

Automating decisions in real time on millions of events per day

High-Integrity Data Pipeline & Automated Reconciliation

A client operating in a regulated, high-frequency environment required a system capable of processing and acting on live operational data at scale. Financial outcomes were directly tied to system behaviour, meaning delays, inconsistencies, or errors could result in immediate and measurable impact.

The platform generated millions of events per day across multiple data streams. The client needed to ingest this data, apply decision logic in real time, and maintain a complete and accurate historical record for audit and analysis. Existing approaches relied on fragmented processing and manual reconciliation, creating risk, delay, and limited traceability.

ENVIRONMENT

Regulated, high-frequency operations

SCALE

Millions of events processed per day

REQUIREMENT

Deterministic processing with full auditability

OUTCOME

Real-time automation with consistent historical replay

The challenge

The system needed to operate continuously under high load, with strict requirements for both accuracy and consistency. Because automated actions were driven directly from incoming data, any divergence between live and historical processing could lead to incorrect outcomes, misaligned records, or failed validation.

- Processing high-volume event streams in real time without delay.
- Executing automated decision logic directly on incoming data.
- Ensuring backfilled historical data matched live system behaviour exactly.
- Maintaining near-zero downtime in a continuously operating environment.

The client therefore needed a system that was not only fast, but provably correct, traceable, and resilient under sustained load.

Our approach

We designed and implemented a unified data processing system that handled real-time and historical data within the same deterministic framework. Rather than treating live processing and backfill as separate concerns, both were executed through the same core logic so the platform behaved identically in production and in replay.

- Event-driven ingestion to process data continuously as it arrived.
- Deterministic processing logic to ensure identical inputs produced identical outputs.
- Reconciliation mechanisms to align live processing with backfilled data.
- Automated workflows to remove manual validation and intervention.
- Structured logging and storage to provide full traceability of every event.

Historical data was stored in a columnar format optimised for analytical access, allowing efficient querying, validation, and reconstruction using standard data tooling.

Results

The resulting system provided a stable and reliable processing layer capable of operating continuously under high data volume. It gave the client confidence that automated decisions were being made against validated data, while also creating an audit trail that could be queried and reviewed end to end.

- Accurate real-time handling of millions of events per day.
 - Consistent outputs across both live and historical data.
 - Elimination of manual reconciliation processes.
 - A complete, queryable audit trail for every processed event.
-

What this enabled

By introducing a unified and deterministic processing model, the client moved from fragmented, reactive workflows to a controlled and repeatable operational framework.

- Reliable automated decision-making based on validated data.
 - Improved system stability under sustained high load.
 - Faster investigation and validation using historical records.
 - A foundation for ongoing optimisation and extension without rework.
-

Technical environment (representative)

- Python-based processing pipelines
- Real-time and event-driven data ingestion
- Columnar data storage for analytical workloads (for example, Parquet)
- Deterministic processing and state management
- Containerised deployment in a cloud environment