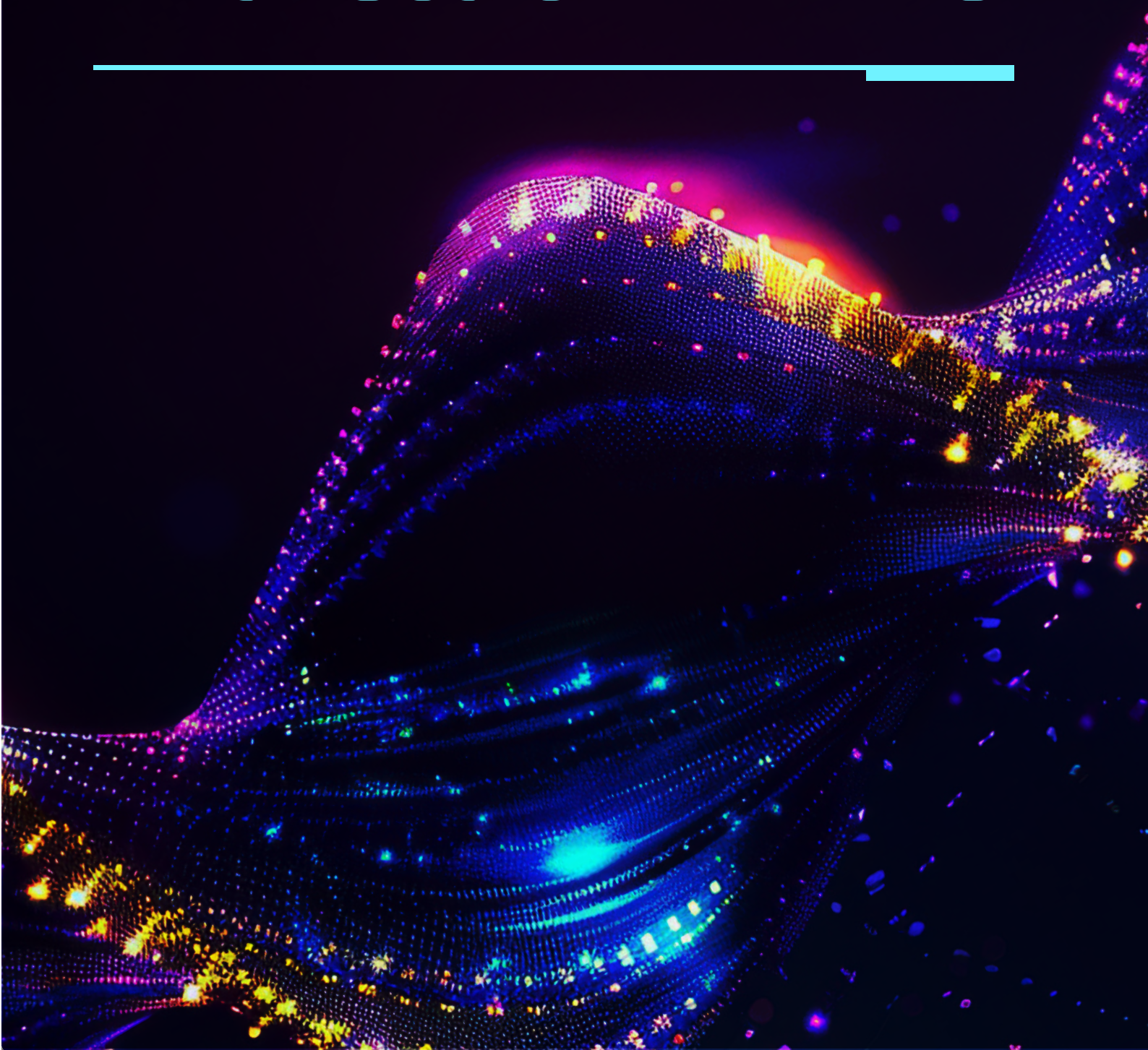
CitiusTech
Shaping Healthcare Possibilities

aws

CASE STUDY

Unlocking Enterprise AI at Scale with AWS



How a Health Plan Organization Streamlined their AI/ML Model Lifecycle Management

