Gartner

Demystifying Semantic Layers for Self-Service Analytics

Published 3 April 2023 - ID G00783855 - 32 min read

By Analyst(s): Christopher Long, Joe Antelmi

Initiatives: Analytics and Artificial Intelligence for Technical Professionals; Evolve Technology and Process Capabilities to Support D&A

Data and analytics technical professionals struggle to deliver a universal semantic layer that balances agility and control in self-service analytics. This research helps technical professionals review options for deploying semantic layers to enable self-service analytics.

Overview

Key Findings

- Implementing a universal semantic layer continues to be difficult due to lack of tool interoperability, poor usability, lagging data governance maturity and organizational inertia. Meanwhile, the pressure from organizations to deliver metric consistency has only grown as consumption channels for analytics have expanded.

- Widespread self-service analytics development fosters innovation and agility at the expense of the governance and consistency once provided through centralized semantic layer architectures.

- Modern A&BI tools are expanding access to their modeling layers for more agnostic enterprise consumption. At the same time, stand-alone semantic layer platforms are extending capabilities to incorporate metrics store concepts. The expansion of these capabilities blurs the lines for organizations looking to implement a universal semantic layer.

Recommendations

As a data and analytics technical professional implementing semantic layers, you should:

- Evaluate and select an appropriate semantic layer for your use case by comparing the benefits, challenges and representative vendors of the three technical options: data layer semantics, stand-alone semantic layer, and A&BI tool semantic layer.

- Design a federated analytics architecture. Include a combination of local and global semantic layer data models based on the use cases, users and desired governance models.

- Develop operationalization processes that leverage A&BI tool capabilities for prototype-to-production semantic layer development. This approach enables business measures, metrics and data models to originate as A&BI prototypes and go into production with appropriate enterprise-level support.

Comparison

As organizations drive toward the goal of becoming data-driven, the demand for analytics increases significantly. Semantic layers are often one part of the solution to deliver the needed analytics across the enterprise. Historically, a centralized, governed semantic layer has been developed and maintained by IT to support BI and enterprise reporting use cases. However, the agility demanded by organizations and the capabilities delivered by modern analytics and BI tools decentralizes this responsibility. No matter how it is implemented, the semantic layer is still an important element of an analytics architecture. But the classic, centralized semantic layer is no longer the only option in this space.

Semantic Layer Implementation Options

Organizations taking a deliberate approach to implementing a semantic data layer will find several primary technology implementation options:

- Data layer semantics

- Stand-alone semantic layer

- Analytics and business intelligence (A&BI) tool semantic layer

These approaches intersect with a variety of locations in the data pipeline where you can place the semantic layer. This dynamic will be familiar to technical professionals who are deploying a logical data warehouse (LDW) or lakehouse architecture. Hence, similar guidance applies to the implementation decisions for semantic layers.

The first two architectural options (below) describe a semantic layer as part of an organization's modern data architecture, such as an LDW or data fabric. The LDW is designed to satisfy the majority of analytics requirements. LDWs support a broad set of analytics engines that can serve a wide variety of users and applications. For that reason, placing the semantic layer in the LDW is often optimal. The LDW itself is composed of multiple component parts, making a semantic layer a logical construct of components that are applied based on the specific technologies implemented. For additional information on the LDW construct, see Adopting a Logical Data Warehouse.

Option 1: Data Layer Semantics

This option describes a semantic layer that is built as an extension of data services within the data layer. Artifacts in this scenario may be made up of a variety of data marts, views (including materialized views), and online analytical processing (OLAP) models. They are presented as the connection point for analytics developers to link with source data. This manifestation of the semantic layer may not always appear to be a single, universal layer due to the potential for random asset development.

Option 2: Stand-Alone Semantic Layer

This option describes a semantic layer that exists as its own architectural element within the data and analytics stack. It is placed between the source data and consumption layers. In this architectural pattern, the tool used to build the semantic layer obscures the source data from analytics developers and consumers and becomes the primary connection point for analytical source data. Depending on the technology implemented, stand-alone semantic layers may manifest in a couple of ways:

- As data virtualization platforms, where source data largely stays in place
- As an abstraction layer, caching data for analytical usage separate from the source

Option 3: A&BI Tool Semantic Layer

This option describes a semantic layer with a local, optimized data store in the A&BI tools implemented. Because of its location inside an A&BI application, it is likely to be a siloed solution. Many organizations employ multiple A&BI tools for a variety of reasons, including:

- By design to take advantage of distinct capabilities within BI tools

- Growth of the organization through acquisition where the acquired company uses a different A&BI solution
- Business unit autonomy to purchase their own A&BI solutions

While multiple tools can bring agility to business user development, they can also make analytical definitions more fragmented throughout the organization. Early reporting tools that implemented semantic layers were very siloed. And while modern A&BI tools allow for greater sharing between developer users, the application silo effect is still quite prevalent.

Table 1 compares the characteristics of the different types of semantic layers.

Table 1: Common Semantic Layer Implementations

(Enlarged table in Appendix)

	Data Layer Semantics	Stand-Alone Semantic Layer	A&BI Tool Semantic Layer
Common Features	▪ Is an extension of data store capabilities ▪ Sits at the edge of the data layer ▪ May include views (including materialized), data marts and graphs	▪ Is an independent platform solution ▪ Sits between the data and consumption layers ▪ May include virtualization, abstraction and data lake enablement	▪ Is a BI tool capability ▪ Is part of the analytics consumption layer ▪ Includes direct query and ingestion-type models
Developer Roles	Data engineers	Data engineers	Business analysts and data analysts
Governance Localization	Centralized	Centralized	Distributed
Key Strengths	▪ Is highly governed ▪ Leverages in-house technology and skills ▪ Can deliver high volumes of data to many concurrent users	▪ Is well-governed ▪ Serves as a centralized source for analytical modeling and metrics ▪ Connects with diverse data formats across varying platforms	▪ Enables flexible and agile implementation and development ▪ Democratizes analytics development ▪ Reduces time to insight
Key Challenges	▪ Is highly reliant on central data engineering resources ▪ Is slow to change ▪ Has limited capabilities with unstructured data	▪ Involves IT-heavy development and implementation ▪ Is traditionally not designed to support data science and machine learning (DSML) and integrated application use cases ▪ Introduces another expensive layer to the data and analytics solution	▪ Is less governed ▪ Offers inconsistent development and deployment of analytical models and metrics ▪ Is often isolated within the vendor stack
Example Vendors*	IBM, Oracle, SAP, Snowflake, Teradata	Apache Druid, Apache Pinot, AtScale, ClickHouse, Data Virtuality, dbt, Denodo, Dremio, Kyligence, Kyvos Insights, Zetaris	MicroStrategy, Pyramid Analytics, Qlik, Salesforce (Tableau), Sisense, ThoughtSpot

Note:
* Major vendors, including Amazon, Google, Microsoft, Oracle and SAP, offer capabilities to support semantic layers in multiple architectural patterns.

Source: Gartner (April 2023)

The remainder of this document will:

- Provide background on semantic layer concepts
- Compare the three types of semantic layers
- Provide guidance on choosing the right type of semantic layer for your needs

Analysis

The efforts made by organizations to become data-driven often push the resource limits of what central data teams are able or willing to provide for A&BI reporting. Many data and analytics teams are overwhelmed by requests for related analytical deliverables. In response, many organizations deploy self-service A&BI tools to deliver the agility and flexibility that traditional (centralized) A&BI development does not provide for.

The problem is that self-service A&BI tools (including their related semantic modeling capabilities) often turn a once-governed data environment into an uncontrolled, fragmented and inconsistent environment. This fragmented environment creates risk in decision making and breeds loss of trust by end users. The need for consistency in business language, analytics calculations, modeling and representation remains a key focus today as organizations strive to become data-driven.

