ClickHouse

Better Together: Postgres and ClickHouse

Kaushik Iska
Tristan Ahmadi

Sep 17th, 2024

What is ClickHouse?

Open source

- 33.4k+ Github stars
- 6.3k Forks
- 1.3k+ Contributors
- 100k+ Commits

column-oriented

- Best for aggregations
- Files per column
- Sorting and indexing
- Background merges

distributed

- Replication
- Sharding
- Multi-master
- Cross-region

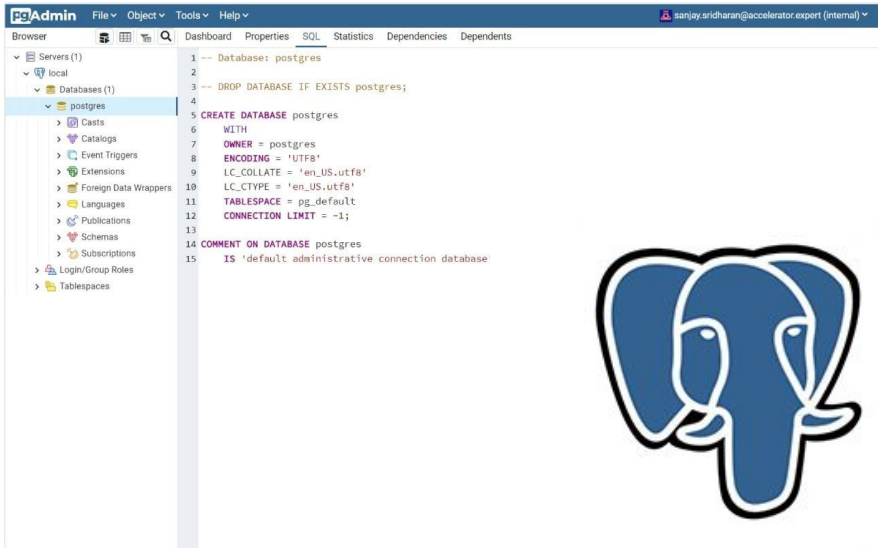
OLAP database

- Analytics use cases
- Aggregations
- Visualizations
- Mostly immutable data

What is PostgreSQL?

Open source relational database management system:

- **ACID Compliance**
- **Data Integrity**
- **Rich ecosystem of tools and extensions**
- Stored Procedures
- Security Auditing
- SQL Standards Compliance
- Comprehensive Documentation
- Active Community
- Open Source



PostgreSQL

Optimized for row-based operations and transactions

- Hits performance limitations with analytical workloads
- Scales inadequately as data volumes increase
- Single master node can become a bottleneck for write-intensive applications
- Inevitably leads to growing operational complexity

Delivers unparalleled performance for analytical workloads at scale

- ✓ The fastest real-time database for analytics
- ✓ Purpose-built to manage massive volumes of data. Scales both vertically and horizontally
- ✓ With ClickHouse, real-time just works. There is no need for the operational complexity that exists when retrofitting another system for these workloads

Company and culture

ClickHouse acquires PeerDB to boost real-time analytics with Postgres CDC integration

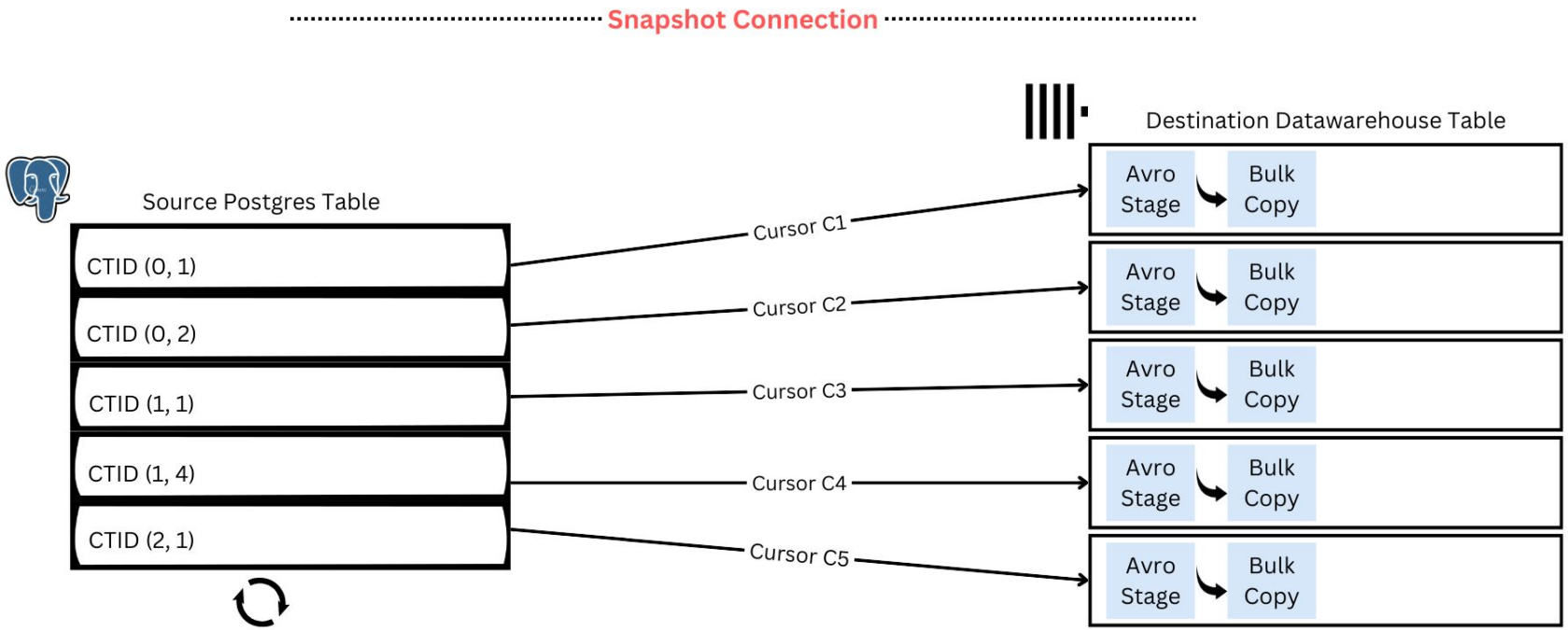
ClickHouse, Inc., the company behind the world's fastest and most popular real-time analytical database, is thrilled to announce the acquisition of PeerDB, a leading provider of change data capture (CDC) solutions

