

|||| ClickHouse

Meetup

WeWork Blue One Square

Gurgaon, India

March 22, 2025 at 10:30 AM IST



Meet the ClickHouse crew



Lalit Mishra

Country Manager, India



Rakesh Puttaswamy

Solution Architect



Anshul Mathur

Account Executive



Piyush Vasudev

Support Engineering Manager



Kunal Gupta

Senior Software Engineer



Amogh Bharadwaj

Software Engineer



Cheryl Tuquib

Head of Field Marketing, APJ

Tech Talks



Introduction to ClickHouse

Rakesh Puttaswamy,
Solution Architect @ ClickHouse



ClickPipes Overview and demo

Kunal Gupta & Amogh Bharadwaj, Software Engineers
@ ClickHouse



Building Intelligent Applications with ClickHouse & LangDB

Matteo Pelati, Co-founder @ LangDB.ai



Optimizing Log Management with ClickHouse

Pushpender Kumar, DevOps Architect @ OLX India



ClickHouse at Physics Wallah

Utkarsh G. Srivastava, Software Development Engineer III
@ Physics Wallah



FabFunnel & ClickHouse: Delivering Real-Time Marketing Analytics

Anmol Jain & Siddhant Gaba, SDE-2 @ Idea Clan



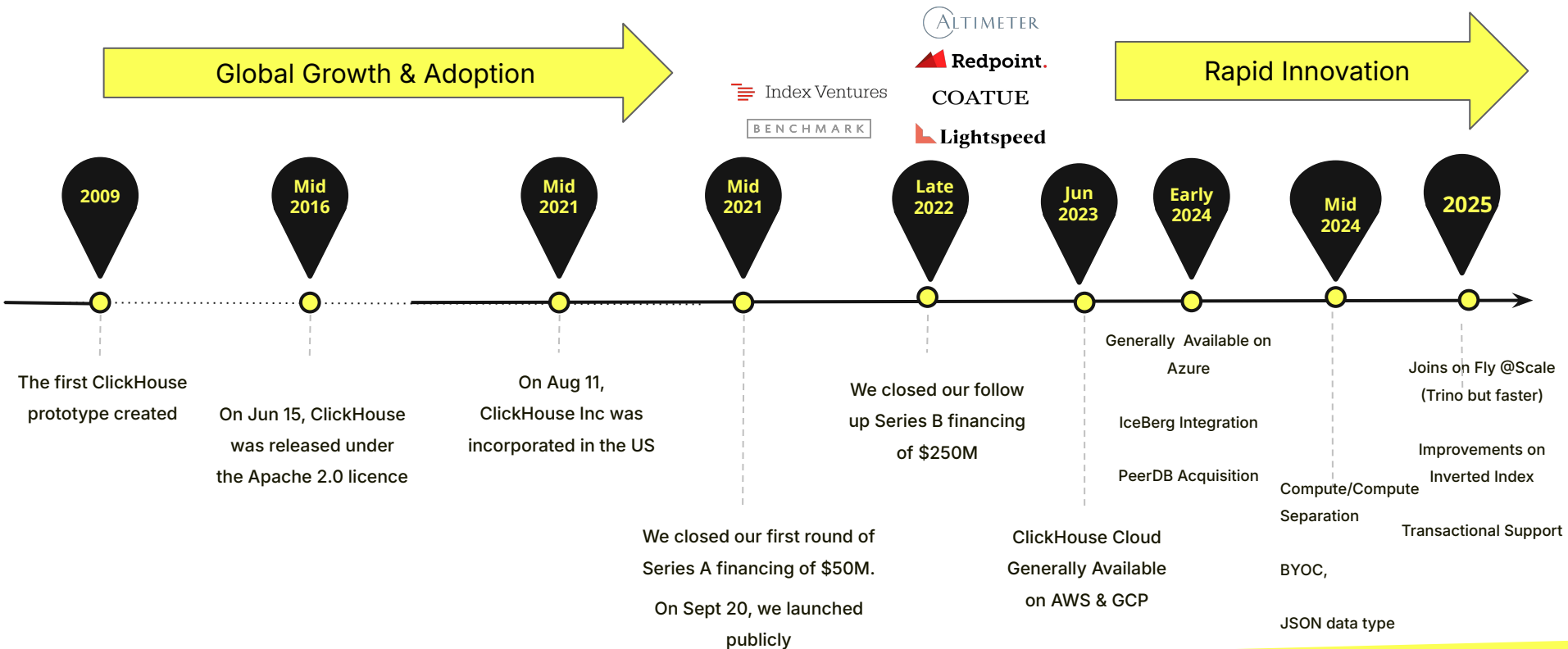
ClickHouse

A Quick Overview

Feb 2025

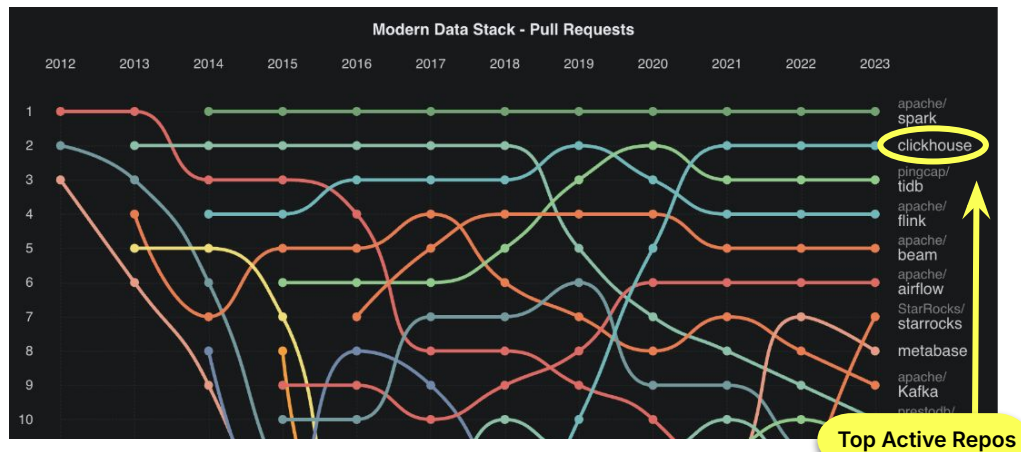
 ClickHouse

The ClickHouse Journey so far.....



ClickHouse Open Source

- ✓ 38.7k Stars
- ✓ 6.1k Forks
- ✓ 1.3k Contributors
- ✓ 100k+ Commits
- ✓ 114k active community members



ClickHouse Cloud



Bloomberg

- ✓ 300+ customers
- ✓ 150 active trials
- ✓ Processed 64+ billion queries, 28 quadrillion records, and 580 PiB of data since launch

NETFLIX

Uber



Spotify

Microsoft

Deutsche Bank



NGINX

