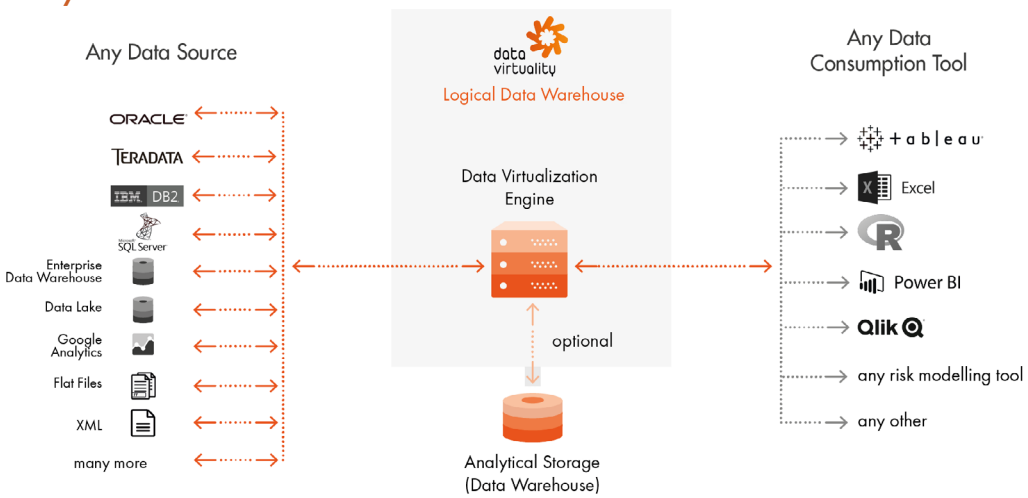
DATA VIRTUALITY

Logical Data Warehouse

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.

How Data Virtuality Works



1. CONNECT TO YOUR DATA SOURCES
   Logical Data Warehouse (LDW) connects to multiple data sources and allows querying data from there by using SQL. Data sources can either be relational or non-relational.

2. CREATE A CENTRAL DATA LOGIC
   LDW also allows you to create a central data logic that covers the business logic as well as the logical connections between the different systems. This layer can easily be implemented by using SQL views and/or stored procedures.

3. GET YOUR DATA IN A DATA WAREHOUSE
   LDW copies/replicates data for you using SQL commands. It transforms these SQL commands into the respective native data source commands, such as SQL commands, API calls, access to various filesystems, or access to CDC (Change Data Capture) interfaces of different systems and databases. Data Virtuality provides predefined jobs for the most common transformations. LDW gathers statistics about the usage of the data and uses these statistics to create a data warehouse in a special database, on a database server of your choice. This is done through automatic recommendations for the internal structure of the data warehouse that are then proposed to the database administrator. The data is then updated in your data warehouse with different time schedules.

4. MAKE YOUR DATA ACCESSIBLE
   Finally, the LDW supports the standard interfaces (JDBC, ODBC, REST) to deliver data to the data consumers. This could be, for example, reporting tools, advanced analytics tools, or custom programs in various programming languages.

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 (Sandboxing)
- 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%.