

Healthcare Data Analytics Framework for the Opioid Crisis

How Amazon Web Services Can Help Agencies and Healthcare
Entities Combat the Opioid Epidemic

March 2020



Notices

Customers are responsible for making their own independent assessment of the information in this document. This document: (a) is for informational purposes only, (b) represents current AWS product offerings and practices, which are subject to change without notice, and (c) does not create any commitments or assurances from AWS and its affiliates, suppliers or licensors. AWS products or services are provided “as is” without warranties, representations, or conditions of any kind, whether express or implied. The responsibilities and liabilities of AWS to its customers are controlled by AWS agreements, and this document is not part of, nor does it modify, any agreement between AWS and its customers.

© 2020 Amazon Web Services, Inc. or its affiliates. All rights reserved.

Contents

Introduction	1
Challenges with Healthcare and Opioid Data.....	2
Data Size, Scale, Formats and Sources	2
Document Processing and Analysis	2
Data Analytics	3
Data Privacy and Security.....	3
Framework Overview.....	3
Components	4
Reference Architecture and Best Practices	8
Conclusion	10
Contributors	11
Further Reading.....	11
Document Revisions.....	11

Abstract

Prescription opioid misuse and opioid use disorder (OUD) are significant health problems in the US and impact stakeholders across the healthcare sector¹. The Center for Medicare and Medicaid Services' (CMS's) [whitepaper](#) summarizes these issues and outlines strategies to handle the crisis and reduce impacts of this epidemic.² Although there are several initiatives at the federal, state, and local level to combat this problem, proactively identifying fraudulent and risky behaviors of various healthcare entities can help with preventive measures. This whitepaper outlines a framework for healthcare data analytics platform using Cloud, Artificial Intelligence (AI) and Machine Learning (ML) technologies to enable real time/predictive analytics and gather deep insights into opioid crisis.

This framework includes capabilities to:

- Enable streaming healthcare data ingestion from multiple sources in disparate formats and store information in centralized data lakes;
- Identify and automatically extract prescription, drug, and other healthcare information from medical documents;
- Provide real-time data analytics and insights into fraudulent and risky behaviors of various healthcare entities including patients, providers, pharmacies, etc.;
- Help proactively identify and predict fraud and abuse of opioid prescriptions based on historical drug usage and other data patterns; and
- Provide a secure platform for storing personal health information (PHI), personally identifiable information (PII) data and guidance to ensure compliance with Health Insurance Portability and Accountability Act (HIPAA) and other healthcare regulatory requirements.

Introduction

Critical aspects of the healthcare ecosystem in the US can be summarized using three important measures: 1) healthcare spending, 2) healthcare fraud, and 3) opioid crisis (*Figure 1*).

