EdTech leader migrates Analytics Workload from AWS to GCP.

18%
Reduction in TCO

67%
Reduction in customer complaints

32%
Faster Queries

CLIENT

The client is an EdTech unicorn that offers a SaaS platform for employee learning, skills and career mobility.

The platform helps companies to attract, develop and retain a high-performance and future-ready workforce.
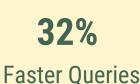
AT-A-GLANCE

Challenges

- Migrating Analytics workload from AWS to GCP

- Migrate to a large scale, near-real-time data platform from AWS to GCP at the same time consolidate business data from multiple sources, dedup, and enrich

Solution

- Migrated from AWS to GCP with 100% serverless components from GCP

- Near real-time data pipeline using Kafka and data enrichment using Spark helped in better reporting

PROJECT CONTEXT

The client wanted to migrate to a large-scale, near-real-time data platform from AWS to GCP to capture and process millions of user events in a near real-time environment.

They needed to consolidate business data from multiple sources, dedup, and enrich for integration with event data.

PROJECT OBJECTIVES

- Build a data lake for reporting using summaries and time-series aggregations with the ability to run large ad-hoc queries and data exports.
- Support ML and advanced analytics use cases.
- Minimise the total cost of ownership.

SOLUTION DELIVERY

- Real-time data pipelines were built from all data sources using Kafka and data was processed in Spark for enrichment.
- Detailed TCO analysis was done for the customer to make them aware of the benefits of GCP serverless architecture
- The data lake was centrally managed by SquareShift's data engineering team.
- First time at the firm, integrated, golden data was available for analysis.
- Data Scientists were able to use live data for their algorithms

TECHNOLOGY STACK

Google Big Query	Google Cloud Functions	Cloud SQL	kafka	Cloud DataFlow	elasticsearch