



Secure, Data-driven, Mobile: AI Delivers Personalized Medicine

Case Study



partner
network

SaaS Architecture
Competency

Data-Driven Personalized Medicine

For chronic disease sufferers – often already under medical care – data-driven personalized medicine provides a promising new application for tailored therapies and disease management.



INDUSTRY

Personalized Medicine

SOLUTION

Blockchain Distributed Ledger, secure HIPAA data at scale, hybrid cloud.

CLIENT

Gali Health

Gali Health

For chronic disease sufferers – often already under medical care – data-driven personalized medicine provides a promising new application for tailored therapies and disease management. Today, the right combination of technology infrastructure, advanced AI technology and analytics provide the power to continuously process infinite amounts of chronic disease-related data.

Translating these sources and insights to make data relevant and timely can make collaboration faster and more productive for patients, providers, and scientific research communities alike. It's a fortuitous bargain: each of these three communities can benefit from the value of each other's data, each one as suited to its unique needs.

CloudGeometry built and deployed a coherent data architecture spanning sources ready for continuous consumption by machine learning models. We have built this out for Gali, an advanced personalized healthcare platform to provide tailored advice to people suffering from chronic disease. Assuring the flow of data to models ensures every single patient always gets the best data to better manage her disease.

Data Ingest from medical research partners
AWS-native data services, including Aurora, RDS-Postgres & Aurora-MySQL

Redshift powered data warehouse & data lake
Amazon Redshift, Amazon S3, Amazon Athena, Amazon EMR

Machine Learning
Amazon SageMaker platform to create a robust process for developing, training and running ML models.

Consumer-facing mobile assistant
Building the app with blockchain technology helps make it both decentralized as well as highly secure.

Continuous Data Integration
Software development environment driven by of "infrastructure as code."

Gali Health

The Challenge

A critical success factor for personalized medicine is protecting patients data privacy. Blockchain is a natural choice to ensure confidentiality, but its most important benefit is reliable data aggregation. In practice, turning the broad and rich mix of data into more effective collaboration requires converging three domains of data: patient-generated health data (PGHD) with data repositories from clinical records (EHRs) and research data. Implementing this approach to data infrastructure and data processing meant an agile setup suited to continuous change. Key requirements were to:

- Pull data updates from research and clinical records and integrate with streaming data sources
- Stage the integrated data from both relational and streaming sources with well-specified APIs
- Provision well-structured data resources for both interactive analytics and machine learning ingestion
- Readily integrate new custom data formats and feeds from all three domains
- Provide for the integration of new applications of the data across all participants: clinical patient management, research organizations, and end-user patients
- Accommodate changes to streaming data ingestion logic so data scientists can expand the scope of machine learning experimentation as applications change

The focus was to ensure a virtuous cycle between the intake of data from changing sources, and changes in consumption of the data to provide better outcomes for researchers, clinicians, and patients.

**Data Ingest
from medical
research
partners**

AWS-native data services, including Aurora, RDS-Postgres & Aurora-MySQL

**Redshift
powered data
warehouse &
data lake**

Amazon Redshift, Amazon S3, Amazon Athena, Amazon EMR

**Machine
Learning**

Amazon SageMaker platform to create a robust process for developing, training and running ML models.

**Consumer-
facing mobile
assistant**

Building the app with blockchain technology helps make it both decentralized as well as highly secure.

**Continuous
Data
Integration**

Software development environment driven by of “infrastructure as code.”

The Solution

Data Ingest from medical research partners

The recent rapid evolution of standardized electronic health records (EHRs) anticipated the radical improvements in data storage costs, often viewed through the lens of “big data”. It would be a mistake to diminish the value of uniform, well-structured information about each patient’s condition and demographics, across a huge spectrum of clinical processes. Across different institutions, those transactions are read and written in that stalwart workhorse, the relational database.

Research organizations working with clinical data take the same approach. Data shared via the Fast Healthcare Interoperability Resources ([FHIR](#), pronounced “fire”) standard, created by the Health Level Seven International (HL7) health-care standards organization, levels the playing field for data management.

