Parth Chauhan

Azure Data Engineer

📞 8425958106 ✉ parthchauhan.data@gmail.com 🔗 linkedin.com/in/chauhan-parth894 ⭘ github.com/Parth420

Summary

Data engineer skilled in building data lakes, **modelling datasets**, and **creating dashboards** to support financial performance insights. Proficient in PySpark, SQL, and BI tools, with **experience in ad-hoc projects**, change management, and **delivering business-aligned solutions**.

Skills

Programming: Python, **PySpark, SQL**.
Technologies: Apache Spark, Microsoft Fabric, Delta Lake, Databricks, Azure Data Lake Gen 2.
Processes: **Data Ingestion**, Data Warehousing, Performance Tuning, **Data Modelling**, Cloud Infrastructure Management, Data Cataloging, Data Pipeline Orchestration, **Data Engineering**, **Data Transformation**.
Cloud: Azure Data Factory, Azure Synapse Analytics.
Soft Skills: Process and delivery mindset, Methodological and operational improvements, Strong presentation skills.

Education

Shah & Anchor (MU) — **July 2019 – December 2023**
*B. Tech **Computer Engineering** (CGPA: 7.2/10)* — *Mumbai, MH*
Relevant Coursework: Distributed Computing, Big Data Analysis, Data Warehousing, Cloud Computing, Python Programming, Project Management.

Projects

Lending Club | *PySpark, Git, Pytest, Agile.* — **September 2024**
- Evaluated **resource requirements** for data processing tasks, creating an optimization plan that **improved team efficiency** by 15%, benefiting project timelines and delivery.
- Engineered and deployed **PySpark transformations** to process over 2 million financial records daily, reducing data processing time by 15% and **enhancing pipeline reliability**.
- Conducted unit testing using Pytest, achieving 95% **accuracy in data processing** and identifying critical defects early, **reducing errors** in production by 20%.
- Created financial dashboards in Power BI to **monitor budget allocations** and expenditure, improving financial transparency by 15%.

Data Pipeline Development on Azure | *Azure Data Factory, Azure Databricks, ADLS Gen2.* — **July 2024**
- Designed and implemented an automated pipeline in Azure Data Factory, **reducing manual intervention** by 50% by automating **file validation and processing** for over 200 files daily in ADLS Gen2.
- Integrated Azure Databricks into the pipeline, improving processing flexibility and **enhancing performance** by 20%, enabling **efficient handling of large datasets**.
- Developed a streamlined data processing workflow, boosting ETL efficiency by 15% and **reducing report generation time** from 2 hours to 30 minutes, accelerating data-driven decision-making.

Earthquake Data Pipeline | *Microsoft Fabric, Python, Power BI.* — **December 2024**
- Built a **data ingestion** pipeline to fetch earthquake event data from the USGS API using Python requests, storing raw data in Microsoft Fabric's Lakehouse bronze layer for initial processing.
- Planned and implemented a **medallion architecture**, transforming and refining data through bronze, silver, and gold layers, **ensuring high-quality datasets** ready for downstream consumption.
- Orchestrated the entire pipeline using Azure Data Factory, automating successive workflows, and created preliminary BI dashboards in Power BI for data visualization, **improving analysis efficiency** by 20%.

Certificates

- Microsoft DP-600: Fabric Analytics Engineer (June 2024).
- Microsoft DP-203: Azure Data Engineer (Pursuing).
- Astronomer - Apache Airflow (July 2024).
- Trendytech - Big Data Masters (October 2024).
- IBM - Agile Explorer (December 2024).
- IBM - Cloud Computing Fundamentals (October 2024).

Achievements

Achieved 95%+ completion rates on certifications, demonstrating a **commitment to continuous learning** and skill advancement.

Authored and published six insightful articles on LinkedIn, enhancing industry knowledge in data engineering and cloud technologies while **providing actionable strategies** that engaged over 1,000 fellow professionals seeking expertise.