**VAMSI TARIGOPULA**

Dallas, TX | 424 (444)-1314 | vamsitarigopula97@gmail.com | linkedin.com/in/vamsitvk/

**TECHNICAL SKILLS**

**Languages:** Python, R, SQL, PL/SQL, SAS.

**Databases:** Microsoft SQL Server, MongoDB, MySQL, Oracle, Snowflake.

**Cloud:** AWS Redshift, Azure Data Lake, Azure Synapse Analytics, Databricks.

**Cloud Computing:** AWS (EMR, EC2, S3, RDS, Athena, Lambda, Glue), Azure Data Factory.

**Data Analysis/BI:** Excel, Tableau, Power BI, Looker.

**Big Data:** Apache Hadoop, Apache Hive, Spark, MapReduce.

**Agile & Project Management:** JIRA, Confluence, SharePoint, ServiceNow.

**WORK EXPERIENCE**

**SR. Data Analysis,** Dataquad Inc, *Houston, TX*Aug 2022 - Present.

* Collaborated with Business Analysts to understand business needs and develop data-driven solutions.
* Developed and implemented **Azure ETL** pipelines (**PySpark**) to extract, transform (dimensional model), and load data (**Azure Data Lake Gen2**) for efficient data warehousing, reducing data loading time **by 45%**.
* Extracted relevant data from **Azure Data Lake Storage** into **PySpark** data frames for further exploration and visualization to identify insights and build prediction models.
* Conducted data collection, data cleaning, feature scaling, feature engineering, visualization, and data manipulation using **Python libraries like NumPy, Pandas, and Seaborn.**
* Created **Tableau** dashboards, **Snowflake** views, and **Python** modules as part of my responsibilities.
* Exposure to NoSQL databases **MongoDB** and **Cassandra**.
* Gathered user requirements through interaction with stakeholders and end-users. Documented user guides and release notes for clear and effective tool usage.
* Use the knowledge of Agile / Scrum methodology as a development process.

**Data Engineer,** Dataquad Inc, *Houston, TX*Jul 2021 - Aug 2022

* Worked as part of a cross-functional team that included Data scientists and Data Analysts to deploy models and work on various projects.
* Developed a data analytics solution on **AWS**, leveraging Amazon **Redshift** for data warehousing. Built data models and implemented reporting dashboards using Tableau for effective communication.
* Created **ETL** jobs using **AWS** **Glue** to extract vendor data from diverse sources. The process included data cleaning, imputation, and mapping, with final results stored in **S3** buckets for analysis using **AWS** **Athena**.
* I also build continuous integration and development pipelines for machine learning models using **Jenkins and Git Actions.**
* Enhanced the execution of current **Databricks** processing using Apache **Spark**.
* Created SQL tables with referential integrity and developed queries using **SQL** and **PL/SQL.**

**Data Engineer,** CMS IT Services, *Bangalore, India*Aug 2018 - Jul 2019

* I was responsible for the Extraction Transformation and Loading (**ETL**) data from multiple sources into Data warehouse.
* Automated data processing and schema validation using **Python**, achieving a **50%** reduction in processing time and eliminating the need for manual data checks, which improved data quality and efficiency.
* Created **Tableau** Dashboard for the top key Performance Indicators for the top management by connecting **various data sources** like Excel, Flat files, and SQL Database.
* Using **Python** and **R** to implement **machine** algorithms to predict and forecast data for better results.

**KEY PROJECTS**

**Facebook Data Analysis**

*Tools and Technologies: Big Data platform, Hadoop ecosystem (HDFS, Yarn, MapReduce, Hive), Apache Hive, HiveQL*

* Developed Hive tables to integrate millions of Facebook records into the Hive warehouse, enabling efficient queries and analysis.
* Leveraged insights from the massive dataset to optimize content and user engagement, resulting in an 85% increase in engagement and a 35% boost in click-through rate.

**Batch Processing Data Pipeline for Sales Data**

*Tools and Technologies: Spark, Apache Airflow, Python, AWS EMR, AWS S3, AWS Redshift, Docker Compose, PostgreSQL*

* Designed and implemented a batch processing pipeline using Spark to classify and load vast sales data from PostgreSQL and CSV files into AWS Redshift for advanced analytics.
* Utilized Docker Compose for containerization, enabling rapid deployment, portability, and consistent pipeline execution.

**CERTIFICATION AND EDUCATION**

**Microsoft Certified: Azure Data Engineer Associate** Nov 2023

**Master of Science (MS) in Applied Computer Science**  Aug 2019 -Jul 2021

Southeast Missouri State University (3.7/4.0)