

# Abhijaya Singh

---

Data Engineer | Azure | Databricks | PySpark | 4 Years Experience

Pune, India | +91-9604135196 | abhijayasingh6@gmail.com | linkedin.com/in/abhijaya6

## Professional Summary

Cloud Data Engineer with 4 years of experience designing and maintaining scalable ETL and real-time streaming pipelines on Azure and Databricks. Proficient in PySpark, SQL, and Azure Data Factory (ADF), delivering 25% faster data processing. Skilled in building near real-time analytics with Spark Structured Streaming, Kafka, and Delta format. Proven ability to deliver production-grade pipelines aligned with business goals.

## Technical Skills

- Languages: Python, SQL
- Cloud Platforms: Azure (ADF, ADLS, Blob Storage, Azure SQL)
- Big Data Tools: Databricks, PySpark, Apache Spark, Delta lake, Delta Live Tables
- Streaming & Messaging: Apache Kafka, Spark Structured Streaming
- Orchestration: Apache Airflow
- DevOps & Version Control: Git, Azure DevOps
- Concepts: ETL Pipelines, Data Quality, Data Warehousing, Batch & Streaming Processing, Data Lakes

## Professional Experience

### Data Engineer – Retail Analytics Project

Tata Consultancy Services (TCS) | Oct 2021 – Present

- Built ETL pipelines using ADF and PySpark on Databricks for 2+ TB of daily data.
- Developed Python scripts for deduplication, validation, and transformation.
- Integrated ADLS, Blob Storage, and Azure SQL with Databricks for unified data flow.
- Improved performance by 25% using partitioning, caching, and optimized joins.
- Automated workflow with Apache Airflow and added failure handling.
- Built near real-time pipelines with Spark Structured Streaming and Kafka.
- Used Delta Live Tables for managed ETL and observability.
- Achieved 99.9% accuracy using schema enforcement and validation layers.
- Coordinated with global teams to align technical delivery with business needs.

## Python Expertise

- Automated schema checks, metadata logging, and format standardization.
- Applied complex business rules in Python before Spark processing.

## PySpark Expertise

- Used complex joins, window functions, and aggregations effectively.
- Enhanced performance with broadcast joins, caching, and partitioning.
- Seamlessly integrated ADF, ADLS, and Databricks for automation.

## SQL Expertise

- Developed advanced SQL for staging, aggregation, and validation.
- Built parameterized procedures for quality checks.
- Maintained data integrity and referential accuracy.

## Education

B.Tech – Computer Science

Bharati Vidyapeeth Deemed University College of Engineering, Pune

Graduated: 2021

## Key Achievements

- Delivered 10+ production pipelines with <1% downtime.
- Optimized architecture to reduce time-to-insight.
- TCS Excellence Award in Data Engineering (2023).
- Gold Medalist – International Maths Olympiad.

## Additional Qualifications

Hands-on experience in end-to-end cloud data migration projects. Strong problem-solving, critical thinking, and adaptability in dynamic, fast-paced environments. Keen interest in expanding into AWS cloud technologies and advanced Data Science techniques.