

# Data modernisation using a logical data management platform

Received (in revised form): 10th May, 2023

## Charles Southwood

Regional VP, Denodo Technologies Ltd, UK



Charles Southwood

**Charles Southwood**, Regional VP at Denodo Technologies, is responsible for the company's business revenues in Northern Europe, Middle East and South Africa. He is passionate about modernizing the data architectures of customers, using data virtualisation, data fabric and data mesh, to help them meet changing business needs. With a degree in engineering from Imperial College London, Charles has over 30 years' experience in data integration, big data, IT infrastructure/IT operations and business analytics.

### ABSTRACT

Like many industries, the capital markets industry relies on reliable access to trustworthy data. For this industry in particular, it is becoming critical for success, if not survival. This type of data access, however, requires modern data infrastructures, and not many organisations are prepared to furnish such infrastructures. This paper introduces ways to modernise the data infrastructure without ripping out and replacing expensive hardware. It covers key capabilities of modern data infrastructures and compares and contrasts emerging technologies such as cloud data warehouses, data lakes, data virtualisation, logical data fabric and data mesh. The paper provides a business case for complementing investments in centralised data management tools, such as cloud data warehouses or data lakes, with a logical data management approach.

**Keywords:** data management, business agility, data lakes, data warehouse, cloud, big data, data fabric

### INTRODUCTION

Data is the lifeblood of any organisation, but in the capital markets industry, this is

probably truer than in most other industries. Capital markets organisations need reliable access to curated, high-quality data, whether for client details, pre-trade or following securities transactions from the time of execution through settlement. Likewise, the industry is seeking to improve back-office operations, including service levels, workflow and third-party processing. The aggregation of effective, accurate and timely data is often easier spoken about than accomplished, however. This is typically due to several challenges:

- Rapidly growing volumes of structured and unstructured data;
- The distribution of data across different systems, with new data sources being regularly added to the mix;
- The pressure to process data — and make use of it — more and more quickly;
- The hesitation to share entire datasets that contain sensitive data, due to the risks of the unapproved data disclosure (with a number of payment protection insurance [PPI] regulations including General Data Protection Regulation [GDPR], Protection of Personal Information Act [POPI], etc.);
- The inability to find workers with the proper skills to address more complex data demands, particularly around artificial intelligence (AI)/machine learning (ML) and advanced data science.

Financial institutions are making huge investments to modernise data architectures to overcome these and many other challenges, yet data is still underutilised. In the Denodo

Denodo Technologies Ltd,  
1st Floor,  
10 Eastbourne Terrace,  
London,  
W2 6LG,  
UK  
Mob: +44 (0)7501  
842965;  
E-mail: csouthwood@  
denodo.com

Journal of Securities Operations  
& Custody  
Vol. 15, No. 4, pp. 336–347  
© Henry Stewart Publications,  
1753–1802

Global Cloud Survey Report 2022,<sup>1</sup> 62 per cent of respondents indicated they could not find, access and analyse over half of their data.

### ARE THE INVESTMENTS BEING MADE IN THE RIGHT PLACES?

Two of the most frequently listed technologies being prioritised currently are cloud data warehouses and data lakes. Depending on their requirements, organisations will likely require both a cloud data warehouse and a data lake, as they serve different needs and use cases.

- *Cloud data warehouses* are the modern versions of the traditional on-premises data warehouse investment that banks have relied on for 20 or more years. A data warehouse is a database optimised to analyse relational (structured) data from transactional systems and line-of-business applications. The key difference between the modern cloud versions of these data warehouses is that they are far less complex to set up and scale, as the cloud provider handles those aspects;
- *Data lakes* store relational data from line-of-business applications and non-relational data from mobile apps, Internet of Things (IoT) devices and social media feeds. The structure of the data is not defined when data is captured, so organisations can store all of their data without careful design or the need-to-know what questions might need to be asked in the future. Data lakes support a wide variety of analytics on data, such as structured query language (SQL) queries, big data analytics, full-text search, real-time analytics and ML, which can all be used to uncover insights.

