

A woman with long brown hair, wearing a beige blazer and gold hoop earrings, is smiling and looking towards the camera while working on a laptop. She is seated at a white desk in a modern office environment. In the background, other office workers are visible, including a man in a blue suit and a woman with curly hair, both working on laptops. The office has large windows and a bright, airy feel.

HEXAWARE

Investment Data Platform Modernization on Google Cloud

A Scalable Foundation for Growth

Case study

Client

| Client

Challenge

Solution

Benefits

Summary

Over a Century of Supporting Financial Futures

Established in the mid-1800s, the enterprise has a long-standing commitment to helping people prepare for the future. Today, it continues to empower its customers to make the most of their financial future, building on its experience and commitment to supporting them through every stage of their financial journey.



Challenge

- Client
- Challenge**
- Solution
- Benefits
- Summary

Enterprise-Scale Data Management and Investment Operations

As part of a highly regulated industry, the enterprise relies on accurate, timely, and governed data to support investment operations, business decision-making, regulatory reporting, and compliance requirements.

The customer operates at an enterprise scale, managing data from multiple upstream systems and serving a broad set of downstream analytics and reporting stakeholders. Investment-related data was received from five different upstream systems in heterogeneous formats including CSV, XLSX, XLSB, and TXT, creating a complex data environment that needed to support daily data delivery requirements and timely availability of curated datasets.

The customer needed a robust, scalable, and automated cloud-native data platform that could support future growth while maintaining cost efficiency.

Complex Investment Data Landscape

The customer faced challenges in managing and processing investment-related data across five different upstream systems and heterogeneous formats including CSV, XLSX, XLSB, and TXT, making standardized processing difficult.

Fragmented Data Architecture

There was no standardized data architecture separating raw, processed, and business-ready datasets. Inconsistent source structures also resulted in data quality and usability challenges, making it difficult to establish clear traceability and increasing effort for downstream reporting and analytics.

Manual Pipeline Dependencies

Data pipelines relied heavily on manual intervention, leading to operational inefficiencies and increased risk of failures. Manual processing also increased engineering effort for data corrections and reprocessing activities.

Limited Pipeline Visibility and SLA Risk

Limited monitoring and observability made it difficult to identify and resolve pipeline issues proactively, increasing the risk of downtime and operational disruption across daily processing workflows. The fragmented processing environment also created higher operational risk in meeting daily SLA commitments, with delays potentially affecting downstream reporting and analytics processes.

Data Governance and Regulatory Requirements

Data quality and processing challenges reduced confidence in data used for business and regulatory decision-making. In a regulated financial environment, this created potential governance and compliance concerns, while increasing overall operational and regulatory risk.



Solution

- Client
- Challenge
- Solution**
- Benefits
- Summary

A Cloud-Native Data Foundation on Google Cloud

Hexaware designed and implemented a cloud-native data engineering platform on Google Cloud Platform, specifically tailored to meet the customer’s regulatory, operational, and scalability requirements. This Google Cloud data engineering-led foundation provided the customer with a robust, scalable, and automated cloud-native data platform.

A Standardized Medallion Architecture

The customer’s investment data was organized through a medallion architecture (ODP > FDP > CDP), creating a clear separation of raw, cleansed, and business-ready datasets. This standardized approach improved data processing, quality, traceability, and data governance, while supporting downstream analytics.

Five Sources, One Standardized Data Flow

Data from five upstream systems in CSV, XLSX, XLSB, and TXT formats was landed into Google Cloud Storage (GCS). Apache Airflow DAGs orchestrated end-to-end ingestion workflows, while Dataflow pipelines built using Apache Beam and Python processed structured and semi-structured source files.

Data was first loaded into staging tables and subsequently ingested into ODP, providing the customer with a consistent raw data layer before further cleansing, standardization, and integration.

ODP, FDP, and CDP for Trusted Data

The ODP layer provided a clear foundation for raw data, enabling source data to be retained before further processing.

In the FDP Layer (Silver), data was cleansed, standardized, and integrated. SQL-based transformation logic was applied on ODP datasets, with data quality and consistency checks incorporated during processing.

In the CDP Layer (Gold), business-ready datasets were created through further aggregation and enrichment. Data models were optimized for reporting, analytics, and downstream consumption.

This layered approach improved traceability and auditability by maintaining a clear separation between raw, cleansed, and business-ready datasets.

