Homelessness and Technology

How Technology Can Help Communities Prevent and Combat Homelessness

March 2019

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 AWS's current 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's products or services are provided "as is" without warranties, representations, or conditions of any kind, whether express or implied. AWS's responsibilities and liabilities 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.

Contents

Abstract

The disparate nature of current homeless information management systems limits a community's ability to identify trends or emerging needs, measure internal performance goals, and make data driven decisions about the effective deployment of limited resources. With the shift in recent years to whole person care, there is increasing demand to connect these disparate systems to affect better outcomes.

In this document, we have outlined four pillars of how AWS technology and services can act as a best practice to organizations looking to leverage the cloud for Homeless Management Information Systems (HMIS). These pillars are as follows:

- Connect disparate data sources using a data lake design pattern.
- Make predictions using data analytics workloads, big data, and machine learning.
- Manage identity and vital records for people experiencing or at risk for experiencing homelessness.
- Leverage the AWS Business Associates Addendum (BAA) and associated services for Health Insurance Portability and Accountability Act (HIPAA) Compliance and NIST based assurance frameworks.

Introduction

Preventing and combatting homelessness depends on a coordinated Continuum of Care (CoC) on the ground locally, sharing information across disparate systems, and collaborating with the public, nonprofit, philanthropic, and private sector partners. The systems that collect this information today (i.e., homelessness services, electronic health records, education, and criminal justice information systems) were designed independently to address particular applications and are managed by different entities with separate IT systems and governance.

The disparate nature of these systems limits a community's ability to identify trends or emerging needs, measure internal performance goals, and make data driven decisions about the effective deployment of limited resources. With the shift in recent years to whole person care, there is increasing demand to connect these disparate systems to affect better outcomes.

Redesigning these systems for interoperability is critical, but it will take time. In the meantime, you can use the best practices in this document to connect disparate information today to develop a comprehensive view for each client to drive better outcomes and enable analytics that support data-driven decision making.

Best Practices for Combatting Homelessness

The following best practices focus on addressing some of the challenges of combatting homelessness, but they are highly applicable to other socioeconomic and healthcare challenges that cross multiple systems.

- Connect disparate data sources using a data lake design pattern.
- Make predictions using data analytics workloads, big data, and machine learning.
- Manage identity and vital records for people experiencing or at risk or experiencing homelessness.
- Leverage the AWS Business Associates Addendum (BAA) and associated services for Health Insurance Portability and Accountability Act (HIPAA) Compliance and NIST based assurance frameworks.

Connect Data Sources with Data Lakes

Connecting disparate data sources to create a comprehensive view of the homeless population and their interactions across numerous service providers and government entities can come with many technical challenges. Schema and structural differences in separate locations can be difficult to combine and query in a single place. Also, some data may be highly structured whereas other datasets may be less structured and involve a smaller signal to noise ratio. For example, data stored in a tabular CSV format from a traditional database combined with a nested JSON schema that may come from a fleet of devices (e.g. personal health records versus real-time medical equipment data) can be difficult to join and query together using a relational database alone.

A data lake is a centralized repository that allows you to store all of your structured and unstructured data at any scale. You can store your data as is, without having to first structure the data, and run different types of analytics. Dashboards, visualizations, big data processing, real-time analytics, and machine learning can all help contribute to better decision making and improve client outcomes.

A data warehouse is a central repository of structured information that can be analyzed to make better informed decisions. Data flows into a data warehouse from transactional systems, relational databases, and other sources, typically on a regular cadence. Business analysts, data scientists, and decision makers access the data through business intelligence (BI) tools, SQL clients, and other analytics applications.

Data warehouses and data lakes complement each other well by allowing separation of concerns and leveraging scalable storage and scalable analytic capability, respectively.

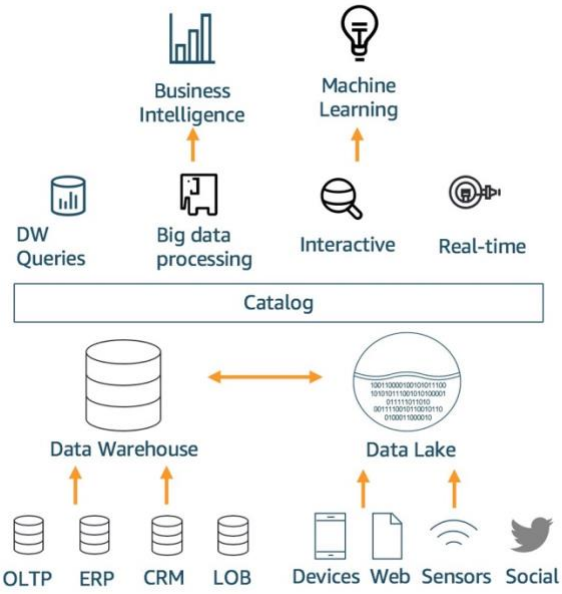

Figure 1: Connecting Disparate Data Sources

A Homeless Management Information System (HMIS) is an information technology system used to collect client-level data and data on the provision of housing and services to homeless individuals and families and persons at risk of homelessness. You can create data lakes to connect disparate HMIS data across CoC and regional boundaries. With a consolidated dataset, you gain a comprehensive and unduplicated understanding of who is served, with which programs, and to what outcomes across a region or state. This depth of understanding reveals patterns that can help care providers rapidly create and tune interventions to the unique needs of homeless groups (e.g., veterans, youth, elders, chronically homeless, and so on) and provides the public, elected officials, and funders with transparency about investments versus outcomes.

