Milan Barot

Healthcare Payer Data Engineer | Claims, Eligibility, EDI 834/837, HRP | SQL Server, SSIS, ADF, Databricks
DMV, VA | 7milan16@gmail.com | (917) 497-6979 | LinkedIn | Portfolio

Summary

Healthcare data engineer with 6 years of experience, including nearly 4 years in US healthcare payer/TPA data at Blue Cross Blue Shield of Arizona and AmeriBen. Builds and supports SQL Server, SSIS, and Azure Data Factory pipelines for claims, eligibility, accumulator, provider, value-based care, and financial data, including EDI 834/837 file delivery to CMS and vendors such as OptumRx and HealthEquity. Hands-on with HealthRules Payer (HRP) migrations, reconciling legacy and HRP-derived data, and supporting mission-critical CMS production releases, using Devin and Windsurf AI coding assistants with human review in day-to-day development. Currently expanding into Databricks, Apache Spark, and Delta Lake lakehouse architectures.

Technical Skills

Healthcare Domain: Claims, eligibility & enrollment, accumulators, provider, value-based care (VBC), EDI 834/837, CMS submissions, HealthRules Payer (HRP), HIPAA
Data Engineering: SQL, Python, PySpark, Apache Spark, SSIS, Azure Data Factory, ETL
Cloud & Lakehouse: Azure, Databricks, Delta Lake, Medallion Architecture
Databases/Data Warehouse: SQL Server, Azure SQL, PostgreSQL, DuckDB
BI & Reporting: Power BI, SSRS
Data Integration: EDI 834/837, REST APIs, FTP, ActiveBatch, SQL Server Agent
Workflow & DevOps: Airflow, Dagster, Docker, Git, Jenkins, Jira
Application & AI: Devin, Windsurf, FastAPI, React, ChromaDB, OpenAI API, Groq API

Work Experience

SQL Server Developer / Data Engineer Jan 2026 – Present
AmeriBen (UST Global — Implementation Partner) *Remote*

- Supporting AmeriBen's move of claims administration and reporting for **self-funded employer groups** from the legacy **GBAS** platform to **HealthRules Payer (HRP)**, protecting the accuracy of the financial and claims reports clients use to manage plan costs.
- Migrated **20+ client-facing reports** across two report families — large claims, financial summaries, contract savings, utilization, lag analysis, and level-of-coverage census — so employer groups kept the same view of plan spend after cutover.
- Delivered combined GBAS + HRP reporting for clients whose data was split across both platforms, including the first client live on both systems, giving them one consistent **stop-loss** and **large claims** picture during the transition.
- Resolved production incidents in migrated claims data, including incorrect lag results on **adjusted claims**, by tracing them back to the source data rather than patching individual reports, preventing repeat errors across clients.
- Identified missing HRP data fields affecting client reports and documented the gaps for business and IT teams, turning reporting requirements into validated data mappings and business rules.
- Validate AI-rebuilt reports (built with **Devin** and **Windsurf**) against existing production results before release, so faster report delivery doesn't come at the cost of accuracy.
- **Tech Stack:** SQL Server 2022, Azure SQL, SSRS, SSIS, ETL, HealthRules Payer (HRP), Devin, Windsurf, TFS.

Data Engineer Sep 2022 – Oct 2025
Blue Cross Blue Shield Arizona (Mastek — Implementation Partner) *Phoenix, AZ*

- Developed a consolidated **EDI framework for 834 and 837** enrollments, claims and dental, using three core SSIS packages to automate the generation and delivery of over **35 distinct file types** to CMS. This architecture ensured secure data exchange and audit readiness while significantly reducing the manual maintenance and operational overhead previously required.
- Provided active production support for mission-critical CMS data exchanges, conducting over **25 successful production deployments** and managing emergency releases while maintaining compliance with federal reporting deadlines.