ClickHouse Team
Jul 30, 2024


Transactional
Row Oriented
OSS
Widely Adopted
CDC
Analytical
Column Oriented
OSS
Widely adopted

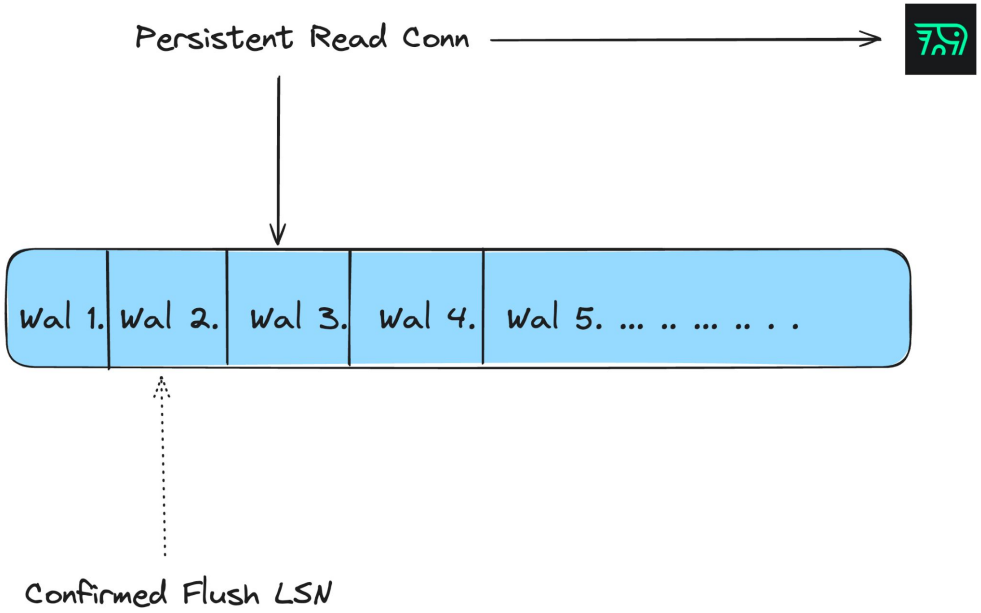
Fast, simple, and cost effective Postgres replication

Parallel Snapshotting



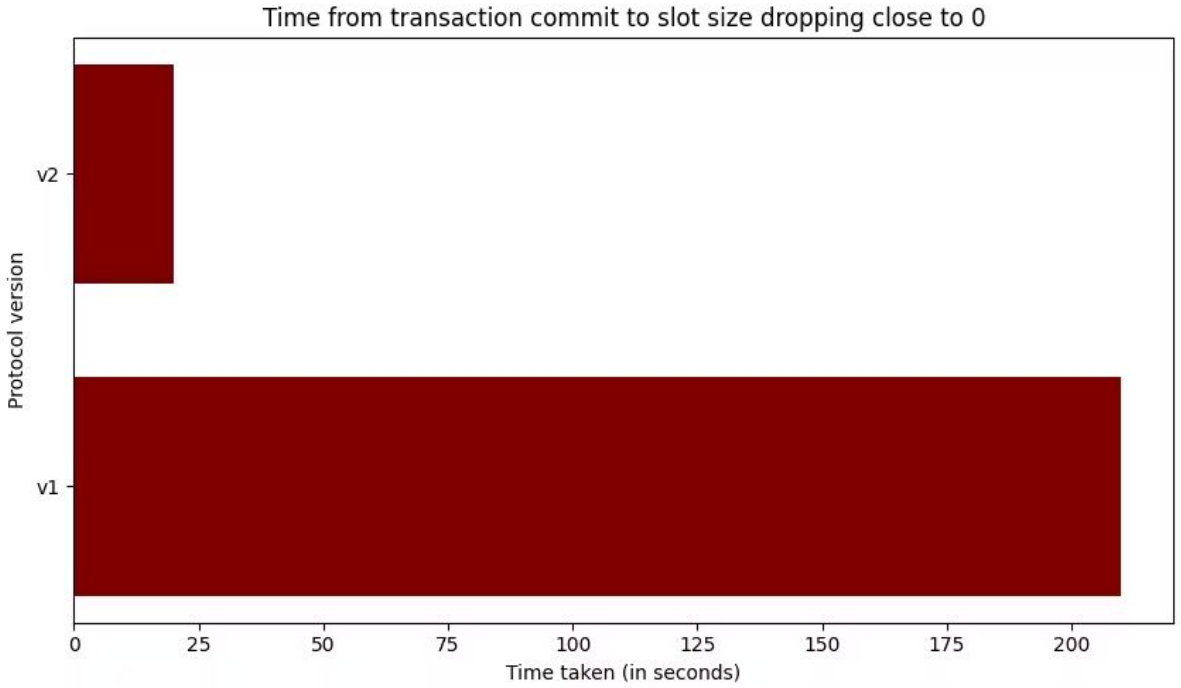
Move Terabytes within Hours instead of Days

