

Top 8 CDC Tools of 2024

Writer's Name: John Potter

Client: Fivetran

Resources:

- [Edify Style Guide](#)
- [Edify Content Instructions](#)

Meta Title Options:

- "Best CDC Tools 2024: Top 10 Picks for Data Sync"
- "Top 10 CDC Tools in 2024: Key Features Reviewed"
- "2024's Leading CDC Tools: Top 10 Comprehensive Guide"

Meta Description:

Explore the top 10 CDC (Change Data Capture) tools of 2024! This guide reviews the best features, benefits and capabilities to help you choose the right tool.

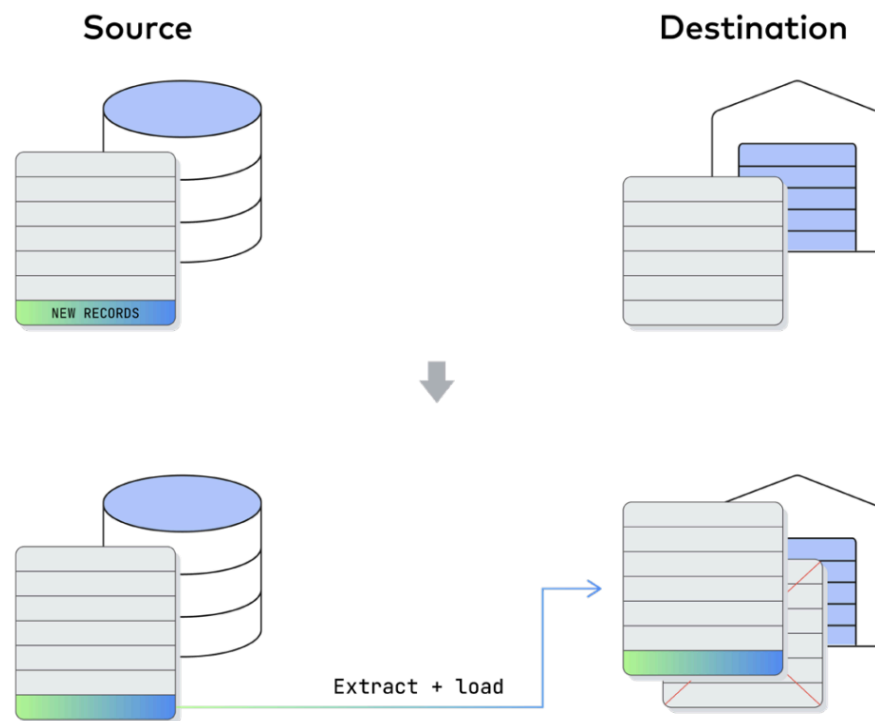
Your business generates data 24 hours a day, seven days a week in many places: first- and third-party applications, employee and customer tracking systems and communication tools like email or chat. If you can't capture this data, you lose the ability to gain insights that could help your business. And if you can't capture the entirety of that data, you could miss out on details that affect the decisions you make.

Change Data Capture (CDC) lets you to track and record data changes in real-time, providing you with the most current and complete information available. For example, if an e-commerce customer adds and removes an item to their cart between incremental data syncs, you may never know they were interested in a purchase. With a detailed data capture method like CDC, you see both the addition and removal of the item and you can decide whether to pursue further action.

Let's explore what CDC tools achieve for a business and what you should look for when selecting one.

H2: What is change data capture (CDC)?

Change Data Capture (CDC) ensures that every change in a source database is replicated to a target system, such as a data warehouse or a data lake. CDC doesn't query data sources directly, but reads from an event log, which minimizes transactional load on the source system. The log data is transferred through data pipelines incrementally, tracking new database events as they occur, rather than extracting or replicating entire datasets. This process makes systems more efficient and reduces bandwidth usage.



H2: Why use CDC tools?

CDC tools streamline manual coding and reduce maintenance costs by automating updates and management tasks. Here are four reasons why choosing CDC tools is a smart choice.

H3: Real-time data integration

Real-time data integration continuously synchronizes data across various systems to keep all information up-to-date and consistent. This capability improves operations that depend on transaction processing and real-time analytics.

The immediate data access CDC provides allows businesses to respond more quickly to market changes and customer needs. For example, Fivetran can replicate CDC data changes as often as every five minutes, allowing you to make on-the-fly business decisions for maximum impact.

H3: Data quality and accuracy

CDC tools offer reliable and accurate data replication, reducing the risk of errors and inconsistencies that may come with other replication methods. CDC captures every data change, including hard and soft deletes, leading to more reliable business insights and strengthens decision-making across the board. It also fosters trust among stakeholders and simplifies compliance with regulatory requirements.

When you use CDC, you can be sure you're getting the highest data quality and accuracy possible.

H3: Advanced analytics and business intelligence

Real-time data access helps businesses leverage analytics for deeper insights and competitive advantage. This steady flow of fresh data integrates seamlessly into business intelligence tools, enabling smarter strategic planning. As data shows up in source systems, it can be visualized in a BI tool within minutes.

