



Vijay Saini

Generative AI Specialist & Cloud Architect with Azure DevOps Expertise

Cloud Architect and Generative AI Specialist with over 10 years of experience in designing and implementing microservices architectures, CI/CD workflows, and automated cloud solutions. Detail-oriented and results-driven, I have a strong track record of enhancing system reliability and operational efficiency through innovative solutions in Generative AI and infrastructure automation, particularly within Microsoft Azure environments.

✉ vijaysainiprofessional@gmail.com

📍 Bengaluru, India

📖 stackoverflow.com/users/5881105/vijay

📞 97844 04866

🌐 linkedin.com/in/vijay-saini-10759a8b

SKILLS

Generative AI

Retrieval-Augmented Generation (RAG)

AIOps

Azure AI

Azure AI Studio

CrewAI - Multi-Agent Platform

LangChain & LangSmith

Custom GPTs & GenAI Agent

Prompt Engineering

LLM

Certified Kubernetes Professional - CKAD & KCNA

Architecture Design

Serverless computing

GitHub Actions

Infrastructure as Code - Terraform

Microservices

CI/CD

PowerShell & Python

Microsoft Azure

Microsoft Certified: DevOps Engineer Expert

WORK EXPERIENCE

Lead Architect – GenAI & DevOps

PwC US

05/2023 - Present

Bengaluru, India

PwC (PricewaterhouseCoopers) is a global professional services firm offering audit, tax, and consulting services. It helps organizations solve complex business challenges, optimize operations, and drive innovation across industries.

Achievements/Tasks

- Developed a multi-region hub and spoke cloud architecture using Infrastructure-as-Code (IaC) principles to manage scalable, distributed environments. Played a key role in onboarding several enterprise clients to Azure, streamlining their operations and supporting their transition to cloud-based services.
- Dived into the world of Generative AI, where I helped create and deploy multiple serverless applications on Azure Function Apps, enhancing our ability to deliver innovative AI solutions.
- As the project gained traction, I helped transform it into a centralized hub for generative AI applications on Kubernetes, enabling efficient communication between apps through Service Bus.
- Implemented Helm for centralized Kubernetes management and KEDA for intelligent pod scaling based on message queue size, optimizing resource usage and processing speed.
- Designed a robust DevOps framework that included CI/CD pipelines and automated database operations, which significantly improved our deployment processes and overall project reliability, using Datadog for effective monitoring.
- Gained valuable experience with tools like Azure AI Search, Document Intelligence, Helm, and OpenAI, which enriched our offerings and kept us ahead of industry trends.
- Leveraged my understanding of generative AI application flow to troubleshoot and resolve issues, tracing requests from custom GPT through Kubernetes ingress and pods. Used tools like Swagger and Datadog to efficiently identify root causes and implement effective solutions.
- Collaborated closely with the security team to identify and mitigate potential vulnerabilities, implementing customer-managed keys for encryption, establishing private endpoints for secure connectivity, and enforcing robust password and key policies, ensuring our architecture met stringent company security standards.
- Implemented an event-driven Retrieval-Augmented Generation (RAG) pipeline from scratch, serving as the backbone of projects. This solution integrated AI services, including Service Bus, Storage Account, Document Intelligence, embeddings, and AI Search.

WORK EXPERIENCE

Technical Lead

MindTree Ltd - Project IDN

02/2021 - 05/2023

Bengaluru, India

Achievements/Tasks

- Developed a scalable Kubernetes infrastructure to develop and deploy Microservices-based platform aimed at bringing multiple partners together to develop plugins based on existing offerings of client products.
- Implemented DevOps practices with GitHub Action CI/CD workflows for Microservices deployment on Azure Kubernetes via Azure Container Registry.
- Achieved one-click end-to-end infra deployment automation. Developed & Maintained Azure Bicep templates for deployment of Azure resources (Infrastructure as Code).
- Troubleshoot and resolved complex production issues related to Kubernetes and Azure. Collaborated with development teams to ensure smooth deployment and operations of applications in the production environment.

DevOps Engineer

o9 Solutions

05/2020 - 02/2021

Bengaluru

Industry leaders transforming its customer's Sales, Supply Chain and Integrated Business Planning with AI solutions .

DevOps Engineer

- Troubleshooting of production issues and updated infrastructure with software patches. Production Support, Upgrades, and monthly maintenance.
- Developed End-to-End Application validation Automation deployed on Jenkins for validating multiple servers parallelly and generate a formal html validation report. Internal Process Automation.
- Recommended & implemented enhancements for improving availability, reliability & performance of our application. o9 platform release & deployment cycle and Product maintenance. Excellent planning and Implementation skill.

Technical Consultant - DevOps

Blue Yonder (Previously JDA)

01/2015 - 05/2020

For over 30 years, Blue Yonder has been the leading provider of end-to-end, integrated retail and supply chain planning and execution solutions.

Azure DevOps Engineer

- Technology Used: Azure DevOps Azure PowerShell, DSC, Azure Repos, Azure Pipelines, ARM Templates, Azure WebApps, Logic Apps.
- Part of the team for planning to migrate our enterprise application from traditional infrastructure to cloud-based infrastructure with minimal disruption in our customer service. End to End Environment provisioning of JDA Products in Microsoft Azure Cloud.
- Developed PowerShell scripts and automated 1000+ hours of manual work/year. This also leads to lesser downtime in our services. Was awarded the second-highest prestigious award(Individual Pillar Award) for this work.
- Troubleshooting of production issues and updated infrastructure with software patches. Production Support, Upgrades, and monthly maintenance.

TECHNICAL SKILLS

Azure Services

Kubernetes Service, Containers, FrontDoor, Application Gateway, App Service, Azure Automation, AppConfig

Generative AI

Prompt Engineering, AIOps, GenAI Agents, LangChain, LangSmith, Custom GPTs, Swagger, FastAPI

Cloud

Serverless Computing, Function App, App Service, Virtual Machine, Monitoring, On-premise to Azure Cloud Migration

DevOps

GitHub Action, Designing Build and Release Pipelines, Secret Management, Azure Boards, Agile

Automation

PowerShell, Python, Azure CLI, Azure PowerShell

Database

Microsoft SQL Server, Oracle, Microsoft Azure SQL Database

PROJECTS

Lead Architect, GenAI Factory Initiative

- Led the design and development of the GenAI Factory platform, creating a unified architecture that supports multiple business use cases, streamlining hosting, operations, and maintenance through a common platform pattern.
- Architected a multi-agent, multi-Generative AI model platform utilizing LangChain, GenAI Agents, LangSmith, and AI Search Index, enabling seamless integration of various generative AI models.
- Implemented a robust Retrieval-Augmented Generation (RAG) pipeline to enhance AI-driven data retrieval and contextual responses, improving model performance across use cases.
- Leveraged Azure Service Bus for decoupled communication, and utilized Helm for Kubernetes management and KEDA for pod scaling, optimizing resource usage and processing speed.
- Led the DevOps strategy, optimizing CI/CD pipelines, automation processes, and infrastructure-as-code deployments to accelerate development and deployment cycles.

Multi-Region Hub & Spoke Platform Automation for a Healthcare Client's Cloud Transformation

- Spearheaded the development of a fully automated multi-region Hub & Spoke platform for a leading U.S. healthcare client, ensuring scalability and resilience in their cloud infrastructure.
- Utilized Terraform for Infrastructure as Code (IaC) and Azure DevOps pipelines to streamline deployment and management processes, enhancing operational efficiency.
- Enabled the client's data transformation journey by delivering a highly available, fault-tolerant platform deployed on Azure Cloud, supporting their transition to a modern cloud-based architecture.

Microservices Based Enterprise Development Platform - IDN

- Developed a scalable and standardized Microservices-based platform aimed at bringing multiple partners together to develop extensions based on existing offerings of client products.
- Achieved one-click end-to-end infra deployment automation. Developed & Maintained Azure Bicep templates for deployment of Azure resources (Infrastructure as Code).
- Developed GitHub Action CI/CD workflows for Microservices deployment on Azure Kubernetes.

CERTIFICATES

Certified Kubernetes Application Developer (CKAD) (04/2024 - Present)

The Certified Kubernetes Application Developer (CKAD) exam certifies that candidates can design, build and deploy cloud-native applications for Kubernetes. The CKAD was created by the Linux Foundation.

Microsoft Certified: DevOps Engineer Expert (01/2022 - Present)

subject matter expertise in working with people, processes, and products to enable the continuous delivery of value in organizations.

Microsoft Certified: Azure Data Fundamentals (11/2023 - Present)

This certification is an opportunity to demonstrate your knowledge of core data concepts and related Microsoft Azure data services.

AZ-900 : Microsoft Certified Azure Fundamentals certification (08/2019 - Present)

This certificate demonstrates to have a foundational knowledge of cloud services and how those services are provided with Microsoft Azure.

Hashicorp Certified Terraform Associate (07/2023 - Present)

The Terraform Associate certification is for Cloud Engineers specializing in operations, IT, or development who know the basic concepts and skills associated with Terraform.

Kubernetes and Cloud Native Associate (KCNA) (12/2022 - Present)

Demonstrates a user's foundational knowledge and skills in Kubernetes and the wider cloud native ecosystem.

AZ-104: Microsoft Certified: Azure Administrator Associate certification (10/2019 - Present)

Expertise in implementing, managing, and monitoring an organization's Microsoft Azure environment.

Microsoft Certified: Azure AI Fundamentals (01/2024 - Present)

Demonstrate fundamental AI concepts related to the development of software and services of Microsoft Azure to create AI solutions.

HONOR AWARDS

Individual Pillar Award (2018)

Blue Yonder

- An annual award to recognize the exceptional performance & individual efforts that have made a significant contribution to enhancing customer satisfaction and timely delivery.

Spot Award (2017)

JDA Software

- Awarded SPOT award for efficiently handling a data center outage potentially causing business disruption of multiple clients.

Team Pillar Award (2019)

JDA Software

- An annual award to recognize the contribution of an individual in helping the team to achieve its targets.

EDUCATION

Master Of Science(hons.)

BITS Pilani, Pilani campus

07/2010 - 06/2014

Pilani, Rajasthan