Definition of "Self-Service" in This Research

Self-service, generally provided through modern A&BI tools, can be divided into two primary categories:

- **Self-service data prep:** *The capabilities and processes that enable data scientists and business users to shape and cleanse data for further analysis. Often, this process results in tables that are shared across multiple analyses.*
- **Self-service analytics:** *The capabilities and features that allow analysts and business users to connect and model data for analysis, without relying on IT or a central developer resource. These models can become part of an organization's semantic layer.*

For the purposes of this research, we will discuss self-service analytics only and refer to it broadly as "self-service."

Today's data and analytics technical professionals are challenged to deliver data and analytics solutions that adhere to the following general principles:

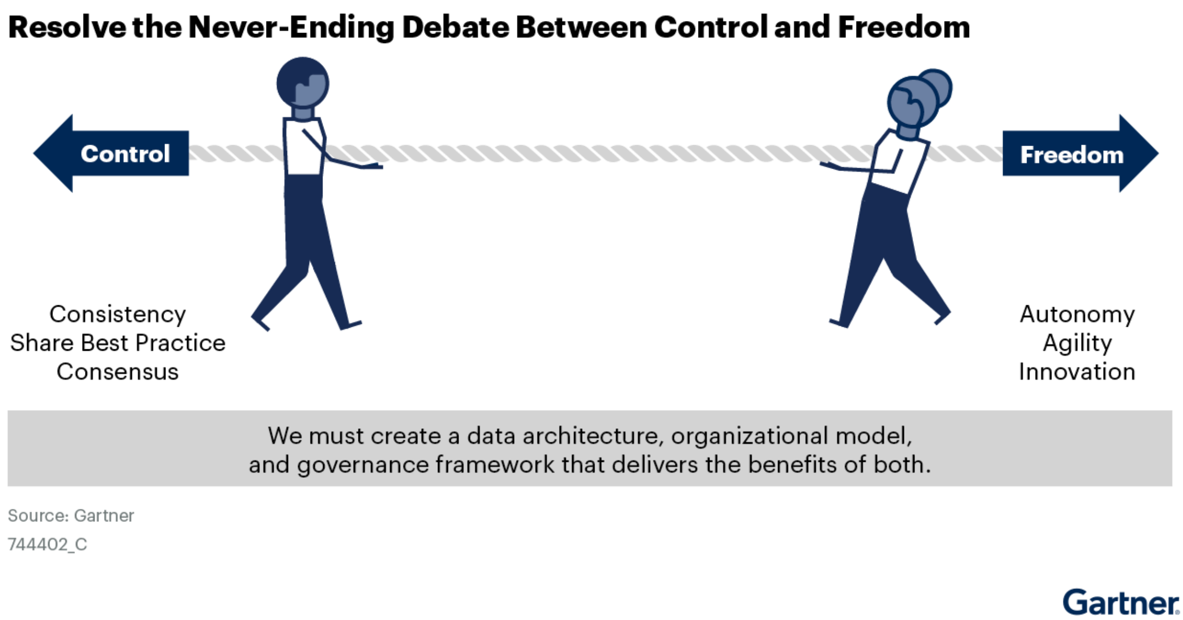
- **Outcome-oriented:** Aligns to the broader organization's goals
- **Valuable:** Provides a benefit to users (e.g., is accessible to a larger population of users, is more flexible, provides more sophisticated analytics or is more affordable)
- **Easy to learn:** Is intuitive to grasp, with training resources available
- **Available and reusable:** Is easily accessed, embedded in applications and workflows, and reusable by multiple systems and users
- **Safe and trusted:**
  - Supports governance policy
  - Includes sophisticated identity, access and security management
  - Is quality-controlled, regulatory-compliant, and delivered by a performant, reliable platform

However, delivering on all of these principles is not easy. And historically, an IT-built semantic layer was a big part of the solution. But organizations struggle to balance analytics development agility with control in order to deliver valuable outcomes to a diverse set of users.

Business units are focused on delivering value in the form of fast, agile analytics in their silos. In the process, they often undermine broader organizational goals for data consistency and a shared set of key performance indicators (KPIs). By contrast, IT departments are occupied with furthering organizational goals for trusted, safe, IT-led data management. Therefore, they don't provide business units with the data or the analytics capabilities that they need in a timely fashion. These competing objectives have spurred a never-ending debate between control and freedom (see Figure 1).

Download All Graphics in This Material

Figure 1: The Never-Ending Debate Between Control and Freedom



> **Because modern tools and use cases offer diverse ways of implementing semantic layers, organizations' views on the ideal semantic layer must evolve.**

Background on Semantic Layers

What Is a Semantic Layer?

Originally introduced as a term in 1991, a semantic layer is a business representation of data. It provides a consistent, unified view of — and access to — organizational data in common business terms. However, the concept of a semantic layer has existed as long as organizations have tried to model and deliver data in end-user terms.

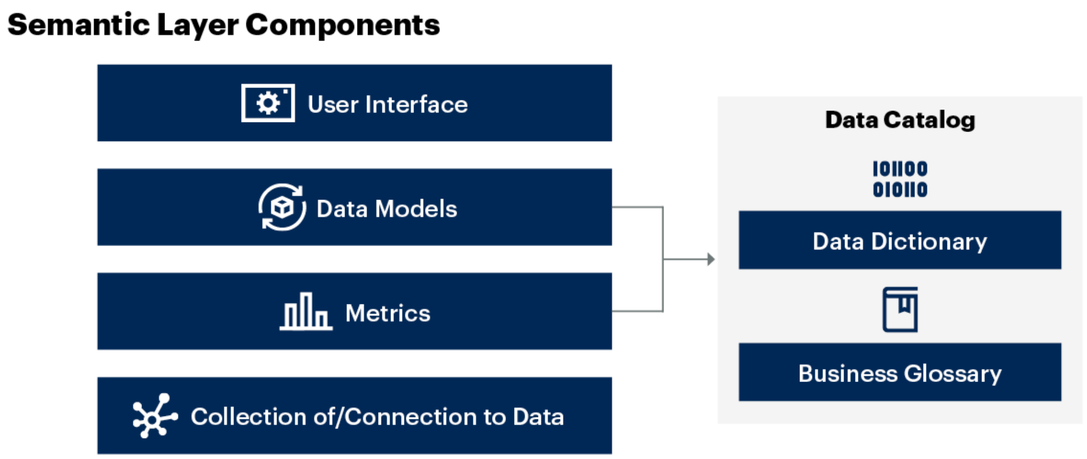
Regardless of how or where a semantic layer is implemented, it should be consumption-tool-agnostic and provide the following core functions:

- A translation of the underlying database structures into business-user-oriented terms and constructs.
- Views of data elements that are intuitive to business users.

- The opportunity to rename data elements so that they make sense to business users.
- An interface to hold business descriptions of data elements.
- A mechanism to define and store calculations and business rules.
- The ability to apply rules and access privileges to KPIs and datasets. (The semantic layer is a junction for role-based access control and auditing.)

With these functions in mind, semantic layers are built upon multiple foundational components, as shown in Figure 2.

