

Rahul Chavan

Pune, India | +91-855-493-8009 | rcchavan663@gmail.com | LinkedIn | GitHub

PROFESSIONAL SUMMARY

Data Engineer with 2+ years of experience in designing, building, and optimizing ETL/ELT pipelines in cloud environments (AWS, Azure, GCP). Specialized in SAS/PLSQL to PySpark migration, LLM/GenAI integration, and real-time ingestion workflows. Skilled in Python, SQL, Databricks, Snowflake, and multi-cloud architecture.

TECHNICAL SKILLS

- Languages:** Python, SQL, PySpark, SAS
- Big Data Tools:** Spark, Hadoop, Databricks, Airflow
- Cloud:** AWS (S3, Lambda, Glue, SageMaker), Azure (ADF, Synapse), GCP (Vertex AI)
- Data Engineering:** ETL/ELT, Modeling, Warehousing, Governance
- AI/ML:** LangChain, RAG Pipelines, LLM Fine-tuning, Chatbots
- Tools:** Snowflake, Power BI, Looker, Terraform, Docker

PROFESSIONAL EXPERIENCE

Data Engineer | *LTIMindtree, Pune* Sep 2023 – Present
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.
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.
Delivered CI/CD pipelines using GitHub Actions, Docker, and Terraform.

PROJECTS

- Pik-Pok – Social Video Sharing** | *AWS, WordPress* 2024
 - Built a serverless video-sharing backend using EC2, S3, Lambda, CloudFront.
 - Integrated CDN caching, pre-signed URLs, and multilingual CMS.
- MediBot – AI-Powered Healthcare Assistant** | *Streamlit, Gemini API, LLM* 2024
 - Developed chatbot for hospital queries using fine-tuned LLMs and Streamlit.
 - Deployed on Hugging Face Spaces, supporting multi-specialty logic.
- Natural Language to SQL Generator** | *Streamlit, Gemini, TogetherAI* 2024
 - Built AI-driven SQL generator handling natural queries and XLSX schema.
 - Used fallback caching, column detection, and query optimization.

CERTIFICATIONS

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

EDUCATION

B.Tech in Civil Engineering | *SKN Sinhgad College of Engineering, Pandharpur* Aug 2022
CGPA: 9.8 / 10