Building a HIPAA Compliant Data Lake on AWS

with Eliza Corporation

About Eliza Corporation

Since 1998, Eliza Corporation has developed healthcare consumer engagement solutions to address some of the industry's greatest challenges – from adherence, to prevention, to condition management, to brand loyalty and retention.

"Pay-for-performance" in healthcare incentivizes payers and providers to keep a population under their care healthier. The Pay-for-performance arrangements provide financial incentives to hospitals, physicians, and other healthcare providers to carry out such improvements and help achieve optimal outcomes for patients. This is a departure from fee-for-service, where payments are for each service used. Eliza Corporation, focuses on Health Engagement Management, and acts on behalf of healthcare organizations (e.g. hospitals, clinics, pharmacies, insurance companies, etc.) in order to engage people at the right time, with the right message, and in the right channel to capture relevant metrics to analyze the overall value provided by Healthcare.

Their Challenge

Eliza Corporation analyzes more than 300 million outreaches per year, primarily through outbound phone calls with Interactive Voice Response (IVR,) technology but rapidly growing channels such as SMS, email, and in-bound IVR.

For Eliza, the outreach results are the questions, the responses form a decision tree, and each question and response are captured as a pair, E.g.

<question, response> = <"Did you visit your physician in the last 30 days?" , "Yes">

This type of data poses challenges in processing and analyzing – for example, you can't have a table with fixed columns to store it.

The majority of data at Eliza takes the form of outreach results captured as a set of <attribute> and <attribute value> pairs. Other data sets at Eliza include structured data. This data is received from various systems that include customers, claims data, pharmacy data, Electronics Medical Records (EMR/EHR) data, and enrichment data. Needless to say, there is a considerable amount of variety and quality considerations within the data that Eliza manages.

Eliza's organic growth resulted in an incumbent data platform and strategy that were architected to store multiple copies of data across multiple systems, and allowed for variation and modification, which in turn created inconsistent versions across the company. This resulted in higher platform costs because multiple systems contained multiple copies of the data. Their architecture created significant scaling and quality control issues for the company because of a lag between the time something happened and the time to derive value. These problems combined limited Eliza's growth and scale as the complexity of their platform meant extended times for evolution.

Building a Data Lake with the Help of NorthBay Solutions

NorthBay, an Advanced AWS Partner Network (APN) Consulting Partner and AWS Big Data Competency Partner, was ultimately chosen as the Big Data partner to architect, implement data storage, process infrastructure and improve the overall performance of the Eliza's process when analyzing massive amounts of data. Eliza originally contacted NorthBay to seek help and guidance from an organization that had deep expertise on AWS and had familiarity and experience working with healthcare organizations. NorthBay was one of the first APN Consulting Partners to hold the AWS Big Data competency and their ability to consider and discuss various best practice approaches to solve the data volume, varity and compliance requirements for Eliza's use case quickly earned them Eliza's trust.

After evaluating Eliza's use case and business needs, NorthBay architected a Data Lake on AWS, adhering to best practices recommendations, such as creating self-documenting and intuitive data paths.

- General benefits of the Data Lake architectural pattern include:
- Decoupling Storage and Compute
- Ingesting & Storing original datasets in native or close to native format
- Allowing for both real-time and batch processing
- Consumption of datasets from storage with variable compute
- Providing metadata, catalog and data discovery for content in the data lake
- Enabling access to data through entitlements and access control.

Solving Key Business Challenges on AWS

The key challenges Eliza needed to address arise out of key compliance requirements like the Health Insurance Portability and Accountability Act (HIPAA), the variety of the data ingested, and the need for a common view of the data.

Meeting HIPAA Requirements

AWS enables customers and partners to build HIPAA-compliant applications. Based on the requirements around Encryption in transit and Encryption at rest and following guidelines mentioned in the whitepaper, as well as various HIPAA compliance guidelines, a number of steps were implemented by NorthBay to ensure the Data Lake was HIPAA-compliant, including spinning up Amazon Elastic MapReduce (EMR) in a dedicated Virtual Private Cloud (VPC), encrypting and decrypting data when needed, and launching a data pipeline orchestration process.

EMR resources are provisioned in dedicated VPCs, and most of the processing is done on transient clusters which leverage spot/reserved and on-demand instances. In addition, a long-running cluster was also provisioned for ad-hoc analysis of data. To make the real-time streaming data ingestion HIPAA-compliant for Eliza, NorthBay leveraged Amazon Kinesis Producer Library to encrypt the data prior to putting it in Amazon Kinesis, and then decrypting it before putting it into Amazon Simple Storage Service (Amazon S3). NorthBay also launched a data pipeline orchestration process, which in turns accesses resources in a dedicated VPC.

Data Obfuscation, Data Cleansing, and Data Mapping

To meet Eliza's interpretation of protecting data under HIPAA, NorthBay established a business rule that when dealing with PII (Personally Identifiable Information) and PHI (Personal Health Information) data. In non-production environments, the PII must be obfuscated or masked before it can be shared with the development teams. Considering the volume and velocity of the data, the obfuscation task itself became a Big Data

problem. To solve this problem NorthBay helped develop an algorithm and data map that reads the data and applies the corresponding obfuscation to protect the data. The data map also provided Eliza a common view across all of their data sources, an issue that they had been struggling with previously.

The data received by Eliza is populated by disparate systems and can include free-form entries by consumers/customers, creating inconsistencies among each entry. NorthBay helped Eliza implement an additional process to cleanse the data and bring it to a common format. The schema structure that was put in places allows Eliza to apply multiple data cleansing rules on the same field and choose the order in which the rules are applied.

Benefits

With NorthBay, Eliza now has the ability to consume any type of data (structured, semi-structured, and unstructured) at any scale. With support for their entire storage lifecycle (frequent, infrequent & archived) they are able to store raw copies of input data and have eliminated the need to enforce schema during data load processes. NorthBay also provided Eliza with a Data Catalog and Seach capability, which enables Eliza to access their data quicker and easier than before. Unlike with a traditional data warehouse, NorthBay on AWS enabled Eliza to decouple the resources and therefore the costs associated with compute and storage.

General benefits of the Data Lake architectural pattern include:

- Decoupling Storage and Compute
- Ingesting & Storing original datasets in native or close to native format
- Allowing for both real-time and batch processing
- Consuming of datasets from storage with variable compute
- Providing metadata, catalog and data discovery for content in the data lake
- Enabling access to data through entitlements and access control.

Conclusion

By building a Data Lake to streamline data management, Eliza can quickly react to their clients' pressing needs to prove results in closing gaps in care, retaining membership, and adhering to medications which support the outcomes base model.

ADVANCED TECHNOLOGY PARTNER

BIG DATA COMPETENCY