How LINE MANGA Uses ClickHouse for Real-Time Analysis
Solving Data Integration Challenges with ClickHouse

(Another way to introduce ClickHouse / ClickHouse を導入するもう一つの方法)

Kazuki Matsuda @ LINE Digital Frontier.

2025-01
ClickHouse Meetup Tokyo

LINE Digital Frontier | ebook japan

Happy New Year 2025

- Wishing great health and success for ClickHouse and everyone here!

Abstract

LINE MANGA relies on numerous MySQL servers, but faced challenges with real-time analysis. Before introducing ClickHouse, we relied on custom scripts for each analysis. This approach was difficult to develop and review, and execution was slow.

In theory, almost all such tasks can be done with simple SQL, which could be naturally parallelized by the query engine. However, due to our vertical and horizontal sharding, this method became impossible.

ClickHouse's integration engine resolves this issue. It allows data stored in different MySQL locations to be joined and aggregated with simple SQL. We believe this will serve as a helpful reference for improving the developer experience, as well as a good first step towards implementing ClickHouse.

Agenda

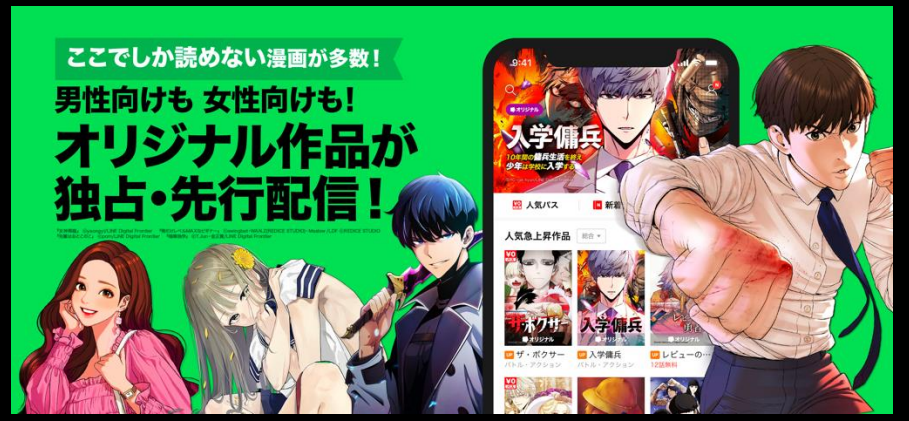
- Introduction – Who we are?
  - LINE Digital Frontier (LINE MANGA)
- The problem – Too many MySQL servers
- Approach by ClickHouse
  - ClickHouse feature: MySQL Table Engine, Merge Table.
  - Architecture for integration search and extraction.
- Conclusion & Closing

Agenda

- Introduction – Who we are?
  - LINE Digital Frontier (LINE MANGA)
- The problem – Too many MySQL servers
- Approach by ClickHouse
  - ClickHouse feature: MySQL Table Engine, Merge Table.
  - Architecture for integration search and extraction.
- Conclusion & Closing

WHO WE ARE?

- We are the team behind **Manga app (LINE MANGA) and Web (ebookjapan)**,
  - Occasionally ranks at the top of app store (Apple / Google) sales in Japan.
  - Part of WEBTOON Entertainment Inc.



Apple: https://www.youtube.com/watch?v=aRKayW4LZCg

概要 | WEBTOON Entertainment

150 カ国以上

COUNTRIES

As of Mar 31, 2024

1億 6900 万

MONTHLY ACTIVE USERS

As of the Quarter Ended Mar 31, 2024

45 万

TOTAL WEB-COMIC TITLES

As of Dec 31, 2023

5500 万

TOTAL WEB-NOVEL TITLES

As of Dec 31, 2023

2400 万人

TOTAL CREATORS

As of Dec 31, 2023

$1.3 億

LTM REVENUE

As of the Quarter Ended Mar 31, 2024

WEBTOON
Web-Comic

世界のプロ及びアマチュアクリエイターにとって主流となるWebマンガサービス（日本を除く）

LINE MANGA
Web-Comic, Digital Manga

スマートフォンやタブレットで気軽にマンガが楽しめる日本で主流となる電子コミックサービス

ebookjapan
E-Book (Novels & Comics)

