playbook

Modern Data Paradigms and Architectures: The Data Mesh and the Semantic Layer

By Markum Reed, Ph.D.

Introduction and Data Mesh Challenges

Today, organizations that want to derive value from their data are modernizing their data environments and organizational structures to support advanced analytics. Some organizations are interested in centralizing their data platforms in the cloud. Others are looking at a distributed architecture. Recently, organizations have started to consider the data mesh. The data mesh is a sociotechnical approach to building a decentralized data architecture by leveraging a domain-oriented, self-service design. With a data mesh, the responsibility for analytical data is shifted from the central data team to the domain teams, supported by a data platform team that provides a domain-agnostic data platform.

The fundamental purpose of the data mesh is to allow organizations to get value from their analytics data and stored, historical data at scale. Organizations can use this approach to handle systematic data landscape changes, the expansion of data sources, and various data transformation and processing use cases. It can also be modified depending on the rate of response to change.

The four fundamental pillars of the data mesh are:

Domain-oriented ownership. Business domains are responsible for their data. Hence, ownership is decentralized, which means that the data owner is the group closest to and that best understands the data. For instance, a product group can be a business domain; it would own its data. Business domains manage their data quality, understandability, and interoperability. Organizations expect them to share data across organizational boundaries.

Data as a product. Data is accessible to those who need it, and data is considered a product, which means the consumers of that product should be satisfied with it. The company should support a product view of data. The data product must be interoperable with—and able to be joined with—other data products owned by other domains. A "data product quantum" expresses the essential information features

such as data, code, policy, and infrastructure dependencies. This idea makes it easy to share that information throughout other domains. Furthermore, data as a product reduces the risk of silos.

Self-service. A self-service platform delegates data access to teams. The platform supports a broad range of user personas, such as data product developers and consumers. Domain teams work autonomously within their domains without depending on a central infrastructure team, which will lower costs and enable the development of data products.

Federated governance. In the data mesh, team members from all domains come together to form the basis for a federated governance model. Each domain owns its governance, but the model balances the autonomy of the domains with the necessity of compliance, interoperability, and security of the mesh.

Does your organization utilize a data mesh paradigm?

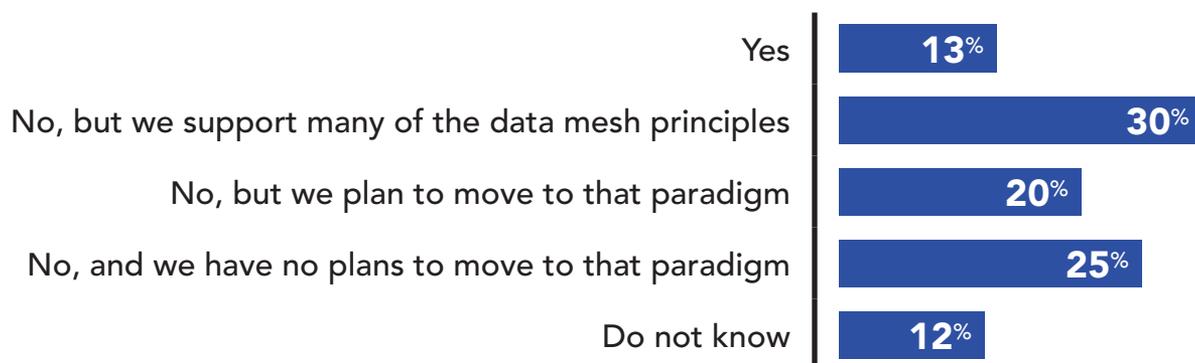|	
Yes	13%
No, but we support many of the data mesh principles	30%
No, but we plan to move to that paradigm	20%
No, and we have no plans to move to that paradigm	25%
Do not know	12%

Figure 1. Based on 198 respondents. See Q2 2022 Best Practices Report: Modernizing the Organization to Support Data and Analytics, *available at* tdwi.org/bpreports.

