



# Advancing public health missions with FHIR data analytics and AI/ML

CDC Health Data Innovation Summit

Henrik Balle

Sr. Solutions Architect  
AWS

Marc Etone Nkwelle

Solutions Architect  
AWS

# Public health challenges

## What we hear from customers:

**Health data is vast, multi-modal, and siloed and non-interoperable.** Gaining knowledge is labor intensive and error prone

Need to build **capacity for a unified public health view** that is ready for analysis and intervention

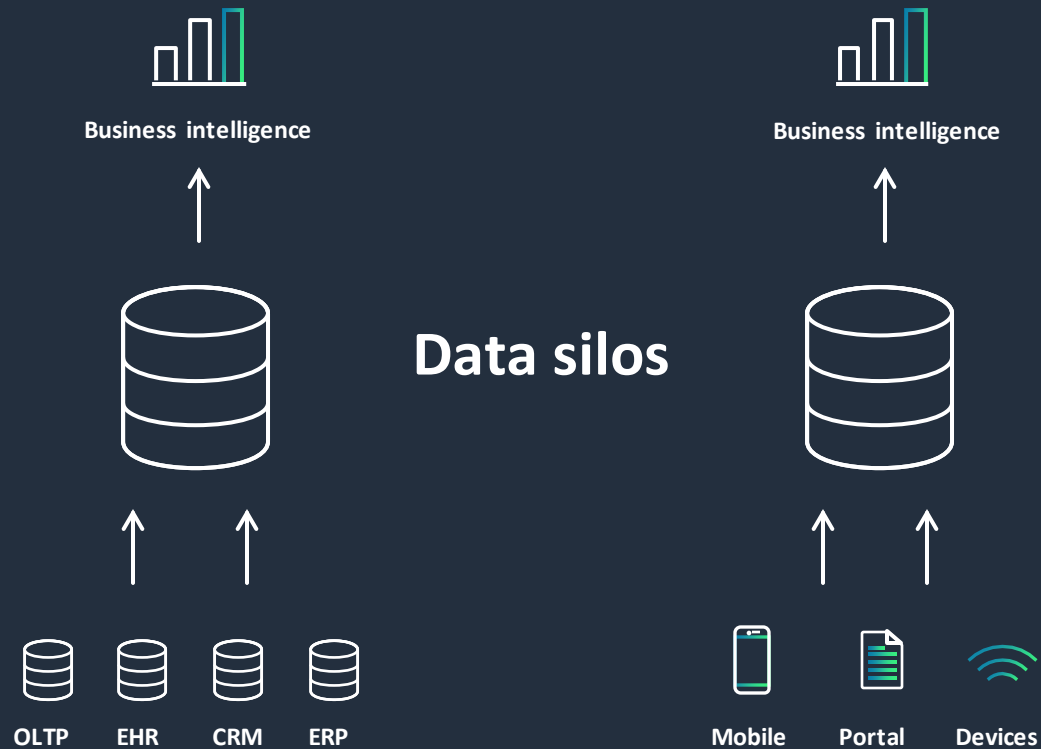
During public health threats **data velocity increases and data volumes can grow rapidly**

**Make data actionable** using AI/ML and advanced analytics on disparate data

Storing and organizing data so that researchers can **access, study, and cross-reference it to facilitate health related breakthroughs**

**Health disparities** continue to widen and health equity remains a challenge

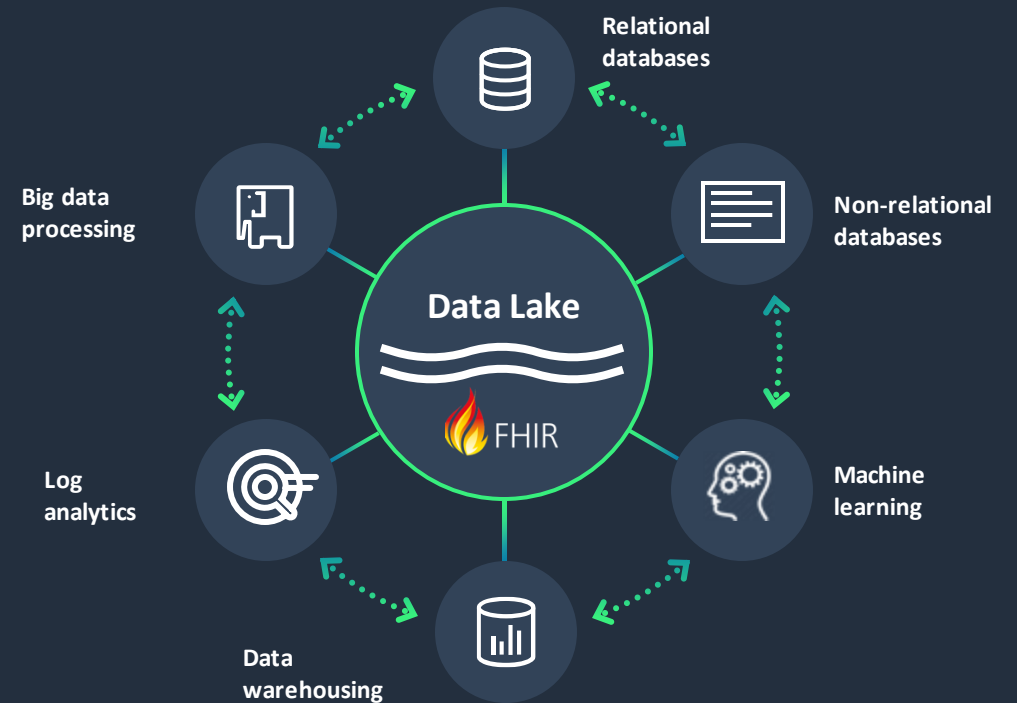
# Breaking down silos and improving interoperability



to



## Modern health data platforms



# AWS HealthLake

A MANAGED, SECURE, HIPAA-ELIGIBLE FHIR STORAGE, TRANSACTIONAL, AND ANALYTICS SERVICE

Transactional FHIR  
server to power  
clinical applications



## Store clinical/claims data using interoperability standards

Store patient medical history from multiple data sources in the normalized common data model (FHIR-based) format



## Build highly scalable interoperability solutions to meet regulatory needs

Leverage HealthLake Patient Access APIs and Bulk FHIR APIs to meet the 21<sup>st</sup> Century Cures Act for patient access and interoperability requirements



## Build secure applications using SMART on FHIR

Build patient360 end user clinical applications by integrating with OAuth2 compliant authorization service and securely access HealthLake data

# AWS HealthLake

A MANAGED, SECURE, HIPAA-ELIGIBLE FHIR STORAGE, TRANSACTIONAL, AND ANALYTICS SERVICE

Unlocking the  
potential of your  
health and science  
data at scale



## Automatic transformation of FHIR data to analytics format

Empower healthcare analysts to query HealthLake FHIR data using SQL



## Simplify extraction of medical information

Built-in Comprehend Medical to extract medical information from unstructured medical text for easier search and query

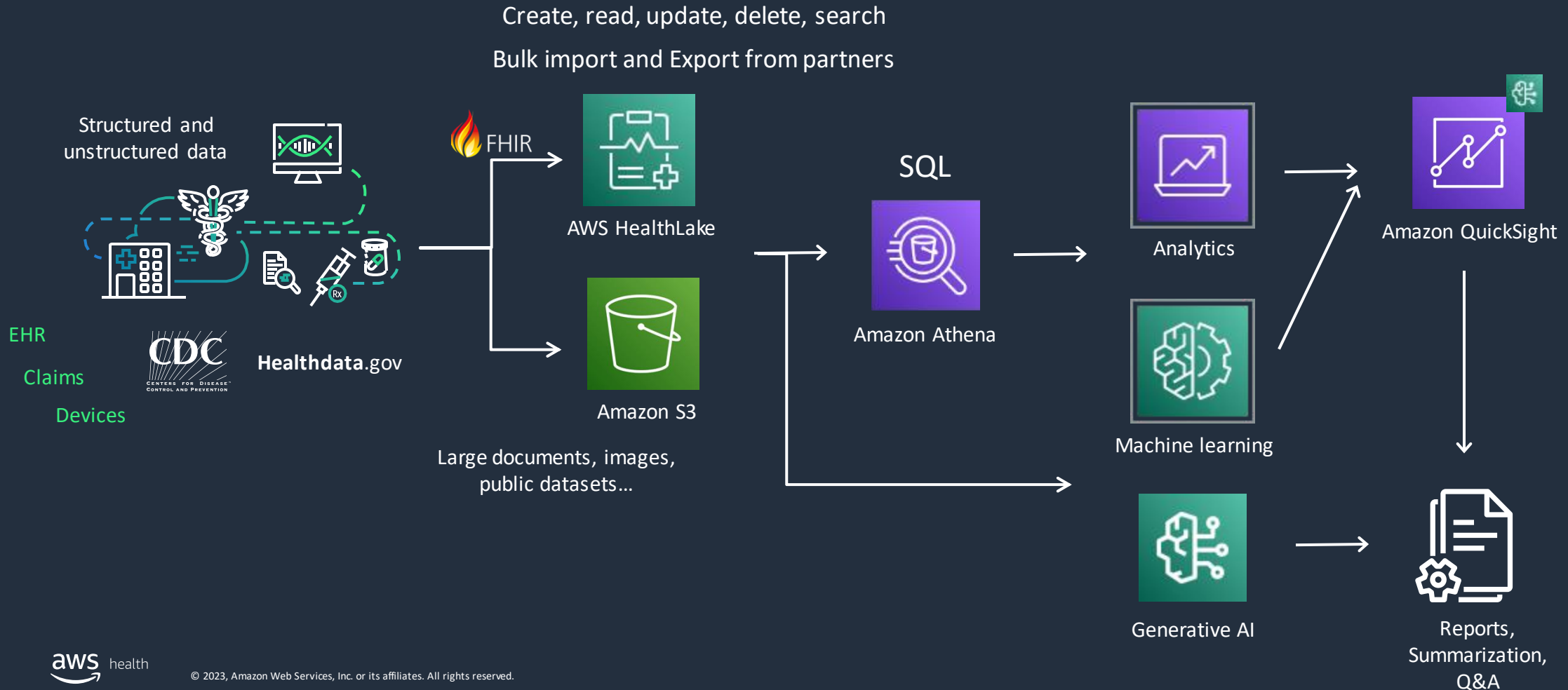


## Apply analytics and ML on health data

Make predictions and identify trends with health data using Amazon SageMaker machine learning models and Amazon QuickSight analytics

