

Dinesh Bali

✉ dineshbali45@gmail.com ☎ +1-660-528-8938 📁 Portfolio **in** LinkedIn

EDUCATION

Northwest Missouri State University
Master of Science in Computer Science

January 2024 – May 2025
GPA: 4.0/4.0

EXPERIENCE

Founding Software Engineer

July 2025 – Present

Alpheva AI - Fintech startup

New York, USA

- Architected a real-time AI-driven financial platform using Go, MCP, and multi-agent pipelines, sustaining 100K+ concurrent users with P95 latency <120ms while delivering instant, personalized financial insights.
- Developed a sharded Redis Streams ingestion layer processing 1M+ events/day with durable delivery, feeding MCP-based agent workers via consumer groups and increasing end-to-end processing throughput 4x.
- Scaled high-throughput Go webhook ingestion for Plaid, Stripe, and Array with idempotency, batched writes, and optimized transaction flows, reducing database pressure by 60% while maintaining <100ms latency during peak webhook spikes.
- Built production RAG and agentic workflows using Pinecone and DynamoDB, enabling semantic retrieval, grounded responses, intent routing, parallel tool execution, and response aggregation, increasing personalized financial insight coverage by 75%.
- Optimized multi-agent financial reasoning through adaptive model selection, concurrent agent execution, caching, and Langfuse-based evaluation, reducing end-to-end reasoning latency by 83%.

Software Developer

May 2024 – May 2025

Northwest Missouri State University

Missouri, USA

- Delivered and operated production backend services in Go and Java (Spring Boot) across 6+ applications, leading a team of 8 engineers to ship reliable weekly releases.
- Migrated chat persistence to Kafka-backed workers, reducing P95 latency by 40% and increasing write throughput by 60%.
- Implemented Redis Pub/Sub fanout for WebSocket chat, supporting 5K+ concurrent connections and 10K+ messages/min while reducing message fanout latency by 35%.
- Established Prometheus and Grafana monitoring for Dockerized services, cutting incident detection time by 50% and resolution time by 30%.

Software Engineer

February 2023 – August 2023

Bobble AI - Series C

Haryana, India

- Engineered high-throughput Go microservices and a sharded Redis Streams pipeline with consumer groups, supporting 50M+ MAU and high-volume event processing.
- Implemented concurrent Go worker pools processing 1M+ events/day, improving resource utilization while maintaining predictable latency during traffic spikes.
- Designed resilient event-driven services with retries, backoff, and load shedding, reducing system errors by 30%.
- Improved MySQL performance through index optimization, query-plan tuning, connection pooling, and caching, cutting P95 latency by 45%.

Software Engineer

January 2022 – January 2023

Cognizant

Telangana, India

- Engineered and deployed highly scalable Spring Boot microservices on AWS, implementing concurrent processing, Redis caching, ETag validation, asynchronous processing, and performance optimizations, reducing API latency by 40%.
- Built secure RESTful APIs using Spring MVC, Spring Security, and Spring Data JPA, enabling authentication for 500K+ users and reducing average response time by 30%.
- Streamlined React-Redux front-end architecture with a centralized state store, boosting page-load speed by 25% and supporting 10K+ concurrent users.

TECHNICAL SKILLS

Backend: Golang (Echo), Python (FastAPI), Java (Spring Boot, Spring MVC, Spring Security, Spring Data JPA), Node.js (Express)

Distributed Systems: Event-Driven Architecture, Microservices, Kafka, Redis Streams, Redis Pub/Sub, WebSockets, gRPC, REST APIs, Webhooks, Concurrency, Idempotency

Databases: PostgreSQL, MySQL, Redis, DynamoDB, Pinecone

Cloud & DevOps: AWS (EC2, VPC, ALB, S3, RDS, SQS, ElastiCache), Docker, Terraform

Frontend: React, TypeScript, JavaScript, Redux, Next.js

Generative AI: Multi-Agent Systems, RAG, AI Orchestration, Agent Evaluation, Model Context Protocol (MCP), Vertex AI, Langfuse, Hugging Face

Developer & AI Tools: Git, Postman, Swagger, Prometheus, Grafana, Claude, Codex

Networking & Systems: Linux, TCP/IP, HTTP/HTTPS, DNS, TLS/SSL, Load Balancing, Reverse Proxies

ACHIEVEMENTS

- Owned end-to-end delivery of 3 production applications, including a consumer app that reached 50,000+ downloads and generated recurring ad revenue.
- Developed a UPI checkout verification flow by parsing bank/UPI SMS confirmations to validate payment status and reconcile against pending orders, reducing manual payment verification by 80% and confirming payments in <2s end-to-end.
- Led technical direction and system architecture at Alpheva AI, driving key design decisions, mentoring engineers, and owning end-to-end execution for a production AI platform.