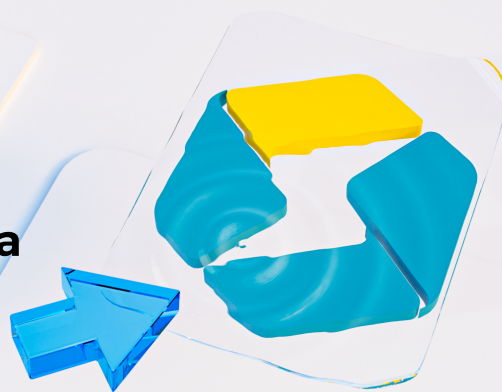




A Top Social App Boosts Query Speeds on Its Exabyte-Scale Data Lake



EBs
of data stored

50%
less compute resource

10x
better P90 query latency

About The Company

This leading social media app operates as a dynamic lifestyle platform serving over 200 million monthly users. A long-time StarRocks user, the company has more than 30,000 CPU cores deployed in production powering its real-time customer-facing workloads. This system processes over 100 billion records of data per day while maintaining 200 ms average latency. Meanwhile, their Iceberg/Hive-based data lake handles data at exabyte-scale and operates as a single source of truth open data store that serves all self-service interactive analyses across the company.

Challenges

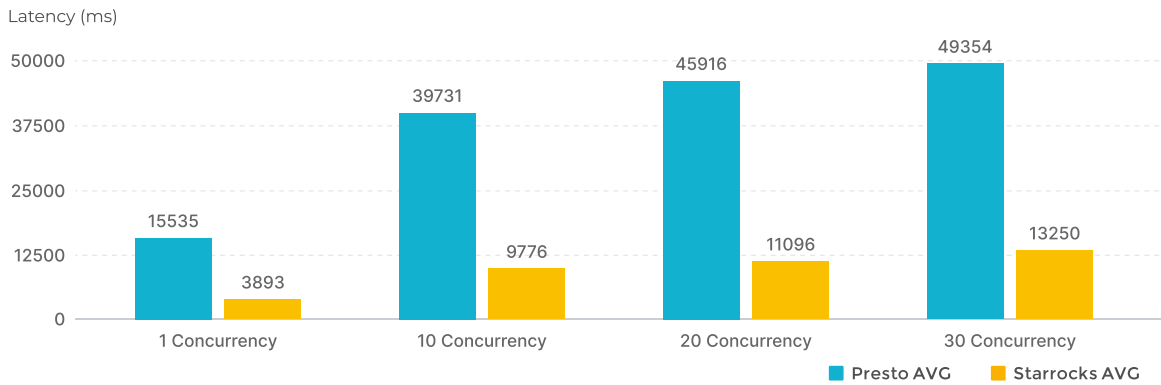
With exabytes of data stored across its 30,000+ Iceberg and Hive tables, it was impractical for the company to use a data warehouse because of the storage cost and maintaining tens of thousands of ingestion pipelines. The company initially chose Presto as its query engine to directly query from its data lake. However, as their self-service analytics workloads and data volumes have grown, Presto became a significant bottleneck to the whole data architecture for several reasons:

- **Performance Lag:** Optimizing Presto to handle growing workloads and larger data sizes posed a significant challenge.
- **Complex Architecture:** Managing both Presto and StarRocks added considerable complexity to development and operations.
- **Availability Risks:** Presto's master-slave setup introduces single-point-of-failure vulnerabilities, which lead to concerns about system stability.

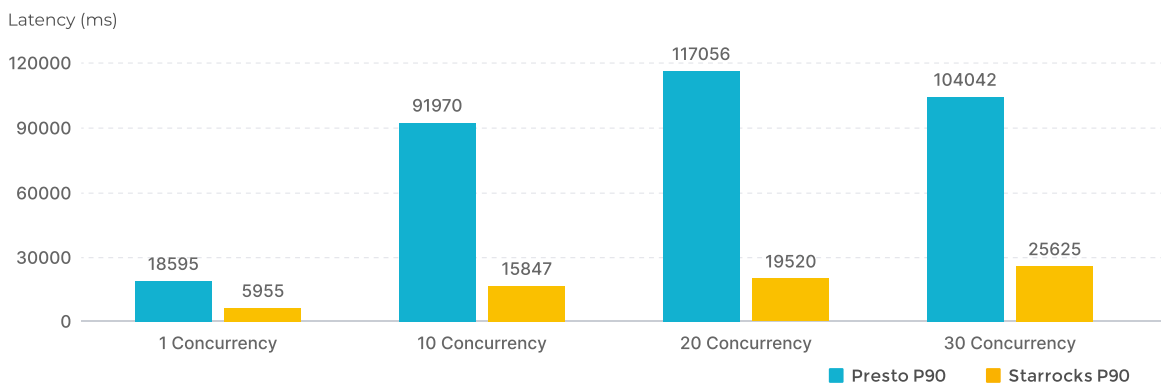
Solution

To overcome these challenges, the company began testing StarRocks as a replacement for Presto. 3000 SQL queries were randomly selected from its production queries, and tested under 1, 10, 20, and 30 concurrencies.

