DATA ECONOMY

Modernizes ETL with AWS DevOps and Cloud Operations

Category

DevOps & CloudOps

Table of Contents

About the Customer

The customer is one of the world's leading healthcare organizations, delivering innovative solutions across pharmaceuticals, medical technologies, and healthcare services. Operating across multiple business units and global regions, the organization manages a vast ecosystem of enterprise applications, data platforms, and business-critical workflows that support research, manufacturing, supply chain, regulatory operations, and commercial functions.

As part of its enterprise modernization strategy, The customer initiated a program to transform legacy Extract, Transform, and Load (ETL) workflows into a scalable, cloud-native, AI-assisted modernization platform. The objective was to accelerate migration from legacy ETL technologies to modern PySpark-based data processing while improving engineering productivity, operational governance, security, and long-term maintainability.

The initiative required a solution capable of automating complex migration activities while maintaining enterprise-grade governance, security, and operational resilience. Beyond code transformation, the platform needed to provide centralized workflow orchestration, metadata extraction, lineage discovery, quality validation, and end-to-end traceability to support large-scale modernization programs across multiple business domains.

To support both application delivery and day-2 cloud operations, the solution was designed on AWS using a highly available, multi-AZ architecture that combines AI-assisted automation, Kubernetes-based microservices, secure infrastructure, centralized monitoring, and operational governance. The resulting platform enables customer to modernize ETL assets through repeatable, governed processes while establishing a scalable operational foundation for future modernization initiatives.

Customer Challenge

The customer's enterprise data landscape consisted of numerous legacy ETL workflows supporting business-critical reporting, analytics, and operational processes across multiple business functions. Many of these workflows were developed using legacy ETL platforms, making modernization a complex and resource-intensive effort. Manual analysis of workflow logic, metadata extraction, dependency mapping, and code conversion required significant engineering effort while increasing the risk of inconsistencies and migration delays.

The organization required a modern solution capable of accelerating ETL modernization without compromising governance, security, or operational stability. The platform needed to automatically extract metadata, identify technical lineage, generate PySpark-based transformation logic, and create supporting engineering artifacts while maintaining traceability throughout the modernization lifecycle.

In addition to development automation, the solution needed to provide enterprise-grade operational capabilities including centralized monitoring, secure secrets management, resilient multi-AZ deployment, workload scalability, operational visibility, and governed execution of AI-assisted workflows. The platform also needed to integrate with existing enterprise systems, enforce secure access controls, and support repeatable modernization programs that could scale across multiple business domains while reducing operational overhead and improving cloud efficiency.

Partner Solution

DataEconomy designed and implemented an AI-powered ETL modernization platform on AWS that combines generative AI, cloud-native application architecture, DevOps automation, and cloud operations best practices to accelerate enterprise ETL modernization.

The solution is built on Amazon EKS using a highly available Multi-AZ architecture where specialized AI agents collaborate to analyze legacy ETL workflows, extract metadata and lineage, generate PySpark code, perform quality validation, orchestrate JIRA work items, and automate modernization activities. The platform integrates Amazon Bedrock Guardrails to help enforce responsible AI usage and applies validation throughout the modernization workflow before generated artifacts are released.

The platform leverages Amazon RDS PostgreSQL for workflow metadata, Redis with Celery for task orchestration, and a vector knowledge base to support AI-assisted processing. Integration with Azure OpenAI, Okta identity services, JIRA, and enterprise repositories enables seamless interaction with existing customer environments while maintaining governed execution and end-to-end traceability.

From a DevOps perspective, the platform incorporates automated CI/CD practices using AWS CodePipeline, AWS CodeBuild, AWS CodeDeploy, Amazon ECR, and AWS CloudFormation to standardize application delivery. Infrastructure is provisioned using Infrastructure as Code, enabling repeatable deployments, environment consistency, controlled releases, and simplified lifecycle management.

