

KALAIYARASU D

Data Engineer | +91 9360522033 | kalaiyarasu210@gmail.com | linkedin.com/in/kalaiyarasu-d

SUMMARY

AWS-focused Data Engineer transitioning from a compliance background, with hands-on experience building scalable ETL pipelines using AWS Glue, Lambda, Step Functions, and Redshift. Skilled in Python scripting for automation and orchestration, with a strong foundation in data validation, access control, and governance from SOX compliance work. Designed and deployed multiple cloud-native data pipelines for e-commerce and analytics use cases. Proficient in Apache Spark and PySpark for distributed data processing.

TECHNICAL SKILLS

- **Programming Languages:** Python, SQL
- **Databases:** PostgreSQL, Amazon Redshift
- **Cloud Platforms:** AWS (Glue, Lambda, Step Functions, S3, Redshift, SNS)
- **Big Data Processing:** Apache Spark, PySpark, Databricks
- **Orchestration Tools:** AWS Step Functions
- **Monitoring Alerts:** SNS
- **others:** SOX compliance, GitHub

PROJECTS

AWS Movie Data Engineering Pipeline

Data Engineer

Jun 2025

Tech Stack: Python, PySpark, Apache Spark, AWS S3, Lambda, Glue, Step Functions, Redshift, SNS

- Designed and deployed a serverless data pipeline to process raw movie data stored in S3.
- Triggered Lambda on S3 events to invoke Step Functions for orchestrating Glue jobs.
- Applied data transformation and validation logic in AWS Glue using Pyspark.
- Stored cleaned data in Parquet format and ingested it into Redshift Spectrum tables.
- Integrated SNS for real-time success/failure alerts and monitoring.
- Project link: GitHub

E-Commerce Data Pipeline using AWS

Data Engineer

Jun 2025

Tech Stack: Python, PySpark, Apache Spark, AWS S3, Lambda, Glue, Step Functions, Glue Crawlers, Redshift Spectrum, Parquet, SNS

- Designed an automated data pipeline to process e-commerce datasets including orders, customers, and product information.
- Ingested raw data into an S3 bucket; used Lambda to trigger Step Functions on file arrival.
- Orchestrated workflow with Step Functions: initiated a Glue job for transformation and stored cleaned data in Parquet format.
- Used a separate Glue Crawler to catalog the processed data in S3.
- Created a Redshift Spectrum external table referencing the Parquet files for analytics and reporting.
- Enabled monitoring through SNS alerts for success/failure notifications at each step.
- Project link: GitHub

PROFESSIONAL EXPERIENCE

Genpact - Technical Associate

SOX Consultant

Nov 2023 - Present

- Led IT and SOX audits by validating critical IT controls (ITAC/SOD), ensuring data integrity and compliance across financial systems.
- Managed user access reviews and Oracle provisioning, reinforcing principles of least privilege and secure data access.
- Collaborated with process owners, internal audit teams, and external auditors to ensure timely remediation of identified deficiencies.
- Tracked control issues, remediation progress, and timelines to support SOX program compliance.
- Documented and communicated compliance risks to senior leadership, ensuring alignment with organizational policies.
- Actively transitioning into data engineering: completed AWS-focused projects involving Glue, Lambda, Step Functions, and Redshift; built serverless pipelines for structured data ingestion and transformation.

EDUCATION

- **Bachelor of Engineering:** Computer Science and Engineering KPR Institute of Engineering & Technology [2019–2023]

RELEVANT COURSEWORK

- Programming for Everybody (Getting Started with Python) – Coursera
- Data Engineering with AWS Bootcamp – Grow Data Skill
- PostgreSQL Bootcamp –Udemy
- Data Warehouse – The Ultimate Guide – Udemy