





Buyers Guide
2021

The Complete Buyer's Guide for a Semantic Layer

10 Things to Consider When Modernizing Your
Analytics Infrastructure

ATSCALE



Unprecedented levels of data scale and distribution are making it almost impossible for organizations to effectively exploit their data assets. Data and analytics leaders must adopt a semantic approach to their enterprise data assets or face losing the battle for competitive advantage.

GARTNER

[“How to Use Semantics to Drive the Business Value of Your Data”](#)

27 November 2018

ABOUT THIS GUIDE

Semantic layers have been around for some time. They were invented as a way to mold relational databases and their SQL dialects into an approachable Interface for business users. In 1992, Business Objects patented the term and formalized their implementation as the Business Objects Universe™. From that point forward, the concept of measure and dimensions as an abstraction of SQL has become the preferred language for business users.

Until recently, however, the semantic layer was always tightly coupled to the business intelligence (BI) platform. As a result, tools like Business Objects had their unique semantic layer, separate and distinct from Cognos' semantic layer, MicroStrategy's semantic layer, Tableau's semantic layer and so forth. As long as enterprises stayed within the walled garden of the BI platform vendor of choice, all was good. Today, there are a variety of ways of analyzing data and long gone are the days where there was one BI platform to rule them. Tightly coupling a semantic layer to one analytics consumption style just no longer makes sense.

To expand on that, the explosion of self-service BI has created some unintended consequences. While business users freed themselves from the chains of IT-prepared analytics, data consistency and trust in analytics' output took a huge hit. Business definitions and terms have become mutable, malleable, and subject to interpretation. It's great that business users have more tools to perform BI themselves, but they need to be working off of consistent, high-quality data because the cost of bad data is enormous. According to IBM, poor data quality costs the US economy around \$3.1 trillion annually.

It's time for a new approach to driving trust in the numbers that better fits better with today's data volume, velocity and variety. In this guide, we will look at the different approaches to selecting and implementing a semantic layer for your analytics stack that will drive consistency, ease of use and trust for a wide variety of analytics consumption types and use cases.

TABLE OF CONTENTS

What Is a Semantic Layer?	1
The Top 5 Signs You Need to Invest in a Semantic Layer	2
Business Units or Groups Have Strong Preferences for Different Analytics Tools	2
Business Analysts and/or Data Scientists Complain About a Lack of Data Access	3
The Slow Pace of Data Integration Drives the Business to Build their Own Solutions	3
Reports from Different BI Tools Use Similar Terms but Show Different Results	4
Business Executives Express Doubts About Their Confidence in the Numbers	4
Getting Started On Your Search	5
Key Considerations	6
Not Tied to a Single Consumption Style	6
Offers Tabular and Multidimensional Views	7
Supports Data Platform Virtualization	7
Easy Model Development and Sharing	8
Ability to Express Business Concepts and Functions	8
Query Performance & Caching	8
Support for Business Intelligence and Data Science Workloads	9
Security & Governance	9
Feature Checklists	11
Conclusion	12
Resources & Further Reading	13



| WHAT IS SEMANTIC LAYER?

In defining a semantic layer, I still haven't found a better definition than that of Wikipedia's:

“A semantic layer is a business representation of corporate data that helps end users access data autonomously using common business terms. A semantic layer maps complex data into familiar business terms such as product, customer, or revenue to offer a unified, consolidated view of data across the organization.

By using common business terms, rather than data language, to access, manipulate, and organize information, a semantic layer simplifies the complexity of business data. Business terms are stored as objects in a semantic layer, which are accessed through business views.

The semantic layer enables business users to have a common "look and feel" when accessing and analyzing data stored in relational databases and OLAP cubes. This is claimed to be core business intelligence (BI) technology that frees users from IT while ensuring correct results.

Business Views is a multi-tier system that is designed to enable companies to build comprehensive and specific business objects that help report designers and end users access the information they require. Business Views is intended to enable people to add the necessary business context to their data islands and link them into a single organized Business View for their organization.”

