

Sujit Shejul

sujitshejul27@gmail.com · +91-7499780221 · Pune, Maharashtra, India

SUMMARY

Data Engineer with 3+ years building scalable ETL pipelines and analytics platforms. Hands-on designing ETL/ELT workflows and Medallion Architecture on Azure and Snowflake; proficient in PySpark, Python, and SQL. Experienced in Airflow orchestration, data quality validations, and creating analytics-ready datasets for Power BI reporting.

SKILLS

Cloud & Azure Data Services: Azure Data Factory, Azure Data Lake Storage (ADLS Gen2), Azure Databricks, Azure Synapse Analytics, Azure Dataflow, Delta Lake

Databricks Concepts: PySpark, Spark SQL, Delta Lake, Notebooks, Cluster Configuration, Job Scheduling, Unity Catalog, Auto Loader, Performance Tuning

Power BI & Reporting: Power BI Desktop & Service, DAX, Power Query, Data Modeling, Dashboards & Visualizations, Row-Level Security

Data Engineering: ETL pipelines, ETL/ELT Development, Data Pipeline Architecture, Data Integration & Migration, Data Transformation, Data Quality & Validation, Data Warehousing, Medallion Architecture, Workflow Automation

Programming Languages: Python, SQL

Databases: Snowflake, PostgreSQL, MySQL, Oracle SQL

Data Governance & Security: Data Quality Checks, Access Control, Consistency & Validation, Auditability

Tools & Technologies: Apache Airflow, DBeaver, Git, GitHub, REST API, JSON, CI/CD, Pandas, NumPy

Concepts: Stored Procedures, SQL Joins, CTEs, Window Functions, Query Optimization, Agile Methodology

Computer Science Fundamentals: Data Structures, Algorithms, Object-Oriented Programming (OOP)

EXPERIENCE

Software Engineer / Data Engineer · Capgemini Feb 2025 – Present · Pune, India

- Built ETL pipelines to ingest batch and CDC data from REST APIs into PostgreSQL using Python; orchestrated jobs with Apache Airflow for reliable scheduling and retries.
- Designed dimensional schemas (fact and dimension tables) and implemented SCD Type 1/2 to preserve historical accuracy for analytics datasets.
- Orchestrated ETL workflows using Airflow DAGs with monitoring and logging; implemented Snowflake data quality checks including null, duplicate, primary-key and row-count validations.
- Optimized SQL queries and stored procedures; maintained CI/CD pipelines and Git-based version control while delivering in Agile sprints.

Data Engineer · Dr. D.Y. Patil Medical College, Hospital & Research Center Jun 2024 – Feb 2025 · Pune, India

- Developed SQL-based ETL pipelines to extract, cleanse, and transform hospital data into analytics-ready tables for reporting and dashboards.
- Improved query performance through execution-plan analysis and targeted indexing to optimize slow-running reports and extracts.

Software Engineer Intern · Persistent Systems Jun 2023 – Jun 2024 · Pune, India

- Applied data structures and algorithms and built Python automation scripts to preprocess and validate datasets for engineering tasks.
- Authored SQL queries for retrieval and aggregation to support debugging and analytics; gained practical Agile teamwork experience.

PROJECTS

End-to-End Azure Sales Data Engineering Pipeline · Azure Data Lake, Azure Databricks, PySpark, Azure Synapse, Power BI

- Implemented Medallion Architecture (Bronze / Silver / Gold) pipelines on Azure Data Lake to standardize raw-to-curated ETL flows for sales data.

- Developed PySpark transformations in Azure Databricks and optimized partitioning and Spark jobs to reduce transformation runtimes by 30%.
- Built curated Gold-layer datasets in Azure Synapse and connected them to Power BI for real-time reporting dashboards.

YouTube Data Extraction ETL Pipeline · Python, SQL, Apache Airflow, PostgreSQL, REST API

- Built an ETL pipeline processing 100K+ records of channel analytics (views, likes, comments) via the YouTube Data API.
- Automated ETL scheduling with Airflow DAGs to replace manual runs and ensure daily, reliable pipeline execution.
- Loaded data into PostgreSQL with validation checks, improving accuracy for downstream analytics and reporting.

EDUCATION

Bachelor of Engineering, Computer Engineering · Dr. D.Y. Patil Institute of Technology	2020 – 2024 · GPA CGPA: 9.16
• Pune, India	

Higher Secondary Certificate (HSC) · R.B.N. Borawake College	2018 – 2020 · GPA Percentage: 79.5%
---	-------------------------------------

Secondary School Certificate (SSC) · Divya Daya Chand School	2018 · GPA Percentage: 92%
---	----------------------------

CERTIFICATIONS

- Databricks Certified Data Engineer Associate
- Microsoft Azure Fundamentals (AZ-900)
- Snowflake - The Complete Masterclass (Udemy)
- Python Programming Certification