Franklin CD Samuel

samuelfranklinjobs23@gmail.com | (571) 587-6747 | LinkedIn |Github|Fairfax, VA, USA (Open to Relocation)

Data Engineer with 5+ years of experience in designing, developing, and maintaining robust data pipelines and infrastructure using a range of AWS services. Adept at leveraging big data technologies and database management systems to deliver scalable, high-performance solutions. Proficient in collaboration and continuous learning, with a proven ability to solve complex problems and drive project success.

AREAS OF EXPERTISE

Programming: Python (NumPy, Pandas, Matplotlib, scikit-learn, TensorFlow), SQL, Bash.
Big Data & NoSQL: Hadoop, Apache Spark (Spark SQL, Spark ML, Spark Streaming), MongoDB, Cassandra.
Data Engineering / Data Warehousing: ETL, SSIS, Snowflake, Data Pipelines, Airflow, Kafka.
Machine Learning: Supervised/Unsupervised Learning, Neural Networks, Deep /Reinforcement Learning.
Databases: MySQL, PostgreSQL, IBM Db2, MS SQL Server.
Cloud: AWS services like S3, Glue, Redshift, EMR, and Lambda, CI/CD, specializing in scalable data engineering, ETL and analytics solutions.
Visualization: Tableau, Power BI - Tools: Jupyter Notebook, Azure, Git, Jira.

CAREER HIGHLIGHTS

Housing and Residence Life, George Mason University, Fairfax, US **Jan 2023 – Present**
Data Analyst and Data Engineer

- Architected and implemented data pipelines using **AWS Glue, S3**, and **Lambda** to enhance data collection and analysis processes.
- Designed and managed data warehousing solutions using AWS **Redshift**, improving data query performance by 30%.
- Utilized Python and **Apache Spark** to process and analyze large datasets, resulting in actionable insights that drove a 50% improvement in user satisfaction.
- Automated deployment and monitoring processes using Jenkins and AWS **CodePipeline**, reducing deployment times by 40%.
- Enhanced data security by implementing encryption and access control mechanisms across AWS services.
- **Integrated Google Analytics data** into AWS-based reporting pipelines, enabling comprehensive analysis of user behavior and website performance, which informed strategic decisions and optimized user experience.

Central Water Park Research Park Station, Pune, India **Jan 2021 – December 2021**
Machine Learning and Data Engineer

- Developed and optimized data pipelines using **AWS EMR and S3** for processing over 100K empirical data points, improving data accuracy by 30%.
- Engineered predictive models using **Apache Spark and Python**, deployed on **AWS Lambda,** achieving a 25% increase in model accuracy.
- Managed large-scale data storage solutions using **AWS Redshift and PostgreSQL,** enabling real-time data access and analytics.
- Collaborated with cross-functional teams to integrate data processing workflows with existing AWS infrastructure, enhancing overall system efficiency.

National Institute of Electronics and Information, Chennai, India **January 2022 – June 2022**
Machine Learning Engineer

- Developed scalable data pipelines using **AWS Glue and Lambda**, improving data processing speed by 25%.
- Integrated AWS services such as **Redshift and Aurora DB** to enhance data warehousing capabilities, supporting advanced analytics.
- Implemented Python-based ETL processes to handle large datasets, ensuring high data quality and consistency.
- Conducted performance tuning and optimization for data workflows, resulting in a 20% improvement in overall system performance.

Make A Difference(NGO) , Vellore, India **October 2020 – June 2022**
Business Analyst and Fundraising head

- Spearheaded data-driven fundraising campaigns, utilizing analytics to target key donor demographics, resulting in a 20% increase in funds raised.
- Conducted business analysis to optimize operational processes, leading to a 15% improvement in project efficiency and resource allocation.
- Collaborated with cross-functional teams to develop and implement strategic initiatives, enhancing organizational impact and outreach.
- Managed relationships with stakeholders and donors, securing partnerships that contributed to long-term sustainability.
- Utilized data visualization tools to present fundraising performance reports to the board, aiding in strategic decision-making.

PROJECTS & RESEARCH

Water Depth Estimation Using Sentinel-2 Satellites for Shallow Water

- Pioneered research and implementation of advanced algorithms, incorporating SVMs and Deep Neural.
- Accurately estimated shallow water depths from Sentinel-2 Satellite image bands, achieving over 90% accuracy.

Developed and Designed Personal Portfolio

- Developed the portfolio using React and Bootstrap, enhancing component reusability and reducing development time by 50%.
- Leveraged React's component-based architecture to create a modular and maintainable codebase, reducing development iterations by 40% .

Research Paper Published in Springer

- Led a groundbreaking initiative focused on data analysis and interpretation within IoT-based systems for critical medical services and healthcare applications, leading to a 50% improvement in data accuracy.

Real Time Stock Market Simulation using Kafka on AWS

- Designed and implemented a real-time stock market simulation using Kafka on AWS, automating ETL processes with a 99% data processing accuracy, handling over 10,000 records per minute, data generation to S3 storage and Athena analysis**.**

Ask PDF

- Designed a PDF processing workflow in AskPDF, leveraging OpenAI's GPT-3.5-turbo to facilitate interactive Q&A, reducing manual document review time by 60%

End-to-End CI/CD Pipeline Implementation Using Jenkins, Docker, Kubernetes, and Ansible

- Engineered a robust CI/CD pipeline using Jenkins, Docker, Kubernetes, and Ansible, automating deployment processes and reducing release times by 70%.
- Deployed containerized builds with Docker and managed microservices orchestration on Kubernetes, seamlessly integrating Jenkins for continuous integration and delivery.
- Automated infrastructure and configuration management through Ansible, integrated with Jenkins, decreasing manual errors by 60%.

Web Data Extraction and Analysis System Feb 2024 – May 2024

- Developed a Python-based web scraper using BeautifulSoup and Pandas for data extraction and processing
- Compiled scraped data into a well-formatted file for further analysis, with 30% reduction in processing time.
- Designed program to handle multiple websites and diverse HTML structures, increasing efficiency by 25%.

Traffic Signal Recognition Sep 2023 – Dec 2023

- Developed a deep learning model using TensorFlow for real-time traffic signal classification (85% accuracy).
- Implemented data preprocessing pipeline using NumPy and Pandas for efficient model training.
- Optimized model size and speed on edge device via quantization and pruning techniques for enhanced inference.

Big Data Analytics Platform Jan 2022 – July 2022

- Built a data pipeline using Apache Spark and Kafka to process and analyze large-scale streaming data.
- Designed and optimized a data warehouse schema in PostgreSQL for efficient analytics queries.
- Implemented machine learning models using Spark ML for predictive analytics on big data.

LEADERSHIP AND ACHIVWMENTS

2024 Virginia Datathon Winners, Virginia Office of Data Governance and Analytics

- Oversaw the development of the innovative Generative AI-driven solution, Commonwealth AI Career Hub, demonstrating its potential to revolutionize workforce challenges.

2022 E Cell Hackathon Runners up

- Conceptualized and developed a healthcare app inspired by Practo, aimed at connecting patients with doctors for online consultations and appointment bookings.

EDUCATION

Master in Computer and Information Science | George Mason University, Fairfax,USA | **May 2024**
Bachelor of Technology, Computer Science (Minor in Bioinformatics) | Vellore Institute of Technology,India|**May 2022**

CERTIFICATIONS

- Machine Learning Specialization - Andrew Ng (Stanford University)
- IBM Data Engineering Professional Certificate
- Google Data Analytics Professional Certificate
- Python By Kaggle
- AWS Certified Data Engineering
- SQL by Udemy
- AWS Cloud Practioner