

Dhanesh Baalaji Srinivasan

ds7636@nyu.edu | +1 9292932179 | [Linkedin](#) | [Github](#) | Seattle, WA (Open to relocate)

EDUCATION

New York University, New York City, NY; Master of Science in Computer Science; **GPA:** 3.75/4; **Honors:** Merit-based scholarship

TECHNICAL SKILLS

Programming Languages: Python, C, C++, C#, JavaScript, SQL. **Full Stack:** Django, .NET Core, .NET Framework, Angular, HTML, CSS. **Cloud & DevOps:** AWS, Azure, Docker, Kubernetes. **Databases & Search:** PostgreSQL, MongoDB, SQL Server, OpenSearch, Elasticsearch. **Distributed Systems & Big Data:** Spark, Hadoop, MapReduce, Cassandra. **Tools:** Linux, Git, Postman. **ML & GPU:** PyTorch, CUDA.

WORK EXPERIENCE

New York University, New York City, United States

Jul 2025 - Present

Software Development and DevOps - Project Lind, NYU Secure Systems Lab | Python, Docker, C.

- Refactored a Python test runner to prevent inconsistent execution states by eliminating unsafe system calls, improving reliability.
- Redesigned test setup with a temp directory and auto-cleanup; removed redundant file operations, reducing I/O overhead by 70%.

LOCOMeX, Inc., New York City, United States (Remote).

Feb 2025 - May 2025

Software Engineer and MLOps Intern | Django, AWS Lambda, DynamoDB, RDS, Postgres, Python, Tailwind CSS, Bootstrap.

- Engineered a full-stack, low-latency autocomplete feature using AWS Lambda and RDS, improving search responsiveness by 70%.
- Built Django APIs and containerized ML models as serverless Lambda functions, enabling scalable, low-latency inference.
- Built a PDF reporting tool with visual analytics in Django, reducing manual report generation time from 1 hour to under 1 minute.

New York University, New York City, United States.

Jan 2024 - May 2025

Graduate Research Assistant - Brooklyn Application, Architecture, Hardware Lab | DARPA Project | C, Python, Assembly, ARM NEON.

- Integrated a Last-level Cache into a spectrum-sensing Processor simulator and created sweeps to obtain the optimal cache size.
- Modeled and introduced variable Common Bus delays to assess signal detection throughput under various latency constraints.
- Devised Power Spectral Density and Match filter kernels using ARM v8.2 NEON for real-time spectrum-sensing computations.

Graduate Teaching Assistant - High-Performance Machine Learning, Big Data | Pytorch, CUDA, C, Pyspark, MapReduce.

- Collaborated with professors to develop lecture materials on Pytorch Distributed training, CUDA kernels, and Apache Cassandra.

Psiog Digital Private Limited, Chennai, India.

Nov 2020 - May 2023

Software Engineer | Angular, ASP .NET Core, ASP .NET Framework, ASP .NET MVC, C#, SQL Server, Azure.

- Crafted RESTful APIs, designed SQL scripts, and responsive User Interfaces (UIs) resulting in a 30% increase in user engagement.
- Devised a scalable Bidding system that incorporated an AI voice assistant which generated 4000+ user registrations within a month.
- Fixed critical Extract, Transform, and Load (ETL) pipeline issues, saving \$200k in potential losses from data downtime.
- Managed CI/CD pipelines across multiple products, reducing deployment times by 25% and hence improving release frequency.

PROJECTS

LlamaLearn - Retriever-Augmented Generation (RAG) flows for Large Language Models (LLMs) | Amazon Web Services (AWS), Python.

- Architected a scalable RAG system using DPR for dense retrieval and NeuralHermes-2.5 (Mistral-7B) for generation, deployed via AWS EKS and ECR with OpenSearch for vector search and DynamoDB for user-specific metadata to enable personalized answering.
- Engineered a modular information retrieval pipeline featuring document chunking, DPR-based vectorization, and OpenSearch k-NN search, enabling low-latency, semantically accurate real-time question answering.
- Improved answer quality and reduced hallucinations by injecting top-k retrieved chunks into the LLM for context-aware generation.

Phintrest - An intelligent image search engine | Amazon Web Services (AWS), Python, Javascript.

- Prototyped an image search engine with semantic understanding using Rekognition and Lex, backed by OpenSearch and Lambda.
- Automated infrastructure via CloudFormation and CodePipeline, ensuring scalable deployments and zero-downtime releases.

Subreddit Recommendations and Sentiment Analysis on Reddit data | Pyspark, DistilBERT, TF-IDF.

- Analyzed 3.8M Reddit posts using PySpark, TextBlob, and DistilBERT to perform large-scale sentiment classification.
- Implemented a content-based subreddit recommendation system using TF-IDF vectorization and cosine similarity to rank subreddits by the volume of posts exceeding a semantic similarity threshold with the user query.

Dining Suggestions Chatbot | Amazon Web Services (AWS), Python, Javascript.

- Constructed a scalable microservices-driven web application's front end using AWS S3, integrating a chatbot interface using LexV2.
- Developed APIs with API Gateway, Lambda, and extracted data from Yelp, indexed it in Elasticsearch, and used Dynamo for storage.