StarRocks

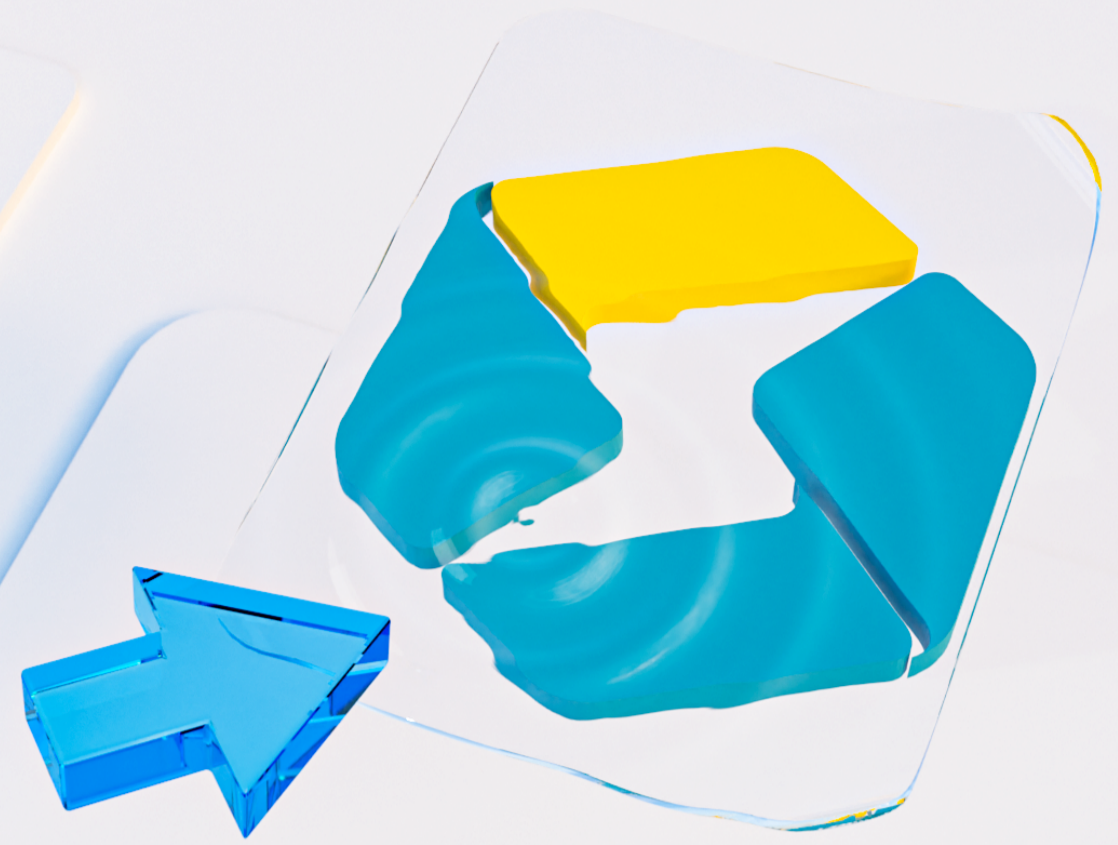# Tencent Unifies Their Gaming Analytics With StarRocks

