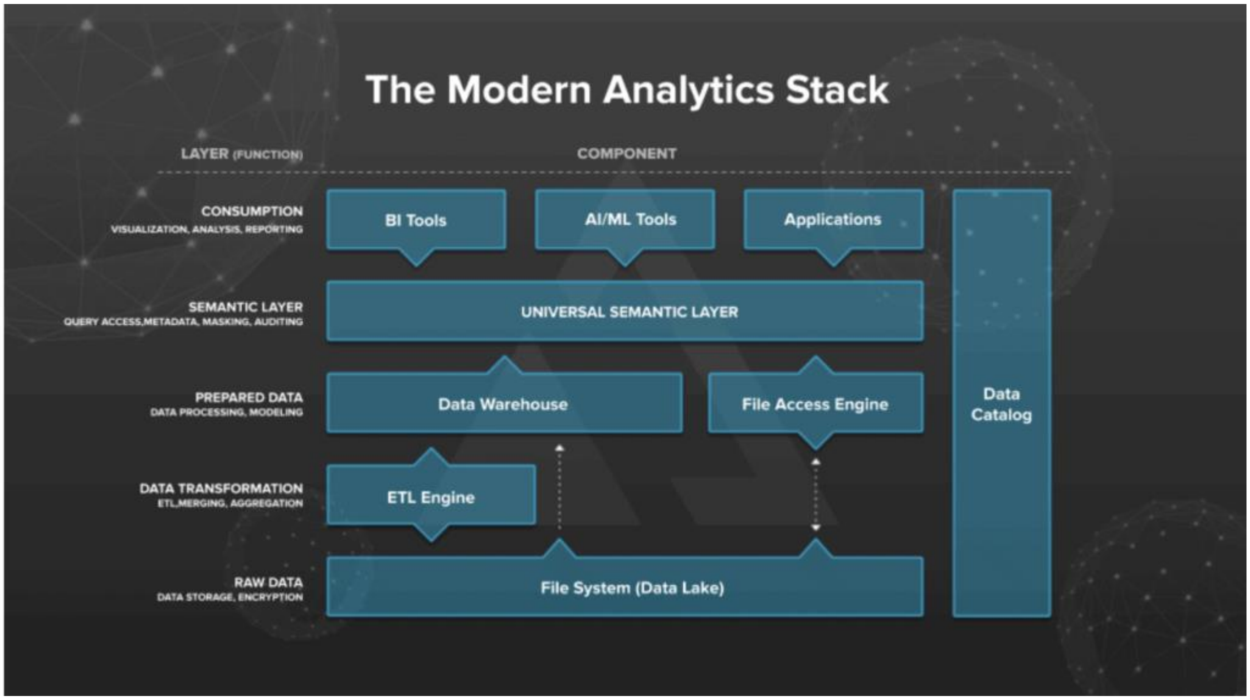
The Universal Semantic Layer (USL)

Definition

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.



What is a Common Data Set/Model?

A collection of predefined schemas including entities, calculations, attributes, and relationships. The schemas represent commonly used concepts and activities, such as Ledger Account or Marketing Campaign to simplify the creation, aggregation, and analysis of data. There will be sub cubes from the larger cubes that may meet a particular business perspective.

What process defines and feeds our Common Data Model?

Multidimensional OLAP (MOLAP) is the traditional classic form. Data Engineers pre-select a subset of dimensions and facts, the system builds a data cube, then users can query the cube. This approach can provide quick query responses once the cube is ready. **Cloud OLAP** (**COLAP**) uses a cloud service to do this.

Building the Business Logic into the Data Warehouse Common Data Model with SSAS

Pros

With this approach we define our common data set/model and then have complete control over that data with one place to secure and govern the data. The single source of truth is realized as opposed to business users having to go to multiple places.

With SSAS we can create a single point for data access. The Semantic Layer can then also serve as a central governance gateway across the enterprise. IT can then secure the data and control its access once and for all. 79% of enterprises rank cloud security and governance critical to their success in the cloud.

Deployment Flexibility. We are cloud ready once we have a semantic layer that is deployable to public clouds like Amazon AWS, Microsoft Azure, Google GCP, even a virtual datacenter that TSTC manages. Our semantic layer is not limited to a vendor.

