

CHINMAYA SRI RAMA SESHU PASUPULETI

DATA ENGINEER

Buffalo, NY, USA | 716-709-1548 | pramaseshu12345@gmail.com | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

PROFESSIONAL SUMMARY

Data Engineer with 4+ years of experience in ETL/ELT, data integration, Cloud platforms, pipeline development, and analytics using **Python, SQL, Informatica, IICS, PySpark, Snowflake, and AWS**. Experienced in building scalable pipelines, data-quality controls, analytics-ready models, and **Tableau/Power BI dashboards**. Skilled in **Reltio MDM, dbt, Airflow, AutoSys, PostgreSQL, Oracle, and Teradata**, with expertise in SQL optimization, troubleshooting, root-cause analysis, and stakeholder collaboration.

CORE SKILLS

Programming Languages & Analysis: Python, SQL, Java, Pandas, NumPy, Excel, Statistical Analysis

BI & Visualization: Tableau, Power BI, Streamlit, Data Visualization, KPI Dashboards

Data Engineering: ETL/ELT, Informatica PowerCenter, IICS, Reltio MDM, dbt, Apache Airflow, Spark, PySpark

Databases: Oracle, Snowflake, Teradata, PostgreSQL, MySQL, MongoDB, Data Modeling, Star Schema

Cloud & Tools: AWS (S3, SQS, EC2), Azure, Docker, Git, GitHub Actions, Jupyter

PROFESSIONAL EXPERIENCE

University at Buffalo | Data Science Cohort

Jan 2026 – May 2026

Client: [QAS.AI](#) | Python, LightGBM, Fast API, Chroma DB, React, RAG, MCP, TypeScript

- Designed and built an end-to-end **CHF clinical decision-support pipeline** with 108 temporal features, multi-label **LightGBM** models, and a RAG-powered physician assistant using **FastAPI and ChromaDB**.
- Automated cohort extraction across **546K+ MIMIC-IV admissions**, covering 60+ ICD codes and 70+ lab analytes, reducing clinical data-preparation time by **~60%** and producing a reproducible 137K-visit training corpus.
- Trained **11 multi-label LightGBM classifiers** on GAN-augmented features, achieving **0.78–0.98 AUROC** for hyperkalemia, AKI, and worsening-HF prediction.
- Engineered a rule-based GDMT orchestration engine supporting **6 drug classes and 4 titration strategies**, with lab-gating, contraindication checks, and clinician-auditable recommendations.
- Deployed a **privacy-preserving clinical AI system** using Streamlit, FastAPI, ChromaDB, Mistral 7B, and MCP, keeping patient data entirely local.
- Built an **XGBoost mortality-risk module with SHAP explainability** to identify key risk drivers and inform GDMT dose-escalation decisions

Data Engineer | Cognizant Technology Solutions

Apr 2021 - Jan 2025

Client: *AbbVie* | Java, SQL, Reltio MDM, Informatica, Tableau, Power BI

Senior Data Engineer — Cognizant

Oct 2023 - Jan 2025

- Engineered and optimized a **Reltio Cloud MDM ecosystem for 9M+ HCP records**, reducing data-processing latency by **25%** and achieving **98%+ data accuracy**.
- Architected end-to-end **ETL/ELT pipelines** using IICS, Informatica PowerCenter, Java, AWS, Oracle SQL, Teradata, and Snowflake across cloud and on-premises systems.
- Reduced enterprise data inconsistency by **15%** through automated validation, historical data reprocessing, and **30+ Reltio LOV configurations** supporting data-governance standards.
- Reduced production downtime by **30%** by optimizing **AutoSys job orchestration**, dependency management, proactive failure detection, and SQL-based root-cause analysis.
- Accelerated sprint delivery by **~20%** by translating business requirements into Reltio configuration specifications, validation rules, reusable data objects, and technical documentation.
- Designed **Tableau and Power BI analytics solutions** for HCP engagement, data quality, and pipeline KPIs, enabling self-service reporting and data-driven decisions across **3+ stakeholder teams**.

Junior Data Engineer — Cognizant

Apr 2021 - Oct 2023

- Supported **Reltio Cloud MDM** pipelines managing **9M+ AbbVie HCP records**, using Informatica PowerCenter, Java, Oracle SQL, AWS, Teradata, and Unix
- Developed and maintained **ETL/ELT ingestion workflows** across Reltio, AWS S3/SQS, Oracle, and Teradata for large-scale healthcare data integration.
- Implemented **30+ Reltio LOV configurations** and historical data reprocessing workflows to improve master-data consistency and governance.
- Performed **data profiling, validation, and quality checks** using SQL and Python-based scripts, identifying and resolving data inconsistencies across HCP datasets.
- Monitored production pipelines through **AutoSys, Unix, SQL, and Postman**, troubleshooting failures and performing root-cause analysis for critical workflows.
- Built **Tableau and Power BI dashboards** tracking HCP data quality, pipeline performance, and operational KPIs for AbbVie business stakeholders.

