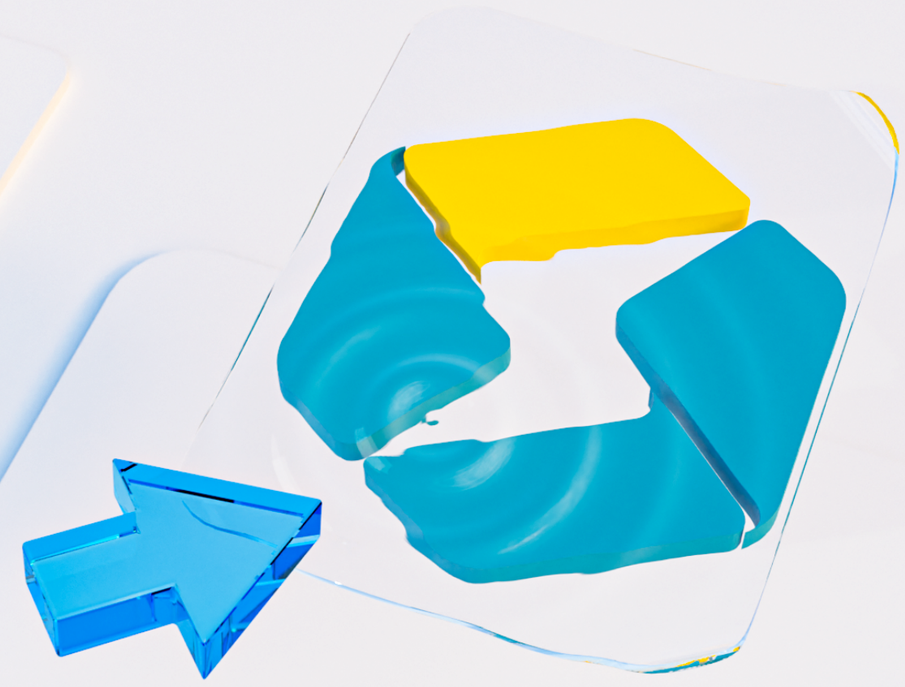
StarRocks

Trip.com Ditches Their Data Warehouse With StarRocks



7.4x
Faster than Trino querying against the same data

100%
slow queries eliminated with 4 materialized views

10x+
performance improvement over the previous solution in production

About Trip.com

Trip.com is one of the world's largest online travel agencies with over 400 million users worldwide. Artnova is trip.com's reporting platform, supporting all business units such as hotels, flights, corporate travel, vacations, marketing, train tickets, etc.