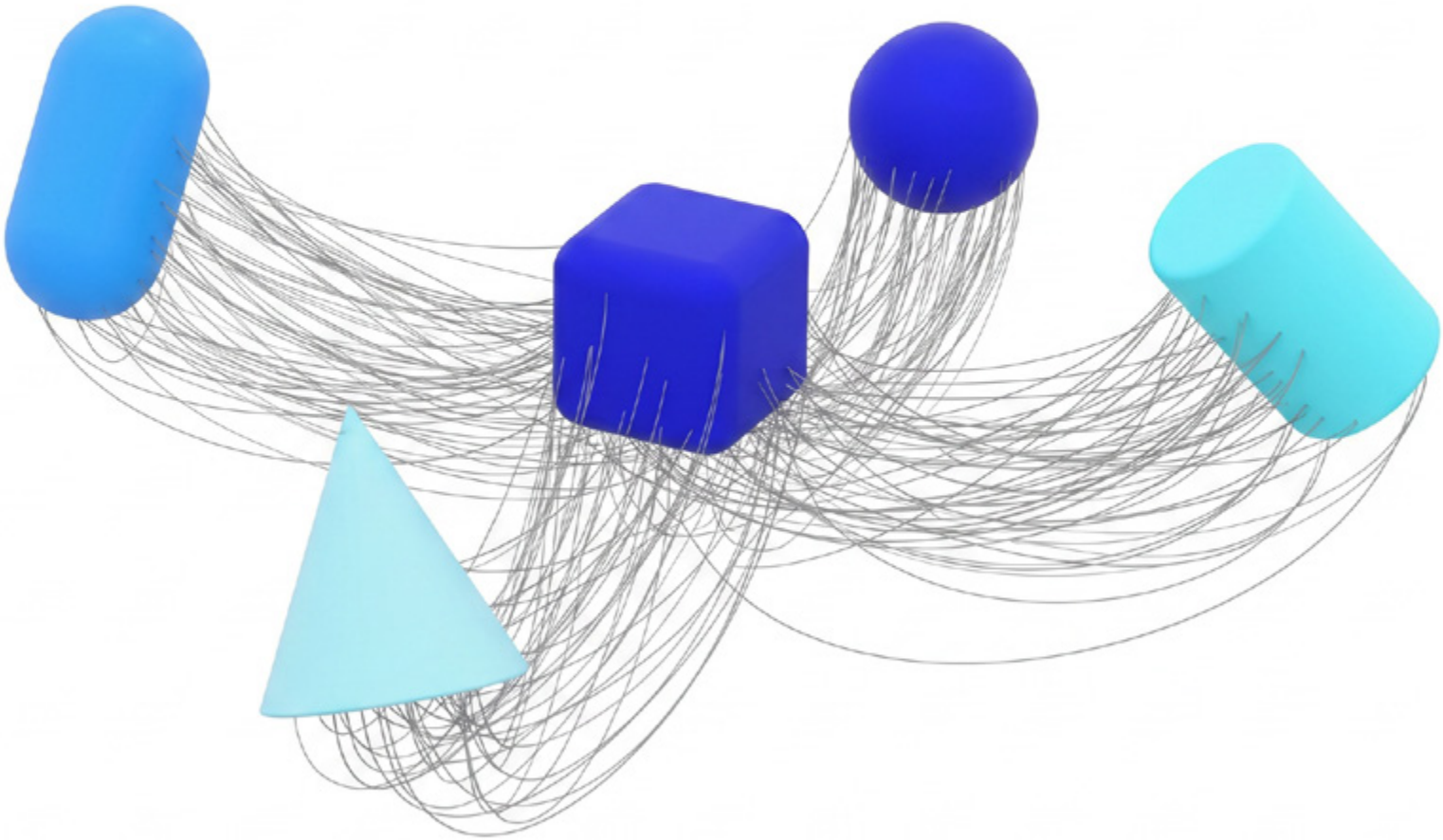
A leading U.S. health plan organization serving members across Oregon, Washington, Idaho, and Utah was on a multiyear journey to become a consumer centric, digitally enabled healthcare ecosystem.

The Challenge

The health plan organization was experiencing fragmentation across their on-premises AI/ML deployments that led to operational complexity. The client lacked standardized infrastructure patterns for AI model landing zones and deployment workflows.

This lack of standardization prevented the health plan organization from maintaining consistency and inhibited scalability of their AI program. To meet enterprise and regulatory requirements for the healthcare industry, the client needed to improve visibility, governance, and control over AI model performance, monitoring, and auditing.

The health plan organization was looking for a unified and standardized AI/ML operating model to centralize and streamline lifecycle management across disparate on-premises deployments.



Copyright © 2026 CitiusTech. All Rights Reserved.

