DATASHEET

DATA VIRTUALITY LOGICAL DATA WAREHOUSE

The High-Performance Data Virtualization Solution

The Data Virtuality Logical Data Warehouse (LDW) is a step up in the evolution of data warehousing, combining the best of ELT/ETL and data virtualization in a single, comprehensive data management solution. This combination allows to retrieve data using a single query language, get speedy query response, and to quickly assemble different data models or views of the data to meet specific needs. Physical data integration is a robust feature of the Logical Data Warehouse that ensures fast query response while decoupling performance from the source data stores and moving it to the logical data warehouse repository. In this manner, the effort-intensive physical transfer of the data is minimized and simplified, effectively removing lengthy data movement delays from the critical path of data integration projects. The final result is easy data access without fundamentally changing the existing environment.

There are three essential elements to the Data Virtuality Logical Data Warehouse: Data virtualization, Caching, and Materialization (automated ETL).

Data virtualization offers a lot of flexibility. It quickly provides initial results and supports rapid prototyping and agile development. Also, real-time data can be queried from various data sources in different data formats without copying and physically moving the data beforehand. However, data virtualization on its own does not scale well for large amounts of data or large number of users.

To compensate for that, many data virtualization solutions, incl. Data Virtuality use **caching** to increase the performance of the queries. But, caching only solves performance challenges for smaller datasets. For larger datasets, caching is not adequate as it leaves with very little control and flexibility to how the data is loaded and stored. Furthermore, caching falls short when it comes to batch data import, data historization, complex multi-step data transformations, and dealing with large amounts of data.

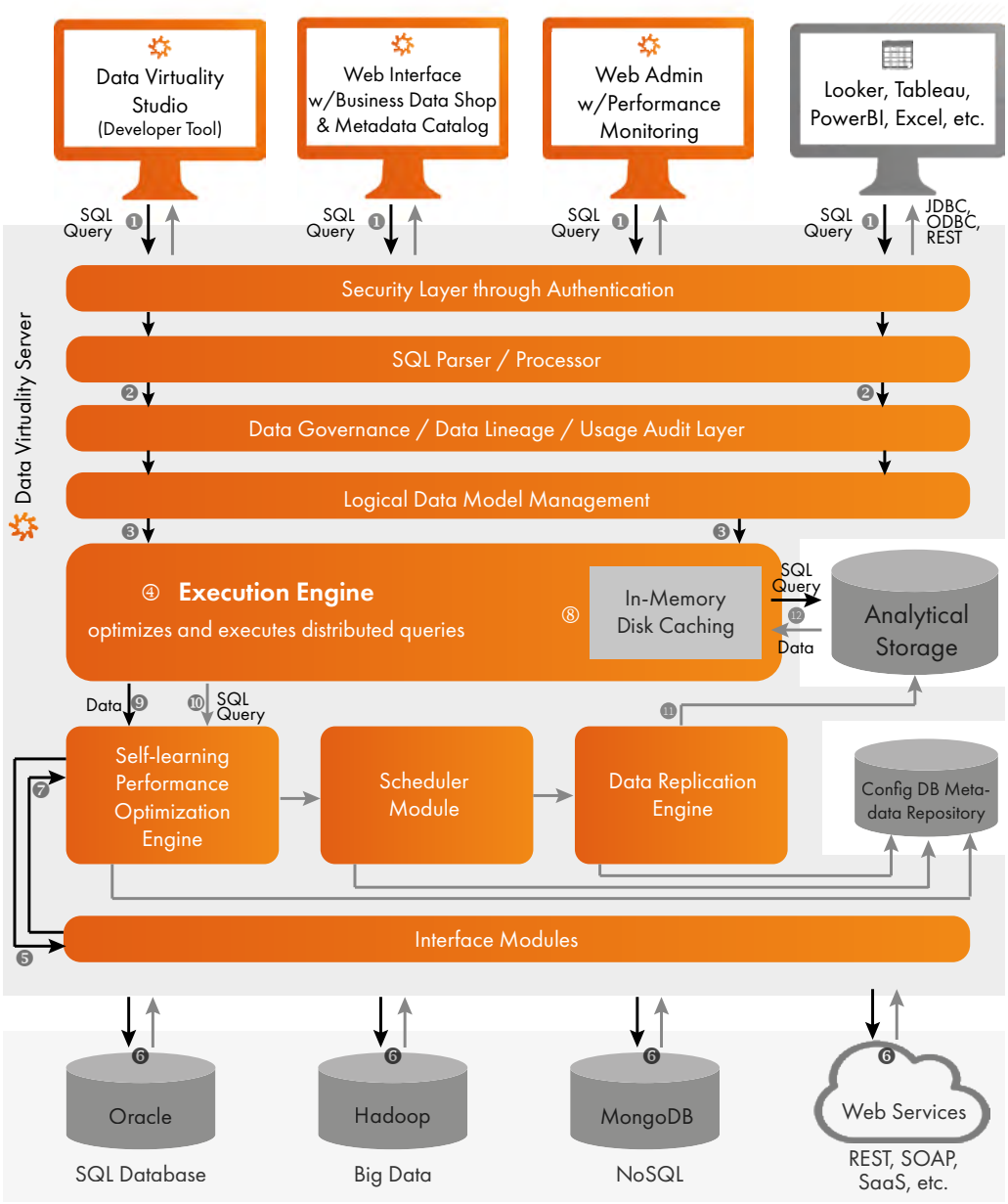
Materialization, enabled by the automated ETL capabilities of Data Virtuality, scales beautifully and provides abilities for semantic business-friendly data element naming and modeling. High performance is ensured and data historization can be facilitated. But, on its own, it lacks agility and real-time data access.