Recent TDWI research indicates that the data mesh is still in an early phase of adoption.

One issue with the data mesh is that each domain platform is a stack. At TDWI, we hear organizations are concerned about each domain becoming its own silo. It is important to understand the potential challenges due to the cultural shift associated with implementing a data mesh.

Additionally, organizational challenges can make or break any data and analytics effort. Organizations will need to weigh the challenges against the benefits of the mesh approach. Each domain will need to think about data as a product, which is a big change requiring a data product manager. The data product manager is responsible for the data and must be able to identify and assess opportunities for deriving value from the data.

Each domain will need its own data team to ensure all data assets within the mesh are shared, published, discoverable, and interoperable. The risk of developing data silos arises when domain teams do not embrace the concept of data as a product.

The semantic layer can be the glue that binds architectures and teams. A semantic layer provides a business representation of data using common business terms that help users access data. In a recent TDWI survey, 59 percent of respondents have deployed or are planning to deploy a semantic layer platform.[1] The semantic layer can enable a data mesh by becoming the access layer connecting databases and BI tools. It allows users to browse the mesh for related data products. Users can traverse semantic relationships to identify the desired data sources.

The modern semantic layer can also be used to deliver data to BI platforms and results from machine learning models to BI. The key is expanding how a semantic layer fits into a modern data architecture by understanding its relationship to the data mesh. This playbook describes the business case for the data mesh and the semantic layer, as well as plays for getting started.

---

[1] See the Q3 2019 *Best Practices Report: Driving Digital Transformation Using AI and Machine Learning*, available at tdwi.org/bpreports

Business Case for the Data Mesh and the Semantic Layer

The semantic layer has several characteristics that make it useful with the data product view called for by the data mesh:

- The semantic model helps data users find meaning, but it does not convey the meaning by itself. The semantic layer makes data addressable. A data product offers a permanent and unique address to the data user to programmatically or manually access it. This addressing system embraces the dynamic nature of the data and the mesh topology.

- The semantic layer makes data understandable. Once a data product is discovered, the next step of the data user's journey is understanding it. Each data product provides semantically coherent data—data with a specific meaning. A data user must understand this meaning—what kind of entities the data product encapsulates, the relationships among the entities, and their adjoining data products.

In addition to understanding the semantics, data users need to know how they syntactically access and query the information, what kind of queries they can execute, or how they can read the data objects. They need to understand the schema of the underlying syntax of data. Sample data sets and example consumer codes ideally accompany this information. Examples accompanied by a formalized description of the data improve data users' understanding.

Over the past several years, organizations have focused their analytics investments on embracing major cloud data platforms. Although modern data platforms simplify operations and solve many problems, they do not address the fundamental problems of getting actionable data to the consumers who want it.

Building a business case for investing in a semantic layer strategy should focus on three important value drivers:

- Increase the value of data by making it easier to combine more data sets, large data sets, and more expansive data windows in a way accessible to more data consumers.

- Reduce the cost of delivering analysis-ready data by optimizing cloud spending, simplifying data engineering and DataOps, and making data consumers more productive.

- Make interacting with data assets easier by increasing the number of data consumers. Improving data literacy in an organization opens up the flow of data-driven insights.

Critical Plays for Getting Started

KNOW YOUR ORGANIZATIONAL MODEL: CENTRALIZED VERSUS DECENTRALIZED

Knowing and understanding your current organizational model will help you understand how your data mesh journey should start. Centralized modeling leads to organizational bottlenecks. The data mesh paradigm moves us away from the centralization of data and its ownership toward a decentralized model.

A decentralized, distributed data architecture means the data from each domain is not copied but kept within the domain. Each domain has its data models.

Moving from a centralized to a decentralized model may make sense for organizations with multiple business units where benefits might outweigh costs and the resources exist to staff each domain.

