

Veera Mukesh Aripaka

+1 (407) 406 3609 - veeramukesharipaka@gmail.com - [linkedin](#) - github.com/mukesh0005

WORK EXPERIENCE

Data Engineer

BizAcuity Solutions Pvt.Ltd

Aug 2020 - Aug 2022

- Automated the building and deployment of **Docker images** into Amazon ECR repositories, orchestrating containerized deployments with **AWS Fargate** for Microservice APIs.
- Engineered **CI/CD** pipelines utilizing **Maven** for build automation, Liquibase , **DBT** for schema management, **GitHub Actions** for workflow automation, which achieved a 20% improvement in build efficiency.
- Designed high-performance CI/CD pipelines leveraging **AWS DMS**, Lambda, and Snowpipe for **batch processing** and **CDC** changes, facilitating seamless migration of data from MYSQL.
- Automated Orchestration of Deployments into AWS RDS, Snowflake by creating **AWS Apache Airflow (MWAA)** environments, integrating GitHub Actions for automation, and **Terraform** (IaC) for provisioning.
- Implemented a serverless streaming data pipeline using **Apache Kafka**, Lambda, API gateway and **DynamoDB** to process high-volume data streams from REST APIs to different environments.
- Integrated **Datadog** across AWS environments to centralize **log management** and monitor key resources like EC2 instances, RDS databases, Lambda functions, and Glue for collection of CPU utilization, network traffic, latency.
- Utilized **Databricks** for orchestrating data transfer workflows from Data Lake Storage to Data warehouses, Databases, leveraging PySpark for efficient data transformation tasks.
- Leveraged both Spark and SQL pools within **Azure Synapse Analytics** to execute big data analytics in **OLAP** databases, for reporting, and **OLTP** databases, such as SQL Db, for handling transactional operations.
- Engaged in **Dimensional modeling** utilizing techniques such as Snowflake schema, incorporating Data Marking for **Normalization(2NF)** and Data Profiling for Data Ware houses like Snowflake and Redshift.
- Developed **RESTful APIs** using Lambda, Api Gateway to facilitate communication between client applications and server-side systems for **No-SQL** Databases such as **MongoDB**, Dynamo Db.

ETL Developer

BizAcuity Solutions Pvt.Ltd

May 2019 - Aug 2020

- Enhanced stored procedures for **AWS PostgreSQL** using PL/pgSQL to orchestrate batch processing of data migrated from on-premises to cloud environments, ensuring efficient data transfer and transformation.
- Architected and executed ETL pipelines leveraging **AWS Glue jobs** integrated with **PySpark** for intricate data transformation tasks, from a data lake(S3) to a **AWS Redshift** , RDS handling datasets upto 2TB size.
- Designed and implemented Data pipelines using **Informatica Cloud** (IICS) for seamless cloud migration to AWS **MySQL**, resulting in a 30% enhancement in data flow efficiency and a 25% reduction in processing time.
- Performed **Data profiling** to analyze source tables and Managed metadata throughout the ETL process, including defining source-to-target mappings, transformation rules, and documenting **data lineage**.

PROJECTS

- Developed an NLP-driven **Chatbot** using **OPENAI** models, enhanced with Pinecone's similarity search. [Try it!](#)
- Developed sentiment analysis project using **PySpark ML** in Azure environment with **Azure Data Factory** for orchestration and **DataBricks** for scalable processing. [Have a Look!](#)

EDUCATION

University of South Florida

Masters in Business Analytics and Information Systems

Tampa,Florida

May 2024

National Institute of Technology

Bachelors in Mechanical Engineering

NIT,Calicut

June 2015 - May 2019

TECHNICAL SKILLS

Cloud Technologies: AWS services, Azure DataBricks ,Snowflake, Informatica, Docker, Power BI, Tableau

Programming Languages: Python, SQL, Shell scripting, Scala, Pyspark,Java Script, Node.js

ML Architectures: [Machine Learning](#) [Text Analytics](#) [Tech of AI](#) [NLP](#) [Deep Learning](#) [Data Science Programming](#)

Demonstrated Capabilities: [Big Data](#) [PySpark](#) [Stastical Programming](#) [MLops](#) [NoSQL](#)

Databases: Postgres, MySQL, MongoDB, Hadoop Eco system.

CERTIFICATIONS

[AWS Data Engineer Associate](#)

[Data Engineering using AWS](#)

[TensorFlow Developer](#)

[DevOps](#)

[AWS Cloud Solutions Architect](#)