



ASAD IKRAM

Full Stack Data Engineer · Chevening Scholar · 6+ Years Experience



asad.ikram53@gmail.com
Lahore, Pakistan
asad-ikram-mc.github.io/portfolio



+44 7482 376417
linkedin.com/in/asad-ikram98



Chevening Scholar 2024–2025 — Top 2–3% of 70,000+ applicants · 160 countries · Fully funded by UK FCDO



PROFILE

Senior Data Engineer · 6+ years building cloud-native ETL/ELT pipelines, scalable data warehouses, and adversarial web scraping platforms at enterprise scale.



WEB SCRAPING

Frameworks

- Scrapy, Scrapy Cluster, Scrapyd
- Playwright, Puppeteer, Selenium
- BeautifulSoup, Splash, Apify
- curl_cffi, httpx, requests

Anti-Bot & Evasion

- TLS / JA3 / JA4 fingerprint spoofing
- curl_cffi, cycle-tls, GoLogin
- undetected-chromedriver
- playwright-stealth, rebrowser-patches
- 2Captcha, puzzle solving, proxy rotation
- Akamai, Cloudflare, Datadome
- PerimeterX, Kasada, F5 Shape Security

API Recon

- Charles Proxy, Burp Suite, HTTP Toolkit
- Mobile emulator MITM (Android Studio)
- Hidden REST / GraphQL discovery
- JS bundle deobfuscation



DATA ENGINEERING

ETL / ELT

- PySpark, AWS Glue, DBT, Matillion
- Apache Spark, EMR Serverless

Orchestration

- Apache Airflow, Step Functions
- EventBridge, GitHub Actions CI/CD

Warehousing

- Redshift, Snowflake, BigQuery
- Hive, Hadoop · Star / Snowflake Schema
- RDL → ODL → ADL layered architecture

Streaming

- Kinesis, Kafka, SNS, SQS



CLOUD &

PROFESSIONAL EXPERIENCE

Data Engineer Consultant

M+C Saatchi Fluency · UK · Remote

Dec 2024 – Present

- Architected **end-to-end AWS data pipeline** (GitHub Actions → ECR → ECS Fargate → EventBridge → Step Functions → S3 → AWS Glue → Athena) from zero to production in under 30 days.
- Delivered **ETL/ELT pipelines** and **SageMaker NLP sentiment analysis** powering campaigns for Amazon, Nike, Ford, Reckitt and the UK Government.
- Eliminated **100% of manual data collection** via Scrapy with CAPTCHA bypass, reverse-engineering, proxy rotation and TLS fingerprint evasion.
- Reduced infrastructure costs **35%** by designing auto-scaling ECS container architecture replacing legacy VMs.
- Built real-time CloudWatch monitoring and Slack alerting with Retool data freshness dashboards; maintained 99%+ pipeline uptime.

Founder & Lead Data Engineer

ArtemisAI · UK · Remote

Feb 2025 – Present

- Built a **Meta social data lakehouse** from scratch (GitHub Actions → ECR → ECS Fargate → DynamoDB → S3), ingesting 30+ brand pages daily at scale.
- Designed **layered Redshift data warehouse** (RDL → ODL → ADL) with Step Functions orchestration; enabled sub-second KPI reporting via QuickSight.
- Deployed ECS-hosted **ML sentiment-tagging** and topic-labeling microservices; packaged ADL outputs as B2B subscription datasets.
- Trained **custom LLMs** on engagement data; predictive trend models improved client content monetization by **40%**.
- Implemented **self-healing infrastructure**: EventBridge dead-letter queues, CloudWatch alarms and Twilio alerts — zero manual intervention.
- Designed multi-tenant Redshift access layer with **row-level security**; enabled secure B2B data sharing across 10+ client accounts.

Senior Data Engineer (Promoted from Data Engineer → Associate SWE)

Dubizzle Labs — **Dubizzle, Bayut, Zameen, OLX** · Pakistan · Onsite Feb 2022 – Aug 2024

- Led 3-engineer team owning **500+ regional web scrapers** across MENA on EKS — CI/CD via GitHub Actions, Matillion ELT / AWS Glue (S3 → Athena → Redshift).
- Redesigned layered data warehouse (RDL → ODL → ADL) for Propforce; **improved ETL/ELT throughput 60%**, directly driving record revenues.
- Built **competitor intelligence platform** combining web scraping and DWH analytics; modelled Star Schema, Snowflake Schema and Dimensional data warehouse schemas.
- Developed **Elasticsearch** data pipelines for property search; optimised query performance **50%** via custom index mapping and sharding.
- Reverse-engineered** competitor mobile apps via Android Studio emulator MITM (HTTP Toolkit + Charles Proxy) to extract **hidden REST and GraphQL APIs**.
- Introduced unit testing and Scrapy contract validation; cut production scraping incidents **45%**.

