

SHREEHARI ANBAZHAGAN

Data & Cloud Engineer | Backend Developer | Applied AI

Birmingham, United Kingdom | +44 7721 046613 | shreeharianbazzhagan@gmail.com

linkedin.com/in/a-shreehari | github.com/ShreehariA | shreeharia.github.io | leetcode.com/u/shreeharianbazzhagan

SUMMARY

Data and Cloud engineer, backend developer and applied-AI builder with 3 years of industry experience, shipping cloud infrastructure, data pipelines, backend services and the AI on top, end to end. Built enterprise ETL pipelines at Comcast (50% cost cut, 60% fewer failures), real-time ML on 17.2M records at MInt Labs (IIT Madras), and a complete UK cloud traceability platform at Kahawa Roots. Completing a Joint MSc in Data Science and AI (University of Birmingham and IIT Madras).

WORK EXPERIENCE

Traceability Platform Developer Intern

May 2026 - Jul 2026

Kahawa Roots (UK) Ltd

United Kingdom

- Designed and delivered an enterprise coffee traceability platform (Kenyan farms to UK buyers and consumers) end to end: system architecture, workflows and data flow.
- Built REST APIs in Python/FastAPI on a PostgreSQL schema, with data modelling and migrations for lot tracking, record correction and approval workflows.
- Developed the Next.js frontend, including customer-facing QR traceability journeys.
- Provisioned the full AWS stack (ECS, RDS, S3, Cognito) as Terraform infrastructure-as-code with GitHub Actions CI/CD and OIDC; implemented Cognito authentication, RBAC, audit logging and secrets management.
- Ran regression and deployment testing, production monitoring and incident debugging; acted as technical lead with stakeholders and mentored a fellow intern, recognised in a reference letter from the Head of Sustainability and Traceability.

Winter Project Intern (Machine Learning)

Nov 2025 - Jan 2026

MInt Labs, IIT Madras

Chennai, India

- Built a real-time bus ETA prediction system for the IITM campus shuttle from streaming GPS data: 17.2M records, 53 spatial-temporal features, and a GNN + XGBoost + LightGBM ensemble served via FastAPI and Docker.
- Delivered 1.88 minute MAE with sub-millisecond inference and no downtime across the evaluation period; built validation, monitoring and fallback strategies for robust performance under noisy GPS data.

Data Engineer (Development Engineer 1)

Jul 2023 - Jun 2025

Comcast, Data Engineering

Chennai, India

- Built and optimised enterprise ETL pipelines with Kafka, Redshift and PySpark, ingesting sources such as Workday and LDAP into the enterprise data warehouse.
- Designed automated data reconciliation frameworks validating source-to-warehouse integrity, improving reporting accuracy by 40%.
- Migrated legacy reporting jobs to modern pipelines, reducing failures and execution time by 60%.
- Reduced infrastructure costs by 50% through warehouse consolidation and optimised snapshot and storage strategies.
- Built EnvSyncer for cross-environment data sync, plus an anomaly-detection proof of concept (AWS Lookout for Metrics) wired into CloudWatch dashboards and Slack alerting.
- Led knowledge transfer and onboarding documentation; supported the Service Reliability Engineering team on deployment strategy through testing and performance tuning.

Data Engineering Intern

Jan 2023 - Jun 2023

Comcast

Chennai, India

- Developed a Slack monitoring system for ETL runs and job status, providing real-time alerts and daily cost breakdowns.
- Contributed to the enterprise data warehouse application, a single source of truth consolidating multiple data sources.

EDUCATION

Joint MSc in Data Science and Artificial Intelligence (Expected 2027)

Jul 2025 - Present

University of Birmingham and Indian Institute of Technology Madras

Birmingham, UK and Chennai, India

- Coursework: Machine Learning, Data Analytics, Natural Language Processing, Data Visualisation, Data Structures and Algorithms for Data Science, Mathematics for Data Science.

Bachelor of Technology in Computer Science and Engineering

Jul 2019 - May 2023

SASTRA Deemed University

Thanjavur, India

- First Class with Distinction; core computer science foundation in data structures, algorithms, DBMS and operating systems.