# AWS HealthLake analytics and AI/ML integration

NATIVE INTEGRATION WITH ANALYTICS SERVICES ENABLED THROUGH SQL ACCESS TO DATA WITHIN AMAZON HEALTHLAKE OR RETRIEVING DATA FROM AMAZON S3

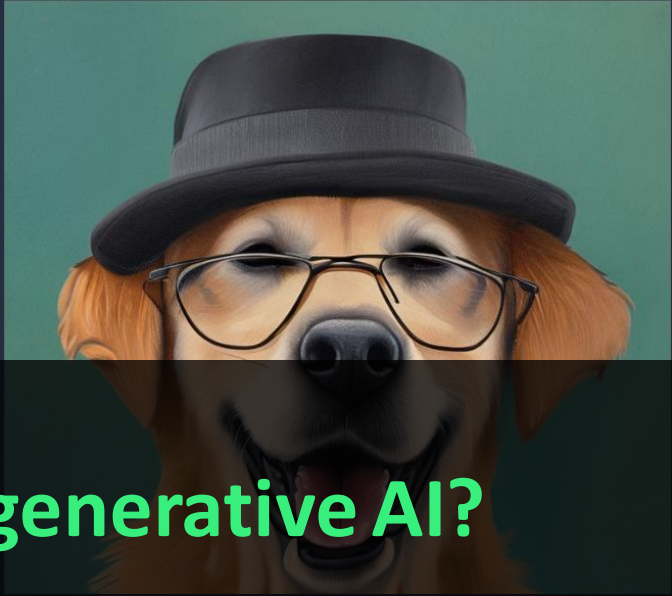




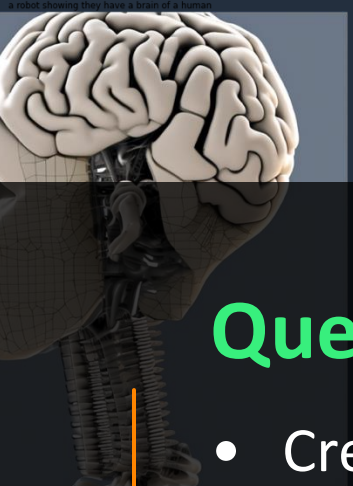
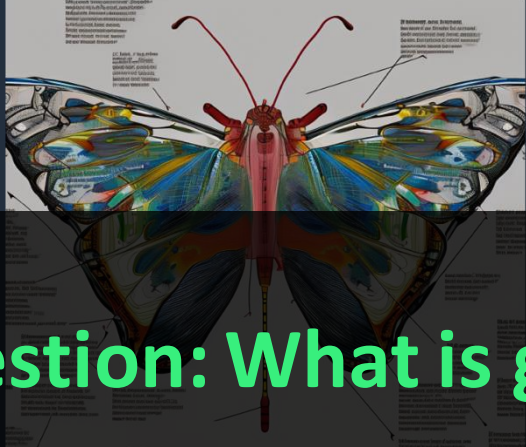
# Generative AI for Public Health



A golden retriever wearing glasses and a hat in a portrait painting



beautiful robotic butterfly anatomy diagram



a robot wearing a brain as a hat



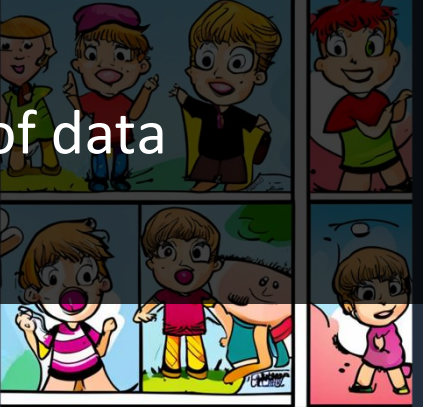
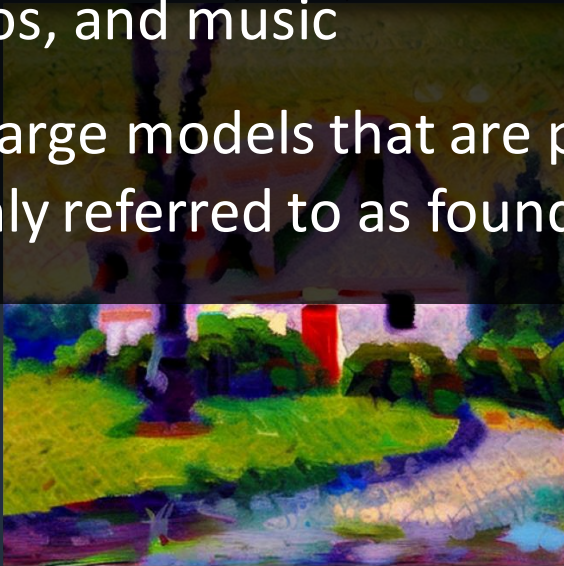
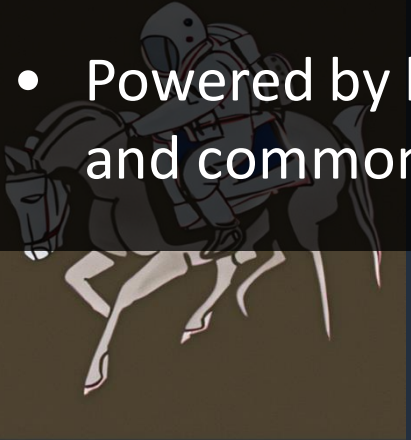
photo of a statue of a robot in university courtyard



astronaut on a horse

# Question: What is generative AI?

- Creates new content and ideas, including conversations, stories, images, videos, and music
- Powered by large models that are pretrained on vast corpora of data and commonly referred to as foundation models (FMs)





