

Rahul Chavan

+91-8554938009 | rcchavan663@gmail.com | [LinkedIn](#) | Pune, Maharashtra

Data Engineer

Experienced in building scalable data pipelines, optimizing ETL/ELT workflows, and cloud data solutions using AWS, Azure, Databricks, and Snowflake. Proficient in Python, SQL, and PySpark for data processing, governance, and performance optimization.

Beyond data engineering, I have hands-on experience in Generative AI, including fine-tuning open-source models and developing advanced RAG pipelines (Agentic & Multi-Agentic RAG) for intelligent data retrieval and automation. Certified in AWS, Databricks, Snowflake, Oracle Generative AI, and Power BI, with expertise in cloud migration and AI-driven data solutions.

Professional Summary

- Experienced Data Engineer with 1.5+ years of expertise in designing and implementing scalable data pipelines and systems.
- Proficient in Python and SQL, with hands-on experience in developing and optimizing data processing algorithms.
- Skilled in big data technologies such as Hadoop and Spark, ensuring efficient data storage, retrieval, and analysis.
- Familiarity with cloud platforms including AWS, Azure, and Google Cloud, adept at deploying and managing data infrastructure in the cloud.
- Successfully migrated SAS-based workflows to PySpark, improving performance and scalability while maintaining data integrity.
- Hands-on experience in Generative AI, including fine-tuning open-source models and developing advanced RAG (Agentic & Multi-Agentic RAG) pipelines for intelligent data retrieval and automation, enhancing AI-driven decision-making processes.

Project Experience

CoE Project

Sept 2023 - Present

LTIMindtree

- Worked on the Scintilla tool, focusing on SAS to PySpark migration and code conversion, including the creation of PySpark scripts to manually test out migrated SAS code.
- Leveraged Databricks for data processing and analytics, utilizing Databricks Notebooks and Catalog for efficient data handling.
- Designed, constructed, and maintained scalable ETL data pipelines for data ingestion, cleaning, and processing using Python and SQL.
- Contributed to various services/modules of the Scintilla tool, including Connection, Analyser, Meta Migrator, and Data Migrator, to ensure seamless migration to Databricks.
- Worked on AWS SageMaker for fine-tuning models and GCP Vertex AI for model deployment, enhancing AI-driven capabilities.
- Created RAG (Retrieval-Augmented Generation) pipelines using open-source models, driving intelligent data retrieval and AI automation.
- Developed a project using AWS services, including AWS CloudFront for content delivery and Elastic Transcoder for media transcoding, ensuring efficient data distribution and processing.
- Created Generative AI applications, including chatbots using Gemini and custom enhancement tools tailored to specific use cases, leveraging the Generative AI framework to meet business needs.
- Executed end-to-end AI deployments in Databricks, managing the full lifecycle of AI solutions from model development to production deployment, optimizing data flows and ensuring seamless integration with cloud services.

Projects

Pik-Pok Social Video Sharing Platform

Tools: AWS EC2, AWS S3, AWS Lambda, AWS CloudFront, WordPress

Roles: Planning, Front End Development, Backend Development, Deployment, Execution

- Developed Pik-Pok, a social video sharing platform utilizing AWS EC2 for scalable compute, AWS S3 for secure storage, AWS Lambda for serverless processing, and AWS CloudFront for global content delivery.
- Led the front-end development using WordPress to create a user-friendly interface.
- Designed and implemented backend architecture to ensure high performance and scalability.
- Managed the deployment and execution of the platform, ensuring smooth integration of AWS services for optimal performance.
- Ensured secure storage and efficient content delivery for users globally, optimizing the platform for video sharing and streaming.

MediBot - AI-Powered Healthcare Assistant

Tools: Streamlit, Google Gemini API

- Developed MediBot, an AI-powered healthcare assistant designed to assist patients with common questions about hospital services.
- Utilized Streamlit to build an interactive user interface, providing an easy-to-use platform for patients.
- Integrated Google's Gemini API to enhance the chatbot's capabilities, enabling advanced natural language processing.
- The chatbot generates responses based on a custom dataset and advanced AI models, ensuring accurate and relevant information.
- Aimed to provide quick, helpful answers, improving the overall user experience in a healthcare environment.

Technical Skills

- Programming Languages → Python, SQL, PySpark
- Cloud Platforms → Azure (Data Factory, Databricks, Synapse, Blob Storage, Data Lake, Azure DevOps), AWS (S3, Lambda, EC2, Glue, EMR)
- Data Engineering → Data Modeling, ETL/ELT, Data Warehousing, Data Pipeline Optimization, Data Cleaning, Data Governance, Data Integration
- AI & ML → Generative AI Deployment, Open-Source Model Fine-Tuning, Agentic & Multi-Agentic RAG Pipelines
- Data Visualization → QuickSight, Looker

Certifications

- | | |
|---|-------------------|
| • AWS Data Engineer Associate | 08/2024 - 18/2027 |
| • Snowflake SnowPro Core Certification | 05/2024 - 05/2026 |
| • Databricks Data Engineer - Associate | 09/2024 - 12/2026 |
| • OCI 2024 Generative AI Certified Professional | 06/2024 - 06/2026 |
| • Fabric Analytics Engineer Associate | 22/2024-22/2025 |

Education

Bachelor of Technology in Civil Engineering

Skn Sinhgad College of Engineering, Pandharpur

07/2018 - 08/2022

- CGPA : 9.8 | First Class with Distinction

Additional Information

- [GitHub-Profile](#)
- [Hugging Face Finetuned model](#)