

SHIVAM DUBEY

Indore, India | +91 7067892564 | shivamdubey2310711@gmail.com

LinkedIn: [linkedin.com/in/shivam-dubey20](https://www.linkedin.com/in/shivam-dubey20) | **GitHub:** github.com/shivamdubey2310

PROFESSIONAL SUMMARY

Certified DP-700 Fabric Data Engineer Associate and M.Tech (Integrated) AI & Data Science student (DAVV, 2027). Hands-on experience building cloud ETL pipelines, real-time streaming systems, and Medallion Delta Lake architectures using PySpark, SQL, Airflow, and Azure/Fabric. Seeking a Data Engineering internship to deliver production-grade data platform solutions.

EDUCATION

Integrated M.Tech - AI & Data Science | *School of Data Science and Forecasting, DAVV, Indore* | **Expected** : May 2027

Relevant coursework: DBMS, Big Data Analytics, Data Warehousing, Statistical Methods, PySpark, Python

TECHNICAL SKILLS

- **Programming Languages:** Python (Pandas, PySpark) · SQL (Window Functions, CTEs, Joins) · Bash
- **Cloud & Platforms:** Microsoft Fabric · Azure Data Factory · Databricks · Azure Synapse Analytics · ADLS Gen2 · Event Hubs · Azure Stream Analytics · Azure Key Vault
- **Databases:** PostgreSQL · MySQL · Azure SQL Database · MongoDB
- **Data Modeling:** Star Schema · Snowflake Schema · Slowly Changing Dimensions (SCD Type 1/2)
- **Big Data & ETL:** PySpark · Apache Spark · Delta Lake · Apache Airflow
- **DevOps & Tools:** Git/GitHub · Docker · Linux

PROJECTS

Near Real-Time CDC Lakehouse Pipeline | *Azure SQL, ADF, Databricks, Delta Lake, ADLS Gen2, Synapse Analytics* | [Repo](#)

- Cut data latency from **24h to sub-10 min** via metadata-driven CDC pipeline using SQL Change Tracking and ADF Tumbling Window Triggers over **1M+ records**.
- Built **Medallion Architecture** with Delta Lake MERGE upserts, PySpark dedup, and idempotent fault-tolerant retry logic.
- Enabled **zero-compute Synapse Serverless** querying on Gold Delta tables by processing only incremental deltas.

End-to-End Medallion ETL Pipeline (AQI Forecasting) | *Microsoft Fabric, PySpark, Delta Lake* | [Repo](#)

- Ingested AirNow API data into a **Bronze → Silver → Gold Medallion Architecture**, processing **10K+ AQI records** with config-driven PySpark and structured logging.
- Enforced Silver-layer quality via type casting, null handling, and composite-key dedup - **zero downstream anomalies**.
- Delivered Gold Delta tables with aggregated AQI metrics, improving **query performance 40%+** over raw reads.

Real-Time Weather Monitoring Pipeline | *Microsoft Fabric, Eventstreams, Eventhouse, Real-Time Dashboard* | [Repo](#)

- Built a real-time streaming pipeline ingesting live Delhi weather telemetry via Fabric Eventstreams into Eventhouse with **sub-second latency**, cutting data-to-insight time **90%+** vs batch polling.
- Configured dual Fabric Activators for high/low temperature threshold alerting, **eliminating 100% manual monitoring**.
- Developed a Real-Time Dashboard with live readings, trend lines, and alert history for **zero-delay stakeholder access**.

CERTIFICATES

- Fabric Data Engineer Associate (DP-700) | Microsoft | Feb 2026 | [Verify](#)
- Astronomer Certification for Apache Airflow 3 Fundamentals | Astronomer | Feb 2026 | [Credly](#)

LEADERSHIP & ACTIVITY

Cloud Lead, GDG on Campus – SDSF Sep 2023 - Aug 2024

- Drove **80+ Google Cloud learning path completions** by leading Cloud Study Jams; delivered Gen AI keynote to **50+ attendees**.
- Elevated college chapter to **Tier 1 status** - ranked among top-performing institutions nationally.