

**Sankar**

**Agentic AI, MCP Architect and Developer**

[sankarb5859@gmail.com](mailto:sankarb5859@gmail.com)

6508610752

**Summary:**

* Seasoned software engineer with **20** years of experience driving innovative solutions across diverse industries, including AI, cloud computing, API Gateway, and payment processing. Proven Agile collaborator with cross-functional teams, bridging technical and business objectives.
* Currently, building MCP, Model Context Protocol servers to support resources and a set of tools for an enterprise business; I also built MCP clients to be associated with LLM using LangGraph agentic workflow.
* Developed skill sets to build RAG applications using managed services such as Amazon Bedrock, Knowledgebase, Vector DB; I am also confident to build these services in Python using LangChain, Pytorch frameworks.
* Investigated replacement of traditional Machine Learning with generative AI models, which could generate new synthetic data, provide more robust forecasting and planning, learn from new data sources, and adapt to changing business scenarios.
* As part of training, I am currently building agents for Ground Air Traffic Controller, Tower ATC and Pilots for their voice communication using voice to text and text to voice Google API on LangChain platform with Swarm Architecture of LangGraph framework. The workflow is orchestrated and automated to support human Air Traffic Controllers.
* In Swarm Architecture the state is maintained and could be shared with multiple agents. For instance, the customer information could be shared with Flight Booking Agent as well as with Hotel Booking Agent.
* At PepsiCo, I was responsible for building the pipeline for data flow from silver layer to gold layer; learning workflow for scenario generation, demand forecasting, and financial evaluation. At the end of the optimization workflow, the model created the planning cycle promotional calendar in excel format with PPG, Pricing and Promotion information for a twelve-week cycle. We deployed the read and download Planning Cycle at Azure API Gateway and created Planning Cycle and model training, data flow was protected within VPN.
* Comfortable and with confidence, I could integrate enterprise data and applications with evolving Gen AI and Learning in a reliable and scalable way using relevant frameworks, such as Fine Tuning, RAG, Agentic AI or MCP, based on use cases.

**Skills:**

**Financial and Banking:** JP Morgan, Key Bank merchant account onboarding, Lowe's implementation of PCI, payment card industry compliance

Natural Language Processing: at High wire, used Stanford NLP for content indexing. PepsiCo Gen AI and RAG, built word2vec for embedding.

**Machine Learning:** Linear Regression, Logistic Regression, k-Nearest Neighbor

**Generative AI:** Natural Language Processing, Learning, Named Entity Recognition, Intent Awareness and Sensing, Large Language Model, Reinforced Learning, Fine Tuning, Demand Forecasting, Financial Evaluation, Model Context Protocol

**Tools and Technologies:** Stanford CoreNLP, LangChain, LangGraph, Hugging Face, Meta AI, Oracle Database 23 ai, Generative Pre-Trained Transformer, Chunking and building Knowledgebase, RAG (Retrieval Augmented Generation), Graph RAG, MCP

**Techniques:** Classification, Linear Regression, Scenario Generation, RLHF (Reinforced Learning with Human Feedback), SFT on Vertex AI platform, MCP for integration of Resources and Tools with Microservices

Programming Languages: Java, Python

**Frameworks:** MCP, TensorFlow, PyTorch, Hugging Face, LangChain, LangGraph Spring Boot, Fast API, LangFuse

**Platforms:** Kubernetes, AWS Bedrock, Sagemaker, Microsoft Azure Cloud, Google Cloud Platform (GCP), API Gateway

**Cybersecurity:** Built secured applications using Fortify and Sync tools

**Other:** Microsoft SQL Server, Red Hat Drools, Google gRPC, Apache Solr

**Data Engineering:**

* **Market Data Processing:** At ValueStox-AI, every night, after the market close, we retrieve the financial data, classify them by sector and industry; we also validate the volatility of each stock and classify them as high volatile, medium and low volatile; that would help investors, especially the Listed Options to decide to invest on PUT or CALL options
* **Competitor’s Sales Data:** At PepsiCo, for Planning Cycle, we process previous year’s sales, marketing and financial data; those are well structured. We also process competitors’ promotion, pricing and sales data; those are not structured and have missing fields; we enhance the data with available information
* **Enterprise Information:** At JP Morgan, every night, as well as on demand, we generate reports for senior management; that involves huge volume of data processing; we adopted ETL and ELT, based on various use cases
* **Digital Publishing Platform**: At Highwire, we built digital publishing platform to support 100 publishers and 4000 journals; we persist the metadata of each journal in NoSQL datastores Cassandra and MongoDB; we built multiple channels in Kafka to process high value or high-speed publishing.
* **Domain Driven Design:** In the last two decades, many organizations have built microservices based on DDD and CQRS patterns; instead of low-level data, we could effectively handle Products, Customers, Pricing Architecture and Promotional Mechanics. For instance, to promote products for Superbowl event at PepsiCo, we have clear bounded context for data processing.

