Lakshmi Sailaja Uppalapati

lakshmisu13@gmail.com ● (513) 462-6437 ● LinkedIn/sailajau

PROFESSIONAL EXPERIENCE

Software Dev Engineer Intern, Amazon.com Services LLC, Dallas, TX   Jun 2024 – Sep 2024

- Integrated **Twilio** alongside **SNS** for **AWS Cognito User Pool** using **Java** and **React** while leveraging **OOP principles** like **Encapsulation**, **Inheritance**, **Abstraction** ensuring scalability, and efficient SMS delivery.
- Collaborated with cross-functional teams to invoke the **UMS Service** using the **AsyncTextInternal API** during user sign-up, streamlining SMS delivery while adhering to **SDLC** design patterns for modular development.
- Contributed to the design and setup of an **Enterprise Data Lake**, supporting analytics, processing, and reporting of over **10 TB** of rapidly changing data, ensuring improved **scalability**.
- Supported **data quality** maintenance by performing **cleaning** and **transformation** in **AWS EMR**, processing large datasets between **S3** and **DynamoDB**, and reducing data discrepancies and transfer time by **25**%.
- Created and managed **Apache Airflow DAGs** to automate **ETL workflows**, using **Spark** and **AWS Glue** to process large datasets, improving data freshness and pipeline efficiency.

Data Engineer, Progressive Insurance, Mayfield, OH   Oct 2023 – May 2024

- Revamped real-time data pipelines using **GCP** services like **Pub/Sub**, **BigQuery**, and **Dataflow**, ensuring **HIPAA** compliance and improving data visualization and reliability.
- Engineered **ETL workflows** using **Apache Airflow** using **Python**, **Pandas**, **Spark** to automate advertising data pipelines, ensuring real-time data availability and reducing manual intervention by 90%.
- Built scalable data architectures using **Terraform(Iac)** and **GCP Cloud** Functions, supporting workflows processing over 1M events daily, decreased deployment times by **35**%, and boosting **team collaboration** by **25**%.
- Designed and implemented scalable **Data Warehouse** and **Data Lake** models using **star** and **snowflake** schemas, optimizing **OLTP/OLAP** data models for structured and unstructured datasets by enhancing performance by 30%.
- Architected real-time streaming data pipelines using **Spark** Streaming and **Apache Kafka**, processing over 750K ad events daily with sub-second query latency, accelerating ad analytics by **35**%.
- Integrated **Large Language Models** (**LLM**s) into data pipelines to automate **data classification** and sentiment **analysis**, improving predictive analytics and decision-making accuracy by **20**%.

Data Engineer, Keka Technologies, India   Oct 2020 – Jun 2023

- Streamlined high-volume **ETL** pipelines using **Azure Data Factory**, processing **1.5 TB** of **data** daily and increasing data extraction and transformation efficiency by **40**%.
- Preprocessed, cleansed, and transformed **1 TB+ datasets** using Azure Databricks, leveraging **Scala** and **Spark** to improve data processing speed by **40**% and optimize performance.
- Worked on Ingestion of data from source(**On-prem SQL**) to target (**ADLS**) using **Azure Synapse Analytics** (Data Warehouse) & **Azure SQL DB**, reducing data processing time by **35**% and enhancing query performance by **25**%.
- Maintained more than 10 databases as **SME** by managing incoming and new data products in **data lake,** while redesigning **NoSQL databases**, including **Cassandra** and **Azure Cosmos DB**, improving data retrieval speeds by **40**%.
- Utilized **Docker** for containerizing applications and **Kubernetes** for orchestration, streamlining deployment and scaling of **data processing** workloads, improving system efficiency by **30**%.

SKILLS

Programming Languages & Data Processing: Python, SQL, Java, Scala, PySpark, Pandas
Cloud Platforms & Databases: AWS (S3, Lambda, Glue, EMR, RedShift, Athena, API Gateway, CloudWatch), GCP (BigQuery, Pub/Sub), Azure (Data Factory, Synapse, Databricks), SQL Server, MySQL, Cassandra, CosmosDB, DynamoDB
ETL & Big Data Tools: Apache Spark, Apache Airflow, Hadoop, Apache Kafka, Hive, HDFS, HBase, Informatica
Frameworks & Development: Spring, Spring Boot, Spring Cloud, React.js, JUnit, Mockito, Coral, Git, GitHub, Maven, Brazil, Jenkins, CI/CD
Data Visualization Tools: Power BI, Looker, Tableau
Workflow Tools & Operating Systems: Apache Airflow, Unix, Linux, Windows, Agile methodology, Agile/Scrum, QuickSight, Asana, Quip, Jira, Unix, Linux, Windows

CERTIFICATES

- Microsoft Certified: Azure Data Engineer Associate (DP-203)

EDUCATION

Master of Engineering in Computer Science, University of Cincinnati   Dec 2024
Coursework: Big Data Analytics, Cloud Computing, Advanced Database Management, Data Engineering, Machine Learning, Software Engineering, Requirements Engineering