Nitesh Thota

nthota2@gmu.edu | (703) 656-3582 | Fairfax, VA
https://www.linkedin.com/in/nitesh1012/
https://nitesh1012.github.io/niteshthota.github.io/

Education

M.S. in Computer Science   Aug 2023 - May 2025(Expected)

George Mason University, Fairfax, VA

GPA: 3.8

Relevant Coursework: Software Engineering, Component-Based Software Development, DevOps Practices, Database Systems, Natural Language Processing, Data Mining, Analysis of Algorithms

B.Tech. in Electronics and Communication Engineering   Aug 2017 – Apr 2021

SRM Institute of Science and Technology, Tamil Nadu, India
GPA: 3.5
Relevant Coursework: Algorithms, Data Structures, Database Management Systems, Computer Networks

Technical Skills

Programming Languages: Java, Python, C, C++, JavaScript, TypeScript
Web technologies & Frameworks: HTML, CSS, SpringBoot, Spring MVC, Microservices, Node.js, React, Angular
Machine Learning: TensorFlow, PyTorch, OpenCV, Scikit-Learn
Data Processing and Analysis: Apache Spark, Apache Hive, Apache Hadoop, Apache Kafka, NumPy, pandas, SciPy
Cloud Computing and DevOps: Amazon Web Services (AWS), Google Cloud Services (GCS), Kubernetes, Docker, Jenkins, Kubernetes, Ansible, Terraform
Tools, Software and OS: Android Studio, Git, Tableau, Postman, VS Code, IntelliJ, Eclipse, Linux, Windows
Databases: MySQL, Oracle, Postgres, SSMS, MongoDB

Professional Experience

Modak Analytics | Hyderabad, India   July 2021 - July 2023
Role   Software Development Engineer
Client   Humana Inc

- **Engineered, designed, and implemented** robust **automated ETL processes** using **Python** and **Java** to streamline **data ingestion, transformation, and hydration** across diverse data sources including **SSMS, MySQL, Big Query, GCS, and Oracle**. Delivered seamless **data delivery** to destinations like **HDFS, ADLS, and Google Cloud Platform (GCP)**, reducing **pipeline latency by 20%**.
- **Optimized complex data pipelines**, improving **performance and scalability** to handle **50% larger data volumes**. Utilized **industry-standard file formats** such as **Parquet, Avro, and CSV** for efficient storage and retrieval in **big data environments**.
- Contributed significantly to the development of **advanced data analytics dashboards**, utilizing tools like **Tableau** and **Python** to enable **dynamic reporting and actionable insights**.
- Enhanced existing codebase to comply with **Greenlight API (GLAPI)** benchmarks, significantly improving **CI/CD pipeline efficiency by 15%** and reducing deployment times using **Azure DevOps**.
- Developed and automated Python scripts to manage the unzipping and secure transfer data to **Google Cloud Storage (GCS)**, ensuring reliable daily execution through scheduled Cron jobs.
- Developed a scalable **Python-based Google Cloud Function**, triggered by **Pub/Sub messages**, capable of processing millions of CSV entries from fragmented files. Achieved **40% faster file processing** and consolidated data into comprehensive files uploaded to **GCS buckets**.
- Designed and implemented logging scripts to maintain comprehensive records of the Humana data engineering platform's operations, enhancing monitoring and troubleshooting capabilities.
- Contributed to **web application development** for hosting the **Nabu product**, delivering tailored solutions to Humana that met their business and operational requirements.

Academic Projects

Dynamic Web Application for User Data Processing

- Developed a **scalable web application** with feedback form pages that processed and stored user data in **MySQL, Oracle, MongoDB, and PostgreSQL** based on user preference.
- Built with **Spring Boot, Node.js, and React, using AWS for hosting** and **Terraform** for infrastructure management.
- Implemented a **Jenkins CI/CD pipeline** for automated testing and deployment, with **Kafka** for **real-time data streaming**.
- Ensured seamless deployment and data management across environments, aligning with business goals and data privacy standards.

PatriotPilot: Virtual GMU Guide

- Designed and developed a **chatbot** that provides quick and accurate information about George Mason University's resources using **Retrieval-Augmented Generation (RAG)**.
- Integrated **NLP techniques** with **FAISS** for efficient context retrieval and **Qwen-2.5-14B-Instruct** for response generation, ensuring highly relevant answers to user queries.
- Implemented **dynamic preprocessing** and overlapping chunk-based retrieval to handle the complex structure of university data, **improving retrieval accuracy by 20%**.
- Recognized as the **best project in class**, showcasing innovative solutions for university-level query answering systems.