Verihubs

Supercharging Verihubs Data Warehousing with Clickhouse

ClickHouse Meetup

December 2025

About Verihubs

Verihubs is a Saas (software as-a-service) based company that provides a complete solution for compliance and security for companies' digital businesses in Indonesia. With Verihubs, the customer experience is smoother, more efficient operations, and well maintained data security.

Verihubs founded in 2019 on a mission to help companies reduce fraud risks through AI solution.

ISO/IEC:27001:2022 - Information Security Management System certified.
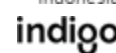
ISO/IEC:30107 - Presentation Attack Detection certified.

Meta — Officially partnered with Meta as Business Solution Partner.

Recommended by OJK as ITSK (Inovasi Teknologi Sektor Keuangan) (S-42/D.07/2024).

Processed over 50 million API/month.

NIST

Verihubs cutting edge computer vision is recognized by NIST for World-Class Technological Advancements in Computer Vision.

➔ Awarded by NIST as #1 Face Recognition technology in Indonesia in terms of performance and accuracy rate and top 100 in the world.

➔ Face Search technology is certified by NIST FRTE 1:N.

➔ One and only NIST Presentation Attack Detection Certified from Indonesia (Liveness).

INSIGNIA VENTURES PARTNERS

MDI ventures by Telkom Indonesia

wavemaker

central capital

Y Combinator

N NUMBER

fintech indonesia

Forbes

indigo

ABDI

400+ Clients have seamlessly onboarded the right customers while preventing fraud - no matter their industry.

Financial Institutions

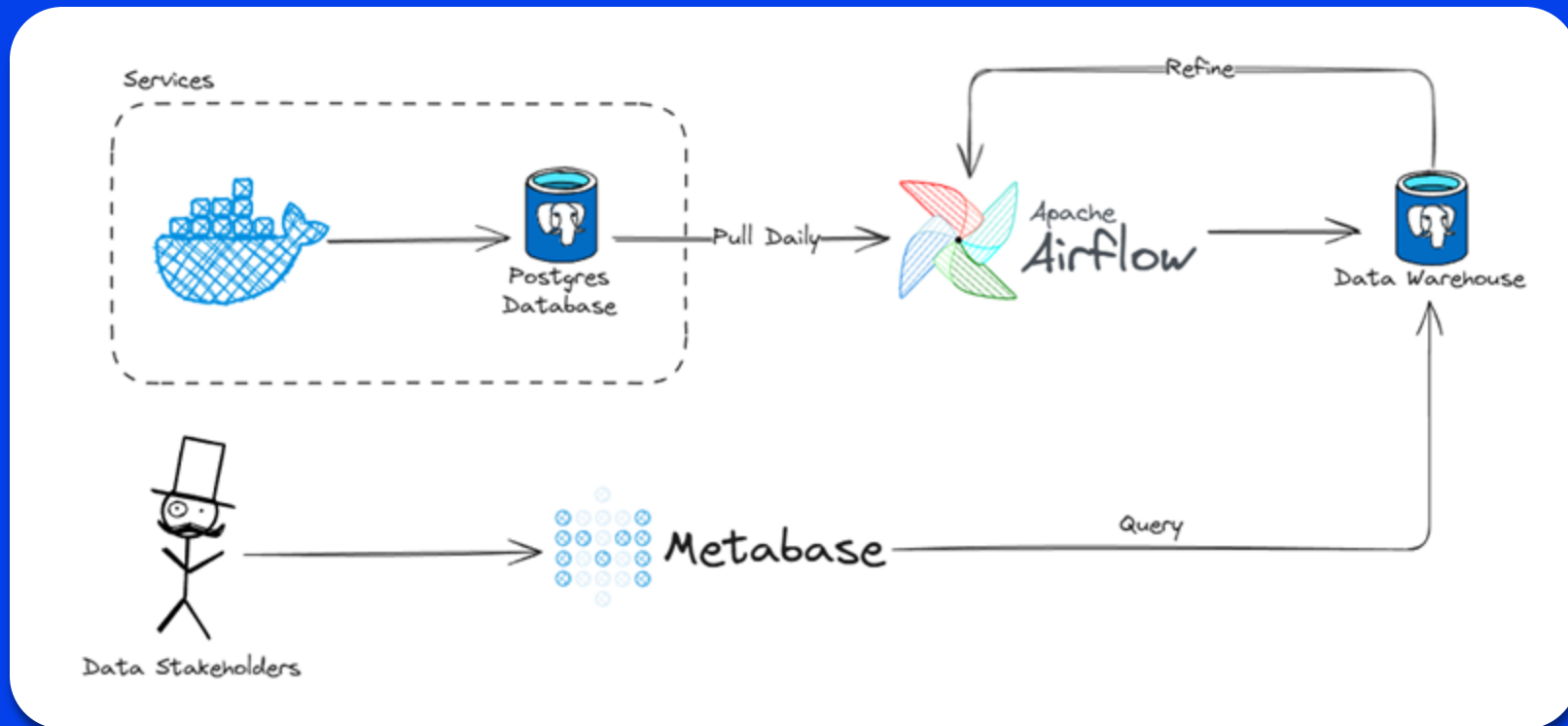

Digital App & Services

Data Stored

- Non-sensitive API Transaction Logs (Traffic & Reconciliation)
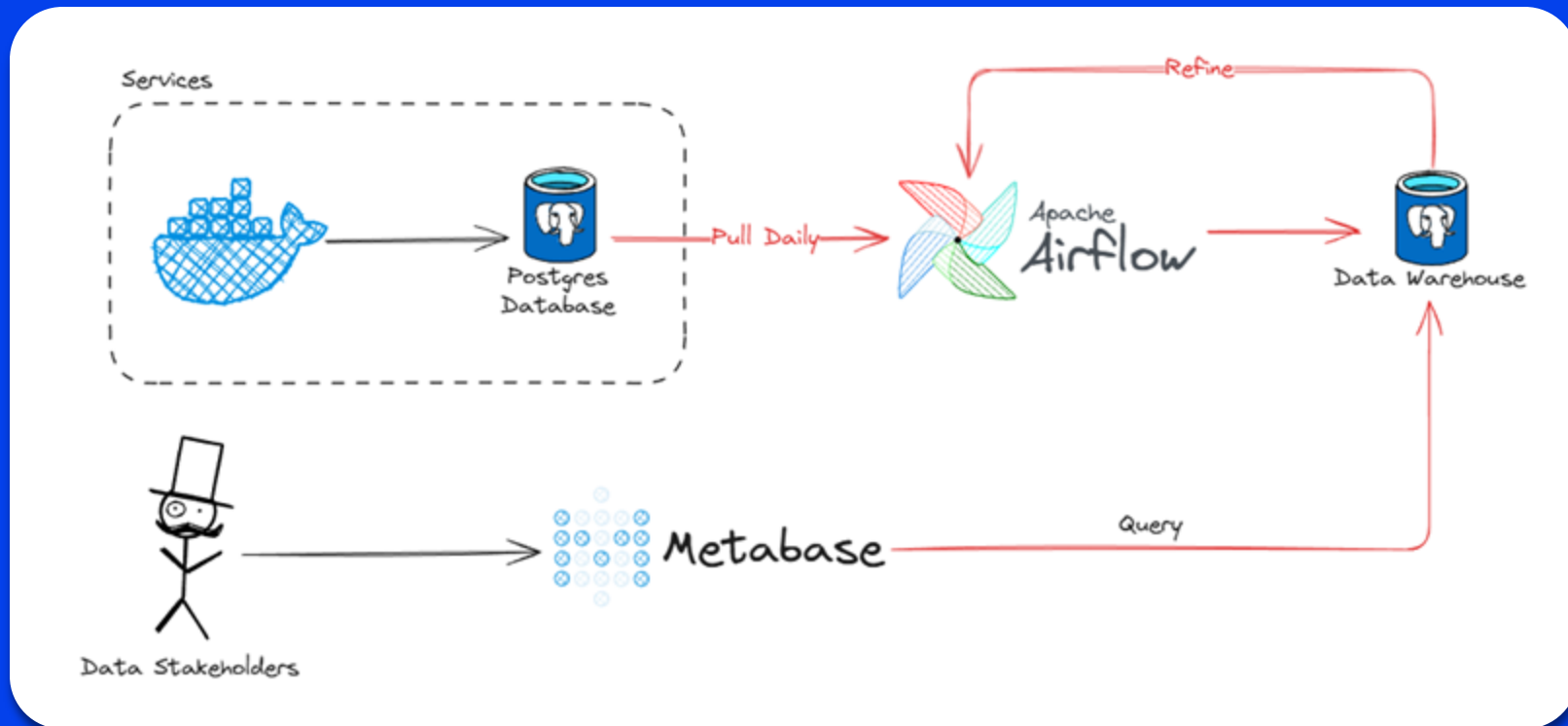- API Pricing Details, Client Company Details

