

Data Engineer with 7+ years of experience designing and maintaining scalable, high-performance data pipelines and ETL workflows using Python, Airflow, and AWS. Proven success in modernizing data infrastructure, building distributed systems, and improving data quality across microservice-based architectures. Experienced with Kafka, Redshift, PostgreSQL, and Snowflake. Strong collaborator with both engineering and business teams, driving data reliability, availability, and actionable insights.

EXPERIENCE

Clover Networks, San Jose, CA - Data Engineer

Feb 2024 - Present

Data engineer responsible for migrating existing legacy data pipelines to new frameworks.

- Built scalable data pipelines to ingest and transform high-volume merchant data into a centralized data warehouse, enabling faster analytics.
- Developed a **Java Spring Boot** application for automating lifecycle management of data pipelines reducing the average time spent on pipeline maintenance from **20 minutes to 8 minutes**.
- Improved stability of **Kafka pipelines** by addressing critical technical debt in sink connectors failure handling, resulting in a **60% reduction in pipeline restarts** and smoother production operations.
- Designed and migrated legacy ETL workflows to GCP-native frameworks using Airflow, and Snowflake, improving data availability and **reducing downtime by 60%**.

FlipT, Jersey City, NJ - Data Engineer

Apr 2022 - Dec 2023

Data engineer responsible for migration of claim adjudication data in pharmacy benefit management systems.

- Developed **REST** endpoints in **Flask** that handles ~2K requests for automated treatment authorization using Python and RabbitMQ **reducing manual intervention from 2 hours to 10 minutes**.
- Implemented asynchronous modules in Python for optimizing database calls and http requests which **reduced service response time by 27% and improved CPU usage by 14%**.
- Engineered fault-tolerant **data pipelines** using **Python, NoSQL** and **Airflow** to process over 50K daily PHI records from SFTP to AWS S3, enhancing compliance and data integrity.
- Orchestrated complex **ETL workflows** in the pharmaceutical insurance domain, proactively adopting industry best practices for scalable pipeline scheduling and staying current with emerging trends.

Alpine Electronics, SJ, CA - ML Engineer

Feb 2022 - Apr 2022

Member of machine learning team responsible for improving the model's ability to generalize unseen gaze patterns and environmental conditions.

- Enhanced feature extraction using multiple image processing techniques in OpenCV like histogram equalization and image segmentation, **boosting the model accuracy by 9%**.

Iqvia, Bangalore, India - Sr. Software Engineer

Feb 2019 - Aug 2019

Contributor to the system migration team involved in transitioning legacy Mainframe code to Python scripts.

- Transformed medical sales data in MySQL DB to delimited files by incorporating SQL queries into Python scripts, resulting in a **40% reduction in client specific report generation time**.
- Migrated legacy JCL batch procedures for pharmaceutical sales reporting to **Python** scripts with MySQL on a client-server architecture, resulting in a **25% improvement in data loading speed** and enhanced maintainability.
- Collaborated with DevOps teams to implement **CI/CD pipelines using Jenkins**, automating testing, deployment, and monitoring processes for a SaaS application, which streamlined release cycles and improved deployment reliability.

Built Mainframe/Python applications for client engagement teams within the life insurance industry.

- Revamped health check activities including server readiness assessment, job failure alert reports and ticket assignments by creating Python modules which saved **1 hour** per day for the team.
- Achieved **40% increase in data retrieval** speed by optimizing DB2 queries using JCL, COBOL, DB2 which led to reduction in daily and monthly **batch processing duration by 40% and 30%** respectively.
- Collaborated closely with product managers and cross-functional teams in the health insurance domain, effectively managing project timelines, driving stakeholder alignment, and ensuring successful feature delivery.

EDUCATION

MS in Data Science	University of the Cumberlands, Kentucky	Jan 2025 - Expected May 2026
MS in Computer Science	San Francisco State University, California	Dec 2021, GPA 3.73
B.Tech in Computer Science	Amrita University, India	May 2015

SKILLS

Languages	Python, Java, SQL, C++
Data Engineering	Airflow, Kafka, Flink, Spark, Hive, Presto, Redshift, Snowflake, PostgreSQL, MySQL
Cloud Platforms	AWS (S3, Lambda, Redshift, EC2), GCP (BigQuery, Dataflow)
ETL Tools	Apache Airflow, Flink, Oozie, Azkaban
CI/CD	Jenkins, Docker, Kubernetes, Git
Reporting	Looker, Tableau, Superset, Prometheus
Methodologies	Agile, Scrum, TDD
Soft Skills	Cross-functional collaboration, Stakeholder communication, Strategic thinking