

Vidit Dhull – Data Engineer

+91 9355512200

vidit.dh@gmail.com

Bengaluru, India

Portfolio: [viditdhull.github.io](https://github.com/viditdhull)

LinkedIn: linkedin.com/in/viditdhull/

LeetCode: leetcode.com/Vidit_Dhull/

SUMMARY

- Data Engineer with **3 years of experience** specializing in scalable **batch and streaming pipelines** using Azure services, **Spark, and Medallion architecture** for enterprise **analytics and ML readiness**.
- **Ranked 1st among 120+ students** in PG-Diploma (Big Data Analytics) at **CDAC Bengaluru**.
- Strong **Python** and **SQL** developer (**HackerRank 5-star; 600+ LeetCode problems**) with hands-on **CI/CD**, code reviews, and agile collaboration.

EXPERIENCE

Data Engineer, Alstom Bengaluru

March 2023 – Present

Dialogs — AI-Driven Log Analytics / Troubleshooting Platform:

- Multi-product log analytics platform that converts raw telemetry into governed, analytics-ready KPIs; reduced troubleshooting time from **months to weeks** and enabled real-time insights.
- Migrated platform from Microsoft Fabric to an **open stack** (Apache Spark, Airflow, PySpark, OneLake) using in-house DevOps tooling; rebuilt pipelines and set up **CI/CD**, reducing setup time by **70%**.
- Scaled ingestion to **10K+ daily events** for CCP2; implemented **partial runs** and optimized Spark decoders, cutting backfill time from **20 min to under 10 min** and improving resource efficiency.
- **Currently leading CCP2 stream:** managing requirements with stakeholders, finalizing KPI models, and coordinating onboarding while supporting platform-wide improvements.

Dialogs – Medallion Lakehouse (POC → MVP):

- Built initial medallion **lakehouse (Bronze/Silver/Gold)** on Microsoft Fabric; standardized heterogeneous logs to **Parquet** and delivered first KPIs and dashboards.
- Developed ingestion pipelines to **standardize raw logs → Parquet (Bronze)**; created Silver tables and Gold Layer KPIs via **SQL stored procedures**, enabling analytics and ML readiness.
- Delivered dashboards that cut manual troubleshooting from **approx. 2 months to under 2 weeks**; implemented Eventhouse for time-series data and Fabric Warehouse for KPI serving.

CCP2 System — Domain & Data Quality:

- Analyzed CCP2 logs with internal DVT; created domain quality rules that accelerated Dialogs ingestion and modeling.
- Authored validation checklists and fault-code glossary; reduced onboarding defects to **under 10%**.
- Built reusable scripts for **data quality checks**, **reducing manual rework by 60%** and enabling analytics readiness.

CERTIFICATIONS

- **DP-700: Fabric Data Engineer Associate (Microsoft)**, Sept 2025
- **DP-203: Azure Data Engineer Associate (Microsoft)**, June 2024
- **DP-900: Azure Data Fundamentals (Microsoft)**, April 2024
- **IBM Data Engineering Specialization (Coursera)**, Oct 2023

SKILLS

- **Programming & Scripting:** Python, SQL, PySpark
- **Data Engineering Tools:** Microsoft Fabric, Eventhouse, Azure Data Factory, ADLS Gen2, Spark on Kubernetes (K8s), Apache Airflow
- **Architecture:** Medallion Architecture, Lakehouse Design, Star Schema, ELT/ETL, Time-series Processing
- **Data Quality & Governance:** Validation Automation, Logging Frameworks, Colibra, KPI Modeling
- **Collaboration & Version Control:** GitLab, Agile/Scrum, CI/CD, Code Reviews, Documentation
- **Additional Skills:** Requirement Analysis, FinOps Awareness, Parquet/Tabular Data Formats, Delta Lake

Education

- **PG-Diploma, Big Data Analytics (Ranked 1st)** – CDAC Bengaluru | 82% (2022–2023)
- **B. Tech in Computer Science** – UIET-MDU, Haryana | (2018–2022)