Conclusion: The **combination of these three technologies** gives the **flexibility and high performance** needed to serve various use cases of financial services institutions.

Further essential functionalities of the Logical Data Warehouse that empower your use cases are:

- **200+ integrated connectors** which give immediate access to any data source or system, even in real-time is a powerful enabler for many modern digital use cases. Data Virtuality provides full maintenance service for all connectors so you can solely focus on your main work.
- Even complex data transformations can be done with **procedural SQL**

How Data Virtuality Works



❶
By just using SQL, the data from all connected data sources can be queried in a business intelligence (BI) tool like Tableau, Cognos, and Looker.

❷
Data Virtuality Server parses the query using a built-in SQL processor and is able to identify which logical data model parts are involved (regardless if query is directed to the logical layer or directly to the data sources).

❸❹
Query plan is built and optimized by effectively distributing the executions across the connected data sources. In the actual execution, the query is split across the interface modules that are responsible for getting partial data from the appropriate data sources and transforming it to the relational format.

❺❻❼❽
The partial results coming from the Interface Modules are then further processed (joined, aggregated, projected, etc.) by the Execution Engine and returned to the requesting BI tool. The Execution Engine can optionally cache data in memory or on disk as needed to improve performance.

❾❿⓫
In addition, the Self-learning Performance Optimization Engine analyzes data and the data sources' statistics as well as the relevance of the query and recognizes performance bottlenecks. Furthermore, it eliminates bottlenecks by automatically creating and managing the physical data structures in the analytic storage (an external source like PostgreSQL, Oracle, SAP Hana, etc.) by utilizing the Data Replication Component. Data can be kept up-to-date by scheduling the replicating on a regular basis, and the amount of replication data can be reduced by running incremental replications, and capturing change data from databases, for example.

⓬
Once data is physically available in the Analytic Database, all of the slow BI query parts are automatically redirected to that database without rewriting the reports.

