Abhay Singh *Data Engineer*

✉ writetoabhaysingh21@gmail.com 📞 +91 9910356085 in LinkedIn ⌨ GitHub 🔗 Portfolio

💼 WORK EXPERIENCE

Data Engineer, *Xebia IT Architects* 02/2023 – Present │ Gurugram, India
Key Development Projects:

1. ADF Pipeline Update Automation and Platform Tooling (Python, Azure DevOps, ADF)
- Owned and redesigned manual Azure Data Factory release workflows that caused slow deployments and frequent configuration errors.
- Designed and built a Python-based automation tool to programmatically update ADF pipelines, eliminating manual JAR deployments and reducing update time by 90 percent.
- Integrated Azure DevOps REST APIs to manage Git branches, update pipeline JSON definitions, and automatically raise pull requests, enabling CI/CD-driven releases.
- Implemented validation and safety checks to prevent redundant updates and configuration drift, reducing release defects by 25 percent and cutting overall release cycle time by 30 percent.

2. Sales Promotion Ingestion Platform Development (Scala, Spark, Delta Lake)
- Led the redesign of a legacy JDBC-based ingestion process that struggled with scale, correctness, and data reconciliation across systems.
- Designed and implemented a distributed Scala ingestion framework using Apache Spark and an SFTP-driven architecture to standardize ingestion and improve fault tolerance.
- Built Spark and Delta Lake reconciliation workflows to enforce schema consistency, primary key alignment, and business rule parity across platforms.
- Developed scalable backfill and comparison jobs to validate over 50 million historical records, ensuring zero duplication and full referential integrity.

3. Legacy OPD Data Platform Modernization (PySpark, Databricks)
- Took ownership of legacy OPD-based reporting datasets affected by inconsistent schemas, redundant pipelines, and tightly coupled transformations.
- Designed standardized PySpark ingestion and transformation modules supporting incremental processing and controlled schema evolution.
- Refactored joins and business logic to eliminate redundancy and improve data mart reliability and query performance for analytics workloads.

Additional Engineering Impact
- Refactored and optimized Spark and Databricks ETL pipelines, improving performance by 25 percent while reducing compute costs.
- Built internal automation utilities that reduced manual operational effort by 16+ engineer hours per week.
- Optimized Azure cloud infrastructure and Spark cluster utilization, contributing to approximately $848K in cost savings.

SKILLS: SQL, **Python**, **Scala, PySpark**, **Apache Spark**(Core, SQL, Streaming), **Apache Kafka, Databricks**, **Apache Airflow, MS Azure**(ADF, Databricks, Azure SQL, Azure DevOps)

🎓 EDUCATION

B.Tech. CSE specialization in AIML, *Sushant University* 07/2019 – 06/2023 │ Gurgaon, India

📑 CERTIFICATIONS

Databricks Certified Data Engineer Associate ⧉ **Databricks Lakehouse Fundamentals** ⧉
11/2023 - 11/2025 01/2025 - 01/2026

🏅 AWARDS AND RECOGNITIONS

Galaxy Best Project Team (Tauras) Award, *Xebia IT Architects*

GEM Award (3x), *Xebia IT Architects*