

# Immar Karim

*DevOps/MLOps*

+92-344-3844854 | [immarunnar@gmail.com](mailto:immarunnar@gmail.com) | [linkedin.com/in/immar-karim](https://www.linkedin.com/in/immar-karim) | Islamabad, PK

## EDUCATION

**National University of Science and Technology - 3.04/4**

*Bachelor in Computer Science*

Islamabad, PK

Aug. 2020 – May 2024

## EXPERIENCE

**CLOUD ENGINEER (KUBERNETES)**

*Chkk*

May 2024 – Present

*Sunnyvale, CA - Remote*

- Developed the Upgrade Copilot feature to help enterprise clients like PayPal, The New York Times, and Swisscom upgrade their Kubernetes clusters with zero downtime.
- Created Automated SHARC, a Go-based tool that talks to Kubernetes API Server and runs series of health checks and API readiness checks on the Kubernetes cluster and its add-ons, ensuring large infrastructures operate without downtime by testing monitoring, metrics, and other Kubernetes components.
- Tested and validated Infrastructure upgrade processes internally to mirror customer setups, ensuring smooth transitions between Kubernetes versions.
- Provided hands-on support for upgrading Kubernetes clusters, managing large-scale deployments with thousands of nodes, and minimizing any potential disruptions.
- Assisted customers with upgrading and optimizing various complex Kubernetes add-ons like Datadog, Prometheus, Istio, Karpenter, etc., ensuring everything runs smoothly post-upgrade.
- Utilized several DevOps tools including Terraform and a range of cloud services across AWS, GCP, and Azure to support diverse infrastructure configurations.
- Participated in 24/7 on-call rotations through PagerDuty, responding to customer-impacting issues and internal application breakdowns, and resolving them promptly using tools like Sentry.

**PRODUCTION DEVOPS ASSOCIATE (MLOPS)**

*Vyro.ai*

Aug 2023 – May 2024

*Islamabad, PK*

- Setup and maintained highly-available, bare-metal, GPUs based Kubernetes cluster, to serve Stable Diffusion text2image text2video ML models and Qwen LLMs to 9 million monthly active users.
- Implemented efficient auto-scaling in the bare-metal cluster using Knative, leveraging its powerful serverless capabilities to cut on server costs.
- Deployed and managed Machine learning models using KServe for scalable and efficient model serving.
- Implemented Kubevirt for running VMs in parallel with containerized workload. This allowed developers to access GPUs from the Kubernetes cluster, reducing the need for dedicated development machines.
- Efficiently deployed GAN models such as esrgan and codeformer utilizing Triton Server.
- Setup EFK stack using Helm charts for logging.
- Setup Argo CD for continuous deployment to the production cluster.
- Successfully ensured 99.9% availability and reliability of all kinds of workloads in production environment, while contributing to best security practices.
- Used my software development experience to automate tasks and run scripts to increase overall productivity.

**INTERN, Software Developer**

*Shadiyana.pk*

Mar 2023 – Oct 2023

*Islamabad, PK*

- Developed and incorporated changes to the company website made in the PERN stack. Helped reduce website load time and achieve better SEO.

## TECHNICAL SKILLS

---

**Languages:** Go, Python, C/C++, SQL, HTML/CSS

**Frameworks:** FastAPI, client-go

**Tools & Technologies:** Kubernetes, KServe, Terraform, Pulumi, Rancher, RKE2, Git, Docker, EKS, GKE, AKS, Ansible, Linux, Github Actions, Temporal, PagerDuty

## CERTIFICATIONS

---

**Certified Kubernetes Administrator (CKA)**

The Linux Foundation