

ARBAZ SIDDIQUI

Senior Software Engineer I
AI Infrastructure &
Distributed Systems

FIND ME HERE

+91-8963874828

arbaz00@gmail.com



SKILLS

- **Programming Languages:** Javascript, Golang, Python
- **Backend Frameworks:** Node.js, Gin, Django, Flask
- **Database:** MongoDB, Redis, Firebase, Neo4j, ElasticSearch, Postgres
- **Tools:** Kafka, MQTT, Media servers, Web3, Blockchain, FFmpeg
- **Cloud:** Docker, Kubernetes, AWS, GCP, ELK stack, Prometheus, Serverless
- **AI:** Diffusers, ComfyUI, n8n, LangChain, Salad

SELECTED HIGHLIGHTS

- Architected a distributed AI engine leveraging **Salad** for decentralized computing, achieving **2s image delivery** for diffusion models (Flux, Wan2.2) while **reducing operational GPU costs by 80%**.
- Built media servers that processed **50K daily streams**, averaging 33 minutes each, with **sub 5s latency**, resulting in 100GB ingestion.
- Implemented audio chat which saw a **traffic of 2M users monthly** with a total of **70M minutes in audio call time**.
- Maintained systems which were able to handle **traffic for a user base of 25M**.
- Integrated monitoring, logging, APM and telemetry to services for better visibility.
- Converted few services to serverless to support on-demand computing which helped us to **reduce the cost by 70%**.
- Interviewed **more than 100 candidates** for backend and DevOps roles.
- Active blog with over **10k monthly users**. Many of the articles are featured in Node weekly.

EXPERIENCE

Senior Software Development Engineer

Turnip

Jul 2020 - Present

- Built a scalable inference engine for diffusion models using ComfyUI and Docker; implemented a dynamic orchestration layer with n8n to handle hot-loading.
- Implemented media servers for mobile game streaming startup on Kubernetes for processing live streams via RTMP and FFmpeg.
- Architected a high-concurrency social platform with dynamic news feeds and 2nd-degree discovery using Neo4j with a distributed event-driven system built on Kafka, ElasticSearch, MongoDB, and Golang for real-time content indexing and engagement.
- Engineered a live audio platform featuring low-latency voice chat using Agora, gated room interactions and a robust real-time state management and messaging layer utilising MQTT, Firebase, and Node.js.
- Collaborated with web3 startups to develop a functional crypto exchange, a blockchain-based community website, and a card trading platform.

Lead Backend Engineer

Paytm Insider

Jun 2017 - Jul 2020

- Led team in developing and implementing robust backend for movie integration with online ticketing, capable of handling 500+ rps flawlessly.
- Designed and implemented logistics architecture which includes opting delivery, order tracking and internal tools for managing shipments.
- Developed the complete e-commerce flow including cart, transaction, inventory management, and taxation.
- Utilized AWS stack to implement microservice architecture (RESTful APIs, queues, lambdas, Kinesis) and deployed on Kubernetes.

PROJECTS

LRU cache for node

Cache manager with least-recently-used policy using doubly-linked list and hashmap for O(1) gets and sets. Featured in Node Weekly for its lightning-fast speed.

AnaBranch

An HTTP load balancer and reverse proxy written in Go. Supports multiple algorithms and health checks.

rSlashVideos

Web app that converts any subreddit into a video playlist, including filters such as hot, top of all time, and top of year.

Node Skeleton

A boilerplate for building node.js apps using the MVC model and best practices. Contains configuration for both local and social authentication, test suite and logging.