CLOUDFLARE

CISCO

COMCAST

Walmart

ROKT

GitLab

ebay

IBM

Contentsquare



What is ClickHouse?

Your (soon-to-be) favorite database!

Open source **column-oriented** **distributed** **OLAP** database

Since 2009
31,000+ GitHub stars
1300+ contributors
500+ releases

Best for aggregations
Files per column
Sorting and indexing
Background merges

Replication
Sharding
Multi-master
Cross-region

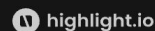
Analytics use cases
Aggregations
Visualization
Mostly immutable data

Use cases



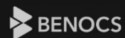
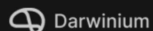
Real-time Analytics

Power interactive applications and dashboards that analyze and aggregate large amounts of data on the fly. Run complex internal analytics in ms, not mins or hrs.



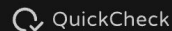
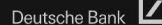
Logs, events, traces

Monitor with confidence your logs, events, and traces. Detect anomalies, fraud, network or infrastructure issues, and more.



Business intelligence

Interactively slice and dice your data for analysis, reporting, and building internal applications. Evaluate user behaviors, ad and media perf, market dynamics, and more.



ML and Gen AI

Execute fast and efficient vector search. Plug-and-play Generative AI models from any provider. Use lightning-fast aggregations to power model training at petabyte scale.



Key Features

Some of the cool things ClickHouse can do

1 Speaks SQL

Most SQL-compatible UIs, editors, applications, frameworks will just work!