Logical Data Warehouse Features

DATA VIRTUALITY SERVER

- Windows Server 64bit
- Linux 64bit (Redhat, CentOS, Ubuntu, and others)

DATA VIRTUALITY STUDIO

- Windows 64bit
- Linux 64bit
- Mac 64bit
- SQL Editor code completion on column level
- Metadata dependency viewer (Data Lineage)
- Metadata catalog and search
- Graphical view builder
- Wizards for easily connecting generic data (files/(S)FTP/S3/Webservices) using formats XML, JSON, CSV, xSV.
- Wizard for querying Google Analytics APIs

BUSINESS DATA SHOP (self-service web interface)

- Metadata catalog and search
- Self-service data access for business users
- Write and run queries
- Download data

DATA FEDERATION

- Cross-database joins
- Nested loop
- Merge join
- Dependent semi-join
- Cross-database unions
- Cross-database SELECT INTO, INSERT INTO
- Dynamic cost-based query optimization

DATA GOVERNANCE

- Automatic data lineage
- Column-level data lineage
- Column masking

SQL DIALECT

- ANSI-92 with extensions
- DDL, DML, procedural SQL
- Nested subqueries
- Common Table Expressions (CTEs)
- Window functions/Framing clauses
- XML/JSON parsing
- Web service access
- Scripting languages (server-side javascript)
- Native query syntax

ACCESS DATA

- Via JDBC
- Via ODBC
  - Windows (32bit/64bit)
  - Linux (unixODBC 32bit/64bit)
  - Mac (unixODBC 32bit/64bit)
- Via REST API (REST-JSON)

DATABASES AND CONNECTORS

- More than 200 ready-to-use connectors. All our connectors can be found here

MOVE AND EXPORT DATA

- One query language: SQL
- Permission-based INSERT, UPDATE, DELETE statements on all relational databases, Salesforce, SAS
- Push-export via FTP, SFTP, SCP, email, S3, Azure Blob storage, web services (REST, SOAP, plain HTTP), file system and others
- Export data using Data Virtuality Studio/SQL

SECURITY, AUTHENTICATION, AUDIT

- Row-based security
- Git integration
- Built-in user/role based permission system
- Permission granularity on schema, table, column level
- LDAP authentication (Active Directory, ForgeRock, etc.)
- History of changes (versioning) for all custom metadata
- Access to audit information and usage statistics using SQL from external tools
- Security protocols: SSL/TLS, HTTPS

STRUCTURE OPTIMIZATION

- Materialized source tables and (virtual) views
- Precalculated joins
- Precalculated aggregations
- Automatic index creation

MATERIALIZATION ALGORITHMS

- Full copy (used with materialized tables, views, joins, aggregations)
- Incremental replication based on timestamp/id fields (used with materialized tables and views)

JOB TYPES

- Full copy with different cleanup options
- Batch update (optionally with overlap cleanup)
- History update (slowly changing dimension type 2)
- Upsert with optional surrogate keys
- Custom SQL jobs
- External programs and scripts

SCHEDULE TYPES

- Once with optional delay
- On time interval (every X minutes, hours etc.)
- Daily at certain times of day
- Weekly on certain weekdays
- Monthly
- Using custom cron expressions
- Depending on other jobs or schedules (on success/failure/always)

IN-MEMORY CACHING FOR EVEN FASTER RESPONSES

- Session scope
- User scope
- Virtual database scope

WHAT ELSE?

- Change Data Capture (CDC) for selected data sources
- Mail notification on job and replication status
- Multi-tenancy
- Graphical web-based performance monitoring
- Password encryption
- Smart data movement approaches (Snowflake & Redshift S3 load, Azure DWH Blob storage load, Salesforce Bulk API)
- Programmatic access to all server functionality using Data Virtuality Management API

About Data Virtuality

Data Virtuality provides data integration solutions that help companies to easily connect and manage their data from multiple data sources such as APIs, databases and flat files. The revolutionary single source of data truth platform combines data virtualization and automated ETL. In this way not only is data management simplified but data integration efforts are significantly reduced - by up to 80%.