Figure 2: Semantic Layer Components



Source: Gartner
783855_C

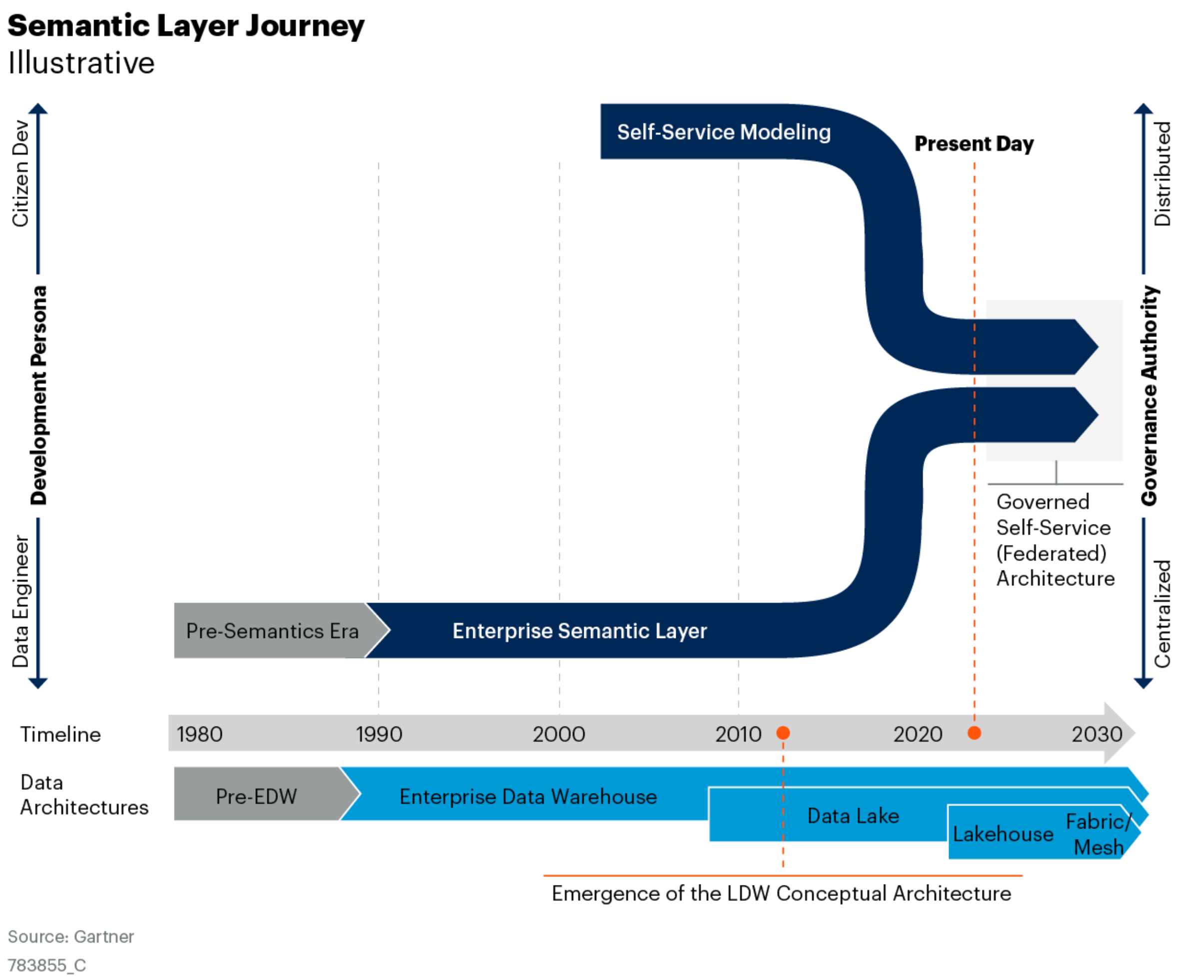
Gartner

Although the broad ambition of the universal semantic layer remains unfulfilled, it is important to understand where the semantic layer has come from and where it sits in today's modern D&A architecture. The following section outlines the history of the semantic layer and the emerging trends data and analytics technical professionals should be aware of.

How Has the Semantic Layer Evolved?

As noted, the concept of the semantic layer is not a new phenomenon and predates the implementation of the enterprise data warehouse commonly found today. Over time, the drive for agility and the rise of self-service capabilities started to overshadow the use of centrally managed semantic layers. Figure 3 shows this movement. Today, we see a convergence of capabilities between independent semantic layer offerings and self-service BI tools.

Figure 3: Semantic Layer Journey



The drive toward this convergence of capabilities has several prevailing themes:

- **Organizations' demands to balance governance and agility:** Organizations that adopted self-service — whether organically or as a reaction to the inflexibility of IT-controlled semantic layers — find themselves straddled with mounting technical debt to maintain fractured views of metrics. They are looking for some enterprise level of governance to guide further development.

- **Organizations' increased demands for integrated analytics:** The growing demand for analytics across use cases, including data science and machine learning (DSML) and integrated applications, has caused many organizations to build dedicated pipelines to serve these needs. Both traditional and self-service semantic models generally have not supported these use cases.

- **Vendor developments to expand use-case support:** Both vendors of stand-alone semantic layers and vendors of self-service A&BI tools are actively developing to achieve the utopian universal semantic layer platform. The emergence of the metrics store concept draws a convergence between the centralized governance of IT-led solutions and the business-user collaboration of self-service platforms.

Between the two primary semantic layer categories (enterprise and self-service), we see two emerging trends:

- The rise of metrics store capabilities as a means of organizing, managing and deploying metrics throughout the organization, regardless of use case

- The expansion of A&BI platform semantic layers into the enterprise space

What Is a Metrics Store?

A metrics store allows users to create and define business metrics as code, govern those metrics from data warehouses, and serve downstream analytics, data science and business applications.

The primary purpose of a metrics store is to capture metric definitions centrally and serve those metrics across any needed analytics use case. In an ideal case, a metrics store would allow business users to create and maintain metric definitions, while enabling IT to act as the infrastructure custodian.

The metrics store broadens stand-alone semantic layers by:

- Enabling business users to contribute and manage metric definitions

- Exposing metrics to use cases beyond general BI and enterprise reporting

The commercialization of metrics stores is still in its infancy. Therefore, it is yet unknown whether metric stores will become their own layer in the analytics stack or be absorbed as a capability of the semantic layer. The implementation of a metrics store offers a compelling capability to define and manage (often disparate) analytics definitions. Data and analytics technical professionals should consider how metrics stores may serve their organization's growing needs as new analytical capabilities are developed. Watch Demystifying the Metrics Store for additional information on metrics stores.

Stand-alone semantic layer vendors that are currently expanding their capabilities with metrics stores include *AtScale, Denodo and Kyligence*. Emerging metrics store vendors include *Cube Dev, GoodData, Metriql, Supergrain, Trace and Transform (acquired by dbt).*

How Are A&BI Platforms Expanding?

Many self-service-focused A&BI tools are also building on the idea of expanding their semantic layer reach. They are doing so in two primary (and related) ways:

1. **By opening access to their semantic modeling layers to third-party A&BI tools.** Common methods for vendors to open access are through:
   - New native connectors
   - APIs
   - Java Database Connectivity/Open Database Connectivity(JDBC/ODBC)
   - JSON
   - XML for Analysis (XMLA)
   - Open Data Protocol (OData)
2. **By incorporating metrics store ideals into their semantic models.** Vendors are adopting the idea of "headless analytics." Headless analytics opens vendors' analytical models to use cases outside of typical A&BI consumption.

This expansion allows A&BI vendors to contend with more stand-alone semantic layer vendors that are A&BI-platform-agnostic.

Thus, these A&BI vendors are challenging stand-alone semantic layer implementations by:

- Becoming more platform-agnostic

- Balancing the (perceived lower or more distributed) costs of A&BI tools
- Leveraging the innovation and collaboration of self-service users already existing in their platform

Table 2 provides examples of expanded platform connection types that A&BI vendors are leveraging.

Table 2: A&BI Vendor Platform Connections

Access Method ↓	*Vendors* ↓
Native connectors	Google
XMLA	Microsoft
ODBC	SAP, Oracle
OData	IBM, Pyramid Analytics
Note: Expanded access to A&BI semantic layers may come with additional licensing fees from the vendor.	

Source: Gartner (April 2023)

Comparing Three Types of Semantic Layers

Data Layer Semantics

Data management is a critical component in the modern digital enterprise, and data warehousing continues to remain the most pragmatic way to process large and complex datasets for timely and trusted insights. Thus, data warehousing continues to be an important component of the logical data warehouse, whether hosted on-premises or in the cloud.