Data Needed

- Traffic Logs Aggregated Daily and Monthly
- Revenue per API product
- Client Performance
- Invoicing for Clients
- Any ad-hoc business data requests

Old Data Warehouse

Services

Docker → Postgres Database → Pull Daily → Apache Airflow → Data Warehouse

Refine

Data Stakeholders → Metabase → Query → Data Warehouse

Bottlenecks

PostgreSQL is fast up until some point

Usual daily data pull
takes around **2-6 hours**
per service

Row table changes for the rows we already pulled for that day.

Created At ⌄	Updated At
November 30, 2025, 3:27 AM	December 6, 2025, 2:03 PM
November 20, 2025, 12:30 AM	December 7, 2025, 12:52 AM
November 9, 2025, 7:24 AM	December 5, 2025, 6:59 AM

ONE MUST IMAGINE LOGISTICS BOTS HAPPY
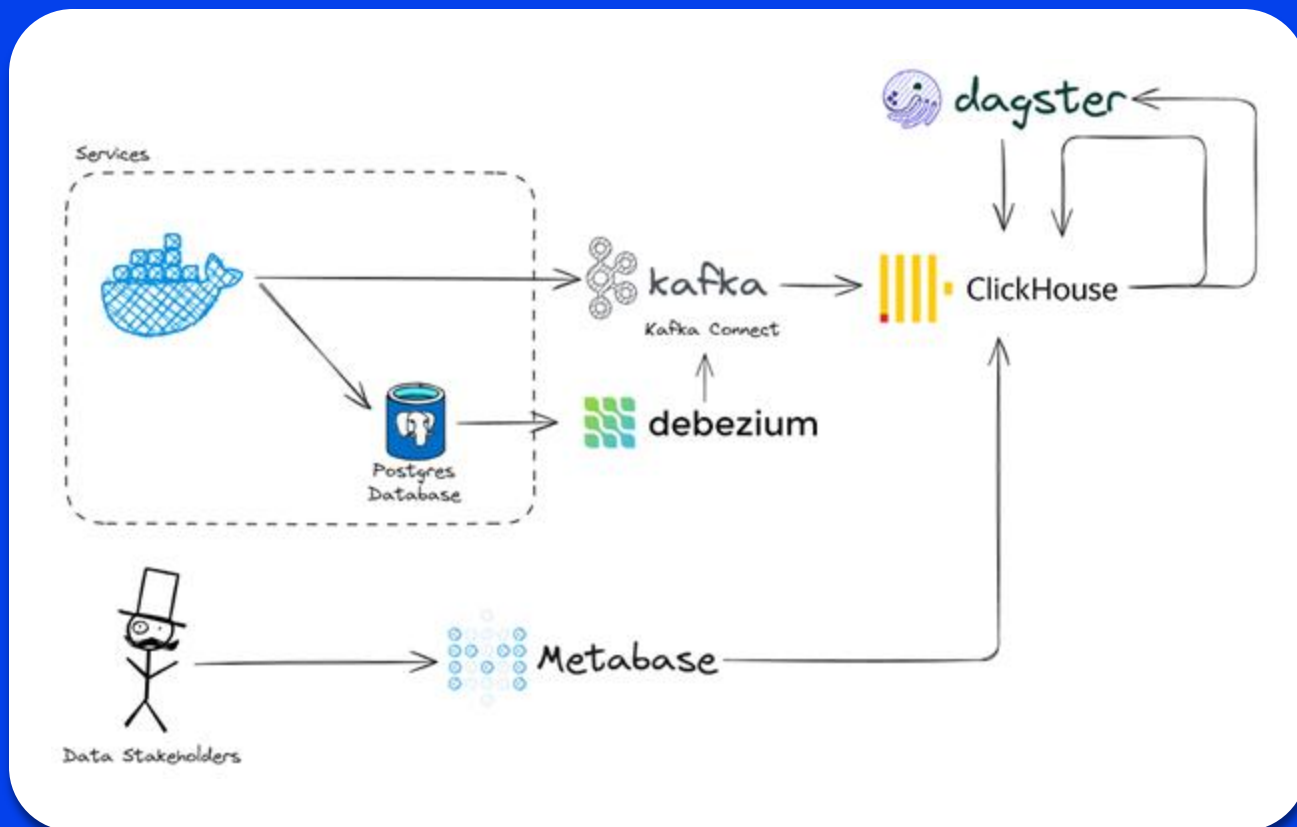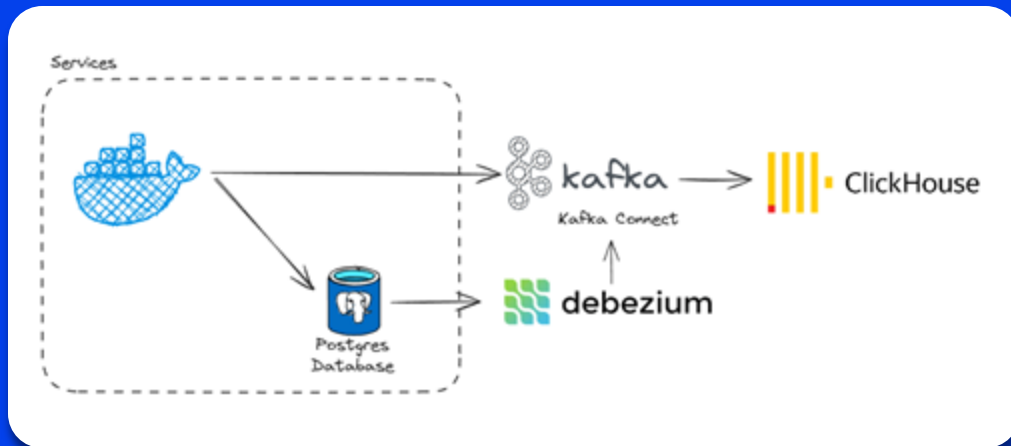
Our Goals

