Kapil Raghav

Backend Engineer | Python | FastAPI | AWS | GenAI

+91 6350246985 | kapilraghav1221@gmail.com | linkedin.com/in/kapil-raghav | github.com/Raghav-2801

PROFESSIONAL SUMMARY

Backend Engineer with 3+ years of experience specializing in **high-concurrency distributed systems** and Generative AI solutions. Proven expertise in **Scalability**, architecting enterprise-grade **RAG pipelines**, and optimizing vector search retrieval. Specialist in developing low-latency **Microservices**, **API Security**, and automating mission-critical ETL workflows that doubled retrieval accuracy for global pharmaceutical leaders.

TECHNICAL SKILLS

Programming Languages: Python, SQL (PL/SQL), C++, HTML/CSS
Backend Frameworks: FastAPI, Flask, Django, Pytest, Pydantic, NumPy
Cloud & AI Infrastructure: AWS (Bedrock, S3, OpenSearch, Knowledge Bases), Docker, Kubernetes, CI/CD, Git
Databases & Foundations: Oracle DB, SQLite, Vector Databases (OpenSearch), System Design, Microservices, OOP

PROFESSIONAL EXPERIENCE

Cognizant Technology Solutions   Noida, India
Associate (Backend / GenAI Engineer)   *Aug 2023 – Present*
- Architected an **End-to-End RAG System** using FastAPI and AWS Bedrock for natural language search across **100k+ pharmaceutical SOPs** with optimized semantic retrieval.
- Engineered a **Two-Stage Retrieval Pipeline** with custom re-ranking (Claude 3 Haiku); increased retrieval precision from **40% to 80%** for enterprise document sets.
- Built a **Python ETL Pipeline** for real-time delta-sync between SharePoint and AWS OpenSearch, ensuring **100% data freshness** via multi-threaded synchronization and metadata tagging.
- Implemented **LLM Guardrails** and Pydantic validation, reducing token costs by filtering irrelevant queries and ensuring high-integrity JSON outputs.
- Optimized ingestion latency via **Titan Text Embedding v2** and custom chunking, accelerating the pipeline from document update to search availability.

Cognizant Technology Solutions   Noida, India
Programmer Analyst   *Nov 2022 – Jul 2023*
- Automated business-critical Oracle reporting workflows, reducing manual operational overhead by **30%** via PL/SQL and BI Publisher solutions.
- Achieved a **15% reduction** in report execution latency through performance tuning on Oracle Discoverer and complex SQL query optimization.
- Facilitated technical training on **Oracle Apps R12** and PL/SQL procedure development for engineering teams.

Cognizant Technology Solutions   Noida, India
Programmer Analyst Trainee (Intern)   *Mar 2022 – Oct 2022*
- Developed a scalable hospital management backend using **FastAPI**, implementing comprehensive CRUD operations and relational data modeling.
- Established best practices for **RESTful API design** and documentation using Python backend fundamentals.

PROJECTS

URL Shortener Microservice | *Flask, Python, SQLite, REST API*   Dec 2024
- Developed a high-performance URL redirection service with unique alphanumeric ID generation logic.
- Designed clean API contracts for GET/POST endpoints with sub-second latency and persistent SQLite storage.

Scalable Task Management API | *FastAPI, SQLite, Pytest, HTML/CSS*   June 2024
- Built a RESTful backend supporting complex task state management, request validation, and structured error handling.
- Achieved high reliability by implementing automated unit and integration tests using Pytest.

Enterprise Backend Service | *Django, Python, ORM, SQLite*   Sep 2023
- Architected a backend using Django ORM for efficient data modeling and clean separation of concerns.
- Integrated basic authentication and request validation to secure sensitive API endpoints.

EDUCATION

ABES Engineering College   Ghaziabad, India
B.Tech in Electronics and Communication Engineering   *2018 – 2022*