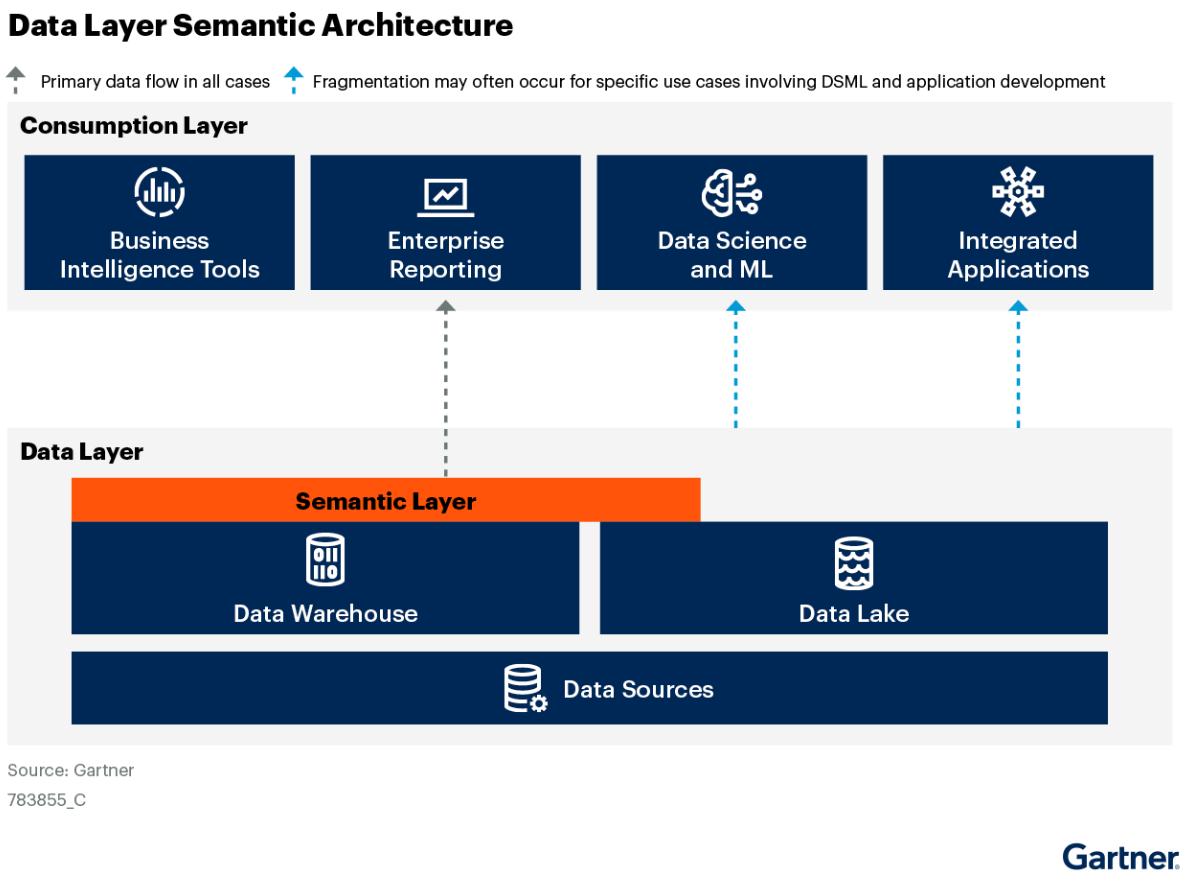
In the context of semantic layers, organizations that are investing in modern DW capabilities are often looking for:

- An efficient approach to semantic layer design through construction data marts, views and materialized views

- Associated business logic and calculations embedded in languages like SQL and Python
- A viable and useful option for a semantic layer that a central data management or analytical development team can maintain

Figure 4 visualizes the general architectural pattern of a semantic layer implemented at the edge of your data layer.

Figure 4: Data Layer Semantic Architecture



Benefits of Semantic Layers as Part of the Data Layer

Building a semantic layer at the edge of your data layer leverages existing technology, architecture and in-house skills to build, maintain and evolve the layer. Because this approach utilizes existing technologies, organizations may see it as a low-cost entry point to implement a semantic layer and deliver trusted data.

Development at this point in the architecture is usually highly governed, and resources are centralized within the organization. This governance and centralization allows organizations to develop consistently, maintain order and provide solution support.

Challenges of Semantic Layers as Part of the Data Layer

Building at the data's edge has several commonly found limitations. Most notably is the pace of change. Business requirements change frequently as markets and conditions fluctuate. The approach is often supported centrally by smaller teams, creating a resource bottleneck for development and support. Additionally, governance requirements at this level tend to control the speed at which changes may go into production. These factors make a semantic layer at the data edge less agile than desired for some organizations.

This approach is not optimal for self-service developers because development in this environment hinges on data engineering skills not commonly found in business users. Data engineers often do not have the context business users consider when designing metrics. Collaboration between business users and data engineers is both necessary and time-consuming to effectively produce analytical artifacts.

In addition to these concerns, the following limitations are common:

- Building many, many views doesn't scale, and can involve a lot of development work by a relatively small team.
- Direct query against models that sit in a data warehouse may encounter performance issues if the queries have not been optimized according to the requirements of both the A&BI tool and the data source. In scenarios where the A&BI tool and the data warehouse are in different clouds, laws of physics still apply, and only limited query optimization may be possible.
- The calculations and business logic required for a semantic layer to work are not always natively available in the data warehouse, necessitating some additional development of this logic, often in a different programming language like Python.
- Analytical data pipelines often become fragmented, as these layers focus primarily on structured data and common reporting and BI use cases. Semistructured or unstructured data residing in the data lake may not be accessible through these platforms. Additionally, the pipelines required for DSML and application development are not always well-supported with this approach.

Representative Vendors

Amazon, Google, IBM, Microsoft, Oracle, SAP, Snowflake, Teradata

---

How Are Semantic Layers Impacted by Data Mesh, Data Fabric and Graph?

Data mesh and fabric are data management concepts that focus on the decentralization of data management through the use of metadata.

The data mesh concept prioritizes decentralization of data management, enabling business units or functions to own, manage and govern their data as a product. Characteristics of data mesh include:

- *It leverages input from subject matter experts (SMEs) to curate metadata.*
- *Data products are designed deliberately, but are subject to SME bias.*

Data fabric is an automation pattern used to derive data products by analyzing metadata and automating data management tasks. Characteristics of data fabric include:

- *It incorporates continuous model learning and evaluation of metadata.*
- *Data products are derived, but may be limited by insufficient metadata.*

Both data mesh and fabric may complement each other in an organization, the outputs of which could be consumed and delivered as part of a semantic data layer. For additional information on data mesh and data fabric, see Quick Answer: Comparing Data Fabric and Data Mesh.

As noted in Building Knowledge Graphs, a knowledge graph is a network of meaningful concepts that are interlinked to build semantic networks. Semantic networks, in turn, help establish relationships between concepts and define how they are interconnected. A knowledge graph then serves as a platform to formally unify knowledge acquisition with data management. Knowledge graphs provide a means to understanding the context of data, and that is the key to comprehending data. As such, like data mesh and fabric, knowledge graphs may provide useful underlying data to inform and further develop an organization's semantic data layer.

Simply put, data mesh, fabric and graph all represent areas of data management that may significantly assist in the development of semantic layers. By themselves, they do not contain the capabilities necessary to provide the components identified for semantic layers outlined in Figure 2.

