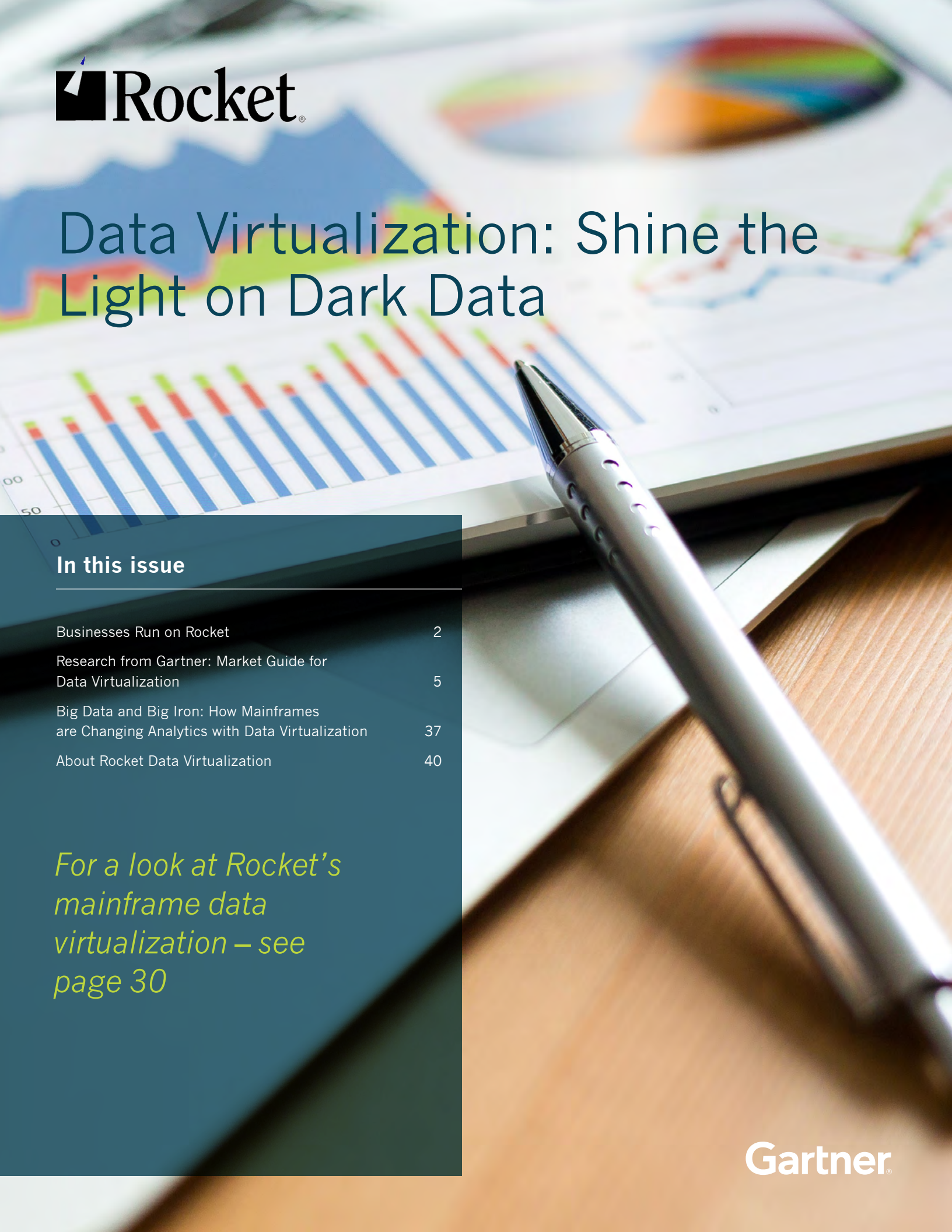


Rocket®

Data Virtualization: Shine the Light on Dark Data

In this issue

Businesses Run on Rocket 2

Research from Gartner: Market Guide for Data Virtualization 5

Big Data and Big Iron: How Mainframes are Changing Analytics with Data Virtualization 37

About Rocket Data Virtualization 40

For a look at Rocket's mainframe data virtualization – see page 30

Gartner®

Businesses Run on Rocket

When it comes to processing power, availability and security, nothing beats the mainframe. But mainframe development isn't stuck in the 20th century; the platform is rapidly evolving to embrace new technologies such as machine learning, predictive analytics and cognitive computing, giving companies with decades of customer and transactional data an edge over less established organizations. The modern mainframe also serves as a source of business intelligence, helping companies identify new revenue opportunities and managing increasingly sophisticated threats, while maintaining its role as the world's pre-eminent transaction processing platform.

For the Rocket Software team, it's all about the data. More data is being created and stored on (or near) the mainframe today than ever before. In fact, the mainframe is the perfect platform for creating and storing this perpetually growing data. We are helping more customers than ever before leverage their mainframe data in place for analytics, application development, modernization, reusable data services, cognitive computing, mobile and web-based applications, and more.

With a 23+ year IBM partnership and unparalleled z Systems technology expertise, companies worldwide trust Rocket to help run their mainframe infrastructure, business processes, and data. We help speed the pace of digital transformation, making it easier for them to take advantage of cloud and mobile computing, advanced analytics, and other innovations that pave the way to new and improved customer experiences.

Shine the Light on Dark Data

Data Virtualization unlocks the power of the mainframe data to deliver better customer insights and drive new revenue opportunities.

We are now living in the information age. Thanks to big data and faster processors, humans are now creating enough content to fill the Library of Congress every 14 minutes. And as technology and networks evolve, the deluge of ones and zeroes will grow every year. Paradoxically this is actually making it harder – not easier – for organizations to find the relevant knowledge they need to succeed. Call it the information-data dichotomy with the main culprit dark data, or information that is being collected but is difficult to access.

This is where Data Virtualization (DV) plays a critical role in helping mainframe users sift through their dark data to uncover the information they need. Turning dark data into real-time business insight represents an economic opportunity to drive new revenue and reduce internal costs, risks, and inefficiencies. DV lets organizations find the proverbial needle in the haystack and perform analytics to drive smart business decisions.

This is critical for organizations that need information immediately because stale or inconsistent data is unacceptable. For example, banks no longer have the luxury of updating their customers' accounts weekly or even daily – people want deposits and withdrawals to be recorded and available instantly. The idea of waiting 24 hours for a check to show up as paid may

have been acceptable in 1997, but in 2017 there is simply no excuse for that kind of lag.

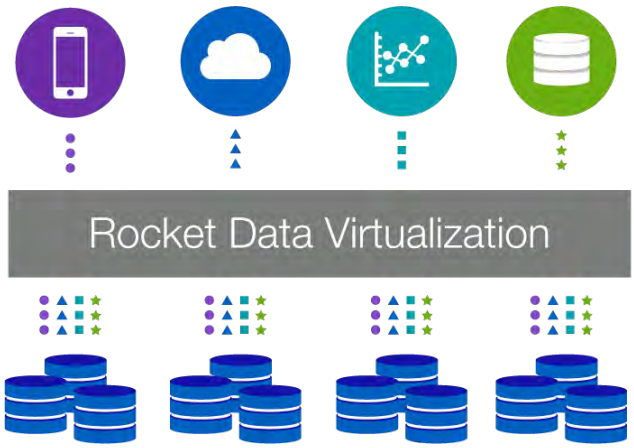
Data Virtualization isn't a theoretical benefit: real companies are using it to increase their profitability and productivity. Financial services organizations are using DV to leverage real-time mainframe data with analytics across multiple platforms (ATM, mobile, Web bill-paying app) to provide more customized customer service. Healthcare providers rely on DV to combine data in real -time from different file formats and databases to decide on-the-spot patient treatments. Insurance companies are combining data on lifetime coverage costs of different patient populations to determine risk and profitability for various groups.

Say Goodbye to ETL

Rocket Software has been working in the mainframe world for more than 25 years, and has seen trends come and go. And one trend that can't end soon enough is technology that extracts, transforms and loads (ETL) data from one platform into an enterprise data warehouse (EDW). Not only is this complex and costly, but it slows analytics to a crawl. And even worse, it opens up the possibility of mistakes, or worse, exposing sensitive data to unauthorized access as it is moved between platforms.

Data Virtualization Snapshot

Data virtualization transforms disparate data into useable information immediately available for mobile, cloud, analytics and Big Data initiatives.



Source: Rocket Software

Rocket DV is different because it actually resides on the mainframe. That means the data never has to move off this highly secure environment. With Rocket DV, data is virtually available and can be joined with other data (on and off the mainframe) to provide business owners with more comprehensive, real-time insight into their customers, business or competition. Users get the information they need, at their fingertips. It really is that easy.

Source: Rocket Software

Research from Gartner:

Market Guide for Data Virtualization

Data virtualization offers data and analytics leaders a data integration strategy to limit data silos and address new analytical and operational opportunities through flexibility in data access. This research provides insight into 16 vendors and their offerings to help in making an informed choice.

Key Findings

- Familiar data integration patterns centered on physical data movement (bulk/batch data movement, for example) are no longer a sufficient solution for enabling a digital business.
- Many organizations report that their existing data warehouse forms yet another data silo in the organization, which leads to a slow degradation of the data warehouse benefits of providing optimized, integrated data delivery.
- Data virtualization offers an attractive alternative to bulk/batch data delivery by allowing flexibility and agility in data access through the creation of a single logical view of data from varied data silos (including transactional systems, RDBMSs, cloud data stores and big data stores).
- Data virtualization offerings are maturing at a steady pace in terms of connectivity options, performance, security and near-real-time data delivery. Current offerings are being increasingly deployed by all major verticals for production-level use cases.

