

Mukesh Durga

md6256@nyu.edu | +1 (929) 726-4108 | LinkedIn | GitHub | New York

EDUCATION

New York University

Master of Science in Computer Science | GPA: 3.93 / 4.00

May 2026

New York, NY, USA

- **Teaching Assistant (Graduate), Software Engineering:** Mentored students through backend system design, API development, authentication, testing & debugging; evaluated assignments and provided technical feedback on correctness, maintainability, and security.
- **Course Assistant, Algorithms:** Guided students on core data structures and algorithmic problem-solving, helping them reason through edge cases, complexity tradeoffs, and solution design while reviewing submissions for correctness and efficiency.

Lovely Professional University

B.Tech (Honours) in Computer Science & Engineering | GPA: 3.85 / 4.00

Jul 2023

India

EXPERIENCE

CAPGEMINI, *Software Engineer* | Bengaluru, India

Jul 2023 – Aug 2024

- Developed Java and Spring Boot microservices for geospatial fleet and logistics APIs on AWS, supporting location lookup, route data access, and operational workflows across 8+ backend services.
- Improved p99 latency on key API request paths by 31% by adding Redis caching, profiling slow service calls, and tuning PostgreSQL queries through indexing and query restructuring.
- Developed Kafka-based asynchronous workflows and gRPC service integrations to decouple distributed backend services, improving backend processing throughput by 26% and reducing failure impact across dependent services.
- Supported migration from EC2-hosted deployments to a self-managed Kubernetes environment by containerizing 6 Spring Boot services and defining deployment, rollout, and configuration manifests.
- Improved production reliability by expanding JUnit and Mockito test coverage, strengthening CI validation, and adding service-level monitoring and alerting, cutting recurring regressions by 20% across release cycles.

CAPGEMINI, *Software Engineering Intern* | Bengaluru, India

Feb 2023 – Jun 2023

- Built an internal AI-assisted support assistant that classified operational requests from marketing and analytics teams, then routed them to the right backend workflow using REST APIs for validation and structured responses.
- Developed Python and Hadoop ETL pipelines to move advertising operations data into PostgreSQL, improving reporting data access by 27% through schema cleanup, query tuning, and caching of frequent lookups.

HIGHRADIUS, *Data Intern* | Remote

Jan 2022 – Apr 2022

- Built data preparation and feature-generation components for a FinTech recommendation workflow, processing customer and transaction data into model-ready inputs surfaced through a React dashboard.
- Supported recommendation-model experimentation by preparing evaluation datasets, comparing model variants, and analyzing offline metrics and A/B test results that informed engagement improvements across tested workflows.

PROJECTS

Stagehand: Multi-Agent Workflow Builder with Live Execution Tracing

- Built a full-stack platform for designing agent workflows on a drag-and-drop canvas, where each run streams live over WebSockets, showing per-step latency, token cost, and tool decisions, with one-click replay to trace any failure.
- Engineered the orchestration engine, handling agent isolation, a message bus, and tool-calling with retry and fallback, then added a UCB bandit that routes each step to a cheap or strong model. Built with React Flow, FastAPI, PostgreSQL, and ClickHouse; adaptive routing cut token cost per workflow by 31% over 60+ runs.

Forecast Studio: Natural-Language Analytics with a SQL Verification Loop

- Built a full-stack app that turns plain-English questions into SQL against a connected Postgres or Snowflake database, runs them inside a read-only sandbox with timeouts and row limits, and verifies that the result answers the question before returning it.
- Designed a multi-step planning agent that decomposes the question, drafts schema-grounded SQL, runs it on a sample, and self-checks, backed by a semantic-dedup cache and per-query cost tracking. On the Spider benchmark, this raised execution accuracy from 57% to 76% over a single-shot baseline and caught roughly 1 in 5 queries that ran successfully but returned the wrong answer.

Onboard.dev: AI Codebase Onboarding with Dependency-Graph Analysis

- Built a tool that ingests any GitHub repo through a background job queue, maps how its modules depend on one another, ranks the ones that matter most, and generates a guided learning path with quizzes plus a Q&A agent that cites code at the line level.
- Parsed repositories into a dependency graph using tree-sitter, scored module importance with a PageRank-style algorithm weighted by commit frequency, and ran RAG over function- and class-level chunks. Built with Next.js, Monaco, D3, Python/FastAPI, PostgreSQL with pgvector, GitHub OAuth, and BullMQ; indexed repos up to 4,800 files in under 3 minutes, with cited answers hitting the correct lines 82% of the time on a hand-labeled set.

SKILLS & CERTIFICATIONS

Languages: Java, Python, C++, SQL, TypeScript

Backend & Full-Stack: Microservices, REST APIs, gRPC, WebSockets, Distributed Systems, System Design, FastAPI, React, Next.js

AI/ML & Agentic Systems: LLM Integration, RAG, Vector Search, Embeddings, Text-to-SQL, Multi-Agent Orchestration, Tool-Calling, Planning Agents, Adaptive Model Routing, LLM & Agent Evaluation, ML Pipelines, Online Inference

Data & Storage: Redis, PostgreSQL, pgvector, MongoDB, ClickHouse, Snowflake, Kafka, Hadoop, S3/MinIO

Cloud & DevOps: AWS (EC2, S3, RDS, Lambda, CloudWatch), Docker, Kubernetes, CI/CD (GitHub Actions, Jenkins), Git

Reliability & Observability: Monitoring, Alerting, Logging, Distributed Tracing, RCA, Load/Regression Testing

Certification: AWS Certified Cloud Practitioner