Sanjay Kumar

📞 +91-7296061111 ✉ trmpsanjay@gmail.com 🔗 linkedin.com/in/trmpsanjay 🐙 github.com/trmpsanjay

Education

Indian Institute of Science (IISc)	2024 – 2026
M.Tech in Artificial Intelligence	*Bengaluru, Karnataka*
Cochin University Of Science And Technology	2018 – 2022
B.Tech in Computer Science and Engineering (GPA: 8.23 / 10)	*Kochi, Kerala*

Experience

Cisco Systems	Jan 2022 – Current
Software Engineer	*Bengaluru, India*

- Led the design and development of **Agentic AI and RAG applications** that execute **action-based workflows** such as license discovery, assignment, expiry checks, and transfers, demo licenses powered by real-time **DB/API tool calls** with **role-based access control**.
- Built an enterprise-grade chatbot (**AskLicense**) integrating Generative AI with enterprise licensing systems, enabling users to query, manage, and transfer licenses conversationally.
- Developed a **multi-database comparison tool**, reducing processing time by **80%** and improving data accuracy by **20%**.
- Orchestrated **batch and streaming data pipelines** (PostgreSQL/GCP → Snowflake), ensuring uninterrupted, scalable data flow.
- Architected and deployed complete **DevOps infrastructure** for GenAI apps using **OpenShift, Docker, Kubernetes**, ensuring scalability, monitoring, and high availability.
- Drove delivery of a new **Centralized Password and Configuration Management product**, eliminating redeployment needs due to config changes and **reducing downtime to zero**.
- Spearheaded caching features for a data pipeline tool, boosting efficiency up to **80%** for very large tables.
- Built multiple POCs using **Java Spring Boot**, demonstrating adaptability and validating architectural concepts.

Projects

AskLicense (Agentic AI Licensing Assistant) | *Python, LangChain, LangGraph, LangSmith, MCP, OpenShift, Kubernetes, SQL*
- Designed and built an **Agentic AI application** enabling users to query and manage software licenses conversationally.
- Implemented **action-based workflows** such as license discovery, expiry checks, transfers, and demo provisioning using real-time **DB/API tool calls with role-based access control**.
- Integrated **Generative AI** with enterprise data systems to answer documentation and compliance queries.
- Deployed with a fully automated **DevOps infrastructure** on OpenShift and Kubernetes, ensuring scalability and monitoring.

Centralized Password Management | *Python, FastAPI, Oracle, Vault, Splunk, Kubernetes*
- Built a microservices-based system with **secure secret storage, role-based access control, and passwordless authentication**.
- Eliminated downtime caused by config changes and improved overall security with **real-time Splunk monitoring and alerts**.

Data Pipeline Orchestration & Integration | *Python, SQL, PostgreSQL, Kafka, GCP, AWS S3, Snowflake, Oracle, CockroachDB*
- Engineered **batch & streaming pipelines**, enabling real-time data flow from PostgreSQL/GCP to Snowflake.
- Optimized pipelines with caching and validation, reducing transfer time by **50%** and improving data accuracy.

Technical Skills

Generative AI: RAG, LangChain, LangGraph, LangSmith, MCP, Azure OpenAI , Gemini, Anthropic,Llama
Programming: Python, Java, C(Basic), C++(Basic), JavaScript(Basic)
Data/Cloud: SQL, PostgreSQL, Kafka, Snowflake, AWS S3, GCP(Basic), CockroachDB
DevOps: Docker, Kubernetes, OpenShift, Jenkins(Basic), Git
Frameworks/Tools: FastAPI, REST API, Pandas, ETL, Spring Boot(Basic), PowerBI(Basic), Tableau(Basic)
Core Concepts: Data Structures and Algorithms, Problem Solving, Artificial Intelligence, Machine Learning, Neural Networks, API, Database