Rithik K
Data Engineer
Email: rithreddy4@gmail.com |**PH**: (940)-843-3349|**Location**: New Jersey
GitHub: github.com/Rithik143-coder

LinkedIn : https://www.linkedin.com/in/rithik-k-333582362

PROFESSIONAL SUMMARY

AWS-certified Data Engineer with 2+ years of hands-on experience designing and optimizing scalable data pipelines across distributed cloud platforms. Proven success in modernizing legacy systems, implementing real-time and batch ETL workflows, and improving data reliability for business-critical decision-making. Proficient in Python, SQL, AWS Glue, Redshift, Apache NiFi, and Databricks. Known for delivering performance-optimized, secure solutions that drive efficiency and cross-functional impact.

TECHNICAL SKILLS

- **Programming & Scripting**: Python, SQL, PySpark
- **Data Processing & ETL**: Apache Spark, Apache NiFi, Apache Kafka, Talend, Airflow
- **Data Warehousing**: Amazon Redshift, Snowflake, PostgreSQL, MySQL, MongoDB
- **Data Integration & Pipelines**: AWS Glue, Step Functions, Informatica, SSIS
- **Databases**: Relational (PostgreSQL, MySQL), NoSQL (MongoDB, Cassandra)
- **DevOps & Automation:** Git, Jenkins, Docker
- **Monitoring & Logging**: Prometheus, Grafana
- **Data Visualization:** Power BI, Tableau
- **Environments**: Linux, Windows, Jupyter Notebooks

CERTIFICATIONS: AWS Certified Data Engineer - Associate. Jan 2025 – Jan 2028
Issued by Amazon Web Services| [View Credential]

EDUCATION - **Master's in Information Systems and Technology**. Aug 2022 – May 2024
University of North Texas, Denton, Texas.

PROJECTS:

1. AWS Redshift Migration Pipeline.
Simulated an enterprise Redshift migration using **AWS Glue**, **Lambda**, **Step Functions**, and **NiFi**, modeled after work at Blue Shield of California.
GitHub: aws-redshift-migration-pipeline

2. Real-Time Retail Streaming Pipeline.
Simulated a real-time retail data pipeline using **Kafka**, **Airflow**, and **Spark**, based on streaming use cases from Infosys.
GitHub: real-time-streaming-kafka-airflow

PROFESSIONAL EXPERIENCE

Data Engineer, Blue Shield of California, CA, USA. March 2024 - present

- Led the successful migration of a legacy on-premise data warehouse to **Amazon Redshift** with less than 1 hour of downtime, ensuring uninterrupted business operations and smooth transition.
- Designed and built scalable **ETL pipelines** using **AWS Glue**, **Apache NiFi**, and **Informatica** to integrate structured and unstructured data from multiple source systems.
- Developed and automated **data cleansing**, validation, and a reusable **data quality framework** using **Talend**, **AWS DataBrew**, **AWS Glue Data Catalog**, and Python scripts, reducing post-migration issues and improving governance.
- Built and maintained **CI/CD pipelines** using **GitLab**, **Apache Airflow**, and **Prometheus**, accelerating deployment cycles and minimizing manual errors.
- Engineered and automated end-to-end data workflows using **AWS Lambda**, **Step Functions**, and **Apache Spark**, streamlining both ingestion and transformation processes.
- Developed and optimized real-time and batch pipelines on **Amazon EMR** using **PySpark**, enabling high-volume processing and faster data availability for analytics.
- Boosted **Amazon Redshift** performance by redesigning schema, applying smart **partitioning and indexing strategies**, and fine-tuning complex queries.
- Managed cross-platform **ETL toolsets** including **ODI**, **SSIS**, **BODI**, and **IBM DataStage** to ensure seamless data flow and compatibility across systems.
- Strengthened security posture by configuring **IAM policies**, **KMS encryption**, and **VPC-based access controls** to protect data during and after migration.

Data Engineer, Infosys, India January2021-August2022

- Installed, configured, and maintained open-source tools including **Databricks**, **JupyterHub**, **PySpark**, **PostgreSQL**, **Prometheus**, and **Grafana**, ensuring robust data operations and real-time system monitoring.
- Designed and optimized **ETL pipelines** using **Apache NiFi** and **Apache Airflow** to seamlessly integrate structured and semi-structured data across varied sources.
- Developed a centralized **data warehouse** on **Azure SQL Data Warehouse**, enhancing reporting capabilities, cross-team collaboration, and governance compliance.
- Built **real-time streaming pipelines** using **Apache Kafka** and Airflow, enabling low-latency ingestion with high data throughput.
- Created and deployed **machine learning models** for demand forecasting using **Scikit-learn** and **TensorFlow**, improving forecasting accuracy and business agility.
- Leveraged **Apache Spark** and **Hadoop** for large-scale data processing, reducing pipeline runtimes and enabling near real-time analytics.
- Delivered actionable insights by building **interactive dashboards** in **Tableau** and **Power BI**, tailored for business and engineering stakeholders.
- Containerized data applications with **Docker** and orchestrated deployments using **Kubernetes** on Azure Kubernetes Service (AKS), ensuring scalability and environment consistency.
- Automated deployment of data workflows using **CI/CD pipelines** with **Git** and **Jenkins**, improving delivery speed and reliability.