Recommendations

- Consider data virtualization capabilities as important components of an overall data integration portfolio, seeking ways in which they can extend existing data integration architectures.
- Evaluate the current state of your data integration architecture. Set proper expectations upfront, select the right use cases and document the agreed-upon service-level agreements to separate out when to collect data versus simply connecting to it, before starting your data virtualization journey.
- Be cautious about assessments and procurement decisions based on information that is more than a year old, because current and emerging data virtualization platforms are maturing fast in terms of performance and security.
- Use this Market Guide as one input to your evaluation when considering data virtualization tools, assessing vendors' support for your current use case, experience in your vertical or domain, provision of flexible deployment models (and licensing options), plus the overall maturity of the tools themselves in terms of security and performance.

Strategic Planning Assumption

Through 2020, 35% of enterprises will implement some form of data virtualization as one enterprise production option for data integration.

Market Definition

Data virtualization technology can be used to create virtualized and integrated views of data in memory (rather than executing data movement and physically storing integrated views in a target data structure), and provides a layer of abstraction above the physical implementation of data. It is based on the execution of distributed queries against multiple data sources, federation of query results into virtual views, and consumption of these views by applications, query/reporting tools or other infrastructure components.

Data virtualization is covered as a segment for a specific style of deployment in the data integration market (the much broader data integration tools market encompasses data access and delivery through bulk/batch data movement, message-oriented movement of data, data federation or virtualization, data replication and synchronization; see Note 1). It is supported by technologies in the data integration tools market.

The data virtualization technology is not new — its origins are in the distributed relational query technology that was developed two decades ago, combined with more contemporary service-oriented virtualization capabilities. Data virtualization is not confined to federation of databases; modern data virtualization tools have the capability to perform advanced transformations of federated queries from different datasets and can also persist the transformed data to the underlying data stores.

Gartner defines data federation as a subset of the data virtualization technology, since the more modern data virtualization tools go beyond just simple federation of queries from data sources. Unlike traditional, first-generation data federation tools, modern data virtualization tools have the ability to federate and abstract data and provide a consistent interface to data distributed across multiple, disparate data sources and repositories. Modern data virtualization tools provide both read and write access to a host of popular data types and sources (such as relational, Hadoop, NoSQL, flat flies and cloud data stores). They provide additional features, including a metadata repository, abilities to persist transformed and federated queries (and subqueries), and advanced security and query processing features, which were previously missing in first-generation data federation tools.

At a conceptual level, the data virtualization tools being delivered by vendors are composed of several key components (see Figure 1).

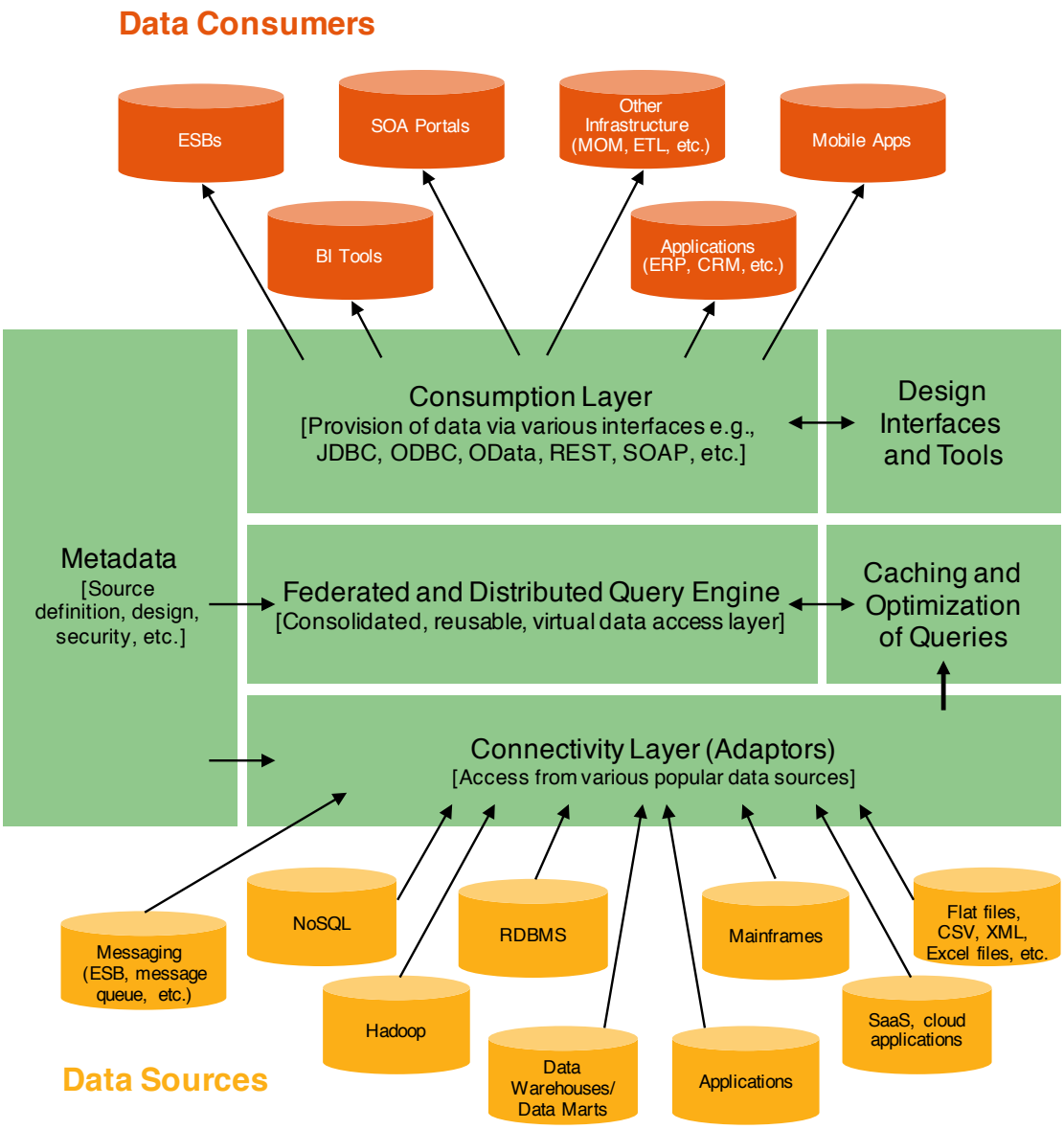
- **Connectivity layer:** Adapters and connectors that enable the access to data from various databases, big data stores (such as Hadoop and NoSQL), applications (such as business intelligence [BI] tools, ERP and CRM), data warehouses and data marts, mainframes, cloud data stores, SaaS applications, message queues and file types. Vendors typically provide a mix of adapters: some they have developed and others they resell from other vendors specializing in connectivity technology. They may also provide capabilities for developing new adapters.

- **Federation and distributed query engine:** The heart of the technology, this component breaks incoming queries into subqueries and executes them, via the connectivity layer, against the respective sources where the desired data resides. The results are returned into memory, where they are joined and mapped to create a composite view of the results.

- **Caching/optimization of queries:** The engine can store whole or partial federated views, as well as controlling where and when source data and federated views are cached. Caching mechanisms range from proprietary file structures to standard relational database management system (RDBMS) tables and also include caching of data in-memory.

- **Consumption layer:** The protocols through which queries are presented and the results of federation can be accessed by users, applications and other infrastructure components. Most vendors support an SQL interface, enabling tools and applications that perform data access via SQL to work easily with their technologies. Support for web services interfaces is also common. The data virtualization tool should have the capability to publish data in a variety of formats and interfaces, including Java Database Connectivity (JDBC), Open Database Connectivity (ODBC), Java Message Service (JMS)-compliant message queues, representational state transfer (REST) and Simple Object Access Protocol (SOAP) web services, JavaScript Object Notification (JSON), XML, portlets and SharePoint web parts. They should also support discovery and the use of data services.

Figure 1. Logical Architecture of Data Virtualization Tools



BI = business intelligence; ESB = enterprise service bus; ETL = extraction, transformation and loading; MOM = message-oriented middleware; SOA = service-oriented architecture; RDBMS = relational database management system; JDBC = Java Database Connectivity; ODBC = Open Database Connectivity; OData = Open Data Protocol; REST = representational state transfer; SOAP = Simple Object Access Protocol.

Source: Gartner (July 2016)

- **Design and administration interfaces:** Additional facilities for configuration, graphical design of integrated views and administration are provided and should be expected. Modern data virtualization tools provide development and administrative/control interfaces and design interfaces with built-in wizards and advisors, as well as recommendation engines that coordinate design issues with performance issues. These also help users to troubleshoot errors when the systems are down.

- **Metadata repository:** Supports other components of the architecture to enable design-time activities and runtime execution. Metadata is generally stored in XML format or a similar proprietary structure. This repository permits the logical mapping of physical data stores to a more application-neutral model.

These tools have the capability to provide a federated data services layer, which allows users to:

- Integrate data from various stand-alone data sources and applications in near real time or batch mode (as required by the use case)

- Perform basic transformations on the data

- Perform transactions that allow persisting the required/updated data to the underlying data store

Market Direction

As an increasingly important part of a comprehensive data integration strategy, data virtualization is attracting renewed interest as organizations recognize its potential for a growing range of use cases. A variety of ways exist in which data virtualization technology can add value as part of the data integration and sharing capabilities of information infrastructure. Most of these opportunities involve augmenting the physically integrated data structures and providing consistent service-oriented approaches for applications and business services to access data. In rendering data resources useful regardless of how they are deployed or where they reside, this technology reads data in place.

Data Virtualization Continues to Grow in Importance

While organizations use multiple styles of data integration technology (bulk/batch data delivery, data replication/synchronization, and message-oriented movement of data, for example) and therefore do not only use data virtualization, this data delivery capability is considered to be an increasingly useful component of an overall data integration technology portfolio.

Data virtualization tools are being actively deployed by organizations to support common use cases, including registry-style master data management (MDM) implementations, the logical data warehouse (LDW) and federating data for pilots and prototypes over data lakes and from development sandboxes. Organizations are expanding their use of this solution beyond “limited” development/test type deployments and are considering it as a real option for enterprise-class projects. This rise in interest for data virtualization is reflected in a steady, continued increase in the volume of Gartner client inquiries about data virtualization concepts and providers

of the technology (see Evidence section). However, Gartner advises organizations to carefully consider performance implications, security concerns, availability requirements for data sources, and data quality and metadata management issues during their design and deployment of federated views of data.

