

PRABHAT PAL

• prabhatpal820@gmail.com • +91 7906202294 • [Linkedin](#)

PROFESSIONAL SUMMARY

Motivated fresher Data Analyst / Data Engineer with strong skills in SQL, Python, ETL processes, data analysis, and Azure Data Factory, experienced in building data pipelines and ensuring data quality. Pursuing MCA (2024–2026) with hands-on exposure to cloud-based data solutions, Agile methodology, and JIRA to support scalable, data-driven decision-making.

SKILLS

- **Programming** Python, SQL, Pyspark
- **Cloud Tech** Microsoft Azure (Databricks, Data Factory, Synapse Analytics)
- **DBMS** MY SQL, Azure Data Lake, Delta Lake, Data Warehousing
- **Process** ETL / ELT, Data Modeling, Batch Processing
- **Others** Git, Github, Jira

INTERNSHIP

LetsGrowMore | Data Science Intern

June 2023 – August 2023

- Led the transition to a paperless business system by implementing an electronic booking and digital workflow solution, reducing labor costs by 30% and office overhead by 10% while improving accuracy and data security.
- Analyzed usage data from 5,000+ monthly active users (MAU) to support data-driven marketing and product decisions, resulting in 2× increase in average engagement time, 30% reduction in drop-off rate, and 3× growth in social media shares.

PROJECT

Brazilian E-commerce Data Pipeline using Medallion Architecture:

- Built a **scalable Databricks Lakehouse data pipeline** for Brazilian e-commerce data using **Medallion Architecture (Bronze, Silver, Gold)**.
- Ingested **transactional, customer, and product data** from **CSV files, APIs, and relational databases**.
- Performed **data cleansing, deduplication, and transformations** using **PySpark and Delta Lake**.
- Created **Gold layer aggregated datasets** for **sales analysis, customer segmentation, and product performance**.
- Automated pipelines with **Apache Airflow** and optimized performance using **Delta Lake (partitioning, Z-ordering, caching)**.
- Enabled **BI reporting** through **Power BI / Tableau** for real-time insights.

Healthcare Revenue Cycle Management (RCM) Data Engineering Project:

- Built an end-to-end **Healthcare RCM data pipeline** using **Azure Data Factory, Databricks, ADLS Gen2, and Delta Lake** with **Medallion Architecture**.
- Ingested data from **EMR (SQL DB), claims flat files, and NPI/ICD APIs** into a centralized data lake.
- Implemented **CDM, SCD Type 2, and data quality checks** in the Silver layer.
- Created **Gold layer fact and dimension tables** for KPIs like **Days in AR and AR > 90 days**.
- Improved security and performance using **Azure Key Vault, audit logging, is_active flags, and parallel execution**.

EDUCATION

- **Madan Mohan Malaviya University of Technology, Gorakhpur, Uttar Pradesh** August 2024 - April 2026
Master of Computer Applications (MCA)
- **Invertis University, Bareilly, Uttar Pradesh** August 2021 – June 2024
Bachelor of Computer Applications (BCA)