While both technologies support analytics use cases, they typically support different ones. Business analysts typically use a data warehouse; with its structured and curated

datasets it provides for batch reporting, business intelligence or data visualisations. On the other hand, data lakes have largely uncurated data and are therefore more commonly used for exploration, for data discovery and for new insights, by data scientists and developers for machine learning and predictive analytics.

Both of these technologies are core capabilities in cloud data modernisation efforts, as they remove the constraints of traditional on-premises deployments and simplify scaling; however, these alone are not enough to address today's data challenges.

### WHY ORGANISATIONS NEED MORE THAN CLOUD DATA WAREHOUSES AND DATA LAKES

The technologies above operate on the same philosophy; organisations must first collect and physically consolidate data in a central repository before they can derive value from it. For many use cases, this may be a viable approach. Still, for several reasons, data will remain distributed:

- *One size never fits all:* There are many types of data platforms, such as operational relational database management systems (RDBMS), noSQL sources, data lake engines and enterprise data warehouses (EDWs). Amazon Web Services (AWS), for example, offers over 15 purpose-built engines to support diverse data models, each specialising in specific use cases. The fact is no one data platform will support all of these use cases;
- *Extract, transform and load (ETL) processes take time and effort:* Data is not organically 'born' in the final system, and data pipelines need to be created, updated and maintained. Each step in the process has multiple sub-steps and can be very complex and time-consuming;
- *Consolidation may be prohibited:* Data privacy and data sovereignty rules are evolving.

Certain countries have limitations on data transmission outside the original country. In addition, some countries have privacy laws restricting personal data disclosure to third parties. By law, companies conducting business in these countries could be prohibited from transferring their data or sending it to a third-party provider for storage or processing.

### **A MODERN DATA ARCHITECTURE MUST EMBRACE A DISTRIBUTED DATA LANDSCAPE**

For over 30 years, capital market companies have been using data warehouses built on centralised data architectures. Over this time, however, few, if any, of these companies have successfully consolidated all of their data in one single platform; most will still evidence this continued ambition among their technology priorities.

This could be for one or more of the reasons above, or for many other reasons. While the modern cloud-based versions of these platforms remove the constraints of traditional on-premises deployments, simplify scaling and are key components of a modern data architecture, the same reasons that made it impractical to consolidate everything to a single platform in the past still hold true today. To be successful in modernizing a data architecture, organisations must add logical data management to their data strategies.

### **THE MISSING COMPONENT: LOGICAL DATA MANAGEMENT**

A modern data architecture must balance the physical *collecting* of data, as occurs in a centralised architecture, with the logical *connection* to data, which occurs in a decentralised architecture. Logical data management provides a unified data access layer across all enterprise data assets that enables immediate access to any dataset without needing to first

copy or replicate it. The benefits of logical data management include:

- *Ease of use:* Consumers can use their existing business intelligence (BI) tools or applications to connect to the single logical data access layer, from which any data can be abstracted and/or combined;
- *Agile data integration options:* One platform supports the full range of data integration options, including full replication and transformations, caching and real-time federation;
- *Centralised security and governance:* Access control and policy implementation are applied at the single logical data access layer, ensuring consistency;
- *Futureproof:* By decoupling access to data from the underlying data sources, IT is free to change underlying data stores without affecting business users. When changes are made, updating a single entry in the logical data access layer automatically directs data consumers to the new location without them even being aware. Business can operate at the speed they need, and IT can make changes at the speed they need.

### **THE EVOLVING ECOSYSTEM OF LOGICAL DATA MANAGEMENT**

The logical data management space is not new, with many established technologies and new product categories. This section covers several key product categories and capabilities.

First, we will define several product categories in the space. We will then compare the capabilities typically offered by products in each category, followed by a description of each capability.

