Santhosh Chandrasekaran

76000 Verano Road, A104, Irvine, California, 92617
650-661-9675 | santhoc@uci.edu | linkedin.com/in/santhosh-chandrasekaran-90410317a
github.com/santhoshCoder98 | dub.sh/santhoshportfolio

Professional Summary

Software Engineer with 3+ years of experience in full-stack web development, building scalable and high-performance distributed systems. Strong background in designing APIs, developing microservices, and optimizing system performance for cloud-based applications. Focused on creating efficient, maintainable, and user-centric web solutions that enhance scalability, reliability and user satisfaction

Education

Master of Computer Science **Sep 2023 – Dec 2024**
University of California Irvine *GPA: 3.77*

B.Tech in Information Technology, **Aug 2016 – May 2020**
Sri Venkateswara College of Engineering *GPA: 3.9*

Technical Skills and Areas of Interest

Programming Languages: C, C++, Java, Python, JavaScript, TypeScript

Frontend Development: HTML5, CSS3, React.js, Next.js, Angular.js, Vue.js, Tailwind CSS

Backend Development: Flask, FastAPI, Node.js, Express.js, GraphQL, PostgreSQL, MySQL, MongoDB, Redis

Machine Learning: Langchain, NLP, Vector Databases, RAG, LLMs (LLaMA, LLaVA), GANs, OpenAI

DevOps & Cloud Technologies: Docker, Jenkins, Azure, AWS (S3, SageMaker, CodePipeline, CodeBuild, CodeDeploy), Kafka, RabbitMQ, Dramatiq, Kubernetes

Areas of Interest: Full-Stack Development, Distributed Systems, Cloud Computing, AI and Machine Learning, DevOps and CI/CD, Containerization, Database Design, REST APIs

Work Experience

Software Developer Intern **June 2024 – Dec 2024**
Mage Data *New York, USA*

- Migrated legacy data storage to **PostgreSQL**, optimizing **query execution** and reducing **data processing** time by **65%**, which enabled **real-time** analytics and faster **decision-making** for stakeholders
- Implemented a **GraphQL API layer** in **Node.js using Express.js** with **PostgreSQL**, optimizing **query execution** and reducing **response times by 40%** for **real-time data retrieval**
- Containerized **GAN** models with **Docker and AWS SageMaker**, reducing manual workload by **50%** and accelerating model updates by **3x**, enabling **faster deployment** of synthetic data models for production use

Senior Software Engineer **Dec 2020 – July 2023**
Aspire Systems *Chennai, India*

- Led the development of **5+** large-scale **web** and **mobile** modules for banking applications, creating **35+** reusable **React** and **Redux** components, increasing feature adoption by **25%** among **2M+** users
- Engineered **scalable distributed backend** services, handling **1M+** daily requests, leveraging **Spring Boot, PostgreSQL, and Redis**, and cutting API response times to **<200ms** through caching and optimizations
- Designed and implemented **CI/CD** using **AWS tools** (**CodePipeline, CodeBuild, and CodeDeploy**), reducing **deployment times by 45%** and ensuring **90% on-time releases**
- Implemented **Agile methodologies (Scrum, sprint planning, Jira)** increasing sprint velocity by **20%**, while collaborating with cross-functional teams to streamline **backlog management** and accelerate releases
- Developed and streamlined **CI/CD pipelines** in a **Service-Oriented Architecture (SOA)** using **Jenkins** and **Azure DevOps**, reducing **deployment failures by 20%** and accelerating **release cycles by 40%**
- Optimized **database queries (Java, Spring Boot)** and implemented **Redis caching**, achieving a **20% reduction** in **API response times** and supporting a **50% traffic increase with zero downtime**

Projects

EmailIQ **Oct 2024 – Dec 2024**

University of California Irvine *Irvine, California*

- Developed an **AI-powered email client** using **Next.js 14 and React**, leveraging **NLP** and the **OpenAI API** to provide intelligent content structuring and enhance user productivity
- Engineered an **AI-driven query system** using **NLP** techniques, improving email search efficiency by **40%** through contextual understanding and response optimization
- Designed a **scalable serverless architecture** with **Neon Database, Prisma ORM, and Pinecone**, reducing infrastructure costs by **30%** while ensuring low-latency email retrieval and a seamless user experience

ChatPDF **July 2024 – Sep 2024**

University of California Irvine *Irvine, California*

- Developed and deployed a full-stack ChatPDF application using **Next.js and FastAPI**, using **NLP and LLM technologies** to enable users to query and extract insights from PDFs via natural language interaction
- Improved document retrieval efficiency by **50%** and enhanced search relevance by integrating **OpenAI embeddings** with **Chroma vector database**, contributing to a more effective **NLP** solution
- Designed an optimized **Retrieval-Augmented Generation (RAG) pipeline** using **LangChain** to enhance the accuracy of **LLM**-powered responses by **35%** for contextual document-based queries
- Fine-tuned a **LLaMA** large language model **(LLM)**, improving its contextual understanding and enabling it to generate more precise answers to complex document queries, demonstrating expertise in **LLM** optimization

Video Indexing and Search Text Analysis | VISTA **Jan 2024 – June 2024**

Amazon Web Services (AWS) Sponsored, University of California Irvine *Irvine, California*

- Designed and optimized **scalable distributed backend systems** using **FastAPI**, achieving a **45%** reduction in API latency through **Amazon S3 integration** and efficient asynchronous processing
- Optimized **SEO and content discoverability** by implementing **server-side rendering (SSR)** in **Next.js** and refining **caching strategies**, leading to faster load times and increasing business impact
- Implemented an AI-driven video analysis pipeline, leveraging **Hugging Face Transformers** for **OpenAI Whisper** speech-to-text processing and **LLaVA** models, achieving **98% transcription accuracy**
- Enhanced the efficiency and reliability of **machine learning** inference with **RabbitMQ and Dramatiq**, enabling **fault-tolerant** and **sequential execution** of large-scale multimodal AI tasks

LogFS File System **Oct 2023 – Nov 2023**

University of California Irvine *Irvine, California*

- Built a **key-value store** on a log-based file system to efficiently read, write, and update key-value pairs in **C**
- Achieved high read-write speeds to disk using **POSIX threads** to asynchronously write data and read data from the queue to flush onto disk. Leveraged **POSIX APIs** to read and write blocks of data for the same.
- Added a direct mapping **caching mechanism** to ensure the latest data could be read quickly to avoid unnecessary flushes to disk

Dynamic Thread Scheduler **Nov 2023 – Dec 2023**

University of California Irvine *Irvine, California*

- Designed and implemented a **custom thread** management system with capabilities for thread creation, **stack allocation**, and **context switching**.
- Improved resource utilization and minimized overhead by integrating signal handling, leveraging **setjmp/longjmp**, and utilizing stacks and queues for fast and lightweight context switching.
- Developed a **round-robin scheduling** algorithm to ensure fair task execution and implemented utility functions for **memory allocation**, sleep functionality, and robust error handling.

Achievements

Role Star Award- Recognized as a top performer at Aspire Systems for outstanding contributions to backend engineering and scalable system development.

Amazon Research Grant ($6,000)- Awarded funding from **Amazon Web Services (AWS)** for developing an AI-driven video indexing and search pipeline, optimizing large-scale ML processing.