By centralizing data and allowing Federated access to a searchable data catalog, you can address pain points around connecting disparate data systems. The data lake can accept data from many different sources. These may include, but are not limited to:

- Existing relational database and data warehouse engines (either on-premises or in the cloud)
- Clickstream data from mobile or web applications
- Internet of Things (IoT) device data
- Flat file imports

- API data
- Media sources, such as video and audio streams

This data should be stored durably and encrypted with industry standard open source tools both at rest and in transit, since the data may contain personally identifiable information (PII) and be subject to compliance controls. Federated access through an Identity provider (e.g. Active Directory, Google, Facebook, etc.) should also be used as a means of authorization to enable different teams to access the correct level of data. Metadata concerning the data should be held within a searchable data catalog to enable fast access to structural and data classification information. This should all be accomplished in a cost-effective and scalable manner with the data held in its native format to facilitate export, further transformation, and analysis.

Data Lake Solution

The Data Lake solution automatically crawls data sources, identifies data formats, and then suggests schemas and transformations, so you don't have to spend time hand-coding data flows. For example, if you upload a series of JSON files to Amazon Simple Storage Service (Amazon S3), AWS Glue, a fully managed extract, transform and load (ETL) tool, can scan these files and work out the schema and data types present within these files. This metadata is then stored in a catalog to be used in subsequent transforms and queries. Additionally, user-defined tags are stored in Amazon DynamoDB, a key-value document database, to add business-relevant context to each dataset. The solution enables you to create simple governance policies that require specific tags when datasets are registered with the data lake. You can browse available datasets or search on dataset attributes and tags to quickly find and access data relevant to your business needs.

AWS Lake Formation

The AWS Lake Formation service builds on the existing data lake 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. You can then access a centralized data catalog which describes available datasets and their appropriate usage. This approach has a number of benefits, from

building out a data lake quickly to simplifying security management and allowing easy and secure self-service access.

Enable Data Analytics Using Big Data and Machine Learning Techniques

Communities want a better understanding of the circumstances that contribute to homelessness, prevent homelessness, and accelerate someone's path out of homelessness. These predictions are critical inputs for the development of interventions across a continuum of care and for disaster response planning. With a data lake, communities can build, train, and tune machine learning models to predict outcomes.

Data Processing and Storage

In today's connected world, a number of data sources are available to be consumed. Some examples include public APIs, sensor/device data, website analytics, imagery as well as traditional forms of data such as relational databases and data warehouses.

Amazon Relational Database Service (Amazon RDS) allows developers to build and migrate existing databases into the cloud. AWS supports a large range of commercial and open-source database engines (e.g. MySQL, PostGres, Amazon Aurora, Oracle, SQL Server) allowing developers freedom to keep their current database or migrate to an open source platform for cost savings and new features. Amazon RDS maintains high-availability through the use of Multi-Availability Zone deployments to ensure that production databases stay operational in the event of a hardware failure.

For customers with data warehousing needs, Amazon Redshift enables developers to query large sets of structured data within Redshift and within Amazon S3. When combined with a business intelligence tool, such as Amazon QuickSight, Tableau, or Microsoft Power BI, you can create powerful data visualizations and gain insights into data that were previously out of reach on legacy IT systems.

Amazon Kinesis makes it easy to collect, process and analyze streaming data. Kinesis enables the construction of real-time data dashboards, video analytics, and stream transformations to filter and query data as it comes into the organization from an array of sources.

Make Predictions with Machine Learning and Analytics

Machine learning can help answer complicated questions by making predictions about future events from past data. Some examples of machine learning models include image classification, regression analysis, personal recommendation systems and time-series forecasting. For a CoC, these capabilities may seem out of reach, but due to the power and scale of the cloud, these capabilities are now within anyone's reach.

Amazon Comprehend Medical, Amazon Forecast and Amazon Personalize put powerful machine learning model creation capabilities into the hands of developers, requiring no machine learning background or servers to manage.

Amazon Comprehend Medical

Amazon Comprehend Medical is a natural language processing service that makes it easy to use natural language processing and machine learning to extract relevant medical information from unstructured text. For example, you can use Comprehend Medical to identify and search for key terms in a large corpus of health records, allowing case officers and medical professionals to look for recurring patterns or key phrases in patient records when providing treatment to homeless individuals.

