Harish Vadaparty

Portfolio | harishvadaparty@gmail.com | linkedin.com/in/harishvadaparty | github.com/Harryalways317

Work Experience

OnFinance.ai — Remote
Backend and AI Engineer — Apr. 2024 – Present

- Led the backend team, working with a cross-tech stack of GCP, Azure, Golang, Python with Gen AI App and multiple copilots
- Built data processing pipelines with queue-based systems for document processing, achieving a 70% increase in speed and making the system fail-proof compared to previous implementations
- Developed and maintained Copilot backends for production clients such as Oister and Avendus
- Implemented RAG workflows and efficient retrieval mechanisms for data extraction from vector databases
- Designed the backend as a template-based system, reducing the configuration time for new client onboarding by 50%
- Worked on DevOps tasks including resource management, CI/CD pipelines, and network and infrastructure setup
- Tech Stack: Python, Golang, Java, MongoDB, PostgreSQL, Kafka, Docker

Hexon Labs — Bangalore, IN
Software Engineer, Backend and Data — Jul. 2022 – Apr. 2024

- Led Backend Development to an established Grocery application with AWS backend, Node,Python and TypeScript
- Build microservices with Python,Node to manage orders/subscriptions, products, and deliveries
- Developed sync scripts to keep multiple services, stock, prices and store Point of Sale systems in sync
- Created Service Applications for Email,SMS,In-App,Notification,Cart Vouchers,Delivery Management, Reports(Apache Superset) to help the end users updated and operations team to run things smoothly
- Integrated Elastic Search that improved search relevancy by 60% and build recommendation engine to improve discovery with an increase of relevant cart items by 40%
- Worked with Tortoise TTS for Voice Cloning from Text and Base Samples and Audio-based Lip Sync
- Intergrated RAG-LLM workflows In-App and Fine-tuned models for better performance and relevancy

Board Infinity — Mumbai, IN
Software Engineer Intern — March 2022 – June. 2022

- Contributed to an established Ed Tech application in rewrite from React Native to Flutter with MVVC
- Developed multiple features on LMS, Blogs, Event and Course Management
- Implemented UXCam for analytics and notifications pipeline for FCM and Imrpoved animations
- Rewrote Modules of the app by caching and data reuse which improved performance by 40%

Education

Lendi Institute of Engineering and Techonology — Vizag, IN
Bachelor of Techonology in Computer Science — June 2022

Projects

Any2Any Convertor — Python, Flask, RabbitMQ, MySQL,MongoDB

- Developed an MP3 extraction and File conversion service with support for pdf to docx/docx to pdf.
- Implemented a Gateway for handling authentication, file storage, and processing requests in the queue.
- Established an email notification system to inform users upon completion of file processing.
- Encapsulated the service in a Kubernetes (K8s) environment for scalability, supporting various conversions and Volume Persistance on failures.

Job Leads Generator and Staffing Assistant — Python,Open AI API,Langchain,Postgres

- Created an end-to-end automated system for identifying potential leads and generating emails.
- Developed web scrapers to extract and clean data about companies with active hiring needs.

- Integrated Langchain & OpenAI to create personalized email templates based on the scraped data, enhancing outreach for staffing services.

CurateSphere - A News Summarizer Platform Python,Fast API,Open AI,Postgres,Qdrant

- Developed a comprehensive news summarization platform, CurateSphere, capable of condensing news articles into concise summaries of under 5 minutes reading time.
- Implemented a robust system for scraping and ingesting news content from diverse sources, utilizing Qdrant for storing article embeddings and appending succinct summaries (TLDR) to each article.
- Ranks news articles on a priority basis and selects top 10 articles based on score generated and summarizes them as top news.

SnipShare Java, Spring Boot, PostgreSQL

- Created a minimalist web application inspired by pastebin, allowing users to save temporary text/snippets to share or reuse.
- Implemented a backend with Spring Boot to manage text storage, generate unique shareable links, and handle link expiry.
- Developed a simple frontend with HTML, CSS, and JavaScript, capturing the Ctrl + S event to trigger backend API calls for saving text.
- Utilized PostgreSQL for efficient storage and retrieval of text entries and metadata.
- Added functionality for generating unique, shareable links for each text entry, enabling access from any location.
- Integrated link expiry mechanisms to enhance security by invalidating links after a specified period.

Spot Instance Manager Python, Fast API, PostgreSQL, boto3 sdk

- Created a minimalist Spot Manager Application that manages spot instances from AWS with auto stop and auto start features.
- Implemented a inferencing proxy for deployed models with VLLM that allocates a spot instance with desired AMI from existing pool of instances and start new instances based on availability
- Uses boto sdk for getting details, status of instanaces and availability group and fleet allocation and synced it with local postgres and redis
- Added functionality for supporting multiple AMI and instance types and fetching instance types from config file under single zone
- Added multiple background cron jobs to stop unused instances and continous data sync between instance pool

Technical Skills

Languages: Python,JavaScript/TypeScript, Go, Dart, Java, C/C++
Frameworks: Flutter, Node.js, Flask, FastAPI, Express
Tools: PostgreSQL, Redis, MongoDB, Elastic Search, AWS

Certifications **AWS Certified Cloud Practitioner - Validation Number**: WL78T7D25MQQ1C30