



meetup

IMDA @ Pasir Panjang Road

Singapore

February 25, 2025 at 6:00 PM SGT



Thank you to our host!



**INFOCOMM
MEDIA
DEVELOPMENT
AUTHORITY**

Tech Talks



Using ClickHouse for AI

Maruthi Lokanathan, Solution Architect @ ClickHouse



Real-Time Data Intelligence for Real Estate Analytics

Yi Sam Lee, Data Lead @ Real Estate Analytics (REA)



From SQL to AI: Building Intelligent Applications with ClickHouse and LangDB

Matteo Pelati, Co-Founder of LangDB



BONUS TRACK - DeepSeek data breach: What Could Have Prevented It?

San Tran, all-around security due @ ClickHouse



ClickHouse Cloud AI/ML

Powering Enterprise AI at Scale

2025

||||· ClickHouse

Maruthi Prithivirajan

Solution Architect @ ClickHouse

Say hi!



Read my blogs!



Agenda

01

Intro to ClickHouse

02

Uniqueness of ClickHouse

03

AI/ML with ClickHouse

ClickHouse

A Quick Overview

Feb 2025

||||· ClickHouse

ClickHouse

2009
Prototype

2012
Production

2016
Open Source

2021
ClickHouse Inc.

2022
ClickHouse Cloud

The Most Popular Analytics Database on the Planet

#1
Analytics DB on DB-Engines

Over
39,000
GitHub Stars

Over
200,000
Community Members

ClickHouse, Inc.

Founded

2021

*by leaders from Elastic, GCP,
Netflix and Salesforce*

Raised

\$300M+

*from Tier 1 venture capital firms:
Benchmark, Index, Lightspeed*

Onboarded

1000+

*customers since launching
ClickHouse Cloud in 2022*

230+

employees

Mountain View, CA
Amsterdam, NL
Dual Headquarters

AWS GCP Azure
*Strong partnership with
all major cloud providers*

What is ClickHouse ?

ClickHouse is an **Open-Source**, columnar **OLAP** database
Designed for **Blazing fast** analytics of massive volumes of data

