



CData Sync Product Overview

MARCH 2023



Table of Contents

- About CData Sync 2
- Key Features 2
- High Availability..... 4
- Incremental Updates 4
- Change Tracking 5
- Capturing Deletes..... 5
- Identifying Unknown Data Types..... 5
- Automatically Detecting Known Data Types for the Schema 5
- Transformations 5
- API Support 5
- Notifications..... 6
- Jobs Monitoring..... 6
- Access Control Policies 6
- Parallel Processing 6
- Firewall Traversal..... 6
- Addressing Schema Alterations..... 6
- Supported Sources and Destinations..... 7
- CData Sync Pricing..... 7
- Next Steps with CData Sync..... 7

About CData Sync

CData Sync is a powerful, no-code/low-code data integration solution. With CData Sync, users can quickly integrate and consolidate organizational data from any source, cloud application, or on-premises system and leverage that data for reporting, operational excellence, customer success, and other business initiatives. CData Sync's simple, point-and-click interface allows users to build data pipelines to cloud and on-prem targets, such as data warehouses, data lakes, databases, and applications easily; move data across and within cloud ecosystems; and replicate that data back to mission-critical on-premises applications.

- Deploy in Cloud, hybrid, multi-cloud, and on-premises environments. Businesses can move all their data to targets they choose – no matter where they reside. CData Sync supports the replication of data from cloud to cloud, on-premises to cloud, and cloud to on-premises.
- With intelligent replication, CData Sync delivers high performance data pipelines based on Change Data Capture (CDC), full, or incremental replication with no impact on source systems. CData Sync only retrieves changes to the system of record, minimizing bandwidth usage and reducing latency of synchronization to the replicated data source.
- Move data from 270+ supported data sources to target destinations of choice. CData Sync replicates data to relational databases (e.g., Microsoft SQL Server, Oracle, PostgreSQL, etc.), cloud data warehouses (e.g., Amazon Redshift, Snowflake, Google Big Query, etc.), and other big data systems (e.g., Amazon S3, Azure Data Lake, Kafka, etc.).
- Perform flexible data transformations before, during, or after data extraction by leveraging SQL, dbt Core, or dbt Cloud. By supporting seamless integration of dbt Cloud into data pipelines, Sync ensures that the custom data transformations and modeling that customers create are automatically triggered whenever data flows through the Sync pipeline. In this way, CData Sync enables IT teams to “set and forget” their transformations and empowers data analysts to standardize and improve their data workflows and reduce burdens on IT.
- Deploy in Kubernetes environments to leverage Kubernetes high availability (no downtime), scalable clustering, and backup-and-restore functionality.

Key Features

Simple point-and-click replication: A no-code/low-code implementation eliminates any need for programming, enabling users to set up a data pipeline in minutes.

Automated data replication to major data warehouses and data lakes: Supports popular databases such as Microsoft SQL Server, PostgreSQL, Snowflake, Amazon, Databricks, etc.

Self-contained: No special administration or installation requirements for the target database.

Flexible mapping, configuration and scheduling capabilities allow enterprises to meet business requirements for ideal replication configuration.

Dynamic schema replication: Ensures that data sources and targets are always in sync. CData Sync monitors the sources and automatically detects and updates the local replicated database with those changes.

Secure backup & archiving: Ensures that critical data is stored safely in the environment of your choice.

Incremental updates: Retrieve only changed data to minimize bandwidth usage and latency of synchronization. Of the 270+ data sources to which CData Sync connects, many support Incremental Updates. This means that each scheduled Job run will find what source data has changed since the last run and update the destination accordingly. Incremental updates make it possible to keep very large data sets synchronized, when replicating everything each time would have been prohibitively expensive.

Powerful data transformation: Users can perform flexible transformations before, during, or after data extraction by leveraging SQL, dbt Core, or dbt Cloud. By supporting seamless integration of dbt Cloud into data pipelines, Sync ensures that the custom data transformations and modeling that customers create are automatically triggered whenever data flows through the Sync pipeline. In this way, CData Sync enables IT teams to “set and forget” their transformations and empowers data analysts to standardize and improve their data workflows and reduce burdens on IT.

Logging and transaction monitoring: Monitor changes made to the database and the data source.

Standard enterprise-class security features: Encrypted client-server communications with TLS/SSL.

REST API and Events: REST API and events allow integration with external applications.

Host on-premises or in the cloud: CData Sync is available as a lightweight ASP.NET Web application and a Java web application both of which come packaged with a stand-alone, embedded server, making it easy to host on a platform of your choice. CData Sync is also available in the AWS Marketplace.

Configure, schedule, and monitor data flow: Enjoy complete visibility into your jobs using the Sync Admin Console. Data flow jobs can be simple flows or incorporate transformations, functions, filtering, etc. that modify the data before or after it is moved to the destination.

High Availability

CData Sync offers High Availability features that enable customers to scale data processing quickly across multiple instances of CData Sync. This means they can easily speed up their data movement, transformation, and replication process to meet the changing needs of the business. In essence, they can have the data they want when they want it.