Standards not with standing, managing separate databases still imposes a lot of overhead, from backup to admin and all steps in between. Amazon Web Services provide a compelling alternative to the traditional standalone RDBMS without compromising the power of semantic compatibility. AWS-native data services, including **Aurora**, **RDS-Postgres** and **Aurora-MySQL**, preserve perfect semantic transparency but eliminate the cost and performance disadvantages stand-alone databases.

In many cases, third-party research partners maintain multiple instances of data that were functionally identical, albeit with different data. For example, different research trials ran the same business processes for different patients, even though higher-level variations between fact tables and schemas were immaterial. This presented an excellent opportunity for database consolidation. CloudGeometry created a single, more cost-effective Aurora instance that faithfully organized and managed data originating from different sources.

**Data Ingest
from medical
research
partners**

AWS-native data services, including Aurora, RDS-Postgres & Aurora-MySQL

**Redshift
powered data
warehouse &
data lake**

Amazon Redshift, Amazon S3, Amazon Athena, Amazon EMR

**Machine
Learning**

Amazon SageMaker platform to create a robust process for developing, training and running ML models.

**Consumer-
facing mobile
assistant**

Building the app with blockchain technology helps make it both decentralized as well as highly secure.

**Continuous
Data
Integration**

Software development environment driven by of “infrastructure as code.”

Gali Health

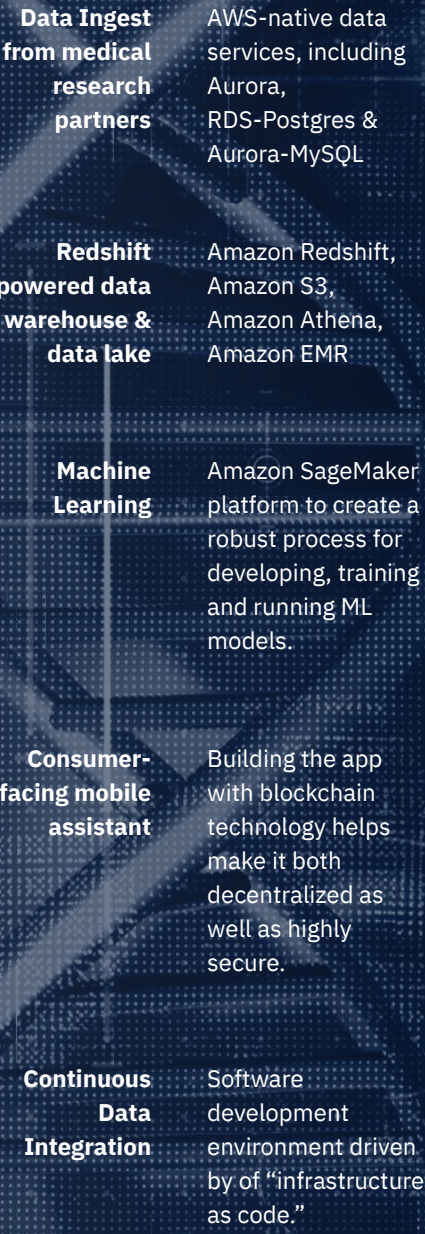
The Solution

Using the [Amazon Database Migration Service](#) simplified setting up and running a consolidated target instance on Aurora. It lets the integrated data platform benefit from running many database instances, each from its different independent third parties, at a fraction of the cost of maintaining a standalone instance for each respective partner. And because the AWS Database Migration Service features continuous data replication, the Aurora instance is always up to date, synchronizing with source data from wherever the 3d party instances run.

Redshift powered data warehouse and data lake

The big change in the era of big data is that not all data lives in a single database controlled by a single well-bounded business process. In addition to the process of collecting and consolidating data from multiple 3rd party SQL databases running on AWS Aurora, the application’s platform can benefit from many other data formats. These include sources originating with a partnered research institution, as well as data generated by the company’s apps and end-users. The ability to quickly merge this data into query-able datasets is at the core of the business.