Source: Wikipedia (https://en.wikipedia.org/wiki/Semantic_layer)



THE TOP 5 SIGNS YOU NEED TO INVEST IN A SEMANTIC LAYER

While working with a variety of customers in a number of different industries, we found that they shared a common set of symptoms resulting from the ailment of a missing semantic layer. If the following situations sound familiar, you should keep reading.

1. BUSINESS UNITS OR GROUPS HAVE STRONG PREFERENCES FOR DIFFERENT ANALYTICS TOOLS

The larger the organization, the tougher it becomes to impose a single standard for consuming and preparing analytics. Whether through acquisitions or just the strong will of business users, forcing a single tool or analytics style is a futile endeavor. The large enterprises we work with are dealing with dozens of BI tools, all with their own versions of the truth. According to the [Dresner's Wisdom of Crowds® Business Intelligence Study](#), over half of enterprises report using three or more BI tools, with over a third using four or more.

On top of that, the advent of the data scientist as yet another analytics consumer creates an even more dire situation. Now, not only do business analysts risk creating bad reports, data scientists risk creating misleading predictions - both have profound implications for business results.

To make matters worse, the pace of innovation in cloud data warehousing, BI and AI/ML has created a constant cycle of upgrades, re-platforms and re-factors.

If you find yourself at the losing end of dictating analytics tools and consumption styles in your organization, don't fret. By providing "analytics-as-a-service" to your business users and data scientists, you can have your cake and eat it too: let your users consume the way that makes sense for their use case while ensuring semantic consistency and data governance.

2. BUSINESS ANALYSTS AND/OR DATA SCIENTISTS COMPLAIN ABOUT A LACK OF DATA ACCESS

There's rarely a lack of data in an enterprise but there's often a lack of understandable data. Data without metadata is practically useless. Whether it's data in log files or data in relational tables, without a business context, it's left to the analyst or data scientist for interpretation. In other words, data without business intelligence is useless and can be even dangerous. This is not an uncommon phenomenon. According to Gartner, [87 percent of organizations have low BI and analytics maturity](#).

If you hear your analytics consumers complaining that they lack data to make decisions, your organization may be suffering from data without metadata. Without a semantic layer powered by a data model, your organization may be slow to respond to changing market conditions. Business analysts and data scientists need a business context to turn raw data into actionable insights.

3. THE SLOW PACE OF DATA INTEGRATION DRIVES THE BUSINESS TO BUILD THEIR OWN SOLUTIONS

Given the fast pace of today's business climate, waiting for a centralized data group to produce reports and dashboards for business users is a thing of the past. According to a recent MIT study, companies in the top three spots in their industry that rely on data-driven decision making were, on average, [5% more productive and 6% profitable than their competitors](#). This incentive to leverage data to compete drove the self-service BI revolution where business users took reporting and data engineering into their own hands.

While business users got their data faster, the unintended consequences of this decentralized approach are obvious. Numerous data platforms, a proliferation of data marts and a large variety of BI tools is a good indicator of the dark side of DIY analytics and proof that your organization may need a semantic layer.



4. REPORTS FROM DIFFERENT BI TOOLS USE SIMILAR TERMS BUT SHOW DIFFERENT RESULTS

If multiple business units or groups are preparing their own reports and dashboards without a common semantic layer, chances are high that different tools will produce different results. Most BI tools include their own modeling layer and all support custom calculations. Whether an error in table relationships or joins, inconsistent use of the company calendar for time based calculations or just mistakes in formulas, you are almost guaranteed to have different numbers for the same data.

If you find inconsistencies in financial reporting from different spreadsheets and reports, your organization is likely suffering from a lack of a common semantic layer. See the next section for the potential consequences.

5. BUSINESS EXECUTIVES EXPRESS DOUBTS ABOUT THEIR CONFIDENCE IN THE NUMBERS

According to Forrester Research's B2B Data Activation Priority report, less than half of firms believe they execute very well in having customer data they fully trust. Once business executives lack confidence in the numbers, every decision is subject to delay. Trust in the data is a major competitive differentiating factor for the best businesses. According to Experian, six in ten companies believe that high-quality data increases efficiency in their business, with a sizable percentage believing that it not only increases customer trust (44%) and enhances customer satisfaction (43%) but also enables more informed decision making (42%) and cuts costs (41%).

