

SOURAV HALDER

DATA SCIENTIST

+91-8777893442 | halder.sourav1996@gmail.com | linkedin.com/in/sourav-halder | github.com/souravhalder1996

Summary

Data Engineer with 4+ years of experience designing scalable data pipelines and cloud-native solutions, now applying advanced analytics and machine learning to solve business problems. Holds a Master's degree in AI and Robotics, with strong proficiency in Python, SQL, AWS, and Machine Learning. Skilled in transforming complex, multi-source data into model-ready assets, building predictive solutions, and driving strategic decisions through experimentation and statistical analysis. Proven ability to collaborate across engineering, product, and business teams to deliver measurable impact in fast-paced environments.

Skills

Programming & Data Engineering: C, Python, SQL, Databricks, ETL/ELT Pipelines, Git

Machine Learning & AI: Scikit-learn, TensorFlow, Keras, XGBoost, MLflow, Feature Engineering, Model Deployment

Cloud & Deployment: AWS (S3, EC2, Lambda, Glue, Redshift, SageMaker), Docker, CI/CD

Analytics & Visualization: Pandas, NumPy, Matplotlib, Seaborn, SQL Analytics, Statistics

Experience

Associate Consultant

October 2021 – Present

Infosys Limited

- Developed and maintained 50+ production ETL pipelines using Python, SQL, and AWS Glue, orchestrating data extraction from multi source SAP ERP, SQL Server, and REST APIs into Redshift data warehouse supporting 14 data marts, ensuring timely data availability for business dashboards and analytics teams.
- Implemented robust pipeline orchestration with event-driven and schedule-based triggers (Lambda, EventBridge, S3), incorporating automated retry logic, data validation checks, eliminated synchronization failures by 85%, resolved data mismatches, and achieved 99.5% pipeline reliability.
- Optimized data delivery cycles to support monthly Sales & Operations Planning (S&OP) processes for leadership and regional sales heads, cutting data delays by 60% and ensuring accurate, timely insights for strategic sales planning and forecasting.
- Re-architected data extraction pipeline by migrating from direct SQL database to Apache Iceberg-based live layer with incremental delta loads, achieving 93.75% reduction in data extraction and eliminating performance bottlenecks caused by high-volume ERP replication traffic.
- Automated incident management by integrating ServiceNow API for pipeline monitoring, enabling automatic ticket creation and intelligent team assignment for job failures—decreased mean time to resolution (MTTR) by 40% and eliminated manual triaging overhead.

Education

M.Tech in Intelligent Automation and Robotics

August 2018 – July 2021

Jadavpur University

Grade: 90.36%

B.Tech in Electrical Engineering

August 2013 – July 2017

Brainware Group of Institutions

CGPA: 8.49/10

Projects

Network Security System | *Python, MLOps, ML Engineering, Cloud Infrastructure*

[GitHub](#) | [Live Demo](#)

- Engineered 4-stage ML pipeline achieving 96.2% accuracy on 100K samples with 99.8% data quality using Python, Scikit-learn, XGBoost, GridSearchCV (5-fold CV)
- Implemented MLflow tracking (150+ experiments), feature engineering via sklearn ColumnTransformer, and automated train-test stratification (80-20)
- Built GitHub Actions CI/CD automating Docker containerization, AWS ECR/EC2 deployment with auto-scaling policies
- Developed FastAPI REST microservice (<85ms inference latency), AWS S3 versioning (150+ backups), MongoDB TLS encryption (1M+ records)
- Reduced debugging time by 60% with custom exception handling, structured logging (4K+ production code), and hierarchical error propagation

Sales Dashboard for Regional Performance | *SQL, Power BI*

[GitHub](#) | [Live Demo](#)

- Queried and joined 3 sales-related tables covering over 25,000 rows across 12 months and 5 regions using optimized SQL joins to create consistent, analysis-ready datasets
- Built 10 calculated measures and KPIs in Power BI to track regional sales, return trends, and product performance
- Designed multipage dashboards with slicers and filters, helping stakeholders drill down by region, product category, and time
- Detected a 14% sales drop post-festival in the South region—insight was used for a marketing recovery plan

Awards & Certifications

- AWS Certified Solutions Architect – Associate:** [View Credential](#)
- AWS Certified Developer – Associate:** [View Credential](#)
- Databricks Certified Data Engineer Associate:** [View Credential](#)