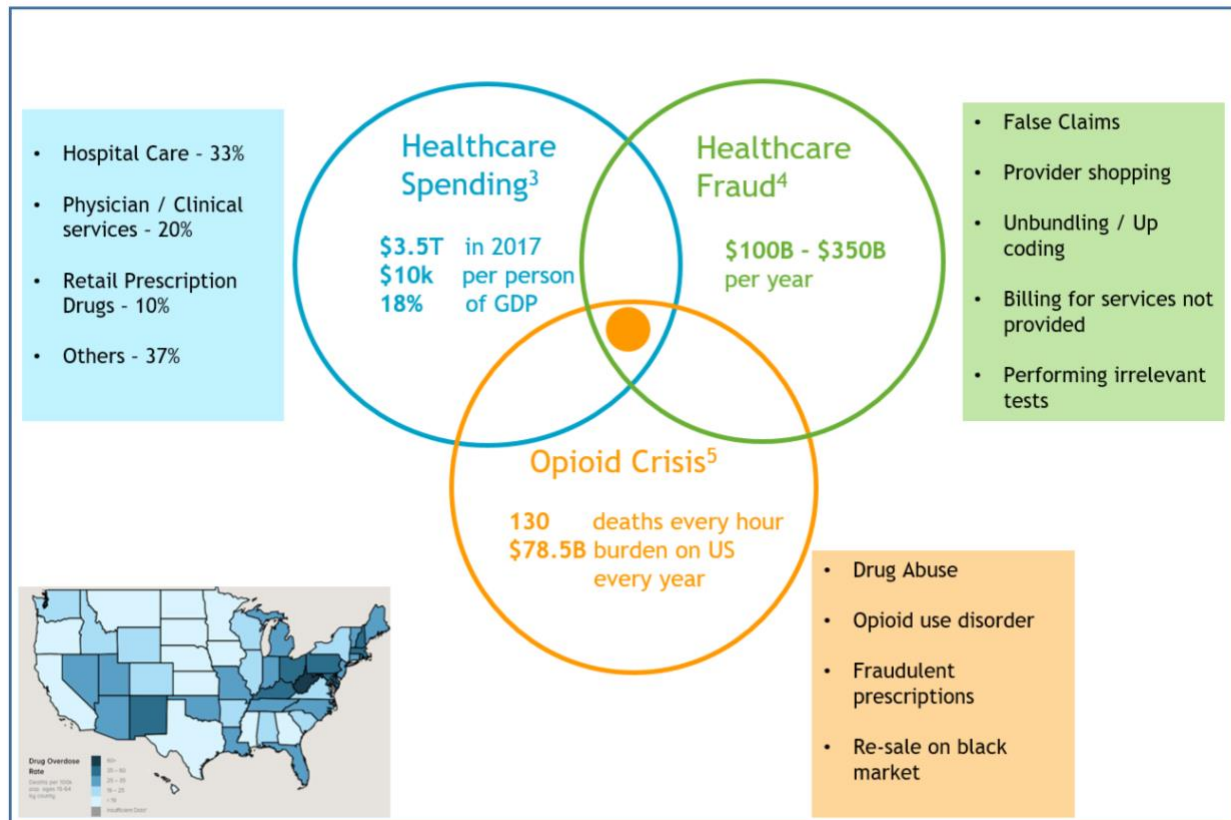


Figure 1: Current state of healthcare in the US^{3,4,5}

These measures are at an all-time high and affect every household in the country with increasing healthcare expenses.⁶ Congress enacted the [SUPPORT for Patients and Communities Act](#) to make medical treatment for opioid addiction more widely available while also reducing illicit drug use.⁷

The challenges in dealing with the opioid crisis include

- Identifying common patterns across healthcare fraud.
- Collecting and analyzing data across providers, payers and beneficiaries/patients.

- Effective campaign management on preventive measures.

Identifying fraudulent behavior through deep insights into data across the entities and proactively dealing with the opioid crisis helps reduce the overall healthcare costs in the US.

Challenges with Healthcare and Opioid Data

There are a number of challenges in dealing with healthcare data, including those challenges listed below. Proposed solutions to these problems are discussed in the next section.

Data Size, Scale, Formats and Sources

Healthcare data comes from multiple sources, including federal, state and local agencies; hospitals and providers; payers and insurance companies. The size and scale of this data is into billions of transactions each year, resulting in 100s of terabytes of data with multiple data formats originating from disparate systems with varying technology capabilities.⁸ These formats include structured and unstructured documents, images, streaming data of individual records, and others. Identifying a common storage option to host these disparate types at scale is a major challenge. AWS provides multiple storage options including [Amazon Simple Storage Service \(Amazon S3\)](#) to address this challenge.

Document Processing and Analysis

Medical documents, such as patient records, claims, administrative data, and prescriptions, have complex terminology that requires time and effort for semantic analysis and processing. Often, this processing is manual, which is error prone and requires healthcare expertise. Conversely, to process this information more seamlessly, organizations have required additional data formats and libraries to ensure data exchange into legacy systems. AWS has multiple services including [Amazon Comprehend Medical](#) and [Amazon Textract](#) to address these challenges through process automation.

Data Analytics

Key CMS focus areas for handling opioid crisis include using data to target prevention and treatment efforts, and identifying fraud, waste, and abuse.⁹ The objectives include data analytics for the following initiatives:

- Monitor success of prevention measures related to reducing overuse/misuse of prescription opioids.
- Improve transparency tools and interoperability for prescription rates.
- Analyze prescription opioid use patterns across CMS programs.
- Support state Medicaid programs to track and report data.