If you find your business sponsors performing their own on-the-fly report reconciliation, you may be suffering a crisis in confidence. Self-service analytics without the foundation of a common semantic layer and data governance makes it difficult to build trust and prove data quality.

You don't need to sacrifice data self service to create trust, though. A universal semantic layer can power data self service while ensuring the consistency, fidelity and explainability of analytic outputs.



GETTING STARTED ON YOUR SEARCH

There are several technical approaches to implementing a semantic layer in your organization. The table below lists each approach's pros and cons.

APPROACH	DESCRIPTION	PROS	CONS	EXAMPLE VENDORS
Business Intelligence Platforms	Traditional BI platforms that bundle data modeling, query management and visualization	<ul style="list-style-type: none"> • No extra technology layer needed • Tight integration • Business user friendly 	<ul style="list-style-type: none"> • Semantic layer specific to BI tool only (not reusable) • Vendor lock in 	Tableau Power BI IBM Cognos SAP Business Objects Looker
Data Virtualization Platforms	Platforms that abstract away the physical source and location of data in a tabular format	<ul style="list-style-type: none"> • Provides flexibility in how/where data is stored • Semantic layer can be used across a variety of tools 	<ul style="list-style-type: none"> • Not friendly for business users (tables, columns) • Data models need to be built before accessing data • Query performance is not guaranteed and/or needs manual tuning 	Denodo Dremio
Data Warehouse/Data Marts	A database of information from a variety of data sources	<ul style="list-style-type: none"> • Single source of truth • Widest array of tool/query access • Easy to secure 	<ul style="list-style-type: none"> • Not friendly for business users (tables, columns) • Slow to integrate new data sources • Dependence on IT 	Snowflake Amazon Redshift Google BigQuery Azure Synapse SQL Analytics
Business Semantic Layers	A platform that presents a business data view that helps users access data autonomously using common business terms	<ul style="list-style-type: none"> • Business user friendly • Single source of truth • Provides flexibility in how/where data is stored • Semantic layer can be used across a variety of tools • Easy to secure 	<ul style="list-style-type: none"> • Extra technology layer required • Data models need to be built before accessing data 	AtScale SQL Server Analysis Services

Illustration 1: Approaches for implementing a semantic layer

As you can see above, a business-oriented semantic layer provides the best tradeoffs given its blend of data virtualization technology and the benefits of traditional BI platforms' semantic and modeling capabilities without the vendor lock in that comes with these tools.

Recommendation: A business oriented semantic layer promotes safe and secure, self service analytics consumption while driving consistency and reducing costs.

KEY CONSIDERATIONS

When choosing a vendor, there are a few core capabilities to keep in mind. Depending on your needs, you can weigh the options accordingly. The following categories are further broken down in our checklist later in this document.

Not Tied to a Single Consumption Style

From the beginning, the BI platform was synonymous with the term “semantic layer”. In recent years, however, the monolithic BI platform has given way to more component-based architectures. As analytics have become more widespread in an organization, relying on a single BI platform to be everything to everyone just isn't realistic. That means that any semantic layer tied to a specific BI tool or platform cannot be a “universal” semantic layer - it's a semantic layer for that tool. In a landscape of many tools and analytics personas, it's essential that your semantic layer be decoupled from a single consumption style. It needs to be truly “universal”.

Recommendation: When choosing a vendor, make sure that the vendor's semantic layer works across a variety of BI and AI/ML consumers - not just their own visualizations layer.

Offers Tabular and Multidimensional Views

There are two types of semantic layers, or models, to consider: a tabular semantic layer and a multidimensional semantic layer.

The tabular or relational model was popularized by modeling gurus like EF Codd and Ralph Kimbal in the 70s and 80s. These modeling techniques rely on concepts like fact and dimension tables and are meant to make a relational database or data warehouse easier to query. The multidimensional data model goes one step further. By defining relationships and aggregation rules, the multidimensional semantic model adds a business friendly context and makes hand writing SQL either unnecessary or substantially more simplistic. For the widest range of uses and consumption styles, a multidimensional semantic layer offers more power in an easier to use package.