# Generative AI is used for a wide range of public health use cases



## Research and discovery

Multi-modal data analysis

Patient to trial matching

Generate synthetic data for testing



## Data analytics

Analytics and visualization

Automated report generation



## Operational Efficiency

Intelligent document processing

Document Q&A and summarization



## Citizens Experience

Patient outcome prediction

Personalize patient discharge instructions and treatment plans

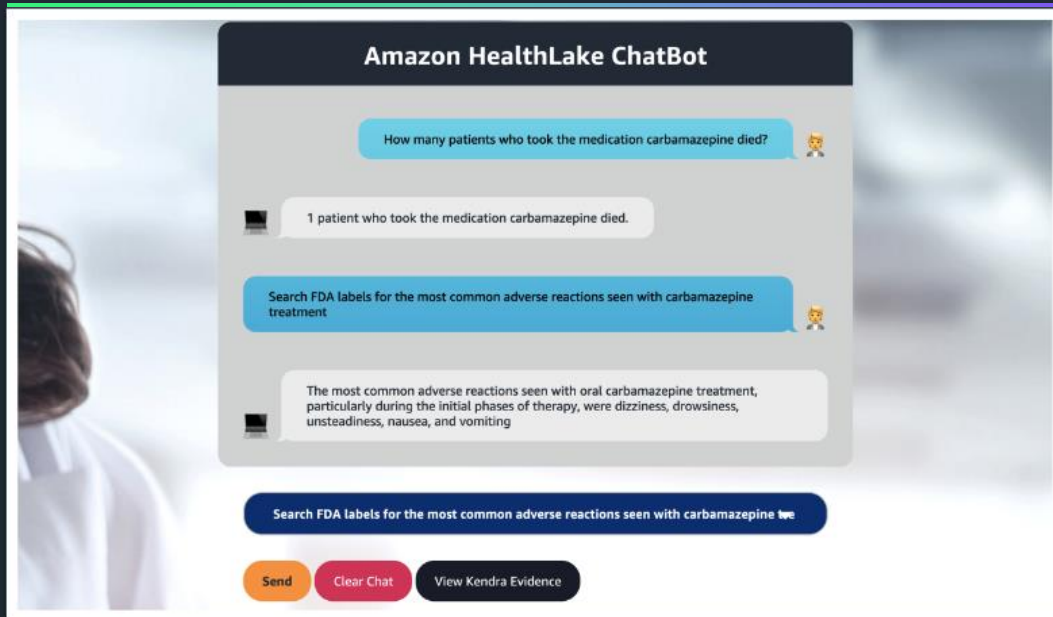


## Digital Health

Patient care concierge

Chatbots

# Q&A with research data



**Amazon HealthLake ChatBot**

How many patients who took the medication carbamazepine died? 🤖

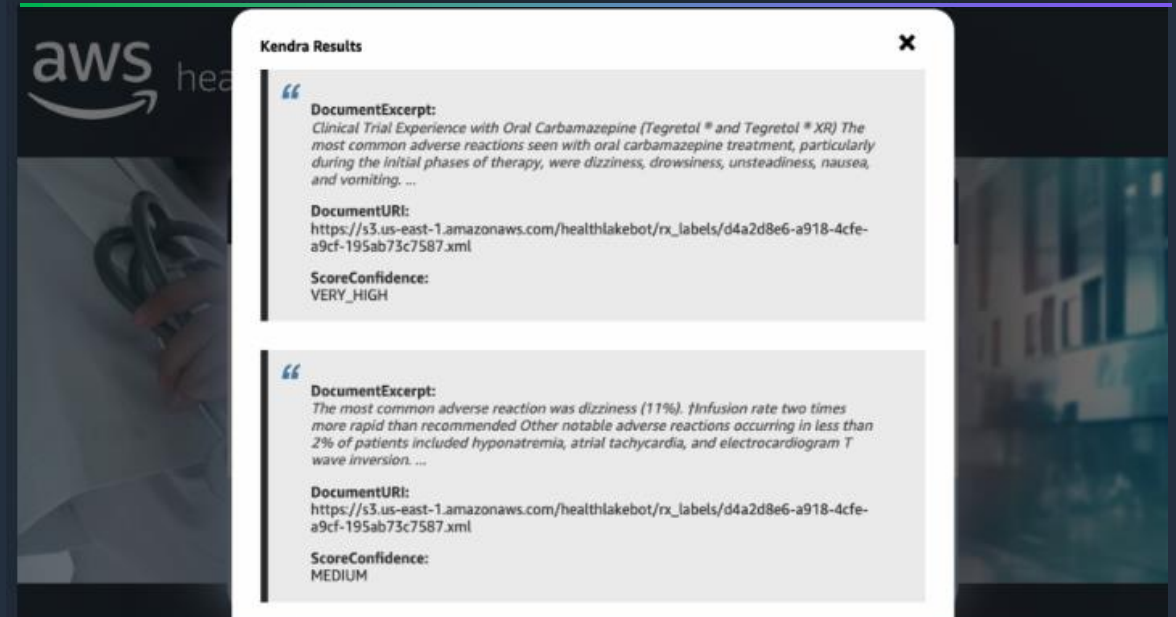
1 patient who took the medication carbamazepine died.

Search FDA labels for the most common adverse reactions seen with carbamazepine treatment 🤖

The most common adverse reactions seen with oral carbamazepine treatment, particularly during the initial phases of therapy, were dizziness, drowsiness, unsteadiness, nausea, and vomiting.

Search FDA labels for the most common adverse reactions seen with carbamazepine 🔍

Send Clear Chat View Kendra Evidence



**Kendra Results** ✕

“ **DocumentExcerpt:**  
*Clinical Trial Experience with Oral Carbamazepine (Tegretol® and Tegretol® XR) The most common adverse reactions seen with oral carbamazepine treatment, particularly during the initial phases of therapy, were dizziness, drowsiness, unsteadiness, nausea, and vomiting. ...*

**DocumentURI:**  
[https://s3.us-east-1.amazonaws.com/healthlakebot/rx\\_labels/d4a2d8e6-a918-4cfe-a9cf-195ab73c7587.xml](https://s3.us-east-1.amazonaws.com/healthlakebot/rx_labels/d4a2d8e6-a918-4cfe-a9cf-195ab73c7587.xml)

**ScoreConfidence:**  
VERY\_HIGH

“ **DocumentExcerpt:**  
*The most common adverse reaction was dizziness (11%). Infusion rate two times more rapid than recommended Other notable adverse reactions occurring in less than 2% of patients included hyponatremia, atrial tachycardia, and electrocardiogram T wave inversion. ...*

**DocumentURI:**  
[https://s3.us-east-1.amazonaws.com/healthlakebot/rx\\_labels/d4a2d8e6-a918-4cfe-a9cf-195ab73c7587.xml](https://s3.us-east-1.amazonaws.com/healthlakebot/rx_labels/d4a2d8e6-a918-4cfe-a9cf-195ab73c7587.xml)

**ScoreConfidence:**  
MEDIUM

Traceability into decision making  
and data generation

# Gather new insights from of unstructured data with RAG

**Patient Discharge Summary**  
Not-A Real Hospital, Department of Family Medicine

**Patient**  
Patient Name: John Doe  
Patient ID: NARH-36640  
Gender: Male

**Visit**  
Attending Physician: Mateo Jackson, PhD  
Admit Date: 07-Sep-2020  
Discharge Date: 08-Sep-2020  
Discharge Disposition: Home with Support Services

**Diagnosis**  
35 yo M c/o stomach problems since 2 monthsh ago. Patient reports epigastric abdominal pain non-radiating. Pain is described as gnawing and burning, intermittent lasting 1-2 hours, and gotten progressively worse. Antacids used to alleviate pain but not anymore; nothing exacerbates pain. Pain unrelated to daytime or to meals. Patient denies constipation or diarrhea. Patient denies blood in stool but have noticed them darker. Patient also reports nausea. Denies recent illness or fever. He also reports fatigue since 2 weeks ago and bloating after eating.  
Pre-existing / Developed Conditions Impacting Hospital Stay: RDS: Negative except for above findings  
Meds: Motrin once/week. Tums previously.  
PMHx: Back pain and muscle spasms. No Hx of surgery. NKDA.  
FHX: Uncle has a bleeding ulcer.  
Social Hx: Smokes since 15 yo. 1/2-1 PPD. No recent EDW use. Denies illicit drug use. Works on high elevation construction. Fast food diet. Exercises 3-4 times/week but stopped 2 weeks ago.

**Discharge:**  
Summary: some activity restrictions suggested, full course of antibiotics, check back with physician in case of relapse, strict diet.

**Report**  
Session Number: AF123456  
Procedure: 03/15/2020  
Attending: Dr. Mateo Jackson, MD

fic mucosa that is surrounded and underlying by a mass appears to have an approximately 1.5-cm ulcer. from the underlying tumor. The underlying mass consist of several different patterns. A single area having a fairly uniform appearance whereas; other appearance. Areas of yellow necrosis are scattered

er neoplasm. The cells are in part arranged in fascicles. In some areas, the fascicles have an interwoven of necrosis with the cells appearing to be surrounded. The lesions appear circumscribed, although not full thickness ulceration. The tumor immediately that spindle growth pattern. Some areas of the appearance. The cells appear to be arranged in. These areas also show a prominent mitotic activity. is numerous relatively open vascular channels which and in some areas appear to be nearly covered.

mentin, Smooth muscle actin, Desmin, Synaptophysin,

**Comment**  
Immunostains were performed on the core biopsy and demonstrate that the tumor cells are positive for CD117. The findings are consistent with the above diagnosis.

## Search across large volumes of documents

Use natural language to identify important sections of relevant documents

## Summarize key points in unstructured records

Allowing analysts to quickly comprehend large amounts of data

## Add insights from unstructured data to improve machine learning

Augment structured data with key insights from unstructured records

# Create BI dashboards with Natural Language

A NEW DASHBOARD BUILDING EXPERIENCE POWERED BY GENERATIVE BI

## Natural language visual creation

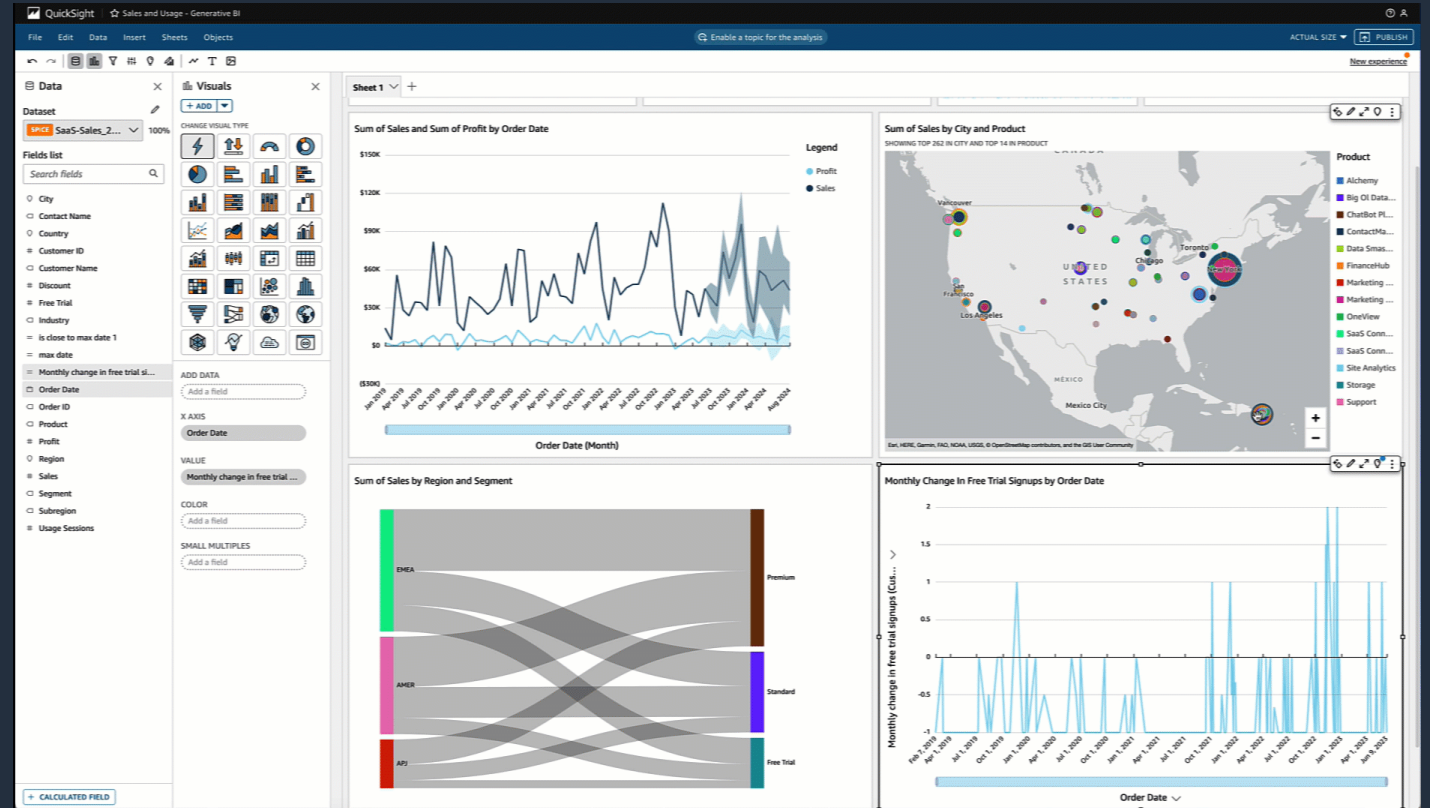
Use vague or precise language to generate and add visuals to a dashboard

