**VISHNU VARDHAN MANDULA**

Jersey City, NJ | (551) 328-5322 |[vishnuvardhanmandula@gmail.com](mailto:vishnuvardhanmandula@gmail.com)| [LinkedIn](http://linkedin.com/in/vishnuvardhanmandula)

# Education

**Stevens Institute of Technology, Hoboken, NJ.**  **Dec 2023**

Master of Science in Computer Science **GPA: 4.0/4.0**

Relevant Coursework: Data Engineering, Advanced Algorithms, Data Mining, MLOps, Statistical Methods

## CVR College of Engineering, India

Bachelor of Engineering in Computer Science Engineering (CSE) **GPA: 7.8/10.0**

Relevant Coursework: Data structures and Algorithms, Operating Systems, Object-oriented Programming, JAVA, DBMS

**Technical Skills**

**Programming Languages:** Python (pandas, NumPy, PySpark), SQL, Java

**Data Engineering Tools:** Apache Airflow, Azure Data Factory, Snowflake, AWS (EC2, S3), Azure (Synapse, Databricks), Docker

**Databases**: SQL (MySQL, Azure Synapse), NoSQL (MongoDB), Data Warehousing (Snowflake, BigQuery)

**Big Data:** PySpark, Kafka, ETL/ELT Pipelines

**MLOps**: MLflow, CI/CD (GitHub Actions), Model Monitoring (Elasticsearch, Kibana)

**Cloud Platforms:** AWS, Azure

# Work Experience

**Qorvo, USA Jan 2024 - Oct 2024**

**Data Engineer**

* Built automated ETL pipelines using PySpark and Azure Databricks, reducing data processing time by 25% and enabling real-time analytics for 5+ data science teams.
* Migrated 10TB+ data from Azure Blob Storage to Synapse Analytics via Azure Data Factory, improving query performance by 40% for financial reporting.
* Designed scalable data infrastructure in Azure (Databricks, Data Lake) to handle 50M+ daily records, ensuring 99.9% pipeline uptime.

**JST Technologies, India Jan 2019 - May 2022**

**Data Engineer**

* Built a distributed data ingestion framework using Apache Kafka and Python to process 5TB+ daily streaming data from IoT devices, reducing latency by 35% and enabling real-time analytics for 10+ client dashboards.
* Optimized Snowflake data warehouse performance by designing star-schema models and tuning queries, reducing average query runtime from 12s to 3s and cutting monthly compute costs by 20%.
* Deployed the data pipeline on Amazon EC2 instances, ensuring high availability and efficient execution for future analytics needs

# Projects

**Olympic Data Analytics | Azure End-to-End Data Engineering Nov 2024 – Dec 2024**

* Designed and deployed an Azure-based ETL pipeline ingesting 50+ years of historical Olympic data (20M+ records) from Azure Blob Storage, transforming it using PySpark in Azure Databricks, and loading it into Azure Synapse Analytics for analysis.
* Optimized data modeling by implementing a star schema in Synapse, reducing query runtime by 35% and enabling seamless integration with Power BI for stakeholder dashboards.

## YouTube Data Analytics Pipeline | AWS Data Lake & ETL Oct 2023 – Nov 2023

* Designed a scalable AWS data pipeline ingesting 10,000+ daily trending videos (CSV/JSON) from Kaggle across 15+ regions, centralizing raw data in S3 and transforming it into analytics-ready formats using AWS Glue and Lambda.
* orchestrated serverless ETL workflows with AWS Glue, reducing data preparation time by 35% and partitioning data by region/category for efficient querying in Athena.

## Stock Market Real-Time Data Engineering Project               May 2023 - June 2023

* Built a real-time stock market data pipeline using Apache Kafka and AWS (EC2, S3, Glue, Athena) to process 10,000+ stock ticks/minute, enabling low-latency analytics for traders and investors.
* Deployed Kafka brokers on AWS EC2 for distributed data ingestion, achieving 99.9% uptime and fault tolerance during high-volume trading hours.

## Twitter Data Pipeline | End-to-End ETL with Airflow & AWS Jan 2023 – Feb 2023

* Built an ETL pipeline to extract 10,000+ tweets/day using the Twitter API, transform raw JSON data using Python (pandas, regex), and load processed data into Amazon S3 for analytics.
* Orchestrated workflows with Apache Airflow deployed on AWS EC2, automating daily data ingestion and reducing manual intervention by 20 hours/month.
* Enabled sentiment analysis by integrating processed data with a downstream ML model, providing actionable insights for brand monitoring.