Always Consume the replication slot



Your Postgres Database is Crash Safe

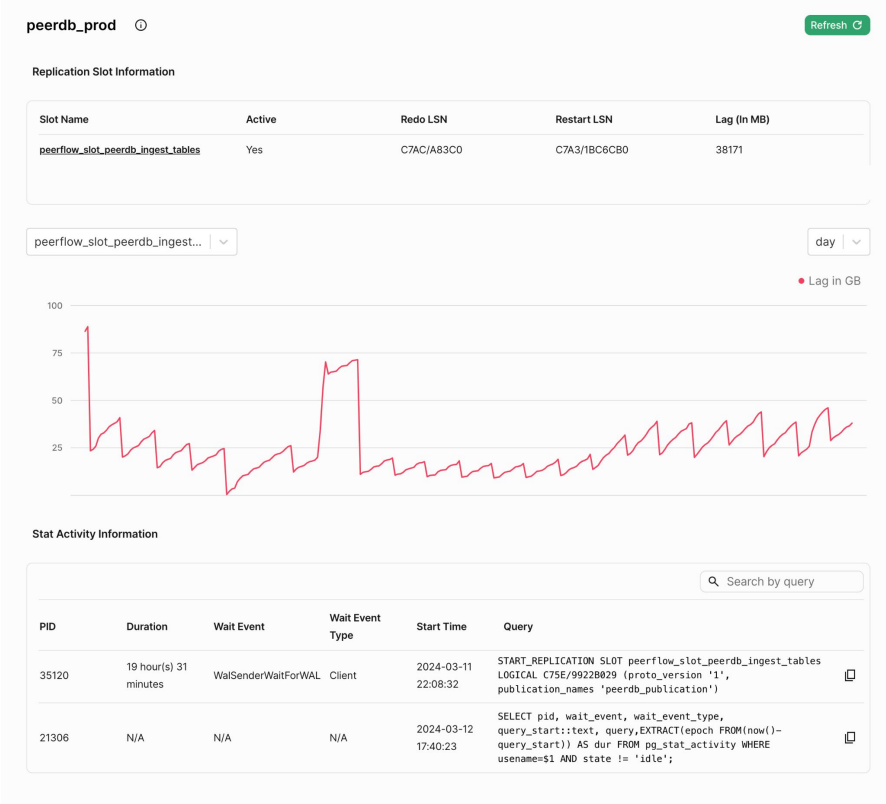
Consume in-flight transactions using version 2



Drastically reduced slot growth and better perf

Postgres Native Monitoring

1. Replication Slot Growth / Lag
2. Number of connections
3. Activity on Postgres - Wait Events



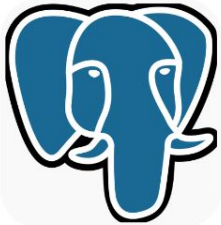
PeerDB is Built For Postgres

Binary Compressed Data Transfer w/ Avro

Native Monitoring and metrics

No Kafka or other external deps

Parallel reading of CDC replication slots



CTID based initial load

Resilient Orchestration

Binary Copy For Homogeneous Transfer

Streaming Server Side Cursors

Query Based Streaming

Parallel Snapshotting

DBT Integration

Toast Column Support

Lossless Unified Datatypes

Demo

PeerDB: 2 main Use-cases

- **One time migrations from Postgres to ClickHouse**
  - One-click migrations
  - 10x faster than other tools
- **Continuous Replication from Postgres to ClickHouse:**

  Postgres (OLTP) and ClickHouse (OLAP) co-exist
  - **Direct connectivity / pipe** - saves months of effort and no need to setup multiple components - kafka, source connector, sink connector
  - 10x faster than existing tools for Initial Load - TBs in hrs vs days
  - High throughput (over 25K TPS) and low latency (<10s)

Q/A

ClickHouse