Near the end of 2011, only 11% of organizations reported that they were utilizing data virtualization in a focused set of use cases (for example, operational application integration, as a semantic tier to multiple datasets that were not permanently stored in operational data stores, warehouses or marts, and a few others). By the end of 2015, however, nearly 25% of organizations reported using data virtualization extensively and its use in creating an independent semantic tier has become a significant one. By way of comparison, 86% of organizations using data integration tools report using a bulk/batch data integration style, 62% actively use data replication/synchronization approaches and only about 28% report using message-oriented movement techniques for data integration.

Data Virtualization Tools Continue to Mature Along Three Distinct Axes

Together with its increased usage, data virtualization has begun to mature along three distinct axes. The renewed interest in data virtualization as a major data integration style (for production-level use cases and mission-critical workloads) can be attributed to the increased maturity of data virtualization tools across these three axes, which have otherwise been inhibitors to its adoption as a data integration standard:

- **Broad connectivity to more diverse data sources.** The most basic level of maturity is attributed to "broad access strategies" with access to more data types, more DBMSs and file systems and the use of native drivers/connectors. The more modern data virtualization tool offerings have native connectors to some of the most popular and critical data sources and applications, which need to be integrated for digital business needs, including relational DBMSs, flat files, XML data, message queues, packaged applications, cloud data stores, SaaS applications, big data stores (Hadoop data stores and NoSQL DBMSs, for example), in-memory DBMSs, social media data and unstructured data sources (PDF files, content management repositories and so on). The ability to now integrate with these diverse data sources for a 360-degree view of the business has given a renewed momentum to data virtualization.

- **Distributed and optimized semantic tier processes.** The major catalyst for broader adoption of data virtualization tools has been its maturity along the second axis — that of distributed and optimized semantic tier processes.

  Historically, data virtualization has suffered from four physical barriers to efficiency:

  - Source system access connections
  - Data volume
  - Network capacity

- Query complexity (composed of frequency, logical processes and number of users)

Data virtualization tools have now matured to efficiently address aspects of these four well-known barriers through distributed and optimized sematic tier processes. Modern data virtualization tools exploit technology-driven approaches to address these issues concerning the physics of, and access to, data. Techniques such as multitiered caching (including incremental tier-to-tier updates), process distribution (some as simple as query predicate push-down to the connected source; others are more complex, such as temporary data redistribution across sources), as well as enhanced utilization of memory grids and processing grids, can all potentially be deployed to varying degrees by the different data virtualization offerings in the current market.

These techniques have not only increased the efficiency of existing data virtualization tools in the market, but have also increased adoption and implementation of these tools for mission-critical workloads.

- Maturity of development and administrative interfaces. The third axis of maturity has been along the development and administrative/control interfaces, which have matured beyond simple developer tools to introduce more "citizen integrator"-friendly design interfaces with built-in wizards and advisors, as well as recommendation engines that coordinate design issues with performance issues and allow non-IT employees to develop and design data integration jobs using modern data virtualization tools.

Different Provider Models for Data Virtualization Technology

As a result of shifts in market adoption and the various features and functionality being incorporated into the vendor solutions, data virtualization has become a technology of interest for the traditional DBMS, data integration and application tool vendors.

- **Data integration tool vendors** have begun incorporating data virtualization capabilities into their existing bulk/batch, message and synchronization/replication capabilities to expand their current portfolio of data integration offerings. The best-case solutions utilize a single development interface and metadata and allow optimization recommendation engines to suggest combining these different data integration approaches.

- **DBMS vendors** have extended their capability to link to external files and tables through functionality such as Oracle's DBLinks.

Today, four distinct provider models for data virtualization technology have emerged:

1 Stand-alone data virtualization vendors (Cisco and Denodo, for example)

2 Data integration tool vendors that also provide data virtualization capabilities, either as a separate product or within a single suite/platform as a complementary capability (Informatica, IBM and SAS, for example)

3 DBMSs with extendable data access through data virtualization (from VirtDB, for example)

4 Application tool vendors (especially BI tools) that have now reintroduced their own capability to perform rudimentary virtual data access. This contrasting approach is a repeat of the previously failed attempt to introduce semantic capability in BI tools.

Each of the four approaches has specific benefits and risks that effectively proceed on a spectrum from neutrality and reuse (that is, data virtualization-specific or broader data integration tools) through familiarity — the semantic virtual tier and BI tools capability for users, and the DBMS for the database administration team that is familiar with broad data architecture approaches).

In particular, this style of technology will be used to deliver federated data services — those that formulate an integrated view of data from multiple databases and enable this view to be accessed via a service interface. The value of this technology to the business is also inherent: in that for approximately 10% of the analyst user base in an organization (see Evidence section), a more agile and iterative development cycle is possible for new analysis and reporting requests.

Major Inhibitors to Data Virtualization Adoption

While data virtualization technology is rapidly gaining impetus in the market, a few factors have acted as major inhibitors to its adoption as an enterprisewide data integration standard.

Common inhibitors to adoption include:

- **Market awareness and appropriate use cases.** Lack of market awareness of and familiarity with the current data virtualization technology, offerings and appropriate use cases (see the Market Analysis section) continue to be the biggest inhibitor to mainstream adoption for data virtualization.

- **Poor performance and scalability of traditional offerings.** Since most data virtualization tools historically suffered (some still do) from data volume, network capacity, query complexity and source system connectivity options, performance of queries, processing and transformations against large data sources across multiple systems was seen as a severe inhibitor to adoption. These have now been tackled by most mature data virtualization offerings through distributed and optimized semantic tier processes (as explained earlier in this section).

- **Lack of self-service data preparation features in traditional data virtualization tools.** Demands for data integration are arising from buyers with diverse personas (not just IT users) and with interests in data that range from centralized IT to citizen integrators. The emergence of self-service data preparation tools and those that offer integration features commonly aligned to analytics support, signals a growing demand for business-user-facing tool capabilities. In the past, most data virtualization tools did not include self-service data preparation capabilities, which became a major inhibitor to their adoption.

- **Inappropriate usage and/or expectations from data virtualization.** Organizations have to be careful not to look on data virtualization as a "silver bullet" to solve all their data integration needs. Data virtualization should complement and not completely replace their data integration architecture. Use cases which require near-real-time data access (where performance is critical and likely to be a problem) across big data scenarios (typically those involving extremely high volumes and varieties of data), still need data virtualization that complements other data integration techniques (including data replication and/or bulk/batch data movement, for example).

- **Metadata management and data quality issues.** Organizations should understand that data virtualization will not address all of their data quality and data governance challenges. The extra, upfront work needed to define and architect a metadata layer across systems, and to incorporate appropriate data quality measures, will still be vital because data virtualization technology does not directly address data quality issues. Eventually, data virtualization vendors will have to evolve to incorporate vital data quality components in order for this technology to be used as an organizationwide data integration standard.

- **Lack of true cloud-based options.** Most data virtualization tools available today provide only on-premises deployment options, with some vendors providing hosted versions of their on-premises offerings on popular cloud service providers such as Amazon Web Services (AWS). Buyers, however, are looking for true platform as a service (PaaS) offerings that provide them with the flexibility and scalability of cloud data integration options. Data virtualization offerings have to mature quickly or risk their market share being taken away by integration PaaS (iPaaS) tools.

- **Perceived high cost and lack of flexible licensing options.** Many data virtualization solutions do not provide flexible subscription-based pricing options and licensing models, and therefore appear to end users to be overpriced — with additional carry-on costs (including additional connectors to data sources, additional hardware and other infrastructure, administration, training and so on).

Overall, the market for data virtualization tools is rapidly gaining momentum and is now in the early mainstream phase of maturity with a market penetration of 5% to 20% of its target audience. Gartner expects this renewed interest and momentum to continue in the upcoming three to five years, as organizations continue to struggle with the physical integration of an increasing number of data silos, and there will be an insatiable need to integrate an ever-increasing number of data sources in a faster and more efficient manner — for reduced time to insights and analytics — in order to achieve a digital business.

Market Analysis

Based on our evaluation of data virtualization technology, Gartner sees two categories of use cases — analytical or operational — that greatly influence SLAs and data virtualization technology selection. Some organizations select a strategic vendor to implement data virtualization for both types

of use case, while others focus on only one type. Below we list and describe a variety of common and upcoming use cases (and usage scenarios) for data virtualization technology that reflect opportunities to enhance traditional practices and meet business requirements. These descriptions should help you to effectively define your data virtualization roadmaps, evaluate vendors and adjust your current information infrastructure to accommodate data virtualization.

Analytical Use Cases

In an analytical context, data virtualization focuses on supporting faster business decision making by allowing users to quickly integrate and resolve data silos. Major analytical use cases include:

- **Logical data warehouse architecture.** Diverse requirements for the enterprise data warehouse impose changes in how data is manipulated, using approaches beyond those that are repository-based. Instead of focusing data warehousing efforts exclusively on the storage of integrated datasets in dedicated repositories, organizations are using data virtualization technology to read data in place; this enables data access services to support the LDW's use of abstracted interfaces for various processing needs.

- **Rapid prototyping for batch data movement.** Data virtualization can be used as a precursor for bulk/batch data movement — extraction, transformation and loading (ETL) or extraction, loading and transformation (ELT). Integrated data is eventually consolidated to improve performance and availability. Rapid prototyping using data virtualization is helpful for developers to "flush out" new requirements.

- **Abstract data access layer for analytics.** Data virtualization provides data abstraction — whether the data is on-premises or in the cloud, structured or unstructured. Data virtualization simplifies data access for analytics, improves reuse, reduces change impact and achieves consistent semantics.

