

Gurpreet Gandhi Engineering Manager

+91 8374333790 — gurpreetgandhi3@gmail.com
[linkedin.com/in/gandhigurpreet](https://www.linkedin.com/in/gandhigurpreet) — gurpret.github.io

PROFESSIONAL SUMMARY

Engineering Manager and Platform Architect with **12+ years** of experience leading the architecture, engineering execution, and delivery of cloud native platforms, distributed systems, enterprise SaaS products, and AI platforms. Proven track record of building and scaling high performing engineering teams, driving technical strategy, owning platform roadmaps, and delivering secure, resilient, highly available, and cost optimized solutions at enterprise scale.

Experienced in leading cross functional engineering organizations of up to **14 engineers** across multiple Agile squads, partnering with Product, Architecture, SRE, Operations, and Customer Success to translate business objectives into scalable platform capabilities. Strong expertise in Platform Engineering, Kubernetes, Distributed Systems, Cloud Architecture, DevSecOps, MLOps, LLMOps, and Generative AI, with a passion for building & improving engineering teams, platform reliability, operational excellence, and AI adoption through reusable platform services and modern engineering practices.

PROFESSIONAL EXPERIENCE

Engineering Manager/Platform Architect

Apr 2020 – Present

Mobius Networks, Gaian Solutions

- **DevSecOps Platform, RunRun** Owned the platform architecture, engineering roadmap, and delivery execution for **RunRun**, a cloud native self service infrastructure provisioning platform for enterprise private cloud environments. Led **3 Engineering Leads and 14 engineers** across **4 Agile squads**, driving quarterly planning, sprint execution, architecture reviews, prioritization, incident management, and production readiness. Partnered with Product Management, Architecture, Operations, QA, and Customer Success to align technical investments with business goals, improve delivery predictability, and accelerate customer onboarding while maintaining platform reliability and operational excellence.
- **Enterprise AI Gateway** Architected and led engineering delivery of a production grade Enterprise AI Gateway supporting **50+ foundation models** with a team of **6 engineers**. Defined the platform architecture, technical roadmap, governance standards, and engineering practices for intelligent model routing, session memory, multi agent orchestration, batching, caching, and fine tuning. Improved GPU utilization by **30 to 50%**, reduced inference costs by **25 to 40%**, and enabled rapid onboarding of enterprise AI applications while ensuring scalability, security, and production reliability.
- **Generative AI Platform** Led the architecture, platform strategy, and engineering execution for an enterprise Generative AI platform powering Image Studio, Video Studio, and 3D Studio with a team of **4 engineers**. Delivered scalable AI services across **1000+ enterprise marketplaces** on NVIDIA H100 Kubernetes clusters. Established engineering standards, sprint planning, design reviews, and production readiness processes while improving engineering velocity. Optimized GPU utilization through intelligent scheduling and scale to zero capabilities, reducing infrastructure costs by **35 to 40%**, increasing GPU utilization by **40%**, and improving throughput by **3x**.
- **Enterprise Agent Platform** Led platform architecture, engineering execution, and roadmap for an enterprise AI Agent platform with **3 engineers**. Delivered reusable agent frameworks supporting **20+ enterprise workflows** with semantic search, knowledge graphs, enterprise integrations, and governance capabilities. Established engineering standards and reusable platform services that reduced new AI agent development time by **40%**, accelerated enterprise onboarding by **1.5x**, and enabled secure AI adoption across multiple business domains.
- **ML Platform Engineering, Nesy Factory** Owned the platform strategy, architecture, and engineering delivery of an enterprise ML Platform with **3 Engineering Leads, 6 ML Engineers, 3 Backend Engineers, 2 UI Engineers, and 1 UX Designer**. Established reusable MLOps and LLMOps capabilities covering data pipelines, model training, serving, deployment, governance, and lifecycle

management. Accelerated AI model delivery by **3x**, enabled **15+ production ML workloads**, and improved engineering productivity across multiple AI product teams.

