Mohammed Saaduddin Ansari

Principal AI Platform Architect | Agentic AI, LLM Integration and AI Infrastructure

linkedin.com/in/md-saaduddin-ansari | mohammed-saaduddin-ansari.pages.dev | Hyderabad, India

PROFESSIONAL SUMMARY

AI platform architect with 11+ years building and running production infrastructure on GCP, AWS and Azure, and the last three years putting AI agents and LLMs into real operations. At LivePerson, as Principal Site Reliability Engineer, built AI agents for alert triage, incident investigation and runbook automation, with MCP tool access, human approval before any action and evals on their output; together with an AIOps pipeline they cut MTTR by 50% and on-call noise by 45%. Connects LLMs to systems through the Anthropic, OpenAI and Gemini APIs with tool use, structured outputs and model routing, and builds RAG pipelines and multi-agent workflows tested with golden datasets. Brings a cloud architect's habits to AI: security controls, cost tracking, tracing and a clear path from prototype to production, proven across 15+ client engagements at a Google Cloud partner.

CORE COMPETENCIES

Agentic AI Architecture • Multi-Agent Workflows and MCP • LLM Integration and Model Routing • RAG Pipelines and Evals • Human-in-the-Loop AI Safety • AI Observability and Cost Control • AI Infrastructure on Kubernetes • AIOps

WORK EXPERIENCE

LivePerson Jun 2025 - Nov 2026
Principal Site Reliability Engineer *Hyderabad | Remote*

- Built AI agents for alert triage, incident investigation and runbook automation, with MCP servers for tool access, human approval before any action and evals on their output
- Connected LLMs to internal systems through the Anthropic, OpenAI and Gemini APIs using tool use, structured outputs and model routing between providers
- Built multi-agent workflows and RAG pipelines, testing their output with custom evals and golden datasets before release
- Built an AIOps pipeline on OpenTelemetry with ML anomaly detection and LLM-assisted incident response that groups 500+ daily alerts into actionable incidents, cutting on-call noise by 45% and MTTR by 50%
- Standardised LLM-assisted Terraform generation and security review across the cloud engineering team, noticeably speeding up infrastructure delivery and reducing configuration drift
- Led an internal developer platform on Backstage and Crossplane for 50+ engineers, giving service teams self-service infrastructure through templates
- Set multi-cloud governance and FinOps controls across AWS, GCP and Azure (tagging policies, reserved-instance planning, AI-assisted right-sizing) for a platform serving millions of customer interactions
- Mentored a distributed team of 8 cloud engineers through architecture reviews, code reviews and pairing on complex changes

SADA (An Insight Company) May 2022 - May 2025
Cloud Architect *Hyderabad | Remote*

- Introduced LLM-assisted infrastructure workflows from 2023 for Terraform generation, security policy validation and PR review, speeding up delivery across 15+ client engagements
- Deployed AIOps incident management with ML anomaly detection and automated runbooks, reducing false alerts by 60% and mean time to detect from 15 minutes to under 3
- Designed multi-cloud AWS and GCP environments and built new GCP foundations from scratch for clients of a Google Cloud partner
- Led client migrations to GCP from on-premises, Azure and AWS, including 100+ on-premises applications, reducing infrastructure costs by about 35%
- Designed the migration strategy that moved 200+ applications to AWS for another SADA client
- Architected multi-region, multi-cloud infrastructure on GCP and AWS for a Fortune 500 client, reaching 99.99% uptime
- Designed disaster recovery for a critical banking system, reducing RTO from 4 hours to under 20 minutes and RPO from 1 hour to under 5 minutes
- Built an AWS landing zone that kept client environments aligned with SOC 2 and HIPAA controls
- Architected a cloud security and compliance framework for a healthcare agency, helping it reach Authority to Operate (ATO)
- Reduced client infrastructure costs by up to 40% through spend analysis, rightsizing and commitment-based discounts

