AANSHI MANAN PATWARI

aanshipatwari@gmail.com | Seattle, WA | linkedin.com/in/aanshi-patwari | aanshipatwari.netlify.app

PROFESSIONAL EXPERIENCE

Software Development Engineer | ***Amazon Web Services, Seattle, Washington*** **August 2025 - Present**

- Qualified a new storage-optimized EC2 instance type for the Kinesis storage fleet through multiple load testing with kernel-level memory tuning, achieving a 2:1 host replacement ratio that unblocked a 4.4 PB fleet-wide memory reduction program. Rebuilt the deprecated load-test harness along the way and documented it as the team's standard procedure.
- Delivered backend streaming infrastructure for 3 new AWS regions in under 3 weeks, owning capacity modeling, service role registration, load balancer and dynamic configuration deployment, host provisioning, and end-to-end verification to unblock 3 downstream teams for launch.
- Implemented configurable callback timeouts in a Rust–Java streaming client, replacing silently dropped requests with typed errors carrying request ID and elapsed time, and extended the fix across the FFI boundary and Java client while raising line coverage to 86%.
- Automated cleanup of 100+ stale environments across 6 service pipelines, cutting security compliance violations from 1,000+ to double digits and converting a multi-day, two-approval manual process into a reusable single-execution template.

Associate Software Engineer, GenAI | ***Fitizen, Chicago, Illinois*** **January 2025 - August 2025**

- Built a high-throughput recommendation engine serving 1K+ QPS at sub-50ms p99 latency using fine-tuned LLMs, Sentence-BERT embeddings, and OpenSearch ANN vector retrieval.
- Cut LLM inference costs by 50% and latency by 35% on AWS SageMaker through LoRA fine-tuning, INT8 quantization, and GPU-aware deployment strategies.
- Scaled distributed inference to 5K+ concurrent requests at 99.9% uptime using GPU batching, SageMaker Serverless, Lambda, and SQS/SNS pipelines with autoscaling and monitoring

AI Engineer Intern | ***DebateBase, Chicago, Illinois*** **June 2024 - December 2024**

- Deployed GPT-4o agent pipelines on Azure AI Services with Dockerized FastAPI microservices and GitHub Actions CI/CD, supporting 1K+ concurrent real-time requests and cutting deployment overhead by 40%.
- Built a production RAG chatbot with LangChain, Azure Cognitive Search, and FAISS using hybrid keyword-semantic retrieval, holding production error rates under 0.5% while reducing hallucinated responses by 25%.

Tech Developer Assistant | ***Edplus - Arizona State University, Scottsdale, Arizona*** **July 2023 - December 2024**

- Engineered a Python question-generation pipeline using graph-driven templating and combinatorial optimization, automating 100+ parametric algebra and calculus problems with 30% higher assessment diversity.
- Developed an AI education platform integrating REST APIs, a BERT-based NLP chatbot, and OpenCV/Tesseract OCR pipelines, reaching 92% accuracy on handwritten equation recognition.

PROJECTS

Minibase using Bigtable Architecture | Java, Bigtable, Minibase, B-tree

- Redesigned Minibase's Java architecture using Bigtable-inspired OOP principles to modularize heapfile, iterator, and DB components; implemented parallelized sort-merge algorithms and version control for job-specific query optimization.

Travoyager - Travel Recommendation Engine | Python, Django, SQLite, Machine Learning, Github

- Led a 9-member team in developing an AI-driven travel recommendation system using collaborative filtering and feature engineering; engineered a Django/SQLite full-stack app with interactive frontend for personalized itineraries based on user behavior analysis.

Share N Learn - Online Learning Platform | AWS (EC2, S3, RDS, Redshift), Java, Spring Boot, MySQL, GraphQL

- Engineered a scalable AWS cloud learning platform reducing downtime by 50%, integrated Redshift for 10k+ records/sec processing, and built Java Spring Boot web apps with caching and asynchronous processing to slash video load times.

EDUCATION

Masters of Science in Computer Science | Arizona State University **January 2023 - December 2024**
Relevant Courses: Software Testing, Data Mining, Database Management Systems Implementation, Machine Learning 4.00/4.00

Bachelor of Technology in Information and Communication Technology | Ahmedabad University **July 2018 - May 2022**
Relevant Courses: Data Structure and Algorithms, Object Oriented Programming, Software Engineering, Cloud Computing 3.52/4.00

TECHNICAL SKILLS

Languages: Python, Java, Rust, SQL, Shell Scripting (Linux/Unix)
Backend & APIs: Spring Boot, Django, Flask, FastAPI, REST APIs, Node.js, React.js
Cloud & Infrastructure: AWS (Kinesis, EC2, Lambda, DynamoDB, SQS/SNS, CloudWatch, S3, SageMaker), Docker, CI/CD (GitHub Actions), Infrastructure as Code (CDK)
Databases: MySQL, MongoDB, DynamoDB, PostgreSQL
Systems & Practices: Distributed Systems, Load & Performance Testing, Observability (Monitoring, Alarming, Logging), Incident Response, Git

AWARDS & RECOGNITION

2x New American University Scholarship Recipient(2023) | Outstanding Performance and Impact on Community Student Award(2022)