Some of the challenges with traditional data analytics include ingestion of data in batches from multiple sources, storing the data in a data warehouse and often waiting for days or even months to identify patterns and gather insights after episodes or encounters occur. Additionally, identifying fraudulent patterns takes time and effort with complex programming models, often contributing to the cost of pursuit. AWS capabilities include real-time analytics, anomaly detection, and predictive analytics using AI/ML technologies and out-of-box models for fraud detection. These capabilities align well to meet CMS' and other state and local healthcare initiatives to combat the opioid misuse.

Data Privacy and Security

Healthcare data consists of sensitive patient, drug, provider and payer information that should be protected and maintain compliance with local, state, and federal laws. Further, any data breach could lead to serious consequences both for patients as well as providers.¹⁰

Framework Overview

The AWS Cloud offers a number of services and capabilities to address the challenges discussed in the previous section. [Figure 2](#) shows a high level architecture and the framework for a healthcare analytics platform. This architecture can be used to gather deep insights into the opioid crisis and provide predictive analytics to proactively identify healthcare fraud within the opioid realm. Additionally, this architecture helps with preventive measures to combat the opioid epidemic by enabling effective campaign management.

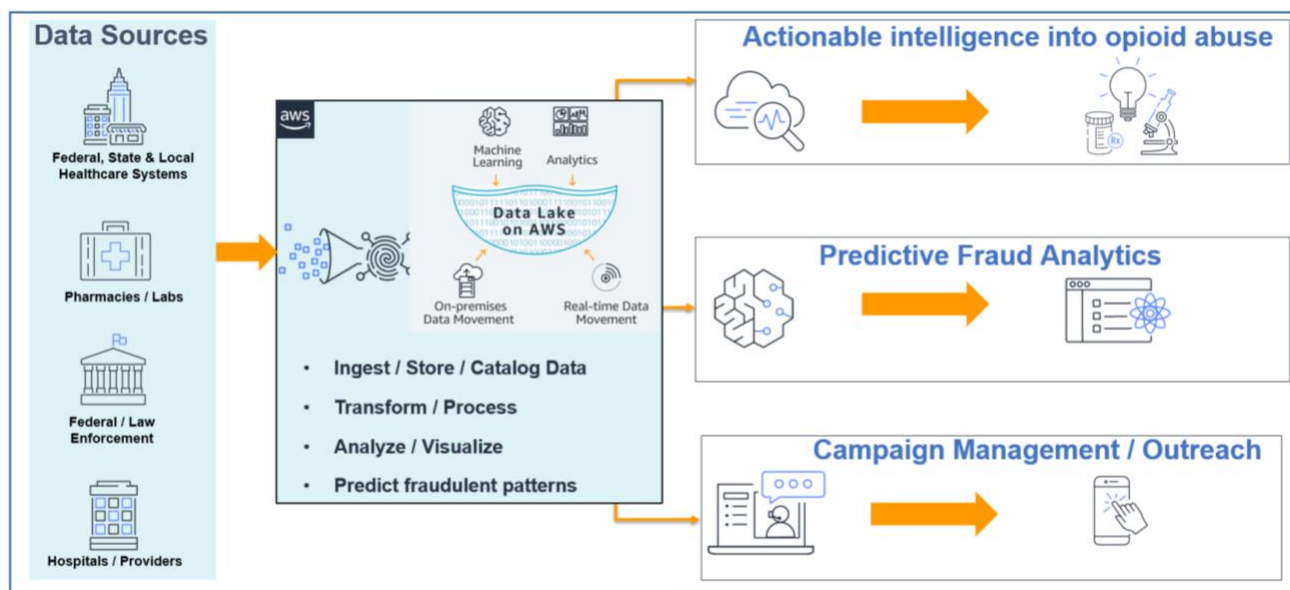


Figure 2: Framework to address challenges with opioid abuse

Components

The following sections provides components of the framework and the relevant AWS services.

Streaming Data Ingestion

The framework enables streaming healthcare data ingestion from multiple sources in disparate formats and storing the information in centralized data lakes. [Amazon Kinesis](#) makes it easy to collect and process streaming data. Streaming data includes opioid claims data originating from each state agency or from other healthcare data sources, such as patient management systems. Amazon Kinesis services include [Amazon Kinesis Data Streams](#) and [Amazon Kinesis Data Firehose](#) for data ingestion and storage into data lakes.