PROJECTS

DeltaEd - Growth-Based Student Leaderboard and AI Coach (github.com/ShreehariA/Google-edu-hackathon) *Google Hackathon*

- Built a RAG-powered AI coach that ranks students by how much they personally improve rather than by absolute standing, deployed on Google Cloud Run.
- Owned the leaderboard and the retrieval-augmented coaching loop that walks each student through their gaps; won Best Pitch and Best Use of Google Tools.

Sentiment Analysis and Topic Modeling for Tweets (github.com/ShreehariA/G20_Library_Summit) *2024*

- Applied pysentimiento and cardiffnlp NLP models to tweet sentiment and topics (94.53% mean probability); built engagement metrics and heat-map visualisations; presented at the G20 Library Summit.

Diabetes Prediction Web App (github.com/ShreehariA/diabetes-pred) *2022*

- Built a Django app comparing six ML models to predict diabetes risk (Random Forest, 0.87 accuracy), generating downloadable PDF reports and SMS notifications.

VAC - Video As Code (github.com/ShreehariA/vac) *Personal project*

- Building a purely algorithmic compiler/decompiler between video and human-readable source code (pixels to code, measure-don't-classify).

Other: ACEest Fitness (Flask, Docker, CI/CD); IITM MLP Project (Kaggle, 87% accuracy); Metric-Learning Data Challenge (Kaggle); Helper Libraries (Python/AWS); IITM Business Data Management. All at github.com/ShreehariA.

TECHNICAL SKILLS

- **Languages:** Python, SQL
- **Backend:** FastAPI, Django, Flask, REST APIs, SQLAlchemy, boto3, Pandas
- **Frontend:** Next.js, React
- **Cloud:** AWS (ECS, RDS, S3, Cognito, Lambda, Glue, Athena, Redshift, IAM, CloudWatch, Secrets Manager, Lookout for Metrics), GCP (Cloud Run)
- **DevOps and IaC:** Terraform, Docker, GitHub Actions, Jenkins, CI/CD, OIDC, Git, Linux, Bash
- **Data Engineering:** ETL/ELT, PySpark, Kafka, Data Warehousing, Data Modelling, PostgreSQL, Redshift, PyArrow
- **AI/ML:** Machine Learning, XGBoost, LightGBM, Graph Neural Networks, RAG, NLP
- **Security and Testing:** IAM, RBAC, authentication, audit logging, secrets management, Pytest, regression and deployment testing

AWARDS

- Best Pitch and Best Use of Google Tools, Google Hackathon (DeltaEd).
- Champ Award, Comcast (Q4 2023 and Q2 2024), for consecutive outstanding performance.
- Good Presentation Award, G20 Library Summit (Feb 2024).
- First Class with Distinction, B.Tech (SASTRA).

PROFESSIONAL DEVELOPMENT

- Presented sentiment analysis and topic modelling on X/Twitter data at the G20 Library Summit (Feb 2024).
- Presented data engineering solutions at Comcast (2024): performance optimisations and key insights.
- Organised a SQL Workshop Fundraiser for Bhumi (Sep 2024), raising funds to support shelter children.
- Attended the Responsible and Safe AI Workshop at IIT Madras (Nov 2023): legal bias, AI attacks and unlearning (CeRAI).
- Attended AWS User Group Chennai on GenAI and serverless (Sep 2024); IIT E-Summit venture-capital talks (Mar 2024); AWS Awesome Day on ML, serverless and big-data analytics (2023).
- Completed a three-day hands-on Web Development Workshop (Django/Python), SASTRA (Aug 2022).

VOLUNTEER EXPERIENCE

Centre Coordinator and LEM Volunteer

Bhumi

Jul 2024 - Present

Chennai, India

- Organised workshops, events and additional classes for foster children; ran sports events to raise funds.
- Taught mathematics to shelter-home children via the Ignite Little Einstein Mathematics program, using interactive hands-on methods.

NSS Leader (2022-24), Volunteer (2019-22)

SASTRA National Service Scheme (NSS)

Aug 2019 - Jun 2024

Thanjavur, India

- Increased annual blood-donation camps and led crowd control and festival management.
- Strengthened community ties through hospital support, temple ethics drives and old-age-home cleanups.

Work authorisation: open to roles across the UK, EU and Dubai (remote-friendly); UK Skilled Worker sponsorship required for long-term UK positions.