

CASE STUDY



Achieving 30% Cost Reduction Through Pipeline Optimization

BUSINESS CHALLENGES

Our client, a leading e-commerce platform, sought to reduce their operational costs associated with their data platform. With over **250 active data pipelines** with **660 TB data**, the existing infrastructure was consuming substantial resources, leading to escalating expenses. The client approached us to identify inefficiencies and implement cost-saving measures

OUR SOLUTION

We conducted a comprehensive audit of all **250+ data pipelines** to pinpoint resource data-intensive jobs.

Our optimization strategy included:

- **Cluster Optimization:** Adjusted cluster sizes and configurations to better align with workload requirements, ensuring efficient resource utilization
- **Auto-Termination Implementation:** Enabled auto-termination for clusters to shut down automatically after periods of inactivity, preventing unnecessary resource consumption
- **Job Scheduling Adjustments:** Rescheduled non-urgent jobs to off-peak hours, taking advantage of lower cloud resource costs during these times
- **Code Optimization:** Refined ETL scripts to enhance performance, reducing execution time and resource usage
- **Data Storage Optimization:** Implemented data retention policies and optimized storage formats to decrease storage costs

SERVICES PROVIDED

- Data Engineering
- Cost Optimization
- Databricks Performance Tuning
- Cloud Resource Management
- FinOps Strategy Implementation

VALUE DELIVERED

- Achieved a **30% reduction** in both storage and Databricks resource costs
- Increased processing speed and time efficiency **by 70%**
- Enhanced pipeline performance and reliability.
- Improved cost visibility and governance through detailed reporting
- Established a foundation for continuous cost optimization practices

