MOHIT TANEJA

Data Engineer | 4+ Years | PySpark · Databricks · Delta Lake · Azure & AWS · SQL

mohittaneja222@gmail.com | +91 8968883759 | LinkedIn | GitHub | Portfolio

PROFESSIONAL SUMMARY

Data Engineer with 4+ years of experience designing, building, and maintaining ETL/ELT pipelines and data platforms in IT services environments. Specialist in Databricks (Delta Lake, Unity Catalog, Lakeflow/DLT, Workflows) for large-scale data processing, with strong SQL and PySpark skills for building scalable, production-grade pipelines. Experienced with medallion architecture (Bronze/Silver/Gold), CI/CD-driven deployments, and using AI-assisted development tools (GitHub Copilot, Claude) to speed up debugging and documentation without compromising code quality.

TECHNICAL SKILLS

- Data Platform & Engineering: Databricks Unified Data Analytics Platform, ETL/ELT Pipeline Design & Maintenance, Apache Spark (PySpark), Data Cleaning & Transformation, Schema Validation, Data Reconciliation, DWH Design
- Databricks: Databricks Workflows, Delta Lake, Unity Catalog (governance & access control), Auto Loader, DLT/Lakeflow Pipelines
- BI & Visualisation: Power BI, Tableau (Calculated Fields, Parameters, Filters)
- Cloud: Microsoft Azure ADLS Gen2, Amazon Web Services S3
- Orchestration: Databricks Jobs, Lakeflow, Airflow
- Analytics & Reporting: KPI Analysis, EDA, Statistical Analysis, Business Intelligence, Data Governance
- Tools & Platforms: Databricks, Git, GitHub, SQL Server, MS Excel, AWS, Azure
- AI-Assisted Development: GitHub Copilot, Claude, ChatGPT, Databricks Genie

WORK EXPERIENCE

Data Engineer — Technology Specialist **June 2025 – Present**

Infosys

- Contributed to data platform blueprint and medallion architecture (Bronze/Silver/Gold) design on Databricks, collaborating with cross-functional teams to align the data model across [X] source systems.
- Implemented Dimension and Fact table transformations supporting data warehousing and analytical reporting for [X] downstream reports/dashboards.
- Applied prompt engineering and agentic-AI techniques to build an internal assistant for pipeline troubleshooting and SQL optimization, cutting time spent on routine debugging by an estimated [X]%.
- Served as the team's go-to resource for Databricks pipeline troubleshooting and dbt lineage documentation, reducing dependency on senior engineers for common issues.
- Used Delta Lake Time Travel and table versioning to recover from and audit erroneous updates in production ETL pipelines, avoiding [X] data-loss incidents.
- Created and maintained dbt documentation and lineage graphs, improving visibility of data dependencies across the platform.
- Deployed changes through Git-based CI/CD pipelines to support safe, collaborative development.
- Published and maintained Power BI and Tableau dashboards on a monthly release cycle, improving reporting turnaround by ~30% and enabling faster stakeholder decisions.
- Conducted root cause analysis (RCA) for recurring platform issues and implemented permanent fixes, reducing repeat incidents and improving platform stability.
- Improved CSAT scores through timely incident resolution and clear stakeholder communication on data platform issues.

Data Analyst / Data Engineer — Domain Expert **Sept 2022 – May 2025**

Wipro Pvt. Ltd.

- Automated data extraction, transformation, and reporting workflows by integrating SQL with Python (Pandas, NumPy), reducing manual processing time by an estimated [X]%.
- Contributed to medallion architecture (Bronze/Silver/Gold) implementation on Databricks, supporting [X] pipelines in production.

- Developed, tested, and maintained ETL pipelines across heterogeneous source systems on Databricks, validating schemas and reconciling records between source and target with zero data loss.
- Wrote complex SQL (MySQL, MSSQL) — including stored procedures — for extraction, transformation, and ad-hoc reporting, streamlining recurring operational data tasks.
- Applied Python (Pandas, NumPy) for large-scale data cleaning, transformation, and aggregation, and used AI-assisted tools to speed up scripting and documentation for a [X]% cut in delivery time on repeat tasks.
- Built visualization charts with Matplotlib and Seaborn to surface insights from cleaned datasets.
- Partnered with business stakeholders to translate reporting and pipeline requirements into data-driven deliverables, providing timely turnaround on ad-hoc requests.
- Performed incident troubleshooting, log analysis, and job recovery, keeping [X] scheduled jobs running within SLA.
- Coordinated with cross-functional teams on daily project status and process-improvement initiatives based on data analysis findings.

KEY PROJECTS

- Databricks – Food Chain Ecommerce | PySpark, Python, SQL, CI/CD, Lakeflow — Built a medallion-architecture pipeline with CI/CD deployment and incremental merge logic for order/inventory data.
- Databricks – Real Estate Insights | PySpark, SQL, Lakeflow — Built an end-to-end Lakeflow Spark Declarative Pipeline with CDC and data-quality validations, sourcing data from an AI agent, with CI/CD via Repo Bundles and Repo Actions.
- Tableau – Sales & Customer Performance Dashboard | Tableau, SQL, Excel — Designed a two-dashboard workbook with 12 worksheets, 30+ calculated fields, and four integrated CSV sources, including a year-over-year comparison parameter.
- SQL Data Warehouse – Retail Analytics Pipeline | SQL Server, ETL, Data Modeling — Designed and implemented a star-schema data warehouse supporting downstream BI reporting.

EDUCATION

- Bachelor of Computer Applications (BCA), Panjab University — 87% (2019 – 2022)
- Senior Secondary (12th), PSEB — 85% (2018 – 2019)

CERTIFICATIONS

- AWS Certified Cloud Practitioner
- AWS Certified Solutions Architect – Associate