50%
better efficiency when developing new data pipeline

15x
storage cost-saving

100%
pre-aggregations are eliminated

About Tencent Games

Tencent Games, the gaming division of the global tech giant Tencent, owns and manages some of the gaming industry's most popular franchises including the blockbuster hit "League of Legends". With 200 million worldwide users and growing, Tencent Games has established themselves as a leader in gaming analytics, developing entirely new ways for their players to engage with the games they love.
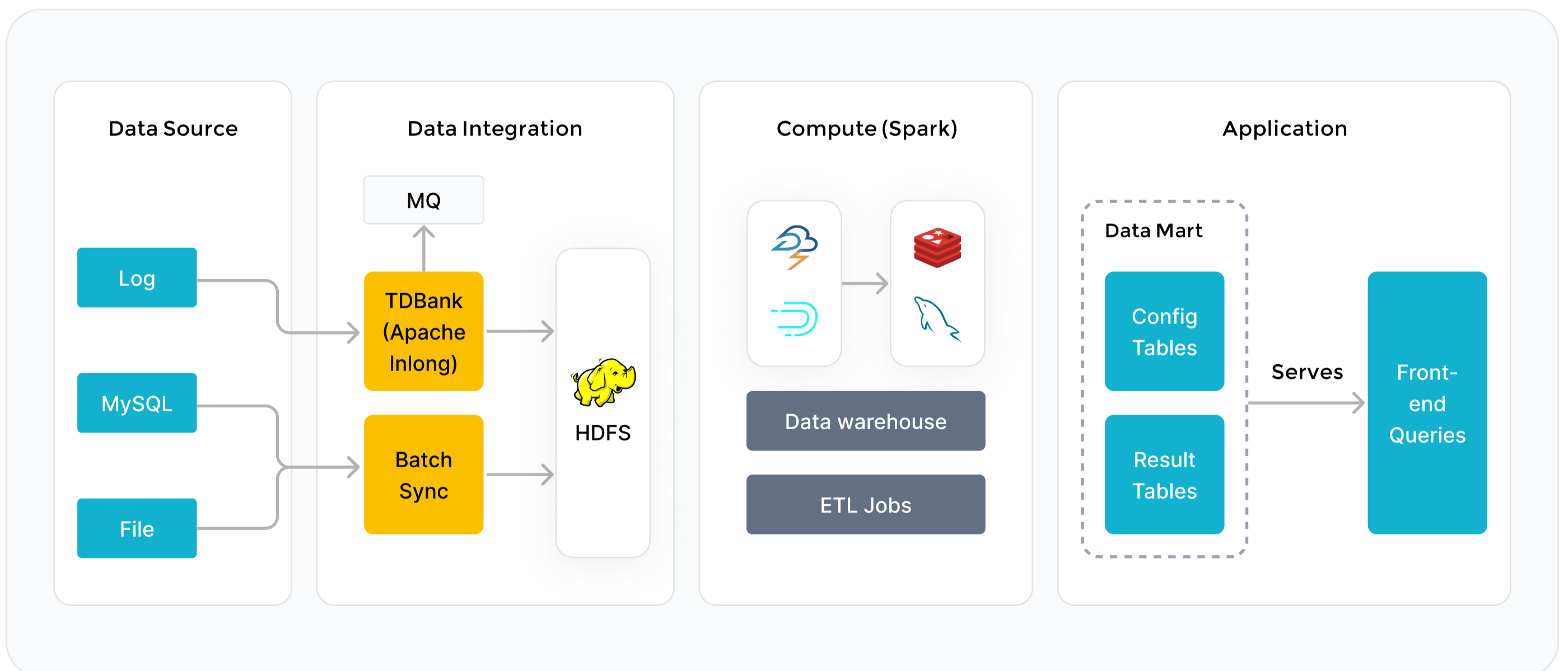
Challenges

Figure 1: Tencent Games' Original Architecture

Tencent Games's titles are developed and operated across multiple studios, resulting in data being siloed within separate companies inside their portfolio. They were looking for a way to evaluate the performance of all their games under one set of metrics. Unfortunately, with their complex old architecture consisting of Hive, Spark, Druid, Redis, MySQL, Postgres, and data being scattered around, Tencent was faced with a number of issues:

- **Scattered data:** Game logs were stored in HDFS, while the application-layer data was dispersed across transactional databases like MySQL and PostgreSQL. Moreover, real-time data is stored in Druid. It is not only the cost and challenges of managing these disparate data sources, this storage scheme made data usage difficult and often created bottlenecks in accessing critical information.

- **Separated Data Systems (Lambda architecture):** The old system used a Lambda setup, which meant one system for real-time data and another for offline/batch data. Managing two separate data paths is complex and costly to maintain.