Amazon Forecast

Amazon Forecast uses machine learning to combine time series data with additional variables to build forecasts. You can use Amazon Forecast to predict changes in a homeless population over time. Forecast can also consider how other correlating external factors affect the population, such as natural disasters or severe weather or the introduction of new programs and initiatives.

Amazon Personalize

Amazon Personalize is a machine learning service that makes it easy for developers to create individualized recommendations for customers using their applications. For example, many times individuals at risk of or experiencing homelessness struggle to find assistance programs. Navigating these many programs and facilities can be daunting and time consuming. By using HMIS data from other individuals in similar situations, you can build a recommendation engine that suggests relevant programs to individuals and families. These recommendations enable them to access programs that they may not be aware of or have the time to research.

Manage Identity and Vital Records

Proof of identity and eligibility are critical to matching the right people at the right time to the right interventions. Copies of vital records, such as social security cards, birth certificates, proof of disability, and copies of utility bills, lease or property title documents are often required by various programs that are designed to help those experiencing or at risk of experiencing homelessness. However, without a secure and reliable place to store and access these documents, the most vulnerable people are often left the worst off. Their lack of documentation can become a barrier to service and extend the length of crisis.

In addition to the need for a secure storage location, customers need a mechanism to control and share documents with authorized parties to evaluate eligibility for various programs and/or to verify authenticity. This mechanism must track who accesses these documents, at what time, and in what manner in a cryptographically verifiable, immutable way. Ledger or blockchain-based applications can meet this requirement by storing the interaction event metadata for a document or set of documents in a verifiable ledger. This ledger creates a verifiable audit trail that can store all of the events that occur during a document's lifetime.

Amazon Simple Storage Service (Amazon S3)

Amazon Simple Storage Service (Amazon S3) stores objects in the cloud reliably and at scale. Using Amazon S3, you can build the substrate for a document storage and retrieval application. Amazon S3 has many pertinent security features, such as multi-factor control of deleting and modifying objects and object versioning. Amazon S3 also uses encryption at rest and in transit using industry standard encryption algorithms and a simple HTTPS-based API. Amazon S3 supports signed URLs so that access to objects can be granted for a limited time. Finally, Amazon S3 offers cost savings with intelligent tiering so that documents can be automatically moved into different storage tiers depending on their usage patterns.

Amazon Quantum Ledger Database (Amazon QLDB)

Amazon Quantum Ledger Database (Amazon QLDB) is a fully managed ledger database that provides a transparent, immutable, and cryptographically verifiable transaction log owned by a central trusted authority. Amazon QLDB tracks each and

every application data change and maintains a complete and verifiable history of changes over time.

Amazon Managed Blockchain

Amazon Managed Blockchain is fully managed blockchain service that makes it easy to create and manage scalable blockchain networks using popular open source frameworks such as Hyperledger Fabric and Ethereum. By combining secure storage in the cloud with a cryptographically verifiable event log, it is possible to build a scalable application that can store documents in a secure manner and be able to verify the contents and access patterns to each individual document during its lifetime.

Leverage AWS for HIPAA Compliance

Health Insurance Portability and Accountability Act (HIPAA) compliance concerns the storage and processing of protected health information (PHI), such as insurance and billing information, diagnosis data, lab results, and so on. 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). AWS offers a standardized Business Associates Addendum (BAA) for business associates. Customers who execute a BAA may process, store, and transmit PHI using HIPAA eligible services defined in the AWS BAA, such as Amazon S3, Amazon QuickSight, AWS Glue, and Amazon DynamoDB. For a complete list of services, see HIPAA Eligible Services Reference.

HMIS Data Privacy and HIPAA

Each CoC is responsible for selecting an HMIS software solution that complies with the Department of Housing and Urban Development's (HUD) standards.

HMIS has a number of privacy and security standards that were developed to protect the confidentiality of personal information, while at the same time allowing limited data disclosure in a responsible manner. These standards were developed after careful review of the HIPAA standards regarding 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)
- The DoD Cloud Computing Security Requirements Guide (SRG)

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.

Conclusion

AWS technology can help communities drive better outcomes for citizens using the technology and services included this paper. However, we understand that homelessness is fundamentally a human problem—all of these initiatives must have strong backing by forward thinking officials and program managers to make an impact in the lives of those at risk or experiencing homelessness.

Contributors

The following individuals and organizations contributed to this document:

- Alistair McLean, Sr. Solutions Architect, AWS
- Jessie Metcalf, Program Manager, AWS

- Casey Burns, Health and Human Services Leader, AWS

Further Reading

For additional information, see the following:

- HMIS Data and Technical Standards
- Reference Architecture for HIPAA on AWS
- Reference Architecture for HIPAA on the AWS Cloud: Quick Start Reference Deployment
- Standardized Architecture for NIST-based Assurance Frameworks on the AWS Cloud: Quick Start Reference Deployment
- AWS Machine Learning Blog: Create a Question and Answer Bot with Amazon Lex and Amazon Alexa
- AWS Government, Education and Non-Profits Blog

Document Revisions

Date	Description
March 2019	Initial document release.