ACR Develops Data Architecture Prototypes for Future Data Lake Implementation

Executive Summary

The American College of Radiology (ACR) is dedicated to advancing the science of radiology for the benefit of patients and society at large. Since its founding in 1923, the organization has grown to represent more than 40,000 radiologists, radiation oncologists, interventional radiologists, medical physicists, and nuclear medicine physicians.

ACR was seeking a long-term partner for a large data architecture project to bolster its data management capabilities and growth potential through cloud utilization. ClearScale stepped in to design a new architecture and templates that would serve as the foundation for follow-on solutions.

"In this new era where 'Data is Gold', having access to all data assets in a single, scalable environment is key to any business's success. ACR was looking for a partner who could help us create a secure data lake infrastructure in the cloud to realize our true analytical capabilities quickly. ClearScale helped us take the first step into our cloud journey by creating a secure landing zone in AWS within weeks. Now this serves as a solid platform not just for our analytics and reporting needs but also for our new applications."

Shree Periakaruppan, Director of Data Engineering and Analytics, ACR

The Challenge

ACR wanted to take its IT data infrastructure to the next level by implementing a data lake solution in the cloud. The organization identified several areas of its legacy data architecture that it wanted to improve.

First, ACR had instances of duplicative data across multiple environments, which meant that conducting data analyses could be time-consuming. Individuals had to perform duplicative tasks, comply with multiple standards, and refer to various data sources for data analytic needs. Multiple copies also meant excess storage and processing costs.

For cross-domain analyses, ACR employees would have to independently contact disparate teams or groups to gain access to data assets, and there was no single data catalog with comprehensive metadata available to facilitate an efficient research and discovery process across all data assets.

From a processing perspective, data integration tasks could require multiple data transformations through ETLs, sometimes combining numerous data sets to produce new repositories. Backtracking from completed analyses in order to understand the origin of field-level and source-level data in these cases could be cumbersome and time-consuming.

On the security front, ACR wanted to implement column-level permissions within various environments to offer increasingly fine-grained access to particularly sensitive data that are governed by a variety of regulatory standards (e.g., PCI DSS, HIPAA, etc.).

To accomplish this next generation solution, the ACR planned to leverage a cloud-based data lake infrastructure. The organization wanted to create a Landing Zone as the first step in the process. Once completed, ACR would then move forward with a data lake MVP.

The organization realized that its best path forward was to partner with a cloud services provider that could provide guidance over a long-term data architecture project. ClearScale, an AWS Premier Consulting Partner, fit the bill perfectly.

The ClearScale Solution

ACR asked ClearScale for a "Cloud Landing Zone" prototype to serve as a foundation for further expansion. ClearScale began the engagement by reviewing ACR's business requirements and existing architecture. ClearScale's experts worked with participants from every function within ACR's IT department to discuss use cases, compliance needs, and the overall cloud roadmap that would need to be supported by the proposed landing zone. Given the nature of ACR's business, HIPAA compliance and robust cloud governance were essential design criteria.

Based on the information collected in these sessions, the ClearScale team was able to design a cloud landing zone according to AWS best practices. ClearScale implemented AWS Control Tower, which enables users to set up and govern multi-account AWS environments. The service makes it easy for builders to provision new accounts without compromising policy compliance, as well as write guardrails that provide ongoing governance.

The design also involved Organization Units (OUs) so that ACR could group accounts that are designated for HIPAA-compliant workloads. That way, the group can easily manage user access to protected data and provide logical separation from other types of workloads.

To help automate governance, ClearScale implemented service control policies (SCPs) and AWS Config rules that allow ACR to apply governance at every level of the landing zone. In addition, the ClearScale team created a custom set of SCPs and config rules to help enforce cloud governance according to ACR's specific cloud objectives.

The Benefits

With ClearScale's help, ACR was able to develop a landing zone that would serve as the starting point for the organization's cloud infrastructure.

ACR was able to logically separate HIPAA-compliant workloads from other workload types to assist in securing the data. The organization's IT team can now apply automated cloud governance controls across its landing zone, saving both time and resources without compromising functionality.

It is now easier to implement user access controls for protected data, thus reducing both the risk and time-intensive nature of conducting complex analyses across multiple datasets. With centralized logging, ACR has a single, integrated source to track activities, which makes it easier to perform diagnostics and trace activity.

Looking ahead, ACR has a powerful foundation on which to grow its cloud infrastructure.