- **Long Data Pipeline:** For query, all data must first be pre-processed, including pre-aggregation and denormalization, in Hive before moving to Postgres for reporting and dashboards. This long process is not only complex and a waste of computing and storage resources, but it also locks the data into a single-view format -- any schema change needs reconfiguring the pipeline and data backfilling

Solution

Facing these challenges, Tencent searched for an upgrade. They conducted a comprehensive evaluation of multiple database systems. Their primary metrics for assessment included query performance, elasticity, LakeHouse integration, data update capabilities, ease of use, integration options, and licensing.

Product	Data update	Ease of use	Integration	License
ClickHouse	Merge-on-read data update only, unstable query performance	Difficult to maintain, no cluster management tool	Internal storage only, no data lake support	Open source
StarRocks	Real-time data ingestion + update support	Easy to manage, easy to use	Works great as a query engine for the data lake	Open source
Presto	Data ingestion with SQL only, no real-time update	No local storage, relatively easier to manage	Good connectivity to many data sources	Open source
Snowflake	Supported	SaaS product, no infra to manage	Rely on S3, proprietary formats	Commercial software
Databricks	Supported	Complex UI, but relatively easy to use	Delta Lake only	Commercial software

Query Performance: Leveraging its true MPP architecture, complemented by the Cost-Based Optimizer (CBO) and a fully vectorized execution framework, StarRocks delivers rapid and scalable performance for both single and multi-table queries. Through its seamless integration with open data lakes—specifically, its impressive query latency with Apache Iceberg in our tests—StarRocks strengthens our LakeHouse capabilities, paving the way for enhanced data governance. In Tencent Games' current production environment with 300 nodes, along with other queries running, a JOIN between a 14 billion row table and a 1 billion row table only took 44 seconds. This showcases StarRocks' complex multi-table query performance at scale.

Data Updates: StarRocks' unique primary key table feature that supports real-time mutable data, we can transition most of our offline metrics to real-time. This not only accelerates our decision-making processes but also offers an up-to-the-minute insight into user behaviors and overall system performance.

Licensing: StarRocks being a Linux foundation and Apache-licensed open-source project, aligns well with our vision of embracing community-driven innovations.
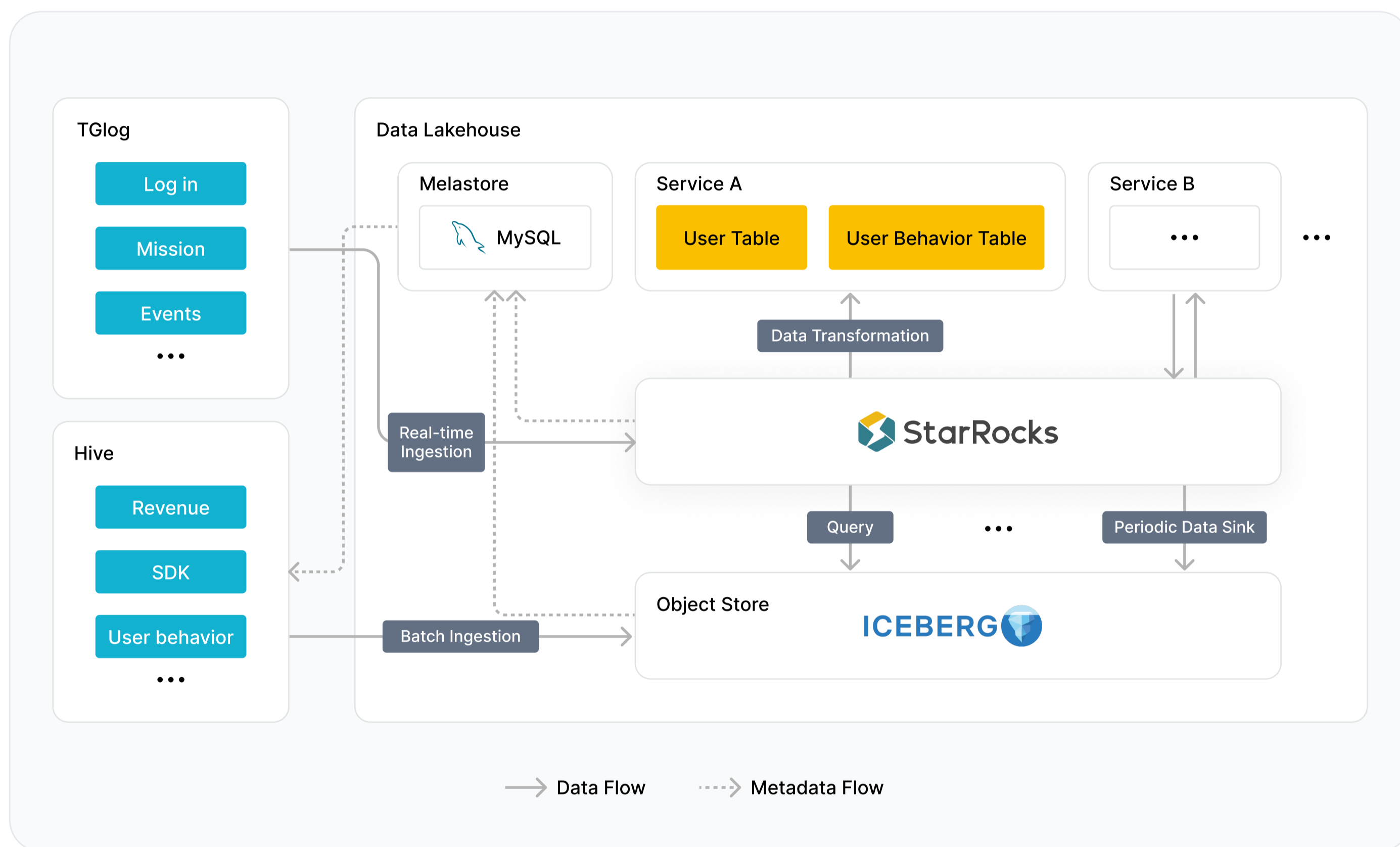

