


# RAHUL CHAVAN

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

## DATA ENGINEER

### Professional Summary

Data Engineer with 2+ years of experience in designing, building, and optimizing data pipelines and ETL/ELT workflows in cloud environments (AWS, Azure, Databricks, Snowflake). Proficient in Python, SQL, and PySpark for data processing, with hands-on experience in SAS to PySpark migrations, Generative AI applications, and real-time data ingestion. Strong knowledge in cloud-native tools, big data frameworks, and AI-driven automation.

### Technical Skills

- Languages: Python, SQL, PySpark, SAS
- Big Data & Processing: Spark, Hadoop, Databricks, Airflow
- Cloud Platforms: AWS (EC2, S3, Lambda, Glue, SageMaker), Azure (Data Factory, Synapse, DevOps), GCP (Vertex AI)
- Data Engineering: ETL, ELT, Data Warehousing, Data Lake, Data Modeling, Data Integration, Data Governance
- AI & ML: Open-Source LLM Fine-tuning, RAG Pipelines, Generative AI, Chatbots
- Tools & Visualization: Snowflake, Power BI, QuickSight, Looker

### Professional Experience

#### **LTIMindtree – Data Engineer**

Sep`2023 – Present

- **Data Engineering & Pipeline Development**
  - Migrated complex SAS and PL/SQL ETL workflows to PySpark using Databricks notebooks and custom analyzers, improving scalability and reducing execution time by 40%.
  - Designed and maintained scalable ETL pipelines in PySpark, SQL, and Python across multi-cloud platforms (AWS, Azure, GCP) for batch and near real-time data processing.
  - Integrated with Snowflake, S3, Delta Lake, and Azure Data Lake to enable secure, efficient data movement aligned with enterprise governance standards.
  - Implemented data quality validation layers using Great Expectations and custom PySpark UDFs to ensure schema consistency and high data accuracy.
- **LLM & GenAI Integration**
  - Built GenAI-powered applications for code translation, lineage extraction, and metadata classification using LangChain, OpenAI, Vertex AI, and SageMaker.
  - Enabled natural language to SQL/PySpark conversion within Streamlit dashboards, allowing business users to perform ad-hoc queries without technical support.
  - Developed agentic systems (Meta Migrator, Analyzer, Validator) for SQL-to-PySpark conversion workflows using multi-agent orchestration and result caching.

- **Platform Engineering & Monitoring**
  - Designed automated logging and exception monitoring pipelines using AWS Lambda, CloudWatch, SNS, and S3 for ETL observability and real-time alerting.
  - Built modular and reusable codebases for PySpark transformations with parameterized configurations and metadata-driven orchestration.
  - Achieved 18% cost reduction in Databricks jobs through job parallelization, dynamic cluster sizing, and Spark performance tuning.
- **Cloud & DevOps**
  - Delivered end-to-end pipeline lifecycle support across AWS (Lambda, SageMaker, Glue), GCP (Vertex AI, BigQuery), and Azure (Blob Storage, ML Studio).
  - Implemented CI/CD for ETL and AI pipelines using GitHub Actions, Docker, and Terraform to streamline deployments and infrastructure provisioning.

## Projects

- **Pik-Pok – Social Video Sharing Platform**
  - Built using AWS EC2, S3, Lambda, and CloudFront for serverless content delivery.
  - Implemented backend in WordPress, ensuring secure media management.
- **MediBot – AI-Powered Healthcare Assistant**
  - Developed using Streamlit and Google Gemini API for patient interaction automation.
  - Used fine-tuned LLMs to generate custom responses for hospital-related queries.
- **Natural Language to SQL Generator (Text-to-SQL Web App)**
  - Developed a Streamlit-based Text-to-SQL generator that translates natural language queries into optimized SQL statements using AI APIs (Gemini, TogetherAI, AgenticAI).
  - Enabled users to upload XLSX datasets, dynamically extract and parse column metadata, and generate context-aware SQL queries for analysis.

## Certifications

- AWS Data Engineer Associate
- Snowflake SnowPro Core
- Databricks Data Engineer Associate
- Oracle Generative AI Certified
- Microsoft Fabric Analytics Associate

## Education

### **Bachelor of Technology in Civil Engineering**

SKN Sinhgad College of Engineering, Pandharpur | 07/2018 – 08/2022 | CGPA: 9.8 (First Class with Distinction)