Jafsal M A

+91-9633062855 | jfzlma@gmail.com | linkedin.com/in/jfzlma | github.com/jfzlma | jafsal.dev | Kerala, India

PROFESSIONAL SUMMARY

Senior Data Engineer and Technical Lead with 7+ years building large-scale web crawling systems and Python ETL pipelines, delivering production-grade ingestion, transformation, and automated delivery at 100Ks-to-millions-of-records scale. Experienced leading data operations teams end-to-end - from crawling through client-format delivery (SAP XML, E4H XML) - and currently moving into agentic AI, building automation agents on the Hermes framework.

WORK EXPERIENCE

PST.AG

Feb 2024 - Present

Python Developer / Team Lead - Data Engineering & Web Scraping

Weimar, Germany (Remote)

- Led design and delivery of Python backend tooling for a scraping ecosystem of 100+ production projects, improving ingestion throughput and downstream data readiness for analytics
- Led the Data Operations team for one year, owning the full ETL project lifecycle - crawling, transformation, validation, and automated delivery - with outputs delivered to clients in SAP XML, E4H XML, and other structured formats
- Built and scaled ingestion pipelines across distributed storage (ArangoDB, Elasticsearch) to process high-volume multi-source feeds with fault tolerance and schema evolution support
- Designed and operated ETL workflows using Apache NiFi for ingestion, transformation, routing, and automated validation of heterogeneous data streams
- Orchestrated ETL workflows with Apache Airflow
- Implemented automated validation and scenario-based verification (BDD) to detect data anomalies earlier and reduce incorrect downstream records
- Optimized the automated file-delivery script used to ship customer deliverables on each release, improving delivery reliability and reducing manual intervention
- Translated client and stakeholder requirements into scoped technical work, contributing to project planning, architecture documentation, and technical specifications
- Completed onboarding for the company's move into agentic AI and joined the team building automation agents on the Hermes framework, integrating MCP and Mattermost for operational communication
- Led a team of 4 engineers, conducting code reviews, mentoring, and driving internal tooling and performance optimization POCs
- Maintained containerized services on Docker, Scrapyd, and Linux infrastructure, integrating releases into CI/CD pipelines for versioned rollouts and monitoring
- Participated in Agile ceremonies and delivery tracking via Jira

Shopalyst Technologies

Apr 2021 - Jan 2024

Software Engineer II - Data Operations and Engineering

Trivandrum, Kerala, India

- Architected Python and Golang backend systems that processed millions of records daily, improving pipeline throughput and operational reliability
- Reverse-engineered Android applications, including SSL-pinned apps, to discover internal endpoints and integrate RESTful APIs into structured pipelines, using BrowserStack and LambdaTest for cross-device testing and validation
- Designed and enforced automated verification pipelines and data quality checks to ensure correctness across extracted datasets
- Operated distributed storage and messaging to support high-throughput workflows, using Elasticsearch, Cassandra, and RabbitMQ
- Used Apache Spark (PySpark) for distributed data processing in the data pipelines
- Automated operational workflows with Bash, Cron, and Jenkins, reducing manual failures and deployment errors
- Mentored junior engineers and maintained system stability through proactive monitoring, debugging, and incident resolution

Tata Consultancy Services

Dec 2020 - Mar 2021

IoT Edge Developer

Kochi, Kerala, India

- Architected modular C++ services for Azure IoT Edge to enable reliable, low-latency processing across distributed edge environments
- Delivered Dockerized services and automated builds using Azure DevOps and Git to support continuous integration and deployment
- Implemented protocol-heavy solutions with strict validation requirements to ensure compliance and system integrity in regulated environments

Tata Consultancy Services

Dec 2018 - Nov 2020

System Engineer

Chennai, Tamil Nadu, India

- Developed Python ETL crawlers using Scrapy and BeautifulSoup to extract, clean, and transform e-commerce data supporting downstream analytics and ML feature pipelines
- Built parsing and feature-extraction logic that improved data completeness and reduced downstream cleaning effort, using regular expressions and pandas
- Documented crawling workflows, contributed to code review processes, and supported knowledge transfer to maintainable scraping codebases

PROJECTS

Rapid Prototyping & AI-Assisted Tooling (Vibe Coding)

Built a full-stack food ordering web app (dev.theplato.in), owning UI, backend logic, data persistence, and deployment

Firebase Studio, Antigravity

EDUCATION

Bachelor of Technology, Computer Science and Engineering NSS College of Engineering, University of Calicut — Aug 2014 - Jul 2018

Palakkad, Kerala, India

PUBLICATIONS

"Different Approaches To Crack Monitoring of Buildings Using IoT," IJRASET Publications, Aug 2018 — IJRASET Publications Aug 2018

SKILLS

Data Ingestion & ETL: ETL/ELT workflows, Apache NiFi, Apache Airflow, large-scale ingestion pipelines, workflow orchestration, multi-source data acquisition, API reverse-engineering, Scrapy spiders, Scrapyd, web crawling, anti-bot countermeasures, headless-browser evasion, proxy tunnelling

Web & Parsing: Scrapy, BeautifulSoup, Selenium, Playwright, curl_cffi, HTML, JavaScript, regular expressions, pandas, RESTful APIs, GraphQL APIs

Agentic AI: Hermes Agent framework, MCP (Model Context Protocol), automation agents

Languages: Python (Advanced), SQL, Golang, Java, C++, Bash

Big Data & Storage: Distributed storage, partitioning, indexing, schema evolution, Elasticsearch, Cassandra, ArangoDB, PostgreSQL, MySQL, MongoDB, RabbitMQ, time-series databases

Infrastructure & DataOps: Docker, Linux, CI/CD (GitLab, Jenkins, Azure DevOps), Git, Cron, Jira

Cloud & Platforms: Azure, Azure IoT Hub, Azure IoT Edge, Firebase

Actively Learning: Databricks (Delta Lake, Unity Catalog), Medallion architecture (bronze/silver/gold), Azure Data Factory, Kubernetes

Data Governance: Data lineage, data contracts, scenario-based validation, QA, PII/compliance, logging & fault analysis

Methodologies: Agile/Scrum, BDD, system design, code reviews, architecture documentation, training sessions, project planning & execution

Spoken Languages: English (Fluent), Malayalam (Native), Tamil (Beginner)