PLAN YOUR STRATEGY AND EXECUTION

Data strategy drives the execution of a data mesh. A data strategy defines how a company utilizes data and creates strategic initiatives. These initiatives lead to executing intelligent touchpoints and applications that improve the user's experience. Each intelligent business initiative can boost user applications with ML. Developing intelligent applications leads to identifying and developing data products that provide the necessary data and insights. One of the core components of a data mesh is a set of related data products. Intelligent applications drive the identification and development of data products.

At the same time, real-world applications prompt the creation of data products. The creation and consumption of data products and intelligent applications drive prioritization and development of platform capabilities. Over time, platform features and services can be reused to deliver new data products. As the organization conducts its data strategy, it implements the organizational and social aspects of the data mesh. However, organizational alignment happens with the delivery and execution of data business initiatives, with explicit intention and organizational planning.

Building the reusable components of the mesh through supportive use cases can help organizations avoid the pitfalls of point solutions. Focusing on both data providers and data users helps prioritize platform capabilities. The platform evolves in response to its users. It cannot fulfill all the features needed by all use cases. Early on, the use cases must be carefully selected.

Early use cases have enough dependencies on the platform, so they drive the prioritization and build of necessary platform services without getting blocked. Although business initiatives and individual use cases can provide investment to develop the mesh, the mesh must be considered a long-term investment and an internal company product.

SEMANTIC LAYER IMPLEMENTATION

Implementing a common semantic layer platform that supports data model sharing, collaboration, and ownership is critical. The central data team can define standard models and definitions with a semantic layer. At the same time, the domain experts own and define their business process models. With the ability to share model assets, business users can combine their models with models from other domain owners to create ensembles that answer more profound questions.

In creating semantic solutions, clearly defined use cases provide the business logic the semantic architecture will answer for the organization. This drives an understanding of the users and stakeholders, communicates the business value or challenge the solution will solve for the organization, and enables the definition of measurable success criteria.

Many organizations' data architectures are based on relational databases, data warehouses, and a wide range of content management clouds that drive data analysis and analytics capabilities. This does not mean organizations must rebuild their functional enterprise architecture to adopt or implement semantic capabilities.

Increasing focus on data modeling and designing efforts is beneficial for these organizations. Adding models and standards will allow model designers to capture business meaning and context to provide the least disruptive starting point. It is important to select an effective approach to model data and map its path from source to target using the relevant transformation and unification processes. Furthermore, model-mapping best practices are established based on an organization's use cases, architecture capabilities, and talent, and they must be flexible enough to support data governance requirements.

Organizations must plan ahead for long-term scale and governance issues despite the skill required to develop mature data management practices. Semantic web standards provide fundamental, standard

frameworks and practices for organizations to embrace when initiating semantic architecture.

Semantic technology features include graph management applications that function as bridges, designating the storage, processing, and retrieval of semantic data.

The architecture for a semantic layer includes:

- A semantic model for holding the knowledge and relationships within data

- Integration with a data catalog for practical application and governance of metadata

- A built-in data governance capability for applying role-based data access and control

- A sophisticated metrics engine for expressing complex business calculations

- An automated performance management engine for delivering "speed-of-thought queries"

Practical and scalable semantic architecture supports upstream customers or employee-facing applications such as enterprise search, data visualization tools, and end services/consuming systems. Organizations can insert semantic components into other enterprise solutions, applications, and services. With this as a basis, an organization can take advantage of AI capabilities to find and use content more easily and automate content categorization to expand data governance practices.

Semantic layers work best as an integrated framework for facilitating knowledge interoperability. Getting started by focusing on business-centric use cases driven by semantic solutions is important. Further, semantic layers complement other technologies that benefit from access and an intuitive representation of content and data, including search and dashboards.

Concluding Thoughts

Although the fundamental purpose of the data mesh is to allow organizations to get value from their analytics data at scale, understanding the potential challenges associated with implementing a data mesh and semantic layer is important. Organizations will need to weigh the challenges against the benefits of the mesh approach. Each domain will need to think about treating their data as a product and building a data team to ensure data assets are shared, published, discoverable, and interoperable.

