THE PRACTICAL GUIDE

How to use a semantic layer for data & analytics

ATSCALE

TABLE OF CONTENTS

Introduction: What is a Semantic Layer?

You may have heard the term semantic layer before; it's been around for some time. People invented semantic layers 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 UniverseTM. From that point on, 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 closely tied to a business intelligence (BI) platform. As long as enterprises remained within the confines of their BI vendor of choice, everything worked well. Today, there are more ways than ever to analyze data. Long gone are the days where there was one BI platform to rule all. Tightly coupling a semantic layer to one analytics consumption style no longer makes sense.

To expand on that, the explosion of self-service BI has freed business users from relying on IT-prepared analytics, but at the expense of data consistency and trust in analytics' output. Business definitions and terms have become mutable, malleable, and subject to interpretation. While it's great that business users now have self-service BI tools, they also need to be working off of consistent, high-quality data. The cost of bad data is enormous; According to IBM, poor data quality costs the U.S. economy a staggering $3.1 trillion annually.

Luckily, a semantic layer that's decoupled from the point of consumption can help ease these problems with data quality and empower self-service analytics. A well-designed semantic layer can lead to better data-driven decisions. It's a critical part of the modern analytics stack.

Making all of this work involves a series of building blocks.

- **Key trends driving a need for a semantic layer**
- **Real use-cases for a semantic layer across industries**
- **Best practices and key considerations for choosing a semantic layer for your business.**

Let's get started!

Trends Driving the Need for a Semantic Layer

Cloud data lakes and cloud data warehouses like Snowflake, BigQuery, Redshift, Databricks and more have become well-accepted data platform architectures. According to the AtScale 2020 Big Data & Analytics Maturity Survey, 61% of respondents currently operate cloud data platforms, and 48% plan on deploying them soon. In the meantime, Hadoop didn't become the be-all end-all data solution but just one solution for managing data.

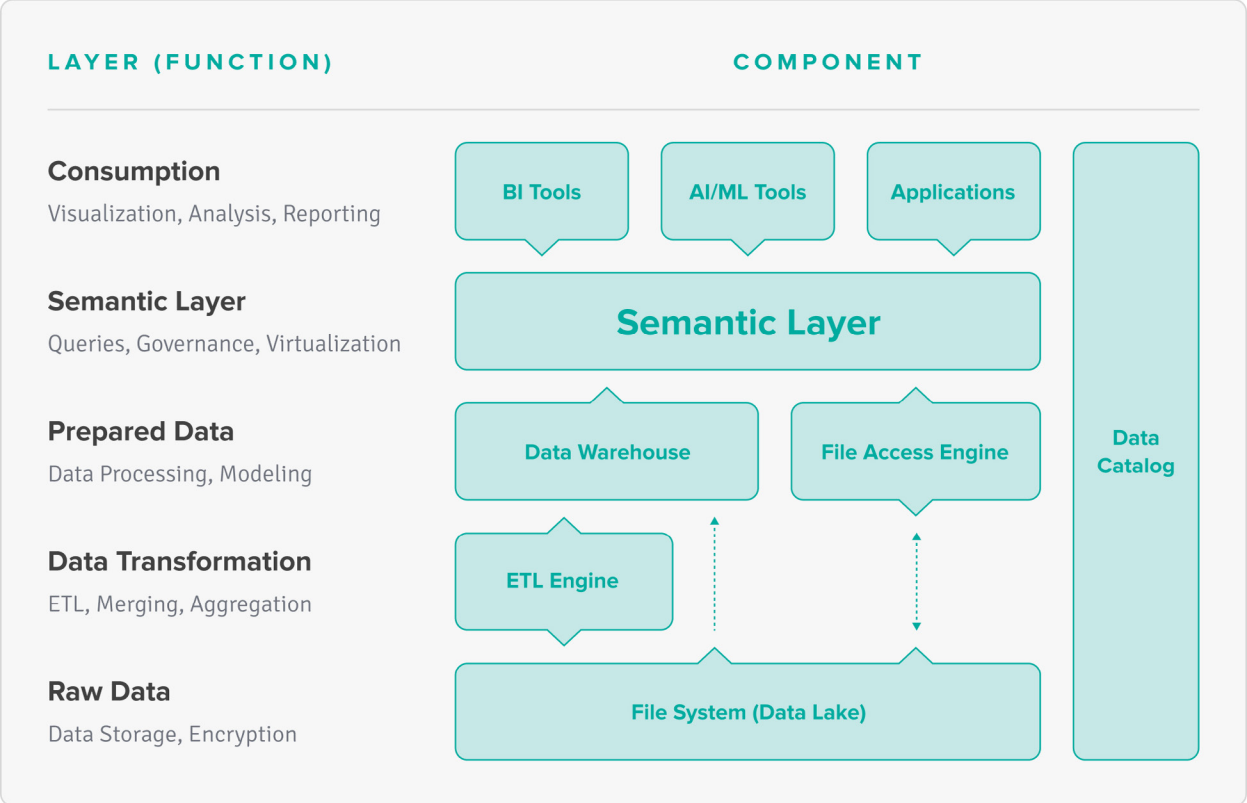
As the volume of data in the cloud grows, data architects are increasingly becoming more comfortable with data living in different locations and in different platform architectures. However, this gives rise to a new challenge for IT: managing data access and quality across multiple silos. A semantic layer becomes a critical piece in a cloud data platform strategy (or a blended cloud and on-prem strategy).

Both data scientists and BI users need access to clean, understandable data. Today's self-service architectures often force analytics consumers to become data wranglers and data engineers. In fact, the average data scientist spends over 45% of their time preparing data rather than modeling it.

Asking business users and data scientists to program their own metrics and business terms is both a massive waste of time and a recipe for chaos and inconsistency. A semantic layer solves this problem by defining business metrics, data access, and transformations in one place. That way, analytics consumers are almost guaranteed to speak the same language, regardless of their use case or toolsets.

Finally, a semantic layer can serve as a central governance gateway across the enterprise, which is crucial as the number of silos and data access points explodes. A semantic layer serves as a single point of access so IT can secure data and control access across the organization. The same Big Data & Analytics Maturity Survey referenced above shows that nearly 80% of enterprises rank security and governance as critical to their success in the cloud.

Top Data and Analytics Leaders Share Their Data Literacy Secrets



As you can see by the diagram above, the semantic layer sits between the point of analytics consumption and the data warehouse and data lake. A semantic layer hides the physical complexity from end users and provides them with understandable business terms and user-friendly data, instead of raw SQL and database schemas. This level of data virtualization makes data access possible for any analytics consumer.