マンガを中心に100万冊以上を取り扱う国内最大級の電子書籍販売サービス

NAVER SERIES
Web-Novel, Web-Comic

Web小説、Webマンガ、電子書籍のプレミアムコンテンツを提供するワン・ストップ・ストア

WEBTOON CANVAS
Web-Comic

アマチュアクリエイターが創作したWebマンガを自由に公開できる場

LINE MANGA INDIES
Web-Comic

アマチュアクリエイターが創作したWebマンガを自由に公開できる場（日本向け）

Wattpad
Web-Novel

アマチュアクリエイターが創作したWeb小説を共有できるコミュニティー

Munpia
Web-Novel

クリエイターが創作したコンテンツを持ってデビューできるWeb小説のコミュニティー及びプラットフォーム

Agenda

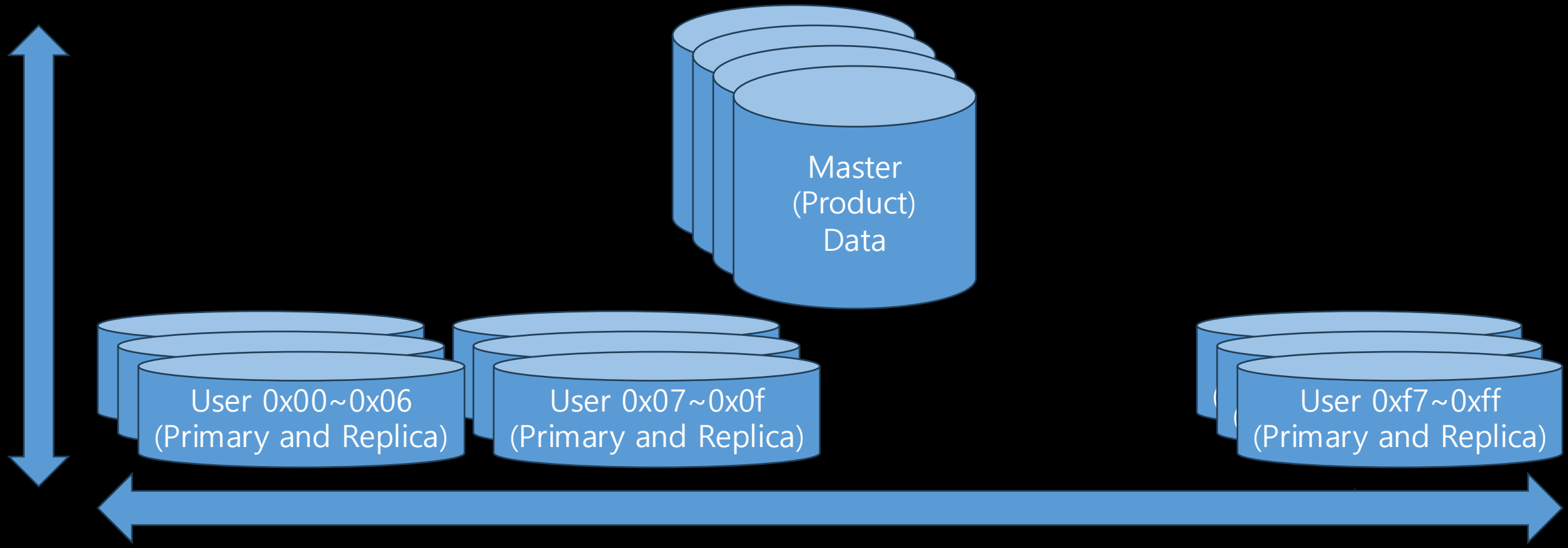
- Introduction – Who we are?
  - LINE Digital Frontier (LINE MANGA)
- The problem – Too many MySQL servers
- Approach by ClickHouse
  - ClickHouse feature: MySQL Table Engine, Merge Table.
  - Architecture for integration search and extraction.
- Conclusion & Closing

The problem: Too many MySQL servers

- LINE MANGA is over 10 years old consumer service.
- Constructed top of so many MySQL servers.

