

Cloud Agronomics Enhances Platform With Modern Data Pipeline on AWS



Executive Summary

Cloud Agronomics is a geospatial and analytics company that uses predictive analytics to support activities related to sustainable food production. The company provides a suite of analytical services, including carbon monitoring, digital agronomy, and yield predictions, to organizations in the agribusiness space. Cloud Agronomics' platform is powered by an advanced hyperspectral imaging capability that can collect granular data about agriculture systems from the air and space.

Recently, Cloud Agronomics wanted to optimize the capacity and speed of one of its data pipelines. As an Amazon Web Services (AWS) [Premier Consulting Partner](#) with extensive cloud data and analytics experience, ClearScale was a perfect choice.

"When we embarked on this complex data analytics project, we knew that we required an expert partner to ensure it was a success. ClearScale had the AWS experience and structured approach we needed to reach all of our objectives. Thanks to ClearScale, our data and analytics platform is now even more impressive. We definitely chose the right partner for our project."

Benjamin W. Leff, VP of Engineering, Cloud Agronomics

The Challenge

Cloud Agronomics' platform gathers information using two different data pipelines: a georeferencing data pipeline that produces latitude and longitude coordinates for imagery, and a reflectance data pipeline that cleans and stores images for further analysis.

To optimize the accuracy of its data, Cloud Agronomics wanted to enhance the reflectance data pipeline. The company aimed to deploy it using a cloud-native approach and implement a serverless framework while following current DevOps best practices. Additionally, the revamped data pipeline needed to allow throughput capacity management and exception handling logic.

Although the end goal was clear, the project was complex. The new reflectance pipeline had to be implemented in stages to ensure that stateflow logic and relationships were set up appropriately every step of the way. Security was also paramount, as clients used the company's data to make major business decisions. No data could leave the cloud without essential protections.

Finally, Cloud Agronomics' cloud-based reflectance data pipeline had to process massive amounts of data quickly; a single flyover produced terabytes of information. The organization needed the data pipeline to load images reliably at a rapid velocity, which is why a cloud-native solution was necessary.

Given the scope and complexity of the project, Cloud Agronomics sought an expert partner to ensure a successful implementation. ClearScale, with its AWS competencies in [DevOps and Data and Analytics](#), had the experience and knowledge to help Cloud Agronomics achieve its goals.

The ClearScale Solution

ClearScale used a multitude of AWS services to design a comprehensive cloud solution around Cloud Agronomics' new reflectance data pipeline. Ultimately, ClearScale's approach enabled the AgTech company to:

- Manage and automate complex business and scientific logic
- Secure all incoming and outgoing data
- Enhance data-related velocity

Orchestrating Complex Business Logic with AWS Step Functions

[AWS Step Functions](#) were foundational to Cloud Agronomics' new data pipeline. The service itself is a serverless function orchestrator that allows developers to define complex relationships between crucial tasks in a clear, comprehensible manner. AWS Step Functions come with an intuitive visual interface that people can use to assemble business-critical applications from other serverless functions and cloud-native services. Overall, users can modernize legacy monoliths, improve resiliency, and automate workflow execution according to unique business logic thanks to the tool's flexibility.

With AWS Step Functions, ClearScale enabled Cloud Agronomics to manage a highly complex distributed IT environment through automation. The team relied on AWS Step Functions to execute the business logic behind the data-gathering process, which included multiple steps of scientific calibration. For example, AWS Step Functions enabled ClearScale to implement a cloud-native semaphore to set concurrency limits. By doing this, the data pipeline can accelerate execution when needed or limit the number of workflows occurring in parallel to maximize performance.

ClearScale also used AWS Step Functions to connect seemingly incompatible tasks across Cloud Agronomics' IT operations. The service's asynchronous and loosely coupled nature allowed ClearScale to connect an application that needed to run on Windows with Windows-specific dependencies to a modern Python-based serverless application. Under the configuration, AWS Step Functions issues a unique token that is supposed to be returned during a designated time window (up to one year). It doesn't matter if data is supplied via SDK methods, REST APIs, or even CLI procedures. The connection also works regardless of whether the services involved are located in the cloud or on-premises, making AWS Step Functions the best solution for bringing together disparate processes.

Moreover, AWS Step Functions provided a way to catch abnormal occurrences in real-time and reprocess failed steps without affecting other tasks. The reflectance data pipeline now uses an exponential backoff algorithm to delay subsequent reattempts to increase the probability of success. If a job fails due to intermittent conditions, there is a reasonable chance that these conditions will eventually disappear and allow for successful execution.

Thanks to AWS Step Functions, ClearScale was able to define and implement the complex business logic needed for Cloud Agronomics' reflectance data pipeline and automate successful data exchanges between workflows that would otherwise be impossible. As Cloud Agronomics' IT infrastructure evolves, the organization can trust AWS Step Functions to handle any level of operational complexity and error handling, thus removing that responsibility from the development team.

