

# ROHIT GUPTA

Pune, India | +91-9419128023 | [rohit.ajaygupta@gmail.com](mailto:rohit.ajaygupta@gmail.com) | [linkedin.com/in/rohit-gupta-001/](https://www.linkedin.com/in/rohit-gupta-001/) | [github.com/rohitgupta29](https://github.com/rohitgupta29) |  
| <https://leetcode.com/u/user1839q/> | <https://medium.com/@RohitAjaygupta> |  
Microsoft Certified Azure Data Engineer

## PROFILE SUMMARY

Results-driven **Data Engineer** with proven expertise in designing and implementing robust ETL pipelines, data models, and scalable data infrastructure. Proficient in **Python, SQL, Apache Spark**, and **Data Structures**, with a strong ability to optimize system performance and ensure seamless data flow across platforms. Experienced in collaborating with **cross-functional teams** to transform business requirements into actionable technical solutions. Committed to continuous learning and staying ahead of emerging technologies to drive **process automation, cost efficiency, and data-driven decision-making**.

## EXPERIENCE

**Cognologix Technologies**, Pune, **Associate Data Engineer** | Sep 2021 - Present

- **Designed and developed end-to-end data pipeline solutions** for the **mobility domain**, focusing on **data modeling, ETL, aggregation for visualizations, security, and logging** using **Databricks, Azure Synapse, ADLS Gen2, Entra ID, and KQL**.
- Built scalable data pipelines on **AWS** leveraging **S3, EMR, and RDS** for an **ed-tech product**, improving processing efficiency and reducing costs.
- Developed modular **data transformation utilities** using **PySpark, Python, SQL**, and integrated with **PostgreSQL, MongoDB, and Delta Lake**.
- Engineered a secure and reusable **Data Pipeline Ecosystem** using **Azure Synapse Analytics**, enhancing data governance and access control.
- Implemented **Airflow-based orchestration** for seamless batch data ingestion from **dbt**, improving pipeline reliability and visibility.
- Built a **critical connector API** in **Flask** to transform nested JSON structures for ingestion into a data warehouse and ODS. Solved complex challenges including dependency resolution, null handling, and performance optimization via **multithreading**.
- Created a **plug-and-play data validation tool** using **Great Expectations** and **Flask**, providing dynamic reports with tables and visualizations. Achieved a **70% reduction in validation time**, significantly accelerating the ETL process.
- Developed a **monthly employee insight reporting tool** for HR and managers using **machine learning** with **scikit-learn, pandas, matplotlib, and seaborn**. Deployed on **AWS EC2**, supporting decision-making in a remote work context.
- Led development of an **NLP-based web app** that analyzed social media posts to surface emotional well-being signals of employees during COVID. Built using **Flask, Pandas, and Plotly**, helping HR with proactive engagement.
- Authored multiple internal technical blogs for the **DataAI department**, sharing insights, best practices, and learnings across teams to foster a strong engineering culture.

**Prepared for Civil Services** (Economy, Indian Constitution, Psychology) June 2014-Jan 2020

**Syamantak- School without Walls** (mentoring and guidance) **Volunteer** Jan 2014- April 2014

- Mentored special children through skill based learning for better cognitive development and self-reliance.

**NTPC Unchahar** (Team work) **Engineering trainee** Nov 2012- Aug 2013

- Received Letter of appreciation from AGM Operations for performance in a plant overhauling project.

## TECHNICAL SKILLS

Data engineering:	System Design, data warehousing, data modeling
Machine Learning:	Data Structures, Algorithms, Sci-kit Learn, Spacy, textblob,NLTK, LLMs
Languages:	Python (Flask,NumPy, Pandas), Pyspark, SQL, Java (spring boot)
DevOps Services:	Docker, Kubernetes, PostgreSQL, Linux, Shell Scripting
AWS Services:	S3, EC2, IAM, EMR, EKS
Azure Services:	Synapse Analytics, Azure Databricks, ADLS , Streaming Services, Entra Id
Tools & Technologies:	Linux, Airflow, kafka, Plotly Dash, Microsoft Excel, draw.io

## CERTIFICATIONS

- Microsoft Certified: Azure Data Engineer Associate | Credential ID: 4F7E535CA55FF2DA
- Advanced SQL ([HackerRank](#)) | Credential ID: 4865cba21b27
- Certified AI Professional ( NIELIT, Calicut) | Credential ID 7729
- Preparing for Databricks Certified Data Engineer Professional

## EDUCATION

### SHRI MATA VAISHNO DEVI UNIVERSITY, J&K

Bachelor of Technology, Mechanical Engineering June 2012