

# PARTH

[parthsingh1253@gmail.com](mailto:parthsingh1253@gmail.com) | [in](#) LinkedIn | [G](#)itHub | [P](#)ortfolio | +91-8527713603

## PROFILE SUMMARY

Data Engineering Enthusiast with hands-on experience in PySpark, Databricks, and Azure cloud services. Proven track record of collaborating with stakeholders to build and optimize ETL pipelines, implementing Delta Lake for efficient data storage, and enabling real-time data processing solutions. Strong analytical and problem-solving abilities with expertise in Agile methodologies, focused on creating scalable, 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**  
*Nagarro, Gurugram*

July 2025 – Present

- 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**  
*Canara HSBC Life Insurance Company, Gurugram*

July 2024 – Aug 2024

- 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

**Batch ETL Pipeline - PySpark and Azure Databricks** [!\[\]\(4f6bf54ae7e4144a72d78316053e412d\_img.jpg\)](#)  
*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.

**NYC Taxi Data Pipeline – Azure Data Factory and Databricks** [!\[\]\(56549452e01ca28bdf2500ced9653143\_img.jpg\)](#)  
*Personal Project (2025)*

- Developed an automated ETL pipeline with ADF, Databricks, and ADLS, ingesting and processing 1M+ NYC Taxi records into optimized Parquet datasets.
- Enhanced dataset reliability by 30% through schema enforcement, removing inconsistent records and consolidating vendor/payment trends for reporting.
- Enabled 40% faster reporting by scheduling pipelines with triggers, delivering insights on vendor performance, payment trends, and trip patterns.

## TECHNICAL SKILLS

**Languages:** Python, SQL, Java.

**Big Data:** Apache PySpark, Hadoop, Delta Lake.

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

**Data Pipelines:** Azure Data Factory, Apache Airflow.

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

## CERTIFICATIONS

AWS Academy Graduate – Cloud Developing: [Link](#)

HackerRank SQL Intermediate Certificate: [Link](#)

HackerRank SQL Basic Certificate: [Link](#)