The problem: Too many MySQL servers

- Horizontally and vertically sharded.



The problem: Too many MySQL servers

Problems – Difficulty in ad-hoc analysis

- It is impossible to…
  - Join master data WITH user data.
  - Aggregate for ALL user data.
- Existing analytical platform solve some point of above issue. But
  - ETL based – not up-to-date analysis.
  - Security (privacy and auditing)
    – can't upload some data due to policy,
    or does not support financial audits.

Problems – Difficulty in ad-hoc analysis

- It is impossible to…
  - Join master data WITH user data.
  - Aggregate for ALL user data.
- Existing analytical platform solve some point of above issue. But
  - ETL based – not up-to-date analysis.
  - Security (privacy and auditing)
    – can't upload some data due to policy,
    or does not support financial audits.

Problems – Difficulty in ad-hoc analysis

- It is impossible to…
  - Join master data WITH user data.
  - Aggregate for ALL user data.
- Existing analytical platform solve some point of above issue. But
  - ETL based – not up-to-date analysis.
  - Security (privacy and auditing)
    – can’t upload some data due to policy,
    or does not support financial audits.

Problems – Difficulty in ad-hoc analysis

i.e.) Hard to maintain persistent state with security.

- Existing analytical platform solve some point of above issue. But
  - ETL based – not up-to-date analysis.
  - Security (privacy and auditing)
    – can't upload some data due to policy,
    or does not support financial audits.

Agenda

- Introduction – Who we are?
  - LINE Digital Frontier (LINE MANGA)
- The problem – Too many MySQL servers
- Approach by ClickHouse
  - ClickHouse feature: MySQL Table Engine, Merge Table.
  - Architecture for integration search and extraction.
- Conclusion & Closing

Approach

- Does not have state. (Data)
- Just reference MySQL on-the-fly.

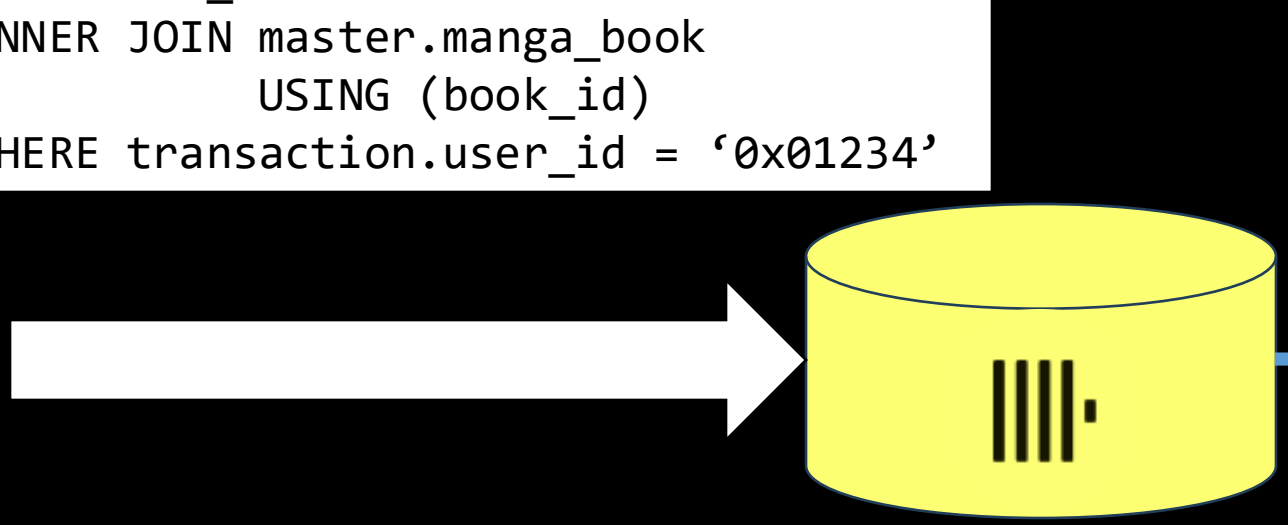
Approach - ClickHouse Features (1/2)

