

Varun Muppalla

716-936-4719 | vmuppalla1210@gmail.com | [LinkedIn](#) | [GitHub](#)

TECHNICAL SKILLS

Languages: Python, SQL, PySpark, Java, R, MATLAB

Technologies: AWS, Azure, Databricks, Snowflake, dbt, Apache Airflow, SQL databases(MySQL, PostgreSQL), NoSQL databases (MongoDB, DynamoDB), Apache Spark, Apache Kafka, Spring Boot, Django, Docker

BI Tools: Tableau, QuickSight, PowerBI

PROFESSIONAL EXPERIENCE

Kaiburr LLC

Feb 2024 – Present

Data Engineer

San Jose, CA, USA

- Led the development of multiple ELT pipelines by collaborating with cross-functional teams of engineers and stakeholders to process and deliver data to various teams.
- Developed Python ingestion scripts to process unstructured application data (JSON) from Azure Blob Storage and MongoDB and load to a Snowflake data warehouse. Orchestrated these scripts using Apache Airflow.
- Programmed and deployed SQL data transformation scripts using dbt (data build tool) and GitHub Actions to create data sets for data science and analytics teams.
- Built a REST API using Django and Azure Functions to deliver aggregated data to app development teams.

Yahoo Inc.

Jun 2023 – Aug 2023

Data Engineer Intern

New York, NY, USA

- Constructed ETL pipeline to process security and vulnerability data from 5 data sources and generate analytics reports using AWS Databricks and Tableau.
- Developed a decision tree to rank the criticality of internal security vulnerabilities, reducing average incident response time from 3 days to 4 hours.
- Refactored data transformation logic and converted Splunk queries to PySpark, reducing ingestion times by 60%.

ZS Associates

Feb 2021 – Aug 2022

Data Analyst Engineer

Pune, MH, India

- Led the development of the SCA-CM (supply chain analytics - contract manufacturing) data analysis module, resulting in an 18% reduction of contract manufacturing costs.
- Architected an ETL Pipeline using AWS Databricks to fetch and compile data from 3 massive geographical data sources (APIs, EDWs, flat files) and generate a 700 GB dataset used by 6 teams.
- Built a big data pipeline using AWS Databricks, and Tableau leading to a reduction of operating cost of a manufacturing supply chain by \$1.1 million.
- Refactored complex SQL queries to PySpark, eliminating data inconsistencies and execution timeout issues.
- Developed a GitLab CI/CD pipeline for migration of deployment of software components.

Winspire Technologies

Feb 2020 – Nov 2020

Business Intelligence Intern

Bangalore, KA, India

- Refactored and optimized 50+ SQL queries; improved system performance and dashboard refresh times.
- Developed dashboards in Power BI to visualize customer metrics for usage on an internal tool.
- Programmed data transformation and aggregation jobs on AWS Glue based on business requirements.

PROJECTS

Olist E-commerce Data Analysis | AWS Lambda, AWS Glue, AWS S3, AWS Redshift, Tableau

[\[GitHub Link\]](#)

- Engineered serverless E2E data pipeline using AWS Lambda, AWS Glue, AWS S3 to fetch and clean data from APIs and store in AWS Redshift data warehouse.
- Developed a Tableau dashboard to visualize sales and logistics metrics for 3 years and derive key business insights.

Buffalo Crime Analysis | Apache Airflow, Docker, AWS Redshift, Athena

[\[GitHub Link\]](#)

- Designed and deployed a data pipeline using Apache Airflow fetch Buffalo CitiStat data from APIs and scrape, clean, and load data to AWS Redshift warehouse.
- Developed comprehensive crime reporting analytics for APD by visualizing key crime metrics. Delivered actionable insights to reduce crime rate in multiple neighborhoods.

EDUCATION

State University of New York at Buffalo

Aug 2022 – Jan 2024

Master of Science in Data Science

Buffalo, NY, USA

Vellore Institute of Technology

Jul 2017 – Jun 2021

Bachelor of Technology in Information Technology

Vellore, TN, India