- **Self-service analytics and virtual sandbox.** Building on top of an abstracted data access layer, data virtualization empowers business users to perform self-service analytics. It provides a user-friendly view and exposes information assets of high quality. Using data virtualization, business analysts can also rapidly experiment with different sets of data. This virtual sandbox can support self-service data preparation requirements for analytics.

- **Regulatory constraints on moving data.** Data virtualization is useful when regulations prohibit the transfer of data. For example, the EU Data Protection Directive restricts movement of personal data outside of the EU, so rather than moving data outside of the EU, data virtualization enables data access "on the fly" (dynamically).

Operational Use Cases

In the operational context, data virtualization primarily supports business operations. It provides a reusable data access layer for operational applications and manages the complexity of diverse data sources. Major operational use cases include:

- **Virtual operational data store (ODS).** Using a data service layer on operational applications, data is federated from transactional applications and databases by creating an integrated view across them to provide a virtualized ODS. Targeted or tactical data stores to support lightweight, short-term analysis and data visibility can be created by federating data at query time from the data sources of operational applications.

- **Reusable data services in service-oriented architecture.** Abstracted, service-oriented interfaces to data sources are increasingly deployed by using the capabilities of data virtualization technology to establish the service interfaces on which application developers can build service-oriented data access. These services can support greater agility from, and freer access to, data structures and enable application service functionality for abstracted access and maintenance of data sources.

- **Simplify application data access or exchange.** Data virtualization provides an abstract layer to diverse data sources. Applications have different requirements regarding data virtualization capabilities; for example, some require two-phase commitment for distributed data sources, whereas others only require create, read, update, delete (CRUD) for a single data source. There are also different requirements for mobile, cloud, and web and B2B exchange.

- **Legacy system migration.** Data virtualization plays an instrumental role as an abstraction layer between old and new systems during legacy system migration. Users can employ data virtualization for prototyping upfront, and integrate both systems during the parallel run architecture.

- **MDM.** Master data, such as customer or product data, tends to be created and stored in diverse systems. Data virtualization helps MDM systems integrate master data.

Representative Vendors

The vendors listed in this Market Guide do not imply an exhaustive list. This section is intended to provide more understanding of the market and its offerings.

The vendors and products listed here are representative because they have achieved some level of visibility and traction in this market. Vendors are widely diverse in their capabilities, although all support the general template described earlier in this analysis. Gartner encourages decision makers to inspect the product under consideration for the detailed functionalities included in each of the core capabilities. Vendors have been listed in alphabetical order.

Capsenta

Capsenta is a spinout of the Department of Computer Science at the University of Texas (in Austin, U.S.). Its Ultrawrap solution enables organizations to integrate and search both structured and unstructured data residing in disparate data sources (without the need to centralize data in a data warehouse), within and outside their organizational boundaries — by virtualizing the data as a data graph and then mapping it to both standard and proprietary

ontologies using machine-learning-based artificial intelligence matching techniques that allow for faster and more-refined search capabilities. Capsenta allows users to enrich data through semantic technology, enabling federated semantic search across data sources simultaneously. In the healthcare, e-commerce and financial domains, Capsenta Ultrawrap technology is enabling data sources to be searched in near real time without the need for performing ETL-based data centralization (into a data warehouse or database); this results in decreased resource utilization and improved agility.

Ultrawrap supports push-down processing in any JDBC-enabled data source. It enriches data with semantics — enabling data stored in numerous heterogeneous data sources to appear as a single homogeneous data source that can be searched, reported on and visualized with popular BI and analytics tools such as Tableau, for example. For performance optimization, Ultrawrap maximizes the existing component optimizers that are inherent in the underlying databases, by translating graph-based queries to SQL.

Ultrawrap provides role-based and schema-level security (see Table 1).

Cisco

Cisco offers Cisco Information Server (CIS), a Java-based data virtualization server that accesses, federates, abstracts and simplifies data, and delivers the results as data services or relational views to consuming applications such as BI, analytics and business process management. It offers an intuitive web-based interface for business users and IT resources.

Table 1. Ultrawrap

Attributes	Details
Product name	Ultrawrap
Headquarters	Austin, Texas, U.S.
Geographic presence	North America, EMEA
Deployment model	On-premises, cloud or hybrid
Licensing model	The Ultrawrap licensing model is based on database size. Pricing can be provided on an annual subscription, or perpetual, basis
Major verticals	e-commerce, financial services, healthcare, life science, manufacturing
Top use cases	■ Healthcare patient stratification/ identification ■ Master data management ■ E-commerce order management
Notable customers (in alphabetical order)	Arbonne, Ohio State University Wexner Medical Center, Providence Health & Services, Schneider Electric, UCB
Number of deployments	15 to 20
Source: Gartner (July 2016)	

Advanced query optimization algorithms and techniques are key components of CIS. Optimizations include a deep SQL implementation with both rule-based and cost-based optimization; multiple join algorithms to process federated data (including automatic data movement); full relational algebra implementation for plan rewriting, constraint propagation, redundant operator pruning; data source capability framework to leverage native data processing and reduction (that is, pushing down joins, predicates, aggregations, sorting, and so on), as well flexible caching techniques (see Table 2).

CIS also supports a full range of transformation logic that can be done natively or in conjunction with third-party tools, custom code, and so on. The platform supports push-down processing and offers a "data-ship join" that automatically moves data from one data source to another to improve the performance of push-down processing.

Security includes authenticating users and enforcing user and group privileges. Additional features include single sign-

Table 2. Cisco Information Server

Attributes	Details
Product name	Cisco Information Server (CIS)
Headquarters	San Jose, California, U.S.
Geographic presence	North America, EMEA, Asia/Pacific and South America
Deployment model	On-premises, cloud or hybrid deployment options
Licensing model	CIS is sold via perpetual or term license. The production licenses are primarily sold per core. Nonproduction instances are priced per server. Options such as adapter bundles are priced per CIS instance. An AWS hosting option is available and is usage-based
Major verticals	Financial services, telecom service providers, life science/healthcare, energy, industrial products, government
Top use cases	■ Federation in the form 360-degree view of "x" data integration projects where x is a key entity such as customer, research project, positions and trades ■ Data architecture initiatives including LDW, enterprise semantic layer, data services, data virtualization layer, data abstraction ■ Migration in various forms including source migration, consumer migration, cloud migration, merger and acquisition, data mart elimination
Notable customers (in alphabetical order)	AT&T, BP, Comcast, ExxonMobil, General Motors, Goldman Sachs Group, Jaguar Land Rover, JPMorgan Chase, Merck & Co., Pfizer, U.S. Department of Defense (DoD)
Number of deployments	More than 300

AWS = Amazon Web Services; LDW = logical data warehouse
Source: Gartner (July 2016)

on, Lightweight Directory Access Protocol (LDAP), row-level security, token-based user authentication, WS-Security standards for web services, plus data encryption on the wire, Base64, Kerberos, NT LAN Manager (NTLM) and Secure Sockets Layer (SSL).

Data Virtuality

Data Virtuality's technology originates from a research project at the University of Leipzig, Germany. Its data virtualization offering — DataVirtuality — incorporates a self-learning algorithm that intelligently decides which data needs to be moved into a repository and how it should be organized (creating federated materialized views, aggregations, indexes, and so on). It offers more than 100 connectors, which allow companies to fully integrate their database and cloud landscape (cloud databases as well as web service APIs). The product allows execution of tasks such as ETL/ELT, batch import, data movement, slowly changing dimensions, change data capture or data-driven application integration through wizards and SQL scripting.

DataVirtuality provides the ability to create custom virtual data models using virtual schemas, virtual views and virtual stored procedures (see Table 3). It includes basic, intermediate and complex SQL-based transformations using extended SQL language, as well as the ability to do a controlled write back of data into the original systems. DataVirtuality provides two types of performance optimization techniques: federated

Table 3. DataVirtuality

Attributes	Details
Product name	DataVirtuality
Headquarters	San Francisco, California, U.S.
Geographic presence	North America and EMEA
Deployment model	On-premises, cloud or hybrid
Licensing model	Data Virtuality offers a subscription-based licensing model depending on the number of server instances, number of data sources, number of connections to data sources and the types of analytical storage connected to DataVirtuality.
Major verticals	E-commerce, digital marketing, telecommunications, financial services, logistics
Top use cases	■ Cloud/on-premises integration to create a hybrid integration platform ■ Managed data access and modeling layer for self-service business intelligence ■ Data warehouse and data lake integration
Notable customers (in alphabetical order)	amaysim, Audi Group, DHL, Home24, Kohler, windeln.de
Number of deployments	28
Source: Gartner (July 2016)	

planning optimization, and persistence optimization using centralized analytical storage. The product offers full push-down processing for relational data sources such as Hadoop and Salesforce, and partial push-down capabilities for NoSQL DBMSs including MongoDB, Cassandra and Neo4j.

Security features include user/role-based permissions, schema/table/column-based permissions, LDAP-based authentication, two-factor authentication and custom plugin-based authentication. Encrypted connections to data sources and to data consumers (for both ODBC and JDBC) are available.

Denodo

Denodo, an independent data virtualization vendor, provides agile and iterative data integration, data abstraction and real-time data services across enterprise, cloud, external, big data and unstructured data sources. The Denodo Platform (see Table 4) connects to disparate data sources, unifies them into logical/virtual data services and provides unified access through a single "virtual" data layer.

The following capabilities are key to the Denodo Platform: advanced dynamic query optimization supported by intelligent caching and scheduled data orchestration; unified data governance and quality; and the ability to deliver data services in multiple formats with managed security and service levels. The Denodo Platform supports both static rule-based (heuristic) and dynamic cost-based techniques of query optimization. It supports advanced query optimization techniques including full and partial aggregation push down, multiple join strategies, data movement optimization and constraint propagation. The Denodo Platform capitalizes on push-down processing to all supported sources, thus reducing data movement and achieving flexible scalability.