1

Speaks SQL fluently



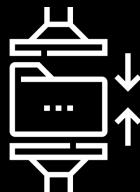
2

Processes data very fast



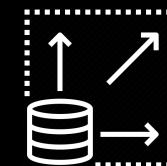
3

Highly efficient storage



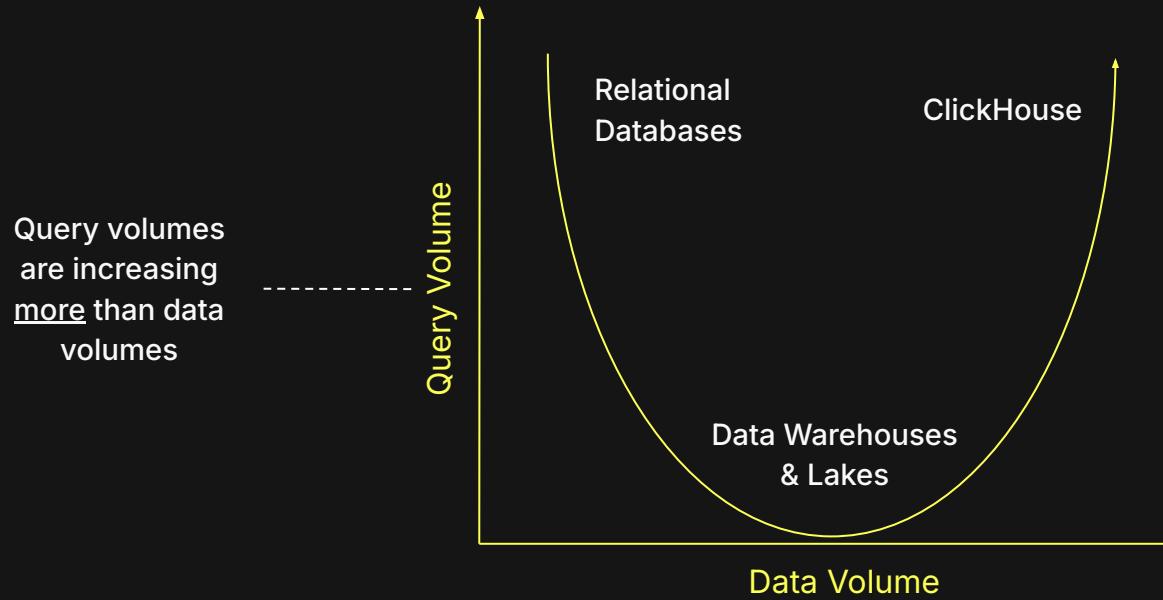
4

Easily scalable to any size



Build for the future

Skating to where the puck is going



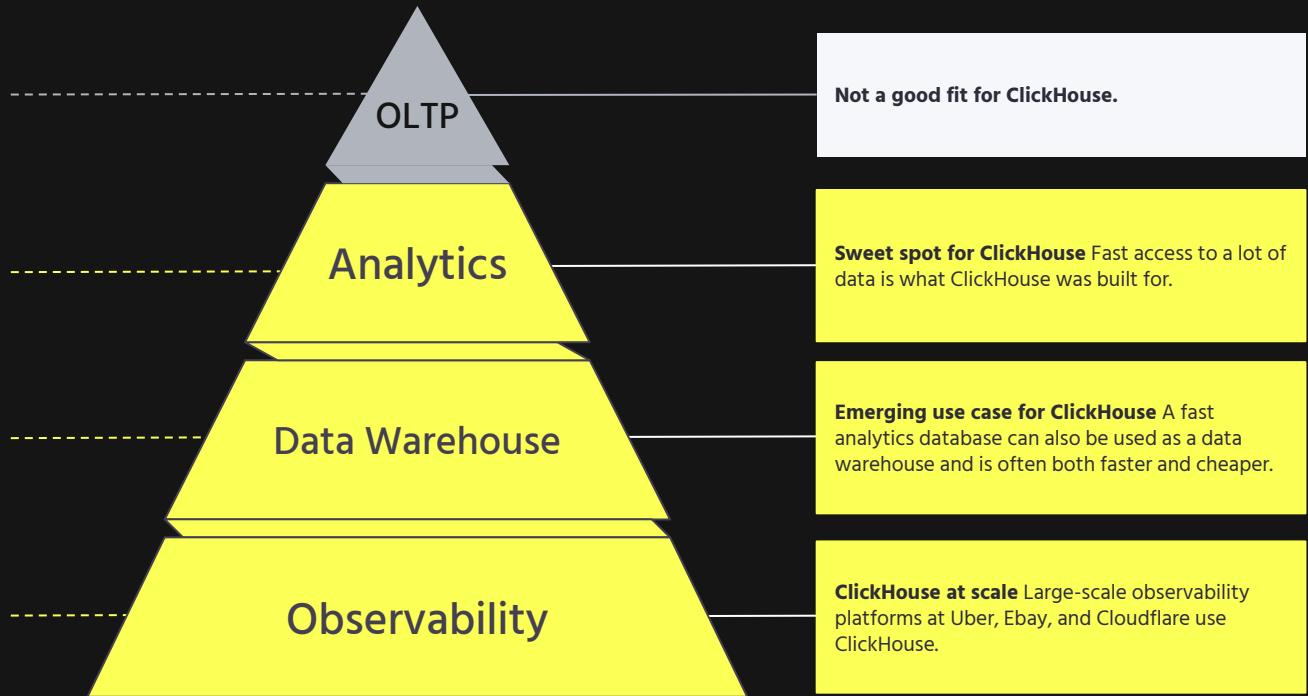
Hierarchy of Data

This is where your most important data lives. Customer names and addresses, payment info, open orders, etc.

Customer facing and internal analytics: Your real-time interactive dashboards and APIs on top of customer transactions and activity, order history, etc.

Where all business data goes, anything that is useful today or maybe will be at some point in the future: All orders, transactions, random datasets, etc.

The “digital exhaust” of all your applications and infrastructure: Logs, metrics, traces. Used for real-time monitoring and debugging.



Who uses ClickHouse?



