RAHUL CHAVAN

CLOUD DATA ENGINEER

+91 8554938009 | rcchavan663@gmail.com | www.linkedin.com/in/rahulchavan01

Summary

Dynamic Data Engineer with around 2+ years of experience designing and optimizing scalable data solutions. Proficient in building efficient data pipelines, improving data quality, and transforming raw data into actionable insights through visualization. Proven ability to drive efficiency and innovation in fast-paced startup environments, contributing to business growth through advanced data engineering practices.

Professional Work Experience

LTIMindtree

Data Engineer : Sep 2023 - Current

- Migrated SAS workflows to PySpark using Databricks custom tools and built robust ETL pipelines with Python, SQL, and Databricks. Automated SAS-to-PySpark lineage extraction using Python-based analyzers and regex parsers.
- Deployed and fine-tuned LLMs using AWS SageMaker and GCP Vertex AI. Enabled text-based data queries using LLMs within Streamlit dashboards.
- Developed and maintained data pipelines on AWS using S3, Glue, Lambda, and Redshift, resulting in a 30% reduction in data processing time.
- Collaborated with cross-functional teams to optimize cloud costs by 18% using job parallelization and cluster tuning in Databricks.
- Developed interactive dashboards and reports using Power BI and Quicksight.
- Contributed to the development of Meta Migrator, Data Migrator, and Analyzer modules.

Skills

- Cloud Platforms: Google Cloud Platform(GCP), Amazon Web Services(AWS), Microsoft Azure
- Data Warehouse: Snowflake, AWS Redshift
- ETL Tools: AWS Glue, AWS athena, Azure data factory(ADF)
- Data Intelligence Platform: Azure Databricks, Microsoft Fabric
- Languages/Framework: Python, SQL, PySpark, SAS
- Business Intelligence Tools: Power BI, AWS quicksight
- AI/ML: Large Language Models(LLMs), RAG, Generative AI

Certification

AWS Data Engineer Associate : Credentials Oracle Generative Ai Certified : Credentials Snowflake SnowPro Core : Credentials
Microsoft Fabric Analytics : Credentials Databricks Data Engineer Assoc. : Credentials

Projects

Google Cloud Advanced Data Analytics

- Designed and automated ETL pipelines using GCP services such as Big Query, Data Proc, and Cloud Functions for data processing.
- Implemented event-driven workflows with Cloud Pub/Sub and Cloud Composer for seamless task scheduling and execution.
- Integrated AI Platform for deploying machine learning models to enhance predictive analytics and insights.
- Utilized Cloud Monitoring, Logging, and Alerting to ensure system reliability and performance tracking.
- Created Power BI dashboards based on Big Query datasets for actionable insights and comprehensive reporting.
- Streamlined collaboration and version control through GitHub for efficient development workflows.

Modern Data Architecture with AWS Glue, S3, Athena, and Redshift

- Designed a modern data architecture leveraging AWS Glue, S3, Athena, and Redshift for scalable data lake and warehousing solutions.
- Engineered a serverless data lake on S3 with AWS Glue for data ingestion, cataloging, and efficient data exploration.
- Developed ETL workflows using AWS Glue with PySpark for cost-effective data transformation and processing.
- Configured Amazon Athena for ad-hoc SQL querying, enabling exploratory data analysis directly from the data lake.
- Optimized Amazon Redshift for high-performance data warehousing and advanced analytics, integrating with Power BI for insights visualization.
- Built data pipelines for seamless integration between S3 and Redshift using AWS Glue.
- Enhanced analytics by incorporating basic AI/ML capabilities with AWS services like SageMaker, Comprehend, and Rekognition.

SmartQuery-AI: Text-to-SQL App using Gemini, Agentic AI

- Natural Language to SQL Generation: Developed a cutting-edge text-to-SQL application, SmartQuery-AI, enabling users to query databases using natural language inputs instead of complex SQL syntax.
- Intelligent Query Generation with Gemini: Integrated Google's Gemini large language model (LLM) to accurately interpret user prompts, generating optimized SQL queries with an understanding of context, schema, and user intent.
- Enhanced Reasoning and Optimization with Agentic AI: Incorporated Agentic AI principles for advanced reasoning, enabling intelligent analysis of user intent, contextualization of queries using database schema, and AI-driven optimization of generated SQL for efficiency and performance.
- User-Friendly Data Exploration: Provided a user-friendly interface, facilitating the upload of XLSX files and enabling interactive data querying in plain English, thereby reducing the barrier to entry for non-technical users.
- Enabled Natural Language Data Querying: Developed an application leveraging Gemini and Agentic AI to convert natural language questions into optimized SQL queries, significantly simplifying data access for non-technical users and accelerating insights.

Education

- B.Tech - 9.82/10 2018-2022
- HSC Science - 77.78% 2018
- SSC - 82.40% 2016