PROMOTION REVIEW 2026

Technical Leadership

Engineering Excellence in Sustainability

Showcasing contributions to enterprise-scale ESG data infrastructure and reliability.

PROJECT SHOWCASE

Vantage

ESG Analytics Platform

> Vantage is GIC's **core sustainability analytics platform**, designed to compute ESG metrics such as WACI, WAGR, CVaR, and PR across large multi-year portfolios.

It acts as a **centralized sustainability datamart** that powers analytics, benchmarking, and reporting workflows used across multiple business functions.

45.5 TB	7+ Years	10+
DATASET SIZE	HISTORICAL ESG DATA	DEPTS SUPPORTED

IMPACT & REACH

Enterprise Dependencies

Serving as the single source of truth for ESG data across the organization.

CONSUMER ECOSYSTEM

Who Depends on Vantage?

Vantage serves as the enterprise-wide golden source for critical ESG data used in client reporting, sustainability analytics and climate risk governance across GIC.

RPMD

Sustainability analytics & reporting

Sustainability Office

ESG governance workflows

Investment Teams

(PE / RE / EMD / TPS)

Risk & Reporting

Regulatory & client reporting workflows

The platform ensures consistent ESG metrics across departments.

GOLDEN SOURCE

PLATFORM STATISTICS

By The Numbers

Engineered for enterprise scale, processing massive datasets with high reliability.

INFRASTRUCTURE SCALE

Scale of the Vantage Platform

A centralized datamart powering sustainability analytics across the entire organization.

TOTAL DATASET

45.5 TB

High-volume storage for granular ESG data points

HISTORICAL DEPTH

7+ Years

Continuous historical computation for trend analysis

ESG MODULES

4 Core

Covering WACI, WAGR, CVaR, and PR metrics

CONSUMERS

10+ Depts

Enterprise-wide consumption (Investment & Sustainability)

AUTOMATED TESTS

1,200+ Cases

Rigorous regression testing ensuring data accuracy

ASSET COVERAGE

Multi Class

Private Equity, Real Estate, EMD & Public Strategies

ROLE DEFINITION

Squad Lead Responsibilities

Balancing technical direction, delivery execution, and team mentorship.

LEADERSHIP & EXECUTION

My Scope & Ownership

Squad lead responsible for **delivery, technical direction, and engineering standards** for the Vantage platform.

STRATEGY & DELIVERY

Execution Planning

Owns timeline commitments and drives execution planning to ensure predictable delivery.

Stakeholder Alignment

Aligns platform changes with downstream consumers to maintain contract stability.

ENGINEERING EXCELLENCE

Architecture Leadership

Leads architectural and performance decisions to ensure system scalability.

Engineering Execution

Drives day-to-day engineering execution and removes technical blockers.

Mentorship & Quality

Mentors junior developers and enforces high engineering quality standards.

SYSTEM DESIGN

Platform Architecture & Engineering Design

PASTE YOUR ARCHITECTURE DIAGRAM HERE

Full-Width Area Available

HIGH-LEVEL ARCHITECTURE

- **Orchestration:** Airflow manages end-to-end pipelines.
- **Compute:** Modular ETL layers for metric processing.
- **Storage:** S3 Data Lake as shared ESG storage layer.
- **Query:** Athena/Presto serve analytics consumers.

ARCHITECTURE OWNERSHIP

Airflow Framework Design

Defined reusable DAG patterns & standards.

Partitioning Strategy

Implemented Hive-style S3 partitioning.

Reliability Engineering

Reduced operational failure risk through pipeline isolation

DESIGN PHILOSOPHY

Engineering Principles

Five core pillars guiding the architecture of Vantage

Scalability

- Partitioned S3 architecture supports multi-year ESG datasets
- Pipelines designed for incremental + historical backfills
- Modular DAGs allow independent scaling of workloads

Maintainability

- Reusable Airflow DAG templates for consistency
- Standardized pipeline structure across all ESG metrics
- Clear ownership boundaries between system components

Automation

- Fully server-side orchestration (no manual EUC workflows)
- Automated retries and continuous monitoring
- Parameterized execution for full reproducibility

Fault Tolerance

- Idempotent pipelines safe for re-run at any stage
- Isolated job failures do not cascade to other modules
- Retry strategy built directly into orchestration layer

Reproducibility

- Hive-style partitioning guarantees deterministic outputs
- Historical datasets can be regenerated consistently
- Strict version-controlled pipeline logic

QUALITY ASSURANCE

Trust Through Testing

Automated validation ensuring data integrity at scale.

Reliability & Testing Architecture

A robust framework combining the Data Validation Engine (DVE) with operational safeguards.

Data Validation Engine

THE FRAMEWORK

An internal validation framework specifically designed for ESG pipelines.

test_config.yaml

```
test_type: "reconciliation"
source: "s3://raw_esg/..."
tolerance: 0.001
# Validates vs SQL Logic
```

Declarative Registry
Tests defined in simple YAML + SQL.

