

Bhaskar Sinha

Data Engineer

Glasgow, United Kingdom | +44 7747194485 | bhaskarsinhauk@gmail.com | [linkedin.com/in/bhaskarsinha4297/](https://www.linkedin.com/in/bhaskarsinha4297/)

SUMMARY

I'm a data engineer who loves turning messy data into clear, useful insights. Over 4 years, I've built cloud-based pipelines, improved data flows and worked with teams to make sure the right data is ready when it's needed. Recently explored the world of Generative AI by developing a personal AI chatbot. Enjoys combining data engineering expertise with modern AI tools to create impactful and real-world solutions. Always curious to learn and apply new technologies to solve problems.

PROFESSIONAL EXPERIENCE

Panigrahi's Civil Engineering Foundation - Durg, India

Associate Analyst (Data Analyst)

Feb 2020 - Feb 2023

- Designed and automated scalable ETL pipelines using Python and MySQL, enhancing data processing efficiency by 15% and ensuring reliable, accessible data for stakeholders.
- Developed and maintained interactive Power BI dashboards, enabling data-driven decisions that improved student performance tracking by 30%.
- Implemented robust data validation workflows and quality checks using SQL, reducing reporting errors by 40% and supporting accurate insights.
- Collaborated with educators and stakeholders to understand requirements, delivering tailored data solutions that boosted student engagement by 15%.

Pylons Technology PVT. LTD. - Durg, India

Jan 2019 - Feb 2020

Software Engineer

- Contributed to development of CoreFab ERP and Institute ERP System using Node.js, Express.js, and MySQL, enhancing API performance by 35% and supporting over 500 daily users.
- Set up CI/CD pipelines using Jenkins, which sped up our deployments by 30% and made it easier to release updates regularly as part of our Agile (Scrum) process.
- Collaborated with QA, analysts and end-users to implement fabrication workflow modules, reducing manual data entry by 25%.
- Participated in user training, bug-fixing and performance tuning, resolving issues to improve system stability.

SKILLS

Technical Proficiency: Python | MySQL | ETL processes | data pipeline development | data validation & quality assurance | relational databases | data warehousing | Data Modeling | Unix Scripting | Azure (Data Lake, Data Factory, Databricks) | AWS (Athena, Lambda, S3, Elastic Beanstalk, RDS) | Application Integration | Data Migration | CI/CD (Jenkins)

Tools & Frameworks: Pyspark | Hadoop | SQL/NoSQL databases | PostgreSQL | Apache Airflow | Docker | Kafka | Kubernetes (basic) | GitHub | Power BI | Snowflake (Basics) | SAP CRM

Soft skills: Cross-Functional Collaboration | Attention to detail | Stakeholder Engagement | Problem Solving | Data Storytelling | Continuous Learning | Compliance Awareness | Project Management

EDUCATION

MSc in Data Analytics, UK

Sep 2023 - Sep 2024

The University of Strathclyde, Grade: 2:1

- **Key Modules:** Advanced Machine Learning, Statistical Analysis, Predictive Modelling, Power BI
- **Spotify Classification Problem:** Forecasted song popularity and genre on Spotify using ML on 13 features, achieved 41.17% accuracy in a Kaggle competition
- **Ersan Duran (IT Consultant):** Predicted Bitcoin price using time series analysis and regression, focusing on data exploration, preprocessing, and model tuning with thorough justification
- **Experian-Reverse Engineering:** Worked with Experian to predict salaries from net to gross, considering tax bracket calculation, National Insurance, and other deductions.

PROJECTS

- [AI Agent Chatbot with FastAPI](#) : Designed and implemented a multi-provider LLM chatbot using FastAPI, Streamlit, and LangGraph with support for various models. Integrated real-time search along with external tools for robustness. Containerized with Docker and adopted DevOps practices with CI/CD pipelines ensuring scalable, reliable and production-ready GenAI applications.
- [Overpass Microservice Architecture Design \(Freelance 2024\)](#) : Built and deployed a scalable geospatial API on AWS Elastic Beanstalk with PostgreSQL/PostGIS, serving highway data extracted from OpenStreetMap.
- [Tokyo Olympic \(2021\) Data Analytics](#) : Built end-to-end pipelines with Azure Data Factory and Databricks to process Olympic datasets and enable predictive analytics on event outcomes.
- [Kafka-Based Stock Market Streaming and Analytics Pipeline](#): Built a real-time data pipeline using Apache Kafka on AWS EC2, streaming stock market data into S3, and enabling analytics via Glue & Athena.
- [Airflow-Orchestrated ETL for Twitter Data on AWS](#): Automated daily scheduling with Airflow DAGs including retries, logging, and monitoring, ensuring reliable and fault-tolerant execution.

PUBLICATIONS

[Object Detection and Optimization of Neural Networks](#)

May 2022

International Journal of All Research Education and Scientific Methods (IJARESM), Vol. 10, Issue 5, ISSN: 2455-6211.

- Developed real-time human detection using TensorFlow, OpenCV, and multiprocessing on embedded systems, optimizing inference FPS with GPU acceleration and efficient model selection.

CERTIFICATIONS

- Big data bootcamp 2.0 (11 months) from ineuron, India
- AWS Certified: AWS Cloud Practitioner (in progress)
- Advanced Google Analytics