Securing OHDSI on AWS for HIPAA and Research Data Management Compliance

Michael Lubke1, Tapati Mazumdar2, Murat Sincan, M.D.1*, Catherine Hajek, M.D.1*
1. Sanford Imagenetics, Sioux Falls, South Dakota
* Equally Contributed

Abstract

Requirements for HIPAA and research protocol compliance should be at the core of a healthcare organization’s implementation practices when storing and analyzing healthcare data in the cloud. Every utilized service must be thoroughly vetted to ensure security best practices and data access permissions are being established properly. This process results in a secure environment where organizations can embrace utilizing the cloud for performing analysis on observational health data.

Methods

The default OHDSI on AWS environment was analyzed and the following were identified as areas needing additional security:
- Restrict access to OHDSI instance to the Sanford trusted network
- Ensure all health data is encrypted at rest and in-transit
- Restrict access to study data to applicable persons and systems outlined in research data management plan

A site-to-site VPN solution was implemented that creates a tunnel between the on-premises trusted network at Sanford Health and the Virtual Private Cloud (VPC) on AWS. This combined with leveraging the AWS Certificate Manager to handle the generation and application of SSL certificates on the OHDSI on AWS environment enforce HIPAA compliance requirements by ensuring that the communication between all system components within the VPC are encrypted in-transit.

Data access policies were configured for each AWS S3 bucket restricting accessing the data to individual members of each research study. Furthermore, in order to meet HIPAA requirements, all S3 buckets are encrypted at-rest. Policies were created for AWS System Roles that allow the components of the custom OMOP CDM ETL pipeline to decrypt the contents of the S3 buckets. The transformed research data is then inserted directly into the corresponding study schema on Amazon Redshift via a JDBC connection, which ensures that the data is once again encrypted in-transit.

Results

Figure 1 illustrates the site-to-site VPN solution that was established between the AWS VPC and a routing device at the Sanford data center. This configuration resulted in the default public-facing applications in the OHDSI on AWS environment stack (Atlas, RStudio, Jupyter) being accessible via the Sanford trusted network only.

Figure 2: Research Study ETL Pipeline Dataflow Diagram

Conclusions

It is imperative that every interaction with patient health data hosted in a cloud environment is thoroughly analyzed to ensure that HIPAA and research study compliance requirements are met. Each component added to the cloud environment must be carefully implemented to ensure that the data is secured regardless of where it exists. This process can be daunting to undertake when an entire data processing pipeline has already been established without initially giving any thought to access controls and encryption practices. For that reason, employing best practices in accordance with HIPAA and data management compliance requirements at the onset of development can provide a safeguard for the cloud environment as well as the data stored within.

References: