

Krishna Sharma

+91-6006035201 | krishnasharma2910.ks@gmail.com | linkedin.com/in/krishna-sharma-711648198

AWS Certified Solutions Architect – Associate

SUMMARY

Backend and AI Engineer with 3.5+ years of experience building scalable Python applications, cloud-native systems, and AI-powered solutions. Experienced in Django, FastAPI, PostgreSQL, AWS, Docker, and microservice architectures. Strong background in designing production-grade APIs, distributed systems, and cloud deployments. Recently focused on Generative AI, Retrieval-Augmented Generation (RAG), vector search, and LLM-powered applications.

EDUCATION

Panjab University

Bachelor of Technology in Computer Science

Chandigarh, India

2019 – 2023

EXPERIENCE

Backend & DevOps Engineer

IT BY DESIGN

2023 – Present

India

- Developed and maintained backend applications using Python, Django, and FastAPI for customer-facing and internal business platforms
- Created a full reporting engine app which ingest live data from PSA tools , store it in MongoDB and process the data to show useful insights which remove the dependency from MSPBots for our application.
- Designed scalable REST APIs and microservices enabling reliable communication across distributed systems
- Migrated social feed services to PostgreSQL, improving application performance by **60%**
- Built asynchronous processing pipelines using RabbitMQ and Celery for scalable background task execution
- Optimized backend workflows, API performance, and database queries supporting high-volume production workloads
- Implemented centralized logging and monitoring solutions enabling **10x faster** root cause analysis and proactive issue detection
- Built CI/CD pipelines and automated deployment workflows using AWS, Jenkins, Docker, and Terraform
- Managed production cloud infrastructure and deployment processes ensuring high availability and zero-downtime releases
- Collaborated with product, QA, and engineering teams to deliver scalable backend features and operational improvements

PROJECTS

Document Intelligence RAG Chatbot | *FastAPI, Python, LLM APIs*

- Built a Retrieval-Augmented Generation (RAG) chatbot enabling users to upload documents and ask context-aware questions
- Developed document ingestion pipelines including chunking, embedding generation, semantic retrieval, and contextual response generation
- Designed FastAPI backend services supporting document management and conversational AI workflows
- Improved response quality through prompt engineering and retrieval optimization techniques

AI Knowledge Assistant | *Python, LLM APIs*

- Developed an AI-powered knowledge retrieval system leveraging Large Language Models and contextual search
- Implemented retrieval workflows and prompt engineering techniques to improve response accuracy and consistency

Cloud-Native CI/CD Platform | *AWS, Jenkins, Docker, Terraform*

- Built automated deployment pipelines improving release consistency and deployment reliability
- Reduced infrastructure provisioning time by **70%** through reusable Infrastructure-as-Code templates
- Implemented zero-downtime deployment strategies for production applications

TECHNICAL SKILLS

Languages: Python, SQL, Bash

Backend Frameworks: Django, FastAPI, REST APIs, Microservices

AI / GenAI: LLM Applications, RAG Pipelines, Prompt Engineering, Vector Search, AI Chatbots

Databases: PostgreSQL, MongoDB, MySQL

Cloud: AWS (EC2, ECS, RDS, S3, IAM, VPC)

Containers: Docker, Kubernetes (Working Knowledge)

Messaging: RabbitMQ, Celery

DevOps: Jenkins, Terraform, CI/CD, Git

Monitoring: Splunk, CloudWatch

CERTIFICATIONS

AWS Certified Solutions Architect – Associate

Amazon Web Services (AWS)

2026

Valid till 2029

ADDITIONAL INFORMATION

Languages: English, Hindi

Interests: AI Systems, Music, Cricket, Exercise