SRI RUSHITHA JANGA

DATA ENGINEER

205-582-8342|rushitha.j@mycvhire.com| Birmingham, AL

SUMMARY

- Experienced Data Engineer with 4 years of expertise in designing and building scalable data solutions using Azure and AWS cloud platforms**.**
- Skilled in leveraging Big Data technologies, including the Hadoop and Spark frameworks, and proficient in associated tools such as MapReduce, Hive, BigQuery, HDFS, Apache Airflow, Kafka, and Apache Flink.
- Proficient in RDBMS concepts, Data Modeling (Facts and Dimensions, Star/Snowflake schemas), Data Migration, Data Cleansing, and ETL Processes.
- Leveraged Apache Spark and Spark SQL within Databricks to develop data pipelines for customer usage analysis and extracted, transformed, and aggregated data from diverse file formats, uncovering valuable insights into customer behavior patterns.

SKILLS

- **Programming Language:** Scala, Python, SQL, Java
- **IDE's:** PyCharm, Jupyter Notebook
- **Big Data Ecosystem:**Hadoop, MapReduce, Hive, Spark, Apache Airflow, Kafka, Apache Flink
- **Machine Learning:** Linear Regression, Logistic Regression, Decision Tree, K mean, Naïve Bayes, Random Forest
- **Cloud Technologies:** AWS (EC2, S3 Bucket, Amazon Redshift, Glue, Lambda,Kinesis, AWS Pipeline), Azure (Azure DevOps, Azure Data Lake, Azure Data Factory, Azure Databricks)
- **Packages:** NumPy, Pandas, Matplotlib, SciPy, Scikit-learn, Seaborn, TensorFlow, PySpark
- **Visualization Tools:** Tableau, Power BI
- **ETL/Database:** MS SQL Server, MongoDB, MySQL, SSIS, Snowflake
- **Operating Systems:** Windows, MacOS

EDUCATION

Master of Science in Computer Science **Dec 2023**
University of Alabama at Birmingham, AL

Bachelor in Mechanical Engineering **Apr 2020**
Vellore Institute of Technology, Tamil Nadu, India

EXPERIENCE

Allstate, AL ***May2023–Current***
Data Engineer

- Implemented AWS Lambda functions triggered by incoming files in S3 to perform data validation and transformation before loading into Amazon Redshift, thereby enhancing data quality and pipeline efficiency.
- Build a serverless ETL pipeline using Glue to extract sales data from an on-premises database, transform it into a format suitable for Redshift, and load it efficiently for analytics.
- Utilized Apache Flink to build real-time data pipelines with low latency and high throughput, enabling real-time fraud detection and anomaly analysis.
- Developing HiveQL queries to extract and transform large datasets from HDFS, supporting data warehousing and ad-hoc analysis.
- Utilized Apache Spark for efficient processing of massive datasets, achieving an approximately 20% reduction in processing time compared to traditional MapReduce jobs.
- Applied Airflow's monitoring and alerting capabilities to proactively identify and troubleshoot data pipeline issues, minimizing data downtime and maximizing data quality.

XLogic Technologies, India*Jan2019 - Dec2021*
Data Engineer

- Employed Azure Databricks, a managed Apache Spark environment, to perform large-scale data transformations and machine learning tasks on data stored in ADLS.
- Used Spark libraries like Spark SQL and MLlib for efficient data analysis and machine learning tasks.
- Implemented Azure DevOps pipelines to automate data ingestion, transformation, and loading processes (ETL), ensuring continuous integration and delivery (CI/CD) for data pipelines, resulting in reduced data pipeline deployment time by 50%, leading to faster data availability for analysis.
- Industrialized Python scripts for efficient data loading into Snowflake, leading to a noticeable decrease in loading time and minimizing ingestion errors.
- Designed and implemented robust data pipelines using Airflow to orchestrate complex workflows, ensuring reliable and timely data delivery for critical business reports.
- Implemented robust Kafka-based streaming pipeline to ensure reliable and timely delivery of real-time data to diverse applications, enabling faster insights and decision-making.
- Built reusable, data-driven workflows that orchestrated seamless data flow between ADLS Gen2, Databricks notebooks, and other Azure data services.