Securing Data with AWS Control Tower and More

Securing Cloud Agronomics' data pipeline required several AWS services, including AWS Control Tower, AWS Organizations, Amazon EFS, and more.

[AWS Control Tower](#) is a valuable governance tool for organizations that have multiple AWS accounts and teams. The service allows users to quickly provision new accounts, automate policy management, and view high-level summaries of the guardrails implemented across AWS environments.

AWS Control Tower was instrumental for accomplishing one of Cloud Agronomics' primary objectives: implementing security isolation via environment segregation. Now, changes in one zone don't affect any other zones. The tool allowed ClearScale to set up multi-account governance according to existing best practices to isolate Cloud Agronomics' workloads at the lowest possible level. Because the company's environments are deployed in different accounts, any human mistakes have minimal impact on the overall IT environment.

[AWS Organizations](#) is a related free service that enables developers to centralize governance across AWS accounts to simplify permissions across different functions. Users can limit the actions or services that certain environments can perform while still allowing access to shared resources. With AWS Organizations, Cloud Agronomics can easily govern its environment and configure workloads as needed without compromising security.

All of Cloud Agronomics' data is encrypted at rest (with [AWS Key Management Service](#) and SSE-S3) and in transit (with Transport Layer Security). ClearScale also created a hardened Amazon Machine Image (AMI) that is used for bastion hosts and VPN access within domain-specific configurations.

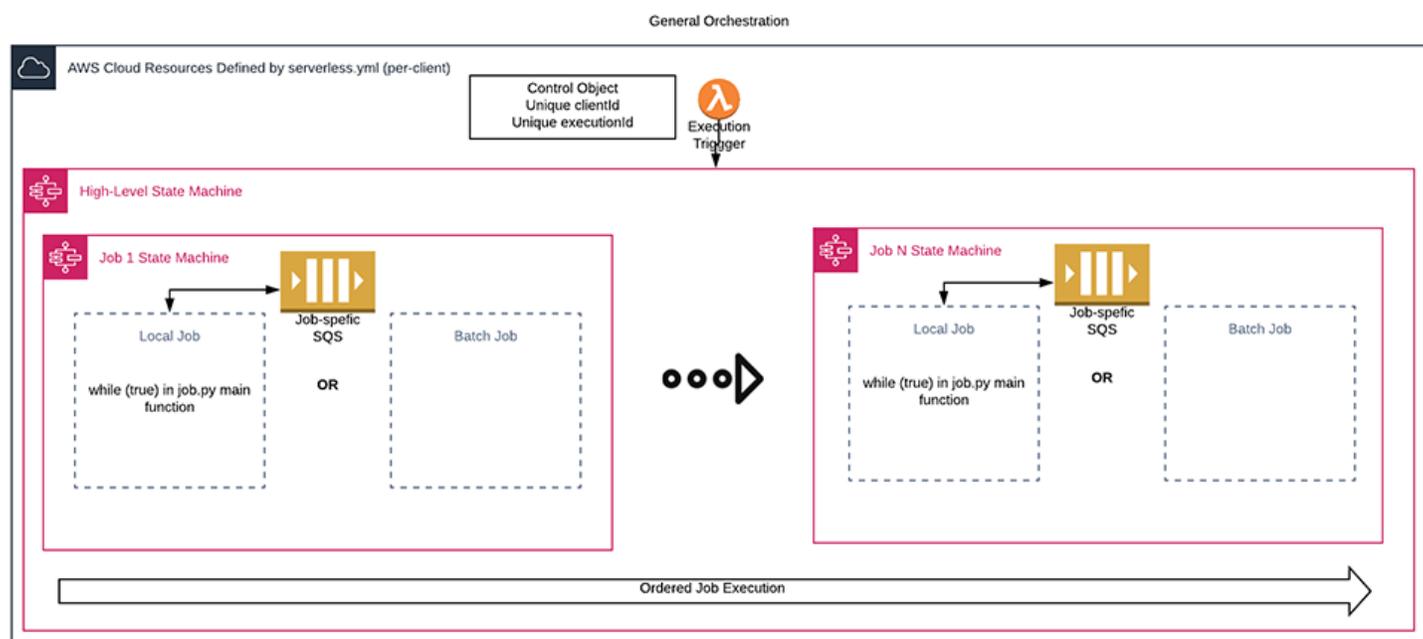
Additionally, the ClearScale team leveraged [Amazon Elastic File System \(EFS\)](#) to allow the sharing of a large scientific library without having to copy it across containers. Amazon EFS is highly reliable and resilient - if one Availability Zone fails, the library is unaffected.