Bloomberg



coinbase



SONY

ROKT

JUSPAY

fastly

Shopee

LangChain

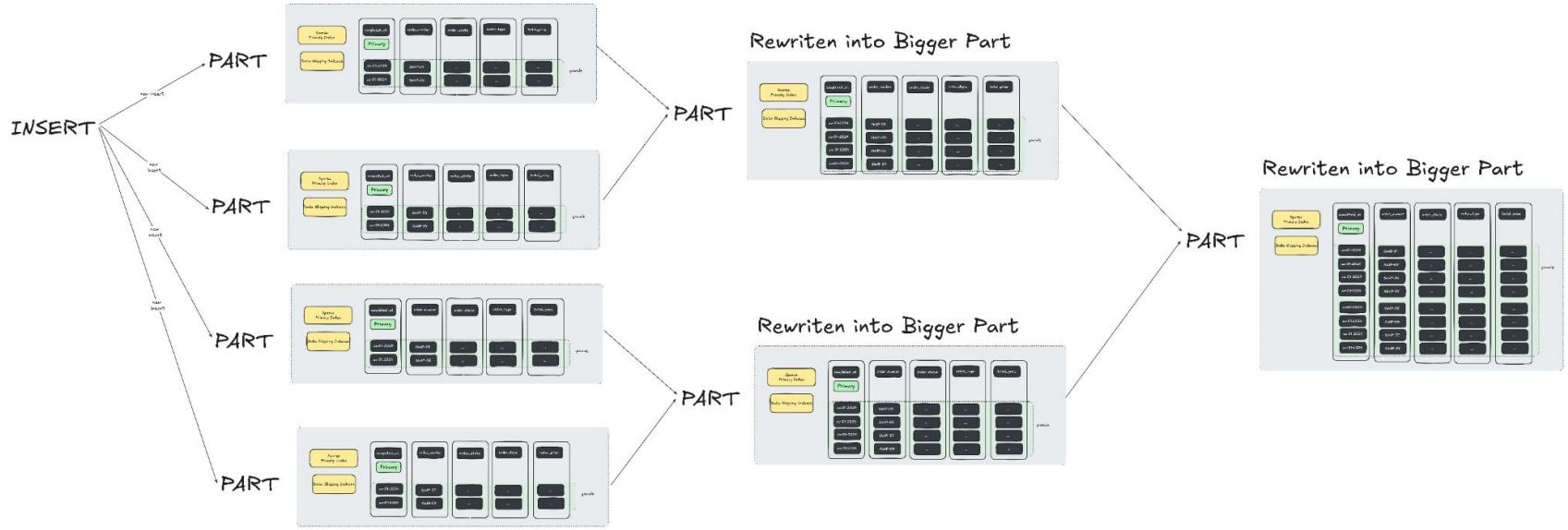
Nansen

zomato

... & thousands more

What Makes ClickHouse Unique?