PROJECTS

360° Enterprise Performance Analytics Dashboard | Tableau - [GitHub](#)

- Built an interactive **Tableau 360° enterprise performance dashboard** analyzing sales, profit, profit margin, customer segments, product categories, and regional trends across the Superstore dataset
- Developed **KPI-driven executive dashboards** and **interactive filters** using Tableau calculated fields to evaluate category, sub-category, regional, and customer-segment performance.
- Analyzed **discount** impact on **profitability** and **loss-making product segments**, identifying regional and category-level opportunities to improve revenue and profit performance.

Healthcare Admissions Analytics System | End-to-End Data Engineering Project | [GitHub](#)

- **Built** an end-to-end healthcare **data pipeline (OLTP -> OLAP)** using **PostgreSQL, Python, dbt, Airflow and Docker** to ingest and transform **100k+ patient appointment records**, enabling no-show behavior analysis across demographic and medical conditions.
- **Designed** a normalized **3NF transactional schema** and analytics-ready **star schema**, enabling advanced SQL analysis with CTEs, window functions, and performance-optimized indexed queries.

Bookshop 360° Business Intelligence Dashboard | Tableau - [GitHub](#)

- Analyzed bookshop sales, revenue, profitability, ratings, licensing, and production data to build a **multi-stakeholder Tableau BI dashboard** supporting data-driven business decisions.
- Created **calculated KPIs, interactive dashboards, and parameter-based What-If analysis** to evaluate discounting, profit margins, ROI, price elasticity, format performance, and seasonal sales trends.
- Translated analytical findings into business recommendations for pricing, marketing ROI, inventory planning, international expansion, and portfolio investment.

E-Commerce Retail Analytics Database | Python, PostgreSQL, SQL, Docker| [GitHub](#)

- **Built** an end-to-end **ETL pipeline** using Python, Pandas, and **SQL Alchemy** to ingest 9 Olist e-commerce datasets into PostgreSQL, processing **~100K orders** and **~3M geolocation records**.
- **Designed** a 3NF relational database schema and developed advanced SQL analytics using **CTEs**, window functions, ranking, and customer RFM segmentation.
- **Optimized** PostgreSQL queries using partial indexing and EXPLAIN ANALYZE, achieving **~60% faster** execution, and containerized the data pipeline and database with **Docker Compose**.

Corporate Financial Performance & Market Analysis | Tableau - [GitHub](#)

- Analyzed **income statement, balance sheet, cash flow, and stock-market** datasets using **Tableau** to assess corporate financial health and market performance.
- Created **calculated financial ratios** and interactive **KPI dashboards** covering **profitability**, liquidity, leverage, efficiency, cash flow, and valuation.
- Applied **time-series analysis** and **data storytelling** to identify relationships between financial fundamentals, risk exposure, valuation changes, and stock-price trends.

Heart Disease Risk Prediction | Machine Learning - [GitHub](#)

- Built an end-to-end **heart disease risk prediction pipeline** using clinical and lifestyle features, applying data cleaning, feature selection, normalization, and exploratory data analysis.
- Trained and compared **5 machine learning classifiers**—Logistic Regression, KNN, Naive Bayes, Random Forest, and XGBoost—using **GridSearchCV** and multiple evaluation metrics.
- Achieved approximately **73% accuracy** and performed feature-importance analysis across **10+ clinical attributes**, identifying **age, systolic blood pressure, and cholesterol** as key cardiovascular risk predictors.

CERTIFICATIONS

Microsoft Certified: Power BI Data Analyst Associate - (July 2026)

Microsoft Certified: Azure Databricks Data Engineer Associate - (Aug 2026)

Informatica MDM Certification - (Aug 2023)

Reltio MDM Configuration Specialist - (Dec 2021)

EDUCATION

M.S. in Engineering Science (Data Science) | GPA – 3.6 | University at Buffalo, Buffalo, NY, USA

Jan 2025 - May 2026

B.Tech in Electronics & Communication Engineering | Amrita University, Bangalore, India

Aug 2017 - May 2021