The semantic layer can be the glue that binds architectures and teams. The key is expanding how a semantic layer fits into a modern data architecture by understanding its relationship to the data mesh. The semantic model helps the data users find meaning, but it does not convey the meaning by itself.

AtScale and the Semantic Layer
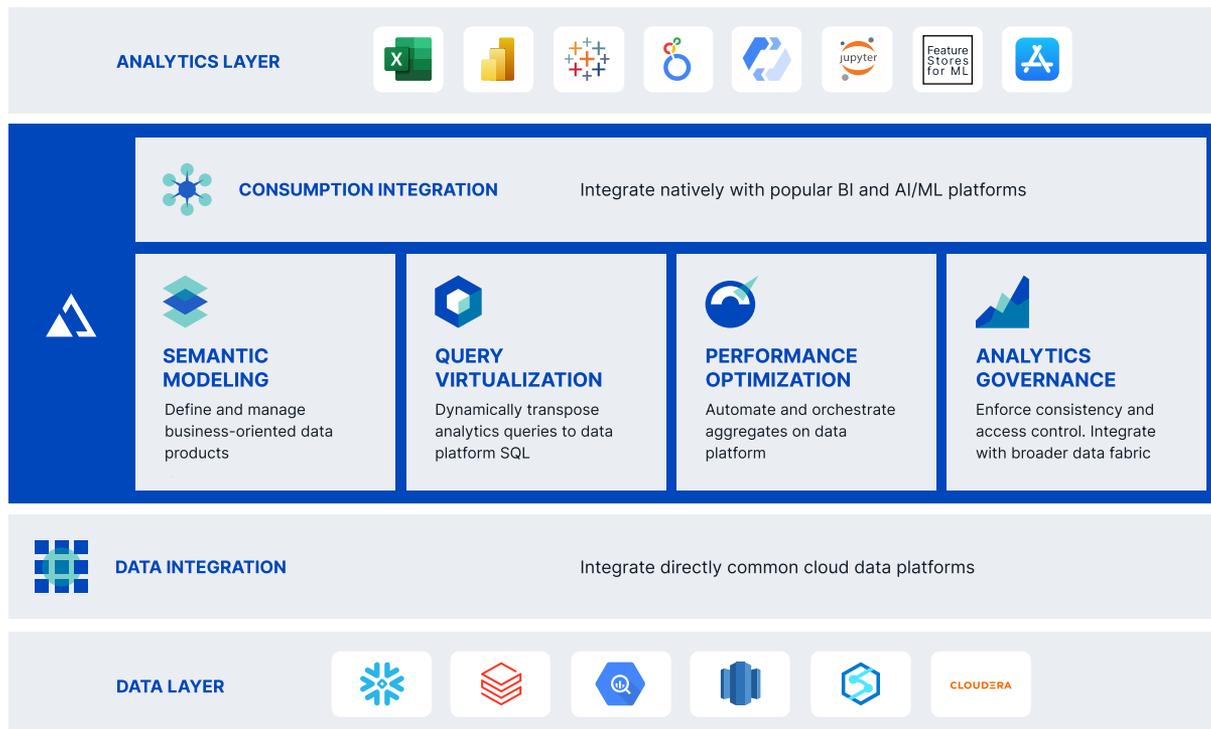## (Content supplied by AtScale)

AtScale is a leading provider of [semantic layer software](#) that sits between your data platform and your BI or AI tools as a centralized place to define business logic, make sure reporting is consistent, improve time to insight, and make data more consumable.

THE HOME DEPOT · Tyson · CardinalHealth · bol.com · Rakuten

TOYOTA · Betclic · J.P.Morgan · VISA

[View AtScale Case Studies and Customer Success Stories](#)

AtScale has a unique approach to accelerating analytics queries on large cloud data sets, delivering speed-of-thought performance without extracting large data sets or maintaining OLAP cubes. We do that without moving data and without complexity, leveraging existing investments in data platforms, applications, and tools.

