

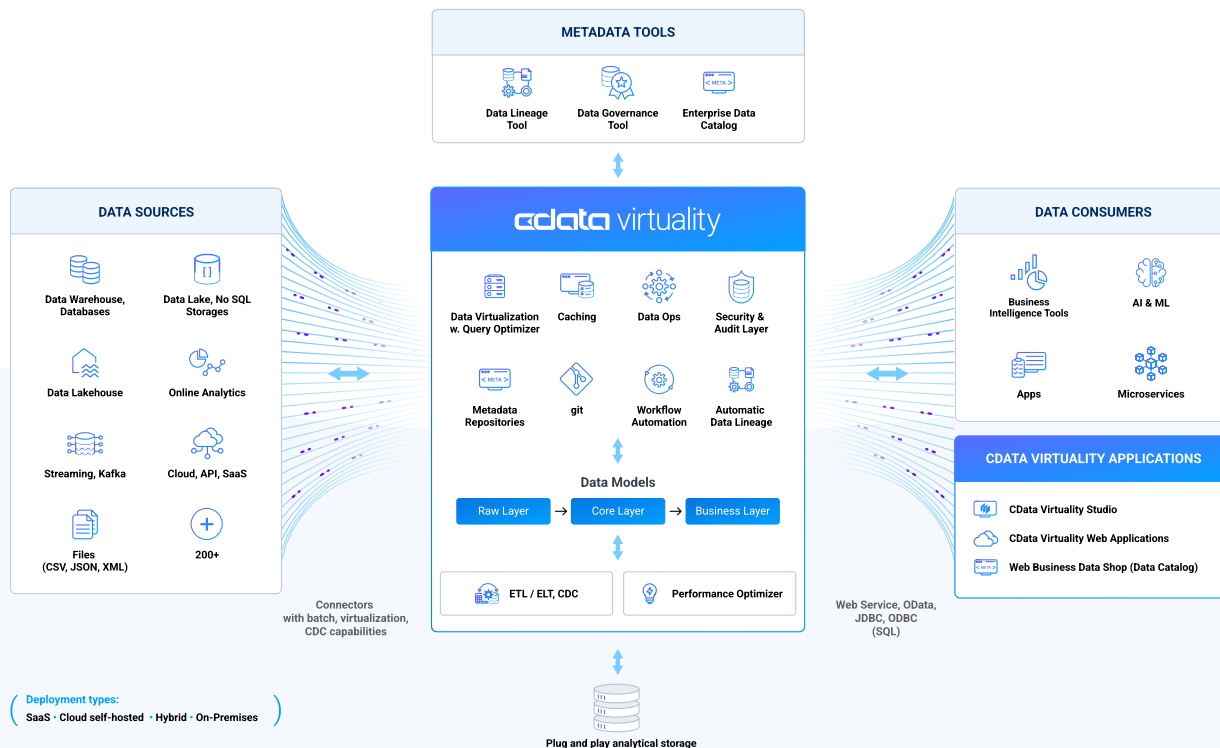
DATASHEET

CData Virtuality

CData Virtuality is a step up in the evolution of data integration, combining the best of data virtualization and ELT/ETL in a single, comprehensive data management solution:

- The data virtualization engine allows to retrieve data using a single query language, get query response, and assemble different data models or views of the data - all in almost real-time.
- Physical data integration is a robust part of the CData Virtuality that ensures fast query response while decoupling performance from the source data stores and moving the data to the physical repository.
- The software provides SaaS (software as a service), on-premises, and hybrid deployment options, enabling you to choose the best-fit for infrastructure management, compliance, and data utilization.

How CData Virtuality Works



1. Connect to your data sources

CData Virtuality 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

CData Virtuality 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

CData Virtuality 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 file systems, or access to CDC (Change Data Capture) interfaces of different systems and databases. CData Virtuality provides predefined jobs for the most common transformations. It 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, CData Virtuality supports the standard interfaces (JDBC, ODBC, REST, OData) to deliver data to the data consumers. This could be, for example, reporting tools, advanced analytics tools, or custom programs in various programming languages.

CData Virtuality features

Deployment types

- SaaS
- On-premises
- Hybrid
- AWS or Azure

Data governance

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

CData Virtuality server

- Windows Server 64bit
- Linux 64bit (Redhat, CentOS, Ubuntu, and others)
- Deployment in Docker Container
- Kubernetes

Business data shop

(self-service web data catalog)

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

CData 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

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)
- Via OData

Move & 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 CData Virtuality Studio/SQL

Data virtualization

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

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)

Streaming

- Kafka connector
- Change Data Capture (CDC) for selected data sources
- Continuous record sets

Databases & connectors

- More than 200 ready-to-use connectors.
- All our connectors can be found: <https://www.cdata.com/drivers/>

Structure optimization

- Self-learning recommendation engine for automatic 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)

Security, authentication, audit

- Kerberos
- 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

In-memory caching for even faster responses

- Session scope
- User scope
- Virtual database scope

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

What else?

- 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 CData Virtuality Management API

CData Software – The leading provider of data access & connectivity solutions

CData Software is a leading provider of data access and connectivity solutions. Our self-service data products and connectivity solutions provide universal access to live data from hundreds of popular on-premises and cloud applications. Millions of users worldwide, including Salesforce, Office Depot, and Holiday Inn, rely on CData to enable advanced analytics, boost cloud adoption, and create a more connected business. Consumable by any user, accessible within any application, and built for all enterprises, CData is redefining data-driven business.

Learn more at www.cdata.com, or reach a representative via email at info@cdata.com.