Cons

It creates a bottleneck where your users are waiting for data. Granted we have already defined a lot of these data points but the creation of new ones now requires SMEs and analysts and Data Engineers all meeting to define the business rules so that the Data Engineer can encode them into the data warehouse and the semantic model. Then the SME has to user test the solution and verify that we all got it right.

Why not let the BI tool be that semantic layer?

Having a semantic layer for Tableau and having a semantic layer for PowerBI, and having a semantic layer for Workday; that does you no good whatsoever, because now you have three semantic layers and of course there's going to be differences. It's really difficult to enforce conformance. I know that all of the BI tools say, "Hey, we have a way of providing governance". The fact is that if you can create a calculation in a BI tool and in a report, then you don't have governance. Using a BI tool's semantic layer definitely doesn't guarantee consistency, because, again, people can use the same words and it can have different meanings. The bad thing here, is that it really forces the business analyst to be a data engineer. You hear terms like "data wrangling", you can see that Tableau started out as a visualization tool and now they have a built in ETL data prep tool. And to me, that's a bad thing. Do you really want your business users who are supposed to be running the business and figuring out how to improve profitability and improve customer experience actually becoming data engineers to do their jobs? Tableau even promised that the Data Management Service would do this. That makes no sense to me.

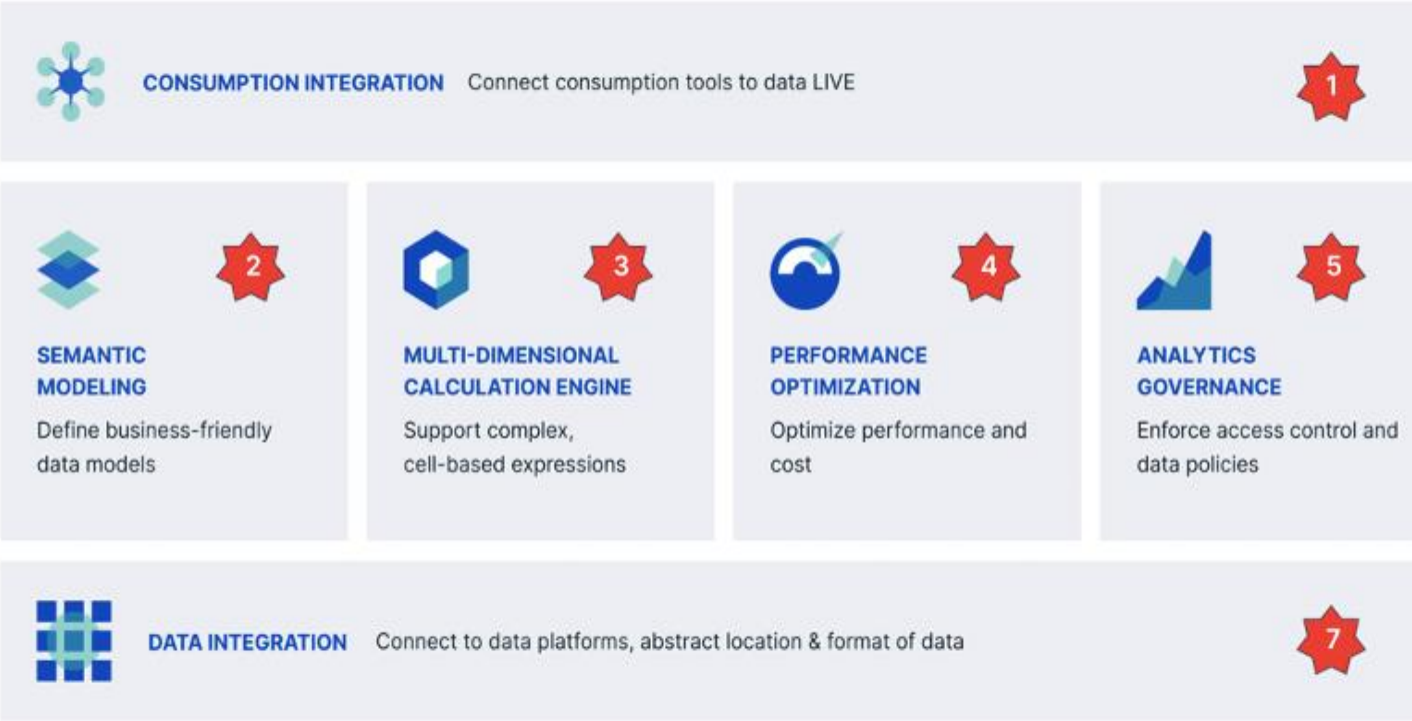
When I arrived at TSTC I was overwhelmed by the data needs initially so I thought I would just provide views and tables and let the analysts mine the data themselves. But this caused our analysts to have to become data engineers and spend all their time in data discovery and in data delivery. What happens when you have conflicting answers to the same business question? Maybe it causes a delay in critical business initiatives, maybe it puts the entire company's financial state at risk. Whatever the consequences may be, it's best to avoid them by keeping business definitions managed in a single source of truth.