**Experience:**

**Value Stox AI February 2025 - Present**

**AI Architect and Python Developer**

**Technology:** LLM, Python, Prompt Engineering, Gen AI, MCP, LangGraph, Google Cloud Platform, NVIDIA GPU, Docker and Kubernetes

* https://valuestox-ai.web.app/
* Since February 2025, I have been consulting at startup, Value Stox AI. I am building MVP, a minimum viable product for value-based investment with Gen AI.
* It has various contexts: stock news; sector analysis; industry analysis; company analysis; portfolio. We have a common chat window for all contexts; based on the current contexts, the backend process would interact with LLM only or RAG or MCP; for instance, portfolio needs to interact with MCP Portfolio server for tools to get the list of current investments

**Public Library (Volunteer/Pro Bono) May 2024 - November 2024**

**AI Architect and Python Developer**

**Technology:** LLM, Python, Prompt Engineering, Gen AI, MCP, LangGraph, Google Cloud Platform, NVIDIA GPU, Docker and Kubernetes

* Designed and developed Library AI, integrating local library catalogs with generative AI using LLM technology to provide users with detailed and contextual responses.
* Implemented metadata retrieval from Z39.50 SRU servers and utilized entity relationships via Graph RAG to enhance user queries, retrieving book information such as title, author, and availability.
* Built and deployed the Library AI platform using LangChain, Gemini, Fast API, and Python, with a scalable inference API and entity relationship management on LangGraph.
* I built the MCP, Model Context Protocol Server and MCP Client in Python for the query and retrieval from Z39.50 SRU servers.

**PepsiCo, Plano, TX July 2023 - Apr 2024**

**Senior AI Solutions Architect and Python Developer**

**Technology:** LLM, Python, Prompt Engineering, Gen AI, Machine Learning, Azure, GPU, Docker and Kubernetes

* At PepsiCo, I have automated onboarding of developers with Azure API and integrated with Fortify, Snyk API for application security. All API outside VPN were protected by Azure API Gateway for authentication and also for authorization
* Based on Domain Driven Design, I built Aggregate, Entity and Value objects in Java version 17 using Spring Boot framework and I also built RESTful API for React frontend integration
* Developed in Python, Planning Cycle AI for promotions of products, using traditional ML models, including classification and linear regression, for predictive analytics, such as scenario generation, demand forecasting, and financial evaluation, but encountered limitations due to stationary data and inability to adapt to changing business conditions
* Investigated replacement of traditional ML with generative AI models, which could generate new synthetic data, provide more robust forecasting and planning, learn from new data sources, and adapt to changing business scenarios, but faced challenges including high demand for data quality, complex models, scalability issues, and significant computational resource requirements.
* Designed and implemented hybrid models, combining traditional learning with Large Language Models (LLMs), to augment datasets, identify new features, recognize and generate outliers and anomalies, and effectively handle missing or incomplete data
* Integrated Retrieval-Augmented Generation (RAG) to enhance traditional ML models, identifying and retrieving relevant data points and features from knowledge graphs, including market, retailer, product, pricing, and promotional mechanics data, ranking and filtering entities based on relevance, and generating output based on ranked data
* At the end of the training, the Planning Cycle Promotional Calendar was generated in excel format. Created the vector space by chunking the excel file; generated embeddings; persisted the embeddings in Oracle 23 ai database. Also created the conversation template for business users; captured their query; retrieved the context from the vector database; augment the query; and generated the response in natural language.
* Built the release pipeline to test and certify the applications on cybersecurity perspective using Fortify and Snyk tools

**PepsiCo, Plano, TX Apr 2021 - July 2023**

**Senior Software Architect for GNRM**

**Technology:** LLM, Python, Prompt Engineering, Artificial Intelligence, Machine Learning, Java, Azure, AKS, Docker and Kubernetes

* Designed and developed a cloud-based GNRM application using ReactJS, Java Spring Boot microservices, and Azure, leveraging continuous integration and continuous delivery (CICD) pipelines
* Spearheaded development of Promotional Calendars for global retailers, including UK Tesco, Kroger, and Albertsons, implementing single sign-on security with Microsoft Identity providers and fine-grained authorization
* Designed AI and implemented the Scenario Generation, Demand Forecasting and Financial Evaluation using Machine Learning algorithms, Linear Regression and Logistic Regression for Classification
* Implemented business rules engine using Drools framework, validating and alerting on key sales requirements, and designed rules engine for expandability across various markets and pricing mechanics
* Adopted the 12-factor methodology for cloud-native application compliance and implemented Command Query Responsibility Segregation (CQRS) for efficient handling of user actions and scalability.
* Built the CQRS and isolation by Market and Retailers in Java version 17 using SpringBoot framework.
* Mentored a global team of 25 developers, providing guidance on business domain and technology, and automated testing using Cucumber framework

