CHECKLIST 2023

# The Semantic Layer's Critical Roles in Modern Data Architectures

By Philip Russom

TRANSFORMING DATA WITH INTELLIGENCE™

ATSCALE

# The Semantic Layer's Critical Roles in Modern Data Architectures

By Philip Russom

Many of the most exciting innovations and advancements in data management today are occurring within the semantic layer of data architectures and data stacks. For example, we are witnessing new or improved approaches to semantic modeling, data cataloging, and data lineage. Even older forms of managing semantics—such as metadata and virtualization—are being infused with new techniques for agile modeling, performance optimization for logical and virtual data environments, and intelligent augmentation (i.e., tool algorithms driven by machine learning and graph analytics).

The innovations of the semantic layer also play a role in improving large-scale data and analytics architectures. For example, the new definition of data fabric is not possible without a modern semantic layer, and the semantic layer can be a backbone for unifying new data and analytics architectures in the

Understanding the semantic layer's critical roles:

1. Understand what a semantic layer should be and do
2. Consider why your data architecture needs a modern semantic layer
3. Recognize where a semantic layer fits in today's data architectures
4. Understand how the semantic layer helps remodel distributed data
5. Appreciate how a semantic layer contributes to common data architectures

cloud. Furthermore, a well-designed semantic layer allows analytics teams to define business metrics, hierarchies, and dimensions on top of big data while providing a means to centrally govern data access and deliver high-performance interactive queries.

This TDWI Checklist educates data and analytics leaders about modern platforms and practices for the semantic layer. It does so by discussing five beneficial characteristics of the modern semantic layer but in the context of the semantic layer's critical roles in modern data architectures.

NOTE: This report assumes the reader is familiar with data architectures. Readers needing more details can read the *TDWI Checklist Report: Six Requirements for the Modern Data and Analytics Cloud Stack*.

## 1 Understand what a semantic layer should be and do

The semantic layer has been around for many years in many guises. The earliest ones originated as functionality embedded and buried within larger tools, typically for reporting, enterprise business intelligence, data integration, and database management. The catch was that such semantic layers were usually optimized for their parent tool and not much else.

Today, a truly modern semantic layer is a standalone tool type that provides data semantics services for multiple tools within a multitool and multiplatform data architecture. This gives the modern semantic layer the ability to serve many architectural layers, tool types, data platforms, use cases, and business departments without favoring one over another.

The data descriptions (or "data about data") created and managed by a semantic layer may take the form of older techniques (such as metadata management, federation, and virtualization) or newer ones (such as data catalogs, data lineages, dimensional models, or automated generation of data descriptions via knowledge graphs). Whether the semantic layer is a unified environment from a vendor or assembled by technical users as a "best of breed" collection of multiple tools from multiple vendors, it should support many semantic tool functions.

A semantic layer platform must go beyond data definitions to provide rich capabilities in semantic modeling and data modeling. In other words, a tool for the semantic layer should actively support the creation of new data structures and data products (whether federated, virtual, logical, or data sets in storage), not just descriptions of source data and its characteristics.

The semantic layer manages the interactions between data consumers (whether they be humans or AIs) and enterprise data assets (typically stored in the cloud). This interaction hinges on a semantic model that maps the language of raw data (e.g., field and table names, file and database formats) with business concepts (e.g., revenue, ship quantity, fiscal quarter). The structure of the semantic model may take different forms, defined using various ways of describing data.

A semantic layer must translate data consumer requests to the flavor of SQL preferred by the source data platform. It must accommodate multiple inbound protocols (not just SQL) because tools themselves support different protocols. For example, to support Excel, the semantic layer should support MDX. For data science, it should support Python. For application developers, it should support REST.

In production, a semantic model solution must deliver "speed of thought," direct query performance through automated performance optimization. Query performance is imperative because without it, many end users will make redundant and non-governed copies of data in the form of data extracts (TDEs in Tableau) or data imports (Power BI, Qlik, etc.). One of the greatest benefits of a modern semantic layer, extended with query optimization and virtualization, is that the semantic layer serves as an abstraction layer for governance, security, and "single source of truth."

## 2 Consider why your data architecture needs a modern semantic layer

The semantic layer provides a collection of data descriptions (and tools to create and maintain them) to make a single, centralized, and standardized architectural layer for most data semantics. Being centralized, a semantic layer delivers a standardized and consistent way of representing enterprise data to different types of users, tools, and data management processes.

This centralized approach simplifies many things, and it delivers important architectural benefits:

- Provides a process that is a governed, standardized, and consistent way of representing distributed enterprise data
- Can be a single point of entry, with single sign-on and other security for systems the semantic layer accesses
- Friendly descriptions of data that simplify and improve modern data practices (e.g., self-service, dashboard customization)
- Reuse of composable data objects and data products listed in the semantic layer
- Automation for data governance, monitoring via data, audit, and data observability
- When done well, can elevate data literacy and democracy

Today's data architectures are trending toward centralized data organization paradigms (databases, data lakes, data warehouses, data science labs) within both cloud and on-premises architectures. Even when physically consolidated, many data environments are not logically organized to support direct analytics use.