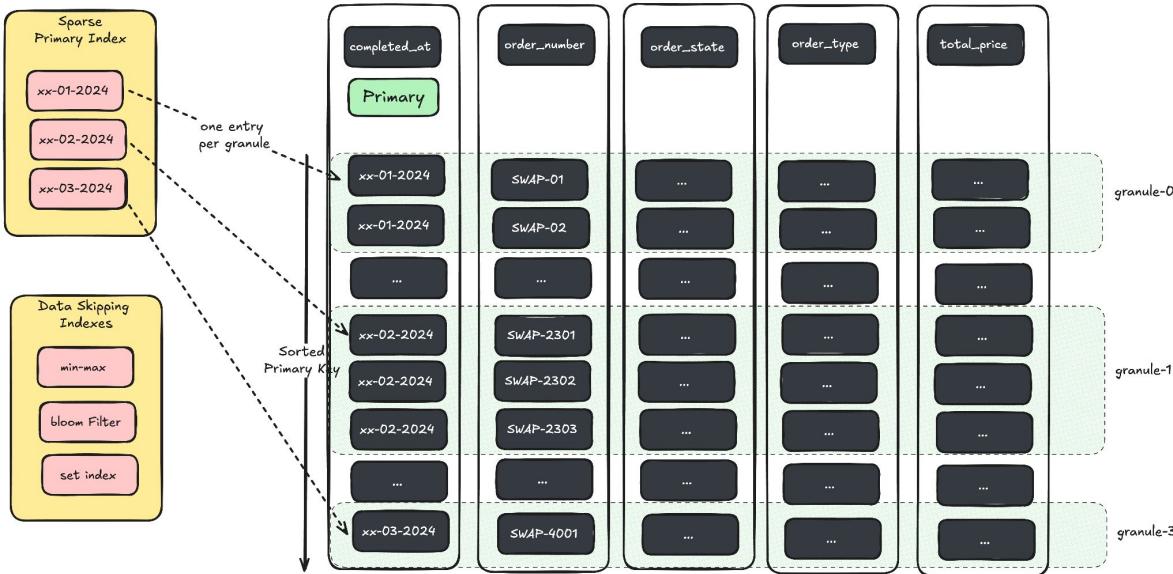
Why: Efficient Merging



Parts: Each time data is inserted, a new part is created. Multiple parts can accumulate over time, which can slow down query performance.

To improve performance and reduce the number of parts, **ClickHouse** periodically runs merging operations.

Why: Multiple Layers of Indexes



- **Granules** are chunks of rows, and **ClickHouse** groups rows into granules based on the index_granularity setting.
- **Sparse Primary Index:** **ClickHouse** uses the sparse primary index to quickly jump to the relevant granules, skipping over large portions of the data that don't need to be read.
- **Skip Indexes:** These allow **ClickHouse** to skip entire granules if it can determine from the index that no rows in the granule satisfy the query's filter.

AI/ML with ClickHouse

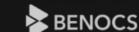
||||· ClickHouse

Use cases



Logs, events, traces

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



zomato

ebay

resmo

runreveal



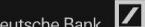
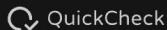
Real-time Analytics

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.



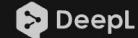
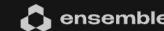
Business intelligence

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.



ClickHouse Cloud: powering enterprise AI/ML at scale

Build an unified data platform to operationalize AI. From real-time data ingestion to high-performance inference and robust observability, ClickHouse Cloud seamlessly delivers data and insights to both users and advanced AI workflows.



Data Infrastructure

Massive ingestion

Ingest billions of events per second using columnar storage and compression.

Petabyte scale

Handle high-throughput data without sacrificing query speed.

Native JSON support

Ingest and query semi-structured data at speed and scale.

Cost-efficiency

Optimize compute and storage to reduce overall spend.



Training and inference

Vector search

Natively store and query high-dimensional embeddings for Semantic Search and RAG.

Feature store

Accelerate model training and reduce development cycles.

Agent support

Power AI Agents efficiently with real-time data.

High-speed inference

Sub-second retrieval for real-time AI applications.



Operations

Live logging and tracing

Pinpoint performance issues instantly.

Granular metrics

Monitor and debug AI models at scale.

Continuous optimization

Ensure reliability and fast resolution of anomalies.



Integrations and delivery

Rich APIs and SDKs Seamless integration into existing AI/ML workflows & data pipelines.

