ClickHouse Singapore Meetup

April 18, 2024

Thank you to our host!

Meet the ClickHouse crew

Paul Davis

AVP Sales APAC
@ ClickHouse

Derek Chia

Senior Support Engineer
@ ClickHouse

Shiqiang Duan

Senior Software Engineer
@ ClickHouse

Johnny Mirza

Solutions Architect
@ ClickHouse

Cheryl Tuquib

Head of Field Marketing @
ClickHouse

Tech Talks

- ClickHouse and The One Billion Row Challenge
  Derek Chia, Senior Support Engineer @ ClickHouse
- Distributed Tracing in ClickHouse
  Frank Chen, Expert OLAP Engineer @ Shopee
- Hunting Non-Optimised Queries in ClickHouse
  Yohann Jardin, Lead Data Engineer and top 3 researchers of the bug bounty program
- Panel Q&A

ClickHouse

A Quick Overview

2024

ClickHouse

What is ClickHouse?

Your (soon-to-be) favorite database!

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

Since 2009	Best for aggregations	Replication	Analytics use cases
31,000+ GitHub stars	Files per column	Sharding	Aggregations
1300+ contributors	Sorting and indexing	Multi-master	Visualization
500+ releases	Background merges	Cross-region	Mostly immutable data

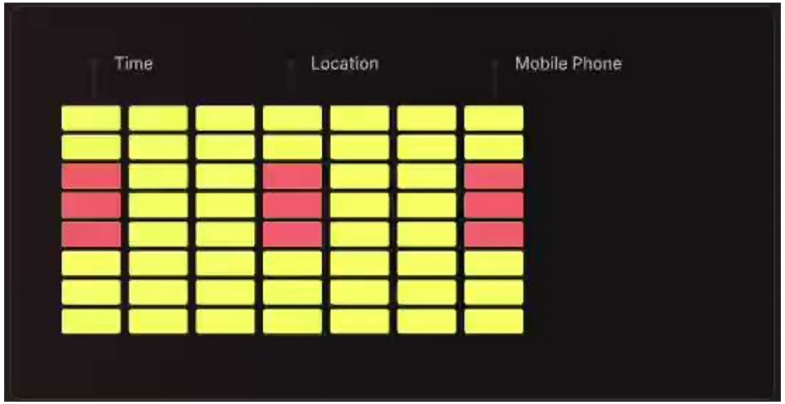
What is OLAP?

Column-oriented databases are better suited to OLAP scenarios. They are at least 100x faster in processing most queries. ClickHouse uses all available system resources to their full potential to process each analytical query as fast as possible.

Row-Oriented

Data is stored in rows, with all the values related to a row physically stored next to each other.

Column-oriented



In ClickHouse, data is stored in columns, with values from the same columns stored together.

Database for Interactive Experiences

Make any data fast

User-facing Analytics	Internal Analytics	Observability
Spotify	Walmart	CLOUDFLARE
Bloomberg	Deutsche Bank	GitLab
twilio	lyft	NETFLIX
CONTENTSQUARE	PostHog	COMCAST

