

Introducing the Onehouse Universal Data Lakehouse™

Combine the scalability and flexibility of data lakes with the stability and accessibility of data warehouses, and open it to your entire ecosystem



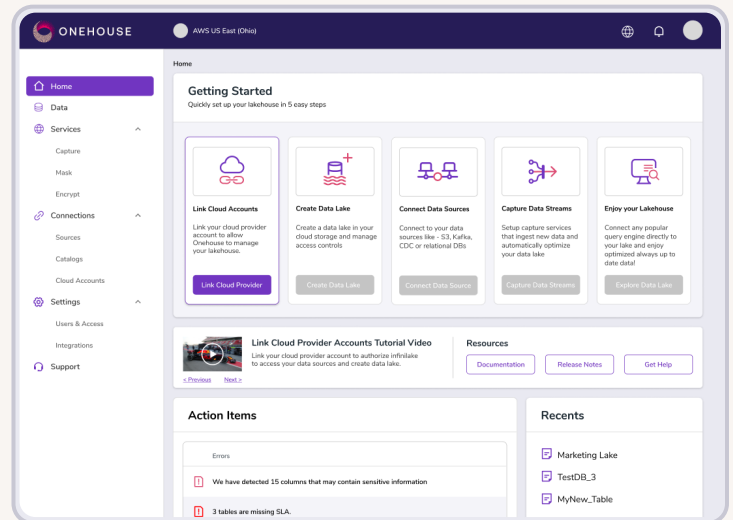
With Onehouse's Universal Data Lakehouse™, you get a fully-managed lakehouse-as-a-service built for:

- Continuous pipelines and incremental processing
- Events-driven and batch-driven workflows in one place
- Automatic optimizations such as cleaning, compaction, and clustering
- Open storage across table formats, query engines, and clouds
- Security, always running in your VPC with SOC II Types 1 and 2 certification