## Quick calculations

Accelerate analysis by easily creating calculations without looking up or learning specific syntax

## Visual fine-tuning

Tailor visuals by describing the formatting changes in natural language



# Data stories and reports generated in and by QuickSight

IMPACTFUL DATA STORYTELLING TO DRIVE ACTIONS AND INSIGHT SHARING WITH PARTNERS

## Generate stories using AI

Generate cohesive, powerful, and insightful narratives analyzing data with only a few words

## Create refined content

Control AI verbosity, customize narrative text, and apply stunning visual themes to bring content to life

## Interpret data for others

Help others derive meaning from data and reach conclusions to drive decisions

## Governed and always up to date

Quickly and easily share with people who already have access to the data and update data any time

The screenshot displays the QuickSight interface for a data story titled "Unlocking Revenue Growth" with the main heading "How to Identify and Convert Your Most Valuable Free-Trial Customers". The story content includes an introduction and a section titled "Understanding Usage Sessions" which features a horizontal bar chart. The chart is titled "Free trial customers with more than 5 usage sessions last month" and shows data for six categories. The "Build for me" panel on the right allows for customizing the narrative and theme.

**Build for me**

Describe the story you want to tell in simple language.

Visuals will be picked from pinboard

Create a narrative that can help sales teams identify the free-trial accounts that are most likely to convert to paying customers. Use number of usage sessions as the driving factor for the strategy but also include sales data and forecasts.

Select the theme for your story

Default Modern DUSK Vibrant

CLOSE BUILD

Customer Segment	Usage Sessions (Last Month)
Enterprise	11
Mid-Market	10
Small Business	9
Startup	8
Freelance	7
Individual	6

# CHOP Accelerates Pediatric Research Using AWS-Powered Data Resources

## Challenge:

The Children's Hospital of Philadelphia (CHOP) wanted to address the challenge of storing and organizing increasing amounts of clinical data so that researchers could access, study, and cross-reference it to facilitate medical breakthroughs.

## Solution:

CHOP used AWS to build the Gabriella Miller Kids First Data Resource Center (KFDRC), a transformative healthcare discovery solution for multi-disciplinary care. The Center harnesses genomic, clinical, and imaging data to derive insights for children globally across a wide spectrum of diseases.

## Benefits:

- Provides the research community with access to genomic and associated clinical data
- Indexed 1.5 PB of genomic, clinical, and imaging data within 1 year
- Increased KFDRC's collaborative potential



**“Our entire system is built on AWS... we went from zero to managing a few petabytes of genomic data within a year using this setup.”**

# Rush University System for Health Creates a Population Health Analytics Platform on AWS

## Challenge:

Rush University System for Health (RUSH) is committed to addressing the underlying causes of the 16-year life expectancy gap among minority and lower-income residents of Chicago's West Side. RUSH sought to build a comprehensive analytics solution to identify and inform scalable interventions for equitable healthcare based on clinical, cardiometabolic, and social needs.

## Solution:

With support from Amazon Web Services (AWS), RUSH developed the Health Equity Care & Analytics Platform (HECAP). This platform transforms, aggregates, and harmonizes data from different sources to reflect the complex interplay of clinical and social factors on patient health.

## Benefits:

- Aggregates data from multiple sources using HIPAA-eligible services
- Builds a complete patient profile using clinical, social, and patient-generated data
- Produces risk scores and prediction modeling to guide clinical and community intervention
- Advances health equity for minority and underserved patient populations



“We have a great opportunity to start bringing in more data from different sources and use the power of AWS to scale massively across our system, significantly benefiting the care of our patients in Chicago.”

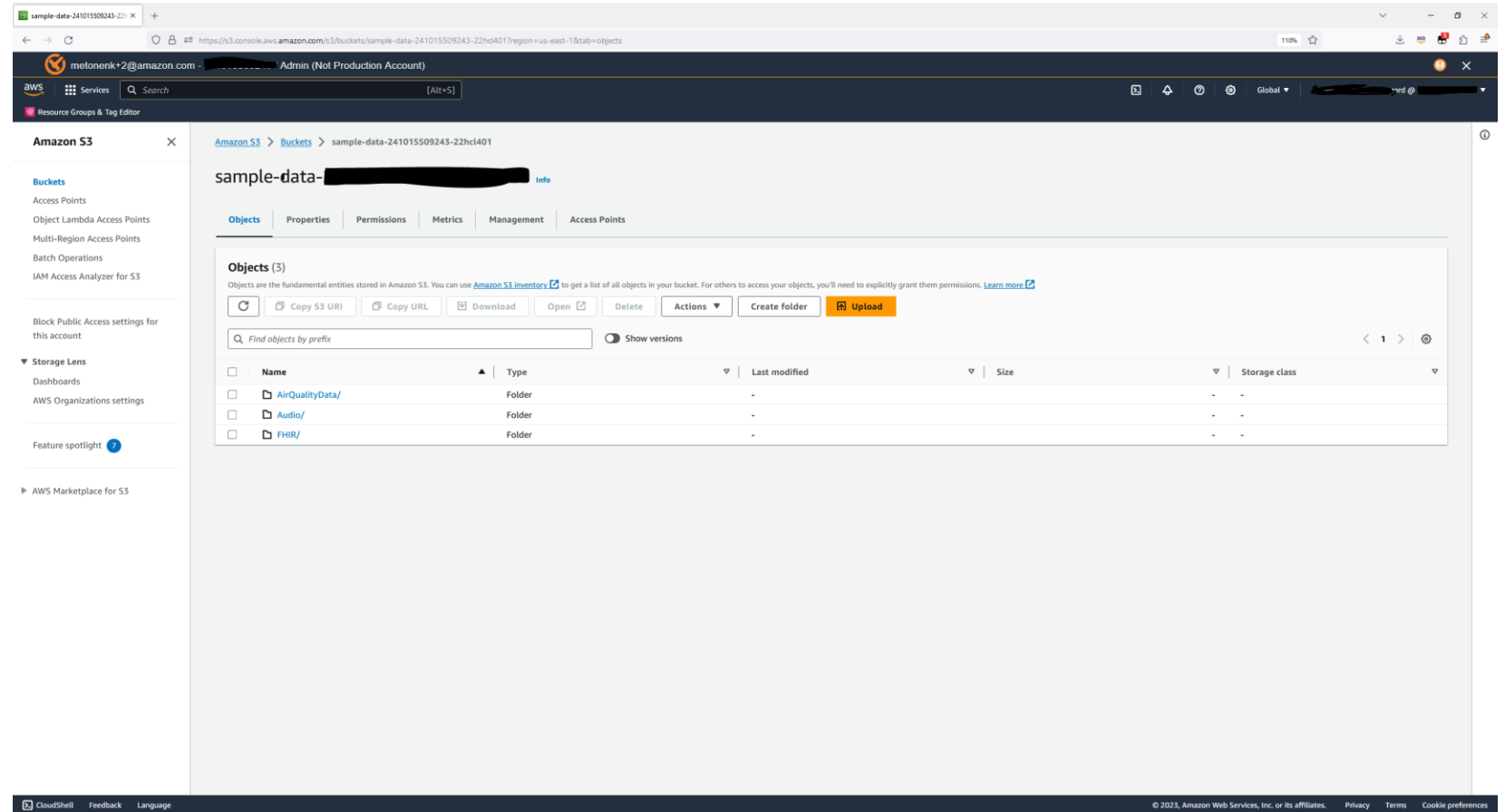
[Learn more](#)