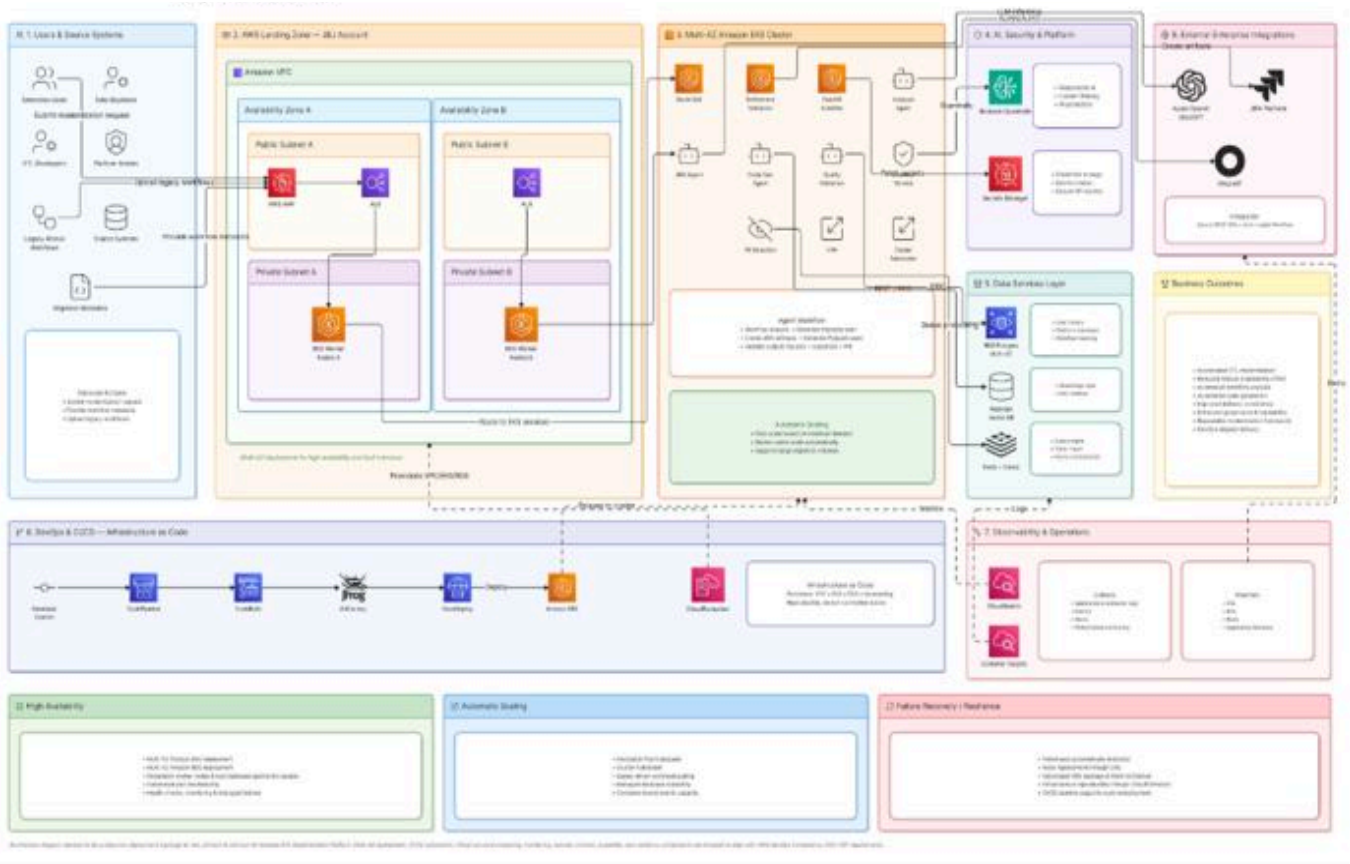
From a Cloud Operations perspective, the solution emphasizes operational excellence through centralized monitoring with Amazon CloudWatch and CloudWatch Logs, secure credential management using AWS Secrets Manager, scalable Kubernetes-based workloads, automated scaling through Kubernetes Horizontal Pod Autoscaler, Multi-AZ resilience, and continuous operational visibility across the platform. These capabilities enable secure day-2 operations while supporting governance, reliability, and cloud operational efficiency.

AWS Services Used

AWS DevOps Competency Services:

- Amazon VPC
- Amazon EKS
- Amazon EC2
- Amazon EBS
- Application Load Balancer (ALB)
- AWS WAF
- Amazon RDS for PostgreSQL (Multi-AZ)
- AWS Secrets Manager
- Amazon Bedrock Guardrails
- Amazon CloudWatch
- Amazon CloudWatch Container Insights
- AWS CodePipeline
- AWS CodeBuild
- AWS CodeDeploy
- AWS CloudFormation
- Amazon ECR