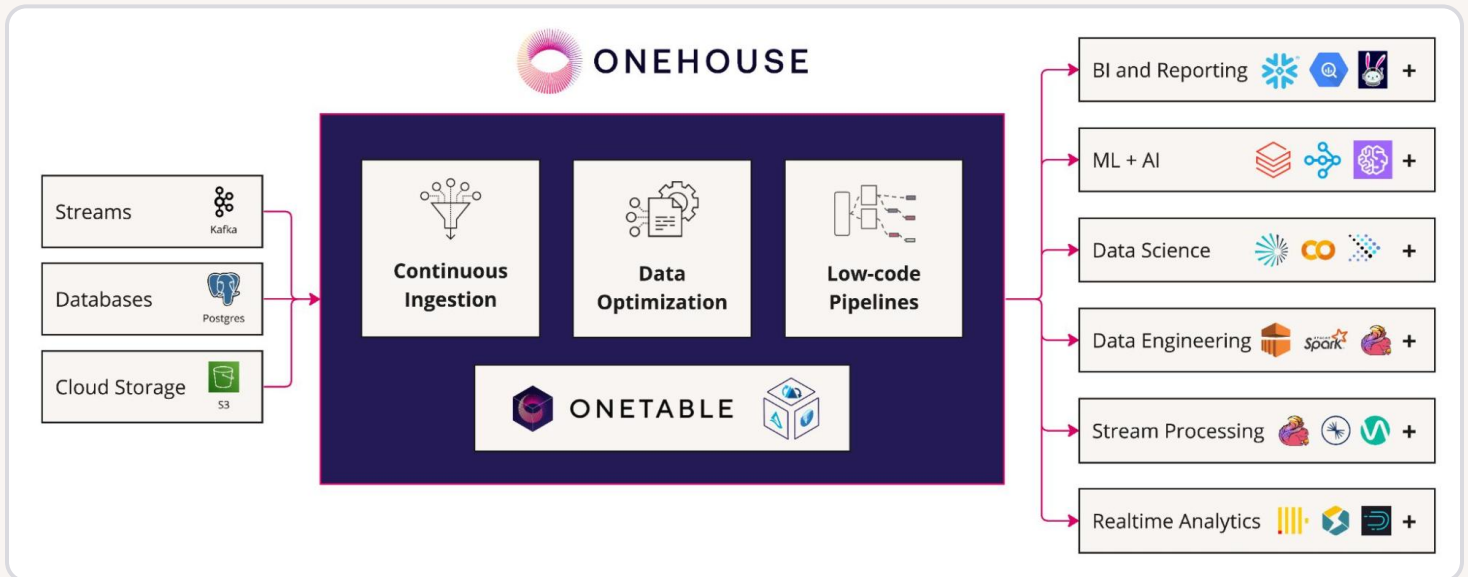
Overview

The Onehouse Universal Data Lakehouse™ delivers a managed data platform-as-a-service featuring the data lakehouse. The data lakehouse combines the scalability and flexibility of data lakes with the stability and accessibility of data warehouses. With Onehouse, data engineers can deliver continuous pipelines with automatic optimizations and open storage across table formats, query engines, and clouds in hours - not months or years.

Onehouse enables you to build limitless near real-time workloads in minutes to power use cases across your entire ecosystem, including change data capture, analytics, AI and ML, and more. You can bridge your events-driven data lake workflows and your batch-driven data warehouse workflows into one near real-time lakehouse, creating one source of truth for all your data at a fraction of the cost and effort.



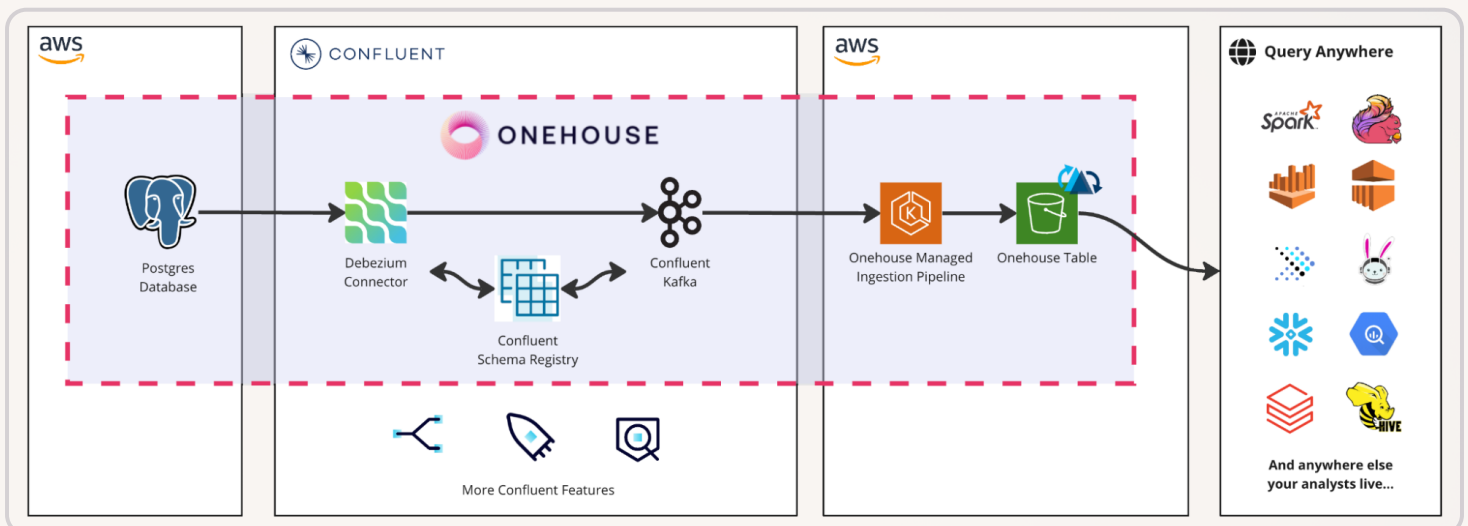
The Onehouse UI makes it dead simple to create a data lake, select data sources, and build transformations on continuous pipelines to the lakehouse in a matter of minutes for use cases such as change data capture, streaming ingestion, and incremental processing.



