

REZA RAJAN

Data Engineer - AWS Specialization

1-407-649-0447
rezarajan@gmail.com
[/in/rezarajan](https://www.linkedin.com/in/rezarajan)
[rezarajan](https://www.github.com/rezarajan)

Experience

myKaarma - Data Engineer (Cloud)

Jul 2023 - Jan 2024

- Designed and contributed to a cloud data lakehouse architecture on AWS to centralize and scale analytics and reporting.
- Built and maintained ETL/ELT pipelines using Python, Apache Airflow, AWS S3, AWS Glue and Redshift to ingest, transform and store cleaned event and transactional data.
- Automated scheduled customer report generation and delivery using AWS Step Functions and Lambda with SLA monitoring and alerting.
- Delivered a PoC batch processing pipeline using Airflow + PySpark that reduced message data processing time by ~70% compared to a stored-procedure based approach; validated performance and reliability metrics before handoff.

myKaarma - Data Engineer (Analytics)

Feb 2022 - Jul 2023

- Collaborated with external stakeholders to capture reporting requirements and translate them into actionable technical specifications for the data team.
- Developed stored procedures for the reporting warehouse, optimizing several with indexing strategies to achieve up to 50% reduction in query execution time.
- Integrated multiple operational and external data sources into Power BI, enabling standardized dashboards that reduced manual reporting workload for the analytics team.
- Partnered with engineers to improve data quality checks and streamline ETL processes feeding the reporting warehouse.

Project Experience

Data Pipeline in AWS using Apache Airflow

Built a modular, production-grade ETL pipeline in Apache Airflow on AWS that supports reusable tasks, backfills, monitoring, and automated data quality checks.

Data Warehouse using AWS Redshift

Developed an ETL process to extract data from S3, stage in Redshift, and transform into dimensional star-schema tables for downstream analytics.

Data Modeling with Apache Cassandra

Designed and implemented an Apache Cassandra data model optimized for query performance and analytics use cases.

Certifications

AWS Certified Data Engineer

Amazon - In-Progress

Data Engineer

Udacity - 2024

Self-Driving Car Engineer

Udacity - 2021

Data Analyst

Udacity - 2020

Education

Accreditation in CS (DSA)

UC Berkeley

2025 - 2025

Bsc. Mechanical Engineering

University of Waterloo

2015 - 2020

Technical Skills

Core Competencies: etl/elt, real-time analytics, data lakehouse, data modeling, apache airflow, kafka, flink, aws s3, glue, redshift

Programming Languages: python, golang, sql, bash, nix

Data: postgresql, mariadb, sqlite, redshift, influxdb, cassandra, mongodb, s3, flink, lambda, dbt, airflow, kafka, spark, elasticsearch, etl/elt

Visualization: grafana, power bi, looker studio, plotly

Data Libraries: pandas, scipy, numpy, tensorflow, pyspark

Infrastructure: linux, docker, kubernetes, aws, proxmox, ansible, terraform, opentofu, dbt

GitOps: git, github actions, ci/cd