



Data Warehousing for Rapido





Data Warehousing

- **Data warehousing** refers to the process of collecting, storing, and managing large volumes of data from various sources in a centralized and structured repository, called a data warehouse. This centralized data store is designed to support decision-making by providing a unified, consistent, and easily accessible view of an organization's data.

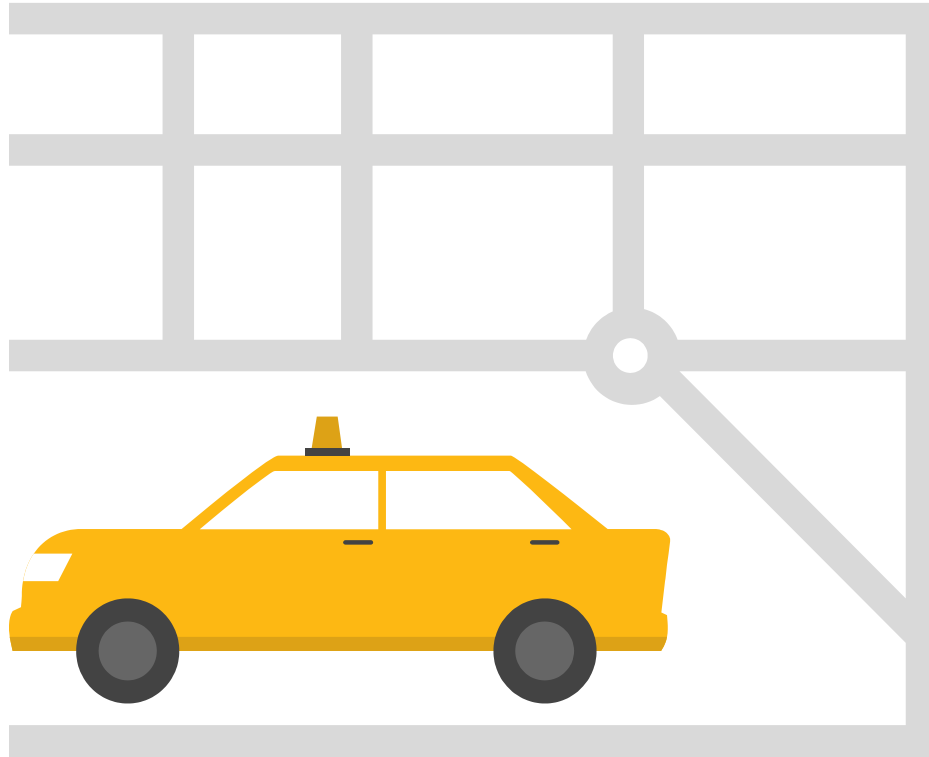


About Rapido

- Rapido, an Indian bike taxi aggregator, deals with massive amounts of data generated from its operations. This includes rider and driver details, trip data, real-time GPS information, payment transactions, and customer feedback. Managing this data efficiently is crucial for providing a seamless customer experience, optimizing routes, improving driver performance, and making data-driven business decisions.



Problem 1



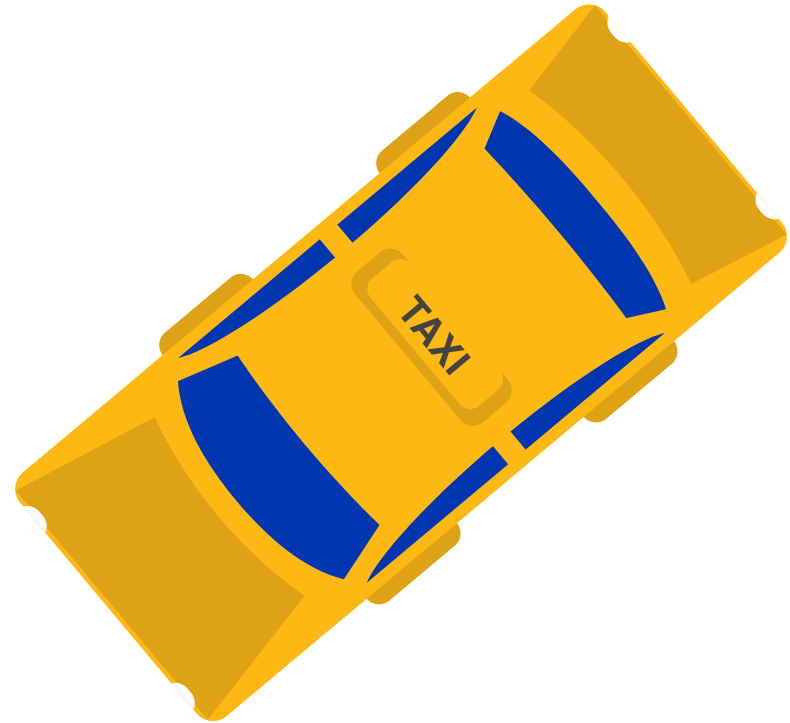
Rapido's data was stored across multiple systems—ride requests and completions in one database, payment information in another, and customer feedback in yet another. This fragmentation made it difficult to gain a holistic view of operations, causing inefficiencies in reporting and decision-making.



