Kaman Yu

Carrollton, Texas | U.S. Citizen | (713) 999-0238 | kamanyu.39@gmail.com

PROFESSIONAL SUMMARY

Data Engineer with hands-on experience building and optimizing ETL pipelines, streaming data solutions, and automated data validation workflows in distributed production environments. Proficient in SQL, Python, Java, Kafka, and CI/CD with a strong record of ensuring data quality and operational integrity. Excellent collaborator with Agile teams, capable of delivering scalable data solutions and continuously improving engineering practices.

PROFESSIONAL EXPERIENCE

Data Engineer

Cognizant Technology Solutions | Jan 2023 – Present

- Designed and implemented ETL pipelines using **Kafka (Java)** and orchestrated CI/CD using **Jenkins**, enabling reliable and repeatable automated data operations.
- Built and maintained automated data validation scripts using **Python** and **SQL** to clean, transform, and load data into heterogeneous target systems including **MongoDB, Cassandra, Oracle DB, IBM DB2,** and **Teradata EDW**, reduce manual data validation effort by approximately **35%**.
- Collaborated with cross-functional stakeholders (product, QA, and business teams) to translate product needs into scalable data engineering deliverables, supporting reporting and dashboard development in **Tableau**.
- Assisted in maintaining data consistency across environments using **K2View**.
- Leveraged AI-assisted development tools (**GitHub Copilot)** to accelerate development of data validation scripts, enhance code readability, and improve maintainability of data validation workflows.

SIDE PROJECTS

Twitter Sentiment Analysis

- Built streaming application using **Java** and **Spark** to process live Twitter API data.
- Indexed processed data into **Elasticsearch** for scalable search and analysis.
- Built interactive **Kibana** dashboards to visualize sentiment trends and provide actionable insights.

TECHNICAL SKILLS

- **Programming Languages:** Java, Python, SQL
- **Data Engineering & Streaming:** Apache Kafka, Spark Streaming, ETL, Data Validation, Data Migration, AWS (S3, EC2 basics), data warehousing, K2View, Apache Airflow
- **Databases:** MongoDB, Cassandra, Oracle DB, IBM DB2, PostgreSQL, Teradata, Snowflake (familiarity)
- **Data Tools & Visualization:** Elasticsearch, Kibana, Pandas, Tableau, Excel
- **DevOps & Tools:** Git, Jenkins(CI/CD), Agile/Scrum

EDUCATION

Bachelor of Science in Computer Science

University of Texas at Dallas | Aug 2018 - May 2022