Sushanth Rangu

(704)363-9144 • sushanthr131@gmail.com • linkedin.com/in/sushanthrangu/ • github.com/sushanthrangu

EDUCATION

University of North Carolina at Charlotte, Charlotte, USA — Aug 2023 – May 2025
Master of Science in Computer Science

- **Relevant Coursework:** Network Based Application Development, Algorithms and Data Structures, Software System Implementation and Design, Big Data, Visual Analytics, Intelligent Systems

Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, India — Jul 2018 – Jun 2022
Bachelor of Technology in Information Technology

- **Relevant Coursework:** Data Structures, Database Management Systems, Algorithms, Object Oriented SD

PROFESSIONAL EXPERIENCE

Accenture — Jul 2022 – Aug 2023
Software Engineer — Hyderabad, IND

- Developed **Java + Spring Boot microservices** with **gRPC/REST APIs** supporting 5K+ daily users on distributed platforms.
- Designed **GPU-accelerated test automation pipelines** integrating **xUnit + CUDA-based simulations**, reducing regression times **40%**.
- Configured **Keycloak** for identity federation enabling **secure SSO** across vendor-facing applications.
- Deployed **Prometheus + Grafana monitoring** improving platform uptime **25%** and reducing incident MTTR **40%**.

INTERNSHIP

UNCC ONEIT — May 2025 – Aug 2025
Software Engineer — Charlotte, NC

- Built a **Golang-based distributed routing engine** optimized for **50+ GPU-backed nodes**, cutting response latency **18%** using **CUDA C++ kernels** for accelerated graph computations.
- Integrated **Azure OpenAI APIs** to enable **real-time ETA predictions** leveraging **GPU-optimized inference pipelines**; boosted routing intelligence for **10K+ concurrent sessions**.
- Scaled microservices on **AKS with GitOps** achieving **99.99% uptime**; automated deployments via **Terraform + GitHub Actions** improving CI/CD efficiency **30%**.
- Processed **2M+ transit records** using **PyTorch + TensorRT FP16** inference, reducing compute overhead **35%** for large-scale routing analytics.

SELECTED TECHNICAL PROJECTS

Face Mask Detection and Person Identification

- Designed a **real-time detection system** using **VGG-16 + FCN** achieving **98.5% accuracy** across 500+ NVIDIA Jetson devices.
- Leveraged **TensorRT FP16** optimizations to accelerate on-device inference by **30%** while reducing memory footprint **25%**.
- Implemented custom **CUDA kernels** for optimized image preprocessing pipelines.

Accessible Event Management Platform (WCAG-compliant)

- **Developed** an **event management platform** using **React, Node.js, Express, and MongoDB** with dynamic data handling.
- **Built** a fully responsive **four-page interface** for event listings, registrations, and user management.
- **Implemented** secure **user authentication** with **JWT + OAuth2**, ensuring safe credential storage and access control.
- **Engineered WCAG-compliant accessibility features** with **ARIA + SSR + VoiceOver**, improving usability for all users.

CurryExpress – Scalable Ordering & Delivery System

- Engineered a **geospatial routing backend** with **Kafka + Kubernetes** achieving **15ms latency** across **50+ nodes**.
- Integrated **GPU-accelerated LLM models** via **Azure OpenAI APIs**, enabling predictive ETA generation in **real-time environments**.

TECHNICAL SKILLS

- **Languages:** C++, Python, Golang, Java, CUDA C/C++, SQL
- **AI & GPU Acceleration:** TensorRT, PyTorch, TensorFlow, CUDA Kernels, OpenCV DNN, Azure OpenAI
- **Frameworks:** gRPC, Spring Boot, Node.js, .NET Core, Express.js
- **Cloud & DevOps:** AWS, Azure, Docker, Kubernetes, Terraform, ArgoCD, GitHub Actions
- **Databases:** PostgreSQL, MongoDB, MySQL, Redis
- **Parallel & Distributed Systems:** GPU Optimization, CUDA Streams, Multithreading, Microservices

ACHIEVEMENTS and CERTIFICATIONS

- **AWS Certified Cloud Practitioner** – Oct 2024
- Practicing data structures and algorithms regularly (**300+ LeetCode problems**); solid grasp of system design principles
- Awarded **Certificate of Appreciation** for winning the Teams Challenge at **Accenture Technology Analyst School**
- Continuously exploring **GPU inference optimizations**, CUDA, and next-gen model parallelization techniques.