Swathi Gaddampally

swathigaddampally13@gmail.com, Mobile: +1(816)-699-9147, LinkedIn

Summary:
Data Engineer around 2 years of experience, armed with MySQL, Python, PySpark, and Azure Data Services. I thrive on wrangling complex datasets, crafting efficient pipelines, and optimizing databases to extract actionable insights. Driven by a Master's in Computer Science. Seeking for full-time roles.

Skills:
Programming Languages: C, Python – (PySpark, Pandas, Matplotlib, Numpy)Java , Java Script, PHP.
Database: MySQL, MongoDB
Others: Git, Data Structures and Algorithms, HTML, CSS, Bootstrap, Microsoft Excel, Microsoft Power Point, SSIS, Microsoft Azure Data Services (like ADF, ADB, Azure Synapse, Azure Datawarehouse, Power BI), SSIS Development, ETL Processes.

Work Experience:
Accenture (Full Time) May 2020 – Dec 2022

- Led the design and implementation of ETL data pipelines using Azure Data Factory, Azure Databricks, PySpark, and SSIS, enhancing data accuracy and processing efficiency. Deployed these solutions via Azure DevOps, streamlining operations.
- Orchestrated the seamless migration of on-premise applications to Azure Cloud, leveraging continuous integration and deployment (CI/CD) practices to improve system reliability and operational efficiency.
- Translated complex datasets into actionable insights using Power BI, enabling data-driven decision-making through comprehensive statistical presentations.
- Developed and optimized end-to-end data flow processes within Azure Data Factory, from extraction to reporting, including data copying to SQL Data Warehouse and executing stored procedures for data integrity.
- Managed database servers, performing configurations and maintenance to ensure performance stability. Utilized Azure Synapse Analytics for SQL scripting and data reconciliation, validating data from source to Power BI dashboards.
- Utilized Databricks Notebooks for data issue identification and resolution in large datasets, performing data cleansing and schema validation to maintain high data quality.

Projects:
Heart Disease Data Analysis:
I collected data from Kaggle and analysed personality traits associated with heart disease using Python libraries like Pandas, NumPy, Matplotlib, and Seaborn. The study was conducted in Jupyter Notebook focusing on common factors leading to the disease.
Movie Recommendation System (Azure):
I created a movie recommendation system that analyses user preferences to suggest relevant movies. This involved gathering movie data, setting up a CI/CD pipeline in Azure Data Factory, and using Power BI for visualizing insights and offering personalized movie suggestions.
YouTube Data Analysis in Azure:
• Gathered daily trending YouTube video data spanning several months from Kaggle.
• Processed and standardized the raw data using an Extract, Transform, Load (ETL) system.
• Securely stored both the processed and raw datasets in an Azure Data Lake, serving as a centralized repository.
• Worked with Azure services such as Azure Active Directory (Azure AD) for security, Azure Data Factory for data integration, Azure Functions for serverless computing, and Azure Synapse Analytics for interactive querying.
• Leveraged Microsoft Power BI to design interactive dashboards for addressing essential inquiries.

Education:

Master of Science in Computer Science – GPA: 3.33/4.0.
University of Central Missouri, Lee's Summit, MO, United states. Jan 2023 – May 2024.
Bachelor of Technology (B. Tech) in Computer Science and Engineering – CGPA: 3.72/4.0.
Malla Reddy College of Engineering for Women, Hyderabad, India. Aug 2017 – Jul 2021.