The Denodo Platform supports drag-and-drop modeling and automatic detection of source schema changes and impact analysis. It provides business users with self-service data discovery and exploration, search capabilities for data and metadata to find relevant information, and graphical representation of entities and relationships.

The Denodo Platform offers robust security features, including a built-in role-based access control (RBAC) security model, and integration with Kerberos and LDAP/Active Directory. It supports integration with external entitlement services through policy-based security, pass-through authentication, fine-grained security access restrictions (up to the "cell" level), granular data masking and encryption (including Secure Sockets Layer, HTTPS, MD5 and Data Encryption Standard).

IBM

IBM's data virtualization offerings include IBM InfoSphere Information Server, IBM InfoSphere Federation Server, IBM BigInsights for Apache Hadoop and IBM InfoSphere Master Data Management (see Table 5). All products support the ability to virtualize multisite federated/virtualized queries and consolidate and virtualize the result sets from distributed resources. IBM InfoSphere Information Server is embedded with built-in transformation functions and a common metadata framework to

Table 4. Denodo Platform

Attributes	Details
Product name	Denodo Platform
Headquarters	Palo Alto, California, U.S.
Geographic presence	North America, EMEA, Asia/Pacific, Latin America
Deployment model	On-premises, cloud and hybrid
Licensing model	■ Server/core basis — Perpetual license for production, development, and QA servers is quad core-based. Subscription licensing model is also offered for on-premises, hosted/private cloud and public cloud deployments. All connectors are included in the pricing and are not charged separately. ■ Enterprise licenses — Unlimited licenses (deployments for three-month to 36-month deployment windows; perpetual license thereafter). ■ The AWS version of the Denodo Platform is available in varying configurations based on the number of servers and the number of concurrent queries and result sets. Pricing starts at less than $1,000/month. ■ Denodo Express — Free download, no time limit, no use restrictions (develop, test, production, demo, training), single user license
Major verticals	Financial services, healthcare and life sciences, manufacturing and high tech, telecom, retail and consumer packaged goods, public sector
Top use cases	■ Logical data warehouse/logical data lakes ■ Enterprise data services layer. ■ Single 360-degree view of customer, product, contract or any other domain.
Notable customers (in alphabetical order)	Autodesk, Boeing, CIT, Intel, Logitech, T-Mobile, Zurich Insurance Group
Number of deployments	More than 300

QA = quality assurance
Source: Gartner (July 2016)

support bulk/batch (extract, transform, load), virtual (federated) or incremental (data replication) data delivery styles, as well as an advanced parallel and distributed processing engine. IBM data quality solutions, including InfoSphere Information Analyzer, allow the user to monitor and analyze data quality, enact data rules and detect data classifications for the standardization of data across the organization. IBM's InfoSphere Information Governance Catalog further allows the ability to search, explore and govern information such as data stores and processes, and allows the visualization of the usage or consumption of information. The IBM virtualization offerings support predicate optimization, in which the optimizer decomposes a query into fragments by choosing the best decomposition based on minimum estimated total resource consumption. It offers encryption using signed certificates and supporting certificate stores. Proxy Controls and Global Security Kit provide encryption services for wrappers and user-defined functions.

Table 5. IBM Data Virtualization Solutions

Attributes	Details
Product name	IBM InfoSphere Information Server; IBM InfoSphere Federation Server; IBM BigInsights for Apache Hadoop; IBM InfoSphere Master Data Management
Headquarters	Armonk, New York, U.S.
Geographic presence	Asia/Pacific, EMEA, North America and South America
Deployment model	On-premises
Licensing model	IBM's virtualization offerings are priced based on Processor Value Unit (PVU) for production environments. For all nonproduction environments (including development, testing and quality assurance), IBM's pricing is based on authorized user (AU) and is charged per AU.
Major verticals	Banking, education, retail or government
Top use cases	■ Support for risk data reporting and compliance — identifying the source and quality of critical data elements for audit control ■ Support self-service analytics, in allowing the search and exploration of well-defined, well-understood, quality and governed information via an enterprise Catalog ■ Business-led development through the sharing and reuse of data stores and processes
Notable customers	Not provided by IBM
Number of deployments	More than 2,500 (deployments of InfoSphere Information Server)
Source: Gartner (July 2016)	

Informatica

Informatica offers its data virtualization solution across the Informatica intelligent data platform, Informatica Data Services and Informatica PowerCenter (see Table 6). Its data virtualization solutions provide a single environment for data integration and data virtualization along with role-based UIs and tools that share common metadata. Analysts can rapidly prototype results, and then quickly collaborate with IT to put them into operation. This forms the basis for an end-to-end agile data integration approach where new data integration projects can be delivered faster than with traditional data integration approaches.

Informatica's data virtualization solution supports push-down processing of queries to relational databases, data warehouse appliances, applications and Hadoop sources. Its multithreaded engine uses a combination of processes and threads to scale up in symmetric multiprocessing and grid environments. It provides native connectivity and supports caching for transformations, logical data objects, virtual tables and results sets for optimal processing. Informatica supports a wide range of rule-based and cost-based query optimization techniques such as early selection and projection, semijoin, data-ship join and predicate optimization.

Table 6. Informatica Data Virtualization Solutions

Attributes	Details
Product name	Informatica Data Services; Informatica PowerCenter
Headquarters	Redwood City, California, U.S.
Geographic presence	Asia/Pacific, EMEA, North America and South America
Deployment model	On-premises, cloud (as integrated platform as a service) or hybrid
Licensing model	Informatica's data virtualization solution is priced per core/CPU. Various complementary solutions of the Informatica Platform are available to be purchased either by core/CPU or by user, or even node. OEM pricing varies; any OEM agreement is customized and negotiated individually.
Major verticals	Financial services, healthcare, oil and gas, the public sector
Top use cases	■ Rapid prototyping ■ Data warehouse augmentation ■ Master data management
Notable customers (in alphabetical order)	Credit Suisse, Devon Energy, Prudential, Quintiles
Number of deployments	Not provided by Informatica
OEM = original equipment manufacturer	
Source: Gartner (July 2016)	

Informatica supports authentication and authorization using a hierarchical security model with fine-grained privileges and permissions for role, user, folder, group and repository. It integrates with LDAP, Active Directory and Kerberos, and supports WS-Security. Informatica provides a full collection of built-in data transformations including data quality rules for cleansing data and data masking rules for sensitive fields, and also provides Federal Information Processing Standard (FIPS)-certified encryption for data in flight. Security for federated SQL access includes pass-through and fine-grained column-level and row-level security.

Information Builders

Information Builders is a privately held company that offers software solutions for BI and analytics and data integration. Its iWay Data Hub (see Table 7) is a general-purpose enterprise information management

Table 7. iWay Data Hub

Attributes	Details
Product name	iWay Data Hub
Headquarters	New York, New York, U.S.
Geographic presence	North America and EMEA
Deployment model	iWay Data Hub is normally deployed on-premises. Information Builders also supports cloud, hybrid, and multiple hosting models, such as IaaS, PaaS, and SaaS, as well as virtual cloud hosting and private (dedicated) cloud hosting.
Licensing model	Licensing for iWay Data Hub is on a per-core basis. For a small number of WebFOCUS applications, it is sold on a per-user basis.
Major verticals	Financial services, healthcare, insurance, state and local government, manufacturing and supply chain
Top use cases	■ Complex reporting environments that include operational reporting, data warehousing, and other data sources ■ Data standardization across multiple applications to create a common data services layer ■ Legacy modernization
Notable customers (in alphabetical order)	BNY Mellon, Nationwide, Nomura Securities, Royal Bank of Canada, Scotiabank
Number of deployments	Not provided by Information Builders

IaaS = infrastructure as a service; PaaS = platform as a service
Source: Gartner (July 2016)

engine that leverages the iWay Universal Adapter Framework to access disparate data sources through a single, federated query. The iWay Data Hub and Universal Adapter Framework together provide location transparency, ensure data integration through data virtualization for both relational and nonrelational data stores (NoSQL, for example), and ensure that databases receive well-tuned SQL queries and deliver well-formed results sets.

The platform can also initiate queries from legacy systems; for example, a query from the mainframes accessing Hadoop and multidimensional data using SQL. The iWay Data Hub has the ability to embed calculations, joins, logical views, hierarchy flattening and other data manipulations into the metadata layer. It supports push-down processing, and in some cases (such as mainframes with flat files) it runs an adapter on the target platform; the iWay Data Hub then pushes processing to the iWay adapter on that target platform. For performance optimization, it employs query optimization, native tuning of individual adapters on target platforms, direct pass-through of SQL (where possible), and query rewrites to take advantage of native SQL extensions and caching.

Robust user, group and field-level security — together with integration with Kerberos, Resource Access Control Facility (RACF), TopSecret and many other security environments — is supported.

K2View's Data Fabric is a distributed data management platform that virtualizes and integrates customers' data in near real time, enabling business to deal with complex integration problems involving diverse data silos by creating a single data layer with unified data from across the organization.

The data within K2View Fabric is organized by logical units, which is the organization's core business entity (for example, a customer, an account or a credit card). For each entity, it creates a unique microdatabase that gathers all data about that entity from all data sources across the organization, providing an holistic view of the entity. From a performance perspective, the microdatabases can be distributed across multiple nodes of data centers, allowing for faster query performance, no single point of failure, and an elastic, linear scale-out architecture. The organization of data in microdatabases, and the in-memory processing capabilities of distributed virtualized queries, can enhance the query performance.

The platform provides these data virtualization capabilities through a combination of distributed data management capabilities, data integration, data encryption, data compression and web services, all accessed via a business-user-friendly configuration studio. In addition, each microdatabase is uniquely encrypted, enabling a level of security that prevents data breaches. This encryption is achieved using a patent-pending proprietary algorithm called Hierarchical Encryption-Key Schema (HEKS).

K2View Fabric (see Table 8) offers full flexibility over roles definition, allowing granular data access control. Administration can restrict users' access to all data levels such as logical unit types, instances, methods, web-services and specific data. K2View Fabric's data masking functionality allows persistent and dynamic data masking and obfuscating data by configuring masking rules on sensitive information as part of the virtualization process.

