

Rahul Chavan

in [LinkedIn](#)  [Github](#)

Phone : [+91-855-493-8009](tel:+91-855-493-8009) | | Email : rcchavan663@gmail.com

Professional Summary

Hello! I'm Rahul, a data engineer with 2+ years of experience in building scalable data pipelines and AI-powered solutions. I've worked on SAS to PySpark migration using Databricks, cloud-native pipelines, and LLM-driven apps. Proficient in Python, SQL, PySpark, and cloud tools like AWS, Azure, and Snowflake. I'm eager to contribute to data-driven innovation and automation.

Qualification

Bachelor of Technology – Civil Engineering
SKN Sinhgad College of Engineering, Pandharpur
CGPA: 9.8/10 (First Class with Distinction)

Aug`2018 – Aug`2022

Experience

Data Engineer at LTIMindtree

Sep`2023 – Present

- Migrated SAS workflows to PySpark using Databricks custom tools.
- Built scalable ETL pipelines with Python, SQL, and Databricks.
- Deployed & fine-tuned LLMs using SageMaker and GCP Vertex AI.
- Contributed to Meta Migrator, Data Migrator, and Analyzer modules.
- Automated SAS-to-PySpark lineage extraction using Python-based analyzers and regex parsers.
- Designed logging and exception monitoring pipeline using AWS Lambda and CloudWatch for ETL workflows.
- Collaborated with cross-functional teams to optimize cloud costs by 18% using job parallelization and cluster tuning in Databricks.
- Enabled text-based data queries using LLMs in Streamlit dashboards.

Projects

- SmartQuery-AI: Text-to-SQL app using Gemini, AgenticAI. XLSX upload → optimized SQL. [\[Link\]](#)
- MediBot: Streamlit + Gemini-powered chatbot for hospital queries. [\[Link\]](#)
- Pik-Pok: AWS-based video sharing app using EC2, S3, Lambda, CloudFront.

Key Courses & Certifications

- AWS Data Engineer Associate [\[Link\]](#)
- Snowflake SnowPro Core [\[Link\]](#)
- Databricks Data Engineer Associate [\[Link\]](#)
- Oracle Generative AI Certified [\[Link\]](#)
- Microsoft Fabric Analytics Associate [\[Link\]](#)

Technical Skills

- Languages: Python, SQL, PySpark, SAS
- Big Data: Spark, Hadoop, Airflow, Databricks
- Cloud: AWS (S3, Glue, Lambda), Azure (Data Factory), GCP (Vertex AI)
- Tools: Snowflake, Power BI, Looker, Git
- AI/ML: LLMs, RAG pipelines, Generative AI, Chatbots