Harika Moole

USA | (919)-656-7543 | mharikareddy1507@gmail.com | GitHub | LinkedIn| Portfolio

EDUCATION

University of North Carolina at Charlotte **Aug 2022 – Dec 2023**

Master of Science in Computer Science

Coursework: Big Data Analytics, Cloud Computing, Database Systems, Visual Analytics, Machine Learning

TECHNICAL SKILLS

Programming Languages:	Python, SQL, Java, Bash
Databases:	PostgreSQL, MySQL, MongoDB, MS SQL Server, Oracle
Data Warehouses:	Google Cloud BigQuery, Amazon RedShift, Snowflake
Big Data Technologies:	Apache Spark, Apache Kafka, Hadoop, Hive, Scala
ETL Tools:	Talend Data Fabric, Informatica Big Data Management, SSIS
Cloud Services:	Amazon Web Services (AWS), Microsoft Azure, Google Cloud Platform (GCP)
Tools:	Databricks, Azure Synapse Analytics, Tableau, Docker, Kubernetes, Git, Agile, Linux
Certifications:	AWS Academy Cloud Foundations, Machine Learning Foundations, Data Analytics

PROFESSIONAL EXPERIENCE

Graduate Research and Teaching Assistant – UNC Charlotte – United Health Group **Jan 2023 – Dec 2023**

- Built a data pipeline in Azure Databricks for processing, analyzing simulation data for a surgical robot development project and transferred the data to the downstream users.
- Analyzed the JSON data from **MongoDB** and selected fields required for analysis by using **Databricks** for efficient data processing, employing Python for data extraction, transformation, and loading (ETL) processes.
- Employed **Apache Kafka** for real-time data streaming and batch processing techniques to ensure handling large volumes of data, timely data ingestion and analysis.
- Developed and maintained **data models** in Parquet format within **Azure Delta Lake**, optimizing data storage and access for analytical processing by saving $16,000-$24,000 annually.
- Utilized **Azure Data Factory (ADF)** to automate, schedule data workflows, using ADF triggers and Databricks' built-in scheduler, reducing manual data handling by **90%.**
- Mentored 25+ students in Big Data using Amazon S3, **RedShift**, Glue, EMR, Athena, Lambda, SageMaker for machine learning model deployment, building predictive models and fine-tuned transformer-based models for sentiment analysis.

Data Engineer Senior Analyst – Accenture **May 2022 – Jul 2022**

- Executed Python functions in **AWS Lambda** to perform quality checks on the raw data in S3, such as column size, row count, data types, and key column checks, achieving a 50% reduction in data processing errors.
- Managed direct loads into the **AWS S3 data lake** and handled data type conversions and slowly changing dimensions (SCD) loads within **Informatica BDM**.
- Built interactive visualizations in **Tableau** to communicate insights to executive-level leadership.

Data Engineer Associate Consultant – Virtusa **Jul 2018 – Apr 2022**

- Migrated critical datasets from an Oracle database to **Snowflake**, utilizing Azure Data Factory (ADF) for efficient ETL processes, data scalability, query performance and data orchestration.
- Engineered and implemented **data pipelines** in Azure Databricks, leveraging **Python** and **SQL** to replicate and enhance SAS-based reporting logic in the cloud, ensuring seamless transition to a cloud-based data warehousing solution.
- Led the migration of **ETL processes** from Data Stage to Talend, improving data processing efficiency by **40%** and reducing job execution time, enabling more timely data analysis and decision-making.
- Worked with **Hive** on top of Hadoop for executing SQL queries and manage large datasets residing in distributed storage.
- Built re-usable libraries in **PySpark** to enhance data quality by **30%** and optimize downstream analytics workflows.
- Developed **Data Ingestion** tool (Data Lake as a Service) that handles full and incremental load using Talend, Tool ingests diverse data objects (Oracle, MongoDB, **PostgreSQL**, CSV, XML files) into Data Lake built on Amazon S3.

ACADEMIC PROJECT EXPERIENCE

Real Time Human Action Recognition | (Python, Tensor Flow, PyTorch, NumPy, OpenCV) **Aug 2023 – Dec 2023**

- Executed a real-time human action recognition system using a pre-trained TimeSformer model, achieving over 90% accurate classification of 60 different actions from the NTU RGB+D dataset. (Github)

FridgesenseApp | (AngularJS, Azure, Python, OpenAI, Firebase, Material UI) **Aug 2023 – Dec 2023**

- Developed a web application that takes user inputted text and submits it to a natural language processing model. The NLP model returns recipes to the web application at the frontend for customers to use for meal ideas. (Github)