Store, Protect, Optimize Your Healthcare Data in AWS

Healthcare reform, increasing patient expectations, exponential data growth, and the threat of cyberattacks are forcing healthcare providers to re-evaluate their data management strategies.

Healthcare providers are responding to these demands by leveraging AWS to enforce lifecycle management, employ cost effective, scalable storage and un-trap data for operational and quality insights.

Key design criteria for managing healthcare data should include:

- Flexible ingestion of data
- Storage for a variety of data types
- Data resiliency including backup and DR

- Ability to protect and secure critical data
- Provide and maintain HIPAA required audit controls
- Secure networking environment

In order to help customers, the following diagrams are reference architectures for managing healthcare data on AWS, including data ingestion and storage management using AWS services.

Reference Architecture

Store, Protect, Optimize Your Healthcare Data in AWS

STEP 1: INGESTION

Data ingestion into AWS can be batch or stream or hybrid.

Batch

- One time large data Migration: Use AWS Snowball to transfer large datasets into AWS
- Use AWS Glue to Extract, Transform and Load batches of data from a database into AWS
- Build a custom data ingestion application using AWS SDK

Stream

- Utilize Amazon Kinesis (Data Streams or Data Firehose) for streaming datasets into AWS

Hybrid

- Use AWS Storage Gateway (file, volume or tape) for continuous asynchronous data integration with AWS

STEP 2: STORAGE

Once the data is ingested, customers have various options to store it on AWS. Data storage on AWS can be block storage, object storage or databases.

- Block Storage: AWS provides Amazon Elastic Block Storage (EBS) and Amazon Elastic File System (EFS) to store block level data on AWS.
- Object Storage: Amazon Simple Storage Service (S3) is an object storage service that is highly durable and scalable for storing large volumes of object level data.
- Databases: Customers can use Amazon Relational Database Service (RDS), Amazon Redshift, Amazon Dynamo DB that caters to multiple database storage requirements.

STEP 3: PROTECTION

To protect the data stored on AWS, customers can utilize multiple security features on AWS:

- Encryption at rest and transit: It's a good practice to encrypt your data at rest and in transit. AWS services such as Amazon S3 and Amazon EBS have native support for encryption. Moreover, you can build encryption in transit by utilizing Encrypt data in transit using IPSec ESP and/or SSL/TLS.
- Monitoring: Customers can monitor their applications using Amazon Cloudwatch metrics and Cloudwatch logs and utilize Cloudwatch events for notifying stakeholders for unwarranted activity.
- Controlled access: Customers can utilize AWS Identity and Access Management (IAM) and S3 bucket policies to grant controlled access to AWS resources.
- Amazon Macie is a security service that allows customers to discover, classify and protect data on Amazon S3. It uses Amazon Machine Learning to find and alert about security threats. For example, customers can use Macie for scanning for PHI leakage or notifications about unusual access to data with PHI.
- For more details about Amazon Macie, please refer to the following link: https://aws.amazon.com/macie/
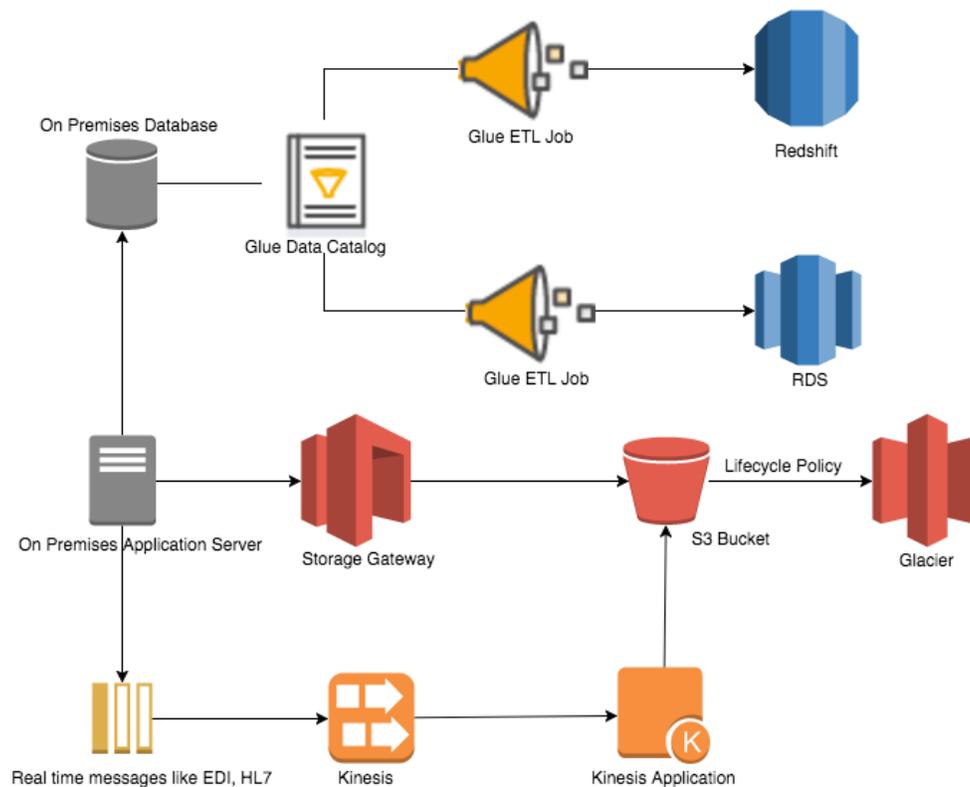
STEP 4: OPTIMIZATION