Table 8. K2View Fabric

Attributes	Details
Product name	K2View Fabric
Headquarters	Atlanta, Georgia, U.S.
Geographic presence	North America and EMEA
Deployment model	On-premises, cloud and hybrid
Licensing model	K2View offers both subscription and perpetual licensing options. Pricing is based on the logical unit count and volume. The solution pricing does not limit the customer on the number of data sources and customers are allowed to add as many data sources as needed for data virtualization.
Major verticals	Communications, financial services (banking, retirement services, insurance) and healthcare.
Top use cases	■ Personalization — 360-degree view of the customer ■ Regulatory compliance ■ Test data management
Notable customers	Not provided by K2View
Number of deployments	More than 50
Source: Gartner (July 2016)	

OpenLink Software

Virtuoso Universal Server from OpenLink Software is an enterprise-grade solution for data access, data integration, data virtualization and multimodel relational database management (see Table 9). The unique hybrid server architecture of Virtuoso enables it to offer traditionally distinct server functionality within a single product offering. As a data virtualization platform, the product supports push-down processing to connected, third-party RDBMSs and other services/sources that support standards-based data access protocols (ODBC, JDBC, and HTTP) and data definition and manipulation languages such as SQL and/or SPARQL.

Security features include unique identification of entities using uniform resource identifiers (URIs; typically HTTP scheme-based, but others such as LDAP are also supported), support for various open standards-based authentication protocols (including OAuth, OpenID, Transport Layer Security [TLS], WebID+TLS, WebID+TLS+Delegation, digest authentication) and fine-grained attribute-based access controls via policies based on entity-relationship-type semantics.

Table 9. Virtuoso Universal Server

Attributes	Details
Product name	Virtuoso Universal Server
Headquarters	Burlington, Massachusetts, U.S.
Geographic presence	North America and EMEA
Deployment model	On-premises, cloud or hybrid
Licensing model	OpenLink offers two licensing options: ■ Open Source Edition — Perpetual license with no licensing cost, optional support subscriptions ■ Commercial Edition — Flexible licensing model charged per host environment (physical or virtual); typically perpetual, may be time limited; typically limited by concurrent usage (client hosts, client sessions, logical and/or physical processor counts, CPU affinity where possible). No license limits on volume, traffic and data sources.
Major verticals	Bioinformatics/pharma/biotech, government and others
Top use cases	■ Integrating older, mission-critical internal systems with newer customer-focused systems ■ Integrating data hosted across multiple vertical solutions (such as HR, manufacturing, cash-register) to present a single, unified conceptual view of the data ■ Ingesting data from external databases, merging, cleansing and then republishing in Linked Open Data form (internally or externally)
Notable customers (in alphabetical order)	Bank of America, Bloomberg, Daimler, Elsevier, EU, Globo, Sanofi, Syngenta Group
Number of deployments	112 across EMEA; 863 across North America and South America
Source: Gartner (July 2016)	

Oracle

Oracle offers three products for data virtualization: Big Data SQL, Oracle Data Service Integrator (ODSI) and Big Data Preparation (see Table 10). Big Data SQL is a solution for SQL and other data APIs on disparate datasets, seamlessly integrating data in Apache Hadoop, NoSQL databases and Oracle Database.

ODSI provides data virtualization by providing data services on top of heterogeneous, distributed data. It allows logical compositions of these sources through joins, filters and other transformations. Big Data SQL supports performance and query optimization. ODSI maps sources into targets with complex transformations of flat or nested data without coding.

It also provides full support of updates through logical views and creates update maps automatically.

Big Data SQL optimizes performance by controlling query execution from end to end. It executes scans and input/output (I/O) on local nodes by deploying an agent on each data node. The agent implements I/O reduction and uses Hadoop, NoSQL formats to further reduce I/O latency. Smart scan optimizes joins locally and reduces data moving to the engine. ODSI supports a distributed cache mechanism, which pushes down transformation logic to data sources to minimize midtier processing and optimizes query results for streaming or buffering. It also applies proprietary join algorithms to midtier results.

Big Data SQL enables Oracle Database security on any data it can query. This functionality ensures that data accessed via Big Data SQL enables a single place for security definitions. Various well-known features — such as virtual private database, redaction, and row and column level security are available across all data virtualized by Big Data SQL.

Progress

Progress provides the Progress DataDirect connectors for data virtualization that offer data connectivity and integration solutions to better manage and integrate data across relational, big data and cloud data sources

Table 10. Oracle Data Virtualization Solutions

Attributes	Details
Product Name	Big Data SQL; Oracle Data Service Integrator (ODSI); Big Data Preparation
Headquarters	Redwood Shores, California, U.S.
Geographic presence	North America and EMEA, Asia/Pacific, South America
Deployment model	On-premises
Licensing model	Big Data SQL is licensed per disk drive, and is available on commodity hardware as well as on Oracle Engineered Systems (Big Data Appliance). ODSI is licensed by server processor count.
Major verticals	Financial services and communications, retail, manufacturing and healthcare
Top use cases	■ Logical data warehouse ■ Active archives for analytical systems ■ Data services layer — providing data abstraction for service-oriented architecture and application developers from changes in a heterogeneous data topology
Notable customers	Not provided by Oracle
Number of deployments	Not provided by Oracle
Source: Gartner (July 2016)	

(see Table 11). A unified API connects to many disparate data sources, allowing connectivity from existing data virtualization and federation tools — to expand their reach. The product features include compensation for web services quality of service limitations, normalization of nonrelational data to relational form, and providing a metadata layer for sources that do not provide one (for example, SaaS and NoSQL).

The Progress DataDirect Cloud product can be used to virtualize external data for SaaS applications such as CRM-federation solutions around Salesforce and Oracle Service Cloud, both of which provide a real-time Open Data Protocol (OData)-based REST consumer. The product enables push-down processing to underlying data sources, where it is supported. It also emulates full SQL functionality for sources that do not natively support certain constructs. Caching, query optimization, bulk data movement and push downs enable performance optimization.

Security features include Open SSL encryption, Kerberos authentication and NTLM authentication.

Red Hat

Red Hat JBoss Data Virtualization is a data integration solution that provides near real time and unified data access across disparate sources to multiple applications and users. The platform enables agile data provisioning and facilitates the creation of

Table 11. Progress Data Virtualization Solutions

Attributes	Details
Product name	Progress DataDirect; Progress DataDirect Cloud
Headquarters	Bedford, Massachusetts, U.S.
Geographic presence	North America and EMEA
Deployment model	On-premises, cloud and hybrid
Licensing model	Licensing options include core, server, per user, number of sources, subscription, perpetual, term unlimited and royalty
Major verticals	Technology and software companies, financial services, healthcare, retail and manufacturing
Top use cases	■ Legacy CRM/ERP integration ■ Standard connectivity for nonrelational data sources ■ Data federation and ingestion for big data projects on Hadoop or Spark
Notable customers (in alphabetical order)	Intuit, MicroStrategy, NetSuite
Number of deployments	Progress connectivity is embedded within the application of independent software vendors and other enterprises worldwide
Source: Gartner (July 2016)	

reusable data models and virtual unified views by combining and transforming data from multiple sources (see Table 12). It makes integrated data available on demand for consumption by external applications through open standards interfaces. An integrated dashboard builder enables quick data visualization and dataset validation, and facilitates collaboration between business and IT.

JBoss Data Virtualization (JDV) has a full stack of capabilities that checks for queries against relational and nonrelational sources, and when these capabilities are matched the query is then either pushed to be evaluated at source or the processing is done at the JDV query engine. JDV supports caching and is cluster-aware. Query optimization is done through capabilities of sources and costing information provided on the source schemas; it supports various advanced techniques such as dependent joins, partition over a join, project pruning and others.

Security features include authentication based on Kerberos, OAuth, Security Assertion Markup Language (SAML), LDAP, WS-Username Token and HTTP Basic, role-based access control, SSL encryption and others.

Table 12. Red Hat JBoss Data Virtualization

Attributes	Details
Product name	Red Hat JBoss Data Virtualization
Headquarters	Raleigh, North Carolina, U.S.
Geographic presence	Asia/Pacific, EMEA North America and South America
Deployment model	On-premises, cloud and hybrid
Licensing model	JDV is sold through a stand-alone subscription model. The subscription model is based on either 16-core or 64-core price bands; cores can be physical or virtual. JDV subscription includes JBoss EAP server runtime, JBoss Developer Studio and 25 development users for a 16-core subscription. It also includes connectors for all supported data sources.
Major verticals	Financial services, telecom, public sector and transportation
Top use cases	■ Unified multidimensional (360-degree) view for analytical and operational use ■ Service enablement of data for agile service-oriented architecture, Data APIs and data-as-a-service ■ Data firewall for improved data security, access and audit control
Notable customers (in alphabetical order)	ANSES, KDDI R&D Laboratories, RBS
Number of deployments	Not provided by Red Hat
Source: Gartner (July 2016)	

Rocket Software

Rocket Software's Rocket Data Virtualization is a mainframe-based data integration solution that virtualizes data from multiple disconnected sources, on and off mainframe, to be combined into a single, logical data source that can be accessed in real-time by a broad range of data consumers utilizing built-in interfaces for SQL, NoSQL, services, web and events. An Eclipse-based integrated development environment is provided for data discovery, creation of virtual tables/views, metadata management, interface testing and diagnostics (see Table 13).

Rocket Data Virtualization includes a multithreaded z/OS runtime to integrate and transform data, performing nearly all of its data processing within the IBM z Systems' Integrated Information Processor for reduced mainframe CPU consumption and reduced mainframe total cost of ownership. It also has the ability to support hybrid cloud or mobile for z Systems environments with a MongoDB API (JSON) providing real-time access between MongoDB and a wide range of mainframe data sources. The product supports IT operations analysis (ITOA) via its catalog of premapped system management file (SMF) types available via SQL interfaces. It supports push-down processing and also supports both systems-level security and native database-level security.

Table 13. Rocket Data Virtualization

Attributes	Details
Product name	Rocket Data Virtualization
Headquarters	Waltham, Massachusetts, U.S.
Geographic presence	North America and EMEA
Deployment model	On-premises within an IBM z Systems z/OS runtime
Licensing model	Rocket Data Virtualization is sold as a stand-alone offering inclusive of all interfaces and data sources. It provides a perpetual licensing model that is priced based on the mainframe model and its MSU rating. The customer has the option to purchase an enterprise license that is tied to the overall mainframe machine rating, license the product based on a mainframe logical partition specific to the data virtualization project, or license the product specific to the consuming application that is interacting with the mainframe data.
Major verticals	Financial services (banking, insurance, credit card services)
Top use cases	■ Analytics, real-time support for operational, discovery and cognitive analytics ■ Optimized data movement — extract, virtualize/transform and load, as a replacement for traditional ETL ■ Transactional data access — in support of web, mobile and the cloud
Notable customers	Not provided by Rocket Software
Number of deployments	More than 100

MSU = millions of service units
Source: Gartner (July 2016)

SAP

The SAP Hana platform (which is an in-memory DBMS column store delivered as an appliance on various hardware platforms) offers enterprise information management services in one native, unified framework that supports all methods of data access and integration (see Table 14). SAP Hana Smart Data Access (SDA) provides an easy-to-deploy solution to get near-real-time data visibility across fragmented data sources. It also allows access to remote data as if the data was stored in local SAP Hana in-memory tables, without copying or physically moving the data.

SAP customers use SAP Hana SDA, leveraging the in-memory processing capabilities of SAP Hana to quickly build and deploy high-performance, data-location-agnostic applications. A federation optimizer delivers query push down (push-down filters, aggregates, semijoins, and so on), the ability to push data to a remote server, functional compensation, functional translation, parallel execution, and remote caching in Hadoop.

The product supports all Hana-native security features; single sign-on via Kerberos constrained delegation for Hana-to-Hana SDA connections is also supported.

Table 14. SAP Hana Platform

Attributes	Details
Product name and version	The SAP Hana platform, which includes key capabilities and applications supporting data virtualization including: SAP Hana Smart Data Access, SAP Hana Smart Data Integration, SAP Hana Smart Data Streaming, SAP Hana Smart Data Quality and SAP Agile Data Preparation (which is a self-service data preparation application that leverages the SAP Hana platform).
Headquarters	Walldorf, Germany
Geographic presence	Asia/Pacific, EMEA, North America and South America
Deployment model	On-premises, cloud and hybrid
Licensing model	SAP Hana SDA is included free of charge with purchase of SAP Hana platform. SAP Hana Smart Data Integration (SDI) is also included with many Hana editions at no extra charge. Customers can extend the license to full-use, bidirectional data movement, as well as SAP Hana smart data quality for a complete integration (bulk/batch, virtualization, real-time replication) and quality solution
Major verticals	Consumer products, retail, oil and gas, banking, high tech, industrial machinery, public sector
Top use cases	■ On-premises to cloud virtualization ■ Data access for analytics purposes from a variety of databases ■ Exploration of big data sources such as Hadoop
Notable customers (in alphabetical order)	Cardinal Health, Shell,
Number of deployments	More than 250
Source: Gartner (July 2016)	

SAS

SAS Federation Server allows clients to virtualize, cleanse, mask and share data from multiple relational and nonrelational data sources in a secure, controlled manner. Though having existing SAS solutions is not a requirement, companies that already use the SAS platform can seamlessly integrate with the product. SAS Federation Server generates near-real-time data quality using SAS Data Quality functions based on prebuilt rules. A centralized, web-based administration console allows users to graphically administer, monitor and maintain connections and data caches. It also enables centralized governance policy enforcement.

SAS Federation Server supports federated queries and facilitates push-down of data integration jobs to back-end data sources via both FedSQL (an ANSI SQL-based query language) and — when extended with the In-Database Code Accelerator — with SAS DS2 (an advanced data transformation language). This includes smart parsing of federated queries and push-down of the corresponding subqueries when data comes from heterogeneous data sources. FedSQL push-down is supported to a different extent across all the platforms for which Federation Server has data drivers. SAS DS2 code can be run multithreaded or pushed down to the database when the additional SAS In-Database Code Accelerator is available (that is, for Hadoop, Teradata and Pivotal Greenplum).

The product provides native support for authentication, authorization, encryption, hashing, randomization, data quality, persistent data masking and dynamic data masking. SAS Federation Server offers native row, column, and session-level security. Extensive logging of data and metadata access is supported for auditing purposes (see Table 15).

Stone Bond Technologies

Stone Bond Technologies offers an enterprise-grade data virtualization platform, the Enterprise Enabler. The platform enables federation of various popular data sources and transformation and packaging for virtualization in a single click. For any sources that can handle queries themselves, the query is pushed to the endpoint for efficient execution. The platform enables data curation and governance with capabilities to trace data lineage visually through its business-user-friendly GUI, which can be easily picked up by citizen integrators. Enterprise Enabler is metadata-driven and all metadata is versioned with time stamp and username. The metadata is automatically generated by the integrated development environment (IDE) and can be exported in an XML format or via an API as desired.

The platform differentiates through its one-click, write-back feature and click optimization, which allows users to select all or part of the data to be stored in-memory or in a physical database for storage and query optimization. Virtual models have the ability to be indexed and keyed and can also be defined based on queries or schemas depending on the nature of the database and the level of optimization that may have been performed on the source (see Table 16).

Security is embedded at different levels within the platform including SSL, RSA, Rijndael, 3DES, single sign-on and Active Directory LDAP integration. Built-in support is available for regulations such as PCI DSS v3.2 sections, HIPAA Subpart C 164.306 and FIPS 140-2, Level 1 — Basic security requirements specified for a cryptographic module.

Table 15. SAS Federation Server

Attributes	Details
Product name	SAS Federation Server
Headquarters	Cary, North Carolina, U.S.
Geographic presence	EMEA, North America, South America, Asia/Pacific
Deployment model	On-premises and cloud
Licensing model	Server-based pricing model charged by the number of processor cores. This includes a set of the most commonly used database drivers: Base SAS, DB2 CS, ODBC, Oracle, PostgreSQL, and SAS HDAT (a proprietary Hadoop file format that speeds read operations on data). Additionally, drivers are available for Apache Hadoop (Apache Hive), Pivotal Greenplum, IBM Netezza, SAP, SAP Hana and Teradata for an additional cost per driver.
Major verticals	Banking, financial services, communications, retail, insurance, government, capital markets, manufacturing, life sciences, healthcare, energy, utilities
Top use cases	■ Data warehouse modernization ■ Dynamic data masking-based use cases ■ Auditing and data abstraction-based use cases
Notable customers	Not provided by SAS
Number of deployments	195
Source: Gartner (July 2016)	

Table 16. Enterprise Enabler

Attributes	Details
Product name	Enterprise Enabler
Headquarters	Houston, Texas, U.S.
Geographic presence	North America, EMEA and Asia/Pacific
Deployment model	On-premises, hybrid and cloud
Licensing model	Provision for both subscription and perpetual pricing. In both cases, the pricing has a server element and a source element. Licensing may also be provided with restrictions to limit the use to specific projects by applying restrictions to numbers of processes, maps and virtual models that can be run in a given time period.
Major verticals	Energy, technology, healthcare, finance, logistics, federal government (FED)/state and local government and education (SLED), manufacturing, hospitality/retail
Top use cases	■ Business intelligence/dashboards/reporting and logical data warehouse ■ Portals ■ Master data management
Notable customers (in alphabetical order)	Baylor College of Medicine, CenterPoint Energy, MD Anderson Cancer Center, Schlumberger, U.S. Department of Defense, Weatherford, WCA
Number of deployments	248
Source: Gartner (July 2016)	

Market Recommendations

Data and analytics leaders must:

- Consider data virtualization capabilities as important components of an overall data integration portfolio, seeking ways in which they can extend existing data integration architectures.
- Evaluate the current state of your data integration architecture state. Set proper expectations upfront, select the right use cases and document the agreed-upon service-level agreements to separate out when to collect data versus simply connecting to it, before starting your data virtualization journey.
- Be cautious about assessments and procurement decisions based on information that is more than a year old, because current and emerging data virtualization platforms are maturing fast in terms of performance and security.
- Use this Market Guide as one input to your evaluation when considering data virtualization tools, and assessing vendors' support for your current use case, experience in your vertical or domain, provision of flexible deployment models (and licensing options), plus the overall maturity of the tools themselves in terms of security and performance.

Acronym Key and Glossary Terms

AWS	Amazon Web Services
BI	business intelligence
ELT	extraction, loading and transformation
ETL	extraction, transformation and loading
iPaaS	integrated PaaS
JDBC	Java Database Connectivity
JMS	Java Message Service
JSON	JavaScript Object Notation
LDAP	Lightweight Directory Access Protocol
LDW	Logical data warehouse
MDM	master data management
NTLM	NT LAN Manager
ODBC	Open Database Connectivity
PaaS	platform as a service
RDBMS	Relational database management system
REST	Representational state transfer
SOAP	Simple Object Access Protocol
SSL	Secure Sockets Layer
SVT	semantic virtual tier

Evidence

The findings in this report draw on:

- A representative vendor survey on data virtualization tools conducted by Gartner during 2Q16.

Of the 22 vendors invited to participate, there were 16 survey submissions describing products that met the definition of a data virtualization tool (as described in the Market Definition section). Vendors who did not respond to the survey in time or that otherwise declined to participate (for various reasons) include Attunity, Cirro, Microsoft, Palantir, Talend and VirtDB. These vendors have therefore not been included in this version of the Market Guide. The information we received through the survey was augmented (where appropriate) by publicly available information (such as that from vendors' websites) and our interactions with Gartner clients and at vendor briefings.