- ClickHouse has some (virtual) engine for integration.
  - e.g.) “MySQL Table Engine”, “MySQL Database Engine”
  - Query to MySQL virtual table on ClickHouse triggers ClickHouse to MySQL query on the fly.
  - ClickHouse can apply arbitrary operation on retrieved data. Including join and aggregation function.

https://clickhouse.com/docs/en/engines/table-engines/integrations/mysql

Approach - ClickHouse Features (1/2)

```
SELECT *
FROM user_00.transaction
INNER JOIN master.manga_book
           USING (book_id)
WHERE transaction.user_id = ‘0x01234’
```



```
SELECT *
FROM manga_book
```

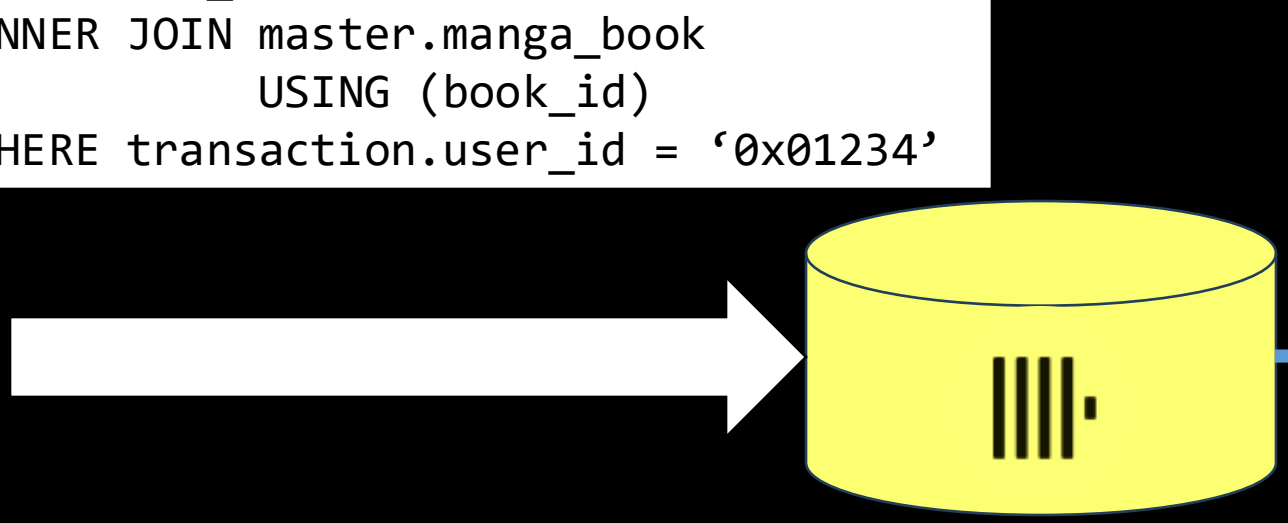
Master
(Product)
Data

```
SELECT *
FROM transaction
WHERE transaction.user_id = ‘0x01234’
-- Condition pushdown.
```

User 0x00~0x06
(Primary and Replica)

Approach - ClickHouse Features (1/2)

```
SELECT *
FROM user_00.transaction
INNER JOIN master.manga_book
           USING (book_id)
WHERE transaction.user_id = ‘0x01234’
```



```
SELECT *
FROM manga_book
```

Master
(Product)
Data

```
SELECT *
FROM transaction
WHERE transaction.user_id = ‘0x01234’
-- Condition pushdown.
```

User 0x00~0x06
(Primary and Replica)

Vertical sharding issue solved!