2 Lots of writes

Up to several million writes per second - per server. "1 billion writes per second"

3 Distributed

Replicated and sharded, largest known cluster is 4000 servers.

4 Highly efficient storage

Lots of encoding and compression options - e.g. 20x from uncompressed CSV.

5 Very fast queries

Scan and process even billions of rows per second and use vectorized query execution.

6 Joins and lookups

Allows separating fact and dimension tables in a star schema.



Creating tables

ClickHouse SQL Basics

```
CREATE TABLE customers
(
  name      String,
  age       UInt8,
  address   Array(String),
  city      LowCardinality(String),
  created   DateTime,
  type      Enum8(...),
  attr      JSON
)
ENGINE = MergeTree
ORDER BY (city, name, type)
```

■ Other engines:

- ◇ ReplacingMergeTree
- ◇ CollapsingMergeTree
- ◇ AggregatingMergeTree

■ Integration engines:

- ◇ Kafka, RabbitMQ
- ◇ MySQL, PostgreSQL, MongoDB
- ◇ JDBC, ODBC
- ◇ S3, HDFS
- ◇ EmbeddedRocksDB

- Adding Replicating- in front makes an engine replicate data (e.g. ReplicatingMergeTree)



Inserting directly

ClickHouse SQL Basics

```
INSERT INTO people VALUES ('Obi-Wan Kenobi', 57, ...) ('Yoda', 900, ...)  
(...)
```

- Batching is important! Either batch yourself or turn on asynchronous inserts:

```
SET async_insert = true  
INSERT INTO people VALUES ('Obi-Wan Kenobi', 57, ...)  
INSERT INTO people VALUES ('Yoda', 900, ...)  
INSERT INTO people VALUES (...)
```



Reading data

ClickHouse SQL Basics

- Saving a query as a view (no data movement):



```
CREATE VIEW view AS SELECT ... FROM table ...
```

- Continuously processing new data from a table into another table:



```
CREATE MATERIALIZED VIEW view AS SELECT avg(...) FROM table GROUP BY ...
```

```
CREATE MATERIALIZED VIEW view REFRESH EVERY 1 MINUTE TO table_name AS ..
```

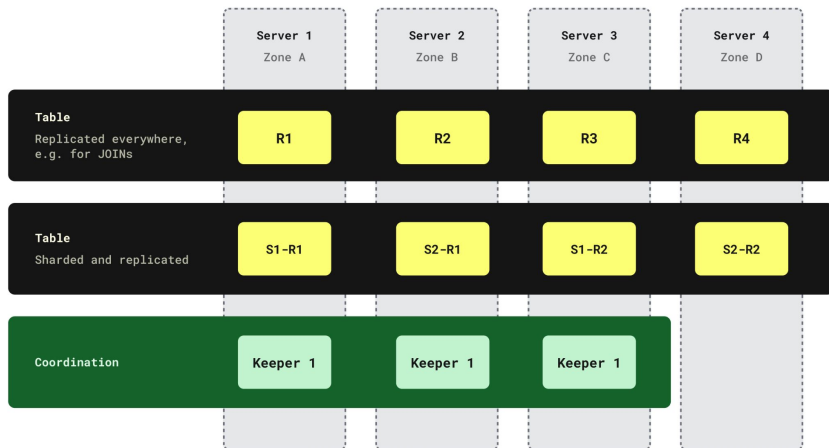


ClickHouse

Self-managed

- ✓ Open-source
- ✓ Flexible architecture
- ✓ Efficient and robust
- ✓ Support contracts available