The possibilities of integrating queries data across multiple streams endless. How many people with a given combination of height, weight, and age reported a change in health when they gained five pounds during the winter holiday season? How much did they exercise? For people who are lactose-intolerant, what should the personalized medicine app recommend as a change in diet? Do users respond more consistently to prompts through the chat interface, or is it more effective to provide scheduled reminders in a smartphone calendar?



Gali Health

The Solution

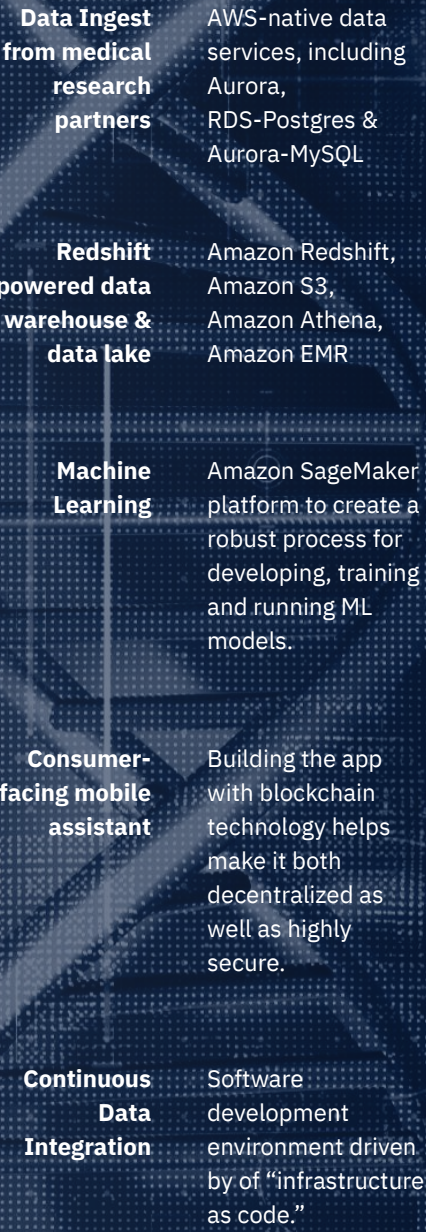
With such a broad landscape of experimentation, the challenge was to create a semantically enriched data warehouse object model. We chose AWS-native **Amazon Redshift** for a number of reasons:

- Powerful and versatile query interface for fast results for complex queries on any scale
- Unique combination of cost-effectiveness, scalability, and performance
- Extensibility to add new sources and grow their data footprint with minimal friction

This extensibility is powerful. For example, streaming data from end-users and application logs are stored in **Amazon S3**. Use of Redshift lets the system query both the data warehouse and S3 objects through a single query interface. Structured data stored in S3 objects can also easily be accessed via ad-hoc SQL querying with **Amazon Athena**.

Another database source implemented by CloudGeometry was the generating of conformed dimensions from streamed data using Stream Sets open source pipelining logic. Differences across data sources are addressed with a programmatic process that preserves field and record level compatibility. For example, research partners whose clinical trials have different nomenclature for steps of the process are aligned with a single uniform field name, creating a global fact table across the range of data sources.

The continuous expansion of data volume and variety, including the creation of brand new data sets, requires continuous data engineering. This data is stored on Amazon S3 in the data lake and available to direct load into downstream systems or interactive exploratory analytics. CloudGeometry established a process using **Amazon EMR** to process the data and create new data sets. The continuous flow of new data and new data sets into a unified model, powering the personalized medicine decision engine and the applications that rely on it.



Gali Health

The Solution

Machine Learning

New data is not restricted to the acquisition and ingestion of external sources. In addition to streamlining access to a wealth of possibilities for query exploration, CloudGeometry helped product operations and data science teams, along with 3d party partners, to continuously deliver new data sets for new experiments. It laid the groundwork for a well-managed process for creating, deploying, training, and optimizing new Machine Learning models.

