Kamalakannan P

Data Engineer

+91 8220933854
kamalakannan200336@gmail.com
Bengaluru, Karnataka, India
kamalakannan200336
Kamalakannan1
Kamalakannan

Professional Summary

Data Engineer with 2 years of experience building data pipelines, supporting enterprise data migration, and developing scalable data processing workflows using Python, SQL, PySpark, Spark SQL, and Databricks. Skilled in ETL/ELT development, lakehouse architecture, data ingestion, transformation, validation, workflow orchestration, and troubleshooting. Focused on data quality, reliable pipeline execution, and delivering analytics-ready datasets for downstream consumption.

Technical Skills

Programming: Python, SQL, PySpark, Spark SQL
Data Engineering: ETL/ELT Pipelines, Data Ingestion, Data Transformation, Data Validation, Data Quality
Big Data & Hadoop: Apache Spark, Hadoop, HDFS, Hive, YARN, MapReduce
Lakehouse & Platforms: Databricks, Delta Lake, Medallion Architecture (Bronze/Silver/Gold), Greenplum
Databases & Cloud: PostgreSQL, MySQL, AWS, Microsoft Fabric, Unity Catalog
Orchestration & Development Tools: Apache Airflow, Talend, Git, Pipeline Monitoring, Debugging
Analytics & Reporting: Pandas, NumPy, Power BI, Advanced Excel, KPI Reporting

Professional Experience

Genpact, Bengaluru *Oct 2024 – Present*
Data Engineer (Data & Analytics)

- Developed batch ETL/ELT pipelines using Python, PySpark, Spark SQL, and Databricks to ingest, transform, validate, and deliver datasets for downstream analytics and reporting.
- Contributed to a greenfield enterprise migration involving 20,000+ tables, 45,000+ database objects, 15,000+ jobs, and 3,000+ reports, supporting a transition designed to minimize user impact.
- Supported Greenplum-to-Databricks migration and Talend EDL job modernization, helping maintain data consistency across data-processing workflows.
- Implemented Medallion Architecture workflows across Bronze, Silver, and Gold layers, applying schema checks and data-quality validations to improve data reliability.
- Orchestrated workflows using Apache Airflow, monitored pipeline execution, investigated failures, and used retries and alerts to support recovery and reliable data delivery.
- Built foundational knowledge of Hadoop ecosystem concepts, including HDFS, Hive, YARN, and MapReduce, to understand distributed storage and large-scale data processing architectures.
- Collaborated with cross-functional teams to clarify requirements, troubleshoot technical issues, and apply reusable engineering practices and code-review feedback.

Genpact, Bengaluru *Feb 2024 – Jul 2024*
Data / AI Intern

- Delivered 5+ analytics and visualization projects using Python, Power BI, and Advanced Excel, developing dashboards covering 10+ business KPIs.
- Organized and validated data and reporting outputs to support reliable analysis and improve the usability of business information.
- Performed data cleaning, transformation, and validation using Python and Excel to improve data consistency and prepare reliable datasets for reporting and business analysis.
- Collaborated with team members to understand reporting requirements, validate KPI calculations, and improve the accuracy and consistency of dashboard outputs.

Project

Enterprise Data Pipeline and Lakehouse Processing

- Developed data-processing workflows using Python, SQL, and PySpark to support data ingestion, transformation, validation, and preparation for downstream analytics.
- Applied a Medallion Architecture approach to organize data into Bronze, Silver, and Gold layers, with validation checks to improve consistency and data quality.
- Used workflow orchestration and monitoring concepts to track pipeline execution, troubleshoot processing failures, and support reliable data delivery.

Education

B.Tech in Computer Science & Engineering *2020 – 2024*
Kalasalingam Academy of Research and Education **CGPA: 9.2/10**

Certifications

- Databricks Data Engineering Professional (2025)
- AWS Certified Data Engineer – Associate (2025)
- Microsoft Fabric Data Engineer Associate (2025)