Applications

Data integration efforts are needed to ensure database-level consistency across applications, both on an internal and an interenterprise basis.

This could involve synchronizing data structures for on-premises applications, or cloud-resident data sources in SaaS, and bidirectional or unidirectional consistency.

When multiple applications store the same data in their databases, that data needs to be consistent. For example, data in an order management system that is also maintained in another system for order fulfillment will need to be appropriately distributed to align the business processes and, where necessary, be consistent.

Interenterprise Data Acquisition and Sharing

Organizations are increasingly required to provide data to, and receive data from, external parties and trading partners (customers, suppliers and others).

Some interenterprise data-sharing requirements involve on-premises or cloud-based environments, or a combination of both. Data integration tools may be used to support data acquisition, sharing and collaborations across business processes and applications, which often consist of the common types of data access, transformation and movement components that are also found in other use cases.

Data Services Orchestration

The ability to deploy all aspects of runtime data integration functionality as data services (for example, deployed functionality can be called via a web services interface).

Data services orchestration includes auditing to assist in service bus management, either internally or bypassing audit metadata to another participating service in the bus.

As acceptance of data service concepts continues to grow, so data integration tools must exhibit service-oriented characteristics and provide support for SOA. This includes:

- The ability to deploy any of the other data integration styles, but with the specific capability to interoperate with application services (logic flows, interfaces, end-user interfaces, and so on); and the ability to pass instructions to, and receive instructions from, those other services on the data service bus
- Management of the publication and testing of data services
- Interaction with service repositories and registries
- Service enablement of development and administration environments, so that external tools and applications can dynamically modify and control the runtime behavior of the tools

Data Migration and Consolidation

Data integration tools increasingly address the data movement and transformation needs of data migration and consolidation, such as the replacement of legacy applications.

Although most of these needs are addressed through custom coding of conversion programs, data integration tools can provide significant support to enterprises undertaking large-scale data migration projects (often due to mergers and acquisitions, modernization or consolidation). However, it should be clear that data integration tools alone do not solve all data migration challenges (see “Take These Four Key Actions to Reduce Data Migration Hazards and Ensure Success”).

Inclusion Criteria

In the context of this Critical Capabilities analysis, we are using the same inclusion criteria as the 2019 “Magic Quadrant for Data Integration Tools.” The inclusion criteria represent the specific attributes that analysts believe are necessary for inclusion in this research.

To be included in this Critical Capabilities research, vendors must possess within their technology portfolio the subset of capabilities identified by Gartner as the most critical from within the overall range of capabilities expected of data integration tools.

Specifically, vendors must deliver the following functional requirements:

- Provide a reference customer base representing any mix of the vendor’s products for use in at least three of these seven key technical data delivery styles: bulk/batch data movement, data services orchestration SOA-style deployments, message-oriented data movement, data replication, data synchronization, stream data integration and data virtualization. Please see the

Critical Capabilities Definition and Use Cases sections for a description of these data delivery styles.

- **Range of connectivity/adaptor support (sources and targets)** — Native access to relational DBMS products, plus access to nonrelational legacy data structures, flat files, XML and message queues and data asset types (such as JSON).
- **Mode of connectivity/adaptor support (against a range of sources and targets)** — Support for change detection, leveraging third-party and native connectors, connection and read error detection, and integrated error handling for production operations.
- **Data transformation support** — It is no longer acceptable to support packaged capabilities for basic transformations only (such as data-type conversions, string manipulations and calculations). Data integration solutions now must support moderately complex needs such as integration with data quality and MDM tools to access changes in data validation directly from the metadata within those solutions.
- **Metadata and data modeling support** — 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. Also, an open metadata repository, including mechanisms for bidirectional sharing of metadata with other tools.
- **Capability to deliver machine-learning-enhanced metadata discovery (and internal analytics)** — Used to enhance human data management and integration requirements.
- **User- or role-specific variations in the development interface** — These are capable of various workflow enhancement mechanisms. These mechanisms may include supporting templates, version modification (via internal library management or other mechanisms), and quality assurance capabilities — either via audit/monitor metadata (manual) or through embedded workflows (administrator tools).
- **Design and development support** — Graphical design/development environment and team development capabilities, such as version control and collaboration. This includes multiple versions running in disparate platforms and multiple instances of services deployments in production environments, as well as alternative or collaborating development environments.
- **Data governance support** — Ability to import, export and directly access metadata with data profiling and/or data quality tools, MDM tools and data discovery tools. Accepting business and data management rule updates from data stewardship workflows, and sharing data profiling information with such tools, are highly desired.
- **Runtime platform support** — Windows, UNIX or Linux operating systems. Demonstrated capability to operate on more than one commercially available cloud environment is desired.
- **Service enablement** — The ability to deploy functionality as services, including multiple operating platforms. The ability to manage and administer operations on multiple platforms and environments is highly desired.