Recommendation: Choose a semantic layer that offers both tabular and multidimensional views to cover the widest range of use cases

Supports Data Platform Virtualization

It seems like just about every five years we see a new data platform style or trend become all the rage. First, it was the mainframe. Then, the relational database, the data warehouse, the MPP database, the data lake and now back to the data warehouse (but in the cloud). If your organization has been around long enough, you probably have one of everything. As technology trends shorten and we see a wider range of options for data platforms and storage, it's essential that your semantic layer future proofs your data platform choice. Data virtualization is an excellent hedge against future platform change and minimizes or eliminates the cost of migrating to those new data platforms. A good semantic layer should offer data virtualization as its core mechanism for querying the underlying data and thereby hide the physical implementation of the data platform to prevent vendor lock in.

Recommendation: Choose a vendor that leverages data virtualization to abstract away data platform differences and minimized platform lock in.

Easy Model Development and Sharing

Raw data is just data. By adding a data model to raw data, we turn that data into consumable information. It's imperative that the platform you choose makes authoring, sharing and collaborating on data models as simply as possible. Choose a semantic layer platform that supports collaborative model development, re-use of common objects and conformed dimensions and the ability to visually model data in addition to opting for a code based approach that's compatible with your organization's software development life cycle (SDLC).

Recommendation: Choose a semantic layer with a multi-user design environment and markup language to promote re-use and enforce standardization.

Ability to Express Business Concepts and Functions

The relational data model is flexible and powerful but it's often difficult or even impossible to express high level business constructs. These constructs run the gamut from simple time-based calculations (period over period, period to date, moving averages, etc.) to more complex (semi-additive metrics, ancestor/predecessor functions). Asking a business user or data scientist to express these computations in SQL is a tall order. The MDX and DAX expression language makes these multidimensional calculations much more approachable. Make sure that your choice in semantic layer supports not just SQL but also more business friendly protocols like MDX and DAX.

Recommendation: Choose a semantic layer that supports business constructs and core analytics requirements around time intelligence and hierarchical rollups.

Query Performance & Caching

When evaluating vendors, this is arguably the area where you should spend most of your time. Without consistent and performant query serving, a semantic layer has little value and end users will avoid using it, which defeats its intended purpose. In analytical use cases, business users are accustomed to interactive query performance since they typically query proprietary analytical

databases or cubes that are designed for fast queries. As a result, a semantic layer needs to deliver even better performance than the native platforms they interact with since the query performance needs to match or beat the existing solutions they are replacing. To make matters worse, many of today's queries often include heterogeneous database joins that further tax query performance.

Semantic layers that simply cache query results or create cached tables are not sufficient for analytical use cases. A proper semantic layer should optimize query performance autonomously, without manual intervention.

Recommendation: Choose a semantic layer vendor that includes a comprehensive performance management system that goes beyond simple caching techniques.

Support for Business Intelligence and Data Science Workloads

A business view of data has been essential to promote self service analytics for business intelligence. However, the need for clean and usable data doesn't end with just the business analyst. Data scientists spend about 45% of their time on data preparation tasks, including loading and cleaning data, according to a [survey of data scientists conducted by Anaconda](#). It is crucial that a semantic layer works for multiple user personas, including the data scientist. With a common data language and business terms, business analysts and data scientists alike are more likely to work off the same assumptions and produce historical results and future predictions that make sense.

Recommendation: Choose a semantic layer that supports a variety of workloads including business intelligence and data science.

Security & Governance

Since the semantic layer serves as middleware for analytical queries, it's imperative that the platform integrates with the enterprise's security infrastructure. There are two main forms of security to consider: authentication & authorization.



First, a semantic layer must integrate with the enterprise's single sign on infrastructure in order to authenticate users, whether that be Active Directory (AD), LDAP, OAuth or other third party authentication platforms. The authorization capabilities must flow through the client applications and the data virtualization platform must synchronize users automatically.

Second, the semantic layer must include the ability to hide or mask sensitive columns, limit data rows based on user access rules and impersonate users when querying the underlying data sources. Impersonation is especially crucial since using a proxy user (instead of the query user) to query underlying data sources may circumvent security policies for those data platforms and force users to duplicate security policies in the virtualization layer.