A semantic layer helps to unify far-flung data architectures in that it:

- Describes data consistently for all layers of the architecture and beyond
- Reaches all platforms: multiple brands, on multiple clouds or on premises, etc.
- Can provide data views that incorporate data from many architectural layers and elsewhere
- Facilitates data access and interfacing for many users and tools
- Can enable virtualization for the logical data warehouse and virtual data lake

Because it is an abstracted layer, the semantic layer creates a kind of future proofing. By decoupling the layers of data consumption tools and data storage platforms, a semantic layer provides IT with the freedom to consolidate, move, or transition its data without disrupting end-user data consumption. A semantic layer also provides an open platform for plugging in new or different data consumption tools as they arise.

As you can see, the business and technology benefits of a semantic layer are many, which is why organizations are turning to it.

## 3 Recognize where a semantic layer fits in today's data architectures

A modern semantic layer can be deployed to serve as an independent layer near the top of today's data architectures (see Figure 1). It fits just below the data consumption layer (for reporting, analytics, and self-service) and just above the data integration and streaming functions of the data fabric. It represents both data traveling into the data fabric and later being processed by the fabric, as well as data at rest in data storage (both on premises and in the cloud), typically organized as data lakes, warehouses, and other storage options.

This is a strategic position for the semantic layer, between the data consumption and data fabric layers. This location enables it to serve multiple functions: users and their tools for BI and analytics get a broader enterprise view of data compared to the limited views from their internal semantic layers; the data fabric gets the advanced semantic functionality and automation it needs; but the whole data architecture still has access to standalone and independent semantic layer tooling. This also creates a unification effect, which is beneficial to multiplatform architectures.

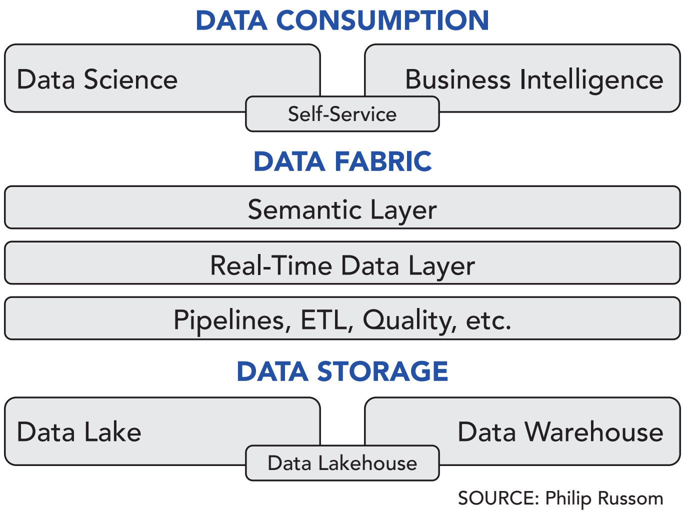


**Figure 1.** The semantic layer's strategic position in the data architecture.

The modern semantic layer can support multiple styles of data architecture, including those that are centralized, distributed, or a combination of both. For example:

- A semantic layer provides centralized data semantics, which in turn can support the centralized data that many users want in new data architectures (especially in the cloud)
- An independent semantic layer also supports decentralized and distributed design patterns, including the data mesh
- The semantic layer can support hybrid combinations of multiple architectural variations, even when in the cloud or on premises

Essentially, a modern, independent semantic layer can support just about any variation of data architecture available today.

## 4 Understand how the semantic layer helps remodel distributed data

As we just saw, a modern semantic layer can be applied to either a physically centralized environment (data lake or warehouse) or to a distributed environment that includes many data sets and platform types (as results from many unplanned or uncontrolled data programs).

In these cases, a semantic layer often acts as an abstraction layer that unifies disparate layers of a data architecture and its diverse tools and platforms. It does this by supporting business metadata and other approaches to creating business-friendly views of data and data products. The resulting business-friendly data models (created at the semantic layer, not physically in storage) hide the complexity of data in the data fabric and storage layers of the architecture, thus making distributed data easier to understand and access for a wide range of users and applications.

Ideally, data views may be constructed as semantic models, relational tables, OLAP dimensions, metric hierarchies, or time series. This means that a modern semantic layer tool should be capable of modeling most of the data structures common in data architecture usage, especially for analytics use cases.

For example, many organizations are hoping for one data source (or collection of data sources) exposed multiple ways to multiple users, teams, and tools. This is so that users can avail themselves of multiple data consumption styles, unlike the limited approaches typical of embedded semantic layers or traditional data warehouses.

As another example, data lakes are known for their massive stores of raw data. The semantic layer can make this data far more useful and valuable by creating a variety of views into the lake. This creates a logical data lake that complements and adds value to the actual lake in storage but without creating redundant copies and without having to maintain data aggregates in storage.