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

- **Revenue** — Generate at least \$30 million of their annual software revenue from data integration tools (perpetual license subscription or maintenance/support). Or, they must maintain at least 300 maintenance-paying/subscription-paying customers for their data integration tools. (Note that the number of downloads without license or maintenance revenue is informative, but not a qualifying piece of information.) Gartner has used as many independent resources for validating this information as possible, specifically to validate provided information.
- **Geography** — Support data integration tool customers in at least two of the following geographic regions or specific national markets: North America, South America, EMEA and Asia/Pacific.
- **Presence** — Demonstrate market presence and was also reviewed and assessed through external search engines, Gartner inquiry interest, technical press presence and activity in user groups or posts. A relative lack of market presence could have been determined as a reason to exclude a product/service offering.

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

- Focused on only one horizontal data subject area; for example, the integration of customer-identifying data.
- Focused on only a single vertical industry.
- Served only their own, internally managed data models and/or architectures (this includes tools that only ingest data to a single proprietary data repository).
- Were used by a single visualization or analytics processing platform.
- Had a DBMS/data management solution for analytics (DMSA)/data lake management vendor that used their data integration tools only to ingest/integrate data into their own repository.

Table 1. Weighting for Critical Capabilities in Use Cases

Critical Capabilities	Data Integration for Optimized Analytics	Master Data Management	Data Consistency Between Operational Applications	Interenterprise Data Acquisition and Sharing	Data Services Orchestration	Data Migration and Consolidation
Access to Multiple Data Sources	20%	10%	15%	10%	20%	10%
Bulk/Batch Data Movement	15%	15%	0%	25%	0%	30%
Data Virtualization	7%	10%	5%	0%	15%	0%
Message-Oriented Data Movement	10%	5%	5%	15%	10%	0%
Data Replication/Synchronization	5%	25%	25%	10%	10%	20%
Stream Data Integration	5%	0%	10%	0%	10%	0%
Data Preparation and Usability	3%	0%	0%	5%	0%	5%
Passive Metadata Support	10%	25%	15%	15%	10%	20%
Active Metadata Support	5%	10%	10%	5%	15%	0%
Scalability/Performance	20%	0%	15%	15%	10%	15%
Total	100%	100%	100%	100%	100%	100%
As of September 2019						

Source: Gartner (October 2019)

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

Critical Capabilities Rating

Each vendor's set of products/services has been evaluated for its critical capabilities on a scale of 1.0 to 5.0, with 1.0 being the lowest score and 5.0 the highest (see Table 2).

Table 2. Product/Service Rating on Critical Capabilities

Critical Capabilities	Action	Adeptia	Denodo	Hitachi Vantara	IBM	Informatica	Information Builders	Microsoft	Oracle	Qlik (Attunity)	SAP	SAS	SnapLogic	Syncsort	Talend	TIBCO
Access to Multiple Data Sources	4.3	4.1	4.6	4.1	4.3	4.7	4.7	4.4	4.4	4.3	4.5	4.6	4.2	4.2	4.6	4.3
Bulk/Batch Data Movement	4.5	4.3	3.7	4.4	4.8	5.0	4.3	4.4	4.2	4.1	4.7	4.7	4.1	4.8	4.7	3.2
Data Virtualization	2.7	2.3	5.0	1.7	4.2	3.5	3.7	3.6	3.2	2.7	3.9	4.0	3.5	1.5	3.0	4.5
Message-Oriented Data Movement	4.1	4.1	3.4	3.0	4.2	4.3	3.9	4.0	3.4	3.8	3.9	3.4	4.0	2.6	4.0	4.7
Data Replication/Synchronization	4.1	2.7	2.0	2.8	4.2	4.5	3.9	3.9	4.9	4.9	4.2	3.9	4.2	3.6	3.5	3.5
Stream Data Integration	3.9	2.5	3.9	3.7	4.3	4.3	4.1	3.6	4.2	4.2	3.9	3.8	4.1	4.2	4.2	3.8
Data Preparation and Usability	3.9	4.3	4.3	4.1	3.9	4.1	2.9	4.4	3.7	3.9	3.4	3.8	3.5	3.1	4.1	3.8
Passive Metadata Support	3.7	2.8	4.1	3.9	4.4	4.3	3.8	3.4	3.6	3.9	3.9	4.2	3.0	2.8	4.2	3.7
Active Metadata Support	3.4	3.1	3.6	3.5	4.2	4.5	3.6	3.2	3.9	3.7	3.8	4.3	3.9	2.5	3.9	3.6
Scalability/Performance	4.0	4.0	4.3	3.8	4.6	4.6	4.1	4.2	4.2	4.2	4.5	4.5	3.9	4.2	4.4	3.8
As of September 2019																

Source: Gartner (October 2019)