Creating tables

ClickHouse SQL Basics

```
CREATE TABLE customers
(
    name      String,
    age       UInt8,
    address   Array(String),
    city      LowCardinality(String),
    created   DateTime,
    type      Enum8(...),
    attr      Map(String, Boolean)
)
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:

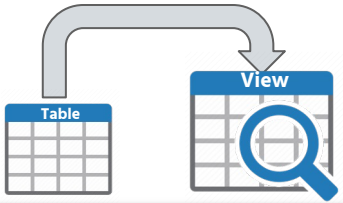
```
SET async_insert = true
INSERT INTO people VALUES ('Obi-Wan Kenobi', 57, ...)
INSERT INTO people VALUES ('Yoda', 900, ...)
INSERT INTO people VALUES (...)
```


Views

ClickHouse SQL Basics

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

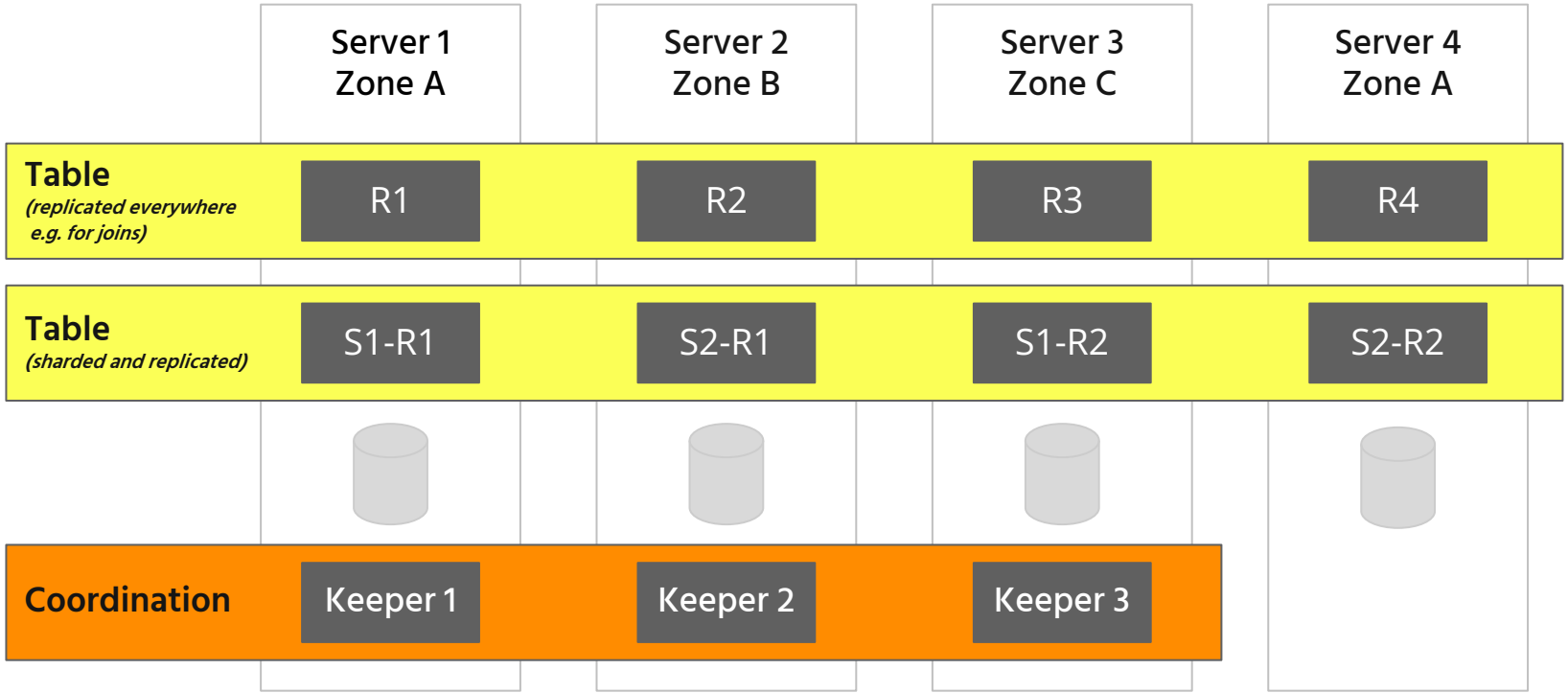
```
CREATE VIEW view AS SELECT ... FROM table ...
```



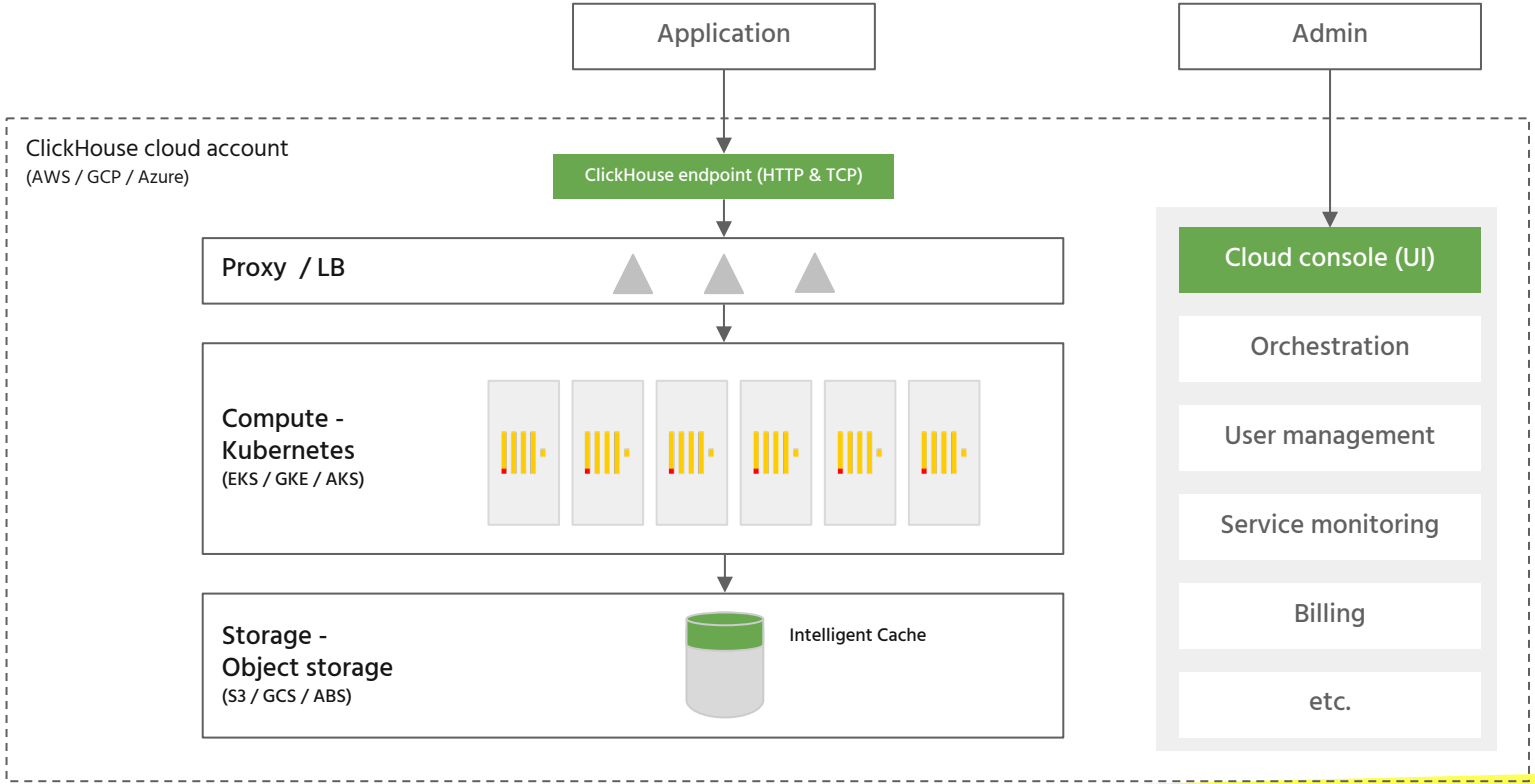
- Continuously processing new data from a table into another table:

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

Architecture



ClickHouse Cloud Architecture



Q&A

Thank you!

Keep in touch!

clickhouse.com/slack

#clickhouseDB @clickhouseinc

clickhouse

ClickHouse