Centralized Storage with Data Lakes

The [Data Lake on AWS](#) Solution automatically crawls data sources, identifies data formats, and then suggests schemas and transformations, so that you don't have to spend time hand-coding data flows. For example, if you upload a series of opioid-related electronic medical records, claims, and prescription documents to [Amazon Simple Storage Service \(Amazon S3\)](#), [AWS Glue](#), a fully managed extract, transform and load (ETL) tool, can scan these documents to identify the schema and data types

present in these files. This metadata is then stored in a catalog to be used in subsequent transforms and queries.

The [AWS Lake Formation](#) service builds on the existing Data Lake on AWS Solution by allowing you to set up a secure data lake within days. Once you define where your lake is located, Lake Formation collects and catalogs this data, moves the data into Amazon S3 for secure access, and finally cleans and classifies the data using machine learning algorithms.

Additionally, user-defined tags, opioid prescription metadata, and claims data is stored in [Amazon DynamoDB](#), a key-value document database, to add business-relevant context to each dataset. You can browse available datasets or search on dataset attributes and tags to quickly find and access the documents in Amazon S3.

To summarize, the S3 storage service together with Glue and Lake Formation act as a centralized data lake for storing documents from multiple sources with disparate data formats. DynamoDB provides fast access to these documents by storing the document metadata (e.g. patientID, providerID, document storage location in Amazon S3, etc.).

Medical Information Extraction from Documents

Medical documents contain complex terminology/relationships that require healthcare expertise and manual analysis. [Amazon Comprehend Medical](#) is a natural language processing service that makes it easy to use machine learning to extract relevant medical information from unstructured text. Using Amazon Comprehend Medical, you can quickly and accurately gather information, such as medical condition, medication, dosage, strength, and frequency from a variety of sources including doctors' notes, clinical reports, and patient health records. In addition to Amazon Comprehend Medical, [Amazon Textract](#) can help extract text and data from scanned documents and images without the need for any custom coding. Together, these two services can form a data extraction pipeline for further analysis into opioid drug usage and fraud patterns.

Real-time Data Analytics, Business Intelligence, and Visualization

The Department of Health and Human Services (HHS) has a five-point Opioid Strategy, and one of the major objectives is to “Strengthen public health data collection and reporting to inform a real-time public health response as the epidemic evolves”.¹¹

Getting insights into opioid use trends and patterns in real time is key to timely responsiveness from federal, state, and local agencies. [Amazon Kinesis Data Analytics](#) provides capabilities to analyze streaming data, gain actionable insights, and respond to

events in real time. Amazon Kinesis Data Analytics is a managed service that scales automatically to match the volume and throughput of your incoming data. Once the data is processed, it can then be used for visualization to gain deeper insights into opioid patterns and trends including drug abuse patterns, high risk patients, high risk providers and pharmacies or fraudulent prescription patterns. [Amazon QuickSight](#) is a cloud-powered business intelligence service that makes it easy to deliver insights into Opioid trends, such as the total beneficiaries in a given state, county or locality; average beneficiary risk scores or opioid prescriber rates using datasets such as the CMS Medicare Part-D.

Predictive Analytics Using AI and ML

Although real-time data analytics are useful to identify opioid usage trends, the ability to predict risk and fraud based on historical data is key to tackling the crisis proactively. As an example, predicting opioid prescription rates by providers or predicting an average beneficiary risk score based on historical trends and current attributes would be helpful for federal and state agencies to identify focus areas and preventive measures. CMS publishes datasets for these types of analyses and predictions.¹²

AWS Artificial Intelligence (AI) services include capabilities to enable fraud detection and risk analysis.

Amazon SageMaker

[Amazon SageMaker](#) is a fully managed service that provides developers and data scientists with the ability to build, train, and deploy machine learning (ML) models quickly. SageMaker removes the heavy lifting from each step of the machine learning process to make it easier to develop high quality models. This is in contrast to the Traditional ML development which is a complex, expensive, iterative process and made even harder as there are no integrated tools for the entire machine learning workflow. SageMaker makes it easy to deploy your trained model into production with a single click so that you can start generating predictions for real-time or batch data.

Amazon Forecast

[Amazon Forecast](#) is a fully managed service that uses machine learning to deliver highly accurate forecasts. Based on the same technology used at Amazon.com, Amazon Forecast uses machine learning to combine time series data with additional variables to build forecasts. Amazon Forecast requires no machine learning experience to get started. You only need to provide historical data, plus any additional data that you

believe may impact your forecasts. As an example, Amazon Forecast can be used to forecast trends on budgets and opioid campaign management.