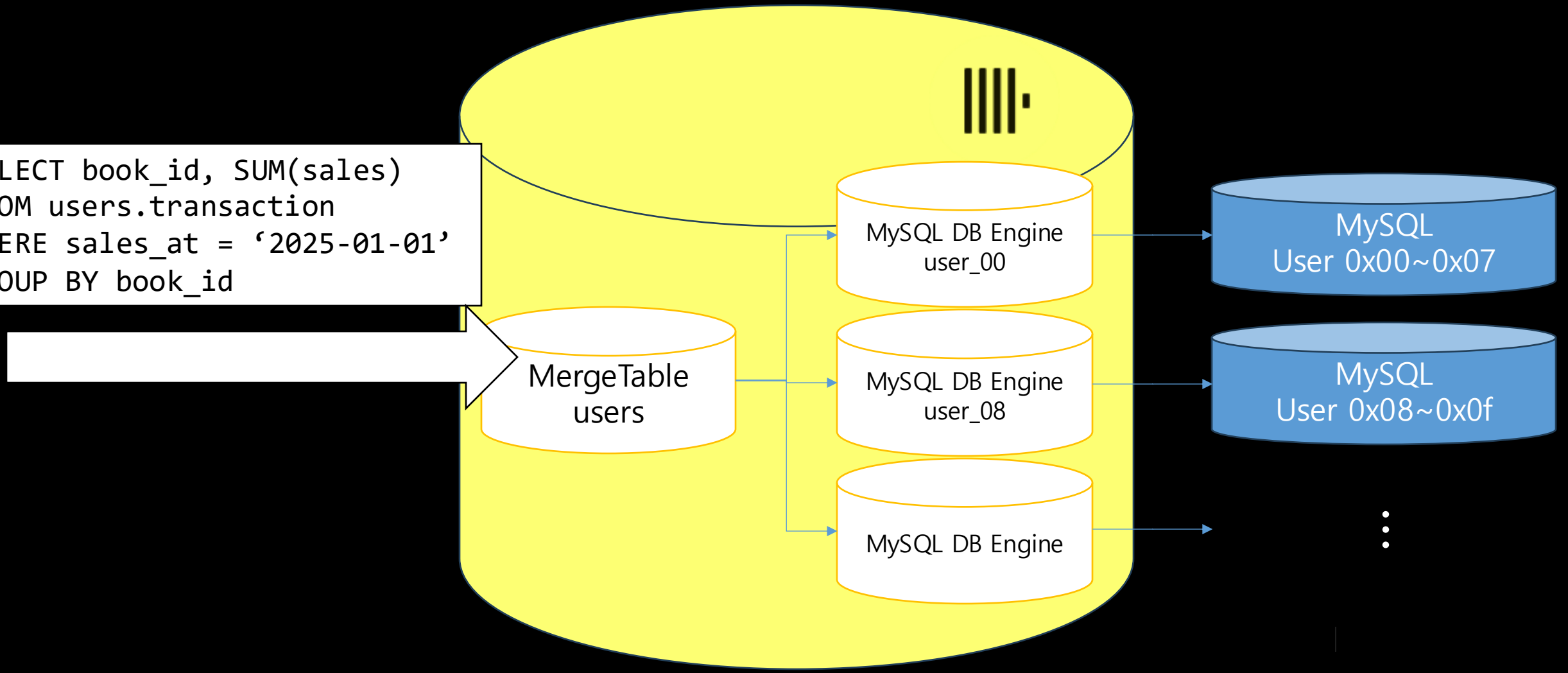
Approach - ClickHouse Features (2/2)

- MergeTable Engine
  - Can aggregate data from multiple horizontally sharded tables.
- Like a UNION View.