Sample self-managed architecture

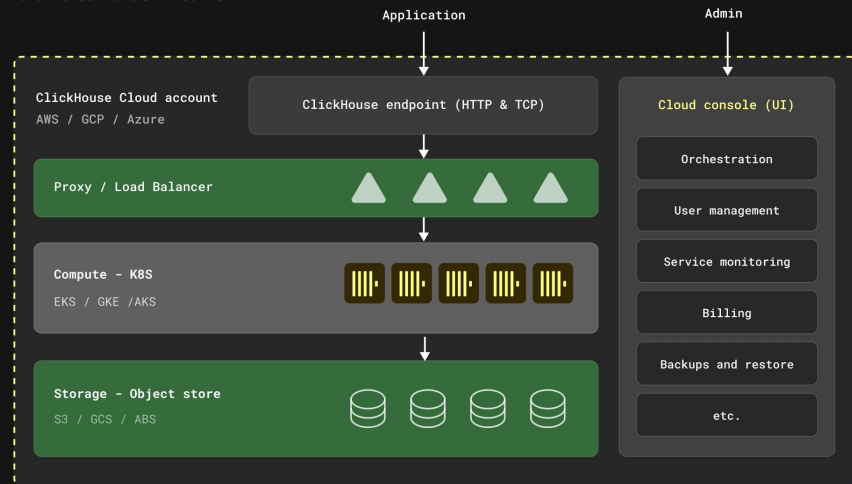


ClickHouse

Cloud

- ✓ Easy to use
 - ✓ Feature-rich
 - ✓ Fast
 - ✓ Scalable
 - ✓ Reliable
 - ✓ PAYG
- Managed for you
Cloud-first features & tooling
Automatically maximizes performance/efficiency
Scale seamlessly
Ensure reliability
SaaS usage and capacity based pricing

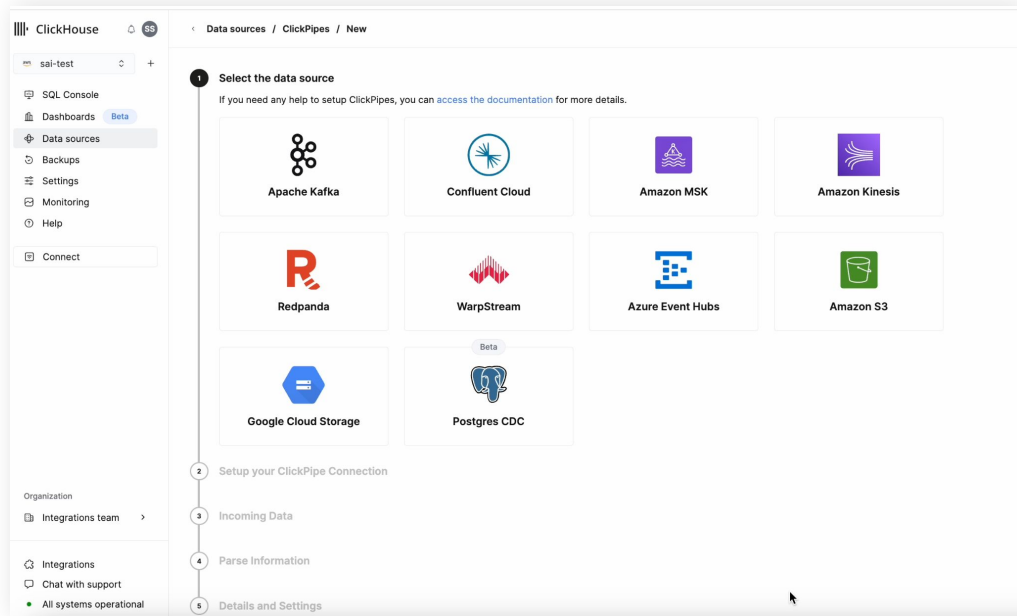
ClickHouse Cloud architecture



ClickPipes Overview & Demo

ClickPipes

- An integration engine that makes ingesting massive volumes of data from a diverse set of sources as simple as clicking a few buttons.
- Simplifies data ingestion from a variety of sources, including Kafka, Kinesis, Postgres, Amazon S3, and Google Cloud Storage.
- Scalable architecture ensures high throughput and low latency, ideal for demanding 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

Fastest and most cost-efficient ETL tool to stream data from Postgres to ClickHouse (and other destinations)