Solution 1

Data Warehousing Implementation

Rapido implemented a centralized data warehouse that integrated data from all these disparate sources. This unified platform allowed for comprehensive reporting and analysis, enabling better business insights. By consolidating data, Rapido could now track the entire customer journey—from booking to payment—in a single view.

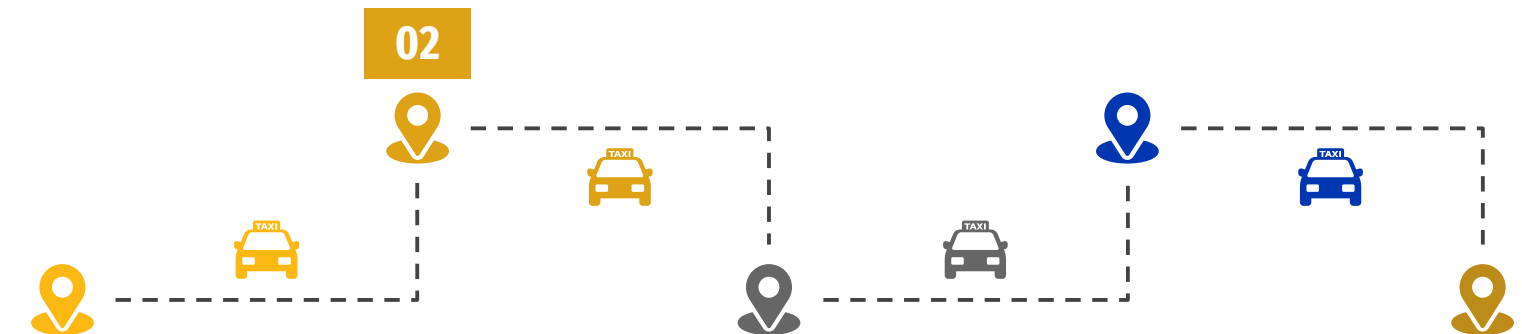




Role of Analyst

Data Mapping

They were responsible for mapping the data to ensure consistency and relevance, making sure that data from different sources aligned correctly in the warehouse.



01

Data Integration

Analysts worked on the Extract, Transform, Load (ETL) process to ensure that data from various sources was accurately combined and cleansed before being loaded into the warehouse.