Effective Communication, Outreach, and Campaign Management

Fighting an epidemic such as the opioid crisis requires efficient communications across federal, state, and local agencies; proactive outreach to providers for safe prescribing practices; and managing call centers to host drug abuse helplines that can handle multiple languages (e.g. English and Spanish).¹³ AWS provides a number of capabilities to make this process efficient using a variety of services.

Amazon Connect

[Amazon Connect](#) is an easy to use Omni channel cloud contact center that helps companies provide superior customer service at a lower cost. Amazon Connect provides a seamless experience across voice and chat for your customers and agents. This includes one set of tools for skills-based routing, powerful real-time and historical analytics, and easy-to-use intuitive management tools - which means Amazon Connect simplifies contact center operations, improves agent efficiency, and lowers costs.

Amazon Transcribe/Amazon Transcribe Medical

[Amazon Transcribe/Amazon Transcribe Medical](#) provide speech-to-text capability applications. Audio data is virtually impossible for computers to search and analyze. Therefore, recorded speech needs to be converted to text before it can be used in applications. As an example, the calls to the opioid crisis helpline can be transcribed and the text transcriptions can be indexed and analyzed to identify opioid use patterns. Additionally, [Amazon Translate](#) can be used to convert text from one language to another (e.g. Spanish to English).

Amazon Pinpoint

[Amazon Pinpoint](#) helps the agencies dealing with opioid crisis to engage the healthcare entities by sending them personalized, timely and relevant communications. As an example, CMS could use Pinpoint for the campaign management initiative to communicate opioid prescription best practices with providers.

Data Security and Privacy

The healthcare industry has highly regulated workloads. Healthcare data storage and processing requires stringent federal regulatory and compliance frameworks, such as the US Health Insurance Portability and Accountability Act (HIPAA) and the Federal

Risk and Authorization Management Program (FedRAMP). AWS provides a number of capabilities to help meet these compliance requirements for data privacy and security to store and process healthcare data.

HIPAA compliance concerns the storage and processing of protected health information (PHI), such as insurance and billing information, diagnosis data, lab results, etc. HIPAA applies to covered entities (e.g., health care providers, health plans and health care clearinghouses) as well as business associates (e.g., entities that provide services to a covered entity involving the processing, storage, and transmission of PHI). The [Reference Architecture for HIPAA on AWS](#) deploys a model environment that can help organizations with workloads that fall within the scope of HIPAA. The reference architecture addresses certain technical requirements in the Privacy, Security, and Breach Notification Rules under the HIPAA Administrative Simplification Regulations (45 C.F.R. Parts 160 and 164).

AWS has also produced a quick start reference deployment for [Standardized Architecture for NIST-based Assurance Frameworks](#) on the AWS Cloud. This quick start focuses on the NIST-based assurance frameworks:

- National Institute of Standards and Technology (NIST) SP 800-53 (Revision 4)
- NIST SP 800-122, NIST SP 800-171
- The OMB Trusted Internet Connection (TIC) Initiative – FedRAMP Overlay (pilot)

This quick start includes [AWS CloudFormation](#) templates, which can be integrated with [AWS Service Catalog](#), to automate building a standardized reference architecture that aligns with the requirements within the controls listed above. It also includes a security controls matrix, which maps the security controls and requirements to architecture decisions, features, and configuration of the baseline to enhance your organization's ability to understand and assess the system security configuration.

Reference Architecture and Best Practices

The following figure outlines a reference architecture for the healthcare data analytics framework on AWS. As discussed earlier, this framework addresses five major areas to:

- Enable streaming healthcare data ingestion from multiple sources in disparate formats and store information in centralized data lakes.
- Identify and automatically extract prescription, drug and other healthcare information from medical documents.

- Provide real time data analytics and insights into fraudulent and risky behaviors of various healthcare entities including patients, providers, pharmacies, etc.
- Help proactively identify and predict fraud and abuse of opioid prescriptions based on historical drug usage and other data patterns.
- Provide a secure option for storing PHI, PII data and guidance to ensure compliance with HIPAA and other healthcare regulatory requirements.