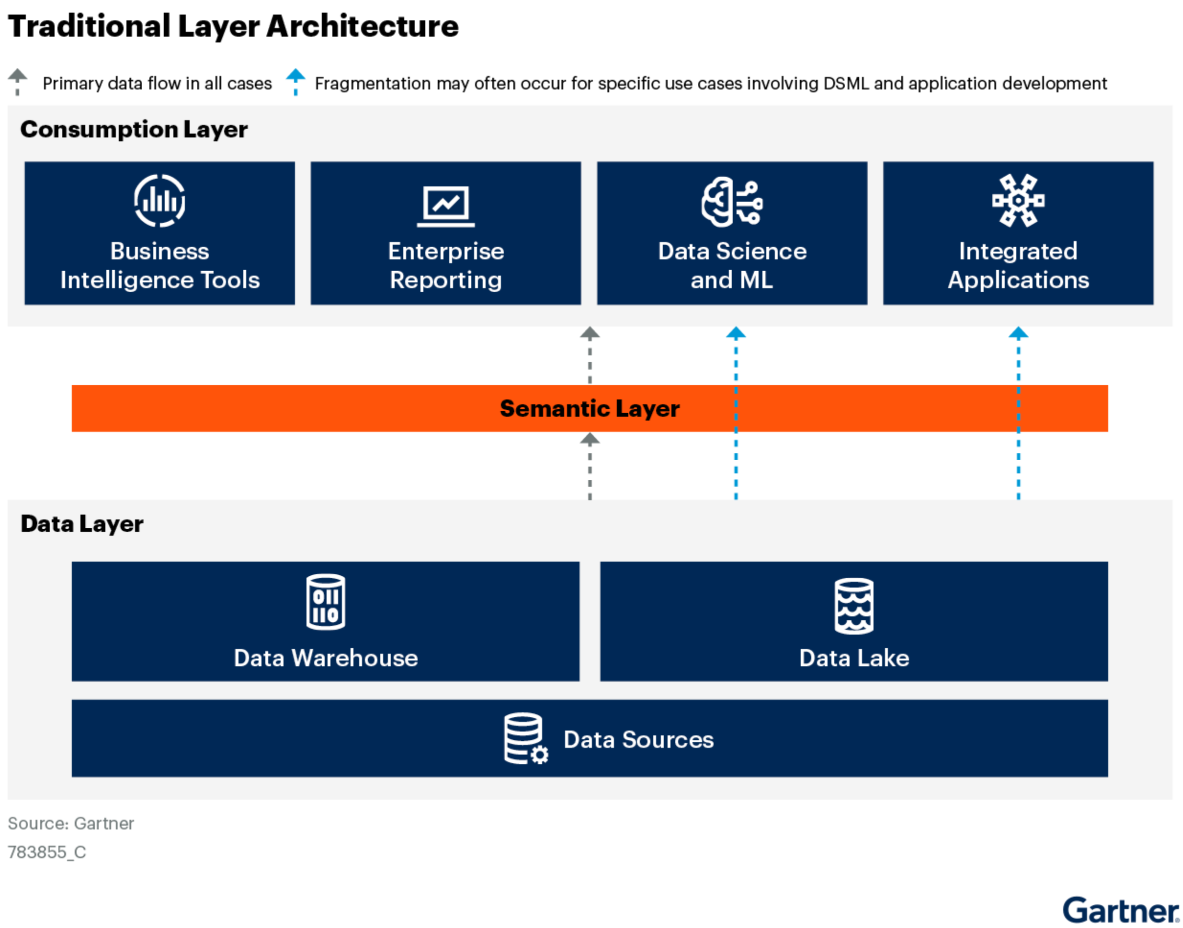
Stand-Alone Semantic Layers

Organizations that collect data in diverse forms and formats find that traditional data warehousing technologies can't meet all their growing business needs. This shortcoming causes organizations to invest in a scale-out data lake architectural pattern. In this approach, the semantic layer can be considered part of the logical data warehouse.

Although similar to data layer semantics in providing governance and centralization, a traditional (or stand-alone) semantic layer offers capabilities to connect disparate data across an organization's data landscape. Additionally, organizations implementing these solutions will find performance optimization options through data virtualization and abstraction capabilities. For additional information on data virtualization, see Assessing the Relevance of Data Virtualization in Modern Data Architectures.

Figure 5 visualizes the general architectural pattern of an independent semantic layer implemented as a stand-alone architectural layer.

Figure 5: Stand-Alone Semantic Layer Architecture



Benefits of Stand-Alone Semantic Layers

Stand-alone semantic layer implementations expand on many of the same benefits of building out layers at the data layer directly. And like deploying semantics in the data layer, implementation at this point in the architecture is highly governed, and development resources are centralized within the organization. This governance and centralization allows organizations to develop consistently, maintain order and provide solution support.

Organizations want to collect diverse forms of data that reside in multiple formats and locations across the enterprise, and then analyze this data for discovery analytics use cases. Traditional data warehousing technologies can't meet these growing business needs, causing organizations to invest in data lake architectural patterns to complement the data warehouse. Building a semantic layer on top of a data lake removes some of the classic problems of lakes: understandability, performance and SQL access. This approach makes a data lake more like a lakehouse, which can perform analytics more easily and reliably.

As a virtualized layer, semantics implemented in this scenario offers the ability to create a virtual model of data that joins relational and nonrelational data, from many sources, on-premises and/or in the cloud. It simplifies data access for analytics, standardizes metric calculations, improves reuse and reduces change impact to achieve a consumption-tool-agnostic, consistent, reusable semantic layer.

Challenges of Stand-Alone Semantic Layers

In this semantic layer implementation scenario, a new component to the organization's architecture is introduced on top of the data layer to extend consistent access to enterprise data warehouses, and often the data lake, thus creating a lakehouse-type architecture. (See Exploring Lakehouse Architecture and Use Cases for additional information on lakehouse architecture.)

This new layer is highly centralized and code-centric. As with semantics implemented at the data layer, implementation, development and support generally rely on IT technical teams. This reliance on IT makes stand-alone semantic layers efficient for development and governance, but often slow to adapt to changing business requirements. Accessibility for business users to collaborate on development of models and metrics is often limited.

In addition to these concerns, the following limitations are common:

- These layers are traditionally developed to support enterprise and self-service reporting and analytics. DSML and application development are not always well-supported with this approach.

- These central tools are generally expensive to implement, thus limiting possible benefits for smaller organizations.
- Using data virtualization to deliver data will not remedy a poorly governed data lake.

Representative Vendors:

Amazon, Apache Druid, Apache Pinot, AtScale, ClickHouse, Data Virtuality, dbt, Denodo, Dremio, Google, Microsoft, Kyligence, Kyvos Insights, Zetaris

A&BI Tool Semantic Layers

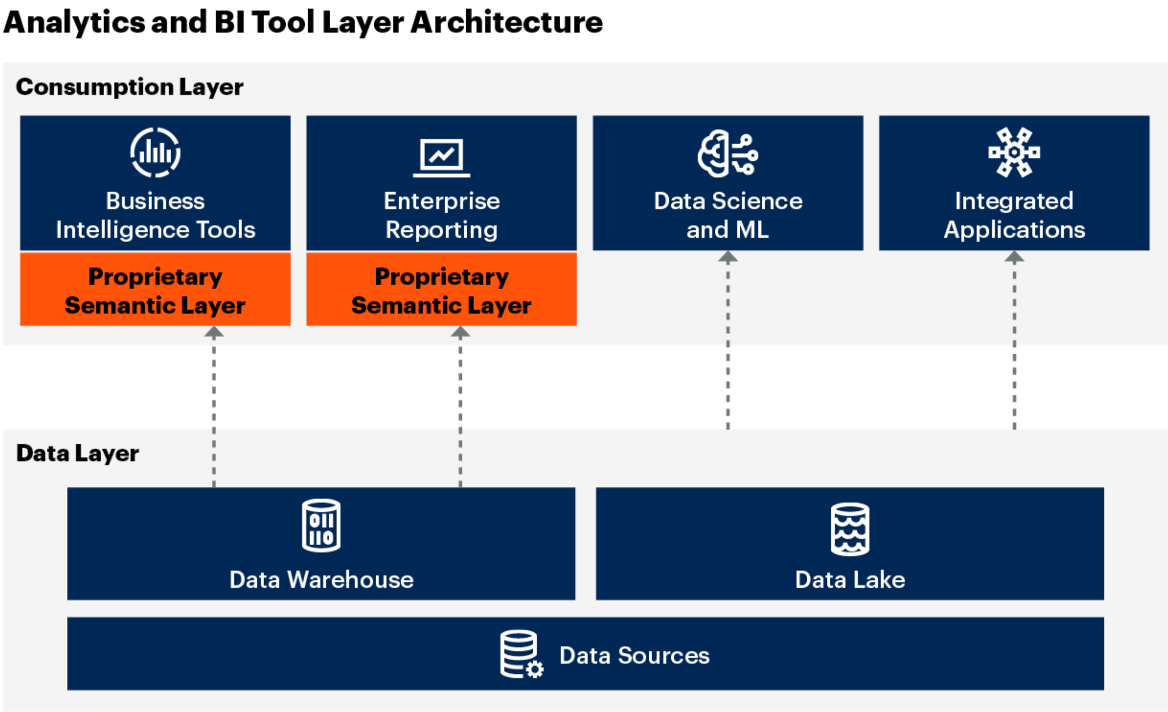
Modern analytics and BI platforms offer their own interpretation of the semantic layer. Users are empowered to load data into the platform, develop analytical models with dimensions and measures, include augmented analytics, and share across the organization. These tools offer the flexibility unavailable with centrally managed semantic layer scenarios.