#### **Product categories**

- *Query accelerators* improve the performance of queries run against large, complex datasets stored in a data lake. Query

accelerators aim to speed up queries by optimising query execution or reducing the amount of data that needs to be scanned. Some query accelerators will support distributed queries, enabling queries to include relational databases as well. Query accelerators typically use a query architecture similar to classic massively parallel processing (MPP) database management systems. Query accelerators optimise query execution and performance at the storage layer, typically integrated with specific data lakes, and they often require specialised skills to operate and maintain. What they do *not* do is enable streamlined management of disparate data sources, by streamlining data access across such environments; for those capabilities, look to the next two bullets;

- *Data virtualisation* enables users to access data from multiple, disparate sources through a unified, logical view, hiding the underlying complexity of the data sources and providing a simplified, standardised view of the data to users. This enables users to query and analyse the data as if it were in a single, integrated data source, even if the data is stored in different formats, structures or locations. Data virtualisation provides a semantic data model across the entire enterprise enabling it to be easily understood and used by data analysts, developers and business users with minimal training;
- *Data fabric* integrates data across hybrid/multi-cloud environments and automates data engineering tasks. Data fabrics provide a comprehensive set of automation capabilities to enable efficient, automated data management and analytics across different data sources and types. These capabilities include data integration, data preparation and processing, data cataloguing and discovery and data governance and security. Note that data virtualisation, by enabling real-time, logical data access rather than

physical data access, turns data fabric into *logical* data fabric, which provides flexible, real-time access to data.

These technologies are closely related, and they offer a progressive collection of capabilities. Let us take a look at the group's collective capabilities.

## THE CAPABILITIES THAT TAKE A LOGICAL APPROACH TO DATA MANAGEMENT

Vendors will often describe their offerings in general terms. This section breaks down the key capabilities of the various product categories, followed by a description of each capability. The goal of this section is to enable you to:

- Understand what capabilities are available;
- Define your evaluation criteria;
- Consider your specific current and future needs.

This should put you in good shape to effectively evaluate the different software technologies from different product categories and make an informed decision about the best fit for your business needs (see Table 1).

- (1) *The data access layer* simplifies accessing and integrating data from multiple sources, typically via SQL, even if the sources may not support SQL natively. It also provides a consistent, unified view of the data-to-data consumers, regardless of the location or format of the underlying data sources. The data access layer includes connectors or adapters for connecting to the various data sources, such as databases, applications, application programming interfaces (APIs) or web services. When evaluating different technologies, look at the available connectors

**Table 1: Key capabilities of product categories**

Capability	Query accelerators and data lake engines	Data virtualisation tools	Data fabric
1. Data access layer	●	●	●
2. Data lake query optimisation	●		●
3. Federated and distributed query engine	●	●	●
4. Data replication (persistence)	●	●	●
5. Metadata repository	●	●	●
6. Security and governance layer	●	●	●
7. Consumption layer	●	●	●
8. Universal semantic layer	●	●	●
9. Design interface and tools	●	●	●
10. Augmented data catalogue		●	●
11. Active metadata			●
12. Recommendations engine			●
13. Data preparation and delivery layer			●
14. Orchestration and DataOps			●

● = Full support, ● = 75 per cent support, ● = 50 per cent support, ● = 25 per cent support

and types of systems they support. You do not want to invest in a solution that may not support all of your current or future data sources;