https://clickhouse.com/docs/en/engines/table-engines/special/merge

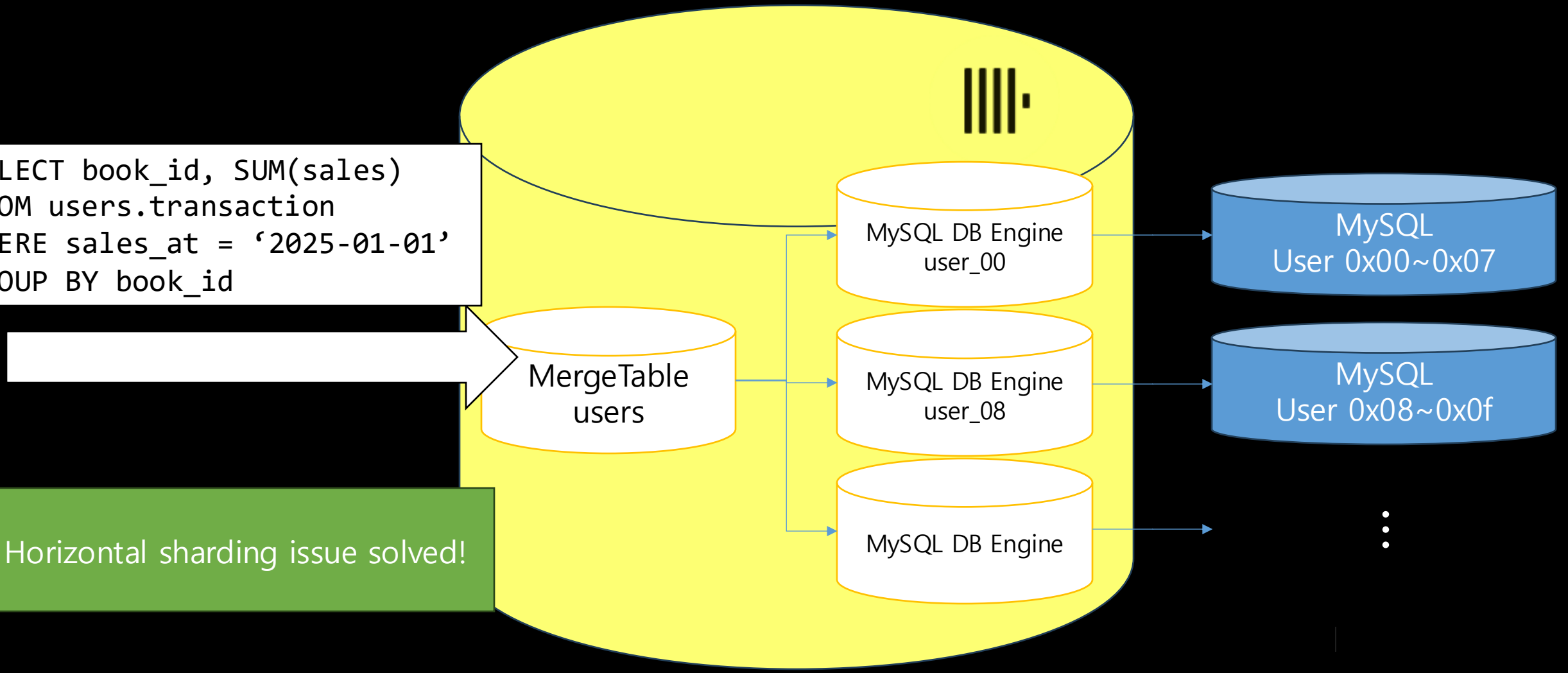
Approach - ClickHouse Features (2/2)

```
SELECT book_id, SUM(sales)
FROM users.transaction
WHERE sales_at = '2025-01-01'
GROUP BY book_id
```