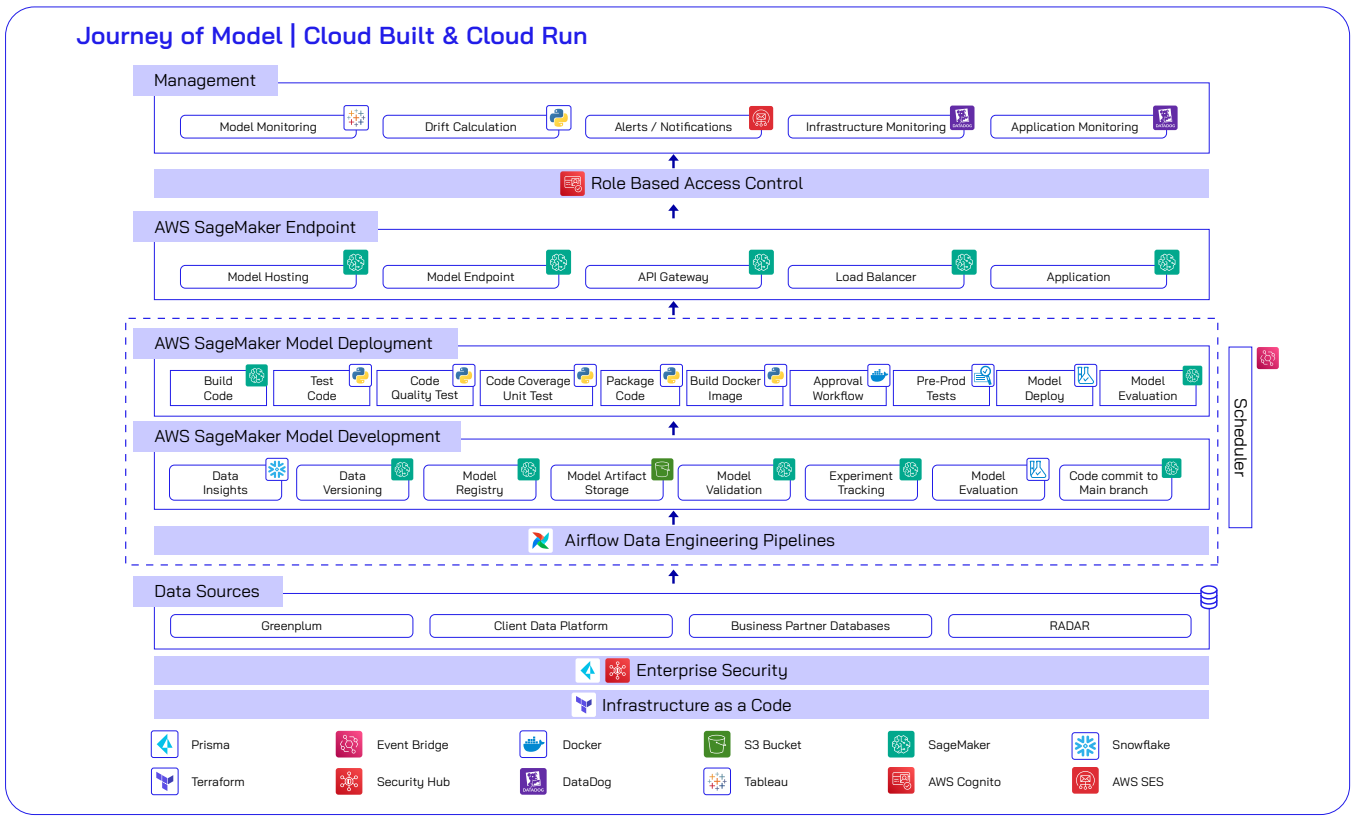
The Solution

CitiusTech and the health plan organization have been working together for over 9 years through more than 30 engagements. For this project, CitiusTech worked with the health plan organization to develop and implement a modular, enterprise-grade, and fully automated account-provisioning framework. This framework was designed to streamline and standardize the onboarding and governance of AI/ML models. By establishing standardized, templated infrastructure provisioning, CitiusTech laid the foundation for consistent and scalable deployments of ML models.

Leveraging cloud-native services, such as Amazon SageMaker and Amazon Bedrock, CitiusTech developed a centralized AWS platform configured to handle the end-to-end lifecycle of machine learning (ML) models, from data ingestion through deployment, monitoring, and governance, using containerization and automated pipelines. A diverse set of enterprise data sources, including Greenplum, client data platforms, business partner databases, and RADAR systems, provide the raw material required for building ML models. Once data is prepared, the model development lifecycle begins within AWS SageMaker.

The deployment phase is governed by a robust CI/CD pipeline built on AWS SageMaker model deployment processes. Once deployed, models are served through an AWS SageMaker endpoint. The architecture incorporates comprehensive monitoring and governance capabilities across both models and infrastructure. The entire pipeline is built on a secure, enterprise-grade foundation. A centralized scheduler coordinates workflows across the entire lifecycle, ensuring synchronization between data pipelines, model development, and deployment processes. This automation reduces manual effort and accelerates time to production.

By identifying relevant use cases, assessing feature alignment, and understanding integration requirements within an existing AWS environment, CitiusTech supported the health plan organization in evaluating Amazon Bedrock AgentCore as a means of building, connecting, and optimizing AI agents.



Copyright © 2026 CitiusTech. All Rights Reserved.