## 5 Appreciate how a semantic layer contributes to common data architectures

A semantic layer can add many types of technical and business value to a variety of local data architectures across an enterprise. This is true whether the semantic layer is retrofitted to a mature deployment or utilized from the beginnings of that architecture. In particular, the semantic layer can be beneficial to data warehouses, data lakes, data fabric, and DataOps.

### THE SEMANTIC LAYER HELPS THE DATA WAREHOUSE (DW)

**The semantic layer modernizes the data semantics that a DW depends on.** For example, most DWs suffer from metadata that is sparse, incorrect, and nonstandard; a semantic layer tool helps to correct these problems and give metadata a higher quality. Similarly, some DWs operate almost exclusively with technical metadata; a semantic layer helps the DW team embrace business metadata and more advanced forms of semantics, such as the data catalog and data lineage.

**The business-friendly semantics of the semantic layer enable DWs to participate in more use cases.** This is particularly useful when less-technical users want to access the data in a DW for self-service or when users need to personalize their management dashboards with metrics and KPIs from a DW.

**The semantic layer is a key enabler for the logical DW.** A logical DW is inherently multiplatform, hybrid, and distributed. To make this complicated microarchitecture seem more unified and usable, DW professionals use data virtualization and data views. Because a semantic layer is inherently logical, virtual, and view-driven, it can be a natural addition to a DW to make it a true logical DW.

## THE SEMANTIC LAYER HELPS THE DATA LAKE

**The semantic layer helps a data lake avoid becoming a data swamp.** A lack of metadata and other data semantics is the leading cause of swamps; a semantic layer provides ample metadata management for this situation.

**The semantic layer provides logical structures for unstructured data lakes.** Most lake data is raw, file-based, and multistructured. The semantic layer gives this environment much-needed structure via logical data views without creating redundant data copies and aggregations. In turn, the views make the raw and unimproved data of a lake far more accessible for self-service, exploration, and analytics.

## THE SEMANTIC LAYER HELPS THE DATA FABRIC

The most recent definition of the data fabric is "an architecture for unifying and governing multiple data management and data semantics disciplines, from data integration and quality to metadata and data cataloging." Among other requirements, a data fabric requires sophisticated semantics that are centralized, standardized, and shared for the many tool types found in modern data fabrics. The semantic layer satisfies this requirement.

## THE SEMANTIC LAYER HELPS DATAOPS

All data-driven development processes benefit from better semantics and DataOps is such a process. In fact, the semantic layer can help DataOps achieve many of its key objectives by providing:

- Centralized, standardized, and shared data semantics for data engineering but with more features than a metadata repository or catalog
- Automation for data semantics to accelerate the delivery of data products
- Semantic modeling (faster than models based on aggregation)
- Reduced data prep and design work
- Quick turnaround for the CI/CD processes of data products

## Concluding Thoughts

Organizations wishing to modernize their data semantics by adopting a semantic layer should heed the following key takeaways and recommendations:

- **Take data semantics more seriously, in general.** Make it a priority, from storage to data engineering to data consumption.
- **Do more than just create and manage technical metadata.** Modernize with business metadata, cataloging, data lineage, and tools optimized for the unified semantic layer.
- **Recognize that today's semantic layer is a standalone architectural layer.** It is not buried inside a single tool. It is standalone, so it can be agnostic and support many tools, users, and use cases.
- **Adopt the semantic layer concept as a best practice and architecture standard.** As a best practice, it improves data access, reuse, standards, and engineering. As an architectural layer, it unifies the whole architecture, enables interoperability, and presents architecture-wide observability.
- **Deploy a modern semantic layer for its business and technical benefits.** For example, it enables new practices, such as the data fabric and next-level data virtualization. It also modernizes and elevates common data solutions, namely the warehouse, lake, fabric, and DataOps.
- **Look for semantic layer tools with innovative and advanced functions.** A comprehensive tool will support multiple forms of data semantics and be open and tool-agnostic. It will also support performance optimization, data virtualization, and multiple approaches to modeling.

## About our sponsor

ATSCALE

atscale.com

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

**Philip Russom** has 25 years of experience as an IT industry analyst researching user best practices, vendor products, and market trends in data management and analytics. This includes data warehousing, data lakes, data integration, data quality, hybrid data architectures, cloud data management, data governance, analytics, and data platforms.

He has worked at most of the world's leading IT analyst firms, namely: Gartner Inc., Forrester Research, Giga Information Group, TDWI, and Hurwitz Group. In those positions and others, he produced over 650 research reports, magazine articles, speeches, and webinars. Before becoming an industry analyst, he worked for database software vendors as a product manager, product marketer, and documentation writer.

## 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, assessments, inquiry services, and topical conferences as well as strategic planning services to user and vendor organizations.

## About TDWI Checklist Reports

TDWI Checklist Reports provide an overview of success factors for a specific project in business intelligence, data warehousing, analytics, or a related data management discipline. Companies may use this overview to get organized before beginning a project or to identify goals and areas of improvement for current projects.

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