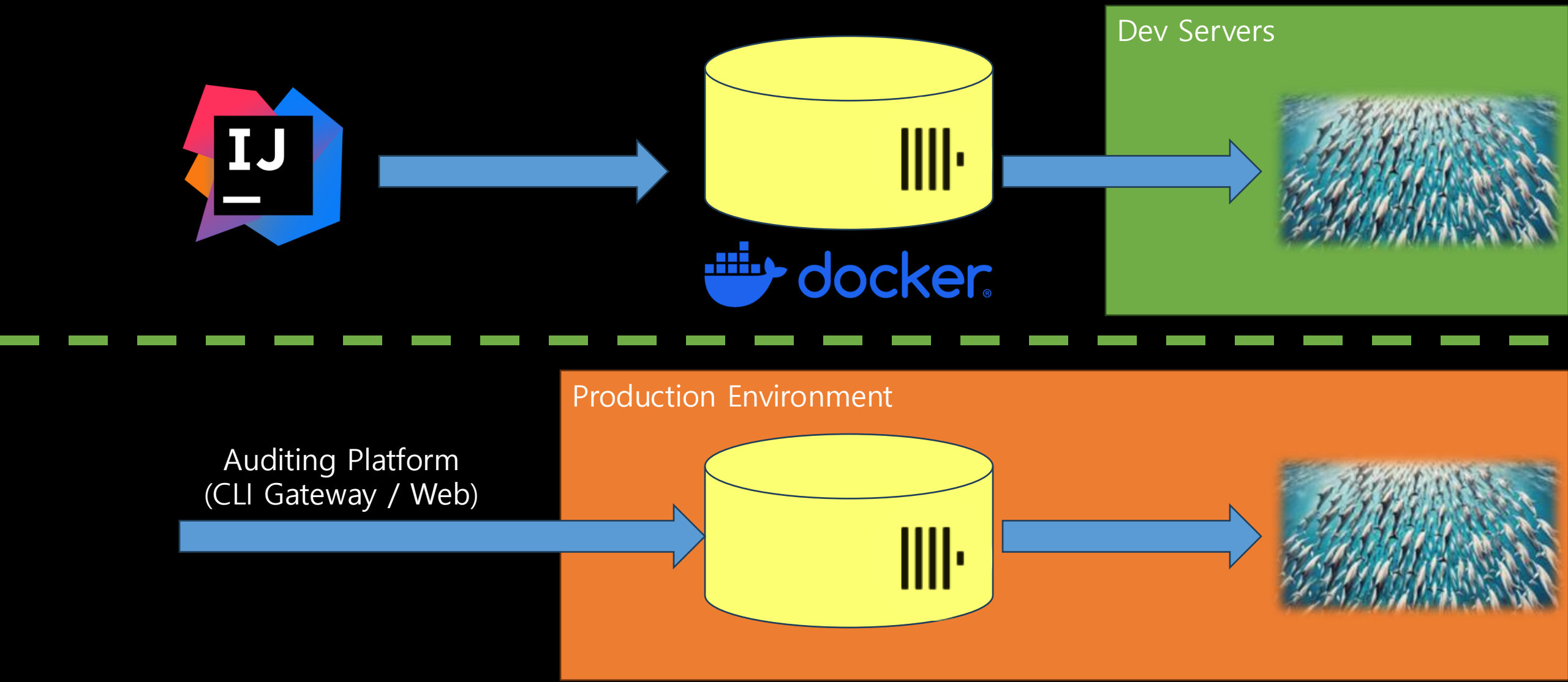
Approach - ClickHouse Features (2/2)

```
SELECT book_id, SUM(sales)
FROM users.transaction
WHERE sales_at = '2025-01-01'
GROUP BY book_id
```



Horizontal sharding issue solved!