**KeyBank, Cleveland, OH Sept 2020 - Jan 2021**

**Architect and Tech Lead for Merchant Onboarding**

**Technology:** Java, Payment Services, Domain Driven Design, AWS

* Led the Merchant Onboarding Automation project technical team, delivering a payments platform on Azure cloud that integrates with KeyBank's Core Services and First Data's (Fiserv) Payments Processor service
* Designed and implemented seamless interactions with supporting services, including Boomtown web chat and Salesforce's Sales Lead, Opportunity, and Case Management modules

**Lowe’s, Charlotte, NC Aug 2019 - Aug 2020**

**Architect and Principal Software Engineer for Payments**

**Technology:** Java, Payment Services, Domain Driven Design, Inhouse data center

* Designed and built PCI-compliant core services for payments applications, supporting multiple payment systems including Credit Card, Google Pay, PayPal, and Apple Pay
* Developed decryption, tokenization, and fraud check services to ensure secure payment processing
* Introduced gRPC for internal microservices communication, improving system efficiency and scalability, as approved by customers
* Deployed services on internal data centers, adhering to PCI requirements for Control Data Environment (CDE)
* Applied 12-Factor methodology to ensure scalable, maintainable, and efficient service architecture

**JPMorgan Chase, Jersey City, NJ Sept 2017 - July 2019**

**Architect and Principal Software Engineer**

**Technology:** Java, Asset Management, Domain Driven Design, Private Cloud

* Developed modern financial security trading applications for Equity, Listed Options, Foreign Exchange, and Fixed Income at JPMC, using technologies like Spring, Cloud Foundry, and ReactJS
* Refactored a legacy application on Gaia cloud, leveraging latest Java, Spring Boot, MyBatis, and Greenplum backend database, without impacting business use cases and performance
* Integrated the refactored application with Microsoft Reporting Server, storing generated reports in Neptune S3 storage bucket, and submitting events to MQ servers
* Designed and developed automation for onboarding Third-Party Applications to Access Gateway, integrating with external systems like IDAnywhere (Authentication), Ping Access (Authorization and Restriction), and DNS (Response Policy Zone)

**HighWire, Los Gatos, CA August 2013 - April 2017**

**Architect and Search Engineer**

**Technology:** Java, Solr, Domain Driven Design, AWS

* HighWire is a leading electronic publishing platform for scholarly publishers. I hosted highly scientific content for university scholars and medical professionals. I was responsible for the replacement of Microsoft FAST search engine with Solr Cloud. Built RESTful web services and automated the indexing process. Also built a highly available, scalable query service to support the demand for quality content with high precision. Here is the list of features:
* Stanford NLP for elimination of stop words, tokenization, stemming Real time indexing process automation

**Cisco. San Jose, CA Aug 2010 - July 2013**

**J2EE Middleware developer**

● This project is all about RMA (Return Material Authorization). It involves application-to-application integration for asset recovery. My responsibilities include, understanding of the RMA and Asset Recovery business processes; architectural design and documentation; technical design; XML Schema finalization; proof of concept development.

**Wells Fargo. Concord, CA June 2010 – Aug 2010**

**C# Developer**

* It was a short-term project at Wells Fargo, Concord developing the backend in C# and, .Net environment

**Web On-The-Go Jan 2009 - May 2010**

**Author**

* I published a book, “Web On-The-Go. It’s all about the potential applications that can be built on mobile platforms. I designed and built prototype applications to validate the concepts and document those observations in the book. Those applications were built on Service Oriented Architecture, with thin clients using Android mobile platform and the server applications on cloud computing platforms.

**Google. Mountain View, CA Aug 2006 - Dec 2008**

**Java Middleware developer**

* I built the dashboard to display the asset in transit from one datacenter to another in Java.

**International Asset Systems. Oakland, CA Aug 2004 - June 2006**

**Java developer**

* I built the shipping container tracking system using Enterprise Java.

**Cisco. San Jose, CA Aug 2000 - June 2004**

**Java developer**

* I built the warehouse system to manage inventory for receiving and shipping with handheld devices and also built the community for network engineers on Tomcat with Java.

**Skill Development:**

* Linear Algebra - Gilbert Strang, MIT Open Courseware YouTube channel
* Statistics - Data Tab YouTube channel
* Learning - Stanford CS229 YouTube channel
* Deep Learning - MIT 6.S191 YouTube channel
* Large Language Model - Andrej Karpathy YouTube channel

**Education:**

* Anna University, India – Bachelor of Engineering in Electronics in 1982