Accelerating Self-Service Analytics with a Unified Semantic Layer

Enabled by Data Virtualization, Unified Semantic Layer Promise Greater Efficiency through Common Definitions.

SOLUTION
Data Virtualization for Unified Semantic Layer.

WEBSITE
www.denodo.com

PRODUCT OVERVIEW
The Denodo Platform offers the broadest access to structured and unstructured data residing in enterprise, big data, and cloud sources, in both batch and real-time, exceeding the performance needs of data-intensive organizations for both analytical and operational use cases, delivered in a much shorter timeframe than traditional data integration tools.

Self-service analytics tools hold a lot of promise for business stakeholders. However, such tools are better suited to savvy power users who are comfortable chasing down disparate data sets and combining them using powerful, ad hoc techniques.

To enable all business users, regardless of technical acumen, to derive value from self-service analytics tools, companies need to be able to provide access to the data through a unified semantic layer, which ensures that no matter the data source, users can access the data using common definitions that are standard, within each particular business culture. A unified semantic layer also acts as a protective layer, preventing users from accessing the incorrect sources.

Fragmented Data

However, companies are unable to establish unified semantic layer because of the age-old challenge: Data is spread across multiple heterogeneous databases, data warehouses, cloud and big data systems, No SQL sources, and flat files, so models and definitions will vary.

This, in turn, results in two related challenges:

1. **Poor Data Integrity.** . When business analysts go directly to the data sources, they may not go to the authoritative sources, resulting in data that is of questionable quality.

2. **Untraceable Data Lineage.** Also, if users collect data from sources directly, they may not keep an accurate record of where the data came from, hindering the ability to determine data quality, eroding trust in the data.

Due to these challenges, Gartner's Rob van der Meulen predicts that by 2018, "most business users will have access to self-service tools, but only one in 10 initiatives will be sufficiently well-governed to avoid data inconsistencies that negatively impact the business."

Data virtualization is a flexible, modern data integration technology that enables companies to seamlessly unify their data sources to enable unified semantic layer for enhancing self-service analytics initiatives. In this brief, we illustrate how data virtualization enables companies to implement unified semantic layer, and we close with two case studies of companies that have leveraged data virtualization to establish unified semantic layer that enhance self-service analytics initiatives for significant results.

What is Universal Semantic Layer (USL)?

A USL provides core business elements and semantics that require analyses and can be shared across multiple consuming applications such as BI tools and operational analytics. It supports virtually any data source, both traditional and emerging, and any consuming application. Enterprises that strive to deliver self-service analytics often need to support multiple user types and BI/Visualization tools. For example, a credit risk department in a bank might use Tableau to consume and visualize the data while the marketing department might use Excel for analyses. This approach necessitates embedding business logic in different BI tools, potentially resulting in inconsistent data being provisioned across these tools. This creates data distrust and reconciliation needs. A Universal Semantic Layer eliminates such issues by provisioning consistent and trusted data, reducing risks, and enabling centralized security and data access control across multiple BI tools.

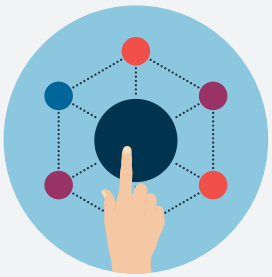
What is Data Virtualization?

Data virtualization is a data integration technology. But whereas most data integration solutions move a copy of the data from disparate sources into a new, consolidated source, data virtualization offers a completely different approach.

Rather than *moving* the data, data virtualization provides a view of the integrated data, leaving the source data exactly where it is. This means that companies do not have to pay the costs of moving and housing the data, and yet they still gain all of the benefits of data integration. Also, virtualized views abstract users from the complexities of access, such as where the data is stored or what type of system it is stored on.

Because data virtualization is deployed as an architectural layer above the disparate sources, serving as an organization's single enterprise-wide data access layer, it enables further benefits: Security, privacy, and governance protocols can be administered across the disparate sources from a single point of control. For example, across the board, organizations can mask social security numbers from all users or from select users, even chosen on a per-user basis, if desired. Finally, and most importantly for the purposes of this brief, this capability enables companies to use data virtualization layers to establish unified semantic layer for accessing the entire enterprise data set across the myriad data sources.

This is how data virtualization overcomes the challenges mentioned at the start of this brief:



1. **Fragmented Data is Seamlessly Unified.** With a data virtualization layer in place, all of the data, across myriad systems in its various formats, appears to users as though it sits in a single, easily accessible repository.

2. **Data Integrity is Preserved.** Because all of the data sources are accessed through the data virtualization layer, companies can use the data virtualization layer to establish strong governance protocols and specify authoritative sources.

