

Onehouse + Confluent = Limitless Real-Time Workloads

Build real-time change data capture pipelines to the Onehouse Universal Data Lakehouse™ in minutes with zero infrastructure management

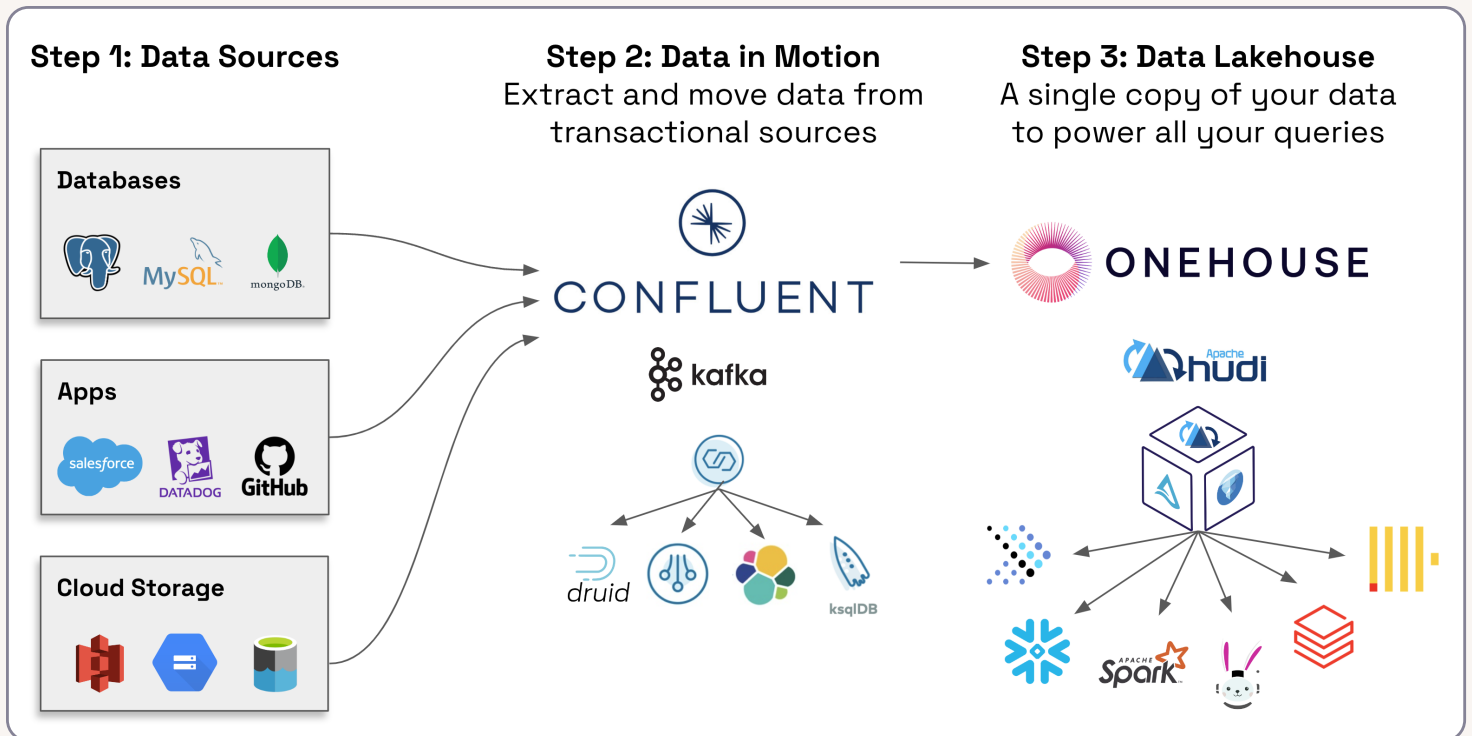


With Onehouse's Universal Data Lakehouse™, you get

- Connect with Confluent certified partner and tight integration with Confluent Cloud
- Fully managed ingestion into an analytics-ready data lakehouse
- Managed services to simplify real-time database replication with change data capture