- (2) *Data lake query optimisation* optimises the queries executed against a data lake, ensuring that queries are executed efficiently and with acceptable performance levels. Techniques include massively parallel processing (MPP) and different flavours of query acceleration. Data lake query optimisation is an important consideration when selecting one of these tools, as it can improve query performance, reduce data transfer costs and improve overall system efficiency. Note that query accelerators are typically designed to run very efficiently when a data lake is the primary source of information, but performance suffers in a distributed environment. Be sure to understand if all or most of your data will be in the data lake and, if not, what might be the implication on performance;
- (3) *Federated and distributed query engines*

enable users to access and query data from multiple sources as if they were part of a single database. Such query engines can process queries that span multiple data sources and combine the results into a single result set that can be returned to the user. They can access and integrate data from different types of sources, including databases, applications, web services and data lakes. Federated query engines are responsible for identifying the relevant data sources, retrieving the required data, and combining the results into a single view. In addition to performing many of the same techniques employed by stand-alone query accelerators, federated query engines will offer other techniques designed and optimised to run across multiple data sources, such as advanced query pushdown, query rewriting, and caching;

- (4) *Data replication (persistence)* refers to the creation and management of copies of data. Replication can take many forms: extract, transform and load (ETL)

processes for point-to-point replication, extract, load transform (ELT) processes for operations within a data lake or data lakehouse, microbatching from external APIs, ingestion from streaming channels and caching and acceleration structures. Each use case is different, and a modern architecture must offer flexibility on the techniques available to create, manage and operate these replication tasks;

- (5) *The metadata repository* is the centralised component that stores metadata about the data sources, such as their schema, data types, data statistics and relationships. It also provides information about the consumer data models, such as their mappings to the source, their lineage and their transformations. The metadata repository can also store activity metadata (See active metadata below). It is the foundation for not only the query optimiser but also for data governance processes such as managing changes, setting up security and certain self-service capabilities that are described in depth in other sections;
- (6) *The security and governance layer* provides advanced mechanisms to manage authentication and access to specific data. For example, if the column is marked as personally identifiable information (PII), permissions can be set not to share that information or to only share it with a subset of authorised users. This is an area in which details are important. Be sure to understand if your considered solution supports advanced features such as an attribute-based access control (ABAC) system to enforce security policies for data access. The policy engine defines and enforces security policies, which can be configured based on various factors, including data sensitivity, user roles and business requirements. It is also

good to understand if the tool you are considering integrates with your organisation's existing security and governance tools;

- (7) *The consumption layer* provides data consumers with centralised access to all data. This layer delivers data to end-users in their preferred forms. Most systems provide Java Database Connectivity (JDBC) or Open Database Connectivity (ODBC) support for SQL queries but consider all potential consumers and whether more options are required. Other options include ADO.NET, MDX, RESTful web services, OData, GraphQL, GeoJSON, exports to Microsoft Excel/SQL, Kafka and JMS message queues;
- (8) *The universal semantic layer* provides capabilities to define data structures that are easier to consume and understand by end users. There are a variety of features related to semantic modelling, such as the definition of derived data models with friendly names and descriptions, definitions of corporate metrics, tagging and classification, governance controls (such as endorsements or deprecations), data quality labels and many more. They all share the common goal of democratising access to data by making it more understandable and easier to use. If this is an important goal in your organisation, semantics should definitely be a key criterion in your evaluation;
- (9) *The design interface and associated tools* assist in managing the system. Capabilities between the vendors vary greatly from command-line-only tools to web UIs that enable data engineers and end-users to easily interact with the system. Consider which type of interface is acceptable for your comfort level, but keep in mind that if you plan on democrating data for all to use across the organization, you should