[Watch Semantic Layer Webinars Featuring Top Data Leaders](#)

ANALYTICS LAYER

CONSUMPTION INTEGRATION — Integrate natively with popular BI and AI/ML platforms

SEMANTIC MODELING	QUERY VIRTUALIZATION	PERFORMANCE OPTIMIZATION	ANALYTICS GOVERNANCE
Define and manage business-oriented data products	Dynamically transpose analytics queries to data platform SQL	Automate and orchestrate aggregates on data platform	Enforce consistency and access control. Integrate with broader data fabric

DATA INTEGRATION — Integrate directly common cloud data platforms

DATA LAYER

Request a Demo to See AtScale's Semantic Layer Software in Action

With AtScale, you get live, high-performance query access to your cloud data warehouse while forming a single source of governed analytics for all data consumers to leverage. This provides speed-of-thought query performance and consistent KPIs across any BI/AI tool such as Excel, Power BI, Tableau, Looker, DataRobot, and more—while lowering analytics costs by ~3x.

8X	**3X**	**10X**	**4X**
INCREASE IN BI QUERY SPEED	REDUCTION IN CLOUD COMPUTE	REDUCTION IN TIME TO BUILD NEW DASHBOARD	REDUCTION IN DATA MODELING TIME

To learn more about using a semantic layer for data and analytics, check out these five resources:

- Download the Compilation of Semantic Layer White Papers by Best Selling Authors & Experts

- Download the Practical Guide to Using a Semantic Layer for Data & Analytics

- Download the Data Leader Study on the Impact of Using A Semantic Layer by DBP Institute

- Download the Semantic Layer White Paper by Andrew Brust

- Sign Up for the Annual Semantic Layer Summit Virtual Conference feat. top thought leaders

About Our Sponsor

ATSCALE

AtScale enables smarter decision-making by accelerating the flow of data-driven insights. The company's semantic layer solutions simplify, accelerate, and extend business intelligence and data science capabilities for enterprise customers across all industries. With AtScale, customers are empowered to democratize data, implement self-service BI, and build a more agile analytics infrastructure for better, more impactful decision-making.

For more information, visit atscale.com.

About the Author

Markum Reed, Ph.D., is the director of research for data management at TDWI. At TDWI, he focuses on data management, data governance, data literacy, data science and analytics, and machine learning. Before joining TDWI, Dr. Reed was an associate professor of data analytics at the School of Business at Henderson State University and part of the research faculty for the Institute of Chief Data Officers. He has also taught at Southeast University, Nanjing, and was a research scholar at Southern Illinois University. He holds a Ph.D. in economics from Southern Illinois University Carbondale. You can reach him by email at mreed@tdwi.org and on LinkedIn (https://www.linkedin.com/in/markumreed/).

About TDWI Research

TDWI Research provides industry-leading research and advice for data and analytics professionals worldwide. TDWI Research focuses on modern data management, analytics, and data science approaches and teams up with industry thought leaders and practitioners to deliver both broad and deep understanding of business and technical challenges surrounding the deployment and use of data and analytics. TDWI Research offers in-depth research reports, commentary, assessment, inquiry services, and topical conferences as well as strategic planning services to user and vendor organizations.

About TDWI Playbooks

TDWI Playbooks provide data professionals with a summary of important key factors about contemporary data-related topics. Playbooks present the issues and challenges facing enterprises about each topic and offer a concise list of proven best practices to succeed in a particular area of analytics, business intelligence, or data management. Playbooks are written by TDWI research analysts and faculty who synthesize their research and experience into easy-to-understand explanations and practical recommendations that enable data professionals to apply the best, most productive approaches and techniques to their projects or initiatives.

tdwi

Transforming Data With Intelligence™

A Division of 1105 Media
6300 Canoga Avenue, Suite 1150
Woodland Hills, CA 91367

E info@tdwi.org

tdwi.org