Organizations implement self-service A&BI tools, in many cases, to:

- Relieve resource pressures on IT
- Empower business users to analyze data ad hoc in response to continued organizational demands
- Provide augmented analysis capabilities to non-DSML-specific analysts

For more information on the capabilities in modern self-service A&BI platforms, see Evolving Capabilities of Analytics and Business Intelligence Platforms. Figure 6 visualizes the general architectural pattern of semantic layers as implemented as part of A&BI tool implementations.

Figure 6: Analytics and BI Tool Semantic Layer Architecture



Source: Gartner
Note: Semantics often built into individual pipelines, resulting in redundant efforts and potential drift in definitions
783855_C

Gartner

Benefits of Semantic Layers in A&BI Tools

Modern analytics and BI tools are not limited to visual presentation of data. Their strength is in their capabilities to allow nontechnical users to connect with and model data, create metrics, and visualize results, thus reducing time to insight for end users. To that end, the strongest benefit of A&BI tools is in the flexibility and agility they bring to the analytics capabilities of an organization.

The semantic modeling capabilities of modern A&BI tools typically leverage a graphical, no-code/low-code interface focused on business users and citizen developers. This shifts the development resources from a centralized, technical model in other semantic layer scenarios to a distributed model where business owners can effectively become analytical product owners. In addition, some vendors provide automodeling and data refinement capabilities, coupled with a networked semantic layer. This means that organizations can skip a lot of the slow manual steps of modeling their data and gain efficiencies to more quickly move their analytics into production.

Challenges of Semantic Layers in A&BI Tools

With great agility comes great governance challenges. Among the chief concerns with self-service A&BI tools is governance. Without controls, analytical models may be developed and shared, thereby creating several key risks:

- Data may be modeled and visualized inconsistently across the organization. Datasets developed in A&BI tools generally have no overarching data model to demonstrate how the different datasets are related to each other. In many cases, the expected same metric within different models may be calculated differently. With an inconsistent understanding of the calculation, these varied definitions will lead to different results. Decision makers could take action on incorrect results with significant consequences.
- Sharing is generally streamlined in these platforms. Without controls or verification, data may be shared with unauthorized users. Although this may be done without malicious intent, inadequate sharing rules may leave the organization open to unnecessary risk of exposing confidential data.
- Ownership over analytical models may also be questioned. As members move around or out of the organization, there is risk that models in use may become unsupported or abandoned entirely. Succession of ownership planning is often not considered as part of enabling self-service developers.

Another significant problem with building a semantic layer inside an A&BI platform is that it is typically proprietary, and not very reusable across other BI tools or other data sources. So the solution may work quite well — until there is a business requirement to use a different A&BI tool or to move the data to a different location or platform. In those scenarios, the organization is typically looking at a time-consuming migration. If this is built inside a cloud vendors' platform, there may also be egress charges for the data that moves out of the particular cloud vendor. As noted in the How Has the Semantic Layer Evolved section, some vendors are addressing this through the implementation of metric-store-type functionality that broadens the capability to act as a stand-alone semantic layer. However, this is still in early stages.

And lastly, as noted in Figure 6, these tools are primarily focused on reporting and analytics use cases. Although many offer native augmented analytics capabilities, these capabilities are not a substitute for more intensive data science and machine learning use cases.

Representative Vendors

Google (Looker), IBM, Microsoft, MicroStrategy, Oracle, Pyramid Analytics, Qlik, Salesforce (Tableau), Sisense, ThoughtSpot

Guidance

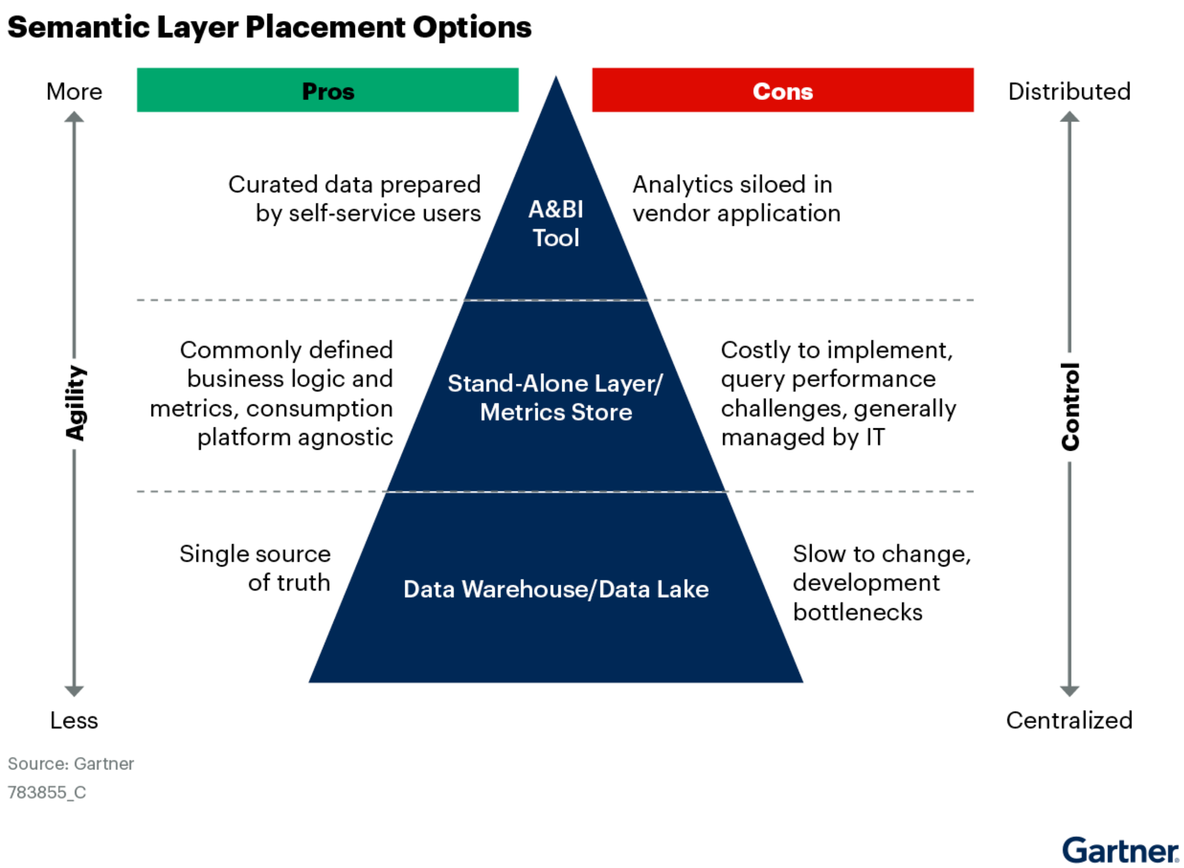
Plan for a Federated Semantic Layer Architecture

The ambitious goal of delivering a single, universal semantic layer across the organization has not yet been achieved by any one technology. As a result, organizations should think about how to implement a semantic layer that resides throughout their A&BI, data integration, data lake, data warehouse and data virtualization (DV) tiers. The federation of this area in the data and analytics architecture must combine the governance of traditional, enterprise-level deployments with the collaborative, distributed ownership found in current A&BI tools.

The move toward federation becomes possible as enterprise semantic approaches evolve to serve increasingly diverse analytics use cases and provide for the collaboration found in common A&BI tools. The ubiquity of common A&BI tools, combined with their expansion to engage third-party BI tools, provides the foundations for innovation and agility in the analytics modeling used across the enterprise.

