

RISHABH SUSHIL JAIN

✉️ rishabhsj23@gmail.com | ☎ +44 7442 640006 | 🌐 London, UK | 💼 linkedin.com/in/rishabhsj23

Data Engineer with around 4 years of experience designing and maintaining scalable data pipelines using Databricks, PySpark, SQL, and AWS. Proven expertise in building CI/CD-driven ETL workflows, data transformations, and production-grade analytics platforms. Databricks Certified Data Engineer Associate with strong stakeholder management and cross-functional collaboration experience. Skilled in Agile methodologies with a track record of delivering high-impact data solutions.

KEY ACHIEVEMENTS:

- Reduced data pipeline processing time by 40% through PySpark optimization and performance tuning
- Achieved zero deployment errors over 12 months through robust CI/CD implementation
- Improved data quality metrics by 35%, directly enhancing business intelligence accuracy across organization

PROFESSIONAL EXPERIENCE:

Data Engineer | STV Group Plc, UK

February 2022 - Present

- Designed and maintained 30+ scalable ETL pipelines on Databricks using PySpark and SQL, processing 15+ million records daily, improving data processing efficiency by 40%.
- Implemented CI/CD pipelines using GitHub Actions to automate deployment workflows, reducing deployment time by 60% and eliminating manual errors.
- Built data transformation and validation frameworks leveraging Delta Lake and Medallion architecture, resulting in 35% improvement in downstream Tableau reporting accuracy.
- Optimized Spark jobs and query performance, reducing cluster costs by 45% while maintaining pipeline SLAs.
- Implemented proactive alerting and monitoring systems using Databricks workflows, reducing mean time to resolution for pipeline failures by 80%.
- Partnered with analytics, BI, and product teams to translate business requirements into scalable technical data solutions.

Associate Developer | RYoshiga Consultancy Ltd, UK

June 2021 - January 2022

- Developed data processing applications using Python and JavaScript to support analytics pipelines for AI and ML projects.
- Built data visualization dashboards and automated data quality checks ensuring data integrity across 5+ client projects.
- Collaborated with clients to gather requirements and deliver data-driven solutions aligned with business objectives.

Content Engineer Intern | Kelkoogroup, UK

June 2020 - May 2021

- Developed Python and XML automation scripts processing 10,000+ merchant data records daily, improving integration efficiency by 30%.
- Supported large-scale Google Shopping product listing ads and sales tracking data pipelines.

TECHNICAL SKILLS:

- **Data Engineering:** Databricks, Apache Spark, PySpark, Delta Lake, ETL/ELT Pipelines, Data Modeling, Data Warehousing, Medallion Architecture, Batch Processing, Data Lakes, Performance Optimization, Query Tuning
- **Cloud & DevOps:** AWS (S3, EC2, Lambda, Athena), Azure, Git, GitHub Actions, CI/CD Automation, Docker, Linux, Shell Scripting
- **Programming & Databases:** Python, SQL, PostgreSQL, MySQL, MongoDB, Java, JavaScript
- **Data Quality & Analytics:** Tableau, Power BI, Data Validation, Unit Testing, Integration Testing, Monitoring & Alerting, Data Profiling, Data Governance
- **Data Integration & Tools:** Airbyte, HubSpot, OneSignal, Postman, REST APIs

MAJOR PROJECTS:

User Authentication using NLP and ML

May 2021 - September 2021

- Built biometric authentication system using facial and speech recognition with Python, TensorFlow, and OpenCV.
- Designed data preprocessing pipelines processing 1+ million audio and image records with feature extraction and transformation workflows.
- Project recognized among top university submissions for technical innovation and implementation quality.

Smart Control System in Automobile

November 2018 - February 2019

- Developed real-time data pipeline processing sensor telemetry using Python and Raspberry Pi for automated vehicle controls with sub-second latency.
- Implemented data-driven obstacle detection and automated braking logic, processing streaming IoT sensor data for safety-critical applications.

Online Examination Portal & College Predictor

February 2018 - May 2018

- Designed normalized MySQL schemas and built SQL-based analytics for exam results and college prediction using statistical algorithms.
- Implemented data integrity controls, transaction management, and query optimization across application workflows.

EDUCATION:

Master of Science - Artificial Intelligence and Robotics | University of Hertfordshire, UK | September 2019 - September 2021

- Key Modules: Practices of Artificial Intelligence, Neural Networks and Machine Learning, Team Research and Data Analysis, AI Systems

Bachelor of Computer Engineering | Savitribai Phule Pune University, India | August 2015 - April 2019

- Key Modules: Artificial Intelligence and Robotics, Machine Learning, Databases and Data Analytics, Web Technology, Computer Networks, Cyber Security

CERTIFICATIONS, TRAINING & ACHIEVEMENTS:

Certifications:

- Databricks Certified Data Engineer Associate (2024)
- Databricks Fundamentals Accreditation
- Microsoft MTA: Security Fundamentals (2018)

Professional Training:

- Machine Learning with Databricks | SQL using PostgreSQL | Git Essential Learning | Artificial Intelligence Foundations: Machine Learning | XML Essential Training | Internship Experience UK

Achievements:

- Judge at Queen Mary University of London for International Greenwich Olympiad (2025)
- Delivered technical presentations on Data Engineering, ML, and Cloud Technologies at industry conferences
- Published research papers on AI and data systems with obtained copyright for technical innovation projects