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

Table 3. Product Score in Use Cases

Use Cases	Action	Adeptia	Denodo	Hitachi Vantara	IBM	Informatica	Information Builders	Microsoft	Oracle	Qlik (Attunity)	SAP	SAS	SnapLogic	Syncsort	Talend	TIBCO
Data Integration for Optimized Analytics	3.99	3.66	4.04	3.67	4.41	4.50	4.10	4.04	4.04	4.03	4.26	4.28	3.90	3.65	4.24	3.91
Master Data Management	3.87	3.18	3.57	3.42	4.35	4.44	3.97	3.81	4.08	4.08	4.15	4.19	3.78	3.27	3.98	3.76
Data Consistency Between Operational Applications	3.90	3.19	3.62	3.43	4.32	4.44	4.02	3.84	4.20	4.20	4.15	4.15	3.89	3.46	4.02	3.85
Interenterprise Data Acquisition and Sharing	4.10	3.76	3.75	3.78	4.44	4.58	4.03	4.05	4.04	4.11	4.25	4.25	3.87	3.70	4.28	3.78
Data Services Orchestration	3.76	3.24	3.98	3.32	4.29	4.34	4.02	3.81	3.98	3.92	4.10	4.15	3.87	3.18	3.99	4.03
Data Migration and Consolidation	4.14	3.62	3.65	3.85	4.48	4.63	4.06	4.07	4.22	4.25	4.33	4.36	3.85	3.93	4.28	3.59
As of September 2019																

Source: Gartner (October 2019)

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

Acronym Key and Glossary Terms

BI	business intelligence
CDC	change data capture
DBMS	database management system
ELT	extraction, loading and transformation
ESB	enterprise service bus
ETL	extraction, transformation and loading
HIP	hybrid integration platform
IoT	Internet of Things
iPaaS	integration platform as a service
JSON	JavaScript Object Notation
iPaaS	integration platform as a service
MDM	master data management
OEM	original equipment manufacturer
SOA	service-oriented architecture

Gartner Recommended Reading

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

“How Products and Services Are Evaluated in Gartner Critical Capabilities”

“Magic Quadrant for Data Integration Tools”

“Toolkit: RFP Template for Data Integration Tools”

“Market Guide for Data Virtualization”

“Market Guide for Data Preparation Tools”

“Use iPaaS to Extend Your Data Integration Strategy to the Cloud in Hybrid Ways”

Evidence

The analysis in this research is based on information from sources including, but not limited to:

- Extensive data on functional capabilities, customer base demographics, financial status, pricing and other quantitative attributes gained via an RFI process engaging vendors in this market.
- Interactive briefings in which the vendors provided Gartner with updates on their product capabilities.
- Gartner’s 2018 Data Integration Customer Reference Survey — An online survey of the reference customers provided by the vendors in this Magic Quadrant. This captured data on usage patterns, levels of satisfaction with major product functionality categories, various nontechnical vendor attributes (such as pricing, product support and overall service delivery), and more. In total, 238 organizations across all major regions provided input on their experiences with vendors and tools in this manner. The survey was conducted during May and June 2018; the results were collated and analysis completed in June 2018.
- Feedback about tools and vendors captured during Gartner’s data integration tools customer reference surveys of 2016, 2017 and 2018, and in reviewing data integration-related Gartner client inquiry topics and discussion notes from June 2016 through June 2018.
- Market share estimates developed by Gartner’s Technology and Service Provider research unit.
- Extensive data on functional capabilities, customer base demographics, financial status, pricing and other quantitative attributes gained via an RFI process engaging vendors in this market.

Critical Capabilities Methodology

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

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

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

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

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

Ratings and summary scores range from 1.0 to 5.0:

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

2 = Fair: some requirements are not achieved

3 = Good: meets requirements

4 = Excellent: meets or exceeds some requirements

5 = Outstanding: significantly exceeds requirements

To determine an overall score for each product in the use cases, the product ratings are multiplied by the weightings to come up with the product score in use cases.

The critical capabilities Gartner has selected do not represent all capabilities for any product; therefore, may not represent those most important for a specific use situation or business objective. Clients should use a critical capabilities analysis as one of several sources of input about a product before making a product/service decision.

GARTNER HEADQUARTERS**Corporate Headquarters**

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

Regional Headquarters

AUSTRALIA
BRAZIL
JAPAN
UNITED KINGDOM

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

© 2019 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates. This publication may not be reproduced or distributed in any form without Gartner's prior written permission. It consists of the opinions of Gartner's research organization, which should not be construed as statements of fact. While the information contained in this publication has been obtained from sources believed to be reliable, Gartner disclaims all warranties as to the accuracy, completeness or adequacy of such information. Although Gartner research may address legal and financial issues, Gartner does not provide legal or investment advice and its research should not be construed or used as such. Your access and use of this publication are governed by [Gartner Usage Policy](#). Gartner prides itself on its reputation for independence and objectivity. Its research is produced independently by its research organization without input or influence from any third party. For further information, see "[Guiding Principles on Independence and Objectivity](#)."