In addition to expanding capabilities in semantic layer approaches, success will be predicated on a mature governance and operationalization strategy for implementing and maintaining semantic layers. Figure 7 provides a simplified view of semantic layer placement options with factors organizations may consider as they build out their architecture.

Figure 7: Semantic Layer Placement Options



Base Semantic Layer Decisions on Analytics Styles

As part of implementing a federated semantic layer, technical professionals should evaluate technologies against the goals of their analytics initiatives. Different semantic layers have affinities with different styles of analytics (see Table 3).

Table 3: Semantic Layer and Analytics Styles

(Enlarged table in Appendix)

	Decentralized Analytics	Centralized Analytics	Federated Analytics
Semantic Layer Priority	Ease of use and user enablement	Control, governance and consistency	A balance of agility and control
Semantic Layer Technology Approach	- Local semantic layer: Modern A&BI platforms that offer maximum usability and power to end users for semantic model creation - Global semantic layer: Data lake enablement platforms that offer significant query and semantic layer capabilities for a broad set of users	- Local semantic layer: Query-focused A&BI platforms that offer access to data warehouse data for model access and temporary tables for creation and customization - For global access to data: Data warehouse and DW automation, which can populate data marts that are centrally created and provisioned	- Local semantic layer: Depending on the use case, a combination of approaches employed, including import mode into an A&BI tool, direct query to the data warehouse, data preparation workflows in a data virtualization or data prep tool, or data that is made accessible via a lake enablement platform - Global semantic layer: The LDW (i.e., a combination of data warehouse views, data lake shared packages of data, and data virtualization) for access to data that isn't in the first two categories; for mature use cases, a data fabric, which adds metadata-driven recommendations and AI/ML insights
Strengths	- Easiest way to cheaply enable access to new data sources	- Easiest way to govern data access - Organizations have built up a semantic virtual tier and minimized the uncontrolled proliferation of data marts	- Flexibility, as the advanced-level LDW is in a constant, dynamic state of change - These changes occur in response to changes in business analytics requirements, moving D&A workloads among the analytics engines in the stack - LDW contains every event, transaction, interaction, sensor reading, customer, employee and supplier — any and every entity
Weaknesses	- Data quality issues are likely to arise with siloed A&BI adoption - Data lake initiatives often fail to reach maturity because of the difficulty of data governance	- So far, only partial success, due to low adoption, dissatisfied end users and high costs - Development processes are rigid and time-consuming - Badly designed batch data movement limits agility and delivery efficiency - New types of data such as IoT, social media, weblogs and geospatial are not supported in existing architecture	- The LDW model, if insufficiently supported by resources for implementation and management, combines the complexity and chaos of decentralization with the cost and slowness of centralization - The risk of failure due to complex implementation increases with the ambitiousness of this project

Source: Gartner (April 2023)

Evaluate User Persona Expectations

In addition to evaluating technical criteria, contextualize what user expectations are when they employ a specific analytical platform. Use the examples of user roles in Table 4 to gain an understanding of the expectations of various personas. These examples provide some insight to the capabilities technical professionals should consider when selecting semantic layer architectures for delivering analytics.

Table 4: User Persona Outcomes and Feature Expectations

(Enlarged table in Appendix)

	Consumer	Explorer	Innovator	Expert
Desired Outcome	View analytics content periodically; use it to make data-driven decisions.	Select from available fields in a semantic layer to seek out diagnostic analytics; discover the answers to "why" questions.	Mashup multiple certified data sources, query against large datasets, create novel visualizations and generate new insights on data that may have an impact on the organization's future.	Introduce data from completely new data sources, create data transformation scripts, and use advanced analytics and ML to build transformational analytics.
Augmented Analytics Features	Natural language text and voice query, and autogenerated insights.	Natural language processing (NLP), SQL generation, automatic visualization generation and jargon-free ML services that provide insights such as key-driver analysis.	Automatic data profiling and data classification join recommendations, visual lineage and impact analysis for data changes, and menu-driven advanced analytics functions.	Rich transformation and query language functions available, and deep capabilities with R, Python, PMML, and augmented ML available to decrease time to production for advanced analytics.
Usability Features	Analytics visualizations should be searchable, prepopulated and customized to the users' needs. Definitions of metrics and measures are easily accessible and linked to dashboard objects.	In addition to consumer features, data should be organized and linked with rich lineage and metadata, with the ability to open data in analytics tools that offer drag-and-drop visualization and analytics capabilities.	In addition to explorer features, sophisticated data preparation capabilities with embedded forecasting, classification and clustering functions should be available.	In addition to explorer and innovator features, advanced data source ingestion and/or connectivity, configuration management and monitoring, and interfaces to and from DSML and augmented ML should be available.
Business Workflow Affinity	Aligned to information portal capabilities and embedded in business applications, not just inside the BI tool, with powerful mobile app-based data access. Reporting capabilities and linkage to productivity applications like Excel are important.	Aligned to analytics workbench capabilities, including the ability to explore and mash up data to deliver new insights.	Aligned to data science hub, with the ability to create features by enriching, joining external sources with semantic layer data. A feedback loop exists to allow more sophisticated users to enrich the semantic layer with more metadata or update fields.	Aligned to artificial intelligence hub, specifically, by automating and augmenting key portions of analytics processes. A&BI functions are programmable, automatable, repeatable, reusable, and integrated into the experts' preferred open-source or packaged toolchain for DSML and AI.
Security and Governance Features	Guardrails have been set that ensure consumers can't access data that they shouldn't.	In addition to consumer capabilities, a data catalog offers a view of what data exists. But actually accessing this data requires explorers to follow the request and approval process.	In addition to consumer and explorer capabilities, when creating datasets in A&BI tools, innovators can apply row-level security, or take advantage of SSO to a trusted, centrally managed data source.	In addition to innovator capabilities, experts are trained in data governance, so that they can enforce governance rules that exist. Additionally, they have a very granular ability to secure, monitor, and measure data analytics capacity, usage and performance.

Source: Gartner (April 2023)

Set Expectations and Prioritize Goals

Enabling an analytics solution that supports diverse use cases requires organizations to balance many goals. Self-service architectures are outcome-oriented and agile. Enterprise semantic layers prioritize governance and provide agnostic access to data.

The ideal semantic layer should have the following characteristics:

- **Access efficiency:** Semantic layers need to be easy for users to access and use. Folder structures are a start, but good semantic layers should be searchable and linked to rich metadata. They should visualize relationships in an intuitive way and offer collaboration capabilities.

- **Integrability:** Semantic layers should allow for integration with different data stores and file formats. A semantic layer should be able to connect to these various layers and allow you to build abstractions, definitions, metrics and measures on top of them to make data more accessible and standardized.

- **Development efficiency:** Semantic layers must make development efficient for technical professionals but easy enough for citizen developers. Semantic layers should allow for integration with different data stores and file formats and be able to connect to these various layers, thus allowing users to build abstractions, definitions, metrics and measures.

- **Platform efficiency:** Semantic layers must deliver sufficient platform efficiency to service many concurrent users and applications. Moreover, the data should be flexible, connectible and consumption-platform-agnostic.

- **Data security:** A good semantic layer should be able to inherit authentication, authorization and access controls, so that users' identities are tied to their data permissions inside the semantic layer. Moreover, data-at-rest and data-in-motion security, as well as row-, role- and column-based security, are important to enable organizations with granular security controls.

As a result, there are no perfect semantic layer solutions — only different sets of trade-offs — and many of these trade-offs are based on the technologies that are used to host the semantic layer. Technical professionals should use these characteristics, combined with the analyses of analytical styles and user personas, as a starting point for their semantic layer architecture decisions.

