

Building a Serverless ETL pipeline for Groove (Carmen) with AWS



Executive Summary

Groove builds cloud-based sales and CPQ toolsets for complex IT and Telecommunication sales teams. Carmen is a core component for Groove Systems that enables complex location awareness for the rest of the Groove products. Carmen is a collection of GIS capabilities that enable its customers to host large cadastral datasets, perform geospatial queries, and host interactive maps for location services.

Challenge

A foundational capability requirement for the Carmen GIS solution is the ability to maintain and process cadastral data sets collected from various telecommunication infrastructure providers. Most developing countries face severe issues with data integrity, cleanliness and accuracy in the cadastral landscape, which makes for a challenge when trying to build value-add geospatial layers onto the existing datasets. Groove required a robust and scalable ETL solution, with stages for data acquisition from multiple sources, data cleaning, standardization, and injection into geospatially enabled databases. Groove and the Carmen teams could not afford to spend time on server maintenance and preferred a fully managed solution.

Why AWS?

AWS offers modern services that allow customers to build highly scalable, performant, reliable, secure and cost optimized ETL pipelines. A purely Serverless architecture can be achieved by using services like AWS Lambda and AWS Glue, with management of servers completely abstracted away from the customer.

Groove chose Tangent Solutions for its strong values, and obsession with customers – a value shared by AWS. Tangent is home to a team of highly skilled technology practitioners, from cloud architects to developers and project managers, offering trust and reliability end to end. Being an Advanced Consulting Partner with a focus on Serverless Technologies, Tangent has the experience and knowledge to solve problems, while simultaneously identifying new opportunities.

About Groove



Groove is a SaaS-model company, addressing the primary needs and painpoints of telcos in the modern age. It is an automated factory which produces high-quality proposals and quotations for multi-site enterprise solutions at revolutionary speeds



Solution

The solution Tangent provided to Groove is comprised of the following AWS resources:

- AWS Lambda
- AWS StepFunctions
- AWS API Gateway
- AWS ECS Fargate
- AWS S3

All these services were used in combination to achieve the solution, with AWS Lambda being the core enabler:

- Lambda: Decoupled systems of Lambdas govern the fetching of data from various different endpoints; processing files of different formats; massaging data into standard formats; adding labels and tags; routing and storing data dynamically to S3 buckets.
- StepFunctions: Used to orchestrate hand-offs between a set of Lambda functions.
- API Gateway: The resulting processed data is exposed to other systems via a REST endpoint wrapper.
- ECS Fargate: Some serverless functionality required long-running compute as well as custom Python run-times and packages. These workloads are maintained in a Docker container, which gets run and hosted by ECS in a serverless fashion.
- S3 serves as the target location for all processed data records, essentially forming a data lake for geospatial data layers.

The result of this solution is a serverless ETL pipeline capable of scaling towards the growing needs of Groove.

Benefits

With a focus on serverless technology, Groove manages to save costs during idle time and low data traffic periods, while being dynamically scalable to service virtually any data load. What would normally have taken months, is now achieved within hours, in a completely automated fashion. Staff members of Groove would traditionally fetch HDDs by driving to different offices, copy data over to local workstations, process the data manually, and deliver processed data via HDDs, emails, or access to their Databases. Now, they only have to care about exception handling, and data validation.

About Tangent

Tangent Solutions (a Division of Jurumani Solutions) is a Born-in-the-Cloud Development and Cloud Consultancy company. We are an energetic company taking an innovative and disruptive approach on digital transformation, from software design to intelligent automation and Cloud. We implement technology, processes, and methodologies that make it easy for our customers to produce high quality products.