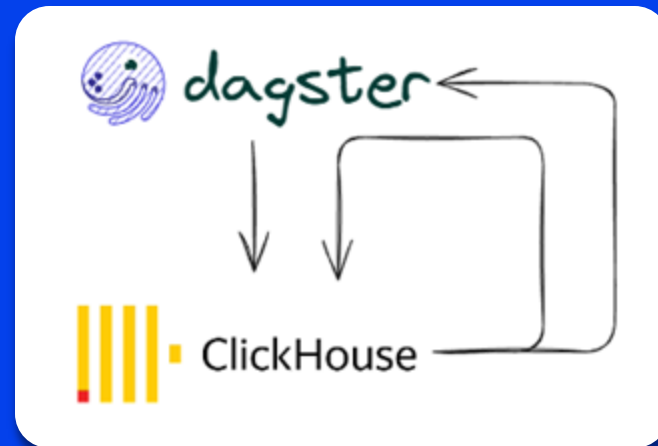Speed

Cost

User Experience

Verihubs

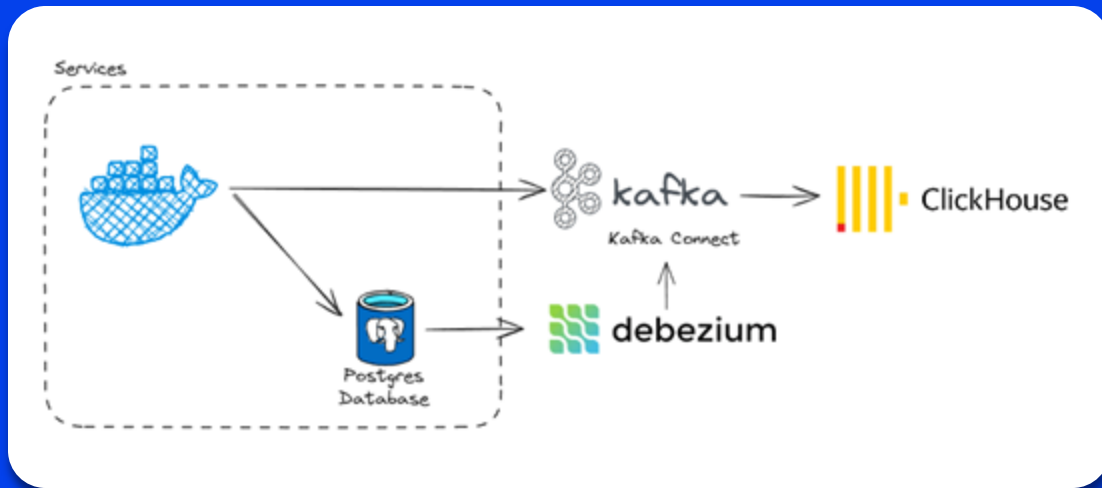# New Data Warehouse

Dissect
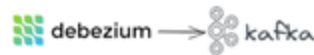

Data Ingestion

Data Post Processing

Data Ingestion

- Data is now **realtime**
- Clickhouse **MergeTree** handles data storing optimization
- Clickhouse have a simple kafka integration via ClickPipes

Data Ingestion

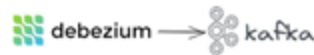**Append Only Logs**

Queued immediately into clickhouse.

