

PARTH

parthsingh1253@gmail.com |  LinkedIn |  GitHub |  Portfolio | +91-8527713603

PROFILE SUMMARY

Data Engineering Enthusiast with hands-on experience in PySpark, Databricks, Snowflake, and Azure cloud services. Proven track record of building scalable ETL pipelines, implementing dbt for analytics engineering, and applying data quality validations to enable reliable downstream analytics. Strong analytical and problem-solving abilities with expertise in Agile methodologies, focused on creating robust, data-driven architectures that deliver actionable business insights.

EDUCATION

B.Tech in Computer Science with Specialization in Cloud Computing
University of Petroleum and Energy Studies, Dehradun

Aug 2021 – June 2025

EXPERIENCE

Associate Data Engineer July 2025 – Present
Nagarro, Gurugram

- Developed ETL pipelines using Azure Databricks, PySpark, and Medallion architecture to ingest, clean, and transform large-scale client data.
- Leveraged Delta Lakehouse storage and Delta Tables for efficient querying, reliability, and optimized data processing.
- Implemented CI/CD automation for data pipelines, ensuring scalable, faster, and error-free deployments.

Cloud and DevOps Intern July 2024 – Aug 2024

Canara HSBC Life Insurance Company, Gurugram

- Built and automated CI/CD pipelines using Jenkins, Docker, and Terraform, integrating shell scripting to streamline deployments and environment setup.
- Managed AWS cloud services (EC2, S3) for infrastructure and storage, improving scalability, reliability, and operational efficiency.

PROJECTS

End-to-End Data Pipeline – Azure Synapse, Snowflake, and dbt 

Personal Project (2025)

- Built a fully parameterized data ingestion pipeline using Azure Synapse to load multiple CSV datasets from ADLS into Snowflake using a single reusable pipeline.
- Implemented dbt models for staging and marts layers, applying data quality tests and transformations to generate analytics-ready fact tables.
- Designed a data validation framework to bifurcate records into clean and error tables, ensuring 100% data reconciliation and reliable downstream reporting.

Batch ETL Pipeline - PySpark and Azure Databricks 

Personal Project (2025)

- Designed and implemented a scalable batch ETL pipeline on Azure Databricks using PySpark, processing 1M+ NYC Taxi Trip records from raw CSV into structured and aggregated datasets.
- Improved data quality by 30% through type casting, filtering invalid rows, and applying daily vendor-level aggregations, enabling accurate insights and reporting.
- Enhanced query performance by 40% by storing processed outputs in Parquet format on Azure Blob Storage, ensuring cost-efficient cloud storage and fast downstream analytics.

TECHNICAL SKILLS

Languages: Python, SQL, Java.

Big Data: Apache PySpark, Hadoop, Delta Lake.

Cloud Platforms: Microsoft Azure (ADF, ADLS, Databricks, Synapse), AWS (EC2, S3).

Data Warehousing: Snowflake, dbt.

Tools: Git, Docker, CI/CD (Jenkins), Power BI.

CERTIFICATIONS

Databricks Certified Data Engineer Associate: [Link](#)

HackerRank SQL Advance Certificate: [Link](#)

HackerRank SQL Intermediate Certificate: [Link](#)