Improved Scalability, Availability, and Reduced Infra Costs of ML Platform Using AWS Serverless Architecture

Customer Overview

The customer, a healthcare technology startup, is focused on building solutions that empower people to proactively manage their health. Through insights derived by boosting available health signals, their solutions help take informed health decisions, driving better outcomes through proactive care.

The Business Context

The customer required a multi-component solution that included a highly scalable platform and a high performing health app for end users. The solution had to collect health data from users across the globe and leverage data science and machine learning to diagnose health issues and generate predictive insights. Since the timing and scale of sample collection and processing was highly unpredictable, keeping the services on at all times would not have been cost-effective. So, the solution had to be highly flexible and scalable based on demand. Also, with large scale global data collection being an ongoing activity, features extracted from the data and meta data were to be used to train the ML models. Other apps from the ecosystem also had to plug into the platform to access the data.

Type of Service Provided

Product and Platform Engineering

Technologies Used

AWS - Lambda Functions, CloudFormation Templates, S3, RDS, PostgreSQL

The Solution

The team of product engineering, cloud, data science, and AI/ML experts from GS Lab |GAVS meticulously crafted the platform to collect health samples and apply machine learning to extract critical information. Considering the need for the platform to cater to unpredictable demand, the team decided to use AWS serverless components, microservices, and SAM (Serverless Application Architecture). This reduced the resources required to develop and maintain the platform and made it highly modular and flexible while also conforming to regional regulatory and data protection standards with minimal effort. Decoupling of ML model training and ML infused solution architectures helped save costs and compliance with data related government regulations.

To find out how GS Lab | GAVS can help your organization, please visit
www.gavstech.com

Challenges

- Early adopter difficulties - learning curve, language specific object/data initialization, lot of experimentation and deep troubleshooting for issues

- The following were needed to meet solution requirements:
  - Modular and services-based architecture for entire pipeline
  - Robust architecture for continuous availability
  - Flexibility and quick scalability since demand was unpredictable
  - Speedy responses from health app
  - Accessibility from anywhere and at any time
  - Solution components created with best suited languages and tools
  - Interdisciplinary team with expertise in multiple languages
  - Tight integration of individual solution components
  - Adherence to local, regional, and global regulatory and law enforcement frameworks
  - Optimal solution with minimal resource investment

Solution Highlights

- Robust platform and high performance health app
- Best-in-class tools and languages to develop individual components
- AWS components used:
  - Lambda functions for almost all components
  - CloudFormation templates
  - S3 for secure storage
  - RDS PostgreSQL for database needs

Solution Impact

- Flexible solution with rapid scalability and high availability
- ~40% lesser infrastructure costs with serverless architecture with medium resource utilization; higher savings for higher loads
- Ops-light solution – only 1 DevOps engineer required for maintenance
- DevOps related savings as patches or upgrades not required for EC2 machines
- Seamless integration of disparate systems made possible by serverless architecture
- Easy adherence to regional and global data protection and regulatory frameworks through modular architecture