StarRocks (cache disabled) vs Presto (double the CPU cores) Average Latency



StarRocks (cache disabled) vs Presto (double the CPU cores) P90 Latency



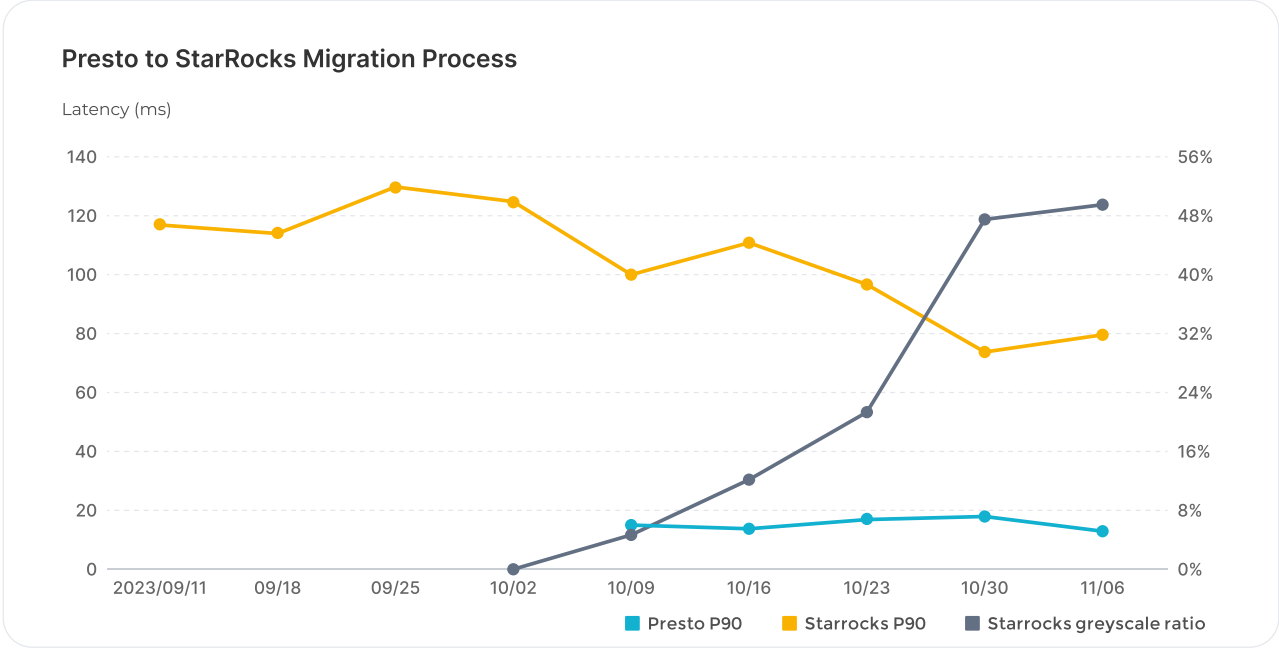
The result? StarRocks, utilizing only half of the CPU cores, showed significant performance improvements for 96% of queries. Across all levels of concurrency, StarRocks achieved an over 4x increase in query performance (and 8x in compute cost-effectiveness) over Presto.

The choice was clear after seeing the benchmark results, and StarRocks replaced Presto as the company's data lake query engine. The migration process was seamless thanks to StarRocks' compatibility with the Trino/Presto dialect that enabled out-of-the-box support for 90% of their production SQL queries.

Result

By eliminating Presto from its data architecture, the company was able to streamline its architecture in a way that not only significantly improved query performance and stability, but also reduced the compute CPU cores by 50%.

During their migration to StarRocks, the company continuously tracked the performance, stability, and accuracy of queries across both their StarRocks and Presto clusters. The company progressively expanded the use of StarRocks only after achieving predefined benchmarks. Even with this rigorous migration process, StarRocks was fully integrated into its data architecture in only 30 days.




After the migration, the company experienced an average query performance increase of 6 to 7x and a 90% decrease in p90 query latency with only half of the CPU cores Presto was utilizing. Furthermore, the migration to StarRocks was executed seamlessly, devoid of any incidents or negative feedback from users.

What's Next for the Company?

The company aims to further expand its lakehouse analytics with StarRocks in the future by:

- Continuing to contribute to the StarRocks community with optimizations for StarRocks' Iceberg data lake integration and UDF support.
- Exploring StarRocks' data cache feature to further enhance query performance.
- Testing StarRock's shared data mode and materialized view to accelerate query performance through on-demand data transformations.



Join the StarRocks Slack Community [↗](#)

Share notes, ask questions, and get feedback from thousands of your peers working at world-class companies.