The AI at the heart of Gali's product is known as *The Gali Brain*. It is made up of three key components:

- Health and Disease Models, which enable her to understand the context of specific health conditions and support users in their health journey;
- Behavioral Model that helps the AI interact intelligently in response to various events, provide advice or connect with a relevant coaching program or service;
- Deep Learning Layer, which ensures that Gali constantly learns from new users

Each of these three components has its data source flows and integrations; the deep learning layer is where the challenge of chronic disease management is most acute. This is how Gali gets “smarter” over time, through streaming machine learning (ML).

By definition, ML is a continuous process. CloudGeometry used **Amazon SageMaker** platform to create a robust process for developing, training and running ML models. Its framework-independent structure enables a consistent workflow for the steps needed to train, tune, and deploy various combinations of data and algorithms. SageMaker manages and automates the full range of sophisticated training and tuning techniques. As a result, ML models can stay ahead of changes to training inputs, despite changing new clients and research data.

**Data Ingest
from medical
research
partners**

AWS-native data services, including Aurora, RDS-Postgres & Aurora-MySQL

**Redshift
powered data
warehouse &
data lake**

Amazon Redshift, Amazon S3, Amazon Athena, Amazon EMR

**Machine
Learning**

Amazon SageMaker platform to create a robust process for developing, training and running ML models.

**Consumer-
facing mobile
assistant**

Building the app with blockchain technology helps make it both decentralized as well as highly secure.

**Continuous
Data
Integration**

Software development environment driven by of “infrastructure as code.”

Gali Health

The Solution

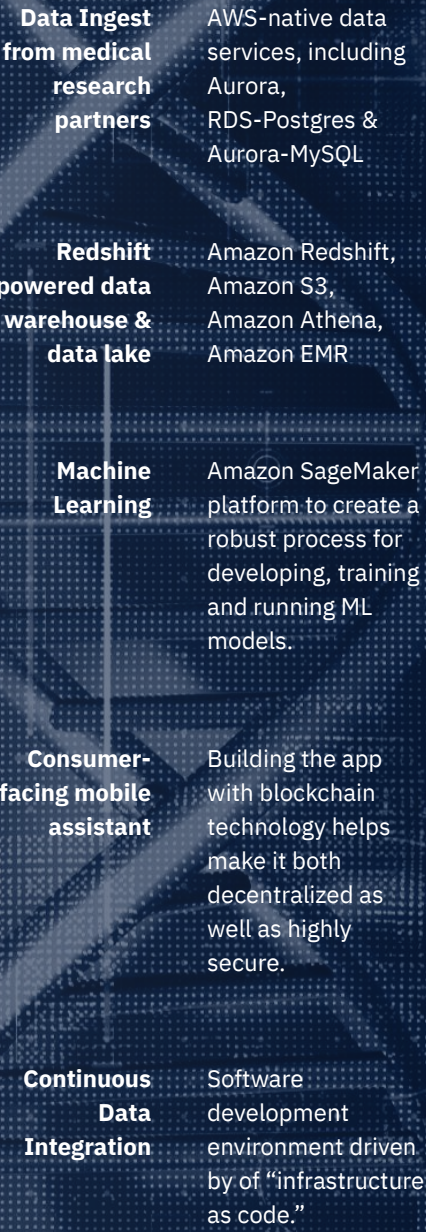
A key step is that when machine learning ingests clinical trial data, model recommendations are reviewed by doctors and scientists prior to approval for production use. CloudGeometry built this into the workflow for the release of new capabilities and features, so no critical functions are implemented without expert human oversight.

Consumer-facing mobile assistant

Thanks to an innovative data-driven mobile app, patients can benefit from the data in new ways without having to become data scientists.

Personalization is front and center throughout the platform: onboarding and ongoing health monitoring, interactive chat, reminders, and health tips. For example, for individuals with Crohn’s disease, it is essential to understand the stages of the disease and its subtypes, including known treatments, procedures, tests, and range of symptoms required for effective chronic disease management. The engine tracks these essential attributes and manages each user’s experience accordingly.

