

# Simplifying the Modern Data Stack



## How Acumen Vega Accelerates Iceberg Adoption and ML-Ready Analytics with no vendor lock-in and world class ML capabilities

Building a scalable and intelligent data platform in today's enterprise environment requires more than just speed. It demands openness, interoperability, governance, and the ability to power AI/ML workloads at scale. Apache Iceberg is the foundation of modern lakehouses for good reason—it decouples storage from compute, supports massive schema evolution, and enables ACID transactions at petabyte scale. But standing up in a production-grade Iceberg environment is often complex.

### That's why Acumen Vega was created in the first place

Acumen Vega is a turnkey accelerator for Iceberg adoption on Google Cloud. It simplifies every layer of the lakehouse—from ingestion to AI model readiness—while maintaining openness and interoperability. In this article, we break down how Acumen Vega unlocks the full power of Apache Iceberg for enterprises building intelligent, ML-powered platforms.

### 1. Open by Design: No Vendor Lock-In, Built on Open Standards

Acumen Vega is built entirely on open technologies. Apache Iceberg is at its core, but the surrounding ecosystem remains modular and standards-driven:

- **Iceberg tables stored in BigLake** (open table format on Google Cloud Storage)
- **Query from anywhere:** BigQuery, Spark, Presto, Dremio, and Vertex AI notebooks
- **Interoperable cataloging:** Vega integrates with REST-compatible catalogs like AWS Glue, Nessie, or Polaris

This means you're never locked into a specific vendor, engine, or tool. You can query your data from any compute engine and even evolve your architecture over time without migration overhead.

### 2. Seamless Data Ingestion and Lakehouse Bootstrapping

Getting data into Iceberg can be a massive hurdle. Vega automates the hardest parts:

- **Streaming ingestion support** from Kafka, Pub/Sub, or Dataflow
- **Batch migration** from Parquet, Delta, Hive, and even CSVs
- **Schema onboarding** via Vega's built-in profiler and converter

Data engineers can quickly onboard legacy and streaming datasets into fully optimized Iceberg tables, partitioned and versioned from day one. That means fewer pipelines, faster time to insight, and drastically lower ETL maintenance.

### 3. ML Interoperability Built-In: Ready for Vertex AI and Beyond

Vega isn't just for dashboards. It's designed for ML workflows from the start:

- **Native integration with Vertex AI** Feature Store and Notebooks
- **Automated snapshot versioning** for reproducible ML training runs
- **Support for synthetic data generation** using BigQuery Data Masking + TFX

Vega transforms your Iceberg tables into ML-ready assets—with full lineage, versioning, and policy control. Analysts can train models directly from Iceberg data, and data scientists can track which version of the dataset was used for any prediction.

### 4. Autonomous Optimization: Let the Platform Tune Itself

Performance tuning a lakehouse is time-consuming. Vega removes the guesswork:

- **File compaction, clustering, and metadata cleanup** run on a policy-driven schedule
- **Smart partition evolution** helps keep queries performant as data grows
- **Monitoring + insights dashboard** shows optimization health and cost impact

You don't need a full-time team just to maintain your lakehouse performance. Vega handles that automatically, helping you scale without scaling your operations team.

### 5. Enterprise-Ready Governance & Observability

With Acumen Vega, governance is not an afterthought. It's embedded:

- **Policy-driven access control** at row, column, and object level
- **Audit logs and version history** for compliance and rollback
- **Integration with Data Catalogs** like Google Data Catalog and Collibra

You can finally enforce consistent policies across your data lakehouse without slowing down users or innovation.

### Conclusion: Acumen Vega Makes Iceberg Easy, Open, and ML-Ready

Acumen Vega is more than an integration tool. It's a production-grade accelerator for organizations ready to embrace Apache Iceberg as the foundation of a modern, intelligent, and open analytics architecture.

Whether you're building dashboards, deploying real-time features, or training AI models, Vega ensures your Iceberg data is governed, performant, and ready for the future.