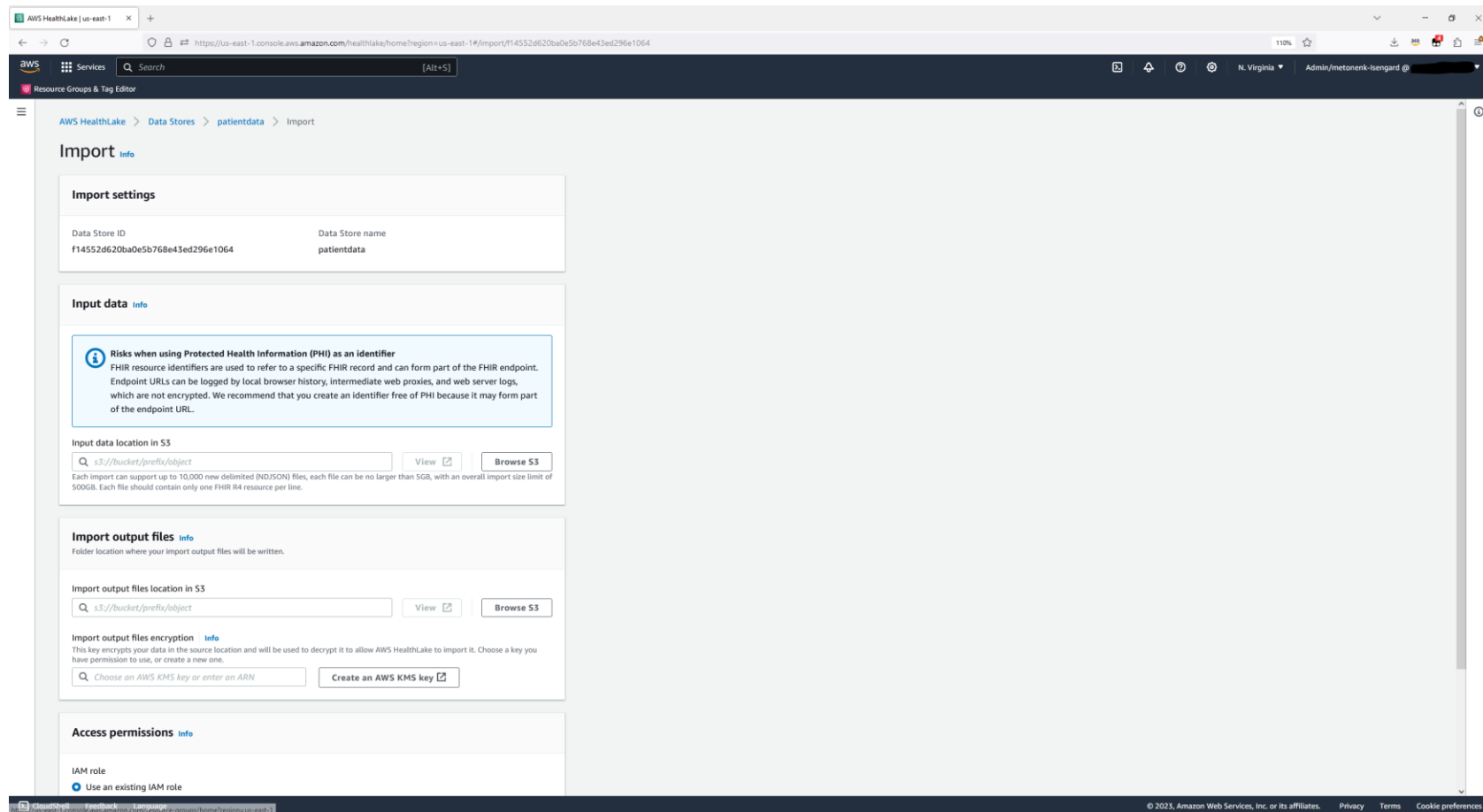
**Let's derive new insights by combining FHIR health data in AWS HealthLake with non-FHIR environmental data**



# Use Amazon S3 as data lake for FHIR and non-FHIR data



# Import files from an Amazon S3 bucket into AWS HealthLake data store



# Query HealthLake data from FHIR data source to get the gender count

The screenshot displays the Amazon Athena Query Editor interface. On the left, the 'Data' panel shows the data source as 'AwsDataCatalog', the database as 'healthlake\_db', and a list of tables and views. The main editor shows a SQL query:

```
1 SELECT gender,
2   gender_count,
3   CAST(gender_count AS REAL) / (
4     select count(id)
5     from "healthlake_db"."patient"
6   ) as percentage
7 FROM (
8   SELECT patient.gender,
9         count(id) as gender_count
10  FROM "healthlake_db"."patient"
11  GROUP BY patient.gender
12 )
13
```

The query has been executed successfully, as indicated by the 'Completed' status. The results are shown in a table with 2 rows:

#	gender	gender_count	percentage
1	male	55	0.5045872
2	female	54	0.49541286

The interface also shows the SQL editor with buttons for 'Run again', 'Explain', 'Cancel', 'Clear', and 'Create'. The bottom status bar indicates the query was completed in 4.589 seconds, scanning 4.96 KB of data.

# Query HealthLake data with Athena to get patient demographics

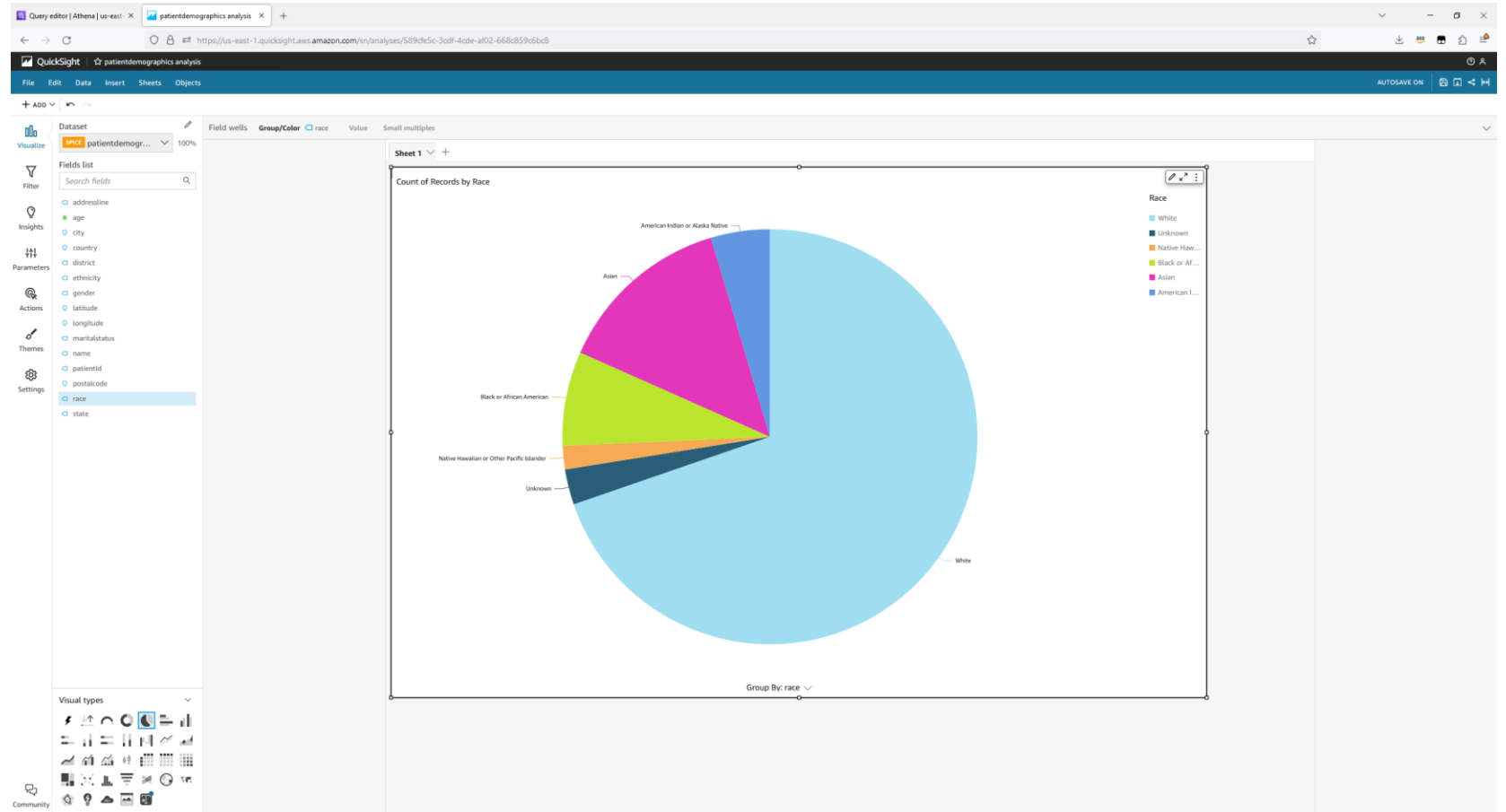
The screenshot displays the Amazon Athena Query Editor interface. The SQL query being executed is:

```
SELECT patient.id as patientId
2 , CONCAT(name [ 1 ],family, ' ', name [ 1 ].given [ 1 ]) as name
3 , (year(current_date) - year(date(birthdate))) AS age
4 , gender as gender
5 , CAST(json_extract(patient.extension[1], '$.extension[0].valueCoding.display') AS VARCHAR) AS race
```