- ensure that the tool is intuitive enough for all potential users;
- (10) *Augmented data catalogues* provide a centralised, searchable inventory of data assets across the organisation, along with metadata, data lineage and other information about the data. Augmented data catalogues automatically classify and tag data assets based on their content and structure. They also enable data stewards and analysts to manually tag and annotate data assets with additional information, such as business definitions, data quality metrics and usage information. They enable users to quickly and easily discover and access data assets across the organisation, for better collaboration and decision-making;
  - (11) *Active metadata* can automate data engineering tasks to reduce complexity and optimise business outcomes and value. Traditionally, when people talk about ‘metadata’, they are referring to data that describes other data. In modern data management, however, additional technical metadata is employed, including elements such as transaction logs, user logins and query optimisation plans, used to create *active* metadata. Active metadata is data that describes such metrics as how the system is performing and how people interoperate with the system;
  - (12) *Recommendation engines* leverage ML algorithms to analyse active metadata such as search queries, data accesses and data downloads, to identify patterns and make recommendations for the more efficient operation of the system, or recommendations on other data assets that may be of interest to the user;
  - (13) *Data preparation and delivery layers* enable users to access the common semantic layer in a sandbox-like environment in which they can explore and transform data to create new datasets. This is a critical capability to enable self-service, ensure that consumers can perform those last-mile transformations to massage datasets into their final consumer form, and to be able to capture their feedback and crowdsource the evolution of the fabric;
  - (14) *Orchestration and DataOps* manages data pipelines and workflows and automates data-related processes. Orchestration refers to the coordination of multiple data-related tasks and workflows, often across multiple systems or platforms, which may involve the management of data pipelines that integrate data from multiple sources, coordinating data processing and analysis tasks. DataOps, on the other hand, refers to applying DevOps principles and practices to data management. DataOps emphasises collaboration and automation, improving the speed, quality and reliability of data-related processes.

### Forrester on data virtualisation

Analyst company Forrester reported on the potential value offered by data virtualisation, a key logical data management technology, in ‘The Total Economic Impact™ of Data Virtualization Using The Denodo Platform’.<sup>2</sup>

Forrester studied and interviewed four companies that implemented data virtualisation, including a global financial services organisation with US\$13bn = in annual revenue, and found that on average they saw 408 per cent return on investment (ROI) and began to see payback in less than six months.

In this report, Forrester says that:

‘Data virtualization helps organizations access data across disparate sources and deliver a unified view of the data faster, cheaper, and using fewer resources than traditional data integration approaches. In this TEI [Total Economic Impact

report], data virtualization delivered 83% reduction in time-to-revenue and 65% decrease in delivery times over extract, transform, and load (ETL) processes. Denodo's logical data fabric goes beyond data virtualization, enabling organizations to effectively integrate and manage their data where it lives without replicating it, allowing both technical and non-technical users to quickly answer key business questions using a data-driven approach.<sup>3</sup>

### **Real-world example: The Johannesburg Stock Exchange**

The Johannesburg Stock Exchange (JSE), founded in 1887, is the largest stock exchange in Africa. As the JSE operates in a broad spectrum of the capital markets, its day-to-day operations involve a large number of core business systems.

JSE's data had to be sourced and integrated from various heterogeneous data systems including relational databases, flat files and APIs, to serve the needs of the various functions. Using traditional ETL processes was cumbersome and inefficient. Additionally, the different versions of data entities across different systems created confusion and scepticism around data authenticity.

The JSE implemented the Denodo Platform, a data integration and data management solution powered by data virtualisation, to build a logical data layer. With data virtualisation capabilities and data catalogue, recommendations engine and other advanced data management capabilities, the JSE used the Denodo platform to build the foundation of a data fabric.

The modern data architecture at the JSE, built around the Denodo Platform's logical data layer, aggregates different forms of data from heterogeneous source systems and makes the integrated data available in real-time to consuming applications. The logical data layer processes approximately 2bn rows per month at the JSE.

'The logical layer enabled us to serve the real-time data needs of our business', said Chester Enslin, head of IT Enterprise Integration and Software Quality Assurance at JSE. 'Since implementing it, we have been able to provide 100% accurate data, 100% of the time for our various functions, particularly post-trade settlements.'<sup>4</sup>

### **LEVERAGE A DATA FABRIC TO ACCELERATE A DATA MESH ARCHITECTURE**

In the data management space, data mesh is a very hot topic. But what is a data mesh?