Figure 2: Tencent Games' Revised Architecture

In our new architecture, we unified all data into our data lake and StarRocks became the entry point for all real-time data and the unified query layer.

All data is ingested into StarRocks in real-time and periodically sinks into the data lake every hour. This way, we can achieve real-time data and a single source of truth on the data lake for better data governance.

RESULTS

By replacing their legacy data systems with StarRocks as their unified analytics platform, Tencent was able to bring their real-time and batch analytics into one system. This was not only cheaper to maintain but also meant there are fewer things that could break, thus improving the system's stability and availability.
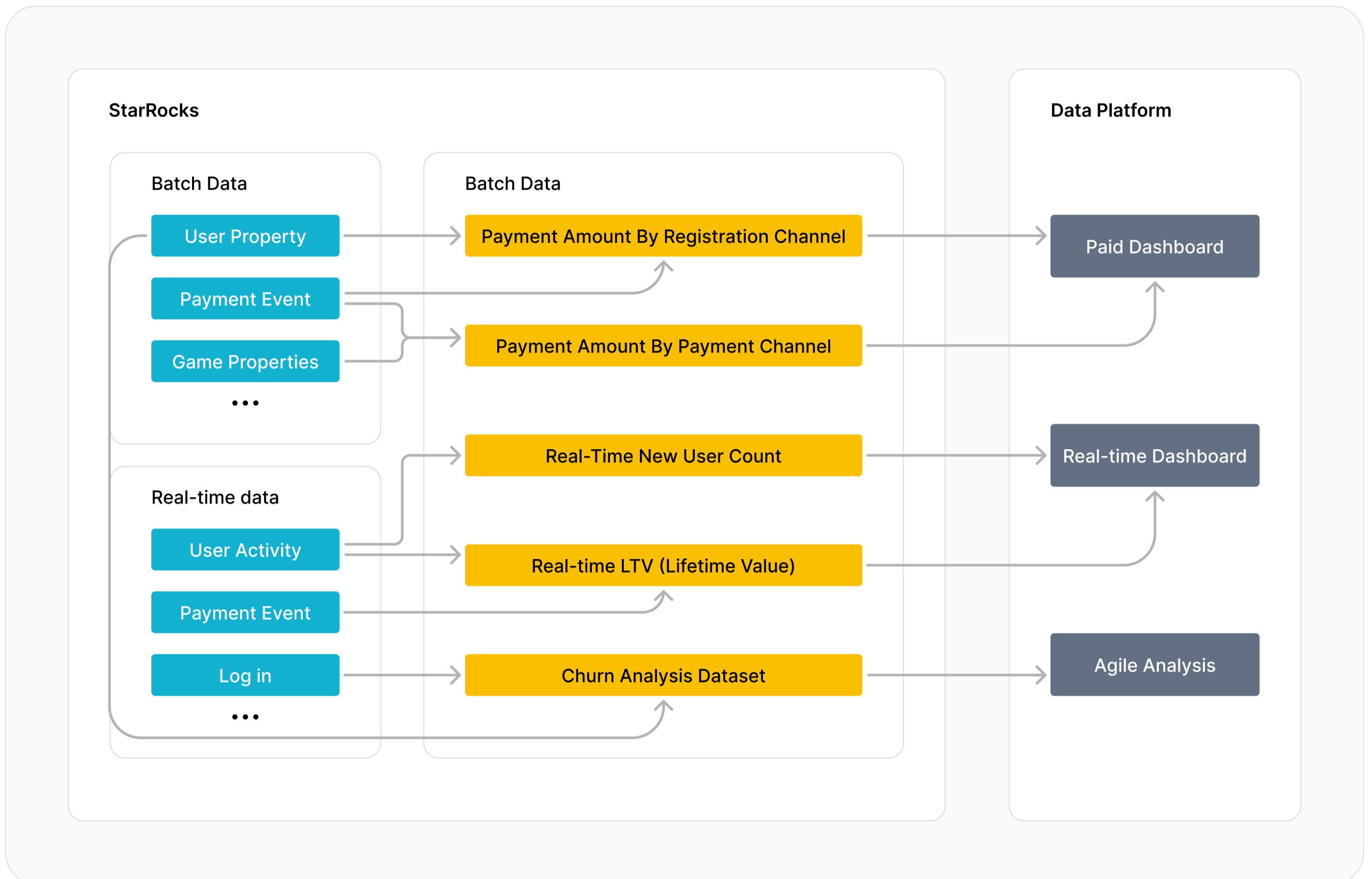

Figure 3: StarRocks-Based Infrastructure

StarRocks went beyond improving Tencent Games' data infrastructure, t also changed the way they used data:

- **Real-Time Metrics:** Tencent's gaming logs are now all being analyzed in real-time. This paves the way for more opportunities, offering insights that were previously out of reach for them.

- **Ditched denormalization:** One of the standout features of StarRocks is its ability to perform on-the-fly JOIN operations. This allows Tencent to sidestep the tedious process of denormalization and realize a remarkable 50% boost in their development efficiency.

