

Rafid Saad

TN Visa Eligible

Principal software engineer with 10 years of experience in building scalable, high-performance APIs and AI systems, driving rapid product development and growth for startups while ensuring low operational costs.

NYC, NY

(+1) 438 885 8021

rafid7776@gmail.com

EXPERIENCE

Unified Vision AI, NYC, NY — Staff Software Engineer

October 2023 - Present

- Partnered with leadership and cross-functional teams to define AI strategy, roadmap, and key technical decisions for next-gen AI products.
- Architected and led the development of a scalable Generative AI framework, increasing AI development speed by 2x and reducing model deployment time by 40%.
- Designed and integrated enterprise-grade AI chatbots capable of handling tens of thousands of requests.
- Designed and implemented AI ML ops - AI model evaluation, cost analysis, and result monitoring systems.

fastn.ai, Austin, TX — Co-Founder

JAN 2023 - October 2023

- Co-founded fastn to use generative AI to automate, and simplify building backend systems.
- Selected for YC interview S'23
- 3X improvement in AI-generated results: Designed and implemented generative AI backend using langchain. The results directly helped clinch multiple sales.

fabric.inc, Vancouver, BC — Lead Engineer

APRIL 2021 - JAN 2023

- Cross-organization leadership: Led CMS and integration teams, and ran API standardization, multi-tenancy, and cost optimization initiatives across
- Reduced failure rate by 5X: Introduced safer deployment and release strategies. Strategies and tools I set up were used to improve product resilience in multiple teams.
- Reduced operational cost 20X: Redesigned and built multi-tenant, microservice-based CMS. Reduced AWS bill by 70%, helped release team resources to focus on critical customer features.
- Reduced latencies by 10X (under 150ms): Using event-based design, read-optimized data, CDN and caching.

Veeva, Seattle, WA — Platform Team Lead

May 2019 - April 2021

- Increased speed of delivery: Updated the infra, and the code to make the system composable using Consul, software practices.
- Reduced deployment time: Automated on-prem deployments with Ansible
- Reduced failure rate: Introduced defensive programming culture - TDD using karate and cucumber, post-mortem culture

SKILLS

Problem-solving, communication

Design: High traffic systems, REST/GraphQL APIs, DynamoDB data modeling, cost optimization, microservices

Frameworks: Spring, Micronaut, NodeJS, LangChain

Monitoring: Datadog, Cloudwatch, Grafana, Influx DB, PagerDuty

Infrastructure: AWS, Kubernetes, Gitlab, Github CI/CD, Kafka, Kinesis

Databases: PostgreSQL, DynamoDB, Redis, Timestream, Elasticsearch

Testing: Karate, JUnit, Mockito, Cucumber

AWARDS/CERTIFICATES

AWS Solutions Architect 2023

RIT Merit Award 2019

Fulbright Scholar 2017

LANGUAGES

Java, Python, JS/TS, Terraform, AWS CDK

ShopDev, Lahore, PK — *Tech Lead*

June 2013 - May 2017

- Production-ready platform in < 6 months: Angular 2 + Node JS-based
- Increased system scale: Decoupled workflows using Kafka
- Reduced customer onboarding time: Simplified ETL processes with Talend

EDUCATION

Rochester Institute of Technology, Rochester, NY — *MS Computing Security*

Aug 2017 - Jun 2019

LUMS, Lahore, PK — *BS Electrical Engineering*

August 2009 - Jun 2013