Document Revision History

Demystifying Semantic Layers for Self-Service Analytics - 7 September 2021

Recommended by the Authors

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

Solution Path for Building Modern Analytics and BI Architectures

Reference Architecture to Enable Self-Service Analytics

Graph Technology Applications and Use Cases

Exploring Lakehouse Architecture and Use Cases

Evolving Capabilities of Analytics and Business Intelligence Platforms

Data Modeling to Support End-to-End Data Architectures

Innovation Insight: Metrics Stores

Table 1: Common Semantic Layer Implementations

	Data Layer Semantics	*Stand-Alone Semantic Layer*	*A&BI Tool Semantic Layer*
Common Features	■ Is an extension of data store capabilities ■ Sits at the edge of the data layer ■ May include views (including materialized), data marts and graphs	■ Is an independent platform solution ■ Sits between the data and consumption layers ■ May include virtualization, abstraction and data lake enablement	■ Is a BI tool capability ■ Is part of the analytics consumption layer ■ Includes direct query and ingestion-type models
Developer Roles	Data engineers	Data engineers	Business analysts and data analysts
Governance Localization	Centralized	Centralized	Distributed
Key Strengths	■ Is highly governed ■ Leverages in-house technology and skills ■ Can deliver high volumes of data to many concurrent users	■ Is well-governed ■ Serves as a centralized source for analytical modeling and metrics ■ Connects with diverse data formats across varying platforms	■ Enables flexible and agile implementation and development ■ Democratizes analytics development ■ Reduces time to insight

	Data Layer Semantics	*Stand-Alone Semantic Layer*	*A&BI Tool Semantic Layer*
Key Challenges	▪ Is highly reliant on central data engineering resources ▪ Is slow to change ▪ Has limited capabilities with unstructured data	▪ Involves IT-heavy development and implementation ▪ Is traditionally not designed to support data science and machine learning (DSML) and integrated application use cases ▪ Introduces another expensive layer to the data and analytics solution	▪ Is less governed ▪ Offers inconsistent development and deployment of analytical models and metrics ▪ Is often isolated within the vendor stack
Example Vendors*	IBM, Oracle, SAP, Snowflake, Teradata	Apache Druid, Apache Pinot, AtScale, ClickHouse, Data Virtuality, dbt, Denodo, Dremio, Kyligence, Kyvos Insights, Zetaris	MicroStrategy, Pyramid Analytics, Qlik, Salesforce (Tableau), Sisense, ThoughtSpot

Note:
* Major vendors, including Amazon, Google, Microsoft, Oracle and SAP, offer capabilities to support semantic layers in multiple architectural patterns.

Source: Gartner (April 2023)

Table 2: A&BI Vendor Platform Connections

Access Method	Vendors
Native connectors	Google
XMLA	Microsoft
ODBC	SAP, Oracle
OData	IBM, Pyramid Analytics
Note: Expanded access to A&BI semantic layers may come with additional licensing fees from the vendor.	

Source: Gartner (April 2023)

Table 3: Semantic Layer and Analytics Styles

	Decentralized Analytics	*Centralized Analytics*	*Federated Analytics*
Semantic Layer Priority	Ease of use and user enablement	Control, governance and consistency	A balance of agility and control

	Decentralized Analytics	Centralized Analytics	Federated Analytics
Semantic Layer Technology Approach	■ Local semantic layer: Modern A&BI platforms that offer maximum usability and power to end users for semantic model creation ■ Global semantic layer: Data lake enablement platforms that offer significant query and semantic layer capabilities for a broad set of users	■ Local semantic layer: Query-focused A&BI platforms that offer access to data warehouse data for model access and temporary tables for creation and customization ■ For global access to data: Data warehouse and DW automation, which can populate data marts that are centrally created and provisioned	■ Local semantic layer: Depending on the use case, a combination of approaches employed, including import mode into an A&BI tool, direct query to the data warehouse, data preparation workflows in a data virtualization or data prep tool, or data that is made accessible via a lake enablement platform ■ Global semantic layer: The LDW (i.e., a combination of data warehouse views, data lake shared packages of data, and data virtualization) for access to data that isn't in the first two categories; for mature use cases, a data fabric, which adds metadata-driven recommendations and AI/ML insights

	Decentralized Analytics	Centralized Analytics	Federated Analytics
Strengths	- Easiest way to cheaply enable access to new data sources	- Easiest way to govern data access - Organizations have built up a semantic virtual tier and minimized the uncontrolled proliferation of data marts	- Flexibility, as the advanced-level LDW is in a constant, dynamic state of change - These changes occur in response to changes in business analytics requirements, moving D&A workloads among the analytics engines in the stack - LDW contains every event, transaction, interaction, sensor reading, customer, employee and supplier — any and every entity

	Decentralized Analytics	Centralized Analytics	Federated Analytics
Weaknesses	■ Data quality issues are likely to arise with siloed A&BI adoption ■ Data lake initiatives often fail to reach maturity because of the difficulty of data governance	■ So far, only partial success, due to low adoption, dissatisfied end users and high costs ■ Development processes are rigid and time-consuming ■ Badly designed batch data movement limits agility and delivery efficiency ■ New types of data such as IoT, social media, weblogs and geospatial are not supported in existing architecture	■ The LDW model, if insufficiently supported by resources for implementation and management, combines the complexity and chaos of decentralization with the cost and slowness of centralization ■ The risk of failure due to complex implementation increases with the ambitiousness of this project

Source: Gartner (April 2023)

Table 4: User Persona Outcomes and Feature Expectations

	Consumer	Explorer	Innovator	Expert
Desired Outcome	View analytics content periodically; use it to make data-driven decisions.	Select from available fields in a semantic layer to seek out diagnostic analytics; discover the answers to "why" questions.	Mashup multiple certified data sources, query against large datasets, create novel visualizations and generate new insights on data that may have an impact on the organization's future.	Introduce data from completely new data sources, create data transformation scripts, and use advanced analytics and ML to build transformational analytics.
Augmented Analytics Features	Natural language text and voice query, and autogenerated insights.	Natural language processing (NLP), SQL generation, automatic visualization generation and jargon-free ML services that provide insights such as key-driver analysis.	Automatic data profiling and data classification join recommendations, visual lineage and impact analysis for data changes, and menu-driven advanced analytics functions.	Rich transformation and query language functions available, and deep capabilities with R, Python, PMML, and augmented ML available to decrease time to production for advanced analytics.

↓	Consumer ↓	Explorer ↓	Innovator ↓	Expert ↓
Usability Features	Analytics visualizations should be searchable, prepopulated and customized to the users' needs. Definitions of metrics and measures are easily accessible and linked to dashboard objects.	In addition to consumer features, data should be organized and linked with rich lineage and metadata, with the ability to open data in analytics tools that offer drag-and-drop visualization and analytics capabilities.	In addition to explorer features, sophisticated data preparation capabilities with embedded forecasting, classification and clustering functions should be available.	In addition to explorer and innovator features, advanced data source ingestion and/or connectivity, configuration management and monitoring, and interfaces to and from DSML and augmented ML should be available.
Business Workflow Affinity	Aligned to information portal capabilities and embedded in business applications, not just inside the BI tool, with powerful mobile app-based data access. Reporting capabilities and linkage to productivity applications like Excel are important.	Aligned to analytics workbench capabilities, including the ability to explore and mash up data to deliver new insights.	Aligned to data science hub, with the ability to create features by enriching, joining external sources with semantic layer data. A feedback loop exists to allow more sophisticated users to enrich the semantic layer with more metadata or update fields.	Aligned to artificial intelligence hub, specifically, by automating and augmenting key portions of analytics processes. A&BI functions are programmable, automatable, repeatable, reusable, and integrated into the experts' preferred open-source or packaged toolchain for DSML and AI.

	Consumer	Explorer	Innovator	Expert
Security and Governance Features	Guardrails have been set that ensure consumers can't access data that they shouldn't.	In addition to consumer capabilities, a data catalog offers a view of what data exists. But actually accessing this data requires explorers to follow the request and approval process.	In addition to consumer and explorer capabilities, when creating datasets in A&BI tools, innovators can apply row-level security, or take advantage of SSO to a trusted, centrally managed data source.	In addition to innovator capabilities, experts are trained in data governance, so that they can enforce governance rules that exist. Additionally, they have a very granular ability to secure, monitor, and measure data analytics capacity, usage and performance.

Source: Gartner (April 2023)