

HARIHARAN NAGASUBRAMANIAM

(929)678-7353 | hnnhariharan12@gmail.com | soester10.github.io/portfolio | linkedin.com/in/hariharan-naga/

EDUCATION

New York University

MS Computer Engineering | GPA: 3.89

New York, NY
August 2022 - May 2024

SRM University

B.Tech Computer Science and Engineering | GPA: 7.9/10.0

Chennai, India
July 2017 - June 2021

Coursework: Data Structures, Algorithm Design Analysis, Machine Learning, Computer System Architecture, Deep Learning, High Performance ML, Software Engineering, Internet Protocols, Network Security, Big Data

EXPERIENCE

Secure Systems Lab, New York University

New York, NY

Research Assistant | CodeQL, AST, Software Engineering

June 2023 – Present

- Examining 'atoms of confusion' classifiers for **software bugs**, analyzing results from **3 static code analysis tools** for evaluating coverage accuracy.
- Generating **Abstract Syntax Trees** for codebases in **C/C++** by conducting **statistical analyses** across open-source repositories such as Git, Emacs, and Linux, achieving over **500%** speed improvement from current tool.
- Enhancing existing classifiers using **CodeQL**, demonstrating a **20%** increase in accuracy and coverage.

Dun & Bradstreet

Chennai, India

Data Scientist | Python, SQL, Spark, Pandas, Selenium, Databricks

August 2021 – July 2022

- Developed and scaled **multiprocessing web crawlers**, automating the extraction of ESG-relevant data for **2.5 million** companies, covering over **60%** of the company's internet-based data.
- Consolidated and optimized **PySpark** workflows with a single efficient User Defined Function (UDF), accelerating Reason Code production by **30%** with the seamless integration of revised logic.
- Orchestrated the deployment and distribution of crawlers across numerous Google Cloud VMs for data extraction, yielding a **1-month time savings** for the company through resource optimization.
- Constructed an end-to-end **NLP pipeline** for the extraction of pertinent data from ESG reports and news for over **200k** companies in 15 global markets.
- Advised the integration of SIC (Standard Industrial Classification) method into production code to shortlist over **5 million** businesses engaged in controversial activities.

PROJECTS

Show Of Hands | Django, JS, AWS, Travis CI/CD, PostgreSQL, S3, Redis

- Developed a **social application** with polling and user interaction capabilities, facilitating discussions and user connections through friend requests, garnering participation from over **500 users**.
- Engineered a robust **chat system** with **Redis Cache memory** to facilitate **real-time user communication** with a **50% reduction** in latency; implemented secure caching and accurate message delivery for group chats.
- Incorporated **Ajax** techniques to enable **live content updates** within webpages, resulting in a seamless user experience and over **30%** performance improvement with optimized page loading times.
- Streamlined deployment processes with **AWS Elastic Beanstalk** and **Travis**, resulting in over **80%** reduction in deployment time; leveraging **PostgreSQL** for scalable management of user profiles and interaction history.

Fire Chat | React, Flask REST API, Firebase, AWS

- Architected a **discourse platform** for movie enthusiasts to access comprehensive data for over **1.5 million** movies and tv shows, including reviews and user opinions, establishing a central hub for movie insights.
- Paginated data storage using Firestore's **NoSQL** structure, resulted in **2x faster** read operations; real-time content updates with **React web hooks** and maintained data security with **Firebase Security Functions**.
- Implemented **Flask REST** framework for the backend, hosted on **AWS EC2**, and **React** frontend on **Firebase**.

BRViT | PyTorch, Computer Vision, ViT

- Designed an end-to-end **Vision Transformer** for Bokeh Rendering, achieving **18%** improvement in subject clarity.
- Integrated **open-source** DPT model for architecture foundation, eliminating the need for a separate depth estimation model, resulting in a **7x faster** computation time compared to other models.
- Achieved **state-of-the-art** results on EBB! Dataset with **1.32%** SSIM metric improvement from existing benchmark.

SKILLS & CERTIFICATIONS

Programming Languages: Python, JavaScript, MATLAB, C/C++, TypeScript, Java

Tools & Frameworks: Django, Flask, FastAPI, React, PyTorch, TensorFlow, Docker, Terraform, Kubernetes, Kafka

Databases & Platforms: SQL, PostgreSQL, Spark, MongoDB, GraphQL, AWS, Google Cloud, Azure, Linux/Unix

Certifications: Google IT Automation Certificate (Coursera, Jun21), Microsoft Certified: Azure Fundamentals (May21), TensorFlow Developer Certificate (Coursera, Sep20)