

Accelerating Insights from EDI & Patient Data on AWS



The Customer

One of the largest US healthcare insurance providers needing to process huge volumes of EDI and patient data quickly and empower data engineers/analysts with self-serve advanced analytics.

The Challenge

- Very complex, siloed, time-consuming data processing with little automation
- Difficult to value data due to limited data-quality assessment
- Hard for analysts, scientists, engineers, and product to collaborate (no single pane of glass)
- Need to scale from local machines to cloud to accelerate solutions and productivity

The Solution

Stand up a secure, cloud-native data/ML stack on AWS with governed workflows:

- Configure secure connections to clusters/databases; connect to S3, DynamoDB, Redshift, RDS
- Orchestrate 100+ daily jobs via Airflow; run distributed jobs on EMR for scale
- Enable teams with low/no-code patterns, 450+ prebuilt functions, and 80+ ML algorithms
- Embed data-quality checks and collaborative project/workflow sharing

The Results

- Faster insights from EDI/patient data pipelines
- Governed, scalable ingestion and transformation at petabyte scale
- Rapid ML model development and deployment
- Reusable patterns accelerating future use cases

Impact Delivered

- 80% Reduction in time-to-market
- 70% Increase in user adoption
- 60% Improvement in cross-team collaboration
- 15% Overall cost reduction

Solution Components

- EMR
- S3
- DynamoDB
- Redshift
- RDS
- Airflow
- AWS Data Lake