Overview

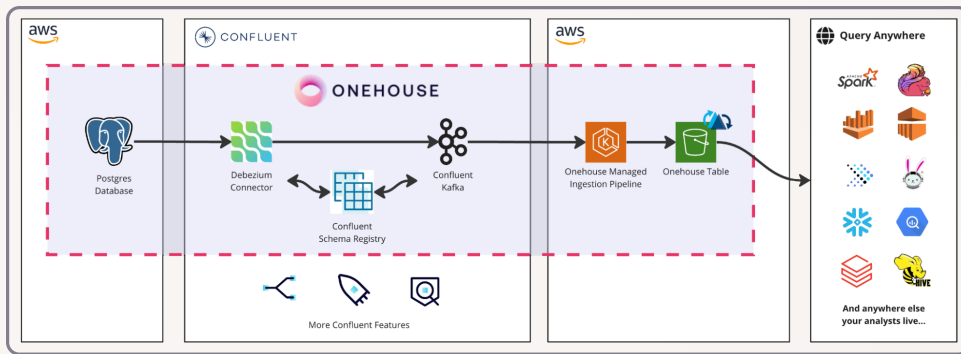
The Onehouse data lakehouse-as-a-service is the trusted source of data for all your workloads, with native support for streaming events. Confluent created the foundational platform for data-in-motion. With Confluent Cloud and Onehouse, you can build limitless real-time workloads in minutes to power use cases across your entire ecosystem, including change data capture, analytics, AI and ML, and more.



It's simple to build real-time pipelines that ingest Apache Kafka topics from Confluent Cloud and then apply transformations with Onehouse's analytics-ready lakehouse. With the Onehouse lakehouse platform's industry-leading interoperability, your trusted data is available to query from a number of popular services, including Snowflake for BI and analytics and Databricks for AI and ML.

Real-time operational database replication via change data capture (CDC) - open and easy

Organizations often need to analyze data from their operational databases to build dashboards and uncover new insights about their customers and operations. However, relational databases such as PostgreSQL and MySQL are not a good fit for analytics at scale because of challenges with costs, storage limitations, and analytical workload performance.



In the architecture diagram above, we are replicating a PostgreSQL database to the Onehouse data lakehouse by capturing change logs with Confluent Cloud.

Onehouse and Confluent Cloud are both designed for streaming data, forming a powerful combination for ingesting CDC data in real-time. As changes happen in your database, you can perform analytics on them live in the data lakehouse. Onehouse manages the lifecycle of the Debezium Confluent Connector, bootstrapping data, restarting connectors, and updating configurations

CDC for a Large-Scale Job Marketplace

A Onehouse customer with large deployments of MySQL has many transactional datasets. With Kafka and Onehouse, they extract changelogs and create low-latency CDC pipelines to enable analytics-ready Hudi tables on S3.

About Onehouse

Onehouse offers a fully automated, cloud-native data lake management solution that significantly reduces the time and resources required to operationalize a production-grade data lakehouse. Leveraging the power of Apache Hudi, it delivers near real-time updates and superior handling of mutable data, making data immediately queryable and available for insights. Its open and interoperable architecture ensures future-proof data management, enabling businesses to use various specialized frameworks on a single instance of their data. Designed for scalability and efficiency, Onehouse empowers businesses to focus on deriving value from their data rather than managing it. Learn more at onehouse.ai.

About Confluent

Confluent is the data streaming platform that is pioneering a fundamentally new category of data infrastructure that sets data in motion. Confluent's cloud-native offering is the foundational platform for data in motion—designed to be the intelligent connective tissue enabling real-time data, from multiple sources, to constantly stream across the organization. With Confluent, organizations can meet the new business imperative of delivering rich, digital front-end customer experiences and transitioning to sophisticated, real-time, software-driven backend operations. To learn more, please visit www.confluent.io.

Seamless Kafka ingestion into the data lakehouse

Onehouse integrates with Confluent Cloud to ingest your messages into a persistent data lakehouse for analytics. With Onehouse, you can:

- Efficiently merge Kafka messages into tables
- Handle late-arriving data
- Store data in open, analytics-optimized columnar formats
- Auto-optimize data for efficient queries
- Enable incremental ETL pipelines to derive new tables from raw data

Source



Destinations