Pairing CDC with advanced analytics platforms enhances data processing and quickly converts raw data into actionable insights. This refined data helps companies forecast trends.

H3: Optimized resource usage

CDC tools help companies reduce the computational demands that traditional data integration methods require. Instead of running a ``select *`` query on your source system every time you run an incremental sync, you read a separate log file and only ingest the changes since your last sync. This process lowers operating costs, making the system more responsive and scalable. It also stabilizes network and database performance, leading to less downtime.

H2: The top 8 CDC tools

Below are the top 8 CDC tools to consider, each offering unique features that cater to various needs for effective data migration:

1. [Fivetran](#)
2. [Qlik Replicate](#)
3. [Hevo Data](#)
4. [Airbyte](#)

5. [Striim](#)
6. [IBM InfoSphere](#)
7. [Debezium](#)
8. [Apache NiFi](#)

Select the tool that allows you to migrate data according to your specific needs for a smooth and successful data migration.

H3: 1. Fivetran

[Fivetran](#) offers a fully managed, scalable CDC replication solution that moves and replicates large volumes of data with minimal latency. This tool supports both [Extract, Load, Transform \(ELT\)](#) and [Extract, Transform, Load \(ETL\)](#) processes, allowing you to establish robust data pipelines that facilitate [real-time analytics](#).

The Fivetran CDC tool, with over [500 pre-built data connectors](#) and [dozens of data destinations](#), centralizes data from various sources and funnels it into your data warehouse. These connectors use [log-based CDC technology](#) to support high-volume data replication across on-premise and cloud-based platforms.

Fivetran also employs a data validation feature that maintains systems accuracy. In the event of an error, such as a broken connector or an incomplete sync, [you'll receive an alert and instructions for resolution](#).

Other Fivetran features include pre-built data models, integration with [dbt Core](#), data access control, single sign-on (SSO), external data logging and compliance with [SOC 2 Type II](#) and [ISO 27001 standards](#). They also offer [24/7 global technical support](#) to help with any issues. With Fivetran, you're never alone in your [data integration journey](#).

Pricing: Fivetran offers a flexible pay-as-you-go model, where you only pay for what you use. Visit the [Fivetran pricing page](#) for pricing estimates. Fivetran offers a [14-day free trial](#) to explore the capabilities of its CDC replication solution.

H3: 2. Qlik Replicate

[Qlik Replicate](#) has a user-friendly interface that provides a streamlined, graphical solution for data integration and ingestion. It empowers users to replicate, ingest and stream data across various platforms such as databases, data warehouses and Hadoop systems. Qlik's design minimizes the impact on source systems and eliminates downtime.

The platform boasts several standout features, including automated schema conversion and validation. It offers comprehensive support for numerous data sources and targets, from major databases to large-scale data warehouses.

Pricing: Qlik offers flexible pricing options to meet businesses' specific needs. Interested parties should contact Qlik directly for a personalized quote. Be aware of the extra costs, however, including a per-user license fee and fee for ongoing software updates and support.

H3: 3. Hevo Data

[Hevo Data](#) is a no-code data pipeline platform that simplifies real-time data integration, transformation and automation. This platform lets businesses consolidate their data for analytics without extensive engineering involvement. It's suitable for companies that value simplicity and need to integrate various data sources.

The platform supports over 150 data sources including databases, SaaS applications and cloud storage. It also offers data transformations using Python code and facilitates real-time data loading, equipped with schema detection and automatic mapping.

Hevo's user-friendly, no-code interface ensures it is accessible to newcomers; its ability to automatically identify and integrate data adds an extra layer of assurance for users. Processing large volumes of data can be expensive, however, and there is less control over custom transformations.

Pricing: Hevo Data offers a 14-day free trial for users to test its capabilities. After the trial, users can choose from several paid plans, starting at \$239 per month. The cost varies based on the volume of events and the features included.

Read our [comparison of Fivetran vs. Hevo Data](#).

H3: 4. Airbyte

[Airbyte](#) is an innovative data integration platform that specializes in replicating data from databases and APIs into data warehouses, lakes, and other destinations. Its open-source framework facilitates real-time data synchronization and offers CDC capabilities. Airbyte supports popular databases like PostgreSQL, MySQL, and MongoDB, making it versatile for businesses needing real-time data updates.

Airbyte retains a library of pre-built connectors that continues to grow. It's able to adapt quickly to new data sources and integration requirements. The platform provides both batch and real-time data processing, offering flexibility depending on the user's needs. Its open-source nature allows for custom development, letting businesses tailor the tool to their data integration challenges.

Pricing: Airbyte offers flexible pricing, including a free tier for smaller projects and scalable plans for larger data integration needs.

Read our [comparison of Fivetran vs. Airbyte](#).

H3: 5. Striim

[Striim](#) is a data integration and streaming platform that minimizes overhead on source systems by using log-based CDC for data ingestion. This platform ensures data consistency across all systems through built-in validation for database sources and targets. Additionally, Striim simplifies change data processing into a usable format while preserving its transactional context using SQL-based queries.