Automated Investment Data Pipelines

Apache Airflow managed dependencies across all processing stages. Modular DAG design enabled reusability, failure isolation, simplified maintenance, and future scalability.

This data pipeline automation reduced manual intervention and operational overhead while providing scalable workflow management with improved failure handling.

Optimized BigQuery Performance and Cost

The customer’s business-ready data models were optimized for reporting, analytics, and downstream consumption. **BigQuery** performance and cost optimization techniques were incorporated to improve processing performance while controlling cloud costs.

Developing Improved Visibility and Operational Reliability

Log-based metrics were implemented within GCP, while automated email notifications were configured for failures and anomalies. The customer gained enhanced observability across data processing workflows, improving operational reliability and reducing downtime.

Solution

- Client
- Challenge
- Solution**
- Benefits
- Summary

How We Built a Consistent, Governed Cloud Infrastructure

Terraform was used to provision and manage all cloud resources, with infrastructure maintained through version-controlled deployments. This enabled consistent deployments across development, testing, and production environments and supported consistency, repeatability, and **data governance** across environments.

Built for Financial Services Data Requirements

The architecture was specifically tailored to meet the customer’s regulatory, operational, and scalability requirements. The layered medallion architecture aligned with financial services data governance standards, while clear separation of data layers improved traceability and auditability.

Automated end-to-end processing reduced manual intervention and operational overhead. Reusable pipeline components supported rapid onboarding of future data sources, providing greater scalability, operational resilience, data quality, and long-term maintainability compared to fragmented and manually managed pipelines.



Benefits

- Client
- Challenge
- Solution
- Benefits**
- Summary

Business Value from a Modern Investment Data Platform

A scalable, automated, and governed Google Cloud data platform improving data quality, operational reliability, reporting readiness, and confidence in enterprise data.

100% Daily SLA Focus

The platform supports daily SLA-driven data delivery requirements through automated end-to-end processing and improved workflow management.

5 Upstream Systems, Standardized Processing

Data from five different upstream systems and heterogeneous file formats can be processed through a standardized data architecture, improving consistency and simplifying future onboarding.

3-Layer Data Foundation

The ODP-FDP-CDP layered architecture provides clear separation of raw, cleansed, and business-ready datasets, improving traceability, auditability, and data governance.

Reduced Manual Intervention

Automated end-to-end processing reduced manual intervention and operational overhead, while modular Airflow workflows improved failure isolation and maintenance.

Improved Data Quality and Consistency

Data quality and consistency checks incorporated during processing support improved data completeness across source systems and reduce data processing errors and reprocessing activities.

Faster Access to Business-Ready Data

Curated datasets are made available for reporting and analytics teams, improving the availability of trusted datasets and providing faster access to business-ready information.

Improved Reliability and Observability

Log-based metrics, automated failure notifications, and enhanced observability improved operational reliability, reduced downtime, and strengthened the ability to detect and resolve pipeline failures.

Scalable and Governed Platform

Terraform-based infrastructure management ensures consistency, repeatability, and governance across environments, while reusable pipeline components support rapid onboarding of future data sources. The Google Cloud platform provides a scalable foundation for future growth.



Summary

- Client
- Challenge
- Solution
- Benefits

| Summary

A Standardized Investment Data Platform

The enterprise is now better positioned to support future growth with a scalable, automated, and governed data foundation on Google Cloud.

With data pipeline automation, standardized ODP-FDP-CDP Medallion architecture, reusable pipeline components, and strong data governance, the enterprise can onboard new data sources faster, expand analytics capabilities, support daily SLA commitments, and build greater confidence in its enterprise data assets.

Its future expansion can support broader analytics capabilities and faster access to business-ready information. Reusable components provide a foundation for continued growth while maintaining governance and operational reliability. The platform is positioned to evolve with the enterprises’ changing investment data needs.



Build for What's Next

Explore how [Hexaware's Google Cloud partnership](#) can help build a scalable, governed data foundation for future growth.

About Hexaware

Hexaware is a global technology and business process services company. Every day, Hexawarians wake up with a singular purpose: to create smiles through great people and technology. With this purpose gaining momentum, we are well on our way to realizing our vision of being the most loved digital transformation partner in the world. We also seek to protect the planet and build a better tomorrow for our customers, employees, partners, investors, and the communities in which we operate.

With offices across the world, we empower enterprises to realize digital transformation at scale and speed by partnering with them to build, transform, run, and optimize their technology and business processes.

Learn more about Hexaware at <https://www.hexaware.com>.

Get in touch