Challenges

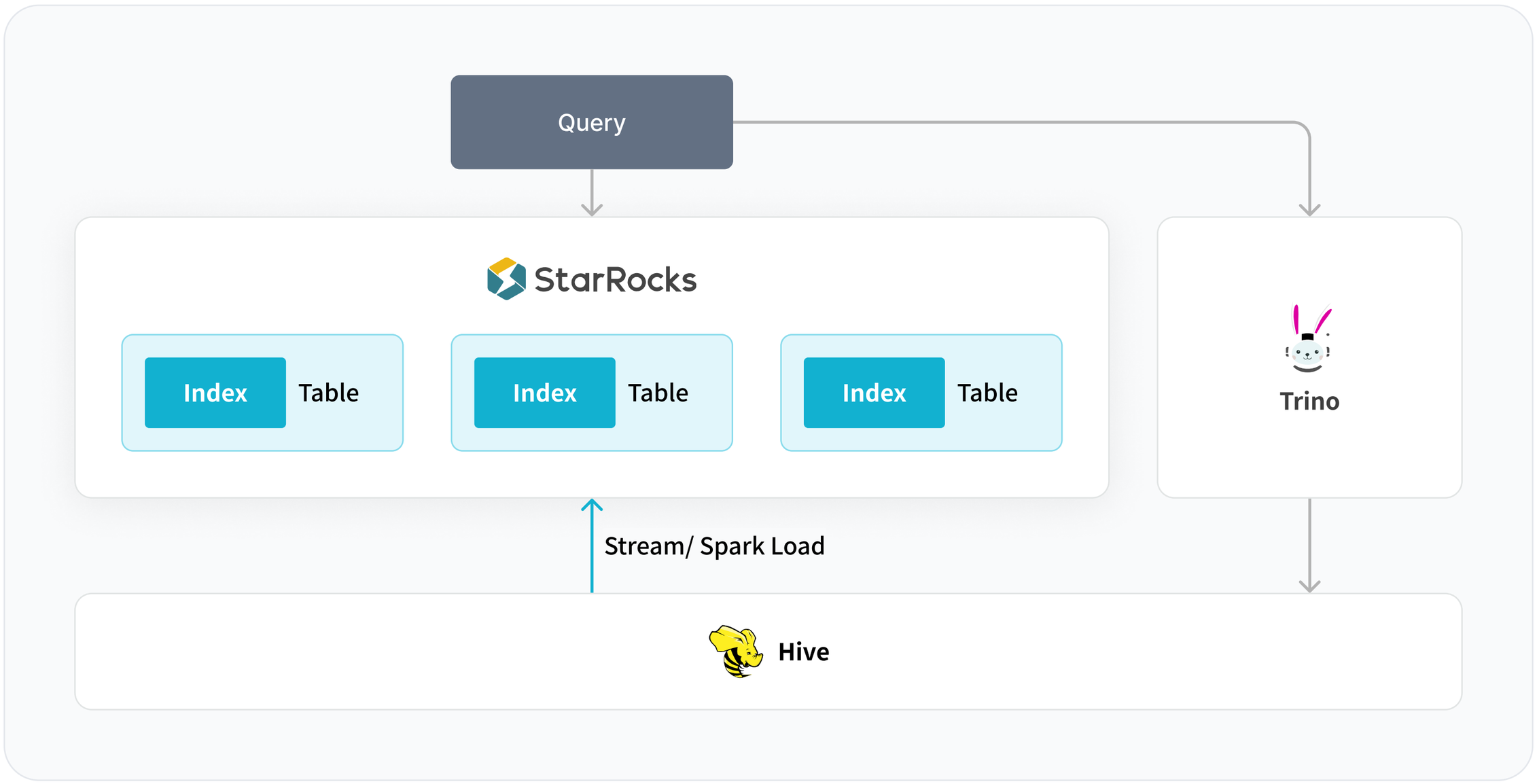


Figure 1. Old Architecture With Data Warehouse As The Query Acceleration Layer

Trip.com's Artnova platform initially used Apache Hive as its data lake with Trino as its query engine. However, due to the vast volume of data involved coupled with the need for low latency and high concurrency, Trino could not meet many of their critical use cases. To address this, Trip.com had to replicate and transfer their data into StarRocks which was serving as their high-performance data warehouse.

This approach, while solving performance issues, unfortunately introduced additional problems:

- Data freshness was still undermined despite StarRocks' ingestion being relatively fast. This affected the flexibility and timeliness of queries.
- Maintaining additional ingestion tasks and designing table schemas and indexes introduced complexity to the entire data pipeline.

Duplicating data to a proprietary data warehouse was complex and expensive, and trip.com was only able to move their most business-critical workloads to StarRocks. This was not acceptable and an architectural overhaul was necessary.

Solution

Fortunately for Trip.com, their solution was right in front of them. Not only can StarRocks persist data and work as a data warehouse, but it can also query data lake queries directly with data warehouse performance. Though their overhaul of Artnova's architecture, Trino was eliminated and replaced by StarRocks as its unified query engine.

