

JAFSAL M A

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

Software Engineer and Technical Lead with 6+ years of experience specializing in **large-scale data extraction from e-commerce platforms, government authority portals, and regulatory websites**. Experienced in building **resilient, high-throughput web crawlers and data ingestion pipelines** for complex, frequently changing sources. Strong background in **Python, Java, Scrapy, distributed data processing, and CI/CD**, delivering **reliable, production-grade systems** in remote, agile environments.

PROFESSIONAL EXPERIENCE

PST.AG, Weimar, Germany

2024 – Present

Python Developer / Team Lead – Data Engineering & Web Scraping (Remote)

- Designed and maintained scalable, production-grade web scraping and data ingestion systems using Python, Scrapy, Playwright, Selenium, and BeautifulSoup across heterogeneous sources.
- Architected and maintained a large-scale scraping ecosystem of ~100 active production projects, crawling 400,000+ pages daily, while ensuring fault tolerance, anti-bot resilience, and data quality.
- Optimized crawler performance, stability, and fault tolerance by handling dynamic content, DOM volatility, rate limiting, and anti-bot protections.
- Deployed, versioned, and monitored spiders using Scrapy and ScrapydWeb, enabling controlled releases and operational visibility.
- Managed high-volume proxy rotation strategies (Oxylabs, Proxyrack, Webshare) to maintain high crawl success rates and minimize blocking.
- Ensured data quality and schema consistency through validation, normalization, error handling, and structured outputs (JSON, XML, XLSX).
- Collaborated with cross-functional stakeholders to deliver custom data acquisition solutions aligned with business requirements.

Technical Leadership & Team Management

- Lead and mentor a team of 4 Java developers, providing technical direction, task allocation, and code reviews.
- Enforced coding standards, design patterns, and best practices across Java and data-processing components.
- Collaborated with architects and product owners to translate requirements into scalable system designs.
- Integrated Java and Python components into CI/CD workflows using GitLab-based version control.
- Participated in Agile/Scrum ceremonies and tracked delivery via Jira.
- Drove continuous improvement initiatives, including POCs, tooling enhancements, and performance optimizations.

Tech Stack: Python, Java, Scrapy, Playwright, Selenium, Scrapy, Pandas, NumPy, XML, GitLab, Jira, ERPNext, Zep, Proxy Managers

Shopalyst Technologies, Trivandrum, Kerala, India

2021 – 2024

Software Engineer II - Data Operations and Engineering

- Led development of scalable crawlers for global e-commerce platforms using Python (Scrapy, Requests, Selenium, BeautifulSoup) and Golang, supporting high-volume data ingestion.
- Extended data extraction beyond the web by reverse-engineering Android applications, including SSL-pinned apps, to identify and consume internal APIs.
- Used Apktool, Burp Suite, mitmproxy, and network inspection techniques to bypass SSL pinning and analyze encrypted traffic.

- Automated mobile app crawling and testing pipelines using Appium, BrowserStack, and LambdaTest across emulators, real devices, and cloud environments.
- Managed large-scale proxy infrastructure with Bright Data (Luminati) to support geographically distributed crawls.
- Implemented data storage and messaging pipelines using Elasticsearch, Cassandra, RabbitMQ, ensuring scalability and fault tolerance.
- Automated operational workflows using Bash scripting, Cron jobs, and Jenkins, reducing manual intervention and improving reliability.
- Mentored junior engineers, enforced coding standards, and reviewed scraping logic to maintain data quality and system stability.

Tech Stack: Python, Golang, Scrapy, Selenium, Requests, Cassandra, Elasticsearch, Kibana, RabbitMQ, Jenkins, Bash, Linux

Tata Consultancy Services, Kochi, Kerala, India

2020 – 2021

IoT Edge Developer

- Developed and maintained C++-based Azure IoT Edge modules for industrial gateways in collaboration with a European defense and shipbuilding client.
- Integrated Azure IoT Hub and Azure IoT Edge for secure device-to-cloud communication.
- Containerized edge workloads using Docker and built modules with CMake for cross-platform deployment.
- Implemented CI/CD pipelines using Azure DevOps and Git, enabling automated builds and deployments.
- Worked with industrial communication protocols (MODBUS) for real-time data acquisition from edge devices.
- Actively participated in Agile/PLM ceremonies, sprint planning, and technical reviews.

Tech Stack: C++, Azure IoT Edge, Azure IoT Hub, Docker, CMake, Azure DevOps, Git, MODBUS

Tata Consultancy Services, Chennai, Tamilnadu, India

2018 – 2020

System Engineer

- Built and maintained web crawlers for e-commerce platforms using Python, Scrapy, BeautifulSoup, and Requests.
- Ensured product attribute enrichment and data accuracy through continuous crawler updates and validation logic.
- Automated reporting, validation, and internal workflows using Python scripts, improving operational efficiency.
- Performed data analysis and cleaning to extract actionable insights from scraped datasets.
- Collaborated with cross-functional teams to resolve data quality issues and maintain pipeline reliability.

EDUCATION

NSS College of Engineering, University of Calicut, Palakkad, Kerala, India

2014-2018

Bachelor of Technology, Computer Science and Engineering

- Developed a remote system to monitor structural health of old buildings using ESP32 microcontroller, Raspberry Pi, Adafruit VL6180X, IP Camera, Telegram Bot, Firebase
- Published a paper Titled "Different Approaches To Crack Monitoring of Buildings Using IoT" in IJRASET Publications

SKILLS

- **Programming:** Python, C++, Golang, SQL, Bash

- **Web Scraping & Automation:** Scrapy, Selenium, Playwright, BeautifulSoup, Requests, XPath, Appium
- **Data Engineering:** Pandas, NumPy, JSON, XML, Elasticsearch, Cassandra, RabbitMQ, Kibana
- **Reverse Engineering & Security:** Burp Suite, mitmproxy, Apktool, SSL Pinning Bypass
- **Cloud & IoT:** Azure IoT Hub, Azure IoT Edge, Azure DevOps, Azure Data Lake Storage, Firebase
- **DevOps & Tooling:** Docker, Jenkins, GitLab, GitHub, CI/CD, CMake, Cron
- **Proxy & Crawling Infra:** Oxylabs, Bright Data, Proxyrack, Webshare
- **Practices:** Agile, ETL Pipelines, Data Quality Assurance, Process Automation, Sprint Planning

Publications

- Different Approaches To Crack Monitoring of Buildings Using IoT
 - IJRASET Publications · Aug 4, 2018
 - <https://www.ijraset.com/fileserve.php?FID=17620>
