VENKATA SAI MANI LAKSHMI KAVYA DARSI
+14706215442 • kavyadarsi3@gmail.com • LinkedIn • GitHub • Portfolio

EDUCATION

Syracuse University, School of Information Studies, Syracuse, NY   May 2024
Master of Science| Masters in Applied Data Science/Data Pipelines and Platforms
- Relevant course work: Data Warehouse | Advance Big Data Management | Big Data Analytics | Database Administration Concepts & Database Management | Text Mining | Natural Language Processing | Data Analysis and Data Modelling

Jawaharlal Nehru Technological University Kakinada, KITS, Guntur, AP   June 2021
Bachelor of Technology| Major in Computer Science and Engineering
- Relevant course work: Agile Methods | Java

TECHNICAL SKILLS

- **Frameworks & Big Data Tools**: SQL, NoSQL, Postgres SQL, Aurora, Redshift, MongoDB (Document Database), Kafka, Cosmos DB, Cassandra, Redis, Neo4j, Presto, Hadoop (MapReduce, YARN, HDFS), Apache Drill, Spark
- **Programming Languages**: Python, R, Java, HTML, Unix, C++, C#
- **Visualizations**: Tableau, PowerBI, Kibana, Looker
- **Cloud**: AWS (Redshift, Dynamo DB, Glue, Athena, QuickSight), Snowflake, DBT Cloud, Azure Synapse Analytics, Azure Data Store, Azure Data Lake Storage, Azure Data Factory, Azure HDInsight, Azure Databricks, Kubernetes, GCP
- **Office Productivity Tools**: MS Excel, MS Word, MS PowerPoint, MS Access
- **Certifications**: Amazon Cloud Practitioner, Azure Data Engineer(DP-203), Tableau Desktop Specialist

PROFESSIONAL EXPERIENCE

Sports Data Engineer, Syracuse University, Syracuse, NY   June 2023 - August 2023
- Developed a pipeline transforming live game streams into **JSON** for real-time sports analytics on a **MongoDB**-based web interface, providing fans with immediate, simplified box scores
- Led in designing and implementing **NoSQL** database schemas, ensuring data integrity, and organizing for real-time, complex queries on a leading sports website
- Leveraged **Apache Spark** for efficient post-game data updates, overcoming Spark limitations by innovating a method to accurately update player and team stats
- Skilled in leading complex data migrations between **NoSQL** and **SQL** databases, maintaining data consistency and reliability across systems

Data Engineer, DXC Technology, Hyderabad, TS   August 2021 - July 2022
- Designed **ETL** pipelines using **Azure Data Factory** to extract data from on-premises sources and load it into **Azure Data Lake Storage**, improving data integration efficiency by 30%
- Implemented data ingestion processes to ensure seamless data flow from on-premises systems to the cloud, maintaining data integrity and security, and reducing data latency by 25%
- Deployed **Azure Databricks** and **Spark** to transform data stored in Azure Data Lake Storage, optimizing it for analysis and reporting, resulting in a 40% improvement in data processing speed
- Collaborated with analytics teams to provide transformed data for generating business insights and reports, leveraging **Azure HDInsight** for processing and analysis, increasing report accuracy by 20%

ACADEMIC PROJECTS

New York Taxi Data Analytics Pipeline Utilizing GCP and Looker   May 2024
- Managed the design and implementation of an **ETL** pipeline on **Google Cloud's Compute Engine** and **MageVM,** processing New York taxi data and significantly enhancing data processing efficiency
- Utilized **chart.io** for data model design and **Google Colab** for implementation, creating a scalable **BigQuery** data warehouse and developing comprehensive **Looker** dashboards to provide real-time insights into ride patterns and customer behavior
- Directed a two-member team, optimizing data processes and modeling practices, aligning efforts with business objectives, and promoting a collaborative and innovative team environment

Stack Overflow Data Transformation Using Power BI   December 2023
- Orchestrated a **Kimball**-inspired **ELT process**, converting a large Stack Overflow dataset into an analytical **Power BI** dashboard to reveal user patterns and operational insights
- Built and managed a Snowflake-based data pipeline, merging data **from Azure Blob Storage** and **Minio** to a central cloud warehouse, enhancing data accessibility
- Applied **DBT Cloud** for sophisticated data modeling of Stack Overflow information, and crafted **PowerBI** reports for strategic stakeholder insights, supporting Snap's analytical needs