EPAM Systems Nov 2021 - May 2022
Lead DevOps Engineer *Hyderabad*

- Led a team of 6 engineers delivering multi-cloud solutions, completing 4 major engagements with no critical production incidents

- Built end-to-end release pipelines with GitHub Actions and ArgoCD, so configuration drift was detected and reverted automatically
- Owned application migration across a cross-cloud, multi-cluster Kubernetes environment
- Created a reusable Terraform module library used across 3 client engagements, cutting environment provisioning from days to under 30 minutes

SAP Labs Jan 2020 - Nov 2021
Sr DevOps Engineer *Hyderabad*

- Owned Kubernetes clusters on GCP as Product Owner, built with Terraform, moving SAP's product suite from on-premises to Google Cloud across 100+ internal environments
- Broke monolithic applications into containerised microservices with Docker as part of the move to Kubernetes
- Built CI/CD with Jenkins, Spinnaker and Ansible for 400+ VMs, cutting the deployment cycle from 2 weeks to under 4 hours with zero-downtime releases
- Standardised monitoring with Prometheus and Grafana, reducing MTTR from 4 hours to 30 minutes

Deloitte USI Oct 2017 - Jan 2020
DevOps Engineer *Hyderabad*

- Automated deployments with Terraform and Ansible for 10+ enterprise clients, cutting provisioning from 2 days to 30 minutes
- Built Jenkins CI/CD pipelines supporting 50+ daily deployments with automated rollback
- Led a team of 3 running monthly platform upgrades, and worked with engineering teams to fix recurring application issues

Infosys Jun 2015 - Oct 2017
DevOps Engineer *Pune*

- Supported the migration of 20+ client applications from on-premises data centres to AWS
- Automated Linux system administration with Bash and Python, saving 30+ hours of manual work a week
- Worked with the release management team to automate releases and troubleshoot deployment failures

EDUCATION

B.Tech, Computer Science and Engineering (CGPA 8.4) *Muffakham Jah College of Engineering and Technology, Osmania University* 2015
Hyderabad

Microsoft Student Ambassador Head and intern (2014-2015); Logistics Head, Computer Society of India; IEEE member

ADDITIONAL INFORMATION

Availability: immediate joiner; open to relocation to the UAE, Saudi Arabia and Qatar
Languages: English and Urdu (fluent)

TECHNICAL SKILLS

AI agents: Custom agents, multi-agent workflows, MCP servers and clients, LangGraph, LangChain, CrewAI, AutoGen, OpenAI Agents SDK, human-in-the-loop approval

LLM integration: Anthropic, OpenAI, Gemini and DeepSeek APIs, tool use, structured outputs, prompt caching, streaming, model routing, LLM gateways

RAG and evals: RAG pipelines, embeddings, pgvector, custom evals, golden datasets, LangSmith, LLM tracing and cost tracking

AI infrastructure: GPU workloads on Kubernetes, model serving, fine-tuning, Ollama

AI coding tools: Claude Code, Cursor, GitHub Copilot, Gemini CLI, OpenCode

AIOps: ML anomaly detection, alert correlation, AI-assisted incident response and RCA, AI-generated runbooks and postmortems

Cloud: GCP (GKE, Compute Engine, Cloud Storage, Pub/Sub, Cloud Build, Cloud SQL, Cloud Functions), AWS (EKS, EC2, VPC, IAM, Lambda, RDS, S3, Route 53, MSK), Azure (AKS, Entra ID)

Kubernetes and containers: Kubernetes (GKE, EKS, AKS), Docker, Helm, Istio, Cilium, Karpenter, Kasten K10

Infrastructure as code: Terraform, OpenTofu, Crossplane, Pulumi, Ansible, Packer, CloudFormation

Observability: Prometheus, Grafana, OpenTelemetry, Datadog, Dynatrace, ELK, CloudWatch, Google Cloud Monitoring

Security and compliance: IAM, OPA/Gatekeeper, Kyverno, HashiCorp Vault, SOPS, Doppler, cloud secret managers, SOC 2, HIPAA

Languages: Python, Go, Bash