AI Assisted ETL Modernization Platform on AWS
DevOps Competency Architecture

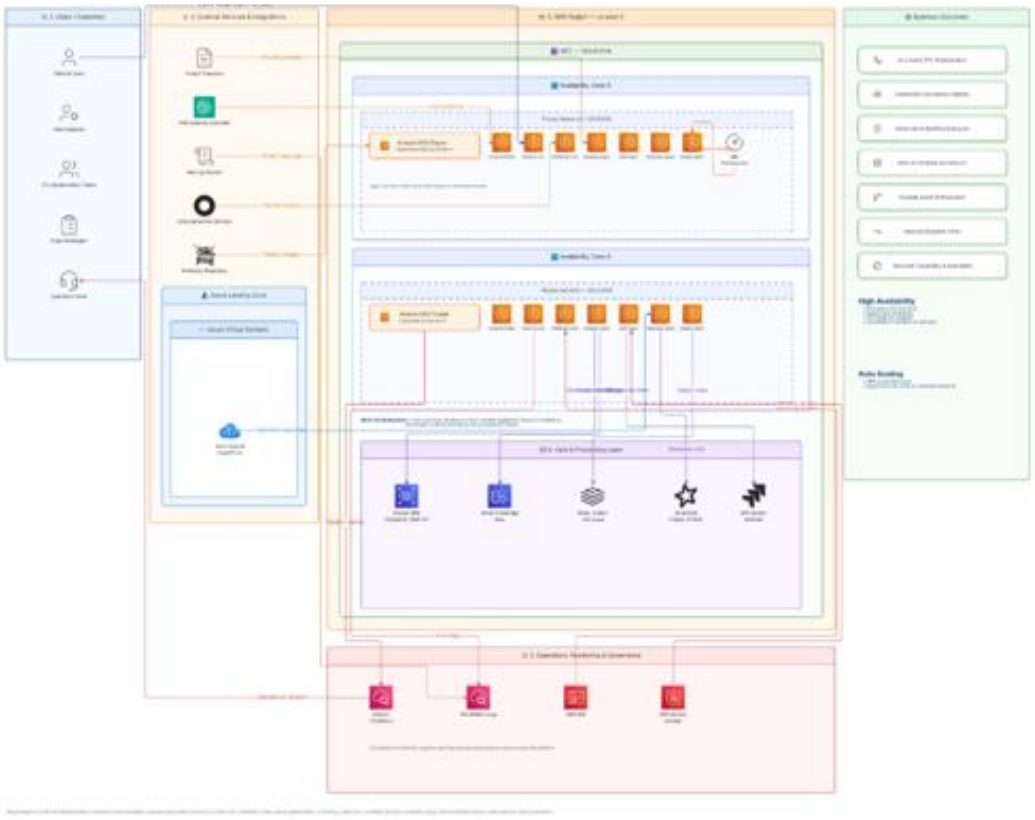


These services support cloud-native application delivery, Infrastructure as Code, automated CI/CD pipelines, secure software deployment, release automation, observability, and platform modernization.

AWS Cloud Operations Competency Services:

- Amazon VPC
- Amazon EKS
- Amazon EC2 (EKS Worker Nodes)
- Amazon RDS PostgreSQL
- Amazon CloudWatch
- Amazon CloudWatch Logs
- AWS Secrets Manager
- Data Transfer
- AWS Bedrock Guardrails
- Application Load Balancer (ALB)
- Amazon ECR

AI-Powered ETL Modernization Platform
Cloud Operations Architecture (AWS Focused)



These services provide the operational foundation for centralized monitoring, workload management, governance, secure operations, scalable infrastructure, and resilient cloud operations across the modernization platform.

Results and Benefits

The AI-assisted modernization platform established a repeatable framework for migrating legacy ETL workloads to a modern PySpark-based architecture while improving both software delivery and cloud operations.

Key business outcomes included:

- Accelerated ETL modernization through AI-assisted metadata extraction, workflow analysis, and automated code generation.
- Significant reduction in manual engineering effort required for migration analysis, documentation, and code development.
- Standardized modernization process that can be reused across future migration initiatives.
- Improved developer productivity through automated workflow orchestration, artifact generation, and CI/CD-enabled delivery.
- Enhanced operational visibility using centralized monitoring, logging, and workload observability.
- Improved governance through secure secrets management, AI guardrails, role-based access integration, and traceable modernization workflows.
- Highly available Multi-AZ deployment supporting resilient platform operations and workload continuity.
- Elastic scalability using Kubernetes orchestration and automated workload scaling to support varying modernization volumes.
- Better operational efficiency through standardized cloud operations, automated infrastructure provisioning, and simplified platform management.
- Estimated modernization cost savings of 40–60%, based on reduced manual effort, repeatable automation, and improved engineering efficiency as identified during the modernization program.

About the Partner

DataEconomy is an AWS consulting partner specializing in cloud modernization, data platforms, DevOps transformation, cloud operations, and generative AI solutions. The company helps enterprises modernize legacy platforms by combining cloud-native architectures, AI-assisted automation, and AWS best practices to accelerate business transformation while improving operational resilience.

DataEconomy's expertise spans application modernization, Infrastructure as Code, CI/CD automation, Kubernetes platforms, enterprise observability, governance, security, and operational excellence. By integrating AWS-native services with AI-powered automation, DataEconomy enables organizations to build secure, scalable, and resilient cloud platforms that improve developer productivity, simplify cloud operations, strengthen governance, and deliver measurable business outcomes.