Vineet Jujjavarapu

+1(240) 461-0053 https://www.linkedin.com/in/vineet-jujjavarapu/ vineetj5@terpmail.umd.edu
https://github.com/vineetj5

SUMMARY

Data Engineering professional with **3+ years** of experience designing scalable **data pipelines**, **ETL workflows**, and cloud-based architectures. Skilled in **Python**, **SQL**, **API** integration, and Agile collaboration to ensure data quality, performance, and reliable insights.

EDUCATION

University of Maryland, College Park **College Park, Maryland, USA**
Master's in Data Science Aug 2024 - May 2026
Relevant Coursework: Statistical Modeling, Big Data Systems, Data Visualization, Machine Learning

AREAS OF EXPERTISE - Data Engineering, Data Analysis, Machine Learning, Data Warehousing, API Developer, NLP

TECHNICAL SKILLS - Python, Java, Spring boot, Spark, Hadoop, Power BI , SQL, Azure, Databricks, FAST API, LangChain, Neo4j, Docker, CI/CD pipelines

EXPERIENCE

Viatris **Canonsburg, Pennsylvania, USA**
Supply chain Data Intern June 2025 – Aug 2025

- **Designed and deploying an end-to-end ETL process** to extract, transform, and load SAP procurement and logistics data into a centralized warehouse, powering a **Power BI analytics layer** for real-time insights into **vendor performance, and logistics anomalies**, thereby **reducing operational risks** and improving data-driven planning decisions.

Lam Research Corporation **Bengaluru, India**
Data Engineer July 2022 – Aug 2024

- Designed and deployed **data-driven APIs** using **Java, Spring boot and Azure**, improving **data availability** and **accessibility** for process engineers to **efficiently analyze and manage** semiconductor recipe files.
- Developed internal microservices using **Python** and **FastAPI,** optimizing the ingestion and retrieval of over **500,000** equipment recipe data entries in semiconductor manufacturing, improving data processing speed by **40**%.
- Developed and optimized **ETL** pipelines in **Azure Databricks** using **Python** to process **1M+ wafer recipe records**, integrating internal APIs for **automated data extraction**, file handling, and parent-child recipe relationship detection. Improved runtime efficiency by **30%** and reduced manual overhead in semiconductor process simulations.
- Wrote and **optimized SQL queries** for API-based file uploads, metadata extraction, and download functionalities; enhanced performance of **analytical operations** on large datasets used in semiconductor process simulations.
- Implemented **CI/CD pipelines** in **Azure DevOps** to **automate** the deployment of **APIs**, ensuring seamless integration, faster releases, and improved reliability of semiconductor data processing solutions.

Data Engineer Intern Jan 2022 – June 2022

- Designed and deployed **backend services** for a **customer-facing application** utilized by process engineers in the **USA**, leveraging **Java, Spring Boot, and REST API's** to streamline **data retrieval** and improve **workflow automation.**
- Implemented a file re-ingestion feature with a timestamp-based filtering mechanism, allowing selective re-processing of files originally stored in **HDFS**. Utilized **Apache Spark** for **data processing** and **Oozie** for **job scheduling**, capturing the job status of processed files.

PROJECTS

Real-time Bitcoin Graph Analysis with RAG-Enabled NLP Chatbot| langchain,Neo4j,docker,llama,FastAPI,Py2neo

- Designed a **real-time data pipeline** using **WebSockets, Py2neo, and Neo4j** to model Bitcoin transactions as a property graph; containerized the system with **Docker** and served insights via **FastAPI.** Built a **Retrieval-Augmented Generation (RAG) chatbot** using **LangChain** and **LLaMA**-based **LLM** to generate **Cypher queries** from natural language, enabling dynamic graph querying and NLP-based fund flow summarization.

Personalized Coursera Recommendation systems Using NLP| sklearn,pandas,Word2Vec,NLTK,TF-IDF,PCA,Clustering

- Developed a complete Course **Recommendation System** using **web scraping** for extracting the data and **NLP** techniques like **TF-IDF** and **Word2Vec embeddings** to analyze and compare courses based on their content. Extracted **key insights** on industry trends and in-demand skills to provide **personalized recommendations.**

Automated Stock Data ETL and predictive Feature Engineering Pipeline using Apache Airflow

- Built a **Dockerized ETL pipeline** using **Apache Airflow and Astro** to automate **incremental stock data ingestion, transformation, and loading** into **PostgreSQL**, with integrated **logging, validation checks, and email alerts** for production reliability.

PUBLICATIONS

Medium Articles on Data Science and Analysis (https://medium.com/@jvineet50**)**

- Authored technical blogs exploring AI in stock analysis, sports prediction, data engineering ETL projects, and data storytelling using ML, NLP, and visualization techniques.