Data mesh is not a technology or a capability but a cultural, organisational shift in how organisations manage their data. A data mesh takes a decentralised approach to data architecture, in which specific business domains, such as core banking, custody, compliance, risk management, order management and reconciliation, organise data. This approach enables teams to take ownership of their data and decide how to manage it best to optimise business outcomes and value.

In contrast with data mesh, data fabric and the other technologies mentioned above are *architectural* or *technological* approaches to data management, so it would not have been appropriate to include data mesh in the capabilities section. This paper would not be complete, however, if it did not cover data mesh and the confusion between it and data fabric.

### **WHEN TO USE DATA FABRIC, WHEN TO USE DATA MESH AND WHEN TO USE BOTH**

You have likely come across articles or blog posts questioning which is the right approach for you, data fabric or data mesh. As mentioned above, the two are very different at their core, so it may be that you would use the two in tandem or one or



the other independently. It is also possible that neither is right for your organisation, but you can take steps to be ready for these architectures down the road.

Data fabric is a flexible, reusable and powerful data management approach critical for any organisation struggling with distributed data silos. One of the core pillars of data fabric is its use of metadata, both passive (data about the data) and active (data about the performance and how the system is being used). Suppose you select a solution that can address the capabilities outlined in the Table 1. In that case, much of the data discovery, translation, enrichment and creation of semantic relationships between metadata will be automated for you. This same metadata will be used to build the augmented data catalogue, accelerating your metadata management projects by enabling business users to understand, enrich and use metadata to inform and further their data and analytics democratisation initiatives. Many highly respected analysts believe most capital market firms will benefit from a data fabric. Gartner has forecasted that data fabric deployments will quadruple efficiency in data utilisation while cutting human-driven data management tasks in half.<sup>5</sup>

On the other hand, a data mesh may not be ideal for all organisations. A federated, localised data management approach may not be needed for centrally structured organisations. This is because they may not have subject matter experts in the lines of business to create the domain expertise to benefit from a data mesh architecture. In the capital markets industry, however, many companies employ more of a decentralised structure, with deep expertise in the lines of business, the data they use and how to make decisions about how to manage it best to optimise business outcomes and value. You need to assess if this is true in your organisation to determine if a data mesh is right for you.

## HOW A DATA FABRIC CAN SIMPLIFY YOUR ADOPTION OF A DATA MESH

A key objective of the data mesh (see Figure 1) approach is to speed up the creation and availability of trusted, high-quality and compliant data. The goal is to train business professionals to produce this data rather than always relying on centralised IT teams who may not understand the data as well as people who work with the data regularly.

Data mesh has four major principles:

- Domain-oriented decentralised data ownership and architecture;
- Data as a product;
- Self-service data infrastructure as a platform;
- Federated computational data governance.

A data fabric is well suited to be the self-service data infrastructure in the third bullet above. The data models can be built using a layered approach, with different levels to facilitate decentralised data ownership while providing a central IT group with the ability to build predefined canonical data models for use by the domain teams, simplifying their tasks. Additionally, this layered infrastructure approach would enable the central definition of security and governance policies so they are consistent throughout the organisation.

Because this is an important aspect of how a logical data fabric provides the infrastructure needed for a data mesh, let us take a closer look at the various layers of a logical data fabric:

- *The data access layer* accesses information from the various data repositories and decouples the heterogeneities of the underlying communication protocol and formats from the upper layers. At this layer, the IT team can generate 'base views' over data sources that represent a normalised schema in a tabular structure. To the upper layers, these base views

