

MANISH GANDOTRA

Gurugram, India ☎ +91-8800463103 ✉ manishgandotra@icloud.com

[in linkedin.com/in/manish-gandotra-b53413b8](https://www.linkedin.com/in/manish-gandotra-b53413b8) github.com/ManishgandotraCoder manishgandotra.com

SUMMARY

Senior Backend-Focused Full Stack Engineer with 8+ years of experience designing scalable SaaS platforms and distributed cloud-native systems across fintech, enterprise, and AI-driven environments.

Expertise in leading backend architecture, building event-driven systems, and scaling high-throughput platforms handling 1M+ records and 10K+ daily users.

Strong in async processing pipelines, AWS serverless systems, and AI-powered backend automation including LLM orchestration, Claude API integrations, and agent-based workflows using Ollama.

Progressed from full-stack development roles to leading backend architecture and distributed system design across enterprise and fintech platforms.

CORE SKILLS

Backend & Architecture: Node.js, TypeScript, Python (FastAPI, AsyncIO), REST, GraphQL, Event-Driven Architecture, Async Task Orchestration, Worker Queues, Circuit Breakers, Modular Monolith → Microservices

Cloud & DevOps: AWS (Lambda, ECS, EC2, SQS, API Gateway, S3, CloudFront, DynamoDB, RDS), Docker, CI/CD, CloudWatch, Autoscaling, Cost Optimization

AI & Intelligent Systems: LLM Integration, Claude API, Claude Code / Claude Cowork, Ollama (Local LLMs), LangChain, AI Agents (Tool Calling, Multi-Step Reasoning, Autonomous Workflows), Embeddings, LangChain, Semantic Search, Structured Output Enforcement, AI Workflow Orchestration, Prompt Management

Databases: PostgreSQL, MySQL, MongoDB, Redis (Indexing, Aggregations, Performance Tuning)

PROFESSIONAL EXPERIENCE

Termgrid

Feb 2024 – Present

Senior Software Engineer (Module Lead – AI & Full Stack)

Remote

Current Project - Termgrid (Fintech)

- Designed hybrid modular-monolith + serverless event-driven architecture improving scalability without microservice overhead.
- Built Python-based financial rule engine using AWS SQS enabling async validation workflows
- Implemented backpressure-aware worker pipelines with DLQ and retry strategies reducing system failures
- Reduced page latency from 900ms → 350ms via SSG, Redis caching, and query optimization
- Developed Lambda-based financial scoring services with horizontal autoscaling
- Segregated API and background workloads improving system stability under high load
- Implemented distributed tracing using CloudWatch for debugging async pipelines
- Built AI-driven financial summarization and recommendation engine using LLM orchestration (Claude API + structured outputs)
- Designed and implemented AI agents for financial workflow automation using tool-calling patterns and multi-step reasoning pipelines
- Mentored 6+ engineers and contributed to architecture review processes

Deloitte

July 2022 – Feb 2024

Senior Software Engineer

Remote

- Developed scalable backend APIs using Node.js and Python for enterprise data systems
- Designed indexing and aggregation strategies improving query performance by 32%
- Built async job processing pipelines for large-scale workflows
- Improved system reliability through monitoring, logging, and failure isolation
- Supported platforms serving 10K+ daily users
- Contributed to internal AI tooling experiments using LLM APIs and automation workflows

- Collaborated across teams and mentored junior developers on backend best practices

Cognizant

May 2021 – Jul 2022

Senior Software Engineer

Remote

- Built real-time monitoring platform using Next.js, Node.js, NestJS, and GraphQL supporting 5,000+ active sites
- Designed scalable backend systems with async processing for high-throughput data pipelines
- Implemented authentication and RBAC reducing access-related incidents by 40%
- Optimized infrastructure and workload distribution reducing cost by 42%
- Improved CI/CD pipelines and deployment stability
- Enhanced observability with structured logging and health monitoring

National Informatics Center

Jun 2017 – Apr 2021

Full Stack Engineer

Chandigarh, India

- Designed and developed scalable backend services using Node.js for government systems
- Built React-based frontend applications focused on performance and usability
- Implemented secure RBAC authorization, ensuring data protection and compliance
- Optimized APIs and database queries, improving system performance under load
- Contributed to architectural decisions and introduced modular design patterns
- Delivered mission-critical systems requiring high availability and reliability

AI PROJECTS

Verita H+ (HIPAA-Compliant AI Platform)

Tech Stacks : Node.js, FastAPI, AWS, Claude API, Ollama, Secure APIs,

- Designed and built HIPAA-compliant AI platform similar to OpenAI for secure medical use cases
- Implemented secure LLM interaction layer with data privacy, encryption, and access control
- Integrated Claude API and local LLMs (Ollama) for flexible and compliant AI processing
- Designed audit logging, request tracing, and data isolation for healthcare-grade compliance
- Built scalable backend architecture supporting secure multi-tenant AI usage

AI SQL Query Generator (Natural Language → SQL)

Tech Stacks : Python, FastAPI, PostgreSQL, React, TypeScript, Claude API / Ollama

- Built LLM-based query generator supporting joins, filters, aggregations
- Integrated Claude API and local models via Ollama for flexible inference
- Implemented validation layer to detect query errors before execution
- Generated structured JSON output for dashboards
- Enabled non-technical users to retrieve insights without SQL

AI Travel Itinerary Generator

Tech Stacks : Python, FastAPI, React, TypeScript, LLM APIs

- Developed AI engine generating optimized travel plans based on constraints
- Implemented budget-aware and preference-based recommendation system
- Generated structured day-wise itineraries via REST APIs

EDUCATION

Punjab Technical University

Bachelor of Technology in Computer Science

Aug 2013 - Jun 2017

Gurdaspur, India