Several key features enhance Striim's capabilities, including pre-packaged applications for initial data loads, in-flight data transformations and reliable built-in checkpoints. These checkpoints reassure users that their data is secure. Striim also offers schema change replication and the ability to deliver changes to cloud and on-premise environments.

Pricing: Striim offers three different plans: the Striim Platform, Striim Cloud and Striim for BigQuery, with costs depending on usage. Custom pricing plans are available upon request from the Striim team.

H3: 6. IBM InfoSphere

[IBM InfoSphere Data Replication](#) is a comprehensive CDC solution that offers real-time data replication capabilities and integrates seamlessly with IBM's analytics suite. This feature makes it particularly beneficial for businesses already integrated into IBM's ecosystem.

InfoSphere captures and delivers database events as they occur, making it well-suited for target sources such as message queues and ETL solutions. It skillfully handles data transformations and enrichments during replication, offers advanced conflict resolution for bidirectional replication and supports a diverse array of data sources.

Its robust front-end functionality facilitates the management of tables and databases in both source and target environments. This suite of features guarantees reliable and efficient data management in complex settings.

Pricing: IBM InfoSphere is available for on-premise installations with custom pricing. There are three plans for cloud deployments—Small, Medium and Large—with pricing starting at \$19,000 per month.

H3: 7. Debezium

[Debezium](#) is a free, open-source change data capture (CDC) platform that seamlessly streams database changes into Apache Kafka. It captures real-time, row-level changes to convert your database into a dynamic event stream. This capacity to respond allows different services to react instantly to updates in shared databases.

Debezium supports a wide range of databases, including PostgreSQL, MySQL, MongoDB and SQL Server, allowing you to flexibly manage data applications. It integrates well with existing Apache Kafka ecosystems, although setting it up requires some Kafka expertise and can be complex.

Pricing: Debezium is free to use since it is open-source. Additional costs may arise that relate to deployment and maintenance.

H3: 8. Apache NiFi

[Apache NiFi](#), available as open-source software, features a web-based interface that simplifies the automation and visual management of data flows across systems. It focuses on simplifying data routing, transformation and system integration. It also supports a wide array of data sources and destinations, boasting over 300 processors to enhance its adaptability.

NiFi controls data flow using backpressure, a mechanism that regulates data flow to prevent overwhelming the system, and prioritization, a method that assigns importance to different data flows, to move data efficiently. The platform also prioritizes data security, protecting sensitive data during routing and transformation processes.

Pricing: NiFi's open-source nature makes it a cost-effective choice. It offers organizations powerful tools for data integration without any licensing fees.

H3: Selecting the best CDC tool

Selecting the ideal CDC tool involves [actively assessing key factors](#) that match your business needs. The tool you select should include necessary data connectors, offer data encryption at rest and in transit, provide real-time delivery and fit your budget. By focusing on these considerations, you can choose a CDC solution that aligns with your operational requirements and security protocols.

H2: Best practices to get the most out of your CDC Tool

Organizations that adopt these best practices can greatly boost the efficiency and effectiveness of their CDC tool. Consistently applying these practices leads to smoother operations and improves data handling capabilities.

- **Define clear objectives:** Consider what you want to achieve with CDC, whether it's real-time data integration, enhanced data warehousing or instant reporting. Knowing your goals is key to selecting the right tools and setting them up effectively.
- **Minimize impact on source systems:** Configure CDC to minimize its impact on the performance of source systems. This process might involve tuning the frequency of data

capture or using methods that are less intrusive, such as capturing data from transaction logs instead of using triggers.

- **Ensure data accuracy:** Set up mechanisms to verify that your CDC tool captures and transfers data completely and accurately. This process may include checksums, count checks or reconciliation processes.
- **Implement robust error handling:** Design your CDC processes with robust error handling to manage data discrepancies, interruptions in data flow or failures in data capture. Implement automatic alerts and retry mechanisms to quickly address and resolve these issues.
- **Monitor and optimize performance:** Regularly monitor the performance of your CDC solution and optimize it as necessary. This process includes adjusting configurations, upgrading hardware or refining the data transformation processes to handle larger volumes of data or to reduce latency.
- **Secure data transmission:** Make sure that any data your CDC tool captures and transmits is secure, particularly when dealing with sensitive information. To safeguard data integrity and privacy, implement encryption, use secure network protocols and set up robust access controls.

Implementing these best practices allows companies to respond swiftly to market changes and customer needs. These practices bolster operational resilience in a data-driven environment.

H2: Start using Fivetran to sync your data

CDC tracks and records updates from source systems, replicating any changes to the target database. This process accelerates analytics by providing real-time data access, reducing the burden on operational systems by handling data more efficiently.

Building on this technology, Fivetran [efficiently replicates](#) large volumes of data in real time with minimal latency. Our platform boasts over 200 pre-built data connectors, all optimized with CDC technology, making it easy to integrate and streamline your data processes. It lets you effortlessly synchronize and manage your data across multiple systems, improving data accessibility and reliability for analytics.

Interested in seeing how Fivetran can transform your data management? [Sign up today](#) to begin your 14-day free trial and experience the benefits firsthand.