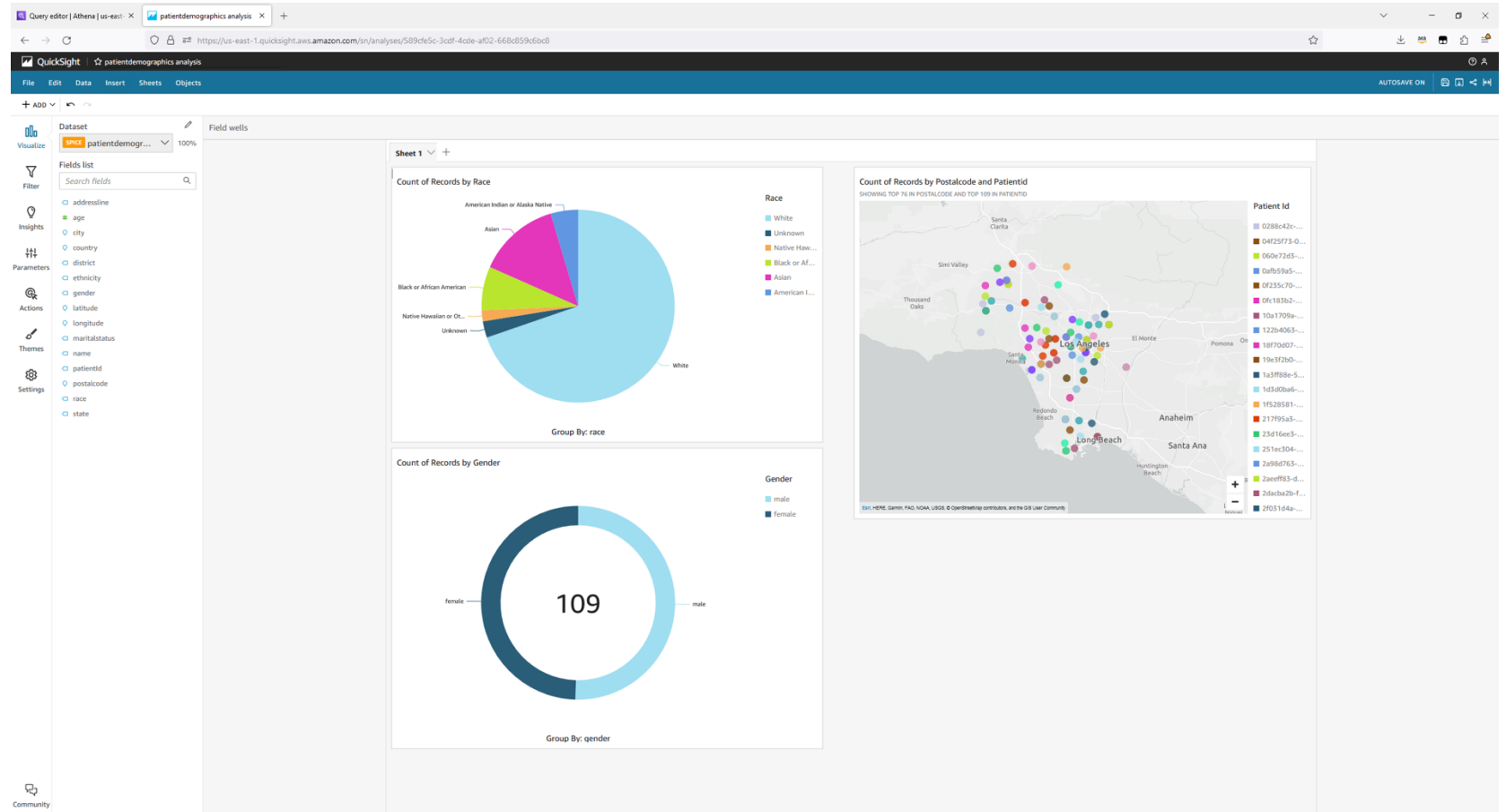
The query results are displayed in a table with 109 rows. The columns are: #, patientid, name, age, gender, race, ethnicity, maritalstatus, addressline, city, district, and state. The results show patient demographics such as name, age, gender, race, ethnicity, marital status, address, city, district, and state.

#	patientid	name	age	gender	race	ethnicity	maritalstatus	addressline	city	district	state
1	a7b903a5-5c6f-3f5e-4035-62d94b0b52b3	Abernathy Lavonia	11	female	White	Not Hispanic or Latino	Never Married	702 Hane Boulevard	Los Angeles		CA
2	f0bfa360-a7b8-a4f1-1ba4-1dc9952c2e05	Hackett Teresa	34	female	Asian	Not Hispanic or Latino	M	224 Grant Bridge Suite 91	Los Angeles		CA
3	9e84e569-7adc-ff42-ccdb-9fe9c23842a6	Hilll Armando	39	male	Black or African American	Not Hispanic or Latino	M	397 Rodriguez Promenade Suite 93	Los Angeles		CA
4	7fd8f9c-cb82-c4e3-109e-775bc8d3482b	Hettinger Werner	53	male	White	Not Hispanic or Latino	M	154 Larson Branch	Los Angeles		CA
5	f066d412-c40e-4d5c-5b11-08b6da289331	Corral Adela	45	female	White	Hispanic or Latino	S	827 Kling Burg Suite 32	Los Angeles		CA
6	8bf6a92-645f-3c82-ddcd-5851496a6aa8	Kling Ethelyn	12	female	White	Hispanic or Latino	Never Married	372 Jakubowski Vista Unit 76	Los Angeles		CA
7	e984f023-1987-15a9-f24b-f980ce800206	Madrid Guillermo	65	male	White	Hispanic or Latino	M	754 Moore Road Apt 77	Los Angeles		CA
8	aade3c61-92bd-d079-9d28-0b2b7fde0fbb	Hoppe Sal	28	male	American Indian or Alaska Native	Not Hispanic or Latino	M	930 Goldner Stravenue Apt 8	Los Angeles		CA
9	b45719be-ad92-ca35-29be-daaac4f258b	Collins Pierre	51	male	Asian	Not Hispanic or Latino	S	568 Stokes Course Unit 49	Los Angeles		CA
10	d488232e-bf14-4bed-08c0-a82f34b6a197	Walter Earnest	2	male	White	Hispanic or Latino	Never Married	677 Ritchie Terrace	Los Angeles		CA
11	5b2f8ddd-408a-884c-00e9-047b13e6cc45	Ratke Refugio	52	male	Black or African American	Not Hispanic or Latino	M	692 Lakin View Unit 66	Los Angeles		CA
12	18f70d07-28a0-17f3-2605-bf21cceb194b	Fahey Jackson	27	male	Asian	Hispanic or Latino	Never Married	631 Muller Junction Apt 77	Los Angeles		CA
13	50aac3ff-d836-adfd-620d-662fa08fd55	Urrutia Eloisa	76	female	White	Hispanic or Latino	M	638 Hamill Orchard	Los Angeles		CA
14	74b7fc1-8223-c477-0ec9-5b2ceea7c28f	Jacobs Ceola	42	female	White	Not Hispanic or Latino	Never Married	856 Gerlach Branch	Los Angeles		CA
15	04f25f73-04b2-469c-3806-540417a0d61c	Satterfield Maye	63	female	White	Hispanic or Latino	M	392 Renner Divide Suite 57	Los Angeles		CA
16	afd63046-5213-0c98-561f-12e6725cd1f5	Cremin Pattie	28	female	White	Not Hispanic or Latino	M	1020 Dibbert Crossroad	Los Angeles		CA
17	9acc871f-b577-5550-b8ad-fa95b58cea25	Maggio Edmund	46	male	White	Not Hispanic or Latino	M	446 Kreiger Route	Los Angeles		CA
18	f731ed0-1a09-5846-fb33-070164b703	Evans Emerson	80	male	White	Not Hispanic or Latino	M	312 Midway Park Apt 81	Los Angeles		CA

Use patient demographics dataset to visualize the demographic snapshot of patients. Get insights on age, race, and gender



# Visualize the location of the patient population from the zip codes table



# Join the FHIR data in HealthLake with non-FHIR data in Amazon S3

The screenshot displays the Amazon QuickSight interface. On the left, the 'Schema Explorer' shows the 'patientasthma' data source with a schema containing 'patientdemographics' and 'Athena'. The main area shows a 'Custom SQL' query named 'patientasthma' with the following code:

```
1 WITH weeklyAsthEncounters(totalAsthmaEncounters, weekNumber) AS (  
2   SELECT COUNT(start_enc) AS totalAsthmaEncounters,  
3     weekNumber  
4   FROM (  
5     SELECT en.period.start AS start_enc,  
6           FLOOR(  
7             day_of_year(from_iso8601_timestamp(en.period.start)) / 7  
8           ) --weekly count of Asthma encounters  
9     AS weekNumber  
10    FROM "healthlake_db"."encounter" AS en  
11    WHERE en.reasonCode [ 1 ].coding [ 1 ].code IN ('233670000', '195967001')  
12    AND en.period.start > '2021-01-01'  
13    AND en.period.start < '2021-12-31'  
14  )  
15  GROUP BY weekNumber  
16  ORDER BY weekNumber  
17 ),  
18 epaAqI(aqiWeekTotal, weekNumber) AS (  
19   SELECT SUM(aqi) AS aqiWeekTotal,  
20     weekNumber  
21   FROM (  
22     SELECT "daily_aqi_value" AS aqi,  
23           CAST(  
24             FLOOR(  
25               DAY_OF_YEAR(DATE_PARSE(date, '%m/%d/%y')) / 7.0
```

Below the query, the 'Datajet' table shows the results:

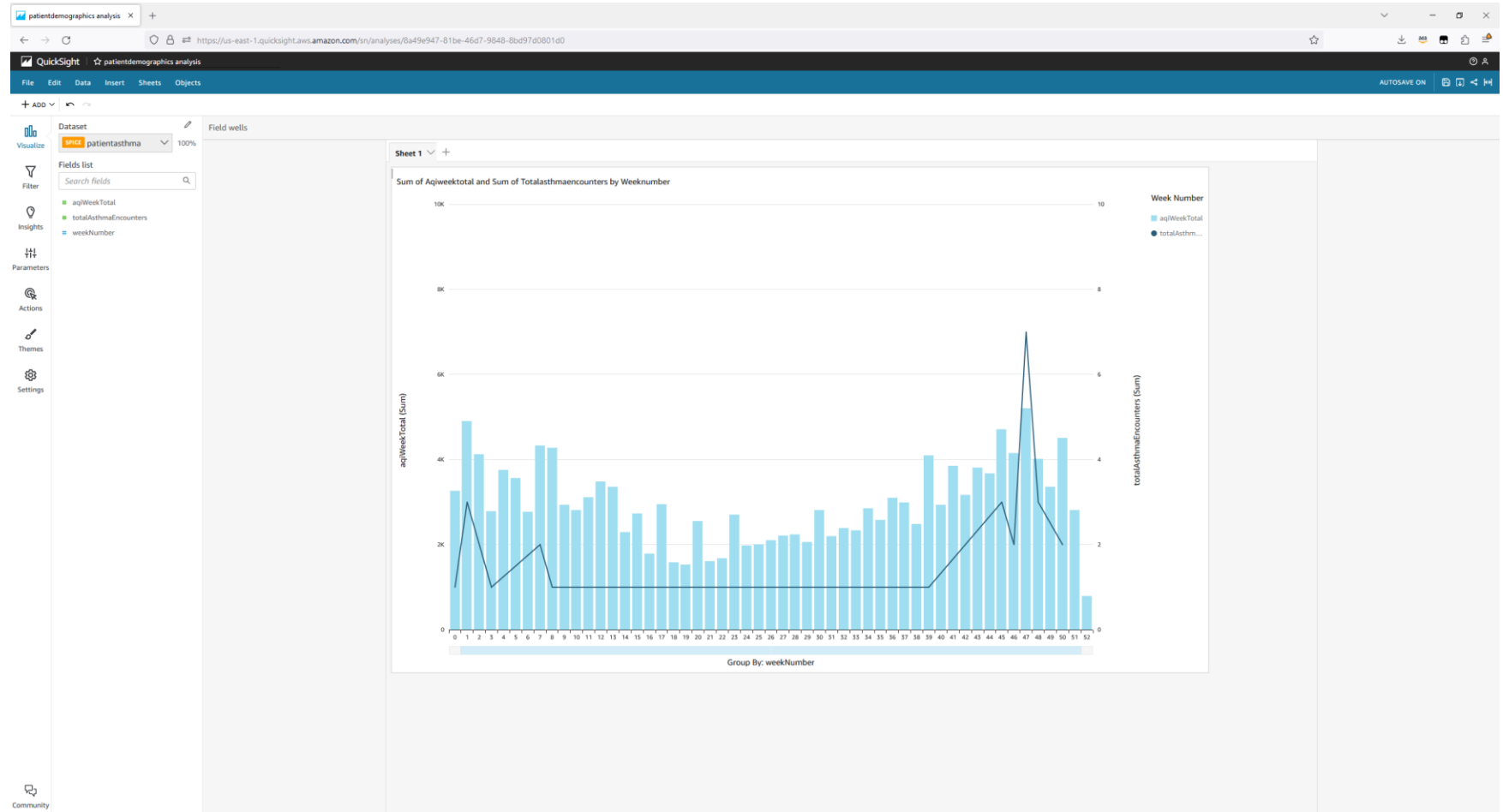
weekNumber	aqiWeekTotal	totalAsthmaEncounters
0	3262	1
1	4891	3
2	4119	2
3	2780	1
4	3749	
5	3564	
6	2775	
7	4323	2
8	4272	1
9	2926	
10	2809	
11	3109	
12	3476	1
13	3356	
14	2282	

# Visualize joined FHIR and non-FHIR data with QuickSight

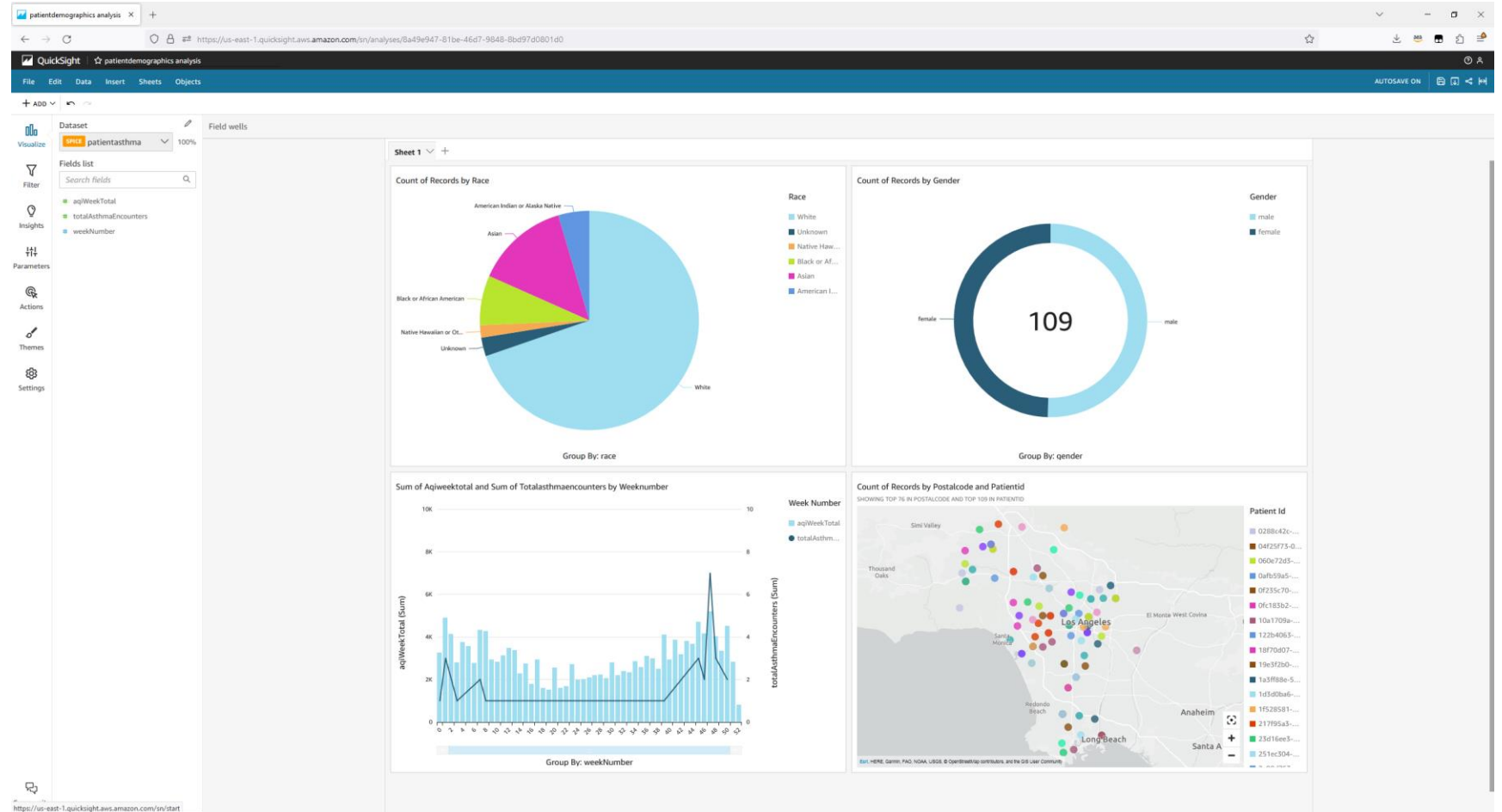




# Visualize the trends for asthma diagnosis cases and the air quality



# Visual of population demographics with combined FHIR and non-FHIR data sources



# Next steps

# How to get started



## Schedule a technical deep dive

with AWS Healthcare SMEs and Solution Architects to enable a well-architected design