Lead Data Engineer

VendueTech · Ireland · Remote

Jun 2025 – Jul 2025 (Contract)

- Led 3-engineer team building **real-time auction data ingestion** using Scrapy, Playwright, AWS Batch and Step Functions; bypassed CAPTCHA and anti-bot stacks.
- Migrated full infrastructure Azure → AWS; implemented **PySpark on EMR Serverless** reducing data latency **40%** across Bronze, Silver and Gold layers.

DEVOPS

AWS

- ECS Fargate, EKS, Batch, Lambda
- S3, Athena, Glue, Redshift Spectrum
- CloudWatch, CloudFormation, ECR, EMR

DevOps

- Docker, Kubernetes, GitHub Actions
- Scrapyd, Scrapy Cluster

DATABASES

- PostgreSQL, MySQL, MS SQL Server
- MongoDB, DynamoDB, Firebase
- Elasticsearch (ELK Stack)

ML / AI

- LangChain, OpenAI API, HuggingFace
- BERT, GPT, LSTM, XGBoost, LightGBM
- RAG Pipelines, Vector Databases
- spaCy, VADER, TextBlob (NLP)
- AWS SageMaker

BI & DASHBOARDS

- Tableau, Power BI, QuickSight
- Sisense, Periscope, Retool, Kibana

EDUCATION

MSc Data Analytics

University of Portsmouth, UK
2024–2025 • Distinction

BSc Computer Sciences

FAST-NUCES, Pakistan • 2017–2022

LANGUAGES

- English (Fluent) • Urdu (Native)

- ▶ Designed idempotent pipeline with DynamoDB deduplication checkpoints ensuring **exactly-once processing**; integrated CloudWatch and SNS monitoring.

Data Engineer Consultant

Fix.com • Canada • Remote (Concurrent)

May 2023 – Present

- ▶ Designed Scrapy + Scrapyd pipelines processing **50M+ data points** from 200+ e-commerce sites; expanded product catalog 50% in 12 months.
- ▶ Optimised distributed Scrapy Cluster on AWS ECS; **cut crawl time 40%** through async I/O tuning, connection pooling and smart retry logic.
- ▶ Implemented residential proxy rotation across Bright Data, Oxylabs and Smartproxy with **CloudWatch burn-rate monitoring** and auto-failover.
- ▶ Introduced Great Expectations data validation; automated schema drift detection, reducing bad data in production by **70%**.

Data Engineer Consultant

Prefero • Italy • Remote

Sep 2024 – Dec 2024

- ▶ Rebuilt legacy scrapers into modern Scrapy architecture with middleware, item pipelines and retry logic; reduced scraper failure rate **60%**.
- ▶ Mentored 2 junior engineers on anti-bot evasion; established code review standards adopted team-wide.

Freelance Data Engineer

CXG • Dubai • Remote

Dec 2025 – Jan 2026

- ▶ Built production crawlers for global luxury fashion brands bypassing **Akamai v3** and **Cloudflare Turnstile**; achieved **99%+** data completeness.

Senior Freelance Data Engineer (Level 2 Seller • 5★ • 300+ Projects)

Fiverr / Independent Consulting • Worldwide

Dec 2018 – Feb 2026

- ▶ Delivered **300+** web scraping and data engineering projects for global clients across 400+ websites; maintained **<1% block rate** through TLS evasion, GoLogin and behavioural mimicry.
- ▶ Bypassed **Akamai, Cloudflare, Datadome, PerimeterX, Kasada** and **F5 Shape Security** for clients across e-commerce, real estate, finance and luxury sectors.
- ▶ Deployed full-stack data products: Scrapy → ECS Fargate → S3 → Redshift DWH → Tableau / QuickSight dashboards for 50+ international clients.

KEY ACHIEVEMENTS

- 🏆 **Chevening Scholar 2024** — Selected from 70,000+ applicants across 160 countries (top ~2–3% globally). Fully funded by UK FCDO.
- 📈 **Promoted Twice in 2 Years** — Associate Software Engineer → Data Engineer → Senior Data Engineer at Dubizzle Labs.
- ★ **300+ Five-Star Projects** — Level 2 Fiverr Seller with 7+ years of delivery across web scraping and data engineering.
- 🛡️ **<1% Block Rate at Scale** — Trusted adversarial scraping expert; built client base through referrals and community contributions.