3. **Data Lineage is Fully Traceable.** Also, because all data flows though the data virtualization layer, data lineage is fully traceable from users to sources.

Case Studies

In this section, we present the case studies of two companies that have successfully leveraged data virtualization to implement universal semantic layer in the service of true self-service analytics.

Large Diversified Global Insurer (The Firm)

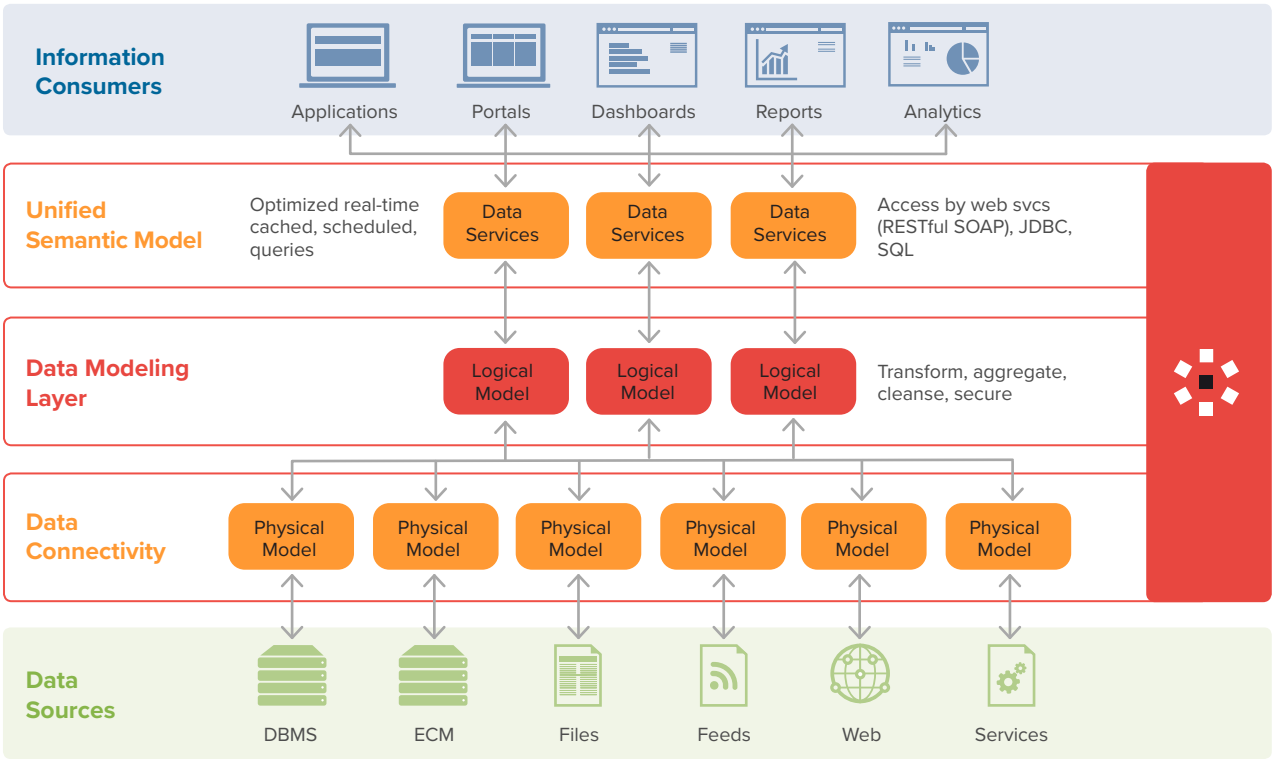
The Firm, headquartered in Boston, Massachusetts, has been in business since 1912. A diversified insurer, The Firm offers a range of insurance products including personal and commercial automobile, homeowners, accident and health, general liability, property, surety, workers compensation, group disability, group life, specialty lines, reinsurance, individual life, and annuity products. Today, the company has operations in 30 countries and economies around the world.

Challenges:

- The Firm lacked common, consistent views of key business metrics.
- Analysts were getting different answers from different tools or reports.
- The company was dedicating too much time discussing the veracity of the data and not enough time addressing business issues.
- Management tasked IT with providing a consistent view of data used to drive the business, regardless of which channel was used to access the data.

Solution

The Firm leveraged the Denodo Platform, which uses data virtualization to unify the company's disparate data sources and establish a unified semantic layer across the enterprise.



Results

With the unified semantic layer enabled by the Denodo Platform, the Firm was able to:

- Gain a unified view into the core business metrics.
- Ensure that all tools drew on common definitions and common data sources.
- Focus squarely on business issues, with trust in the integrity of all data.
- Access a consistent, cross-channel view of data across the enterprise.

