Adwiteey Mauriya
ML Engineering & Gen AI | LLM | Fintech

✉ send@amauriya.com

📞 +44 2032-862-669, +971 58-569-1933

🏴 Portuguese, UAE golden visa

🌐 www.amauriya.com

Interview me

I have 10 years of experience in machine learning and 7 years in big data engineering, 4 years of experience in LLM, specialising in scalable data pipelines, automation, and AI-driven insights. I've worked with Snowflake Cloud, Apache Airflow, and Python to process large datasets efficiently. Recently, I've focused on generative AI and LLMs in fintech, optimizing financial analytics and risk assessment.

Work Experience

Founder (Self Employed) , Flaredog, Zurich, Switzerland   *Mar/2024 - Feb/2025*

- Launched Flaredog, a business solutions company specializing in AI/ML infrastructure.
- Developed and deployed 5+ AI/ML solutions across finance, healthcare, climate, and blockchain, leading to an average 30% efficiency boost in client workflows.
- Led business development initiatives, client acquisition, and cross-functional project delivery.
- Implemented cost-efficient ML infrastructure, reducing cloud computing expenses by 25% while improving model inference speed by 50%.
- **Tech stacks: Machine learning, AWS, Large Language models(LLM), Langchain, Databricks, Python, PySpark, Torch, Langchain, Snowflake, BigQuery, Clickhouse**

Senior Data Science Engineer, Liquidity Capital Inc., Abu Dhabi, UAE   *May/2023 - Feb/2024*

- Developed and deployed a financial chatbot powered by LLMs & LangChain, handling 5,000+ document-based queries monthly, reducing manual financial analysis time by 40%.
- Optimised probability of default (PD) model, increasing prediction accuracy by 15%, enabling data-driven risk assessments and enhancing underwriting efficiency by 30%.
- Built a centralised feature store, consolidating 100+ financial indicators from structured & unstructured data, improving model retraining efficiency by 50%.
- Automated active vs. out-of-business classification, increasing classification accuracy by 92%, reducing manual verification efforts by 60%.
- Deployed a BI tool on EKS, enabling real-time insights for 50+ stakeholders, reducing report generation time by 35% and enhancing decision-making efficiency.
- Enhanced model inference speeds by 3X through optimization in Databricks & Torch, cutting down compute costs by 25%.
- Collaborated with cross-functional teams, integrating AI-driven insights that led to a 20% improvement in investment decision accuracy.
- **Tech stacks: Machine learning, Natural language processing(NLP), Large Language models(LLM), Langchain, Databricks, Python, PySpark, Torch, Langchain, BigQuery, Snowflake, Clickhouse**

Senior Data Engineer, Liquidity Capital Inc., Abu Dhabi, UAE   *Aug/2022 - May/2023*

- Led the development of a bank account aggregator pipeline, processing 2M+ transactions monthly, improving data ingestion efficiency by 90%.
- Automated ETL workflows, reducing data processing time from 16 hours to 10 minutes, enhancing reporting speeds by 80%.
- Implemented an ML-driven anomaly detection system, cutting down fraudulent transaction alerts by 50%, increasing accuracy in risk assessment.
- Designed a CI/CD pipeline for ETL repositories, reducing deployment time by 60% and minimising errors in production releases.
- Managed AWS EC2 instances & AWS Lambda functions, reducing infrastructure costs by 30% through auto-scaling optimizations.
- Developed and maintained Delta Lake architecture, enabling 99.9% data availability across financial datasets.
- Optimized Spark-based transformations, achieving 2.5X faster query execution, reducing cloud computing costs by 20%.
- Established backend code review best practices, improving deployment efficiency and reducing rollback incidents by 60%.
- **Tech stacks: Airflow, Dagster, Prefect, PySpark, Databricks, Deltalake, AWS, AWS Lambda, PySpark, Snowflake, BigQuery, Step Functions, Python, SQL, Selenium, openAI, Large Language Models(LLM), Langchain, Clickhouse.**

Data Engineer, Tripadvisor Inc. Lisbon, Portugal   *Nov/2020 - Jul/2022*

- Maintained and monitored datasets across 3+ major cloud platforms (GCP, Snowflake, Hadoop), ensuring 99.98% data availability for analytics and reporting.
- Optimised ETL pipelines, reducing data processing time by 50%, which improved dashboard refresh speeds for business intelligence teams.

