

Building a Scalable Data Analytics Platform

SUMMARY

A global agricultural services company needed a secure, consistent way to bring data from multiple products together for analytics and other downstream use cases. Existing data was distributed across operational databases and APIs, limiting historical analysis and requiring analytics applications to access production systems directly. Source Allies helped create a governed analytics platform that centralized historical data, documented its meaning, enabled secure sharing, and established reusable patterns for onboarding data products consistently.

APPROACH

Our team worked across multiple product teams to establish common architecture and onboarding patterns while creating a consistent experience for data consumers.

- **Reusable Data Architecture:** Designed common patterns for moving static and continuously changing data from operational systems into a centralized analytics environment.
- **Automated Data Management:** Created automation to build and maintain databases, tables, and columns from data-model definitions, reducing manual setup.
- **Governed Data Access:** Established cataloging, metadata, and secure cross-account sharing patterns to make data easier to discover while protecting sensitive information.
- **Distributed Ownership:** Developed reusable patterns and documentation that enabled individual product teams to onboard and maintain their own data.

SOLUTION

Source Allies developed a data mesh architecture that brings operational data into a centralized AWS analytics platform without requiring reporting applications to query production systems directly. Reusable ingestion patterns capture both historical and changing data, while automation simplifies the creation and maintenance of analytics structures. A documented data catalog and consistent governance model make information easier to discover, share, and use securely across teams. The platform also established a foundation for new capabilities, including an AI-powered data catalog that helps developers explore schemas, relationships, data classifications, and gaps.

Results

8

Data products onboarded

168

Analytics-ready tables

87% Reduction

Data product onboarding time

<6 Seconds

Complex analytics queries