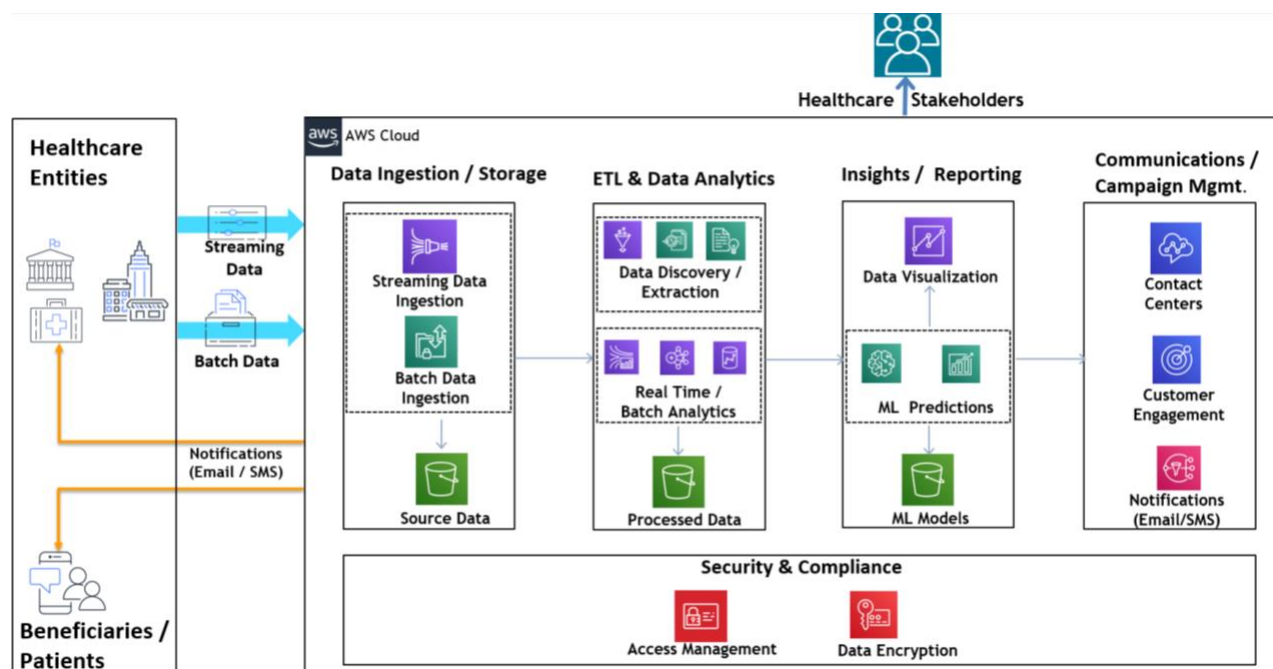


Figure 3: Healthcare data analytics platform reference architecture

Table 1 – Healthcare data analytics platform objectives / Service options mapping

Objective	Services & Options
Streaming, Batch Data Ingestion	Amazon Kinesis Data Firehose (real time) Amazon S3 SFTP (batch/file transfer) Amazon Snowball (data transfer for 10s of TBs of data) Amazon Database Migration Service (Migrate data from on premises databases)
Data Lake, Storage, Databases, Data warehouse	Amazon S3 (Object storage) Amazon DynamoDB (NoSQL Database) Amazon Redshift (Data warehouse)

Objective	Services & Options
ETL / Data Analytics	Amazon Glue (Schema discovery, ETL) Amazon Textract (Extract information from documents) Amazon Comprehend Medical (relationships within text) Amazon Kinesis Data Analytics (Real-time analytics) Amazon EMR (batch analytics) Amazon Redshift (Data warehouse)
Data Insights, AI/ML, Reporting	Amazon SageMaker (ML model build, train and deploy) Amazon QuickSight (Data visualization) Amazon Forecast (Time Series forecasting) Amazon Fraud Detector
Security Management	AWS Security Hub (Comprehensive view for compliance and security alerts) AWS Key Management Service (AWS KMS) AWS Identity and Access Management (IAM) AWS GuardDuty (Threat detection and monitoring)
Campaign Management	Amazon Connect (Contact Centers) Amazon Pinpoint (Campaign management, outreach, communications)

Conclusion

The US healthcare ecosystem is plagued with a number of challenges including prescription opioid misuse and abuse. Additionally, healthcare spending, waste, abuse and fraud are at all-time high¹⁴. There are several initiatives at the federal, state, and local levels to combat these challenges. All of these initiatives require large data storage, analytics and deep insights, in real time to accelerate decision making for preventive measures. This white paper introduced a framework and a reference architecture for healthcare data analytics using AWS to store, protect and optimize healthcare data; this paper also outlined a reference architecture to help unlock the potential for gaining deep insights into opioid misuse/abuse using AI/ML capabilities.

