Nitesh Thota

Fairfax, VA | (703) 656-3582 | thotanitesh5@gmail.com | LinkedIn | GitHub | Website

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, Web Development

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, 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**

Client - Humana Inc. Role - Software Development Engineer

- **Designed, developed, and optimized automated ETL pipelines** using **Python, Java, Apache Airflow**, and **Apache Spark**, reducing data ingestion and transformation time by **20%**. Integrated diverse data sources such as **SQL Server (SSMS), MySQL, BigQuery, GCS, and Oracle** while ensuring efficient data delivery to **HDFS, ADLS, and Google Cloud Platform (GCP).**
- **Enhanced data pipeline performance and scalability** to process **50% larger data volumes**, leveraging **Parquet, Avro, and CSV** for optimized storage and retrieval in **big data environments**.
- **Developed real-time data streaming solutions using Apache Kafka and Google Cloud Pub/Sub, achieving 40% faster structured and semi-structured data processing.**
- **Automated CI/CD processes** using **Azure DevOps, Jenkins, and Terraform**, improving deployment efficiency by **15%** and ensuring compliance with **Greenlight API (GLAPI) benchmarks**.
- **Implemented RESTful APIs with Spring Boot and Flask**, integrating with **AWS Lambda** and **Google Cloud Functions** to support scalable backend services, improving API response time by **25%**.
- **Developed scalable Python-based Google Cloud Functions**, triggered by **Pub/Sub messages**, to process millions of fragmented CSV records. Achieved **40% faster file processing** and automated ingestion into **GCS buckets**.
- **Automated data processing workflows** by scheduling **Cron jobs** to unzip, validate, and transfer large datasets to **Google Cloud Storage (GCS)**, ensuring **100% job reliability**.
- **Deployed containerized applications** using **Docker and Kubernetes**, ensuring microservices scalability and high availability in cloud environments like **AWS, GCP, and Azure**.
- **Optimized complex SQL queries and NoSQL database operations**, improving data retrieval performance for **MongoDB and PostgreSQL** by **30%**.

Academic Projects

Survey Management System – Full Stack Web Application **Sep 2024 - Dec 2024**

- Built a scalable survey management application using **Spring Boot** for the backend and frontend frameworks (**Vue.js and Angular**) to create interactive and responsive **user interfaces**.
- Implemented **RESTful APIs** and integrated **JPA/Hibernate** with **MySQL** for efficient storage, ensuring seamless **CRUD operations**.
- Unified deployment by integrating the front and backend into a single Spring Boot JAR, containerized with **Docker** and hosted locally for verification.
- Optimized a **CI/CD pipeline** using **Jenkins** for automated testing and enhanced scalability with **Kafka** for data streaming and **Terraform** for infrastructure management.

PatriotPilot: NLP-Based University Chatbot **Sep 2024 - Dec 2024**

- Developed and integrated a chatbot for George Mason University's online resources using a **Retrieval-Augmented Generation (RAG)** framework, deployed as a pop-up on a prototype GMU CS website for seamless user interaction.
- Scraped, structured, and preprocessed data from university websites into JSON pairs, optimizing preprocessing to reduce noise and retain semantic relevancy.
- Embedded text using **E5-Large-v2**, stored embeddings in a **FAISS index**, and fine-tuned **Qwen 2.5 14B Instruct LLM** with instruction-response pairs.
- Achieved recall and **precision scores >70%**, ensuring accurate retrieval and response generation.