ivinco

REALTIME REPLICATION FROM MYSQL TO CLICKHOUSE IN PRACTICE

Industry: Technology | **Location:** United Kingdom | **Solution:** MySQL, Replication

KEY HIGHLIGHTS

Challenges

- Difficulty to extend data retention, MySQL was hard to scale
- HA and fail-over issues
- MySQL requires many indexes, leading to dozens of indexes per table.
- Using daily MySQL tables resulted in bad cross-day queries

Requirements

- Seamless and transparent. Current system should be minimally affected
- Development resources are limited, so the less development is needed -- the better
- The core logging and data collection functionality should not be touched at all to minimize risks

Solutions

- Use ClickHouse cluster from the beginning for replication and sharing
- Use ClickHouse-MySQL for data migration

Impact / Benefits

- Seamless data migration from MySQL into ClickHouse
- Time saving for dev team on this stage

All ClickHouse power is available for analytics and monitoring

CUSTOMER STORY

Ivinco

Ivinco is a technology company whose mission is to become the technical experts to their partners, make their ideas a reality, and to help established companies by building systems that scale along with the growth of a business.

The Problem

Ivinco is working on a large scale search engine project which provides REST API for near-realtime full-text search.

The data is news articles, blog posts, social media messages, and it is acquired from multiple data sources (project partners). Data size is around 200TB in MySQL and ~25TB in Manticore Search.

Customers use REST API search service to analyze this data for marketing researches, PR campaigns effectiveness evaluation, and business awareness.

Performance data is used by Ivinco admins for service performance analysis and monitoring. For example, following typical queries may be asked:

- Number of queries per day, grouped by keys
- Number of queries per day dynamics
- Query time distribution over subsystems (PHP time, MySQL time, Manticore search time, Redis time, etc)
- System performance degradation

- How hardware and software updates impact performance in general and each subsystem personally (PHP time, MySQL time, Manticore search time, Redis time, etc)
- Timings comparison per period (day, week, month, etc)
- Failed queries dynamics and many, many others

ClickHouse Solution

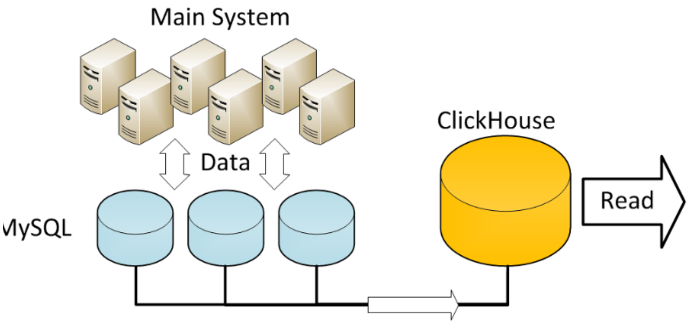
At first, Ivinco team tried the straightforward approach, installing ClickHouse as a direct replacement of MySQL and mirroring the schema all log records there, that did not work very well. Even the first approach failed Ivinco decided to change it to the following:

- Use ClickHouse cluster from the beginning for replication and sharding
- Proper Primary Key selection. ClickHouse's speed has been impressive!
- Use ClickHouse-MySQL for data migration

That approach worked perfectly!

ClickHouse MySQL

Ivinco case is not unique. It is pretty common that some old mature systems are built with MySQL. Legacy systems are often hard to modify, and developers may be busy with other projects. The initiative to migrate ClickHouse may come from DevOps team that has to deal with all MySQL problems. So whatever changes are made to the system has to be done on DevOps side with minimum involvement of core developers. The is where ClickHouse-MySQL helps. So let's see how it works in more detail.



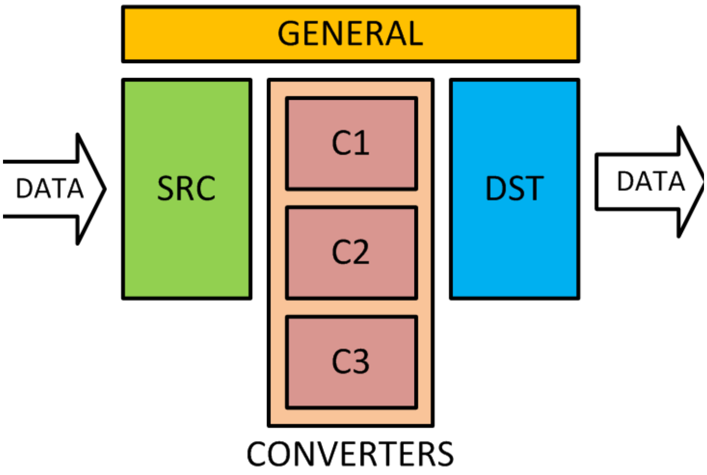
The main challenge is to move data from MySQL to ClickHouse. There are several ways to do that:

- Run something like SELECT FROM MySQL -> INSERT INTO ClickHouse. Polling as it is.
- ClickHouse provides MySQL storage engine, so it is possible to access MySQL data directly.
- Introduce replication slave for MySQL that writes to ClickHouse.

The latter approach seems to be the most convenient since it allows for maximum flexibility and customization:

- No polling at all - MySQL is active and send all the data as soon as it gets
- Awareness of other data events, such as UPDATE/DELETE, etc.
- Data transformations are possible before writing to ClickHouse

ClickHouse MySQL can be logically separated into 4 main parts: application backbone, source handlers, destination handlers, and converters that provide flexibility in configuration.



In the end, this project was a good collaboration for both Ivinco and Altinity. It demonstrated how easily ClickHouse can be used as an analytical backend for existing MySQL solution.

SOLUTION COMPONENTS

Solution Service

ClickHouse-MySQL has been developed with those ideas in mind, it is feature rich and provides the following functionality:

- Migrate existing data (bootstrapping)
- Migrate incoming data
- Schema migration and template, which helps to create ClickHouse schema based on the schema in MySQL.
- Data filters and transformers
- Plugins - in case something custom is needed, there is an option to process data stream with additional code.