**Open
Source**

2.4k ★ on
GitHub

**ClickHouse
Cloud**

Postgres CDC
connector for
ClickPipes

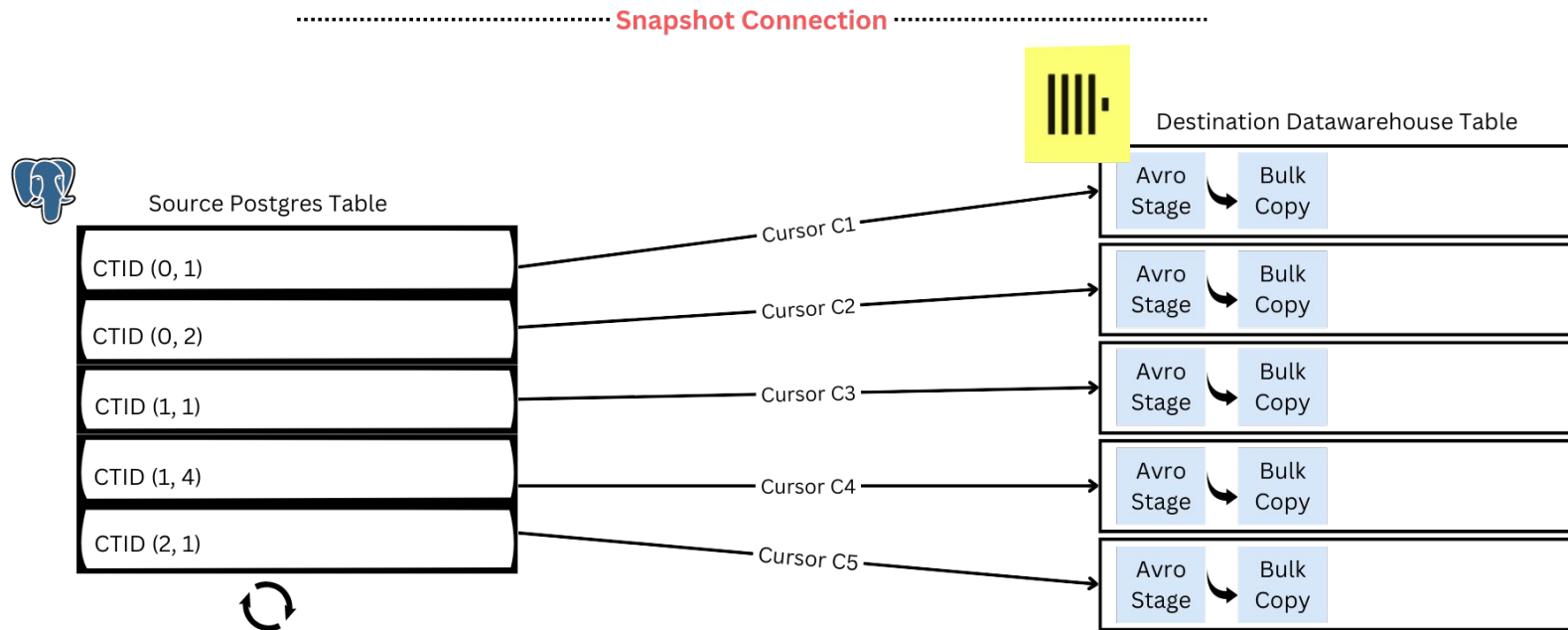


Postgres ClickPipes Supported Features

- Specify custom order keys on ClickHouse
- Support for Read Replicas
- Partitioned Tables
- Nullable Columns
- Column exclusion
- Automatic Heartbeating for PG14+
- Multiple metrics via OpenTelemetry
- Geospatial Data (PostGIS)



Parallelized Initial Snapshot For CDC



The Postgres CDC connector for ClickPipes

- Native integration with ClickHouse Cloud
- Supports CDC and 1-time migrations
- Connection to private Postgres instances via SSH Tunneling/AWS PrivateLink
- Out of the Box alerts for slot lag and ClickPipe errors
- Comprehensive 24×7 on-call support



ClickPipes Demo

Upcoming Features

- MySQL as a Source
- ClickHouse new JSON Type
- Postgres Logical Replication V2

Available in Public Preview!



clickpipes.peerdb.io

KEEP IN TOUCH!



clickhouse.com/slack



[@clickhousedb](https://twitter.com/clickhousedb)



[@clickhouseinc](https://www.linkedin.com/company/clickhouseinc) #clickhouseDB



[clickhouse](https://github.com/clickhouse)

 ClickHouse

OBSERVABILITY WORKSHOP

April 15, 2025 - 9:00 AM IST



clickhou.se/clickhouse-o11y-apr2025

Thank you!