Rana Aurangzaib

Senior Backend Engineer/Senior Software Engineer

Email : aurangzaib048@gmail.com
LinkedIn: https://www.linkedin.com/in/rana-aurangzaib-45557b135
Github: https://github.com/aurangzaib048
Medium: https://aurangzaib048.medium.com
Cell: +92-304-6031309

Summary

Results-oriented Senior Backend Engineer with 6+ years of experience in Python and data-pipelines (ETL). Specialized in developing, deploying, and optimizing ETL, micro-services, and web scraping tools. Proficient in AWS, Docker, Kubernetes, PostgreSQL, MS-SQL Server, and more. Proven ability to lead cross-functional teams, manage multiple projects, and deliver innovative solutions. Seeking a challenging role to leverage technical skills and leadership experience to drive efficiency and innovation.

Programming Skills

- **Languages, Frameworks and Tools** : Python, Django, Luigi, FastAPI, Flask, Scrapy, AsyncIO, Machine Learning, Multithreading, Multiprocessing, Pytest, Unittest, Selenium, BeautifulSoup, MatPlotLib, Git, GitHub, data-pipelines (ETL), Restful APIs, GraphQL, ElasticSearch, Linux, LaTeX
- **Database Systems**: PostgreSQL, Redis, MS-SQL Server, Cassandra, Snowflake, DynamoDB
- **Big data technologies**: Apache Airflow, Apache Beam, Apache Flink, Apache Kafka
- **Cloud Technologies**: AWS (S3, EKS, EC2 and a lot more), GitHub Actions, CI/CD, Docker, Kubernetes, Prometheus
- **Methodologies**: Agile, Scrum, Kanban, System analysis and Design, Microservices, Event Driven Development, Test Driven Development

Experience

- **Brave Software** — San Francisco, CA (Remote, Pakistan)
  Backend Engineer — *Nov 2021 - Present*
  - **Brave News pipeline**:
    - Developed and deployed a Python-based news aggregation pipeline at Brave Software.
    - Increased content update rate by almost 40% through strategic use of multi-processing and multithreading techniques.
    - Utilized Docker for containerization and Kubernetes (EKS) for orchestrated deployment.
    - Achieved over 200K+ Monthly Active Users (MAU), highlighting project effectiveness.
    - Demonstrated adaptability and scalability as the service expanded to 14+ global regions.
  - **Brave News Source Suggestion**:
    - Developed a Python service for source embedding representations and similarity matrix in Brave News.
    - Elevated news source diversity by 30% and prioritized user privacy and provided users with more precise and varied news content suggestions.
    - Utilized Natural Language Processing (NLP) techniques for high-quality embeddings and accurate similarity scores.
    - Deployed the service via Docker and Kubernetes (EKS) for scalability.

    - Project details and code available at: https://github.com/brave/source-suggestions.
  - **Brave News Topic Clustering**:
    - Collaborated with the research team to enhance the Brave News Pipeline.
    - Added the "Brave News Topic Clustering" sub-project for real-time critical news delivery accuracy.
    - Integrated public sentence-transformers models and mini LLM.
    - Boosted user engagement by 40% as a result of this enhancement.
    - Combined collaborative effort and advanced technology for content relevance and improved user experience.
  - **Brave News Article Categorization**:
    - Trained and optimized BERT-based NLP model using Python and FastAPI, achieving 95% accuracy in categorizing articles.
    - Analyzed titles and descriptions for precise categorization.
    - Resulted in a 30% improvement in user content discovery.
    - Enabled readers to access and engage with personalized articles.