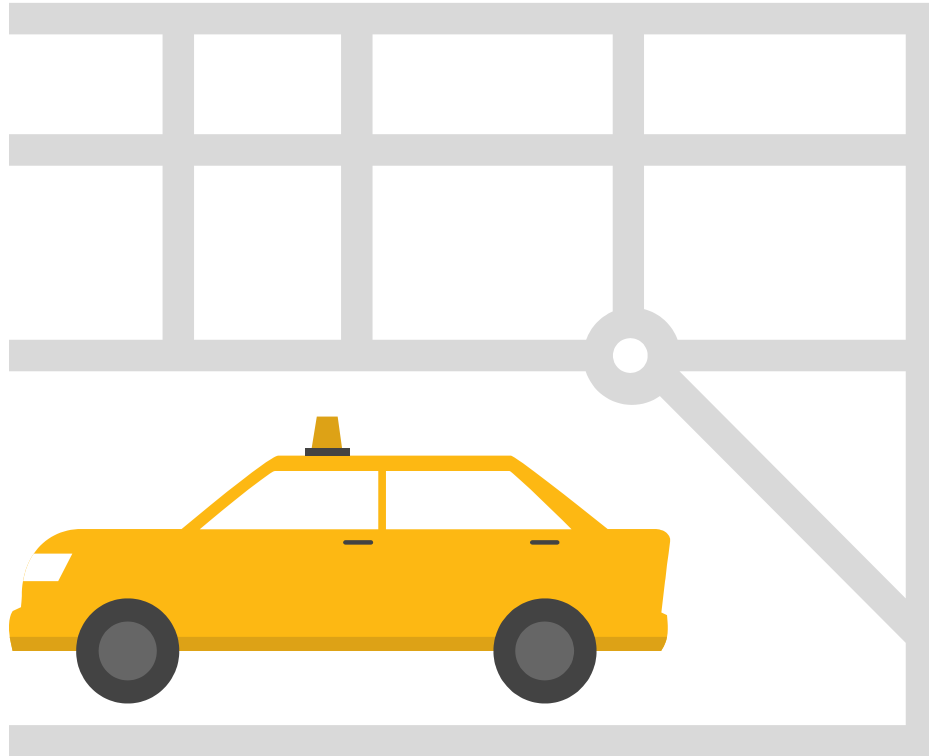
03

Reporting & Insights

Post-integration, analysts created dashboards and reports that provided actionable insights to various departments, helping them optimize processes like route planning, customer support, and marketing strategies.



Problem 2



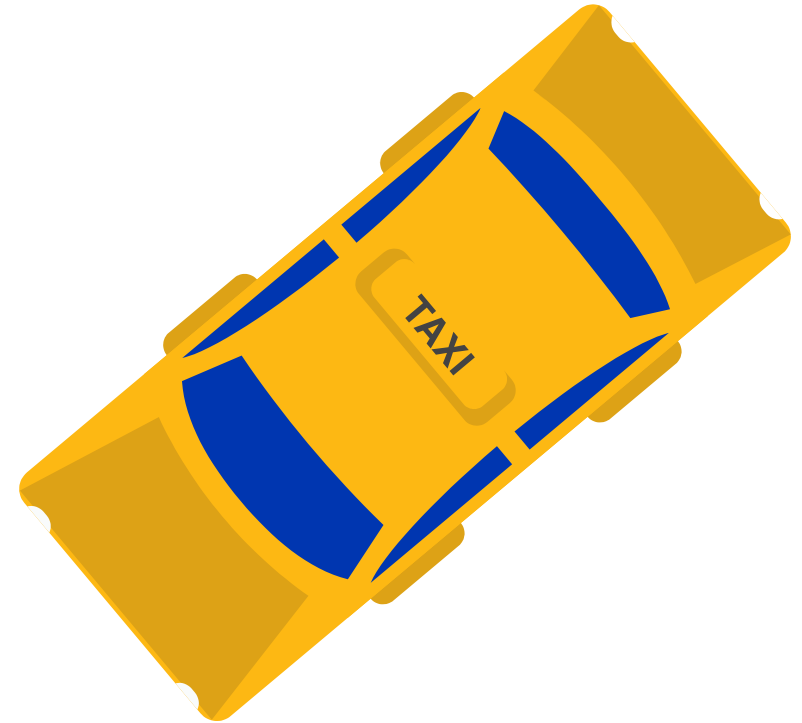
With the growing size of Rapido's data, running complex queries, especially those involving large datasets, became increasingly time-consuming. This slowdown impacted the ability to generate real-time reports, crucial for quick decision-making in a fast-paced environment like ride-hailing.



Solution 2

Optimization of Data Warehouse

Rapido's data warehouse was optimized using techniques such as indexing, partitioning, and materialized views. Additionally, they implemented a columnar storage format, which significantly improved the speed of query processing for analytical workloads.





Role of Analyst

Query Optimization

Analysts worked on optimizing SQL queries, ensuring they were efficient and made full use of the data warehouse's features.

02



01



03

Performance Monitoring

They continuously monitored the performance of the data warehouse, identifying slow queries and recommending further optimizations.

To streamline operations, analysts developed automated scripts that ran common queries at regular intervals, ensuring that decision-makers had access to the latest data without delay.

Automation



Summary

- The implementation of a centralized data warehouse and subsequent optimizations solved critical issues related to data fragmentation and query performance for Rapido. Analysts played a pivotal role in these processes, from integrating and mapping data to optimizing queries and monitoring system performance, ultimately enabling Rapido to make faster, more informed decisions.