Automated Suites
Accuracy, reconcile, and regression tests.

Historical Coverage
Runs against full multi-year ESG datasets.

Operational Reliability

THE OUTCOME

Transforming testing from a manual bottleneck into an automated safety net.

TESTING SPEED	+80% Faster
CONFIDENCE	High

Repeatable UAT
Execution without manual intervention.

Automated Tracking
Result logging and diff reporting.

Safe Refactoring
Changes deployed with regression guarantees.

LEGACY STATE

Critical Bottlenecks

Initial architecture limitations preventing scalability and rapid iteration.

DEEP DIVE TOPIC 1 | PERFORMANCE TUNING

Initial Architecture Constraints

Baseline analysis: Severe performance limitations identified during the initial assessment.

~24 Hours

Extreme Processing Latency

Single data date run took ~24 hours per module, making daily reporting impossible.

~2 Months

Prohibitive Historical Processing

60 historical month-end runs required ~2 months to process, blocking methodology changes.

~6 / Year

Limited Iteration Velocity

Only ~6 full reruns possible per year, severely limiting agility and bug fixing capacity.

OOM Errors

Memory Exhaustion

Pandas-based in-memory processing frequently exceeded Glue memory limits.

ARCHITECTURE 2.0

THE SOLUTION

Architectural Redesign for Scalability

Five key engineering moves that unlocked performance and maintainability.

STRUCTURAL IMPROVEMENTS

01 Step-Based Modular Architecture

Refactored monolithic jobs into discrete steps. Pipelines are now composed of small, manageable units.

02 Idempotent Modules

Each module is Parquet-output driven. Failures can be retried safely without data duplication or cleanup scripts.

03 Logical Class Separation

Code structure realigned to business methodology, making logic easier to trace and audit for Sustainability teams.

COMPUTE ENGINE UPGRADE

04 SQL-on-Parquet (DuckDB)

High Impact

Moved transformation engine from iterative code to vectorized SQL-on-Parquet using DuckDB.

- Vectorized execution speed
- Direct S3 querying capabilities

05 Eliminated Memory Bottlenecks

Removed large in-memory Pandas workloads that caused OOM errors.

Scalable Processing

> Transforming the engine for speed, reliability, and massive scale.

PERFORMANCE TUNING

Smart Storage

Decoupling ingestion from computation to accelerate iterations.

OPTIMIZATION SPOTLIGHT

S3 Caching Layer Optimization

Removing the ingestion bottleneck to enable rapid historical reprocessing.

SOLUTION - ITERATION 2

Prevent Redundant Downloads

Introduced an intermediate S3 caching layer. Upstream data is downloaded once and stored as Parquet.

Reuse Across Methodologies

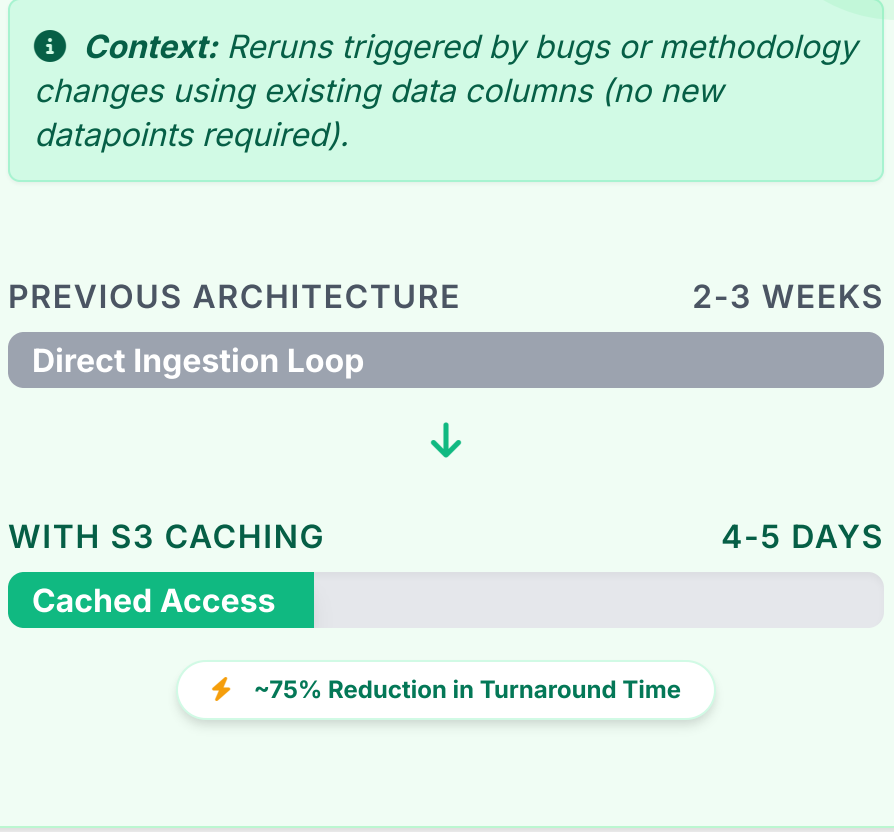
Cached datasets are methodology-agnostic. When logic changes, we re-run compute immediately against the cache without waiting for ingestion.