- Designed and implemented end-to-end **ETL** pipelines using **SSIS** and **SQL Server** to process eligibility, claims, accumulator, and value-based care (VBC) files, ensuring compliance with business requirements and vendor-specific standards. Collaborated with vendors such as OptumRx, HealthEquity, Equality Health, and others to deliver high-quality data solutions.
- Utilized **Azure Data Factory (ADF)** to design and orchestrate scalable data integration pipelines for migrating legacy member and claims data to **HealthRules Payer (HRP)** on Azure Cloud.
- Rebuilt critical ETL pipelines using SSIS and SQL Server after the **CHC system hack (2024)**, restoring high-priority data workflows under tight deadlines with minimal downtime.
- Led the rapid resolution of critical production failures, ensuring minimal data loss during high-volume enrollment periods where direct connectivity with CMS was required.
- Collaborated closely with **Deloitte** on the HealthEquity project to develop eligibility and claims files and automated it end to end. Responsibilities included gathering business requirements, creating technical documentation, optimizing ETL pipelines, conducting performance testing, and ensuring smooth deployment into production environments.
- Leveraged **Azure SQL Database** for staging, transformation, and persistent storage of high-volume healthcare data. Configured ADF linked services, pipelines, and triggers for automated, scheduled data ingestion from on-premises SQL Server to Azure.
- Developed complex data extracts that combined on-prem warehouse data with select cloud-based sources, ensuring consistency, accuracy, and seamless integration across hybrid environments, reducing execution time by **25%** on average across the board.
- Automated data workflows by integrating SQL Server Agent jobs with **ActiveBatch**, setting up FTP connectivity with strict naming conventions and cadence-driven schedules for file delivery to vendors.
- Led offshore engineering resources across multiple ETL initiatives, managing task assignments, technical dependencies, and delivery timelines.
- Partnered with business stakeholders to gather requirements for new projects, conducting detailed assessments of complexity, technical effort, development timelines, and warranty periods. Worked with Mastek to draft SOWs for approval by Blue Cross Blue Shield, ensuring alignment with business objectives.
- **Tech Stack:** Azure Data Factory (ADF), Azure SQL Database, C#, SQL Server, SSIS, SSRS, ActiveBatch, SQL Server Agent, FTP, EDI (834, 837), HealthRules Payer (HRP), Agile.

Data Engineer Feb 2021 – Sep 2022

CBRE *Dallas, TX*

- Designed, developed, and maintained robust ETL pipelines using **Python, SQL, and PostgreSQL** via CBRE's internal ETL framework (Lymbic). Translated complex business logic from real estate analysts into SQL/Python transformations for property usage, tenant risk, and leasing performance.
- Ingested diverse external data sources such as Google Mobility Reports and BREAM energy efficiency assessments, transforming them into structured datasets in Parquet and CSV formats that fed Power BI dashboards used by leadership during the COVID-19 pandemic.
- Applied data profiling and transformation using **Pandas** and **PySpark** for efficient in-memory processing of large CSV/Parquet files, enabling faster QA cycles and stakeholder feedback loops.
- Supported **Databricks** orchestration workflows — monitoring job execution, rerunning failed tasks, and assisting engineers with job recovery — ensuring daily and weekly batch pipelines remained consistent and reliable.
- **Tech Stack:** Python, SQL, PostgreSQL, Power BI, Airflow, Pandas, Parquet, Agile.

ETL / SQL Server Developer Oct 2019 – Aug 2020

AmNet Services Inc. *Fairfield, NJ*

- Developed **SSIS** packages to automate extraction and ingestion of network test data from Excel, flat files, and MS Access sources into a SQL Server staging environment.
- Designed and implemented a dimensional data warehouse model using **ERwin Data Modeler** to support analytical queries on network KPIs.
- Wrote complex **T-SQL** stored procedures for data cleansing, aggregation, and validation, and built **SSRS** reports and Power BI dashboards for signal strength, call drop rates, and E911 test call success metrics.
- **Tech Stack:** SSIS/SSRS, T-SQL, Power BI, Agile.

Selected Projects

Kalshi-Prediction-Engine (PredictIQ) — Market Efficiency Research | *PySpark, Delta Lake, RAG, LLMs*

- **Question:** can any public signal — news, sentiment, event data, or LLM reasoning — predict Kalshi political-market outcomes better than the market's own price, out-of-sample?
- Built a **Medallion (Bronze/Silver/Gold) lakehouse** on PySpark and Delta Lake ingesting **16 news RSS feeds** (10 mainstream, 6 right-of-center), GDELT event streams, Congress.gov, and live Kalshi market data, with 4 Silver and 3 Gold transforms on a 5-minute cycle across **20+ Docker containers**.
- Layered a **RAG pipeline** (ChromaDB, MiniLM embeddings) and LLM analysis (Llama 3.3 70B, GPT models) that produced bull/bear cases and probability estimates, served through a FastAPI + React human-in-the-loop trade terminal with Kelly-criterion sizing.
- Tested each signal with **no-lookahead walk-forward validation and bootstrapping**. Sentiment, momentum, reversion, GDELT, and gpt-4o-mini all lost to price; a promising **+7.2% LLM edge at 108 markets shrank to +1.1% at 274 settled markets**, and a +45% ROI cohort turned out to be a longshot bias, not information.
- Showed rule-based risk controls hurt: stop-losses cut the win rate from **74% to 43%** across 171 out-of-sample trades, because mean-reverting prices triggered exits on temporary dips.
- **Conclusion:** Kalshi political markets are efficiently priced at weeks-long horizons relative to cheap public signals. Limits: one venue and one regime (2026), possible LLM training-cutoff leakage, and intra-market arbitrage not yet tested.

Education

Washington University of Science & Technology — Vienna, VA
B.S. in Information Technology — Software Engineering — *Mar 2022*

The City University of New York — New York, NY
B.A.S. in Computer Information Systems — *Dec 2019*

Certifications

Databricks Certified Data Engineer Associate — Nov 2026