

Scaling Enterprise Data Operations

SUMMARY

A financial services organization relied on a centralized AWS data lake to securely ingest, store, and distribute financial data across the business. As usage grew, manual processes, integration failures, and resource constraints created scalability challenges. Source Allies improved key ingestion and processing workflows to increase reliability, support larger data volumes, and reduce operational overhead.

APPROACH

Our team focused on targeted improvements to established data workflows, addressing recurring failures and scalability constraints while making the platform easier to operate and maintain.

- **Streamlined Data Delivery:** Replaced an unreliable streaming-based process with direct S3 uploads, eliminating unnecessary complexity and recurring reconciliation issues.
- **Scalable Data Processing:** Introduced sharding, partitioning, and batch processing to handle multi-gigabyte files and increase concurrency.
- **Integration Reliability:** Refactored Box API interactions and introduced caching to reduce redundant calls and prevent rate-limiting failures.
- **Operational Sustainability:** Improved production documentation, created runbooks, and established documentation standards to make recurring issues easier to investigate and resolve.

SOLUTION

Source Allies modernized critical workflows within the existing AWS data platform to improve scalability and reliability. Direct S3 uploads replaced a Kinesis-based process that regularly required reconciliation and manual intervention, while sharding and partitioning enabled large files to be processed efficiently. Batch interactions with DynamoDB and HashiCorp Vault improved sensitive-data tokenization at scale, and improvements to Box integration reduced recurring ingestion failures. The team also strengthened documentation and operational practices to make the platform easier to support long term.

Results

Zero

Reconciliation mismatches after S3 migration

100%

Reduction

Daily API rate-limiting failures

~70

Repositories migrated to standardized pipelines