

Kathirvel Rajmohan

✉ kathirvelrajmohan9498@gmail.com | 📞 +91 9790472974

🌐 github.com/kathirvelrajmohan | 🌐 kathirvel.in | 🔗 linkedin.com/in/kathirvelrajmohan

Professional Summary

Data Engineer with around 2 years of experience writing complex SQL and building PySpark pipelines in Databricks — including cluster configuration, job orchestration, and AWS S3 integration — alongside enterprise application development. Databricks Certified Data Engineer Associate, with hands-on experience extending pipelines beyond Databricks using Docker, Airflow, and CI/CD through an independent end-to-end ETL project.

Skills

Languages: Python, SQL, Java, JavaScript, C/C++

Data Engineering: Apache Spark (PySpark), Databricks (Jobs, Workflows), Delta Lake, ETL/ELT Pipeline Design, Data Quality & Validation, Star Schema / Dimensional Modeling, Medallion Architecture (Bronze/Silver/Gold), Apache Airflow

Cloud & DevOps: AWS (S3, IAM), Docker, CI/CD (GitHub Actions)

Databases: PostgreSQL, MySQL, MongoDB, Spark SQL

Tools & Other: Git, REST APIs, Data Structures & Algorithms, Agile/Scrum

Certifications

Databricks Certified Data Engineer Associate 2026

Validates expertise in lakehouse architecture, Delta Lake, ETL pipelines, Apache Spark, and data engineering best practices.

Work Experience

Optimum Solutions, Chennai

Oct 2024 – Present

Associate Consultant

- Wrote and optimized **complex SQL queries** involving multi-table joins, CTEs, aggregations, and conditional logic to support application data operations, validation, and reporting.
- Built and tested **Python/PySpark data pipeline notebooks in Databricks**, configuring compute **clusters** and scheduling jobs via **Databricks Workflows**, and reading/writing data to/from **AWS S3** for downstream reporting and analytics.
- Identified a performance bottleneck in a SQL-based transformation processing 1M+ rows (50-minute runtime); redesigned it in PySpark, cutting runtime to 13 minutes (**~74% faster**) and improving pipeline turnaround.
- Developed and delivered new features, batch processes, and bug fixes across multiple project repositories using **Git** for version control and branch-based development workflows.
- Performed **unit testing (UT)** and **functional testing (FT)** to validate feature correctness and ensure quality before production deployment.
- Collaborated with cross-functional teammates and stakeholders in an **Agile/Scrum** environment, providing daily stand-up updates and aligning on sprint goals.

Education

IFET College of Engineering, Villupuram, Tamilnadu

Sep 2021 – Apr 2024

B.E. in Computer Science and Engineering

CGPA: 8.4/10

Relevant Coursework: Data Structures and Algorithms, Database Management Systems, Data Warehousing and Data Mining, Machine Learning, Object Oriented Programming, Operating Systems, Computer Networks

Seshasayee Institute of Technology

July 2018 – Jun 2021

Diploma in Electrical and Electronics Engineering

CGPA: 8.2/10

Relevant Coursework: C++ Programming, Programmable Logic Controller, Logic Gates

Project Work

- Sales ETL Pipeline — Databricks, PySpark, Delta Lake, AWS, Docker, Airflow:** Built an end-to-end data engineering pipeline on Databricks using Medallion Architecture (Bronze/Silver/Gold) on a retail sales dataset, with Delta Lake ACID tables, PySpark transformations and window functions, a 53-check automated data quality framework, and a star schema scheduled via Databricks Jobs. Extended the Gold layer with a secure AWS S3 export (IAM least-privilege, Databricks Secrets), and built a standalone, Dockerized version of the Silver transform validated against the Databricks pipeline, with GitHub Actions CI and an Apache Airflow DAG for orchestration decoupled from a managed platform.
Tools and Technology Used: Python, PySpark, Databricks, Delta Lake, Spark SQL, Star Schema Modeling, AWS S3, Docker, GitHub Actions, Apache Airflow.