Sandhya Mamadi

+1 (816)-291-0021 | sandhyamamadi@gmail.com | linkedin.com/in/sandhya | github.com/mamadisandhya

EDUCATION

University of Missouri - Kansas City — Kansas City, MO
Master's in Computer Science — *Aug 2022 - Dec 2023*

Bhoj Reddy Engineering College For Women — Telangana, India
Bachelor's in Information Technology — *Jun 2016 - Sep 2020*

TECHNICAL SKILLS

Cloud Technologies: Azure Data Factory (ADF), Databricks, Azure SQL database, Azure Cosmos DB, Azure Event Hub, Logic Apps, Function Apps, Azure Data Lake Storage (ADLS) Gen2, Azure Blob Storage, Azure HDInsight, Azure Synapse Analytics
Big Data Technologies: Hadoop, Spark, HDFS, Kafka
Databases and Data warehouse: MS SQL Server, MySQL, MongoDB, Snowflake
Languages: C, JavaScript, Java, Python, SQL,PL/SQL, shell script
Tools: Git, Visual Studio, PyCharm, IntelliJ, Eclipse, Informatica, Airflow, Power BI, tableau
Libraries: Pandas, NumPy, Matplotlib
Methodologies: Agile

CERTIFICATIONS

- Microsoft Certified Azure Data Engineer Associate | Link
- Snowflake-The Complete MasterClass Certified | Link
- 2X SQL Certified

EXPERIENCE

Data Engineer — Mar 2021 - Jul 2022
Tata Consultancy Services — *India*

- Developed and maintained data pipelines using Azure Data Factory to ingest, transform, and load data from various sources into Azure Data Lake Storage,enhanced data processing speed by 50 percent and reduced latency by 25 Percent.
- Designed and implemented data models for efficient storage and retrieval of structured and unstructured data in Azure SQL Data Warehouse.
- Migrated data from on-premises servers and legacy databases to Delta Lake on Databricks, enhancing data quality, and significantly reducing data latency for analytical workloads.
- Implemented basic transformations in Databricks using pyspark and pySQL scripts.
- Created Pyspark scripts for ETL process and utilized a variety of task operators to execute different types of tasks creating a DAG cycle using Python Airflow library.
- Implemented incremental data loading strategies to minimize downtime during migrations, resulting in 30 percent reduction in migration time.
- Collaborated with DevOps team to integrate data migration processes into CI/CD pipelines using Azure DevOps, enabling automated deployments and ensuring continuous delivery of data solutions.
- Experience involves handling diverse file formats such as JSON, XML, CSV, and Parquet.
- Proficient in developing PowerBI reports and dashboards and used version control tools such as GIT, tracked issues on platforms like JIRA.

Junior Data Engineer — Jan 2018 - Feb 2021
PeopleCluster — *India*

- Designed and implemented end-to-end data pipelines using Azure Data Factory to enable efficient ETL processes from various data sources into Snowflake data warehouse.
- Deployed Azure Data Lake Storage for scalable data lake management, implementing partitioning and retention strategies for raw and processed data
- Developed real-time data streaming into Snowflake via Azure Event Hubs and Functions for reliable data ingestion.

- Orchestrated data collection processes using Apache Spark and Azure Databricks, increasing processing speed by 16 Percent.
- Implemented Snowflake features like Time Travel for data versioning and conducting point-in-time queries.
- Reduced data inconsistencies by 90 Percent through robust replication with Snowflake and other platforms via Azure Data Factory.
- Expertise in Creating dashboards and reports through Tableau.
- Managed a diverse customer base, demonstrating good communication, organizational, and problem-solving skills.

PROJECTS

Olympic Data Analysis | *Azure Services, Databricks, PowerBI*   Jan 2024 - Feb 2024
- Orchestrated end-to-end development of a scalable data pipeline utilizing Azure Data Factory, Azure Databricks, and Power BI for Olympic data processing and analysis.
- Implemented data transformation techniques within Azure Databricks, including handling missing values and deduplication, ensuring data integrity and reliability.
- Developed interactive dashboards and reports in Power BI to visualize athlete performance trends, medal standings, and sports popularity.

Tourists Behavior Analysis | *Python, Jupyter Notebook*   May 2023 - Jun 2023
- Objective of the project is to analyze the tourist's behavior and forecast future tourism demands.
- Implemented time series modeling techniques to create a monthly count of tourists and seasonal patterns.

House Price Prediction | *Python, GCP, Jupyter Notebook*   Jan 2023 - Feb 2023
- Developed a Python regression model utilizing linear regression techniques to accurately predict house prices based on area.
- Successfully deployed the regression model on the Google Cloud Platform (GCP), enabling efficient and scalable predictions.
- Leveraged Python programming and Jupyter Notebook for model development and analysis.