Due to the fluid nature of this market, no references were contacted for this research. Due to the timing of the survey, the results submitted are considered to provide a snapshot as of the end of June 2016.

Vendors received a copy of the information for factual review prior to publication.

- Gartner client inquiry data (we take more than 700 inquiries on data virtualization annually). During the past two years (from 2014 through 2016, so far), the number of inquiries about data virtualization/data federation that Gartner has received from clients has been consistently high. In both years, they have exceeded 700, with more than 150 inquiries per quarter. Inquiries have come from a very wide range of industries, and from organizations with varying levels of maturity.

- Gartner analysts' assessment of product capabilities across a range of data integration, DBMS and data virtualization tools, from 2013 to 2016.

- Calls with product reference customers conducted by Gartner analysts.

- A survey of reference customers conducted for Gartner's 2015 study of data integration tool adoption and usage (the 2016 data integration tool survey was also being conducted in parallel, but was not yet ready at the time of writing).

Note 1. Data Integration Delivery Styles Other Than Data Federation/Virtualization

Data integration tools typically support a combination of the following main data delivery styles:

- Bulk/batch data movement involves bulk and/or batch data extraction and delivery approaches (such as support for ETL processes) to consolidate data from primary databases and formats. This delivery style draws on data from across system and organizational boundaries. It can play a role in all the use cases examined in this report.

- Message-oriented movement encapsulates data in messages that various applications can read, so that they can exchange data in real time. This delivery style can play a role in all the use cases examined in this report. 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 enterprise service bus.

- Data replication and synchronization synchronizes data to enable, for example, change data capture between two or more DBMSs, schemas and other data structures, whether they are of the same type or different. This capability supports high-volume and mission-critical scenarios by keeping operational data current in multiple systems.

Source: Gartner Research Note G00299299, Ehtisham Zaidi, Mark A. Beyer, Shubhangi Vashisth, 25 July 2016

Big Data and Big Iron: How Mainframes are Changing Analytics with Data Virtualization

Ready or not, Big Data is transforming business. The world's most successful companies have already figured out how to leverage data and analytics to dominate their industries, and many others are in the process of catching up. As for organizations that aren't...well, they might as well get ready to join the extinct ranks of dinosaurs and Betamax machines. That's because the advantages of analytics – predictive, preventive, and proscriptive – are simply too important for any enterprise to pass up.

Big Data may already be a reality in many industries, but it faces significant challenges as it emerges into the forefront of strategic planning. Business leaders need to be able to **access massive amounts of data easily** – no matter where it is – and access only pertinent information rather than comb through every bit and byte looking for the needle in the haystack. In an on-demand world **speed is critical** because business partners and customers demand instant response times.

Security is another major issue because massive amounts of data are the motherlode for hackers looking to do everything from steal identities to illegally transfer digital assets.

Big Data and Mainframes

Organizations that rely on mainframes – including most banks – have been living in the world of Big Data for years. Long before other industries had to crunch huge amounts of information, banks were already combatting a tsunami of information that had to be processed and analyzed quickly. That's why they turned to mainframes, which offer unparalleled power,

Source: IBM z13 mainframe

speed, and security. That's still the case today, more than 50 years after the first mainframe was deployed.

No other hardware in the world has the sheer storage capacity and processing power of a mainframe. These devices are able to churn a truly unprecedented amount of information in a very short time, reducing long projects to short ones by performing millions upon millions of operations per second. For this reason, mainframes are not just one solution for data, they're the solution, with speed, power, and toughness combining to solve the key problems that keep CIOs up at night.

That's where data virtualization (DV) comes into play.

We no longer have the luxury of waiting for data. We live in a world where people expect instant access to everything from restaurant reservations to sports scores, and they won't settle for delays. Long gone are the days of depositing a check at the local bank and waiting a week for it to show up in our bank accounts: today, transactions need to happen at the speed of a mouse click. Unfortunately, the traditional method of doing analytics can slow everything to a crawl because data needs to be taken off the main system, fed into another system, crunched, and then get fed back into the original system. Not only is this Extract, Transform, Load (ETL) process inefficient, but it opens the door for errors and security breaches. That's exactly what Data Virtualization eliminates because it allows analytics to be done natively on the mainframe.

Why Data Virtualization Matters

A mainframe can perform countless operations per second, crunching data at an astounding rate. An operation that would take hours on many systems takes only seconds, and a project that usually takes months takes days, turning what might've been a long-term program into a readily available piece of information whose findings can be plugged into company operations with a truly rapid turnaround. And storing all of a company's information in a single mainframe bank means that information is immediately available for analysis, without the time necessary to download it from the cloud or transfer it from another location.

Data virtualization is a major part of what makes mainframes such a powerful home for analytics. By pulling only the necessary information out of a sea of data, virtualization drastically shortens the amount of time every process requires. It's like trying to cook five artichoke hearts when you have a billion artichokes:

virtualization instantly plucks the hearts out of the artichokes without bothering with the rest of the plant. By paring down information only to what is pertinent to an operation, data virtualization saves time that would otherwise spent sifting through the dross. Useful on any system, on a mainframe this operation sets the processor's muscle free to work on important data analytics, not moving bits and bytes around.

Safe as Houses

There are many security advantages to housing all a company's data on a mainframe. This advice might sound like being told to put all of one's eggs in one basket. But when the basket is a bank of incredibly powerful computers, the prospect becomes much more tempting. Banks and financial organizations have long known that the speed of a mainframe is the best way to stop security fraudsters who rely on slow processing times to take the money and run before anyone is the wiser. But it's against the greater threat of data breaches and identity theft that Big Iron, powered by DV, truly shines.

Big Data is already generating enough information to fill up the Library of Congress every 14 minutes, and black-hat hackers have been able to breach dozens of major companies to invade the privacy of millions of consumers. Data spread around a series of systems can be picked off piecemeal through minor, occluded breaches that are easy to overlook. But data stored on a single mainframe (or bank of mainframes) is far more difficult to breach. Eliminating the "open doors" that ETL creates makes DV the right choice for improving security and reducing the risk of unauthorized access to critical data.

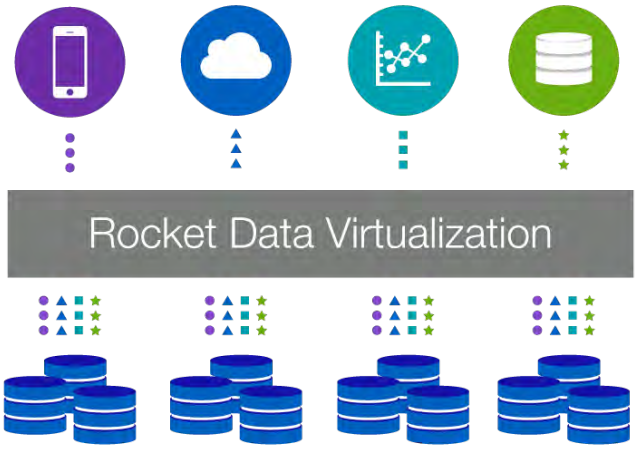
Analytics requires lots of big brains – or more specifically, a great many powerful processes – as we get deeper into the Information Age every day. All of the data we're creating needs to be processed quickly and kept secure, and new techniques like data virtualization are the best approach to make that a reality.

Source: Rocket Software

About Rocket Data Virtualization

Without real-time access to mainframe data, companies don't have a comprehensive picture of their customers, which limits offering enhanced customer service and real-time target marketing. Every credit card swipe or online purchase provides real-time customer data. When enriched with sources such as social media, location and brand preference, it can enable firms to anticipate customer preferences.

Rocket® Data Virtualization (RDV) creates virtual, integrated views of data, and enables users to access mainframe data in place without having to move, replicate, or transform it. By bringing analytics closer to the data, RDV saves companies time and money. They can immediately combine mainframe data with data from other enterprise sources to gain real-time insight into potential risks, customer needs, and market opportunities.



RDV offers a cost-efficient option to either replace ETL altogether, or serve as a "data utility" that optimizes existing ETL processes to provide real-time data in place of large batch jobs. Because RDV runs almost exclusively on the z Systems Integrated Information Processor (zIIP), it doesn't consume mainframe MIPS capacity, and can significantly reduce mainframe cost associated with ETL.

Most millennial software developers lack the familiarity and programming skills needed to work with mainframe data. Working in combination with IBM® z/OS® Connect Enterprise Edition (EE), Rocket Data Virtualization (DV) gives mobile and cloud application developers the ability to integrate z/OS data into their applications without the need for any mainframe expertise or changes to mainframe code.

Rocket®

Contact us

Rocket Software
77 Fourth Avenue
Waltham, MA 02451
USA
+1 781 577 4321

www.rocketsoftware.com
info@rocketsoftware.com
sales@rocketsoftware.com

Data Virtualization: Shine the Light on Dark Data is published by Rocket Software. Editorial content supplied by Rocket Software is independent of Gartner analysis. All Gartner research is used with Gartner's permission, and was originally published as part of Gartner's syndicated research service available to all entitled Gartner clients. © 2017 Gartner, Inc. and/or its affiliates. All rights reserved. The use of Gartner research in this publication does not indicate Gartner's endorsement of Rocket Software's products and/or strategies. Reproduction or distribution of this publication in any form without Gartner's prior written permission is forbidden. The information contained herein has been obtained from sources believed to be reliable. Gartner disclaims all warranties as to the accuracy, completeness or adequacy of such information. The opinions expressed herein are subject to change without notice. Although Gartner research may include a discussion of related legal issues, Gartner does not provide legal advice or services and its research should not be construed or used as such. Gartner is a public company, and its shareholders may include firms and funds that have financial interests in entities covered in Gartner research. Gartner's Board of Directors may include senior managers of these firms or funds. Gartner research is produced independently by its research organization without input or influence from these firms, funds or their managers. For further information on the independence and integrity of Gartner research, see "Guiding Principles on Independence and Objectivity" on its website.