- **Arbisoft** Lahore, Pakistan
  Senior Software Engineer *Aug 2018 - Nov 2021*
  - **Multi-Source Data Integration for Hotel Pipeline with Python**:
    - Contributed to enhancing a hotel data pipeline using Python, Luigi, and Airflow.
    - Achieved a 70% increase in data integration efficiency.
    - Designed a system for seamless collection and integration of data from diverse sources (S3, Snowflake, raw files).
    - Processed and consolidated information from providers such as Sabre, EAN, Priceline, Booking, Gaita, and Trust-you.
    - Stored data effectively in Postgres, Cassandra, and DynamoDB databases for easy accessibility and analysis.
  - **Micro-Service with Cloud Integration and Real-Time Monitoring**:
    - Developed Python-based micro-service using FastAPI and GraphQL for receipt parsing.
    - Integrated cloud APIs, achieving automatic reimbursement in the back-end application.
    - Utilized Prometheus for real-time monitoring, ensuring system performance metrics:
      - Improved throughput by 25%.
      - Reduced latency by 20%.
      - Managed a 15% increase in request volume efficiency.
  - **LogP - Web-based Log Parsing and Visualization Application**:
    - Led the development and launch of LogP, a web-based log parsing app using Django, React, and Matplotlib.
    - Designed and developed the application to monitor ETL logs, achieving. Improved data extraction efficiency by 30%. And enhanced visualization using Matplotlib for better insights.
    - Coordinated with cross-functional teams, gathering requirements and defining features.
    - Ensured timely delivery of a high-quality application.
  - **Email Parser for Travel and Hotel Bookings**:
    - Collaborated on a travel and hotel email parsing team using Python, Beautiful Soup, and regex.
    - Extracted essential information from emails, improving parsing accuracy and efficiency.
    - Designed and implemented a deep learning model based on LSTM networks.
    - Built an application converting unstructured email messages into structured JSON.

    - Stored structured data in Cassandra databases for efficient retrieval.
  - **Data-gen-2000**:
    - Spearheaded the development and deployment of Data-gen-2000, a Python-based tool for generating dummy data.
    - Designed and implemented various data generation algorithms for effective testing.
    - Simulated realistic scenarios and tested edge cases to ensure comprehensive coverage.
    - Utilized Python libraries and techniques for optimized data generation speed and accuracy.
    - Minimized resource consumption while facilitating automation testing of databases.
  - **Web Scraping Tool with Scrapy Framework**:
    - Collaborated on developing a web scraping tool using Python and Scrapy framework.
    - Extracted data from diverse platforms: China Coach, Damart, Eloquii, Jackjones, Twitter, and Instagram.
    - Designed and implemented Scrapy spiders to crawl and scrape web pages, extracting relevant information.
    - Stored data in a structured format for easy retrieval and analysis.
    - Utilized techniques like proxy rotation and user-agent spoofing for detection avoidance and enhanced scraping efficiency.

Publications

- Our study, **"Scalable Containerized Pipeline for Real-time Big Data Analytics"**, published in the IEEE International Conference on Cloud Computing Technology and Science (CloudCom) 2022, introduced an advanced data processing pipeline deployed on Kubernetes. It marked a significant enhancement in real-time big data handling, delivering up to 2.4× increased throughput and 80× reduced latency. Available on IEEE Xplore https://ieeexplore.ieee.org/document/10005498.

Research Experience

- **Cloud Computing and Big Data Lab @ FCIT (ex PUCIT)** Lahore, Pakistan
  Researcher *Mar 2022 - Dec 2022*
  - During my tenure at the Cloud Computing and Big Data Lab at FCIT, Lahore, Collaborated with esteemed professionals like **Dr. W. Iqbal** and **Dr. M. Abdullah**. We enhanced real-time anomaly detection in IoT data up to almost 80% using advanced deep learning techniques, marking a significant advancement in adaptive big data pipelines and efficient data processing algorithms.

Education

- **University of the Punjab** Lahore, Pakistan
  M.Phil (MS) in Computer Science *Oct. 2018 – Jul. 2020*
- **University of the Punjab** Lahore, Pakistan
  Bachelors of Computer Science *Oct. 2014 – Aug. 2018*