Recommendation: Choose a semantic layer that integrates with your single sign on standards and supports column level security, row level security and impersonation.

FEATURE CHECKLISTS

The following checklist is a tool for evaluating different vendors along the capability categories described above. Use a number between 1 and 5 (5 being best) to score the vendor's capabilities for each feature. You may also use the weighting column to personalize the scoring results based on your most important priorities.

FEATURE CATEGORY	FEATURE	SCORE (1-5, 5 = BEST)	WEIGH (1-5, 5 = BEST)	WEIGHTED SCORE (CALC)
Use Cases	Supports analytical workloads			
	Supports data science workloads			
Connectivity (northbound & southbound)	Supports legacy, on-premise data warehouses			
	Supports cloud data warehouses			
	Supports on-premise and cloud data lakes			
	Supports SaaS data sources (Salesforce, Workday)			
	Supports tools that speak SQL via JDBC or ODBC			
	Supports tools that speak MDX or DAX and live Excel connections			
	Supports custom applications via REST or Python interfaces			
	Supports zero client install for data consumers			
Development Environment	Supports web based development (versus client application)			
	Supports multiple, simultaneous editors for virtual view development			
	Supports reusable objects and model component sharing			
	Supports development lifecycle (dev/test/prod)			
Calculations and Analytical Functions (OLAP)	Supports Time Intelligence (period over period, period to date)			
	Supports MDX, DAX, pre and post query calculations			
	Supports aggregation functions (SUM, AVG, MAX, MIN)			
	Supports non-additive metrics (Distinct Count, First, Last)			
	Supports live Excel pivot tables and Excel CUBE functions			
Query Performance & Caching	Supports automated query performance management			
	Supports dialect specific optimizations			
Security & Governance	Supports single sign on for all data consumers			
	Supports user impersonation and delegated authorization			
	Support and respects native data platform security constructs			
	Supports row level security for users and groups			
	Supports column hiding and masking for users and groups			
TOTAL				

Illustration 2: Semantic Layer feature checklist



CONCLUSION

To summarize, here are the key recommendations to keep in mind as you choose your vendor:

1. Choose a “universal” semantic layer that is compatible with a large number of northbound interfaces (i.e. BI tools) and southbound interfaces (i.e. data warehouses and data lakes).
2. Choose a semantic layer that specializes in business style analytics.
3. Choose a semantic layer that supports a variety of analytics consumption styles and tools. Avoid “closed garden” platforms that seek to tie their semantic layer to their own visualization or analytics tooling .
4. Choose a semantic layer that will scale with your data growth to meet the demanding requirements of the business analyst and data scientist..
5. Choose a semantic layer that supports data virtualization to simplify data access and prevent vendor lock in.
6. Choose a semantic layer with a rich, customizable development environment that includes a rich markup language..
7. Choose a semantic layer that supports the expression of complex business constructs and definitions including conformed dimensions and hierarchical relationships..
8. Choose a semantic layer with automated performance management that delivers OLAP style query performance.
9. Choose a semantic layer that supports multiple use cases and user personas including the business analyst and data scientists.
10. Choose a semantic layer that integrates into your single sign on infrastructure and supports logical and physical data governance and security.

As you can see, there’s a lot to consider when choosing a semantic layer, but the investment is well worth it. With a universal semantic layer, you can have your cake and eat it too. You can continue to support the analytics self service trend, but do so with consistency, governance and control. A universal semantic layer is key to making your organization agile and data driven in an era where data makes the difference.



RESOURCES & FURTHER READING

Gartner Research: [“How to Use Semantics to Drive the Business Value of Your Data”](#) (27 November, 2018)

AtScale Blog: [“What is a Universal Semantic Layer? Why Would You Want One?”](#)

Gartner: [Make Financial Data Decision-Ready](#) (24 January 2020)

ABOUT ATSCALE

AtScale powers the analysis used by the Global 2000 to make million dollar business decisions. The company's Intelligent Data Virtualization™ platform provides Cloud OLAP, Autonomous Data Engineering™ and a Universal Semantic Layer™ for fast, accurate data-driven business intelligence and machine learning analysis at scale. For more information, visit www.atscale.com.

ATSCALE