Building the app with blockchain technology helps make it both decentralized as well as highly secure. Daily health and lifestyle information combined with medical history and lab and genetic data create an enormous and valuable cloud with billions of data points from each person. Through the exchange of health data for tokens, patients not only get more tools to manage their disease. This strikes a new balance between providers and consumers of the data. Partners can benefit from access to patient community data, and offer additional services to community members with specific health backgrounds without conflicts in privacy interests.



The Solution

Continuous Data Integration

The data integration platform that drives these multiple application engines relies on a complex, integrated distributed infrastructure. The infrastructure relies on a process of continuous integration of changes to the software logic that drives the data processing at every layer of the stack.

CloudGeometry built out a software development environment driven by of “infrastructure as code.” It applies equally well to the underlying IaaS cloud infrastructure services as it does to data repositories and the data processing logic. All artifacts reside on a single consistent repository infrastructure. Each change to software lives exclusively in the source code, rather than through standard operating procedures and manual processes. It automates the infrastructure deployment process, be it data configuration or machine learning logic, in a repeatable, consistent manner.

The same approach applies to every step of development, deployment, test automation, release management, and production operations. With the rate of change on all the moving parts, the CloudGeometry DevOps team implemented a CI/CD process, based on the [CloudGeometry Solution](#).

The most immediate advantage of the CI/CD is in eliminating delays in coding and testing improvements by the platform development team. With a transparent, predictable release process, developers could readily push new software to manage the ongoing changes to all data and application interfaces.

Another key benefit of the CI/CD process was to eliminate disconnects and test escapes that software test automation can encounter in a system of such complexity. The built-in continuous testing approach is about controllable consistency. This guarantee that the same changes tested on in development are applied in production: the same DB changes, same app changes, in the same order, byte to byte, query by query.

**Data Ingest
from medical
research
partners**

AWS-native data services, including Aurora, RDS-Postgres & Aurora-MySQL

**Redshift
powered data
warehouse &
data lake**

Amazon Redshift, Amazon S3, Amazon Athena, Amazon EMR

**Machine
Learning**

Amazon SageMaker platform to create a robust process for developing, training and running ML models.

**Consumer-
facing mobile
assistant**

Building the app with blockchain technology helps make it both decentralized as well as highly secure.

**Continuous
Data
Integration**

Software development environment driven by of “infrastructure as code.”

Gali Health

The Benefits

Data engineering and cloud expertise delivered by CloudGeometry was a key success factor in introducing the revolutionary patient-centered approach that Gali Health and its AI bring to the challenges of chronic disease management, starting with Inflammatory Bowel Disease (IBD) — including Crohn’s and Ulcerative Colitis. This realizes the vision of secure data convergence across patients, clinicians, and researchers.

- Deliver on the promise of a virtuous cycle of data that attracted new research institutions, hospitals, domain experts and leading medical organizations to collaborate within what has historically been a siloed healthcare ecosystem
- Leverage the Amazon native data technology stack for data at any volume, velocity, variety, and value
- Raise the bar for data leverage by providing a rich, productive dataflow of consistent data for query access in ad-hoc research, as well as ML models that enable both advanced medical data science as well as innovative new applications driven by that data
- Give people who suffer chronic diseases a new way to improve their health with direct control of their data – and a more effective way to collaborate with the professionals who treat them.

Data Ingest from medical research partners	AWS-native data services, including Aurora, RDS-Postgres & Aurora-MySQL
Redshift powered data warehouse & data lake	Amazon Redshift, Amazon S3, Amazon Athena, Amazon EMR
Machine Learning	Amazon SageMaker platform to create a robust process for developing, training and running ML models.
Consumer-facing mobile assistant	Building the app with blockchain technology helps make it both decentralized as well as highly secure.
Continuous Data Integration	Software development environment driven by of “infrastructure as code.”

Technology Solutions & Services

to build, optimize and run your cloud software and operations.



