PvX Partners

Case study on how KeyValue built a financial services platform providing non-dilutive capital to support user acquisition strategies for gaming companies

Report type:	Domain:	Market
Case study	Investment	🇸🇬 Singapore

KEYVALUE
Let's build your idea

Problem statement

Build an intelligence & insights platform to understand the performance of all invested companies on a daily basis. This should enable PVX to monitor everything from marketing performance, revenue, and all other KPIs & business metrics of the invested companies that really matter. The platform uses predictive modelling and advanced machine learning models to forecast future performances, aiding proactive decision-making for PVX.

Underwriting   Benchmarking   Visualisation

Monitoring   Financial projections

Key metrics

50+
Different games

60+
Data pipelines

350+ charts to visualise data

800GB
processed in 20 minutes.

2M+
Transactions per game

10+ Data sources

Business impact

Improved investment decisions

Enabled PvX Partners to make data-driven underwriting and monitoring decisions, reducing risk and enhancing investment outcomes.

Predictive capabilities

Established a robust data foundation for machine learning models to forecast future performances, aiding proactive decision-making.

Operational efficiency

Automated data workflows with minimal manual intervention – lowering operational costs and errors.
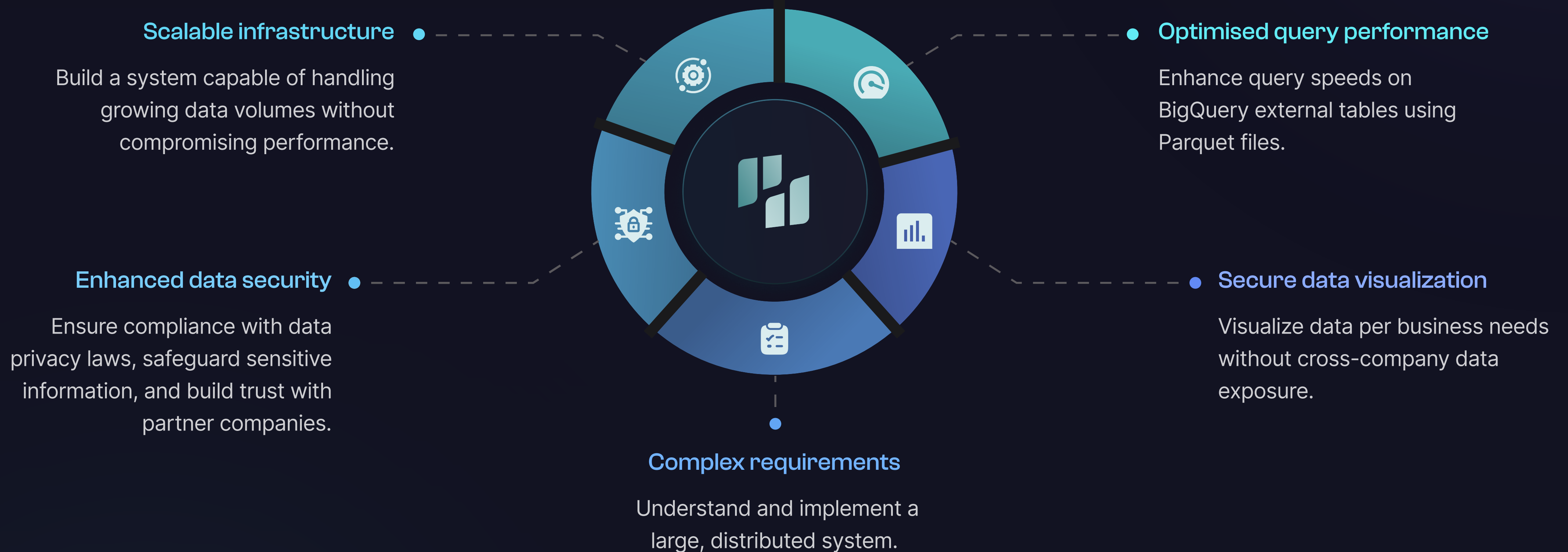
Competitive insights

Provided companies with benchmarking tools to gauge performance against industry peers, fostering strategic growth.