Real-time dashboards

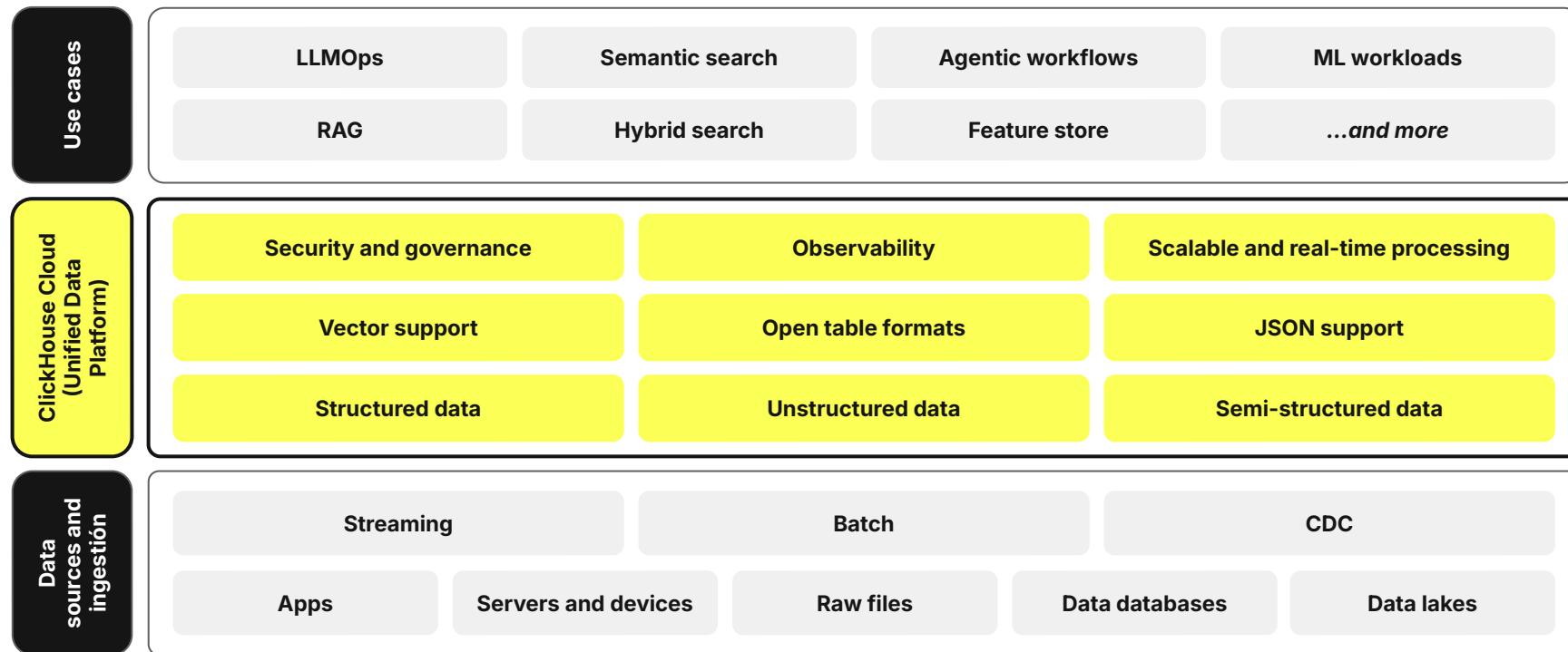
Transform analytics into business intelligence instantly.

Third-party connectors

Support for popular BI, data transformation, and AI frameworks.

Unified Data Platform for AI/ML services

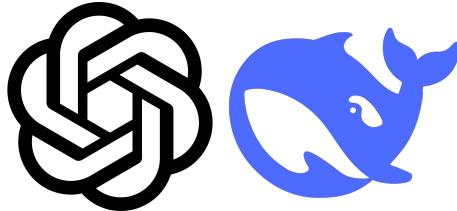
From real-time ingestion to advanced AI/ML outcomes with ClickHouse



Scaling GenAI: The Evolution of Models, Infrastructure, and Data

Foundation models are bigger, inference is faster, and data must keep pace.

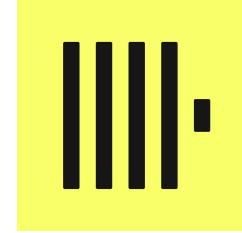
Foundation Models



Training & Inference Infrastructure



Data Infrastructure



Massive Growth: Model sizes ballooning to trillions of parameters

Specialized Domains: Vertical-specific LLMs for precise outcomes

On-Demand Results: Low latency for seamless user experiences

Accelerated Compute: GPUs, LPUs, specialized chips for parallel processing

Fast Token Generation: Minimizing inference time for real-time AI apps

Ultra-Low Latency: Users expect near-instant results

Scale Without Limits: Manage petabytes of data with ease

Continuous Streaming: Real-time ingestion to keep models updated

Instant Queries: Millisecond-latency analytics for interactive AI



Adopters



Weights & Biases

Langtrace

The braintrust logo is set against a purple background featuring faint, abstract brain-like patterns.

helicone



Langfuse



Autoblocks

...and more!



We're expanding our AI/ML Coverage



Engage and learn with us and the community!

CHC Academy



**CHC
Community**



Thank you!