Dhanesh Baalaji Srinivasan

ds7636@nyu.edu | +1 9292932179 | Linkedin | Github | New York City, NY (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

Languages: Python, Java, C#, SQL, C, C++, JavaScript. **Full Stack:** Django, .NET Core, .NET Framework, Angular, Tailwind CSS, Bootstrap.
Big Data: Hadoop MapReduce, Spark, Dask. **ML/AI:** PyTorch, CUDA. **Databases:** Postgres, SQL Server, MongoDB, AWS DynamoDB.
Vector Databases: AWS OpenSearch, Elasticsearch. **OS:** Linux, Windows. **DevOps:** Docker, Kubernetes. **API Testing:** Postman.
AWS Services: AWS EC2, S3, Lambda, RDS, SQS, Lex, Cognito, VPC, Textract, EKS, ECR, Rekognition, Codepipeline, Cloudformation.

WORK EXPERIENCE

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 RESTful APIs with Django and prepared normalized, high-performance PostgreSQL schemas for business-critical features.
- Built a PDF reporting tool with visual analytics in Django, reducing manual report generation time from 1 hour to under 1 minute.
- Containerized XGBoost sanctions risk prediction model as a serverless Lambda function, enabling scalable, low-latency inference.

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.
- Developed a Spanish language translation engine tailored from scratch, increasing activity among Spanish-speaking users by 45%.
- Delivered Technical presentations to help non-tech recruits gain a solid foundation in .NET Core and Software design patterns.

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.
- Crafted a chatbot powered by Amazon LexV2 for natural language processing, providing robust user interaction.