The Results

60% reduction in RAG development time

70% improvement in accuracy of RAG contextual retrieval

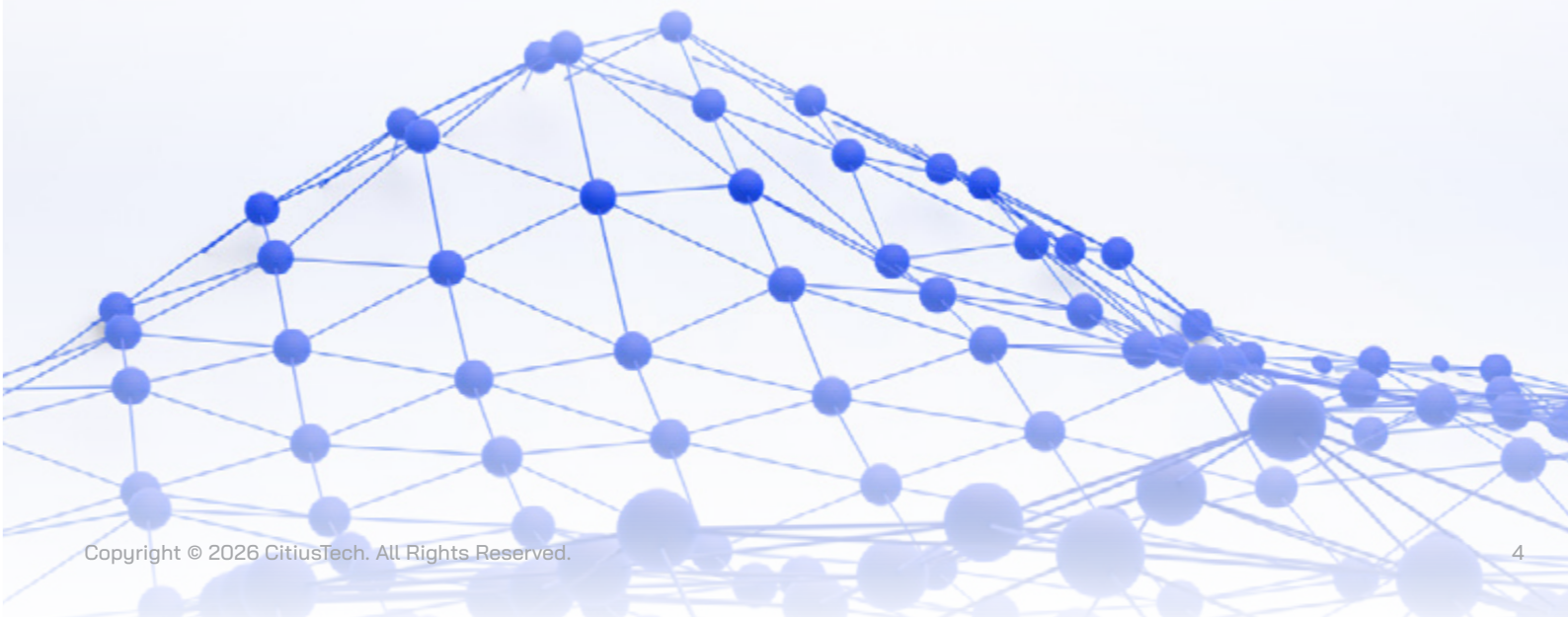
99% availability of production Amazon SageMaker

The Impact

Through a centralized AWS platform enhanced with Amazon SageMaker and Bedrock services, the health plan organization was able to take a more scalable, user-friendly, and cost-effective approach to AI/ML model lifecycle management. Now, the client can deploy ML models consistently and govern them to meet compliance requirements. The AWS platform supports advanced GenAI and Agentic AI capabilities to automate account provisioning and enhance member service.

By developing standardized and reusable Retrieval-Augmented Generation (RAG) patterns, CitiusTech reduced RAG development time by 60%. The client was also able to improve the accuracy of RAG contextual retrieval by 70% for optimized training of Large Language Models (LLMs).

Through the 99.9% availaility of production SageMaker, the health plan organization can architect endpoints to avoid downtime and ensure that the business keeps running.



Copyright © 2026 CitiusTech. All Rights Reserved.

CitiusTech

Shaping Healthcare Possibilities

About CitiusTech

CitiusTech, the digital vanguard, enables MedTech, Payers, Providers, and Life Sciences organizations to power healthcare digital innovation, business transformation, and industry-wide convergence through next-gen technology and human ingenuity. We are an AWS Premier Service Tier Partner and winner of 2024 AWS Partner of the Year.

Explore how to leverage CitiusTech's extensive Healthcare expertise and AWS's scalable cloud solutions to address today's digital challenges.

Reach out today

Copyright © 2026 CitiusTech. All Rights Reserved.