

Usman Shahid

+971-52-932-8265 | u.manshahid@gmail.com | linkedin.com/in/u-manshahid | github.com/codemug |

Dubai, United Arab Emirates

PROFESSIONAL SUMMARY

Staff **Site Reliability Engineer** and **Technical Lead** with **11+ years** scaling resilient, cost-effective cloud infrastructure for high-traffic services. Currently running a **self-service AI Agent Platform** with **Model Context Protocol (MCP)** tool-calling, **multi-agent orchestration**, and **human-in-the-loop (HITL) approval flows** - **~200K requests/day**, **~800 registered tools**, **~60 production agents** - bridging classic SRE discipline with the agentic-AI era. Track record: **40% EKS cost reduction** via Karpenter / ARM / Spot / VPA, **99.99% uptime** consolidating 12 API gateways to Kong at ~10k RPS, **300+ microservices** migrated to EKS. Polyglot - production **Go, Python, Java**, and **C#**. Daily AI-tooling user (Claude, Cursor, Copilot, Gemini).

CORE COMPETENCIES

Site Reliability Engineering

Platform Engineering

AI Agent Platforms (MCP)

Kubernetes & EKS

FinOps (cost engineering)

GitOps (Terraform/ArgoCD)

Observability (Prometheus/Thanos/Grafana)

Polyglot (Go/Python/Java/C#)

WORK EXPERIENCE

Careem

January 2023 - Present

Staff Site Reliability Engineer

- Technical Lead and SRE Architect mentoring a team focused on infrastructure provisioning and developer experience.
 - Designed & built a **self-service AI Agent Platform** with **MCP tool-calling**, **multi-agent orchestration**, and **HITL approval flows** for internal infrastructure-support workflows - **~200K requests/day**, **~800 registered tools**, **~60 agents in production**. Engineering teams contribute their own tools and agents.
 - **Slashed EKS cluster costs by 40%** via resource-optimization program - **Karpenter** consolidation, migration to **ARM & Spot** nodes, **VPA**, and custom scheduling (FinOps at scale).
 - Consolidated **12 legacy API gateways (~10k RPS) onto a single Kong gateway with 99.99% uptime** during cutover; built a GitOps-driven bespoke control plane for API route management.
 - Designed and deployed a fully automated cloud-resource provisioning platform with **GitOps (Terraform + Terragrunt)** - provisioning time for new services from **days to minutes**, eliminating manual configuration errors.

Careem

December 2020 - January 2023

Senior Site Reliability Engineer

- Led migration of **300+ microservices** from legacy AWS infrastructure to a scalable, cost-effective **AWS EKS** platform:
 - Engineered automated provisioning of multiple EKS clusters with **Terraform + Terragrunt** - repeatable, auditable infrastructure foundation.
 - Provisioned centralized **API Gateways** using Kong with a GitOps-based bespoke control plane for API route management.
 - Built the observability stack with **Prometheus**, **Thanos**, and **Grafana** - high-cardinality metric storage, SLO dashboards, and cost tracking across the EKS fleet.
 - Improved deployment velocity and safety via **service mesh (Linkerd)** with automated **canary rollouts via Flagger**.
 - On-call 24x7 on a bi-weekly rotation for the cloud-infrastructure platform.

Infrastructure Architect & Team Lead

- Designed & implemented the infrastructure of a **Cloud IIoT Platform** while **leading a team of 6 polyglot engineers** in a **KanBan** model - workload distribution, technology selection, mentoring, customer integration.
- Built a bespoke **Kubernetes Ingress Operator** for bare-metal Kubernetes clusters using **RedHat operator-sdk** (custom control-plane work, not just operator consumption).
- Secured & streamlined the software delivery lifecycle by integrating **ArgoCD** with **Keycloak** - a secure, auditable **GitOps** workflow for **Helm**-based deployments.
- Implemented **Google SRE golden signals** (latency / traffic / errors / saturation) on bare-metal Kubernetes clusters using **Prometheus**.

Intech IIS

January 2017 - June 2019

Senior Software Developer / DevOps Engineer

- Containerized all services of a distributed Industrial IoT platform with **Docker**; orchestrated via **Kubernetes** and **D2iQ** stacks.
- Automated deployment of the IIoT platform using **Ansible Playbooks**; provisioned production-ready Kubernetes and D2iQ clusters.
- Built a **serverless framework in Python** running code on containers via Kubernetes / Marathon / Metronome (one-off, cron-scheduled, long-running configurations).
- Built a parallel metric aggregation publisher for **StatsD** in **Golang**.
- Built an **OPC to MQTT** data pipeline service in C# using **Avro RPC** and **REST**.
- Built a generic client API using Java **Streams** for **CRUD** operations on **IBM Maximo** and **SAP EAM** systems via Avro schemas, exposed via Avro RPC.
- Built an **OpenID Connect Identity Provider** microservice using **Spring OAuth2** with JDBC, Active Directory, LDAP, and Mongo storage backends (Spring JPA + Spring LDAP).
- Implemented **dynamic proxying** and **JWT bearer authentication** using **OpenResty LuaJIT**.

Diyatech (Development Hub for Alachisoft)

June 2014 - January 2017

Software Developer

- Core developer for **NCache Enterprise** - distributed caching server (Redis-class) native to the .NET ecosystem, written in C#:
 - Reimplemented Microsoft's Collections (data structures) to prevent **Large Object Heap (LOH) leakage**.
 - Implemented the **ASP.NET Core** Session State Management middleware with NCache backend storage and full session-locking support.
 - Implemented a reliable socket protocol for message delivery using **Protobuf**.
- Core developer for **NosDB Enterprise** - NoSQL JSON data store (CouchDB / MongoDB-class) native to .NET:
 - Built file-storage-based **B+ Tree indexing** with support for single-attribute and compound indexes.
 - Built the query execution system using **Streams** and nested Enumerables in C#.

EDUCATION**Bachelors (Honors) in Computer Engineering, National University of Sciences and Technology (NUST)**

Graduated 2014

CGPA: 3.40

SKILLS

Site Reliability / Platform Engineering: SRE practices, SLO/SLI/error budgets, on-call & incident response, GitOps, Internal Developer Platforms, FinOps, capacity planning

Kubernetes ecosystem: Kubernetes (managed + bare-metal), Karpenter, VPA, custom scheduling, Linkerd (service mesh), Flagger (progressive delivery / canaries), Helm, Kustomize, ArgoCD, operator-sdk (custom Go operators)

Cloud: AWS (EKS, EC2, S3, VPC, Route 53, IAM), bare-metal Kubernetes, multi-cluster patterns

Observability: Prometheus, Thanos, Grafana, Google SRE golden signals, SLO dashboards

Infrastructure & CI/CD: Terraform, Terragrunt, Ansible, Docker, D2iQ, GitLab CI

API & Gateway: Kong (12 to 1 consolidation, ~10k RPS, 99.99% uptime), OpenResty / LuaJIT, OpenID Connect, JWT, dynamic proxying

AI / LLMops: Model Context Protocol (MCP), multi-agent orchestration, HITL approval flows, AI Agent Platform engineering; daily use of Claude, Cursor, GitHub Copilot, Gemini

Languages: Go (production), Python (production), Java (production), C# / .NET (production)

Data / Storage internals: B+ Tree indexing (NosDB), distributed-cache internals (NCache), Mongo, LDAP, JDBC, Avro / Protobuf

AWARDS

Excellence Award, [Alachisoft Inc.](#) - outstanding performance recognition

November 2015