What Does a Semantic Layer Do?

A semantic layer makes data usable for everyone and presents a consistent, business-friendly interface to corporate data so that you can evaluate a vendor. A semantic layer also:

1. Connects users to live data, of any shape and size, wherever it landed
2. Delivers queries at the "speed of thought" on any size of data
3. Governs user access to sensitive data for *every* query, regardless of the tool used
4. Connects and blends data across silos from on-premise to cloud to SaaS applications
5. Bridges the business and data science teams by integrating historical and predictive data

A semantic layer platform needs to deliver on seven main vectors of value. The following diagram illustrates the core capabilities of a semantic layer:


CONSUMPTION INTEGRATION
Connect consumption tools to data LIVE
1
2
SEMANTIC MODELING
Define business-friendly data models
3
MULTI-DIMENSIONAL CALCULATION ENGINE
Support complex, cell-based expressions
4
PERFORMANCE OPTIMIZATION
Optimize performance and cost
5
ANALYTICS GOVERNANCE
Enforce access control and data policies
DATA INTEGRATION
Connect to data platforms, abstract location & format of data
7

1. Consumption Integration

A semantic layer needs to be truly universal. This means a semantic layer must support a variety of use cases and personas including business analysts, data scientists, and application developers. It also needs to support a wide range of query tools using their native protocols including SQL, MDX, DAX, Python REST, JDBC, and ODBC.

2. Semantic Modeling

The core of the semantic layer is the data model. A semantic layer maps the logical elements (dimensions, metrics, hierarchies, KPIs) to the physical entities of databases, tables, and relationships. In order to deliver a digital twin of the business, a semantic layer must support reusable models and components to drive a hub and spoke (data mesh) analytics management style backed by a CI/CD compatible markup language and GUI-based modeling environment.

3. Multi-Dimensional Calculation Engine

The semantic layer data model must be backed by a scalable, multi-dimensional engine to express a wide range of business concepts in a variety of contexts. The semantic layer engine must support matrix-style calculations (time intelligence, multi-pass, etc.) using a multidimensional expression language like MDX or DAX and query underlying cloud data platforms "live" without data movement or a separate data store.

4. Performance Optimization

Without query acceleration, a semantic layer will likely be bypassed using BI tool extracts and imports, which defeats the purpose of a semantic layer. As such, a semantic layer must automatically tune and improve performance using machine learning and user query patterns without moving data outside the native cloud data platform or requiring a separate cluster for managing aggregates.

5. Analytics Governance

A semantic layer needs to satisfy a wide range of data governance scenarios. It must integrate with corporate directory services (i.e., AD, LDAP, Okta) for user identity management, apply row-level security to every query and be able to hide and mask data columns based on user, group, and role-based (RBAC) access data rules.

6. Data Integration

Data lives in multiple silos, including on-premise, legacy data warehouses, data lakes, cloud data warehouses, and SaaS applications. A semantic layer must be capable of accessing and modeling data across these multiple sources and support a variety of data types including nested data like JSON.

A universal semantic layer is quickly becoming a critical component in a modern data and analytics stack. However, when evaluating semantic layer options, it's important to keep one thing in mind: If any of the above requirements is missing, a semantic layer is unusable. In other words, it's binary – it either works 100% or it doesn't work at all. Don't let this be an impediment, though, because a universal semantic layer makes everyone a data-driven decision-maker.