Harshil Jhaveri

Dubai, UAE | +1 5138884348/+971 508169076 | hjhaveri@usc.edu | LinkedIn | GitHub | Google Scholar

Full-stack and backend engineer with 4+ years of experience building large-scale, cloud-native, microservices-driven applications across distributed systems. Expert in Java, Javascript and Python with deep experience designing complex systems end-to-end—from architecture to production. Proven ability to build scalable APIs, distributed event-driven platforms, CI/CD pipelines, and secure, reliable infrastructure on AWS. Strong cross-functional collaborator who mentors teams, drives engineering best practices, and ships high-quality software at speed.

Skilled at translating complex requirements into clean design, writing optimized and secure code, and resolving production incidents in on-call environments.

EXPERIENCE

ORACLE CLOUD INFRASTRUCTURE, SANTA CLARA, CA, USA | Software Engineer – II **July 2023-Present**

- Project lead for Zero Trust Packet Routing (ZPR) mechanism for Load Balancer Control Plane, reducing unwarranted access by 100% for thousands of enterprise customers through intelligent policy enforcement.
- Built LLM-based intelligent incident detection system using retrieval-augmented generation (RAG) to analyze operational logs, identify root causes, and recommend mitigation strategies—reducing mean time to resolution (MTTR) by 40% and automating diagnosis previously requiring manual investigation.
- Implemented a Data Warehousing Architecture for Load Balancer Service for assessment of security risks and business analytics.
- Increased system availability from 99.95% to 99.995% by architecting resilient systems with predictive failure detection, leveraging ML models to anticipate capacity issues before they impact customers.
- Lead automation of 2.5 weeks of developer work for monthly host OS patching through intelligent vulnerability detection and prioritization using ML-based risk scoring.
- Designed adaptive load balancing policies incorporating real-time performance predictions, optimizing resource allocation across distributed infrastructure.

CLOUDERA, SANTA CLARA, CA, USA | Software Engineer Intern **May 2022-August 2022**

- Created services to bolster authorization and authentication for SSL interactions with Hadoop clusters on Apache Knox.
- Automated client certificate authentication and configured composite provider authorization. Executed a header-based identity maintenance plugin, and authorization bearer services, and deployed packages to production via Gerrit.

FINOVATE CAPITAL, MUMBAI, INDIA | Software Development Intern **June 2021-July 2021**

- Introduced AI based generation of invoices, and transfers to and from FTP servers. Reduced manpower required by 300%, and increased auditing accuracy twofold.
- Led E2E backend and frontend stack development of an interactive web app used by MSME's for supply chain financing.

CMS COMPUTERS INDIA, MUMBAI, INDIA | Data Science Intern **December 2019–January 2020**

- Pioneered a Vision-based traffic control system, detecting and estimating vehicle speed from live feed transitioning from traditional ultrasonic sensors. It also characterized automobile's coordinates and type at offence time.
- Executed automatic red signal violation detection and penalty generation with 87% accuracy trained on a YOLO model.

ATOM TECHNOLOGIES INDIA, MUMBAI, INDIA | Software Development Intern **June 2019-July 2019**

- Developed a full stack E-Commerce platform to offer purchase and online sale of payment models to 150,000+ vendors.

SELECTIVE PROJECTS

1. **securebot.io**

- Built a security first MCP marketplace, evolving to an automated MCP server generation on demand.
- Built secure AI gateway platform that manages AI agent permissions and observability, enabling developers to secure agentic systems with 97% less instrumentation effort.
- Architected production system allowing AI agent security implementation in just 3 lines of code—dramatically simplifying the integration of authentication, authorization, and monitoring for autonomous AI systems.
- Designed intelligent permission management system for AI agents, implementing fine-grained access controls, policy enforcement, and real-time authorization decisions for agent-to-service interactions.

- Built comprehensive observability infrastructure capturing AI agent behavior, decision traces, and security events—enabling developers to understand and debug autonomous agent actions in production.
- Developed backend infrastructure using Python and TypeScript, implementing RESTful APIs, real-time event processing, and scalable microservices architecture deployed on cloud infrastructure.

2. **Smart Disaster Rescue and Mitigation System** (*Made using PyTorch, Angular, Docker, NodeJS and MongoDB)*

   Disaster Management Platform, with dedicated real-time chatroom for updates, offline SOS triggering, safest route computation to nearest police station, and toll payment through app. Available as a web and mobile app, and a dedicated API.

SKILLS

- **Backend & Systems:** Java, JavaScript, Python, Kotlin, Node.js, Spring Boot, FastAPI, Django, Flask, RESTful APIs, GraphQL, gRPC
- **Frontend:** React, Angular, TypeScript, Tailwind CSS, Redux
- **Databases:** PostgreSQL, MySQL, MongoDB, Redis, DynamoDB, Elasticsearch
- **Cloud & Infrastructure:** AWS (ECS, Lambda, S3, RDS, API Gateway, SQS/SNS), Docker, Kubernetes, Terraform
- **Distributed Systems:** Apache Kafka, RabbitMQ, Microservices, Event-Driven Architecture
- **AI/ML:** LLMs (GPT-4, Claude), RAG Systems, PyTorch, TensorFlow, LangChain, LangGraph, Vector Databases (Pinecone, FAISS)
- **DevOps:** GitHub Actions, Jenkins, CI/CD Pipelines, Prometheus, Grafana

EDUCATION

University of Southern California **August 2021–May 2023**
Master's in Computer Science (Coursework in Algorithms, Artificial Intelligence, Machine Learning, Databases, etc)

Dwarkadas J. Sanghvi College of Engineering, University of Mumbai **August 2017–June 2021**
Bachelor of Engineering, Computer Engineering, CGPA (9.67/10)

SELECTIVE RESEARCH EXPERIENCE

CHIP: Contrastive Hierarchical Image Pretraining

Few-shot, three-level hierarchical classification model for animal species using variational autoencoders.

LEADERSHIP & INVOLVEMENT

- **Apache Knox Open Source:** Improved auth providers and client certificate authentication.
- **Shrimad Rajchandra Love and Care:** Built and maintained main website, created visual dashboards to raise and track funds of over $3 million for Covid-19, created a portal to assess and store data about volunteering efforts worldwide