

RAJESH GIRI

DATA ENGINEER

Pune, Maharashtra | +91-8605733751 | rajeshgiri.work@gmail.com | www.linkedin.com/in/rajesh-giri

PROFESSIONAL SUMMARY

Data Engineer with 3.6 years of total IT experience, including nearly 3 years in data engineering. Experienced in developing and maintaining ETL/ELT pipelines using Python, PySpark, Apache Spark, SQL, and Azure Databricks. Hands-on experience with Medallion Architecture (Bronze, Silver, Gold), Delta Lake, AWS S3, ADLS Gen2, data ingestion and transformation, data quality and schema validation, source-to-target reconciliation, Spark optimization, schema evolution, workflow orchestration, and CI/CD using Azure DevOps and Git.

TECHNICAL SKILLS

- **Programming & Data Analysis:** Python (OOP), SQL, Pandas, MySQL, Core Java
- **Big Data & Lakehouse:** Apache Spark, PySpark, Spark SQL, Delta Lake, Medallion Architecture (Bronze/Silver/Gold)
- **Cloud & Data Platforms:** Azure Databricks, Azure Data Factory (ADF), ADLS Gen2, AWS S3, AWS IAM
- **Data Engineering:** ETL/ELT, Data Pipelines, Data Quality, Data Validation, Schema Validation, Data Warehousing, Data Modeling, Schema Evolution, Source-to-Target Reconciliation
- **Tools & Delivery:** Databricks Workflows, Azure DevOps (CI/CD), Git, GitHub, PyCharm, Agile/Scrum

PROFESSIONAL EXPERIENCE

2GBR Software Pvt Ltd

Pune India

Data Engineer (Project: Best Buy – E-commerce & Retail)

Dec 2025 – Present

Technologies: Python, PySpark, Databricks, Delta Lake, AWS S3, MYSQL, ADLS Gen2, Databricks Workflows.

- Designed and developed end-to-end ETL pipelines using PySpark and Azure Databricks for T-1 data processing, handling approximately 1 million raw records (~1 GB) daily and approximately 30 million records (~30 GB) monthly.
- Built and maintained a Medallion Architecture (Bronze, Silver, Gold) using Delta Lake on AWS S3, ingesting approximately 1 GB of raw data daily into curated, analytics-ready layers.
- Developed Databricks-based PySpark transformations for sales analytical datasets stored in AWS S3, including cleansing, standardization, joins, and aggregations.
- Performed data quality, data validation, and schema validation checks to identify missing values, duplicates, inconsistencies, and structural issues.
- Developed business-ready datasets across Sales, Product, and Geographical domains for downstream analytics and decision-making.
- Optimized Spark-based ETL processing using partitioning and Spark configuration techniques.
- Automated workflow orchestration and monitored ETL pipeline executions using Databricks Workflows, tracked job statuses, investigated failed runs, and managed schema evolution and job failure alerts to support reliable data processing.
- Developed reusable Python modules and implemented CI/CD pipelines using Azure DevOps and Git within an Agile/Scrum team

Synergie Technik Pvt Ltd**Pune, India****Associate Data Engineer (Project: Kohl's – E-commerce)**

Oct 2023 – Nov 2025

Technologies: Python, PySpark, Azure Databricks, Azure Data Factory, SQL, Pandas, Delta Lake, Git.

- Validated source datasets to identify nulls, duplicates, missing records, and data inconsistencies.
- Cleansed and preprocessed data using PySpark and SQL to improve data quality.
- Performed source-to-target validation and reconciliation to verify ETL outputs against source data.
- Investigated data-quality issues and pipeline failures, performed root-cause analysis, and coordinated with the team on resolution.
- Created and configured Azure Data Factory pipelines for batch orchestration and integration with Azure Databricks.

Software Engineer (Project: Manufacturing ERP System)

Apr 2023 – Oct 2023

Technologies: HTML, CSS, JavaScript, Bootstrap.

- Developed responsive front-end pages for a manufacturing ERP application covering Inventory, Quality, Machining Operations, Production Planning, Final Quality Check, and Invoicing.
- Built interfaces for inventory and material inward entry, machining operations tracking, final quality inspection, and report viewing.
- Developed production planning, quality tracking, invoicing, and dispatch handoff screens.
- Collaborated with the backend team to integrate frontend pages with backend services and resolve UI bugs during testing.

EDUCATION

- | | |
|--|----------------------------|
| • PG-Diploma in Advance Computing (PG-DAC)
Central for Development of Advance Computing (CDAC) | Trivandrum, Kerala
2023 |
| • Bachelor of Engineering (B.E.)
Sinhgad Institute of Technology (SPPU) | Pune, Maharashtra
2021 |