- **Cloud Platform** Built and scaled a **12 member Platform Engineering team** from the ground up across Platform, DevOps, and Infrastructure disciplines. Mentored engineers, grew technical leaders, and supported the promotion of engineers into lead roles while establishing engineering processes, ownership, and career development frameworks. Owned the platform architecture, roadmap, and engineering delivery of a cloud native platform running on Azure Kubernetes Service. Established platform standards for Kubernetes, networking, security, deployment automation, disaster recovery, and operational governance supporting **150+ microservices** and **12+ databases** across multi region environments. Achieved **99.5% platform availability** while significantly reducing application onboarding effort.
- **Platform Observability** Architected and delivered a centralized observability platform supporting **150+ production services**. Partnered with Platform Engineering, SRE, and application teams to establish enterprise monitoring, distributed tracing, logging, capacity planning, and operational dashboards. Reduced Mean Time to Resolution by **30%** and improved production visibility by **80%**.
- **Database as a Service, PI Platform** Served as a member of the architecture committee defining platform architecture for streaming and batch data services in a multi tenant SaaS platform integrating **12+ database technologies**. Delivered horizontally scalable ingestion services benchmarked at **500K to 1M events per second** while establishing architectural standards for scalability and high availability.
- **Cloud Cost Optimization** Led cross functional initiatives with Platform, Infrastructure, and Finance teams to establish cloud governance, workload optimization, GPU sharing, autoscaling, and resource rightsizing standards. Reduced Azure infrastructure costs by **35%** while maintaining production reliability and service level objectives.
- **Content Delivery System, Pitcher Catcher** Architected and implemented a high performance communication layer using RSocket over TCP to improve cloud to edge data transfer for large files. Designed resilient communication patterns using circuit breakers, retry strategies, and adaptive bandwidth management to improve reliability across unstable network conditions.
- **Elastic2 Environment, SnapLogic** Worked as an individual contributor with distributed DevOps teams for spinning up a new environment supporting enterprise customer onboarding. Delivered infrastructure automation, deployment improvements, and Kubernetes platform fixes for AWS based environments across **20+ production environments**.

Software Engineer

Dec 2017 – Apr 2020

Planon Software

- Delivered planon integration solutions connecting **Azure, SAP, Leverton, & Planon applications** using REST APIs, asynchronous messaging, and scalable cloud integration patterns.
- Developed a Cordova Plugin for operating RFID scanners, enhancing data collection capabilities for cross-platform mobile applications using **Zebra & TSL SDK**.
- Designed predictive asset maintenance capabilities leveraging machine learning models, improving maintenance recommendation accuracy by **25%** while reducing operational downtime.
- Developed & deployed multiple **Tailored Made Solutions** using Planon as a Platform **PAAP** for International Team
- Improved software quality by driving SOLID architecture, automated testing, CI/CD, code reviews, and Agile engineering practices across multiple enterprise implementations.

Software Engineer

Jun 2016 – Dec 2017

Tech Mahindra

- Designed and delivered enterprise Java applications, RESTful services, middleware integrations, and tax-compliance modules supporting mission-critical financial systems.
- Worked on database optimization and application performance tuning initiatives that reduced a critical reporting workflow from **48 hours to under 4 hours (91% improvement)**, eliminating SLA breaches.
- Designed and implemented an enterprise effort and billing automation solution adopted across multiple delivery accounts, improving reporting accuracy and operational efficiency.

Associate Technical Consultant

Jul 2014 - Jun 2016

IFS Solutions

- Developed & maintained Oracle ERP modules, enterprise workflows, reports, and integrations using **Java, PL/SQL, SQL, REST APIs, and SOAP services**.
- Collaborated with enterprise customers to analyze requirements, architect product customizations, resolve production issues, and deliver successful application releases.

EDUCATION

Bachelor of Technology in Computer Science Engineering

2010 - 2014

CERTIFICATIONS

- **Google Cloud Generative AI Leader**, Google Cloud
Credential ID: 2db7f826218b41b3803c9ad5d71c9012
Issued Feb 2026
- **Certified Kubernetes Application Developer (CKAD)**, The Linux Foundation
Credential ID: LF-6xrpcn6ujy
Issued Dec 2024
- **Leadership skills**, IIM Ahmedabad
Ongoing
- **Pre-MBA Statics**, IIM Ahmedabad
Ongoing