TECHNOLOGY SOLUTIONS

APPLICATION MODERNIZATION

Transform Legacy Systems
for Future-Ready Innovation

CLOUD COST MANAGEMENT

Optimize Cloud Spending
for Maximum ROI and Efficiency

AI & DATA

Achieve breakthrough automation and analytic
insights via cutting-edge data strategies

SECURITY AND COMPLIANCE

Reinforce your cloud infra, workload,
operations, and development end-to-end

FOUNDATION SERVICES

CLOUD INFRA & OPS

Robust Infrastructure Solutions for Seamless
Operations

- DevOps as a Service
- Managed Cloud Operations
- Cloud Spend Optimization
- Resilience, Continuity & Backup
- AWS Well-Architected

MODERNIZATION & MIGRATION

Smooth Transitions to Modern Architectures
with Minimal Disruption

- Application Modernization
- Kubernetes Adoption
- AWS Database Migration
- Data Engineering Operations
- Data Integration
- Data Migration
- Cloud Migration & Adoption

ADVANCED SERVICES

CLOUD-NATIVE DEVELOPMENT

Build Scalable, Resilient Applications with
Cloud-Native Technologies

- CI/CD
- Multi- Platform App Design & Development
- Cloud-Ready Teams
- Enterprise SaaS Modernization
- Multi-Tenancy SaaS
- B2B Customer Success Engineering

CLOUD-NATIVE OPERATIONS

Streamline Operations with Advanced
Cloud-Native Practices

- Platform Engineering
- Workload Management
- Monitoring & Observability
- Infrastructure Management

AI/ML & DATA SERVICES

AI/ML & DATA

Leverage Advanced Analytics and Machine
Learning to achieve exponential acceleration in
the ROI of your data assets

- AI/ML Engineering for Data Analytics
- Generative AI
- Traditional ML for Data Analytics
- AI/ML Development and and Data Science
- Data Engineering for MLOps


San Francisco


New York


London


Berlin
Prague


Cluj-Napoca
Montenegro


Tel Aviv

CloudGeometry delivers expert technical services, helping our clients unlock the full potential of cloud-native open source tooling and commercial platform technologies.

With roots in Silicon Valley, we've seen firsthand what works (and what doesn't). Count on CloudGeometry to accelerate application modernization, Kubernetes adoption, developer enablement, secure multi-tenancy, AI/MLOps, DevOps automation and more.

- As **AWS Advanced Consulting** partners, our certified solution architects and platform engineers help address the range of challenges facing enterprise innovators and venture funded startups alike.
- The **Cloud Native Computing Foundation** has accredited us as a **Kubernetes Certified Service Provider**.
- We serve as charter contributors to the **Linux Foundation Data & AI Commons** (LF Data & AI), supporting a diverse, sustainable ecosystem for open source data and AI technologies.

Over the last decade, we've built and deployed hundreds of big, fast full-stack apps with well-engineered cloud infrastructure across industries: Financial Services, Industrial Automation, Healthcare, AdTech, Consumer-grade Mobile, smart devices, and more.

From enterprise upgrades to data engineering to cloud-native scale-out, CloudGeometry helps you plot the shortest path across all dimensions of modern cloud software engineering.

 partner
network

Advanced
Consulting
Partner

SaaS

Amazon Kinesis

AWS Lambda

Amazon Redshift

AWS Database
Migration Service



CLOUD NATIVE
COMPUTING FOUNDATION

GENERAL MEMBER
LF AI & DATA
GENERATIVE AI COMMONS



Expert
Platform Engineering
& DevOps

[CloudGeometry.io](https://cloudgeometry.io)
[+1 408 444-7061](tel:+14084447061)
info@cloudgeometry.io
github.com/cloudgeometry

Cloud Infrastructure
& Application
Modernization

100 S Murphy Ave #200
Sunnyvale, CA
94086 USA

AI & Machine Learning
Data Services

Open Source
Tooling & Integration



partner
network

Advanced
Consulting Partner

SaaS Architecture
Competency