- Implemented data compression techniques, lowering storage costs by 40% and contributing to $10M+ annual cost savings.
- Designed and built 30+ new data pipelines, automating ingestion and transformation for 200M+ records per day.
- Improved query performance by 3x in Snowflake and BigQuery, enabling near real-time analytics for marketing and product teams.
- Led the migration of legacy ETL workflows to Airflow, reducing manual interventions by 80%, increasing job success rates to 99.5%.
- Enhanced data quality checks, reducing data discrepancies by 90%, improving reporting accuracy for finance and customer teams.
- Championed best practices in code reviews, leading to a 50% reduction in deployment errors and improved team efficiency.
- Trained and mentored 2+ junior data engineers, accelerating their onboarding process and reducing ramp-up time by 30%.
- **Tech Stack: Python, Scala, SQL, Airflow, AWS, GCP, Snowflake, Postgres, PySpark, Hive, BigQuery, Clickhouse**

Business data analyst, <u>Netjets Inc.</u>, Oeiras, Portugal   *Oct/2019 - Jul/2020*

- Automated pilot training data warehouse, reducing manual data entry workload by 70%, improving accuracy, and increasing operational efficiency.
- Developed 20+ dynamic reports and dashboards using Tableau, Power BI, and SSRS, enabling executives to make data-driven decisions in real time.
- Increased data processing speed by 3x, improving insights delivery for training performance analysis.
- Designed KPI tracking framework, helping the company optimise pilot training costs by 15% while maintaining compliance with aviation standards.
- Achieved 110% of the targeted qualified leads from strategic training workshops, exceeding conversion expectations.
- Implemented data validation workflows, reducing data inconsistencies by 90%, ensuring high-quality analytics.
- Improved gross margin tracking, leading to a 5% increase in profitability for the training department by identifying inefficiencies.
- Optimized SQL queries and ETL pipelines, reducing report generation time by 50%, enabling faster decision-making for leadership.
- **Tech stacks: Python, SQL, PySpark, Snowflake, AWS, Tableau, SSRS, Excel, Power BI, SSIS.**

Researcher, <u>Culham Center for Fusion Energy</u>, Oxfordshire, UK   *Aug/2016 - Sep/2018*

- Developed and optimized numerical tools to simulate plasma behavior in nuclear fusion reactors, improving prediction accuracy by 35% and reducing computational time by 40% on high-performance clusters.
- Scaled large-scale simulations on supercomputers with 100,000+ CPU cores and GPU nodes, enhancing processing efficiency 3x while reducing energy consumption by 20% through optimized resource allocation.
- Designed an automated pipeline for importing and analyzing fusion experiment data, reducing manual processing time by 60% and improving data throughput by 50% using advanced compression techniques.
- Reduced computational errors by 30% by implementing improved numerical solvers and stability corrections in simulation codes.
- **Expertise required: Python, PySpark, Databricks, Scala, Mathematics, Statistics, Programming, Distributed computing, High performance computing, OpenMP, MPI, Slurm, Fortran, Physics**

Education

<u>Instituto Superior Tecnico</u>, Lisbon, Portugal. <u>University of Oxford</u>, U.K   *Jan/2015 - Sep/2020*

- Ph.D. in Technological Physics Engineering.
- Title: Effect of impurities in intrinsic rotation of tokamak plasmas.

<u>IISER Pune</u>, India   *Aug/2009 - Jun/2014*
- B.S. - M.S Dual Degree.
- Physics(Major), Mathematics(Minor).

Technical Skills

- **Big data**: Hadoop, Spark, Databricks
- **ETL**: Airflow, Dagster, dbt, N8N
- **Database**: Snowflake, BigQuery, Postgres, Mysql, Clickhouse
- **Devops**: Kubernetes, Docker, EKS, Jenkins, Gitlab CI/CD
- **ML**: Scikit-learn, Keras, Torch, Tensorflow, MLflow, Langchain
- **BI**: PowerBI, Tableau, SRSS, Apache Superset
- **Dev**: Intellij, Vscode, Git, SVN
- **Cloud**: GCP, AWS, Supercomputers
- **PM**: Microsoft 365, Jira, Confluence

Soft Skills
- Team Player
- Problem Solver
- Public Speaking
- Presentation

Programming languages
- Advanced: Python, Scala, SQL, Typescript
- Intermediate: Kotlin, C++, C

Languages
- English
- Portuguese
- Hindi

*Underlined texts are hyperlinks