

RosettaHealth achieves cost-effective Storage Scalability on AWS with Stelligent

RosettaHealth is a software company providing solutions for secure, standards-based exchange of healthcare records between organizations using HIPAA compliant processes and technologies. Stelligent, a Division of Hosting, is a company that provides its customers with automated Continuous Delivery solutions in Amazon Web Services (AWS). Working together, Stelligent designed an infrastructure platform with Amazon Simple Storage Service (S3) using Lifecycles for storage and Amazon Redshift for analytics, designed to match business needs with HIPAA security and do so in a scalable, cost-effective manner.

Challenge

Structured health records (medical and patient) may contain Protected Health Information (PHI), and this data format represents the bulk of RosettaHealth's total data stream. These records must be maintained, both at rest and in transit, in a HIPAA compliant way yet remain both searchable and rapidly retrievable. The standards-based format for said documents is XML and the original source document must be stored as-received alongside transformed versions created to support search indices.

In addition to health records, the RosettaHealth solution stores application support data and logs, in the form of transactional event data, in the same datastore as the health records. The datastore is currently a 3rd-Party hosted MongoDB solution, providing both storage and search options. Log data is not typically used after 3 months but must be archived and retained.

RosettaHealth needs to balance features with costs of storage in a scalable way, while also providing archival processes (including removal) that meet operational and regulatory requirements. To stay focused on providing customer value, RosettaHealth prefers to use a [hosted MongoDB database solution](#) instead of architecting the infrastructure for, and administering, the database.

Finally, two distinct data storage models exist that we can refer to as, for simplicity, the 'bus' and 'repository' models. The bus model describes short-term (on the order of 30 days), intermediary storage purely for health record exchange. The evolving repository model describes long-term storage providing for enhanced search and analytics capabilities. Both models should be supportable by any proposed solution, with the repository model driving this case study.

Solution

RosettaHealth performs custom data transformational processes upon the ingestion of clinical data records, including conversion from a complex standards-based XML format to a JSON format, as well as the creation of custom indices to support probabilistic searches for clinical data in MongoDB. To reduce infrastructure cost and complexity, the combined use of Elasticsearch for enhanced search and Amazon DynamoDB for persistence was discussed as a replacement for MongoDB. However, since Amazon Elastisearch is not HIPAA compliant in AWS at this time, that combination isn't operable for Personal Health Information (PHI). Instead, we recommended staying with MongoDB and utilizing the [AWS MongoDB reference architecture](#).

Given the logical separation of the original XML documents from the searchable data in MongoDB, we recommended storing all such documents of origin in Amazon S3 buckets instead of MongoDB. The proposed mechanism enforces encryption at rest via Amazon S3 bucket policies, and provides an Amazon S3 object's Amazon S3 URL to the MongoDB record as point of reference for rapid retrieval. A AWS CloudFormation template is provided demonstrating the general principles for both bus and repository storage models, but Stelligent recommends that RosettaHealth's production implementation include the following:

- Creation of a KMS key with which Amazon [S3 encryption-at-rest](#) should be performed
- Bucket policy enhancements [limiting access to only specific IAM Roles](#)
- AWS IAM Role modifications for Amazon EC2 instance roles requiring access to List*, Get and Put actions on the bucket.

Document storage solutions were recommended to match both bus and repository models. With the clinical data bus model, documents are stored in a bucket such that when business logic dictates that RosettaHealth delete its copies of the data, corresponding documents stored in the bus bucket are deleted in parallel via the AWS SDK. Additionally, by leveraging Amazon S3 lifecycle policies, bus bucket objects can expire after 30 days or other pre-defined period.

With the clinical data repository model, documents of origin are stored in warm storage via Amazon S3 for as long as necessary, with the option to configure Amazon S3 lifecycle policies for both Infrequent Access and Amazon Glacier storage, and configure aging policies independently for different clients in a manner consistent with HIPAA regulations.

In addition to clinical data, the log data in MongoDB represented an ever-growing load of potentially archivable data. We have recommended periodic export of log data to S3, with query support for this non-PHI data using Athena and lifecycles policies to transition the log data to either cold storage or deleted via expiration.

Finally, enhancements to the analytical query capabilities were proposed using a dimensional data model and importing structured files into Amazon Redshift. As this was tangential to the storage solutions, infrastructure and data-process details were not provided at this time.

Results/Benefits

The proposed solution offers several key benefits for RosettaHealth:

- Reduction in large-volume data hosted by 3rd party MongoDB database provider
 - Said data is instead stored in Amazon S3, leveraging S3-native archival capabilities for warm and cold storage via lifecycle policies
- Reduction in storage costs for clinical data repository needs
- Analytics capability for enhanced query mechanics via Amazon Redshift using a dimensional model
 - Supports both patient search and extended research needs

About Stelligent

Stelligent is an APN Advanced Consulting Partner and hold the AWS DevOps Competency. As a technology services company that provides DevOps Automation on Amazon Web Services (AWS) Cloud, we aim for “one-click deployment.” Our reason for being is to help our customers gain the ability to continuously deploy their software, when they want to, and with confidence. We’ve been providing DevOps Automation solutions on AWS since 2009. For DevOps best practices, Follow @Stelligent on Twitter. Learn more at www.stelligent.com or contact Stelligent at 888-924-4539.