

Dr Shashidhar Ameenpur

Data & Research Engineer

London, United Kingdom | [LinkedIn](#)

Data & Research Engineer with an MSc in Health Data Science and Microsoft Fabric certification, specializing in clinical data pipelines, wearables analytics, and signal processing.

EXPERIENCE

Data Engineering Apprenticeship

Feb 2025 — Aug 2026

Cambridge Spark

London

Developed scalable data architecture, streaming systems, and cloud-native engineering capability through applied delivery. Built ingestion frameworks, data lakes, real-time pipelines, and automated orchestration workflows. Applied Python, SQL, Spark, and Docker to production-oriented data engineering projects.

Clinical Data Analyst

Feb 2022 — Present

South London and Maudsley NHS Foundation Trust

London, England, United Kingdom

Built clinical datasets and automated analytics pipelines using SQL, Python, and Azure services. Designed and deployed an end-to-end automated report generation pipeline for Empatica E4 wearable data—processing accelerometer, BVP, EDA, HR, and temperature signals from AWS S3, applying signal cleaning and feature extraction in Python, orchestrating via Luigi, and generating patient-level clinical reports in R. Reduced per-patient processing from hours of manual work to a fully automated workflow.

Sr. Content Lead - Payment Policy & EMR Data

Oct 2014 — Feb 2020

Cotiviti

Hyderabad, Telangana, India

Maintained clinical payment rules and developed new rules per regulatory guidelines and client requests. Created and managed SQL Server tables for EMR health data across EPIC and MedTech systems. Ran stored procedures and updated tables to maintain data quality across large-scale clinical datasets. Collaborated with clinical and engineering stakeholders to translate policy requirements into data-layer logic.

HCC Medical Coding Specialist

Apr 2012 — Jun 2013

UnitedHealth Group

London

Performed Hierarchical Condition Category (HCC) medical coding to support risk adjustment and CMS reimbursement accuracy. Reviewed clinical documentation to assign accurate ICD-10 codes aligned with HCC model requirements. Supported quality audits and coding reconciliation processes to ensure compliance with CMS guidelines.

EDUCATION

Cambridge Spark

Feb 2025 — Aug 2026

Data Engineer Level 5

Distinction

Swansea University

Sep 2021 — Jan 2022

Msc Health data science in Data Science

Health Data Science, R, Python, SQL

Rajiv Gandhi University of Health Sciences

Jan 2004 — Jan 2009

bachelor of ayurveda medicine and surgery in Ayurvedic Medicine/Ayurveda

PROJECTS

End-to-end wearable data processing and clinical reporting system, NHS Automated Report Generation Pipeline

Designed and deployed automated pipeline processing Empatica E4 wearable data (accelerometer, BVP, EDA, HR, temperature) from AWS S3. Applied signal cleaning and feature extraction using Python (pandas, scipy), orchestrated task workflows via Luigi, and generated patient-level PDF reports using R and RHRV for HRV analysis. Implemented PHI redaction layer using Presidio and NuExtract with AWS Secrets Manager for security.

Cloud-native data ingestion and processing framework, AWS S3 Data Engineering Pipeline

Built scalable ingestion framework and data pipeline leveraging AWS S3 for storage and Python/Spark for processing. Implemented automated orchestration and validation workflows as part of Cambridge Spark apprenticeship.

Signal processing and feature extraction from wearable devices, Time-Series Feature Engineering - Empatica Wearable Data

Developed feature extraction pipeline for physiological time-series data from Empatica E4 wearables. Applied signal processing techniques to clean raw sensor data and generate clinical-grade features for downstream analytics and research applications.

On-device sensitive data masking for clinical documents, Local PHI Redaction Tool

Built local PHI redaction tool combining Presidio, OCR, and NuExtract (via Ollama) to identify and mask sensitive information in clinical documents. Designed for on-device processing to maintain data privacy without external API calls.

Batch extraction and processing of structured data from presentations, Automated Multi-File PowerPoint Table Extractor

Developed Python utility to extract tables from multiple PowerPoint files, enabling automated data consolidation and transformation from presentation formats into structured datasets.

CERTIFICATIONS

AWS AI Practitioner Challenge, Udacity	Mar 2026
Associate Data Engineer in Snowflake, DataCamp	Nov 2025
Collaboration, Marketplace & Cost Estimation Workshop, Snowflake	Oct 2025
Hands-On Essentials: Data Engineering Workshop, Snowflake	Oct 2025
Hands-On Essentials: Data Warehousing Workshop, Snowflake	Oct 2025
Generative AI: The Evolution of Thoughtful Online Search, LinkedIn	Jun 2023
Microsoft Certified: Azure AI Fundamentals, Microsoft	Jun 2021
Microsoft Certified: Azure Data Fundamentals, Microsoft	Jun 2021
Microsoft Certified: Azure Fundamentals, Microsoft	Jun 2021
Microsoft Excel (Microsoft 365 Apps and Office 2019), Microsoft	Jun 2021
SAS Programming 1: Essentials, SAS	Jun 2021
Data Fluency: Exploring and Describing Data, LinkedIn	Jun 2020
Learning Data Analytics, LinkedIn	Jun 2020
Microsoft Certified: Fabric Data Engineer Associate, Microsoft	

SKILLS

Data Engineering · Data Pipelines · Extract, Transform, Load (ETL) · Data Governance · Data Integration · Ingest and transform data · Implement and manage an analytics solution · Monitor and optimize an analytics solution · R (Programming Language) · Transact-SQL (T-SQL) · SQL · Microsoft Power BI · Electronic Medical Record (EMR) · Microsoft Azure · Data Analysis · Python (Programming Language) · Healthcare Management · Healthcare Consulting · Medical Coding · Healthcare · Revenue Cycle · Healthcare Industry · ICD-9 · Medical Billing · Medicaid · Working with Physicians · ICD-10 · Data Analytics · Search Engine Technology · Generative AI · PySpark · Spark · AWS S3 · Secrets Manager · Snowflake · ELT · Data Lakes · Luigi · Docker · Data Validation · Warehousing · Time-Series Analysis · Wearables · HRV Analysis · Signal Processing · Clinical Data · EPIC · MedTech · HCC Coding · CMS Policy · Presidio · NuExtract · Ollama · pandas · scipy · RHRV · Data Architecture · Orchestration · Cloud Services