The high availability feature enables an “always on” environment, so customers never have to worry about downtime or slow data processes. It also provides the ability to do more in less time, which accelerates BI and analytical processes and enables customers to be more nimble as they respond to changing business conditions.

Incremental Updates

CData Sync supports [Incremental Updates](#). Rather than querying the entirety of source data every time it needs to be replicated, CData Sync incrementally updates a destination, such as a data warehouse, by only querying for data that has been added or changed since the last time a job ran and then merging that data into the data warehouse.

CData Sync uses two methods for incremental updates:

Incremental Check Columns (ICC): Updates only records that have been modified since the last update using the tables’ Last Modified Column. The two different data type ICC uses are **DateTime** Incremental Check Column (pre-configured for most sources, and other sources can select which column will be used) and **Integer-based** Incremental Column Check (use integer-based for column checks, like auto-incrementing ID).

Log-Based Replication: Some data sources support Log-Based Replication, where a source uses a logfile to log events (Insert, Update or Delete) that cause changes in the database. Rather than querying the source table for changes, Sync reads the logfile for any change events. Sync then extracts those changes for replication and stores the current log position for the next replication.

Data sources listed below support Log-Based Replication:

- **Microsoft SQL Server** uses either [Change Tracking or Change Data Capture](#)
- **Oracle** uses [Oracle Flashback](#)
- **PostgreSQL** uses [Logical Replication](#)
- **MySQL** log-based replication is coming soon.

Change Tracking

CData Sync queries the 'last modified' column within the database. [Change tracking](#) applies to Microsoft SQL Server and differs from Change Data Capture. Please note that not all databases have the 'last modified' column as part of their environment.

Capturing Deletes

CData Sync automatically captures deletes by retrieving a list of deleted records from the source using either the API or the Change Tracking feature (where applicable).

Identifying Unknown Data Types

CData Sync can recognize many undefined data types and can infer the data type based on the data itself. CData Sync recognizes the following data types:

Boolean, Data, Time, TimeStamp, Decimal, Float, Double, SmallInt, Integer, Long, Binary, Varchar, and GUID.

Automatically Detecting Known Data Types for the Schema

CData Sync automatically detects the data types for the schema in most relational databases, such as Oracle, Salesforce, Microsoft SQL, and some APIs. Whenever source column data types are known, Sync automatically creates the matching data type in the destination system(s).

Transformations

CData Sync supports common approaches to data transformations:

- [In-flight ETL](#) (Extract, Transform, Load)
- [In-destination ELT](#) (Extract Load Transform)
- For more detail regarding full capabilities, please refer to CData Sync [Transformation Documentation](#)

API Support

CData Sync has [REST API](#) built-in support, providing a flexible way to manage applications. The job management API allows users to create, update, and execute jobs.

For more information, refer to the [REST API Documentation](#)

Notifications

CData Sync supports email, [Slack](#), and [Microsoft Teams](#) notifications upon completion of a job or whenever there is an error associated with a particular job. For example, if user credentials for a source connection expire, then CData Sync sends a notification with the relevant details that prompt users to make appropriate changes within the console.

Jobs Monitoring

CData Sync provides granular job-level logging that captures all the HTTP activity associated with the source and destination connections of each job, along with job orchestration activities, such as in-flight transformations. There is also application-level logging and an audit log that captures the activities of every CData Sync user within the CData Sync admin console.

Access Control Policies

CData Sync supports three levels of users with varying degrees of privilege. You can find more information here: [CData Sync - User Management and Roles](#)

Parallel Processing

Tasks and jobs are executed in parallel to increase efficiency, and users can control the level of parallel execution. [Learn more](#)

Firewall Traversal

Because CData Sync can run anywhere, it is an ideal solution for customers with hybrid cloud architectures, where some systems run in the cloud and others reside on internal networks. In these hybrid scenarios, customers can install CData Sync to run inside their network, alleviating any need to expose ports to the internet, open firewalls, or create VPN connections.

CData Sync's ability to run anywhere also greatly reduces latency, because it can run close to a source or destination, improving the performance of ETL or ELT jobs.

Addressing Schema Alterations

Data is constantly changing; CData Sync ensures these changes are always reflected. For every run, Sync compares the source schema with the destination schema to look for differences. Should Sync detect a difference in structure between the two, it will modify the destination schema by adding columns or increasing column size to ensure that source data fits.

CData Sync accomplishes this in several ways:

- If a column is detected in the source table that does not exist in the destination table, the Sync will alter the destination table by adding the column.
- If the datatype in the source increases in size, then Sync will alter the destination table by updating the column. This includes increasing the column size of a string column or the byte size of non-string columns.
- **Note:** CData Sync will never delete columns from the destination table if the column was removed from the source table AND will never shrink the size of the destination column if the datatype has been updated in the source.

Supported Sources and Destinations

Please visit this [page](#) to view supported sources and targets.

CData Sync Pricing

CData Sync pricing can be found on the CData [web site](#).

Next Steps with CData Sync

- Sign up for a free CData Sync [trial](#)
- Take a look at CData Sync [documentation](#)
- [Book a demo](#)

CData Software is the real-time data connectivity company. 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 rely on CData to enable advanced analytics, boost cloud adoption, streamline operations, 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.