AWS services help improve citizen engagement, enhances communication and marketing campaigns for agencies dealing with the opioid crisis. Overall, AWS services

can help address a number of issues with opioid abuse/misuse and fraud, waste and abuse within the US healthcare ecosystem.

Contributors

Contributors to this document include:

- Srinath Godavarthi, Sr. Solutions Architect, WWPS/SLG
- Eric Schwenter, Principal Solutions Architect, WWPS/SLG
- Jason McNamara, HHS Strategic Programs Manager, WWPS/SLG

Further Reading

For additional information, see:

- [AWS Healthcare & Life Sciences](#)
- [AWS HIPAA Compliance](#)
- [Healthcare on the AWS Big Data Blog](#)
- [AWS Cloud Security](#)
- [AWS Whitepapers and Guides](#)
- [Flagging suspicious healthcare claims with Amazon SageMaker](#) on the AWS Machine Learning Blog

Document Revisions

Date	Description
March 2020	Initial publication.

Notes

- ¹ *Healthcare Payer Strategies to Reduce the Harm of Opioids* retrieved from <https://hfpp.cms.gov/news/hfpp-opioid-whitepaper.pdf>
- ² *Healthcare Payer Strategies to Reduce the Harm of Opioids* retrieved from <https://hfpp.cms.gov/news/hfpp-opioid-whitepaper.pdf>
- ³ *National Health Expenditures 2018 Highlights* retrieved from <https://www.cms.gov/files/document/highlights.pdf>
- ⁴ *Medicare Fraud Strike Force Charges 107 Individuals for Approximately \$452 Million in False Billing* retrieved from <https://archives.fbi.gov/archives/neworleans/press-releases/2012/medicare-fraud-strike-force-charges-107-individuals-for-approximately-452-million-in-false-billing>
- ⁵ *Opioid Overdose Crisis* retrieved from <https://www.drugabuse.gov/drugs-abuse/opioids/opioid-overdose-crisis#one>
- ⁶ *National Health Expenditures 2018 Highlights* retrieved from <https://www.cms.gov/files/document/highlights.pdf>
- ⁷ *Substance Use-Disorder Prevention that Promotes Opioid Recovery and Treatment for Patients and Communities Act or the SUPPORT for Patients and Communities Act* retrieved from <https://www.congress.gov/bill/115th-congress/house-bill/6>
- ⁸ *2013 U.S. Healthcare Efficiency Index Electronic Administrative Transaction Adoption and Savings* retrieved from <https://www.caqh.org/sites/default/files/explorations/index/report/2013Index.pdf>
- ⁹ *CMS Roadmap: Fighting the Opioid Crisis* retrieved from <https://www.cms.gov/About-CMS/Agency-Information/Emergency/Downloads/Opioid-epidemic-roadmap.pdf>
- ¹⁰ *Tennessee diagnostic medical imaging services company pays \$3,000,000 to settle breach exposing over 300,000 patients' protected health information* retrieved from <https://www.hhs.gov/about/news/2019/05/06/tennessee-diagnostic-medical-imaging-services-company-pays-3000000-settle-breach.html>
- ¹¹ *5-Point Strategy To Combat the Opioid Crisis* retrieved from <https://www.hhs.gov/opioids/about-the-epidemic/hhs-response/index.html>
- ¹² *CMS Medicare data sets on opioids* retrieved from <https://data.cms.gov/browse?tags=opioid>

- ¹³ *Substance Abuse and Mental Health Services Administration National Helpline* retrieved from <https://www.samhsa.gov/find-help/national-helpline>
- ¹⁴ *National Health Expenditures 2017 Highlights* retrieved from <https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/Downloads/highlights.pdf>