Proof of Concept - Automated Masking Pipeline Increases CareCentrix Productivity

carecentrix.

Executive Summary

Connecticut-based CareCentrix, Inc. ("CareCentrix") is a provider of health-at-home solutions that help members transition home after an acute episode, or who have long-term needs to heal and age at home. In addition to improving health outcomes, CareCentrix reduces costs by coordinating each individual's care journey using purpose-built technology and analytics to optimize care. CareCentrix health-at-home solutions also include home health, infusion therapy, sleep testing, and durable medical equipment (DME) management.

As a service provider, CareCentrix handles protected health information (PHI) and personally identifiable information (PII) on a daily basis. In connection with its services, in certain situations, CareCentrix needs to share records that contain PHI or PII with third parties but needs to remove the PHI/PII prior to sharing. CareCentrix teams manually review these records to ensure PHI and PII is not present before sharing. CareCentrix engaged in this ClearScale proof of concept to potentially identify a more streamlined process. After evaluating three potential partners, CareCentrix identified ClearScale as the partner due to its cloud, healthcare, and machine learning expertise.

> "ClearScale was a great partner throughout this Proof of Concept. They took the time to understand our use cases, challenges on the ground, and recommended a solid, scalable solution. The ClearScale team was solid in their understanding of the technical solution and ensured all milestones were met on time. Our organization was pleased with the team and execution of the Proof of Concept".
>
> **Cynathia Foreman,** CCX Senior Director IT Delivery, CareCentrix

The Challenge

In connection with its services, in certain situations, CareCentrix needs to share records that contain PHI or PII with third parties but needs to remove the PHI/PII prior to sharing. CareCentrix employees manually redact the PHI / PII information from the records to ensure the PHI/PII is masked appropriately before sharing. This process may create performance and scalability challenges as volumes increase.

CareCentrix wanted to test a more streamlined approach for information sharing and challenged the team to identify a Proof of Concept (POC) that introduced efficiency, scalability, and automation for manually transcribing, masking, and sharing records. Rather than build the functionality in-house, CareCentrix chose to bring in a cloud expert who could complete a POC to demonstrate alternatives to manual redaction of PHI / PII data. After review of several Amazon preferred partners, CareCentrix selected ClearScale due to its extensive cloud experience and AWS Competency in Healthcare.

The ClearScale Solution

Based on CareCentrix's business requirements, ClearScale determined the best approach for the POC would be to leverage serverless solutions and managed services as much as possible. Doing so would maximize scalability and free CareCentrix engineers from having to worry about infrastructure management down the road.

After reviewing CareCentrix PHI / PII data masking requirements and objectives, ClearScale identified AWS Step Functions, a serverless microservices orchestrator service, for the heavy lifting. This service makes it easy for users to sequence AWS Lambda functions and integrate other AWS features into existing workflows to build highly sophisticated applications as needed. With AWS Step Functions, ClearScale could replace multiple manual processes with AI-powered algorithms without compromising accuracy.

ClearScale and CareCentrix worked together to create a set of mock (fake) test data that was representative of an appropriate sample size/complexity for file processing.

Next, ClearScale set up a pipeline that executed the steps below in order:
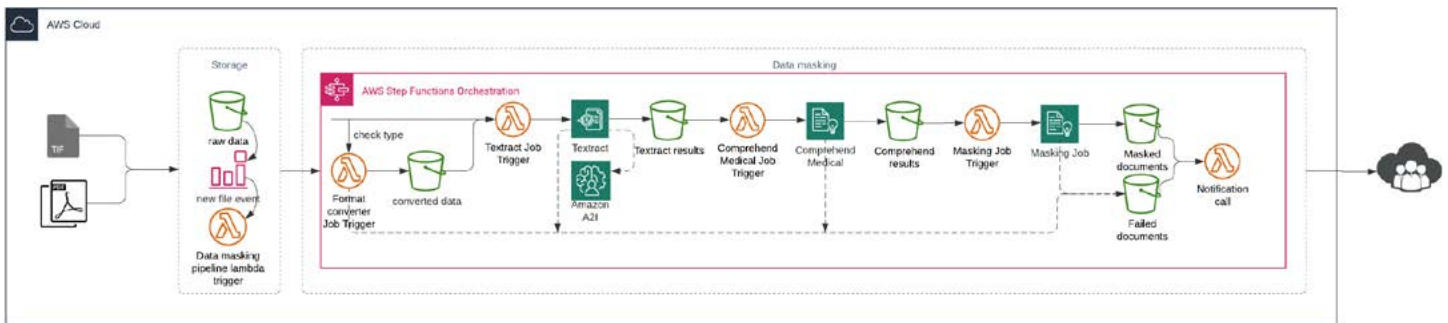- Process a new file event
- Convert the new test file, if necessary
- Extract text from the test file with Amazon Textract
- Identify PHI/PII in the extracted text using Amazon Comprehend Medical
- Mask all PHI/PII based on text entry coordinates located by Textract
- Save new test file version to Amazon S3
- Notify user when finished

The two AWS services at the center of the masking pipeline, Amazon Textract and Amazon Comprehend Medical, use machine learning to output probabilistic results with confidence scores.

To mitigate any risk of Amazon Comprehend Medical missing PHI/PII, ClearScale masked all mock test data flagged as potentially sensitive, even if it returned low confidence scores. On the Textract side, ClearScale's developers created opportunities for manual user intervention. If corrections need to be made, CareCentrix employees create an Amazon Augmented AI (A2I) job to review the Textract prediction (i.e., output), make manual corrections, and then resume the pipeline. By setting up the pipeline in this way, ClearScale created an opportunity for CareCentrix's team to review AI outputs for accuracy whenever appropriate.

Between the managed machine learning services and serverless pipeline orchestrator, CareCentrix gained a POC capable of automatically scanning records, identifying sensitive information, masking protected content, and saving a new version for downstream distribution. The upshot is that CareCentrix could meet compliance obligations by automating processing thousands of documents daily.

Architecture Diagram

The Benefits

ClearScale helped CareCentrix validate the benefits of automated data masking workflow for sensitive documents, and proved the new pipeline could reduce the time associated to manual transcribing and masking PHI/PII. ClearScale estimates that CareCentrix can achieve 3x efficiency improvement in the clinical documentation redaction process with an automated masking pipeline to achieve greater cost savings.