CIT Group

Founded in 1908, CIT is a financial holding company with more than $65 billion in assets. Its principal bank subsidiary, CIT Bank, N.A., has more than $30 billion of deposits and more than $40 billion of assets. CIT became a Systemically Important Financial Institution (SIFI) or “too big to fail” bank after it acquired a large retail bank.

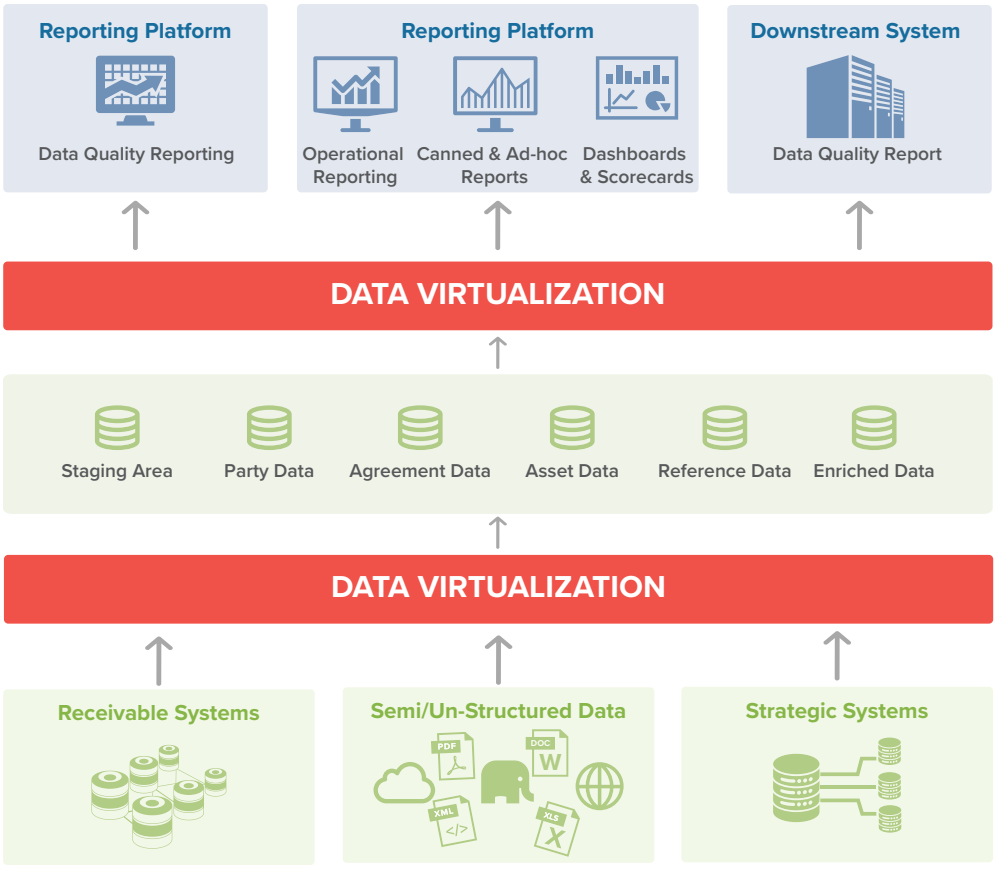
CIT needed a controlled data environment to support the intense regulatory scrutiny. In the legacy architecture, consumers were pulling data directly from source systems. As a result, information that was modified in one system was not always tied across to the other systems, and the company lacked a unified view into risk. To avoid this problem, CIT needed a common data access layer to link across the various silos. The bank also needed smart data governance processes in place, to ensure that stewards were accountable for their data and can efficiently manage its quality.

Solution

CIT Group implemented a data services layer (DSL), which is comprised of two sub-layers:

1. **A data provisioning layer**, which acts as a common provisioning point for all consumers, integrating the data from the company’s disparate primary data sources.
2. **A data access layer,** which presents the data to consuming applications through a unified interface.

The Denodo Platform is core to both layers, leveraging its data virtualization capabilities to manage and orchestrate the flow of information throughout the controlled data environment. The Denodo Platform enables the DSL to expose data from upstream sources using a universal semantic model, streamlining processes with greater standardization.



Results

- By exposing the disparate data sources using a universal semantic layer, the data services layer greatly enhanced the efficiency of multiple processes across the organization.
- The Denodo Platform enabled faster time-to-market and incremental information delivery.
- The data services layer became the common provisioning point for all of CIT’s data instead of the legacy point-to-point integration.
- Data virtualization reduces data replication and unnecessary copies of data.
- The Denodo Platform enables smart data governance through enforcement of policies, standards, and procedures, and capabilities such as efficient data lineage, metadata management, and monitoring of data quality before consumption.