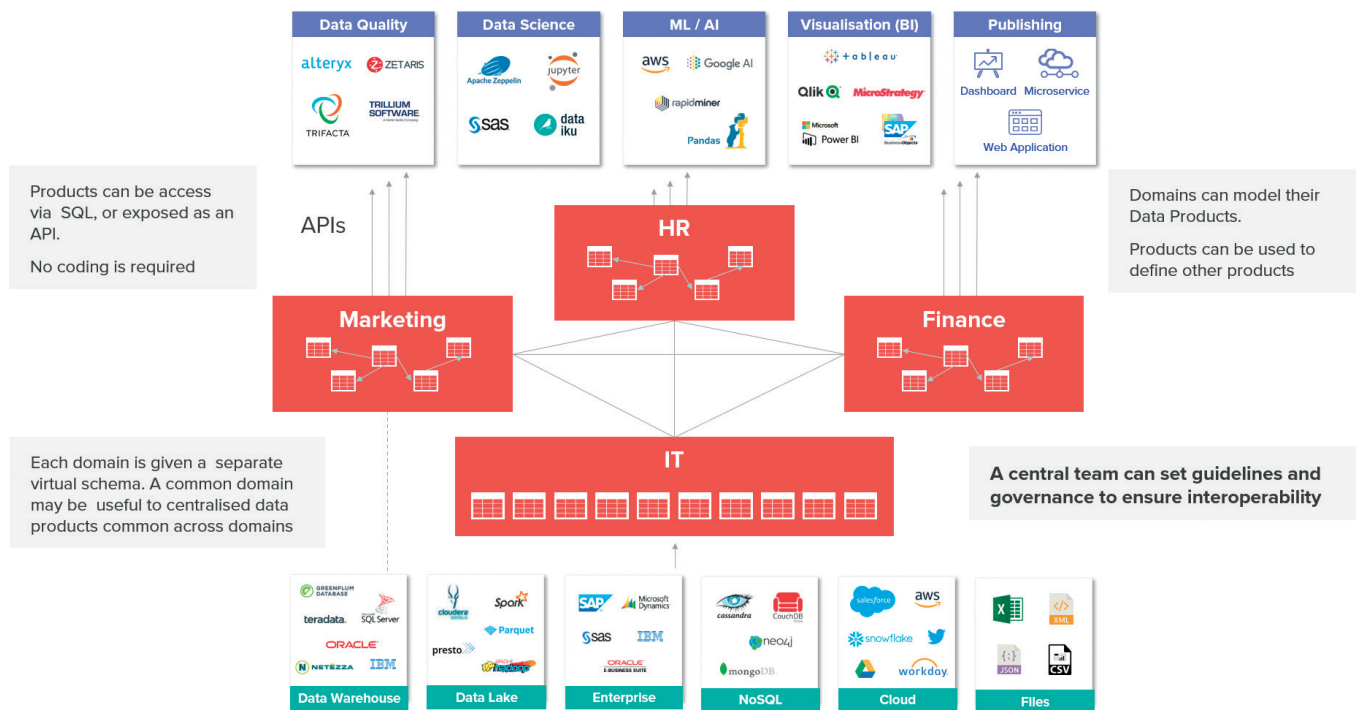


Figure 1 Foundation for data mesh

appear as relational views of the data, regardless of the underlying data source technology;

- *The data integration or modelling layer* offers data combination and transformation capabilities with logical operators for the seamless creation of composite data views on top of the base data views delivered by the data access layer. In this layer, users can perform complex data transformation, metadata modelling and data quality and semantic matching operations using SQL and other relational tools that they are already familiar with;
- *The publishing layer* facilitates the creation of a single point of access and interaction with the underlying data sources and abstracted data views in a standard way, such as JDBC or ODBC. In this layer, administrators can create a clear, simple representation of key business entities or domains to satisfy business information needs.

This enables organisations to leverage the subject matter expertise of individual domain teams to build the views that will be shared with the rest of the business. Stakeholders can start engaging with the data at their preferred level, based on their technical skill.

Data fabric also provides a way for consumers to find and access the data products they need via a data catalogue, which enables consumers to search for data products just as they would search for a commercial product online. Data catalogues typically provide data lineage information, enabling data consumers to easily see the source systems from which the data originated. Data catalogues also enable streamlined collaboration, as comments can be shared on the data and its intended use. These features help build trust in the data in the eyes of the data consumers.