## Develop and deploy a proof of concept

implementation with AWS and APN partners



## Training

to educate your workforce in supporting population health infrastructure



# Thank you!

Henrik Balle

henballe@amazon.com

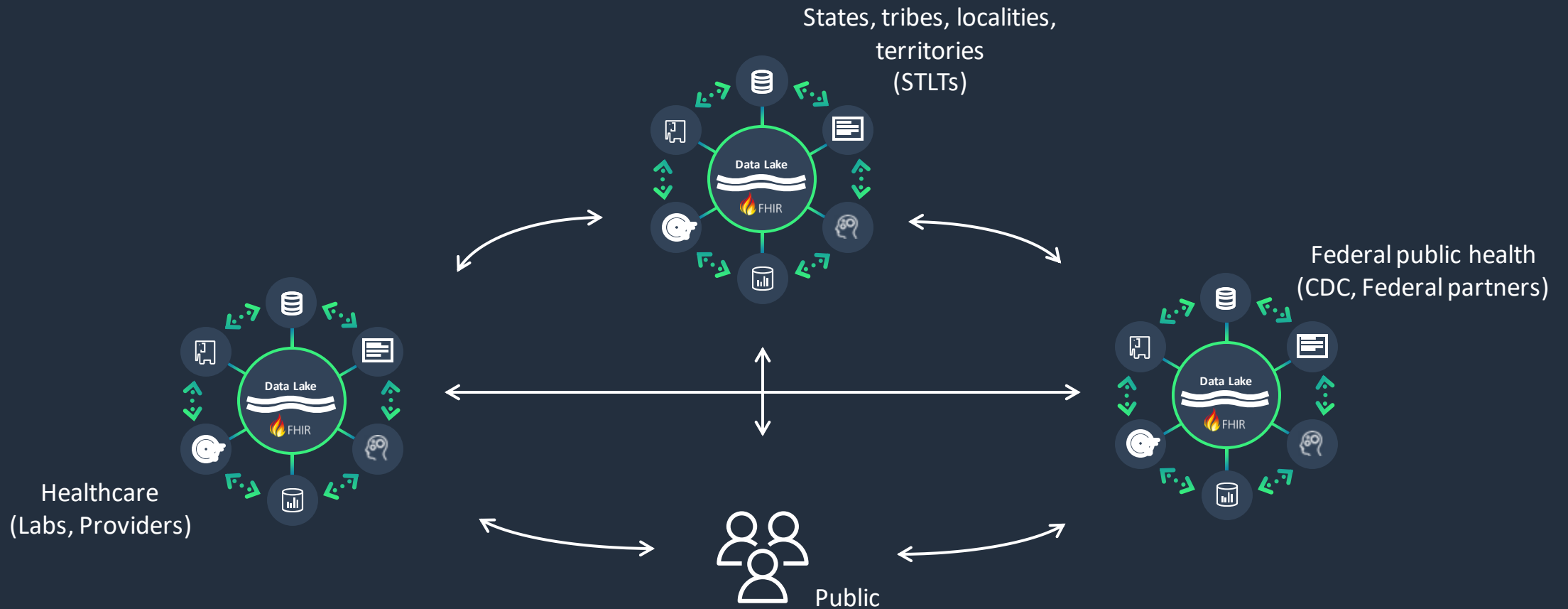
Marc Etone Nkwelle

metonenk@amazon.com

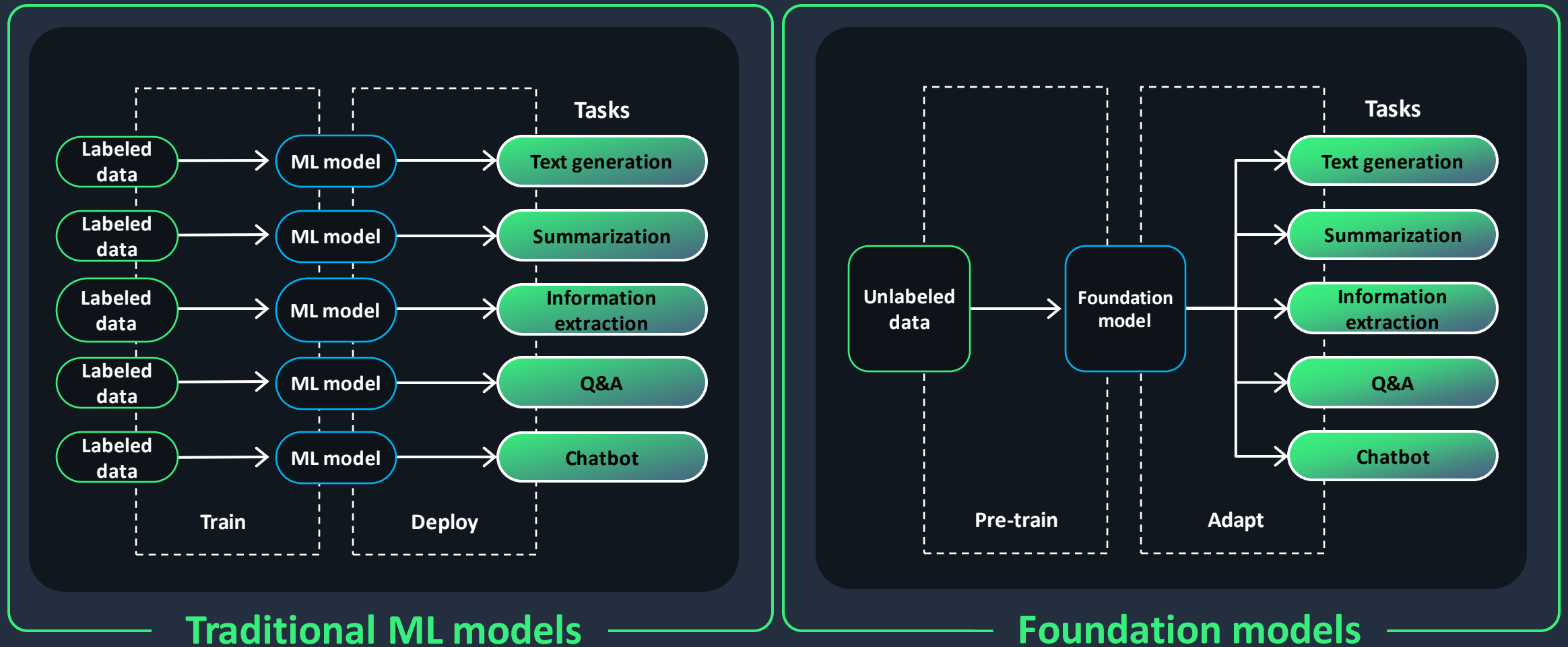
Ena Holmes

enaholme@amazon.com

# FHIR based data sharing across the Interconnected Public Health Ecosystem

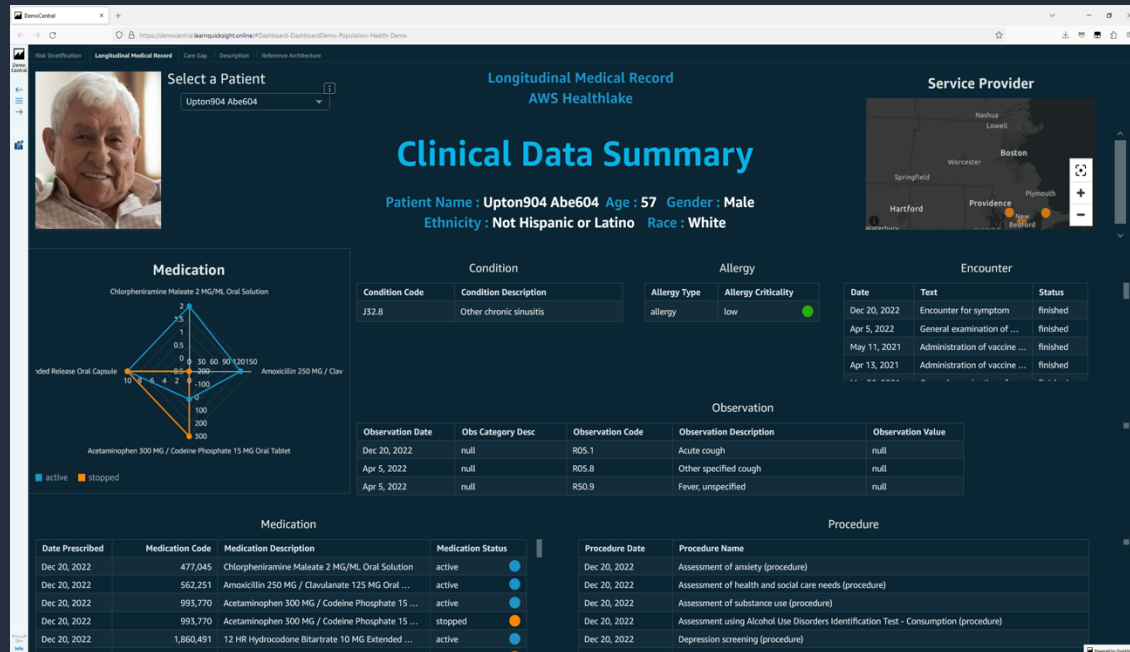


# How foundation models differ from other ML models

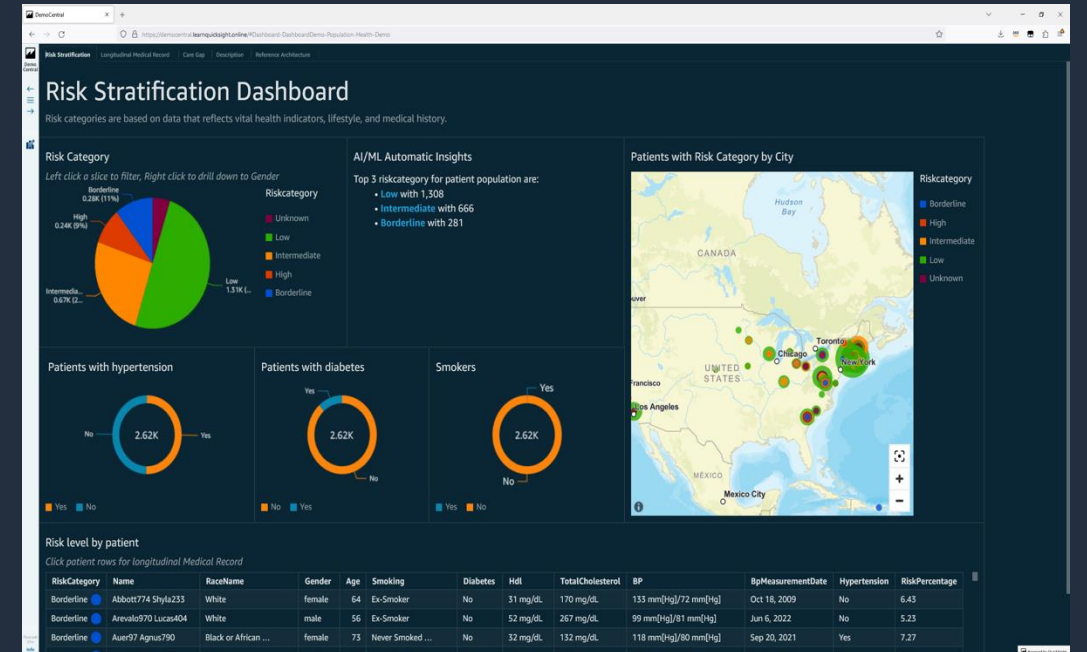


# Visualize patient and population data in QuickSight

## Longitudinal patient dashboard



## Risk stratification for population





# Public Health with AWS HealthLake Demo Architecture

