Prachi Baranwal

+91 7607702662 | prachibaranwal27@gmail.com | https://www.linkedin.com/in/prachi-baranwal | Portfolio

PROFILE

Data Engineer (2.6+ years) with Azure Databricks (PySpark, SQL), Auto Loader, Delta Lake, and ADLS experience. Focused on scalable ETL pipelines, data quality, and RBAC governance. Strong foundation in Python and SQL. Familiar with Kafka fundamentals and Snowflake exposure.

SKILLS

- **Platforms:** Databricks, Azure Databricks
- **Processing:** Apache Spark, PySpark, Spark SQL, Delta Lake, Databricks Auto Loader
- **Data Engineering:** ETL/ELT, Data Pipelines, Incremental Loads, Schema Evolution, Data Quality, Reconciliation, Audit Logging, Monitoring/Alerting, SLA Validation, Lineage, RBAC
- **Programming:** Python, SQL
- **Orchestration/DevOps:** AutoSys, Azure Data Factory, Azure DevOps, Git, CI/CD, Jira
- **Storage/Warehousing:** ADLS, Delta Tables, Snowflake (exposure)
- **Concepts:** Data Modelling, Performance/Pipeline Optimization, Hadoop Ecosystem, Kafka (fundamentals)

EXPERIENCE

Infinite Computer Solutions, *Data Engineer*
Aug 2023 – Present
Bangalore

- Built **Azure Databricks** ETL/ELT (PySpark, SQL) processing **2M+ records/day**; improved **throughput 30**.
- Implemented config-driven DQ checks; improved validation **pass rate 25%**.
- Built incremental ingestion with **Auto Loader** (schema evolution, checkpointing, auto discovery) for scalable loads.
- Automated monitoring + retries via **AutoSys; cut failures 5%→1%** and **saved 8+ hrs/week**.
- Optimized Delta/Spark (Bronze–Silver–Gold, pruning, OPTIMIZE, **Z-ORDER**); cut **runtime 40 min/batch** and **cost ~18%**.
- Worked with stakeholders to deliver validated datasets for reporting and analytics.

Mobiloitte Technologies Pvt. Ltd., *Software Engineer Intern*
Feb 2023 – Apr 2023
Delhi

- Supported data extraction and validation and built small Python/SQL scripts to clean and analyze datasets; collaborated with the team to deliver outputs.

EDUCATION

Dr. A. P. J. Abdul Kalam Technical University, Lucknow,
Jul 2023
Lucknow, Uttar Pradesh
Bachelor of Technology in Computer Science (CGPA: 8.18 / 10.00)

- Relevant Coursework: Big Data, Machine Learning in CS (Python)

PROJECTS

WI MyChoice Implementation — Medicaid Program, *Databricks, PySpark, SQL, Azure Data Factory, AutoSys*

- •Built end-to-end ETL/ELT pipelines in Databricks (PySpark/Spark SQL) to produce curated datasets for downstream reporting.
- Implemented DQ framework (completeness/uniqueness/accuracy + reconciliation) with audit logging and failure routing to increase trust in datasets.
- Automated 100+ workflows using AutoSys scheduling, dependency handling, and retries to meet pipeline SLAs.

Healthcare Data Project, *Azure Databricks, PySpark, SQL, SSMS, ADF, AutoSys*

- Built pipelines to clean, transform, and load data into curated reporting tables.
- Added early checks (missing values, duplicates, row-count match) to keep data consistent across runs.
- Performed regression testing in SQL Server (SSMS) **across multiple tables** by comparing before vs after outputs to ensure changes didn't break results.
- Maintained runbooks/KT notes for validations and common failures to speed up support and handovers.

CERTIFICATES

- Databricks Certified Data Engineer Associate — Databricks
- Azure DP-900: Azure Data Fundamentals — Microsoft
- Foundations: Data, Data, Everywhere — Google
- Azure Databricks & Spark for Data Engineers — Udemy