Eliminate Bottlenecks

Removed the multi-hour ingestion step from the critical path of reprocessing jobs, transforming IO-bound jobs into CPU-bound ones.

Historical Take-on Velocity

Context: *Reruns triggered by bugs or methodology changes using existing data columns (no new datapoints required).*



BUSINESS VALUE

Tangible Outcomes

Quantifiable improvements in delivery velocity and operational efficiency.

IMPACT ANALYSIS

Business Impact of Performance Redesign

Transforming technical optimizations into strategic business advantages.

TIME-TO-DELIVERY

28% Faster

End-to-end delivery timeline significantly compressed, accelerating time-to-insight for investment teams.

Faster Iteration Cycles

Rapid feedback loops allow for quicker validation of new ESG methodologies.

Metric Refinement

Responsiveness & Scope

Supports expanding ESG scope and enables instant response to ad-hoc Sustainability Office requests.

Service Agility

Higher Throughput

Delivering more metric modules without increasing team headcount or infrastructure costs.

Capacity Flat

PREVIOUS STATE

Manual Bottlenecks

Legacy testing limitations that existed prior to automated validation framework.

DEEP DIVE TOPIC 2 | TESTING AUTOMATION

Testing Bottlenecks Limiting Velocity

Legacy SIT & UAT processes were manual, labor-intensive, and error-prone, creating friction in the release cycle.

High Latency

Slow EUC Query Execution

Queries were fired manually against pre-prod/prod environments, resulting in long turnaround times for large datasets.

High Risk

Manual Parameterization

Date switching across test ranges was performed by hand, introducing a high risk of human error during configuration.

Limited Scale

No Scalable Regression

Impossible to validate wide historical ranges efficiently; overall testing scope was strictly limited by human capacity.

No Audit Trail

Fragmented Results Tracking

Test results were tracked outside the system in disparate files, making it difficult to compare runs or audit historical quality.

ARCHITECTURE

System Design

Automated validation framework ensuring data integrity at scale.

DVE FRAMEWORK

DVE Architecture: Automated Data Validation

Standardizing how data validations are defined, executed, and tracked across the enterprise.

EXECUTION FLOW

Sequential Pipeline

01

Test Registry

Test definitions stored as `YAML + SQL`. Fully version-controlled and reusable across environments.

02

Source Loading

Identifies unique datasets needed. Extracts once and **caches as Parquet** to minimize I/O overhead.

03

Execution Engine

Parameterized SQL execution across date ranges. Applies preprocessing and **tolerance rules** consistently.

04

Comparison Layer

Automated diff detection for missing, extra, or mismatch records. Outputs structured audit logs.

KEY ENGINEERING PROPERTIES

Deduplicated

Smart extraction avoids redundant data pulls across tests.

Deterministic

Comparison rules guarantee reproducible results every run.

Config-Driven

Onboard new tests via YAML without code changes.

Scalable

Scales to massive datasets without linear manual effort.

IMPACT ANALYSIS

Operational Excellence

> Accelerating delivery velocity while simultaneously reducing operational risk.

DVE FRAMEWORK IMPACT

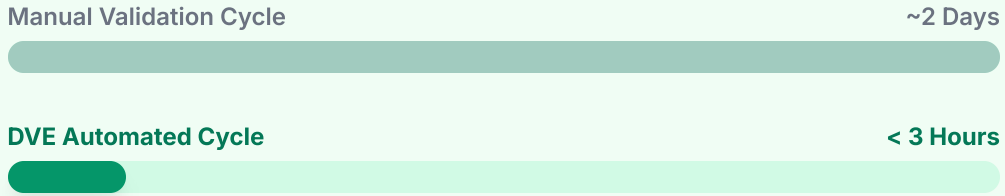
Accelerating Delivery & Reducing Risk

Transforming manual validation into an automated strategic advantage.

DELIVERY ACCELERATION

80% Faster

Validation cycles compressed from multiple days to hours, enabling faster UAT sign-off and shorter release timelines.



Massive Scalability

Validates tens of millions of records on demand. Supports growing dataset sizes without linear manual effort.

- Volume Independent

Risk Reduction

Eliminates manual parameterization errors. Ensures standardized, auditable, and repeatable test runs.

- Audit Ready

Team Productivity

Engineers focus on feature delivery instead of manual testing. Achieves higher testing coverage with less effort.

- High Velocity