Architecture

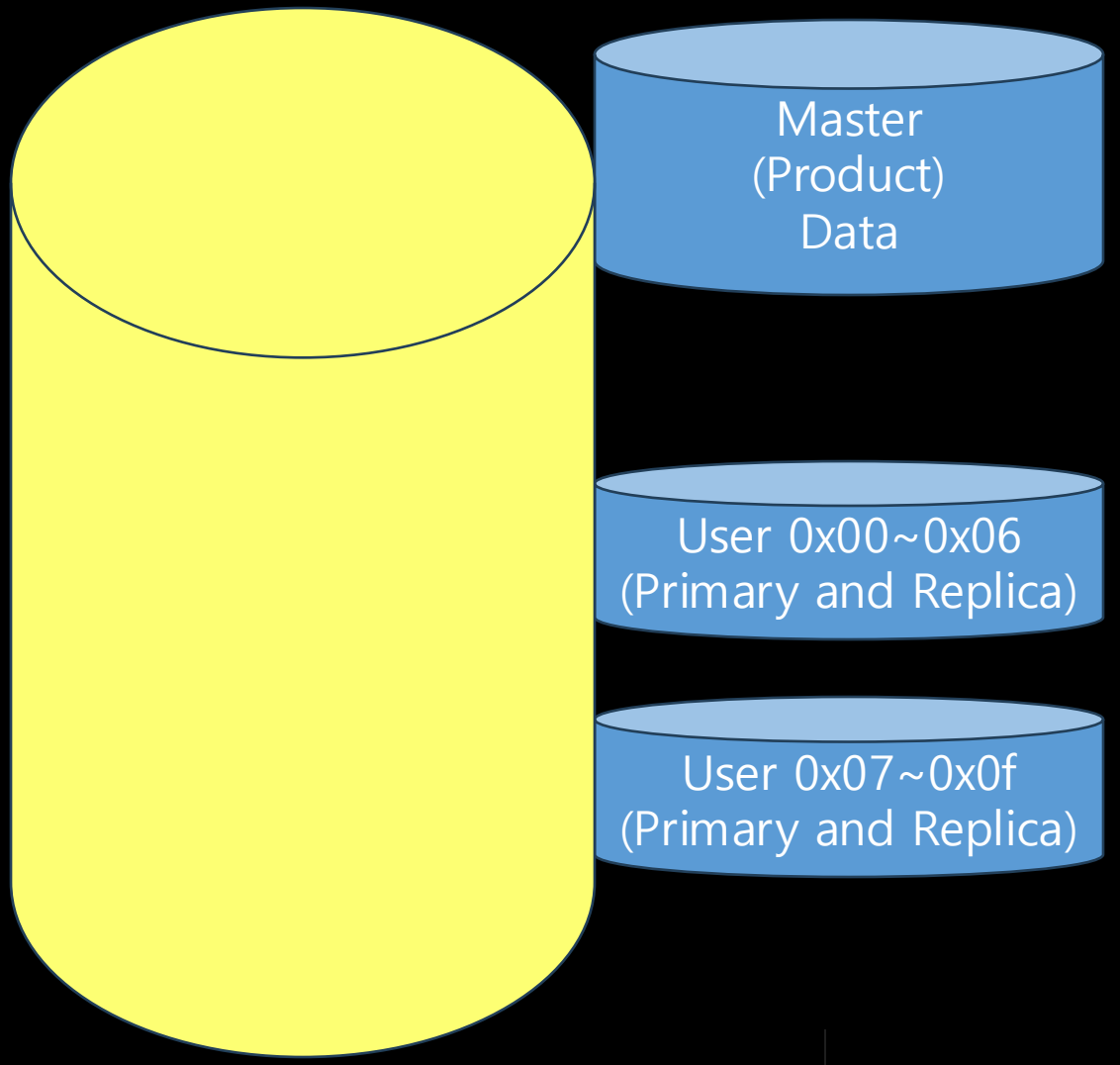


Demo

JOIN/Aggregate distributed My SQL instances.

- Vertical Sharding
  - Master data MySQL instance
  - User data MySQL instances
    - 2 horizontally shard.

Just a simple SQL.



Closing

- In closing, the Integration Engine can be very useful even if you don't store data natively in ClickHouse.
- Can join multi MySQL servers, other DBs, many formats in local disk and over http.
- Can try with single binary called clickhouse-local. Let’s try.

https://clickhouse.com/docs/en/operations/utilities/clickhouse-local

Thank you!

- For your attention and time today.
- And to my colleagues who contributed ideas and feedback that formed the basis of today’s presentation.
  - Special thanks to Okada-san, who introduced me to ClickHouse. Unfortunately, he couldn’t join us today, but he’s one of the key people driving large-scale adoption of ClickHouse.
  - And to everyone at LINE Digital Frontier who collaborated with me to take on this idea together.

End Of doc.

Q&A

- Comparison with other solution.
  - Trino, Spark etc…

    (Query Engine which has integration connector)
    - Relatively hard to setup locally.
    - Lack of “Export to LocalFile”, “Query file as Table”