

Souvik De

India | souvikde.tech@gmail.com | +91 9073302976 | souvikns.com | linkedin | github/Souvikns

Summary

Software Developer with 4+ years of experience building backend systems and AI-powered applications. Specialized in RAG architectures, vector databases, and LLM integrations using Go, Node.js, and Python. Experienced in designing scalable, high-performance APIs and developer tools. Maintainer at AsyncAPI and mentor at Google Summer of Code (GSoC).

Experience

Software Engineer, XANE.AI – Gurgaon

July 2025 – Present

- Designed and developed scalable, high-performance GenAI solutions for Maruti Suzuki's Customer Assistant System, ensuring reliability and seamless user experience.
- Designed and implemented a backend system to OCR PDF documents and convert them into RAG-ready knowledge sources, enabling context-aware AI assistant agents.

Lead Software Engineer, CodeMate.AI – Noida

December 2024 – May 2025

- Engineered an enhanced search algorithm for Swagger files, optimizing RAG system performance and search precision.
- Implemented CI/CD pipeline utilizing GitHub Actions to automate application release cycles, enhancing deployment efficiency and reliability.
- Optimized a local server utilized by our VS Code extension, achieving over 30% improvement in startup time and enhancing overall user experience.
- Implemented automated deployment from GitHub Actions to a virtual machine (VM) and configured branch protection rules, improving code quality, deployment stability, and increasing team productivity by eliminating manual deployment efforts.

Software Engineer, Postman – Bangalore

Feb 2022 – June 2024

- Developed and maintained a CLI application that integrated all official tools supported by the AsyncAPI Initiative, providing a unified solution for developers working with the AsyncAPI specification.
- Initiated and developed AsyncAPI Bundler, an npm package that intelligently resolves \$ref references and merges multiple AsyncAPI specification files. The tool has gained widespread adoption with over 16,000 downloads.
- Served as a maintainer for over three open-source projects, leading development efforts, reviewing pull requests, and driving project direction within the community.
- Served as a mentor for Google Summer of Code (GSoC) under Postman, successfully guiding participants to complete their projects and achieve program goals.
- Initiated and led a working group with a colleague to define a common governance framework for building parsers and to identify tooling that minimizes development time and effort across multiple programming languages.

Backend Developer Intern, Mage – Remote

Nov 2021 – Feb 2022

- Developed microservices to automate cloud infrastructure provisioning using Go's templating engine, improving scalability and deployment efficiency.
- Built a React-based web application to deliver a low-code solution for generating and provisioning cloud infrastructure on demand.
- Gained hands-on experience with Docker and Kubernetes to deploy, manage, and scale microservices effectively.

Projects

GitHub Action to sync GitHub Issues to Notion Database

Notion-Board

- Developed a GitHub Action to auto sync and update github issues and it's state with Notion Pages.
- Earned recognition from the open-source community through project stars and valuable user feedback.
- Tools Used: Typescript, GH Action.

FastAPI application that scrapes AsyncAPI documentation to create a RAG system and provides a chat API to answer user queries.

AsyncAPI-RAG

- Built data scrapper to scrape AsyncAPI spec and tools documentation from github.
- Used LangChain to chunk Markdown data, generated embeddings with the Nomic embed model from Ollama, stored embeddings in a Qdrant database, and exposed a REST API to handle user queries and return contextual answers.
- Tools Used: Python, qdrant, Langchain, ollama, docker

Technologies

Languages: Javascript, Typescript, Go, Python, SQL

Technologies: NodeJs, FastAPI, Gin, Postgress, Firebase, NextJS, React, Langchain, Oclif

Education

Chandigarh University, BE in Computer Science

June 2018 – June 2022

- **Coursework:** Computer Architecture, Comparison of Learning Algorithms, Computational Theory