

RISHAV DUTTA

Data Engineer | Python Developer | Cloud & Automation Enthusiast

Email: rishavdutta.kgp@gmail.com | Phone: +91 7584954316 | Location: Bengaluru, India

Total Experience: 4.5 Years

Digital Portfolio: <https://rishav0123.github.io/portfolio/>

PROFILE SUMMARY

Innovative and results-driven Data Engineer with 4+ years of experience designing and building scalable ETL pipelines, real-time data streaming solutions, and cloud-native architectures using PySpark, AWS, and automation frameworks. Skilled in developing data ingestion, transformation, and visualization workflows to enable actionable insights. Passionate about optimizing data systems for performance, cost efficiency, and reliability.

KEY SKILLS

Languages: Python, SQL, JavaScript, Java, C#

Data Engineering & Big Data: PySpark, Apache Kafka, Airflow, AWS Glue, Spark SQL, ETL/ELT, Batch & Streaming Pipelines

Databases: Amazon Redshift, PostgreSQL, MySQL, MongoDB, Teradata

Cloud & DevOps: AWS (Lambda, EC2, S3, Step Functions, CloudTrail, IAM), Docker

Frameworks & Tools: Django, Playwright, Pandas, NumPy, Jupyter, Git, REST APIs

BI & Visualization: Power BI, Tableau

Other Tools: Supabase, Frappe, ERPNext, WordPress Plugin Development

Soft Skills: Leadership, Communication, Critical Thinking, Agile Methodologies

WORK EXPERIENCE

Tata Consultancy Services — Data Engineer

Bengaluru, India | June 2021 – Present

- Designed and implemented ETL pipelines using PySpark and AWS Glue, optimizing data ingestion from diverse sources (S3, APIs, and databases).
- Developed real-time data streaming pipelines using Kafka and Spark Structured Streaming, enabling event-driven analytics.
- Built automated Teradata access management solutions with Python for role synchronization and stale ID cleanup.
- Migrated legacy cron jobs to AWS UAC (Unified Automation Console), improving reliability and centralized monitoring.
- Automated Tableau to Power BI migration with reusable scripts and seamless data model integration.
- Collaborated with DevOps teams to design serverless data workflows with AWS Lambda and Step Functions.

- Reduced pipeline execution time by 40% via query optimization, partitioning, and parallel job orchestration.

PROJECTS

Sentimatix — AI-Powered Financial Intelligence Platform

GitHub: <https://github.com/Rishav0123/sentimatix.git>

Site Link: <https://sentimatix.onrender.com/>

AI-Powered Financial Intelligence Platform Developed an end-to-end financial analytics platform connecting a React/TypeScript frontend with a FastAPI backend. Implemented the Model Context Protocol (MCP) to create an agentic workflow that orchestrates RAG (Retrieval-Augmented Generation) searches, real-time news scraping, and LLM-based sentiment analysis. Deployed to production on Render with a Dockerized CI/CD pipeline, solving complex dependency challenges to deliver scalable, context-aware investment insights.

Option 3: Technical Deep-Dive (Metric/Scale Implied)

- **Agentic AI Architecture:** Designed a modular MCP Server to decouple AI logic from tool execution, enabling dynamic tool calling for stock analysis.
- **Data Pipeline & Search:** Built a scalable ingestion pipeline using PySpark and Supabase (pgvector), creating a semantic search engine for financial news.
- **Modern Interactive UI:** Delivered a responsive "Glassmorphism" dashboard using React 18 and Tailwind CSS with real-time WebSocket updates.

Event Management System

- Designed backend architecture using Supabase for database and authentication.
- Developed Edge Functions for real-time triggers, user actions, and API requests.
- Built REST APIs for event creation, registration, and notifications with role-based access control.

Tech Stack: Supabase, JavaScript, PostgreSQL, Edge Functions, REST APIs

EDUCATION

B.Tech/B.E. — National Institute of Technology, Durgapur (2017 – 2021)

COURSES & CERTIFICATIONS

- AWS Certified Data Engineer – Associate (DEA-C01)
- Big Data Processing with PySpark — Coursera
- Apache Kafka Fundamentals — Confluent
- Power BI for Data Engineers — Udemy