- **Seamless Metric Adjustments:** Forgoing denormalization also gave Tencent a new level of flexibility in metric changes. They no longer need to reconfigure their data pipeline or undertake resource-intensive data backfilling whenever they decide to make changes.

- **Empowering Agile Analysis:** Since aggregations and JOINs can now be done on the fly, datasets are readily accessible through fixed SQL templates. This is a game-changer for Tencent's data users. Whether it's a tweak in dimensions or a change in statistical logic, users can make ad hoc adjustments, ensuring that their analyses remain agile and relevant.

WHAT'S NEXT FOR TENCENT GAMES?

After experiencing the initial success of StarRocks, Tencent Games plans to roll out usage across several other areas:

- **Enrich data assets in StarRocks:** Develop richer genre-specific and feature-based datasets in StarRocks for detailed game analyses and operational insights.

- **Keep pushing query performance on the data lake:** Integrating <u>StarRocks Materialized Views</u> for local data caching and optimizing Iceberg metadata storage in StarRocks for quicker filtering and less serialization overhead.

- **Enhancing Data Ingestion Performance:** Introducing compute node clusters to optimize data ingestion, reduce bottlenecks, and lighten the load on the main cluster.

- **Keep contributing to the StarRocks project:** Previously Tencent contributed to the <u>compute node</u> and participated in the development of the <u>StarRocks external catalog's Iceberg V2 table support</u>, and they plan to keep contributing to the StarRocks project in the future.