Enhancing Data Velocity With Amazon S3

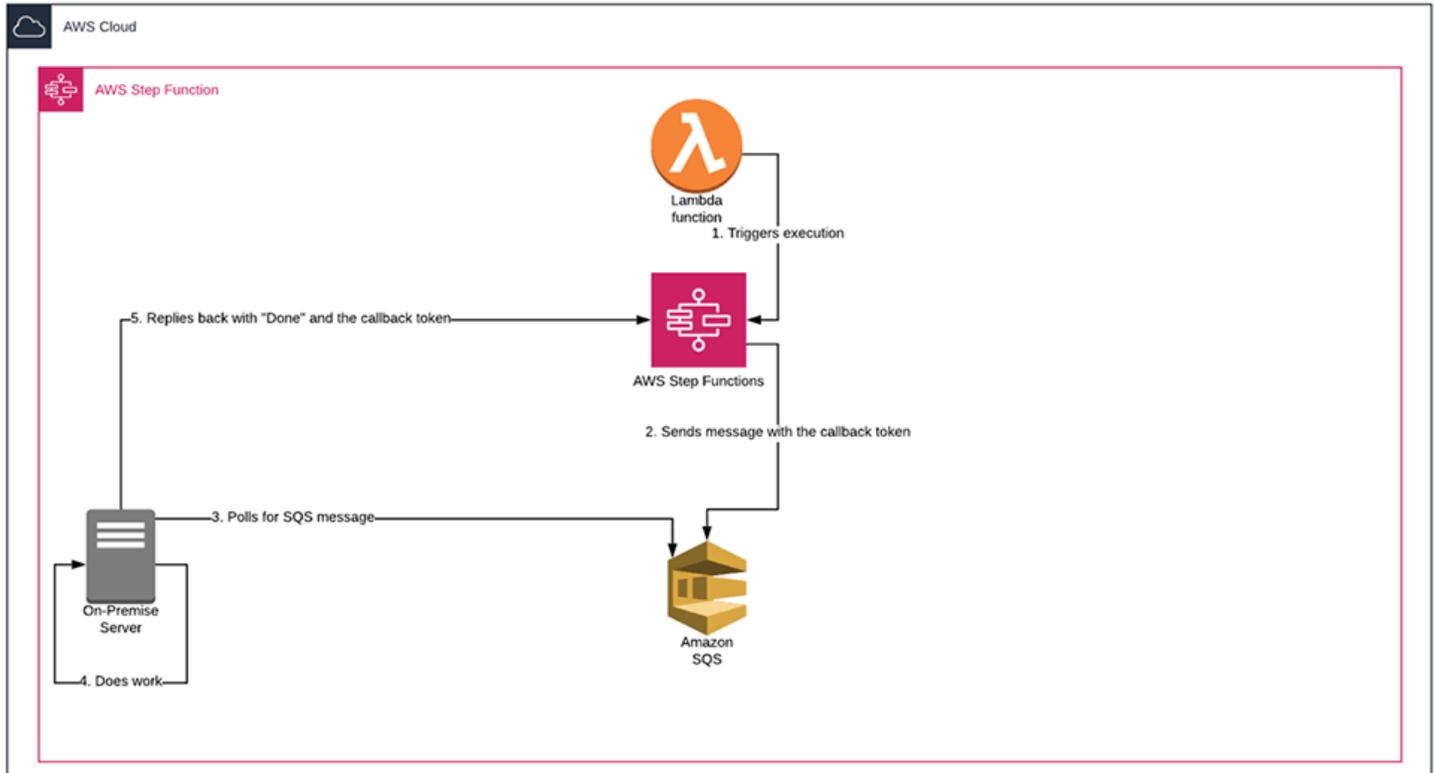
On the data velocity front, ClearScale's experts took advantage of AWS's available-on-demand cloud resources. Whenever clients need to process more data without affecting execution times, Cloud Agronomics' architecture will provision additional computational units on a per-second basis. On top of that, when Cloud Agronomics needs to transfer data to customers quickly, it can use the [Amazon S3 Transfer Acceleration](#) tool to deliver information efficiently.

Architecture Diagrams

In-Cloud Orchestration Architecture Diagram



On-Premise Orchestration Architecture Diagram



The Benefits

Today, Cloud Agronomics' platform is even more impressive, thanks to ClearScale's help. The AgTech company's reflectance data pipeline relies on AWS Step Functions to automatically manage and execute complex scientific logic to achieve business goals. The solution is also much more reliable as it benefits from AWS Availability Zone redundancies and on-demand resource provisioning.

Additionally, Cloud Agronomics benefits from pay-as-you-go computing that frees the team from having to estimate capacity needs in advance or buy more than necessary. By moving away from on-premises infrastructure, Cloud Agronomics was able to save money, even while gaining modern, cloud-native capabilities.

Finally, Cloud Agronomics' data is much more secure in the cloud than in on-premises infrastructure. The company can rely on AWS to mitigate cyber attacks and encrypt all data loaded from its two data pipelines.

With its new reflectance data pipeline, the first of its kind, Cloud Agronomics can deliver higher quality data to customers in the agribusiness space and further differentiate itself from the competition. The company can trust AWS to execute the complex business logic behind its innovative solution and optimize performance over the long term.