Challenges

Scalable infrastructure

Build a system capable of handling growing data volumes without compromising performance.

Enhanced data security

Ensure compliance with data privacy laws, safeguard sensitive information, and build trust with partner companies.

Optimised query performance

Enhance query speeds on BigQuery external tables using Parquet files.

Secure data visualization

Visualize data per business needs without cross-company data exposure.

Complex requirements

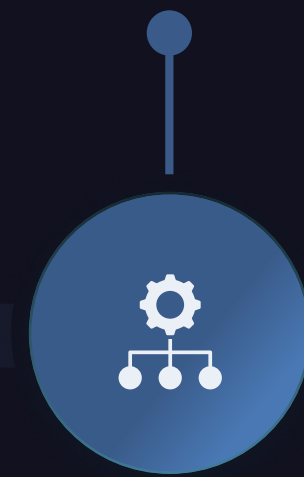Understand and implement a large, distributed system.

Implementation

Orchestration
Implemented Apache Airflow (self-hosted) for job scheduling to reduce costs.
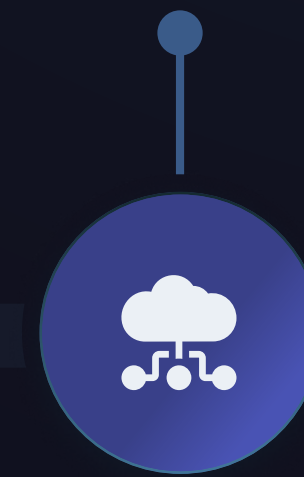
Query optimization
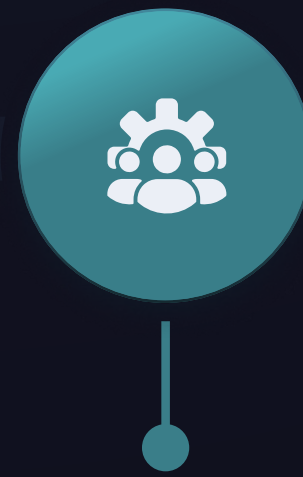Partitioned Parquet files on frequently filtered columns to minimize data scanning.

Cloud platform
Chose Google Cloud Platform (GCP) for its services like Dataproc, GCS, BigQuery, and Vertex AI.
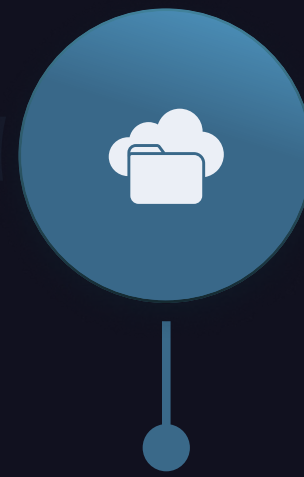
Spark job management
Employed Dataproc batches for serverless, auto-scaling Spark jobs triggered as needed.
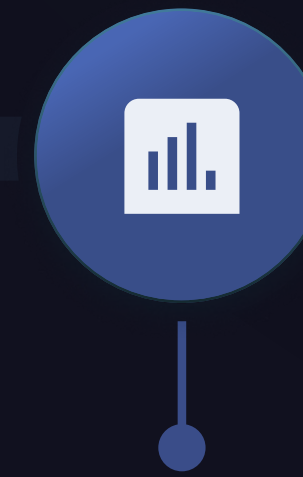
Data storage format
Used Parquet format in Google Cloud Storage (GCS) and added as external tables in BigQuery for querying.

Data visualization
Adopted Apache Superset (self-hosted) for its extensive chart options and ability to handle data security with roles, permissions, and row-level security.

Technology stack

Machine learning — Vertex AI

Google Cloud Platform

Data sources
- Google BigQuery
- Google Cloud Storage
- CSV
- amazon S3
- PostgreSQL
- snowflake

Workflow & orchestration — Apache Airflow

Extract
python

Validate →

Transform
APACHE Spark

Aggregate
APACHE Spark

Datalake
Parquet

Data warehouse
Google BigQuery

Data mart
Google BigQuery · PostgreSQL

Visualize
- Apache Superset
- Chart.js

WebApp
- React
- spring boot

Identity management
KEYCLOAK