- Once the data is stored and protected on AWS, customers can utilize AWS analytical services to optimize the data and build analytics.
- Services like Amazon Redshift, Amazon Athena, AWS Glue, and Amazon S3 allow you to build robust analytical software on structured datasets while Amazon Elastic Map Reduce (EMR) provides a managed Hadoop environment for processing large volumes of unstructured data for machine learning applications.

Data Ingestion

On Premises Database

Glue Data Catalog

Glue ETL Job

Redshift

Glue ETL Job

RDS

On Premises Application Server

Storage Gateway

Lifecycle Policy

S3 Bucket

Glacier

Real time messages like EDI, HL7

Kinesis

Kinesis Application

STEP 1:

Connect on-prem databases to AWS Glue to ingest data using ETL operation. This data can be ingested into a database on AWS or stored as files on Amazon S3.

STEP 2:

On prem servers can utilize storage gateway to asynchronously upload datasets into AWS S3.

STEP 3:

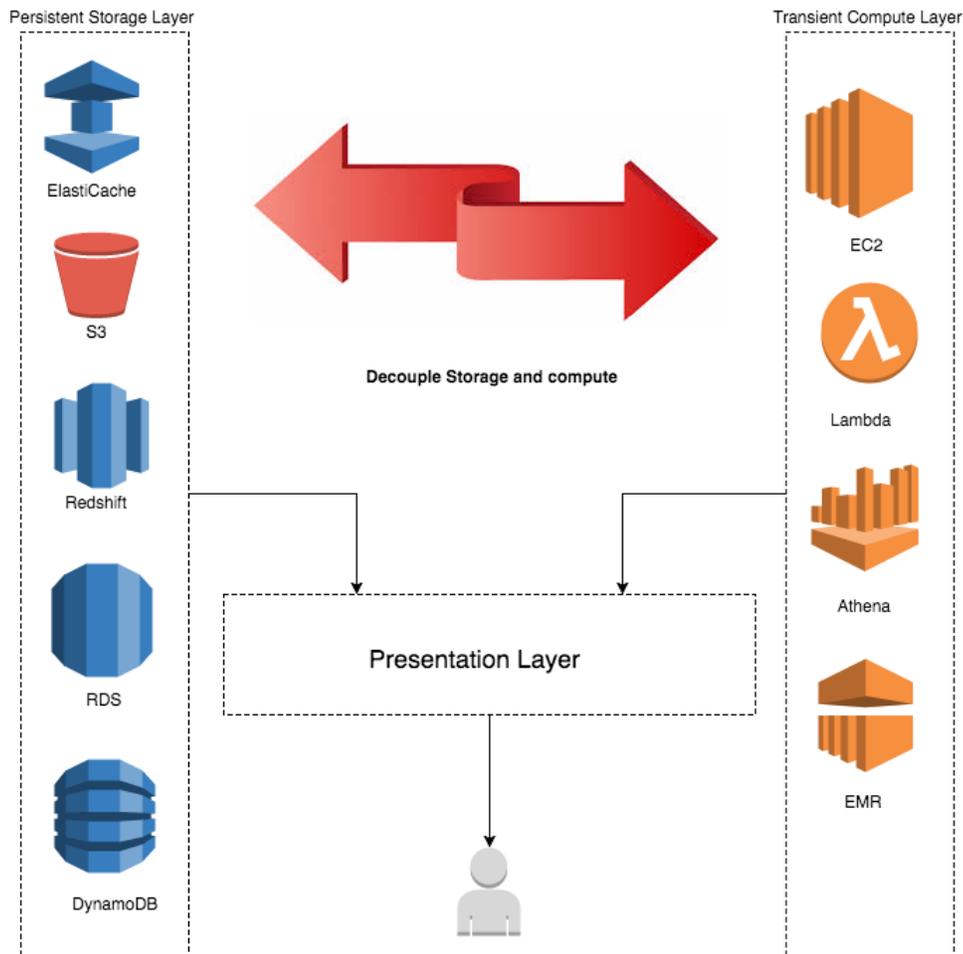Utilize Amazon Kinesis and Kinesis applications to ingest and process streaming records into AWS.

