DEBRAJ MONDAL

DATA ENGINEER (PYSPARK)

+91 9874220939 | debraj123mondal@gmail.com | Kolkata, WB, INDIA | LinkedIn | GitHub

---

CAREER SUMMARY

Skilled PySpark Developer with 3+ years of experience in Python, PySpark, SQL, Shell Scripting, and other Big Data tools. Developed high-performance ETL data pipelines, ensuring data quality at scale. Recognized for strong leadership in guiding teams and taking ownership of critical deliverables during challenging periods, ensuring stability and scalability. Highly valued as a collaborative team player for supporting and mentoring colleagues on processes and implementing innovative ideas that enhance overall team performance. Certified across multiple cloud platforms.

TECHNICAL SKILLS

PySpark | Python | SQL | Hive | Bash | CDP | Gitlab | Zena | Django | Informatica PowerCenter | TensorFlow

WORK EXPERIENCE & PROJECTS

PySpark Developer | Associate | Cognizant (Client: Country Financial)   2023- Present

- Contributed as a key developer for a US-based insurance client, collaborated across multiple agile teams on diverse client projects, supported team through challenging phases, and helped improve efficiency and morale.
- Migrated jobs from Spark 2 to 3 and legacy tables from Hive to Iceberg for schema-resilient parsing, optimized wide tables processing through checkpointing and lineage breaking to achieve 30% faster job runtimes.
- Developed CDC ingestion and curation pipelines (historical & active) for business-critical datasets (Ratabase, marketing) from mainframe, structured and semi-structured sources, enabling downstream AI/ML workloads.
- Led a team in developing an automated PySpark testing framework with PyTest, integrated into the DevOps CI/CD pipeline, cutting data-validation effort by 80%.
- Conceived and led development of award-winning innovative tools—automated schema validator, source data profiler, HQL analyzer, and fuzzy based PII detector—during IP sprints, cutting analysis/validation cycles from 5 days to 1 and enhancing data quality across ingestion pipelines.
- Upgraded legacy HQL batch jobs to PySpark for the claims data model (GWCC and Siebel), reloaded historical data, and implemented incremental loading to support ongoing curation, achieving 25% faster job runtimes.

Data Integration Full Stack Intern | Cognizant   2022

- Completed comprehensive training in SQL, Java, Python, UNIX, Big Data tools, DW, and AWS; developed hands-on expertise in ETL tools (Informatica PowerCenter, IICS, Talend, DataStage, Ab Initio).
- Implemented data integration using Informatica PowerCenter to ingest data into a star-schema data warehouse for a university analytics system, managing SCDs (Type I and II).

EDUCATION

B.Tech in Computer Science & Engineering | FIT, MAKAUT, Kolkata | CGPA: 9.2/10   2022

CERTIFICATIONS   2023 – 2025

- **AWS**: Machine Learning – Specialty, Solutions Architect – Associate, Cloud Practitioner
- **Microsoft**: Azure Data Fundamentals, Azure Fundamentals
- **Databricks**: Data Engineer Associate
- **Snowflake**: SnowPro Core
- **RELTIO**: Technical Consultant Intermediate
- **Informatica**: MDM Implementation Practitioner