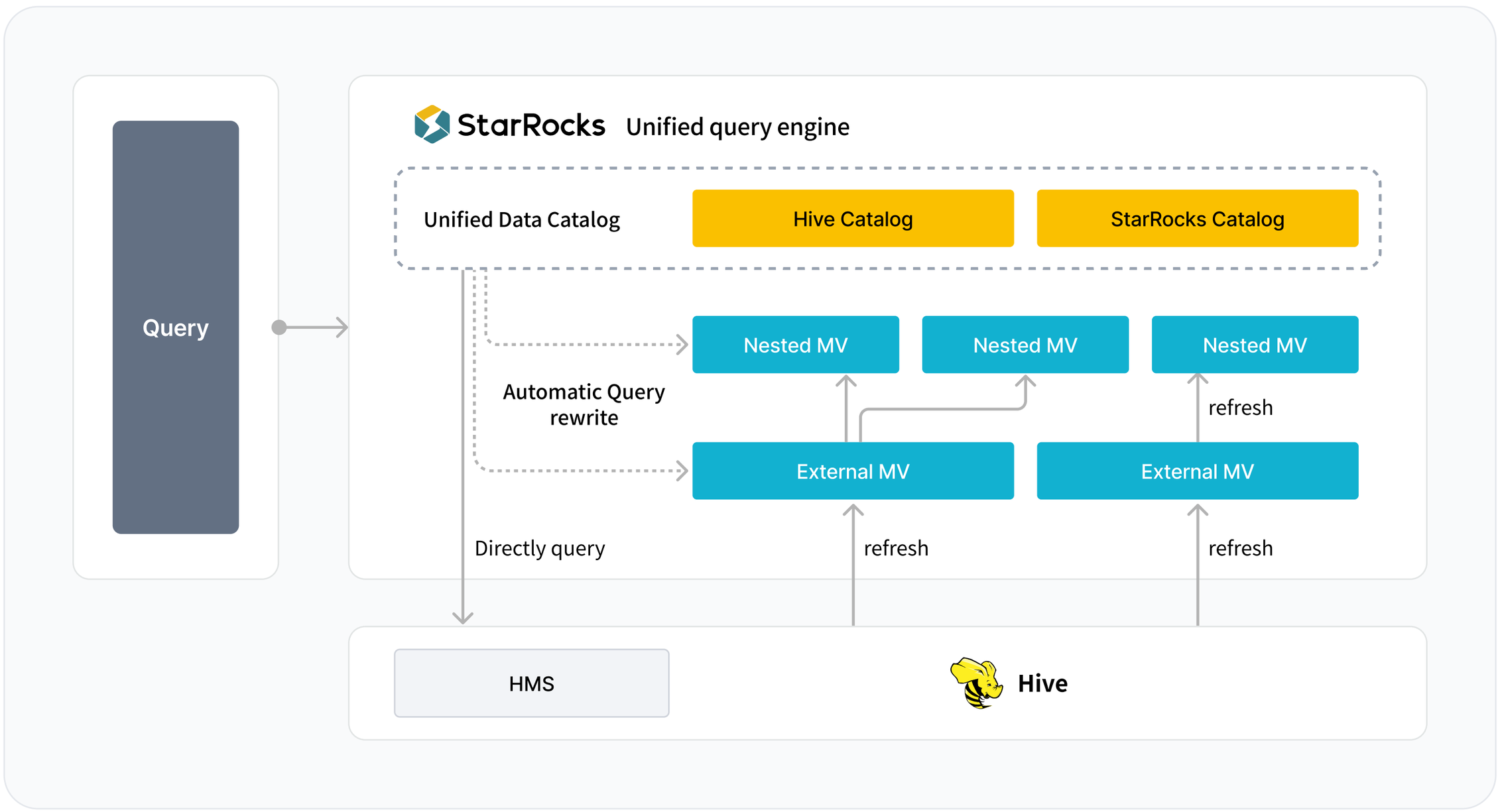


Figure 2. New Architecture With StarRocks As A Unified Query Engine

To accelerate slow queries and further enhance workloads, StarRocks' Materialized View (MV) is used to perform pre-computation including denormalization and pre-aggregation. StarRocks MV automatically rewrites queries to the most optimal materialized view,

speeding up query performance without requiring SQL modifications by users. Paired with StarRocks' superior data lake performance, Artnova's queries now default to raw data stored on the data lake, and pre-computation is only performed on demand.

Result

The shift from Trino to StarRocks didn't just translate to a performance improvement, it completely changed the way Trip.com approached querying data on the data lake:

- Ditched their data warehouse: By replacing Trino with StarRocks, Artnova simplified its architecture by ditching its data warehouse alongside its ingestion pipelines and data model designs.
- Simplified the pre-computation pipeline: On-demand pre-computation pipelines reduced Artnova engineers' workload by eliminating the unnecessary planning and development of pre-computed tables. As a result, all slow queries are solved with only four on-demand Materialized Views deployed.
- A 10x performance boost: Tests against the same data with production queries show StarRocks is 7.4x faster than Trino. With business-critical workloads and slow queries further accelerated by materialized views, the overall performance improvements add up to more than 10x.

What's Next For Trip.com

After experiencing the initial success of StarRocks as the query engine for their Artnova platform, Trip.com has plans to further expand its StarRocks usage:

- Keep exploring StarRocks' capabilities as a query engine and migrate more workloads to StarRocks.
- Explore StarRocks in batch processing scenarios such as ETL (extract, transform, load) jobs for business logic transformations.
- Continue contributing to the StarRocks project, including Apache Iceberg UPDATE support and indexes support on StarRocks Materialized Views.
- Explore StarRocks' shared data architecture.