STEP 4:

Customers can use AWS Snowball for large one time migration of data into AWS.

aws

Analytics

Persistent Storage Layer

- ElastiCache
- S3
- Redshift
- RDS
- DynamoDB

Decouple Storage and compute

Transient Compute Layer

- EC2
- Lambda
- Athena
- EMR

Presentation Layer

STEP 1:

Store persistent data in Amazon S3 or AWS databases like Amazon RDS, Amazon Redshift or Amazon DynamoDB. You can also utilize Amazon Elasticache to caching frequently accessed data.

STEP 2:

Utilize AWS compute services like Amazon EC2 and Amazon EMR to process that data and scale independently of the storage layer.

STEP 3:

Utilize AWS serverless services like Lambda and Athena to analize data without the need of maintaining servers.
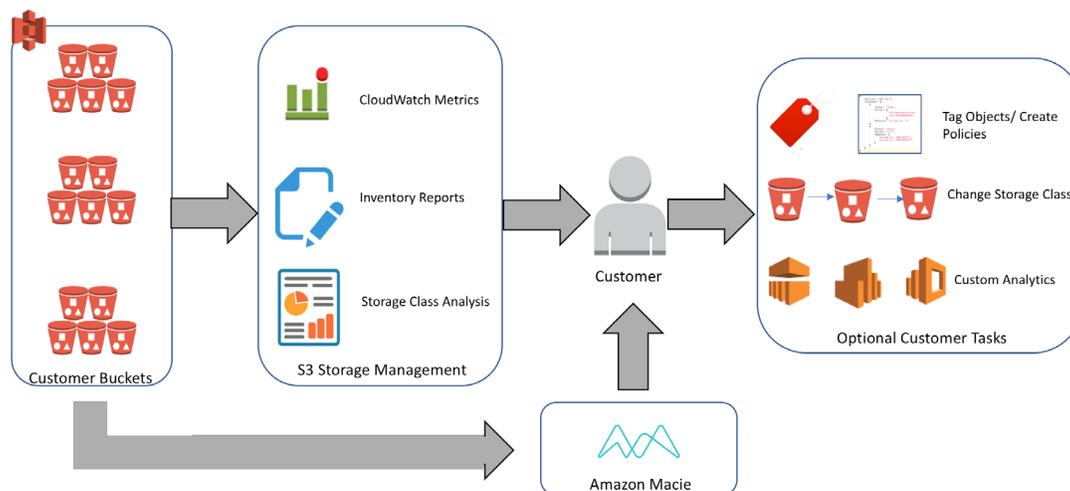
STEP 4:

Merge the outputs of the compute and persistent storage layer into a presentation layer which is used to serve the output to the users.

Store, Protect, Optimize Your Healthcare Data in AWS

Storage Management

STEP 1:

Tag objects that belong to a certain customer and use the tag as a filter in an IAM policy for controlling access to those objects.

STEP 2:

Transition infrequently accessed objects to Amazon S3 IA and objects that are never accessed to Amazon Glacier.

STEP 3:

Utilize notifications and reports from Amazon Macie to identify security issues with data on Amazon S3

STEP 4:

Build custom analytics using Amazon EMR, Amazon Redshift, Amazon Athena, Amazon QuickSight, or Amazon ElasticSearch that uses Amazon S3 inventory reports data as a source.