

PARTH

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

PROFILE SUMMARY

Data Engineer with hands-on experience in PySpark, Databricks, and Azure cloud services. Skilled in building and optimizing ETL pipelines, implementing Delta Lake for efficient data storage, and enabling real-time data processing solutions. Strong problem-solving abilities with a focus on creating scalable, data-driven architectures that deliver business insights and operational efficiency.

EDUCATION

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

CGPA: 7.40
2021 – 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 – August 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 using PySpark on 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.

NYC Taxi Data Pipeline – Azure Data Factory and Databricks
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, filtering invalid rows, and aggregating vendor/payment insights for analytics.
- Enabled 40% faster reporting by scheduling pipelines with triggers, delivering insights on vendor performance, payment trends, and trip patterns.

TECHNICAL SKILLS

Languages: Python, SQL, Shell.

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.

HackerRank SQL Intermediate Certificate.

HackerRank SQL Basic Certificate.