Onehouse is architected to simplify the near real-time ingestion of data from streams, databases and cloud storage and process that data for a number of use cases such as reporting, ML + AI, and data science.

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

One of the most common use cases for Onehouse is database replication via change data capture. Organizations often need to analyze data in 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.

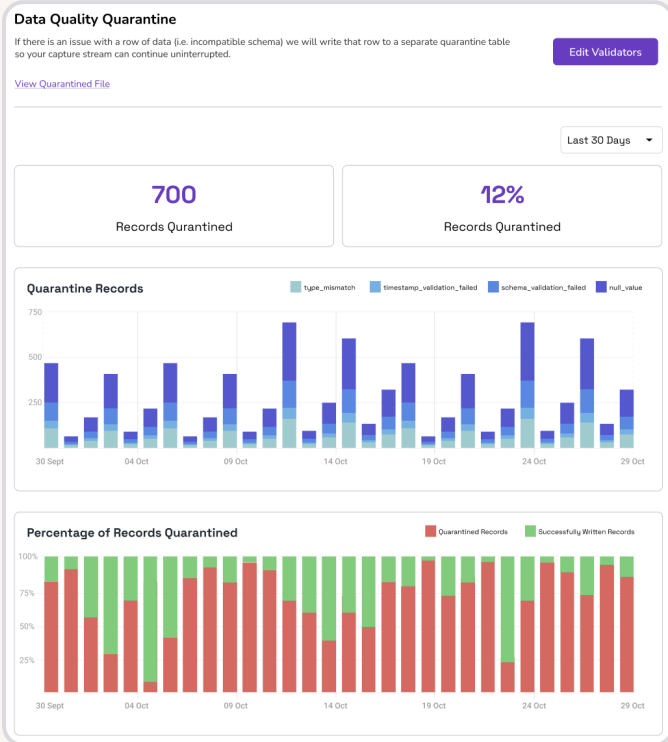
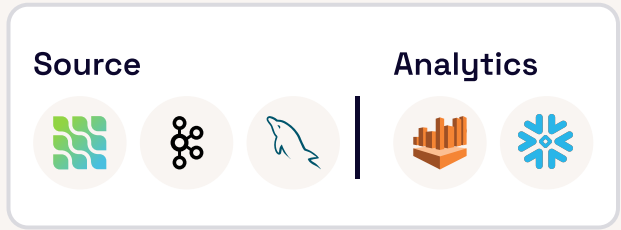


In the architecture diagram above, Onehouse is replicating a PostgreSQL database to the Onehouse data lakehouse by capturing change logs with Confluent Cloud.

Onehouse can work with Debezium and Apache Kafka or Confluent Cloud to replicate operational databases, 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.

CDC at International Professional Networking Organization

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, which they also analyze on Snowflake with Onehouse's interoperability with Apache Iceberg.



Proactive Data Quality Checks Keep Your Pipelines Running Smooth

Capture unexpected data into quarantine tables instead of halting your entire pipeline.

With its Data Quality Quarantine feature, Onehouse ensures your pipelines are always delivering accurate data, while preserving every bit of data that enters your lakehouse. Data Quality Quarantine protects you from upstream schema changes, malformed records, and data that falls outside expected data ranges.

Onehouse can be deployed in minutes



Additional Common Use Cases for Onehouse

Data ingestion: A customer previously using Fivetran for ingestion to Snowflake was able to cut their Fivetran bill by 75% by moving most of their ingestion pipelines to Onehouse.

Incremental processing: A large SaaS company was previously using Kafka and cloud storage for batch processing. By moving to incremental processing with Onehouse, they improved data freshness from 3+ hours to less than 15 minutes and reduced infrastructure costs 40%.

Real-time machine learning pipelines: An insurance company uses Onehouse to generate real-time quotes for customers on their website. Onehouse helped access untapped datasets and reduced the time to generate an insurance quote from days to less than an hour.

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 and the openness of Onetable, 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.