

Vinay Kudari

Tech expertise in E-commerce, Healthcare, Gaming Industries & Startups

vinay.kudari30@gmail.com

+1-469-943-6778

RELEVANT SKILLS

Python, SQL, JAVA, FastAPI, LangGraph, PySpark, Kubernetes, Git, Redis, Kafka, Terraform, GCP, AWS

EXPERIENCE

- Apple** Cupertino, CA
Software Engineer | **llm, generative-ai, agents, api-design** July 2024 - Present
 - Built a multi-agent system leveraging LangGraph and LLMs to convert natural language text into optimized SQL
 - Developing a general-purpose agentic system to streamline analysis of documents and databases, significantly reducing manual review efforts
 - Engineered a scalable full-stack financial analytics dashboard supporting real-time data streaming, enabling dynamic visualization and reporting
- Gap Inc** San Francisco, CA
Software Engineer | **java, spring, graphql, azure, kafka, kubernetes, llm** July 2023 - June 2024
 - Spearheaded a real-time personalization engine development, leveraging extensive datasets to optimize user experiences and drive significant engagement and sales enhancements.
 - Facilitated migration of legacy cloud foundry systems to kubernetes leveraging github actions
 - Engineering a low-latency LLM-based information retrieval system to extract data from multi-source document corpus
- Amazon Web Services** Palo Alto, CA
Software Development Engineer | **spark, athena, dynamo, glue, lambda, cdk** Dec 2022 - June 2023
 - Designed and published an internal query builder python package (ZenQL) tailored for AWS Healthlake
 - Designed an event-driven server-less system using AWS services, allowing for parallel execution of tasks, resulting in seamless scaling to handle concurrent requests and improved processing throughput
 - Designed and optimized complex SQL queries in Athena resulting in a 40% reduction in query execution times
 - Architected near real time data processing system using AWS CDK, leveraging PySpark for ETL transforms on multi-terabyte healthcare data sets
- Playstation** San Francisco, CA
Software Development Intern | **kafka, concurrency, async, no-sql, cassandra** May 2022 - Dec 2022
 - Worked on auto follow feature for PS5 Explore Hub. Message queues and REST endpoints were developed
 - Learned custom framework and leveraged asynchronous programming pattern to optimize the latency by 60%
 - Developed a cross platform CLI tool to automate generation of grafana analytical dashboards from logs
- Brave Orbit (Early stage startup - Founding engineer)** Capetown, South Africa
Full Stack Developer | **rest-api, django, flutter, pytest, redis, docker, gcp** Jan 2020 - Aug 2021
 - Programmed a cross-platform app using BLoC pattern, implemented database, network and state management layers
 - Engineered HIPAA compliant scalable async push notification service for medication reminder application
 - Developed a task scheduler for clinical surveys that delivers email, SMS alerts and gathers response on a recurring basis
 - Architected a custom ERP system with multi-product substitutions and combo-product life cycle tracking features
 - Setup Kubernetes CI/CD workflows to automatically test, build and deploy backend services on cloud
 - Streamlined automated testing pipelines and improved test coverage for both frontend and backend services
- Accenture** Bangalore, India
Big Data Developer | **redshift, pyspark, emr, etl, s3** Sep 2018 - Dec 2019
 - Performed time series analysis on drug assay data to identify raw materials that effect the drug quality
 - Designed ETL data validation pipeline using micro service architecture; Manual testing resources were cut by 70%
 - Developed PL/SQL procedures to generate parent child hierarchies using temporal drug life cycle data

PROJECTS

- Personal AI assistant (genai):** Building a multi-agent AI companion with human-like memory simulation, capable of personalized conversations, context-aware recall, and tool use
- Open-source contributions (nlp, open-source):** Core contributor to HuggingFace Datasets official repository
- Distributed key-value store (distributed systems):** Developed redis-like distributed key-value store with a focus on achieving strong consistency and fault tolerance using the RAFT consensus algorithm
- Full text search engine (search, nlp, information-retrieval):** Developed a query parser to transform search query into boolean expression; Optimized the text retrieval by creating inverted index and performed batch processing.
- Unordered image stitching and deghosting (vision, image matching):** Estimated approximate overlap area by minimum variance technique on binarized images; stitched images based on binary tree model; removed deghosting from the overlap areas by selecting regions based on minimum local gradients

EDUCATION

- University at Buffalo [3.9/4.0 GPA]** Buffalo, NY
Master of Science in Computer Science (AI/ML Specialization) Sep 2021 - Dec 2022
Research: Finding inconsistency between text and related images, advised by Prof. David Doermann
- National Institute of Technology, Durgapur [3.3/4.0 GPA]** West Bengal, India
Bachelor of Technology in Computer Science and Engineering July 2014 - May 2018