Mrunaldhar Bathula

+1 540-429-3619 mrunaldharb23@vt.edu linkedin.com/in/mrunaldhar-bathula

EDUCATION

Virginia Tech **Virginia, USA**
Master's in Computer Science *Aug 2023 – present*
Relevant Coursework: Distributed Systems, Machine Learning, Data Structures, Statistics, Web Development.

Indian Institute of Technology, Roorkee **India**
Bachelor of Technology *Jul 2017 – May 2021*

Experience

Software Development Engineer [Link] **Aug 2024 – Present**
Proctor360 *Richmond, VA*

- Led the integration of **YOLOv10**, replacing AWS Rekognition. Directed a team to optimize model size and inference speed through pruning and quantization on **AWS UltraClusters**, achieving a 90% cost reduction.
- Architected a fault-tolerant AWS backend with **Python**, **SQS**, **DLQ**, **Lambda**, **MongoDB**, **S3**, handling up to **1 Million requests** daily with dynamic provisioned concurrency auto-scaled via **CloudWatch** event triggers.
- Developed a serverless **RESTful** microservice for resource validation using **Python**, **Go**, **Lambda**, reducing deployment failures by 30% and integrating with **AWS DynamoDB** and **S3** for data persistence.
- Designed and automated cloud infrastructure provisioning using **Terraform** and **AWS CloudFormation**, achieving a 60% improvement in infrastructure scalability and reducing manual deployment efforts by 50%.
- Collaborated with the CTO and team to design and develop a Manifest v3 **browser extension** using **Javascript**, enabling real-time tab monitoring and API-driven session management. The extension supports over **30,000 users**.

Software Development Engineer [Link] **Nov 2022 – Aug 2023**
Tiger Analytics *Chennai, India*

- Developed a **GraphQL**-based Model Tracking Service using **Python FastAPI** and **PostgreSQL**, enabling version control, metadata storage, and artifact management (**AWS S3**) for 50+ models.
- Built a real-time Data Drift Detection Service leveraging **Apache Kafka** and **Cassandra**, reducing latency by 40% with Databricks. Integrated alerting pipelines via **AWS SQS and SNS** for automated notifications.
- Containerized **8+ microservices** using **Docker** and orchestrated via **Kubernetes (EKS)**, reducing deployment overhead by 60%. Secured infrastructure with network policies and secrets management.
- Architected a Visualization Service using **Flask** and **WebSockets**, powering dashboards with real-time metrics.
- Designed and automated **CI/CD** pipelines using **Jenkins** and **GitLab CI**, achieving 95% automated test coverage **pytest** and zero-downtime deployments via blue-green strategies. Integrated **SonarQube** for code quality gates.
- **Mentored** a team of **5 developers**, standardizing coding practices and conducting code reviews, leading to a 40% reduction in troubleshooting time and ensuring seamless deployment processes.

Software Development Engineer **Jun 2021 – Nov 2022**
Sterlite Technologies *Pune, India*

- Designed a metrics collection framework for Kong **API Gateway** using **Python**, extending Lua plugins with **Prometheus** exporters. Reduced latency by 25% by offloading metric aggregation to async workers (**Celery**).
- Built a Kubernetes Operator in **Python** (Kopf) to automate Kong Gateway scaling, dynamically adjusting pod counts based on **gRPC** stream concurrency and Upstream API error rates. Reduced cloud costs by 25%.
- Optimized Kong's request processing by developing Python-based custom plugins for **JWT** validation and **rate limiting**, replacing Lua logic with **asyncio** driven workflows. Cut CPU usage by 30% while handling 20K RPS.

Projects

HookieEats [Link] **Aug 2024 – Sep 2024**

- Developed a meal recommendation system using multi-RAG architecture with **LlamaIndex**, **Pinecone**, vector embeddings, semantic search, and Azure OpenAI GPT-4o LLM. Engineered a serverless backend with **Azure Functions** and **MongoDB**, handling live streaming data. Optimized query response times by 40%.

Scalable AI Fraud Detection **Jan 2025 – Feb 2025**

- Developed a scalable AI pipeline for fraud detection in insurance documents using **PyTorch**, **FAISS** for similarity search, and ResNet-50 for feature extraction. Optimized inference with **ONNX** Runtime, **TensorRT**, and **Triton** reducing latency by 40% and deployed models on **AWS SageMaker**.

Efficient ABSA Using Pruned and Quantized Transformers **Aug 2024 – Dec 2024**

- Implemented transformer-based Aspect-Based Sentiment Analysis (ABSA) using **PyTorch** and **Hugging Face**, optimizing BERT and DistilBERT models through structured pruning and **TensorRT**-based quantization. Achieved a 20% reduction in model size while maintaining an F1-score of 0.80, enhancing computational efficiency and enabling scalable real-world deployment.

Technical Skills

Programming Languages: Java, Python, Go, Kotlin, C#, Rust, C/C++, JavaScript, TypeScript, SQL
Frameworks/Tools: Scikit-Learn, Flask, Django, PySpark, TensorFlow, PyTorch, Docker, Kubernetes, AWS, Azure, RAG, LangChain, FIASS, Kafka, RabbitMQ, SQL Server, DynamoDB, HDFS, Hive, MapReduce, Spark
Experience: Microservices, WebSockets, MongoDB, Cloud Systems, Image Processing, Mobile Development, Distributed Systems, Microservice Architecture, Computer Vision, Deep Learning, NLP

Achievements

Winner VT Hacks (Cloudforce x Microsoft) for contributing to business-focused AI-driven meal planning solutions. Promoted technical education by mentoring participants in GIT AI hackathon and **Judging** innovative solutions.