MUHAMMED FARIS

Senior Data Engineer

Kerala • 8111863658 • muhammedfaris.um@gmail.com • [LinkedIn](#) • [GitHub](#) • [Website](#)

SUMMARY

Data Engineer with 5+ years of experience building scalable ETL pipelines, data warehouses, and lakehouse architectures on Databricks and AWS. Proficient in PySpark, Airflow, and Python for data processing and orchestration. Experienced in designing data models, optimizing SQL queries, and ensuring data reliability through automation and CI/CD workflows.

SKILLS

- **Data Engineering:** Data Modeling, ETL Pipelines, Lakehouse Architecture, Backend Development
- **Programming & Frameworks:** Python, PySpark, FastAPI, Flask, Playwright
- **Databases:** PostgreSQL, Databricks Delta Tables, Redis
- **Cloud Platforms:** Amazon Web Services (AWS), Databricks
- **Orchestration & Messaging:** Airflow, TaskFlow API, Kafka, RabbitMQ, Linux Cron
- **DevOps & CI/CD:** Linux, Docker, Git, GitHub Actions, AWS SAM
- **Libraries:** Pandas, BeautifulSoup, PDFMiner, Locust, Memray, Presidio
- **Languages:** English, Malayalam

WORK EXPERIENCE

Senior Data Engineer, Chi-Square labs   **Nov 2021 - Present**

- Designed and developed an Enterprise *Data Warehouse* (EDW) on *Databricks* using *PySpark*, implementing *star schema* with fact and dimension tables to support business intelligence and analytics.
- Designed *data models* and implemented schema design best practices (*star and snowflake*) to support analytics scalability and maintain data integrity.
- Experienced in using Databricks for *big data processing*, collaborative analytics, and *notebook-based development*.
- Built a unified *Lakehouse architecture* on Databricks, integrating structured and unstructured data stored in *AWS S3 Data Lake* for analytics and ML workloads.
- Experienced in maintaining and optimizing *ETL pipelines* to ensure seamless data flow and transformation across multiple systems.
- Developed multiple *Python-based scripts* and *backend services* for data extraction, transformation, and automation.
- Designed and managed *Airflow DAGs* for data deduplication, job scheduling, and complex *orchestration workflows*.
- *Containerized* data processing and API applications using *Docker*, deploying them on *AWS Lambda* and *ECS* for scalable execution.
- Utilized *PostgreSQL* for building and maintaining a *knowledge base service*, ensuring efficient data storage and retrieval.
- Set up *monitoring and logging* for ETL pipelines using AWS *CloudWatch* and *Airflow task logs* to proactively detect failures and optimize performance

- Skilled in extracting and transforming structured and unstructured data using Python libraries, including **BeautifulSoup**, **Playwright**, **PDFMiner, pdfplumber, Camelot, JSON, Pandas,** and **ocrmypdf**.
- Hands-on experience in deploying and managing data solutions using **AWS services** such as **S3, Lambda, ECS, AWS MQ, SQS** and **EC2**.
- Implemented automated **CI/CD** workflows using **GitHub Actions** and **AWS SAM**, streamlining deployment processes and ensuring consistent delivery.
- Developed an **ML-based chatbot** leveraging **spaCy** and a rule-based NLP approach, integrated with WhatsApp using **Twilio** and **FastAPI** for real-time communication.
- Performed concurrent load simulations with Locust and memory usage analysis via Memray to enhance scalability and efficiency of Python-based services.
- Skilled in **troubleshooting** and resolving issues related to data processing, ensuring minimal disruption to business operations, and maintaining high data availability.
- Implemented and managed inter-service communication using **FastAPI, Kafka,** and **RabbitMQ** to ensure reliable, **real-time data exchange** across distributed systems.

Software Engineer, Xanthron E-Solutions   **Jul 2020 - Oct 2021**

- Collaborated with a team focused on developing applications for the **education domain**.
- Designed and implemented scalable **RESTful** APIs using **Flask**, enhancing backend efficiency and reliability.
- Utilized a tech stack including **Python, Flask,** and **MySQL** for robust backend development.
- Deployed and maintained web applications on **AWS EC2** instances, leveraging **Celery** for **asynchronous task execution** and improved system performance through parallel processing.
- Worked closely with front-end teams to design and refine APIs, ensuring seamless system integration.

ACHIEVEMENTS AND CERTIFICATIONS

- **Databricks Fundamentals Accreditation [E-VP9791]**
- Certified in **HIPAA (Health Insurance Portability and Accountability Act)** compliance.
- Diagnosed and rectified a critical **memory leak** within the pdfplumber Python library by using memory profilers; enhanced document processing speeds by 40% and improved overall system stability.
- **Automated** the process of extracting data from medical records, *tripling processing speed* and freeing up two team members for key compliance activities.

EDUCATION

Bachelor of Science in Computer Science   **2017 - 2020**

University of Calicut

- Studied programming, database systems, software engineering, web development, and data structures.
- Key Areas: C, Java, PHP, Bash, Android, DBMS, Algorithms, Operating Systems, Computer Networks, and Applied Mathematics.