### Getting started

With this paper, we hope to have demonstrated the business case for complementing

investments in centralised data management tools, such as cloud data warehouses or data lakes, with a logical data management approach. This combination provides the flexibility and scale of the cloud with the ease-of-use of a single location for users to find, access, integrate and share data securely for timely, more informed business decisions.

To take the first steps in a modernisation journey of this nature, begin by forming a team composed of stakeholders from both the business and technical sides of the organisation, and engage in the following activities:

- (1) List all of the data-access challenges that affect your organisation;
- (2) Identify the ‘Nirvana’ state, in which all of these challenges are solved;
- (3) List out data governance and security protocols that will need to be established, in order to reach that state;
- (4) Determine if a data mesh is appropriate for your organisation, or if the modernisation will solely focus on technology;
- (5) List potential solutions, including specific technologies and architectural designs;
- (6) If possible, given the appropriate vendor(s), participate in proof(s)-of-concept of the solution(s);
- (7) Implement the solution(s) on a limited scale, such as a single use case;
- (8) Evaluate the change, measuring performance against initial requirements, and decide to continue company-wide or go back to step 6, to try a different solution;
- (9) If desired, engage with a consulting company with experience in logical data management and the capital markets. Such an engagement can begin at any stage of the process, even before step 1. These companies can assist with detailed design, technology selection and evaluation and ongoing advice;
- (10) Enjoy the benefits of logical data management, including real-time access to trustworthy, curated data.

Clearly, data modernisation initiatives are not trivial; however, they do not need to be prohibitively time-consuming or expensive, and many, if implemented correctly, can be implemented without a single second of downtime. Feedback suggests that greatest successes are typically achieved where the supplier and client work closely from the outset to establish a data vision and strategy, ensuring rigor around operations and governance, while addressing the priority use cases for the business, to maximise time to value.

## CONCLUSION

Data will always be distributed, but logical data management solutions provide a way to connect to the data in real time, as needed, without having to physically move it into a centralised repository. As we have seen, the three main logical data management product categories are progressive: data virtualisation offers what the stand-alone query accelerators offer, and logical data fabric builds on what data virtualisation offers.

This paper covered 14 broad capabilities of logical data management solutions, and each of today’s available solutions contains a slightly different mix of these capabilities. Some organisations with highly specific needs — those that also do not anticipate much change down the road — might be well served by a data lake query accelerator or a pure-play data virtualisation tool. To maximise flexibility and futureproof an organisation’s investments, however, consider looking for a solution that supports all of the 14 capabilities above, in the form of a logical data fabric. This would provide the most options for current and future needs and all of the benefits of a modernised data infrastructure. In the capital markets, this

can mean the difference between success and obscurity in the years to come.

## REFERENCES

- (1) Denodo (2022), 'Denodo Global Cloud Survey Report 2022', available at <https://www.denodo.com/en/document/whitepaper/denodo-global-cloud-survey-report-2022> (accessed 30th May, 2023).
- (2) Denodo, 'Forrester Study: The Total Economic Impact™ of Data Virtualization Using the Denodo Platform', available at <https://www.denodo.com/en/document/analyst-report/forrester-study-total-economic-impact-data-virtualization-using-denodo> (accessed 30th May, 2023).
- (3) *Ibid.*
- (4) Denodo (2022), 'Johannesburg Stock Exchange Leverages the Denodo Platform to Verify Billions of Dollars in Trades Every Day', available at <https://www.denodo.com/en/system/files/document-attachments/cs-jse-02-web.pdf> (accessed 30th May, 2023).
- (5) Gartner, 'Essential Guide to Data Fabric', available at <https://www.gartner.com/en/publications/essential-guide-to-data-fabric> (accessed 30th May, 2023).

Copyright of Journal of Securities Operations & Custody is the property of Henry Stewart Publications LLP and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.