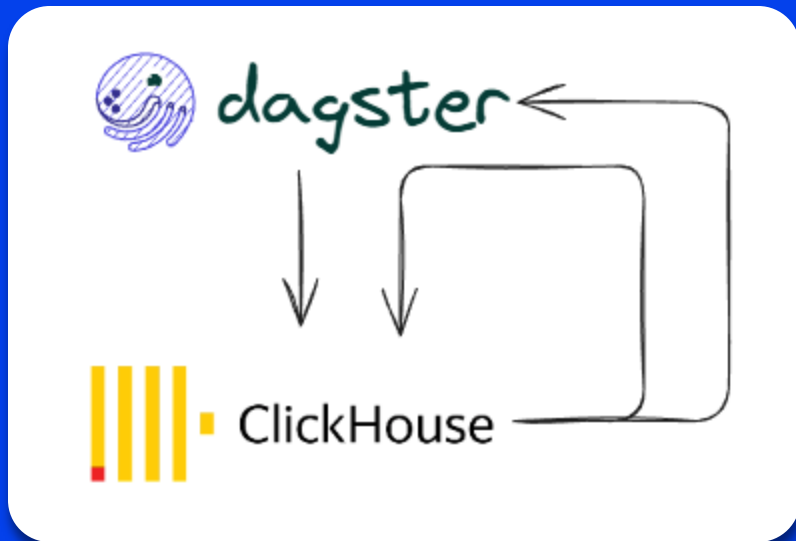**Updatable Data**

Captures every row upserts (CDC) and send the affected row.

Data Post Processing

- Data is further refined in Clickhouse through **Materialized Views & Projections**
- Dagster is used for complex data processing (e.g. tracking suspicious behavior user, data mismatches between OLTP) and alert them to our slack
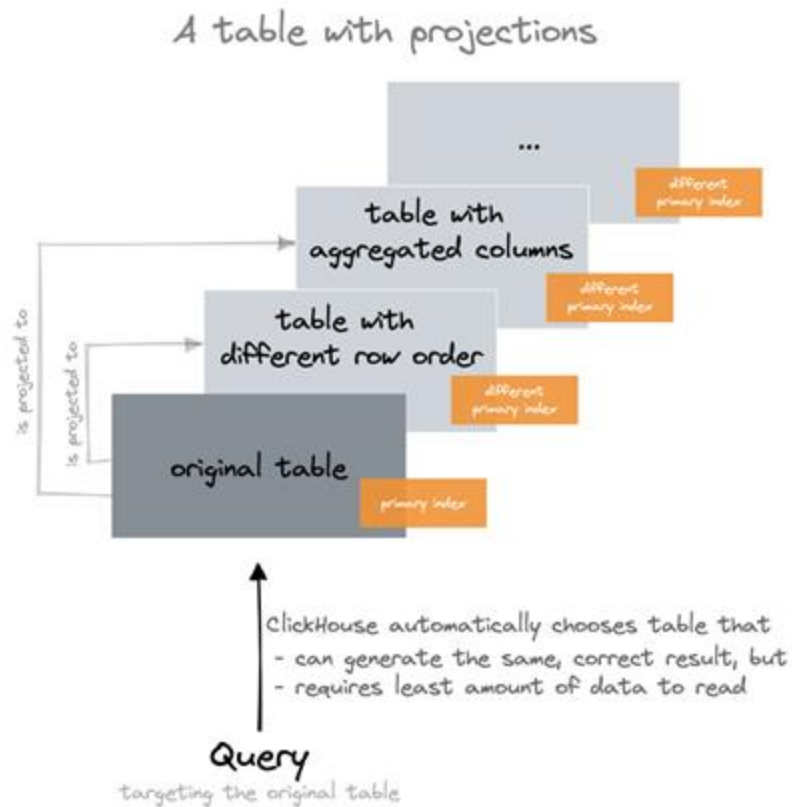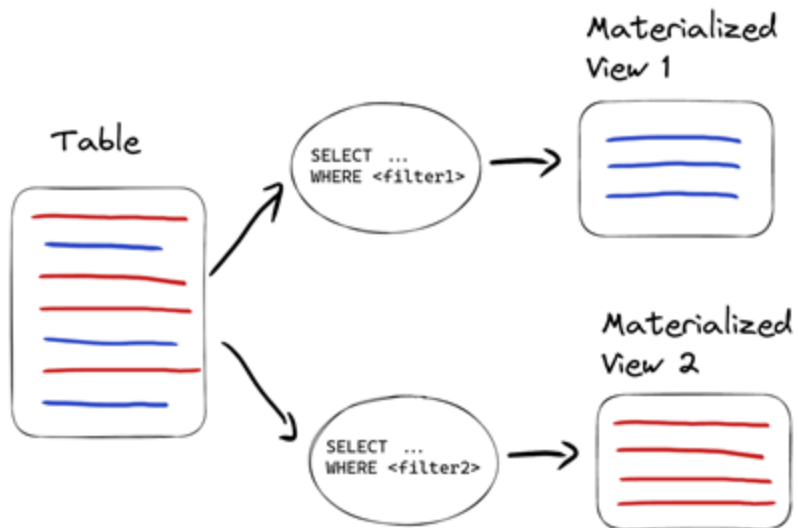
Verihubs

