LAVANYA REDDY T

+1 330-983-6613 | lavanyart7@gmail.com | https://www.linkedin.com/in/lavanyareddyt/ | https://github.com/Lavanya-97

Professional Experience

Data Engineer, Fidelity Investments, Boston, USA (August 2023- Present)

- Optimized a legacy streaming data processing pipeline in Azure Event hub to reduce **processing time by 40%** optimizing efficiency.
- Designed and deployed 3+ streaming data processing pipelines leveraging Azure Data Factory for data cataloging and schema evolution, alongside Azure Logic Apps for orchestrating the real-time data processing workflows, facilitating streamlined analysis of streaming data.
- Utilized Azure Functions, VM, and Event Grid for scheduled data transformations, cutting pipeline **run-time by 2 hours weekly.**
- Implemented continuous aggregates in Azure Synapse Analytics to optimize real-time data analysis, with scheduled refresh policies for streaming updates and automated data archival for efficient storage management.

Data Engineer, TCS, Hyderabad, India. (February 2022 – August 2022)

- Implemented Azure Database Migration Service (DMS) to migrate financial data incrementally from Oracle database to Azure Cosmos database, reducing dependency on legacy systems by 80% and improving scalability and accessibility of financial data.
- Utilized Python scripts to extract financial data from REST APIs **expanding the scope by 50%** and providing deeper insights into market trends and investment opportunities.
- Developed a suite of management-focused Power BI dashboards offering insights into financial headcount data. Leveraged Azure Data Factory for data cataloging and Azure Functions for orchestration enabling automated ETL pipelines.
- Implemented Agile methodology, specifically Scrum for project management, resulting in regular collaboration and iteration among cross-functional teams to deliver incremental value to stakeholders.

Data Engineer/Modeler, Pfizer, Hyderabad, India (November 2018 –January 2021)

- Developed an ETL data pipeline utilizing Alteryx queries and tools like Azure Data Factory, GitHub, Power BI, and Apache Kafka to automate the HCAHPS dashboard and incorporate 10+ operational KPIs, **improving patient satisfaction by 15%.**
- Established a robust data governance framework utilizing Azure Data Catalog, Azure Data Lake Storage, Azure Active Directory(AAD) to **enhance data quality** and accessibility for R&D initiatives **by 25%.**
- Engineered Azure Data Factory-based ETL workflows leveraging Python for advanced data manipulation, leveraging Azure Functions for seamless execution of transformation scripts, yielding actionable insights and **achieving a 15%** boost in customer satisfaction metrics.
- Designed and implemented a DataMart using Azure Synapse Analytics and optimized data retrieval with Apache Airflow, **achieving a 50% reduction** in Tableau reporting turnaround time by automating ETL processes and streamlining data workflow.

Big Data Intern, Prodapt, Hyderabad, India (September 2017 –November2018)

- Developed data ingestion workflows using Sqoop to collect and process real-time data from sources including ServiceNow, JIRA improving system monitoring and reducing incident **response times by 30%.**
- Assisted in optimizing Hadoop-based data storage solutions, utilizing HDFS and Hive to enhance the efficiency of storing and querying IT service management data resulting in a **20% reduction in processing time** for large datasets.
- Utilized Erwin tool to design data models, reducing database maintenance time by 30% and improving query performance by 20%
- Collaborated with senior data engineers to design and deploy MapReduce jobs for processing and analyzing large-scale datasets contributing to the generation of actionable insights for improving service delivery and operational efficiency.

Education

Master of Applied Science - Computer Science (August 2022 – Dec 2023)
Southeast Missouri State University, Missouri **GPA: - 3.83/4.0**
Bachelor of Technology **(**August 2014 – May 2018)
VNRVJIET, Hyderabad, India **CGPA: - 8.5/10**

Skills

Languages: Python, SQL**,**HTML,CSS, UNIX Shell Script, Scala
Technical: Machine Learning (supervised and unsupervised learning), Data Visualization, Big Data Analysis.
Cloud Platforms: Azure (Data Factory**,** Databricks, Logic Apps, Synapse Analytics, ADLS Gen 2) and AWS

Big Data Technologies: Apache Airflow, Apache Kafka, DataStage, Hadoop, Databricks, Apache Beam, SSIS, Databricks.
Tools: Docker, Jenkins, GitHub, Toad, Tableau, PowerBI.
Databases: Terraform, PostgreSQL, SQL server, Airflow, Mongo Db, , IBM DB2 **.**
Certifications: Azure Data Associate, Azure AI Fundamentals, Azure Fundamentals

Projects

Book Recommendation System

- Developed a Django-based application providing book recommendations using collaborative filtering and content-based filtering, customized to user preferences and search contexts.
- Utilized BeautifulSoup for scraping Goodreads website, obtaining book summaries and cover images to enrich user experience

Advanced Data Processing Pipeline - MongoDB, Docker, Kafka, and OpenShift

- Architected and deployed a robust and scalable data processing pipeline leveraging OpenShift, Kafka, Docker, and MongoDB.
- Managed the configuration and deployment of Kafka and MongoDB using OpenShift, ensuring high availability and efficiency.
- Optimized data ingestion and processing workflows into MongoDB, utilizing MapReduce and the Aggregation Framework for comprehensive data analysis.