ClickHouse

Materialized View

- Listens to a clickhouse table, transforms, and writes into **another physical clickhouse table**.
- Good for pre-aggregating **heavy metrics**

ClickHouse

Projection

- **Optimizes data layouts** that are physically stored **alongside the main table** data and are invisible to the user.
- Good for accelerating data querying

Verihubs

Left panel:

Table

SELECT ...
WHERE <filter1>

Materialized View 1

SELECT ...
WHERE <filter2>

Materialized View 2

Right panel:

A table with projections

...

table with aggregated columns

different primary index

table with different row order

different primary index

original table

different primary index

primary index

is projected to

is projected to

ClickHouse automatically chooses table that
- can generate the same, correct result, but
- requires least amount of data to read

Query

targeting the original table

Results:

Verihubs

Elapsed: 1.260s   Read: 9,253,464 rows (433.00 MB)

Elapsed: 2.947s   Read: 18,535,783 rows (982.27 MB)

98%   Realtime   50%

Query speed improvement from minutes to seconds

Data warehouse ingestion

Cost saved from cloud expenses

Important Things to Know Before Migrating to Clickhouse

- **Your database design matters. A lot.**
    - Denormalize schemas, fewer joins, and choose your sorting keys wisely.
    - Table Engines are commitments. Different engines behaves very differently.
- **INSERTs are fast, but deletes & updates are not.**
    - Deletes & Updates are labeled as "mutations" in Clickhouse
    - Clickhouse is Eventually Consistent (your updates might propagate later).
- **Clickhouse SQL <> PostgreSQL**
    - They have similar query syntax, but some differ and Clickhouse also have their own built-in functions
- **Backfilling needs more planning**
    - Materialized views do not backfill automatically from their source tables (projections do).
    - Think before you create cascading Materialized Views.

Key Takeaways

- PostgreSQL is great until your workload outgrows it
  - Our workload couldn't scale with Postgres
- Clickhouse lets you redesign for speed while still being cost effective
- Streaming data is now possible thanks to Kafka
- Clickhouse offers many table engines to optimize your workload.
- Materialized Views and Projections do the heavy lifting inside Clickhouse.

References & Further Reading/Watching

- https://clickhouse.com/docs/engines/table-engines [Table Engines]
- https://clickhouse.com/blog/clickhouse-faster-queries-with-projections-and-primary-indexes [Projections]
- https://clickhouse.com/blog/using-materialized-views-in-clickhouse [Materialized Views]
- https://clickhouse.com/docs/managing-data/materialized-views-versus-projections [Comparing materialized view and projections]
- https://clickhouse.com/docs/sql-reference [SQL Reference]
- https://www.youtube.com/watch?v=6mmQUOmA-T0&list=PL0Z2YDlm0b3gtIdcZI3B_8bMJclDOvY8s [Clickhouse "How To" Playlist]

Verihubs

Thank you!

rayantonius
rayantonius
Rocksus