

Avinash Ramesh

Remote, Tx | (408) 338-5862 | avinashramesh2312@gmail.com | www.linkedin.com/in/avinash94/ | Github: github.com/rameshavinash94/

SUMMARY

7+ years building scalable backend systems. Expertise in API design, microservices, devops, cloud and big data frameworks.

SKILLS

Languages: Python, Java, TypeScript, JavaScript, SQL
Cloud & DevOps: Azure, AWS, Kubernetes, Terraform, GitHub Actions, Argo CD, helm charts.
Web & frameworks: Node.js, ExpressJs, Flask, FastAPI, React, Java Spring Boot, HTML, CSS
Testing: Cucumber, Playwright, Mocha
Data Storage & Processing: PostgreSQL, Blob, DynamoDB, Redis, Hadoop, HDFS, YARN, Sqoop, Kafka, Spark, Hive

EXPERIENCE

Amazon Web Services, Seattle, WA
Software Development Engineer II - GenAI / ML infrastructure Dec 2025 - Present

- Architected and implemented scalable automation frameworks/systems for NVIDIA GB200 and GB300 UltraServers within AWS Hardware Engineering Services, enabling streamlined lifecycle management of high-density AI systems.

Carl Zeiss Meditec, Inc. Dublin, CA (Remote, Tx)
Senior Software Engineer (Product - Zeiss Veracity - Zeiss Health Data Platform) June 2023 - Nov 2025

- Architected and maintained a high-performance Tier 1 service interfacing with Zeiss FHIR servers, processing millions of requests daily to provide real-time access to patient information for applications and devices.
- Led the migration of core EMR communication microservices from AWS to Azure, consolidating five services into two. This initiative optimized inter-service communication and simplified deployments.
- Designed and deployed a dockerized EMR agent system that listens to network share paths and MLLP servers to efficiently sync files and HL7 messages from EHR systems to our cloud, reducing data update latency by 40%.
- Built and managed a comprehensive Azure infrastructure for the project. I established an automated CI/CD pipeline and designed secure certificate management to boost deployment efficiency.
- Developed a custom observability lib on top of Winston Logger to centralize logging and metrics, streaming data to App Insights to significantly improve system monitoring and streamline debugging.
- Engineered an automated schema migration tool and integrated it into the CI/CD pipeline with GitHub Actions, ensuring database integrity and minimizing downtime by executing schema changes before service deployments.
- Developed a scalable log service to retrieve and process payloads and service responses from storage queues, storing the data in our encrypted PostgreSQL database for internal error and log analytics.

Software Engineer Intern - Full Stack (Team: Zeiss MedAI - Internal Tools) May 2022 - December 2022

- Architected a comprehensive DICOM data management solution, streamlining clinical research and analytics, resulting in a more than 50% reduction in data preprocessing time.
- Spearheaded the end-to-end design and development of the clinical data management internal product, encompassing backend services, micro frontend, infrastructure provisioning, and data pipelines.
- Modeled a hierarchical database for DICOM (Patient, Study, Series, etc.) with PostgreSQL, significantly improving imaging data ingestion and retrieval speed.

ITI DATA - EXL Company (Client: CitiGroup), Chennai, India
Associate - Senior Software Engineer (Big Data Platforms) July 2019 - July 2021

- Designed and implemented a service to detect data anomalies and generate automated daily reports, which reduced issue resolution time by 50%.
- Owned a robust backend system to support internal & external vendor data feeds ingestion, creating a centralized 'golden' data store with a prioritized data selection hierarchy.
- Migrated data pipelines to Spark, achieving a 40% performance improvement and substantial cost savings in tool licensing.

Analyst - Software Engineer August 2016 - June 2019

- Designed and developed a microservice to facilitate seamless data creation, retrieval, updating, and deletion of vendor data, providing comprehensive control and flexibility for managing information.
- Automated manual tasks using shell scripting and python, redesigned processes for efficiency, and enhanced existing service performance by implementing a caching layer.

EDUCATION

San José State University San Jose, CA, USA <i>Master of Science in Computer Software Engineering (Specializing in Data science)</i>	August 2021 - May 2023
Anna University Chennai, TN, INDIA <i>Bachelor of Technology in Information Technology</i>	June 2012 - April 2016

CAREER HIGHLIGHTS

- ❖ Strong Engineering background with 7+ years of experience and a Ms in Software Engineering(Data science track).
- ❖ Proficient in Python, Typescript(Nodejs, Express), & SQL with expertise in REST APIs and microservices design.
- ❖ Have a proven track record of building customer-centric solutions that are robust, scalable, and easy to use.
- ❖ Follow object oriented design patterns and dependency ingestion techniques to develop loosely coupled applications.
- ❖ Designed and deployed various AI/ML projects during my master's and gained profound knowledge of the ML lifecycle.

PROJECTS

- [EditScope \(Transcription based video editor with AI features\)](#)
Innovated an AI-powered audio and video editing tool, streamlining collaboration and accessibility for users of varying expertise. Seamlessly integrating cutting-edge features like real-time voice cloning, automatic transcription, and text-to-speech conversion for swift and intuitive content modification.
- [Wiki Question & Answering \(QA\) App.](#)
Designed a dynamic Wiki QA Web Application, harnessing NLP to swiftly answer user queries by intelligently sourcing and analyzing Wikipedia content, streamlining information retrieval and enhancing user experience.
- [App for Detection & Classification Of ECG Images](#)
This product aims to leverage image processing and machine learning to transform ECG images into a 1-D signal, extracting key heart activity elements like P, QRS, and T waves. These insights facilitate diagnosing a range of cardiac conditions, enhancing medical assessments with advanced techniques.
- [Patient Management System \(Book/Cancel Appointments with Doctors\)](#)
Developed a Patient Management System enabling efficient booking and cancellation of appointments with doctors, enhancing healthcare access and scheduling convenience.
- [Intention based ML/ Deep learning approach for fake news detection.](#)
The project focuses on enhancing fake news detection using a novel approach that captures the underlying deceptive intentions. Utilizing deep learning techniques, it combines linguistic features like cognitive complexity, verbal immediacy, emotional experience, and sentence embeddings to classify news articles into categories of real, fake with deceptions, and fake without deceptions.
- [DL Framework for Facemask Detection using CNN, Facemask-Removal using GANs and Gender Classification](#)
This project leverages ML to classify masked faces, predict masked individuals' appearance, and determine gender, addressing privacy concerns in surveillance. It offers potential for biometric authentication and will be accessible through a user-friendly web application.
- [Company Analysis - Knowledge Graph](#)
This project creates a knowledge graph for investment decisions by combining data from APIs like FinancialModelingPrep, PeoplesData, NewsAPI, and scraping Crunchbase. After NLP preprocessing, the graph is stored in Neo4j, providing users with summarized insights.
- [Real Estate Analysis.](#)
We've innovated a strategy to pinpoint the most profitable and secure real estate investments through a 'golden cluster' approach, integrating latent variables like crime rates and walkability metrics. Our classification and regression techniques further refine investment decisions and property price predictions.
- [Retail Store Analysis](#)
Utilizing Olist's Brazilian retail dataset, this project employs clustering for churn risk, classification by Customer Lifetime Value, and regression to predict delivery dates for customer satisfaction. The dataset of 100k anonymized orders between 2016 and 2018 empowers data-driven retail insights.
- [Poem Generation](#)
The project involves building a Poem class that leverages web scraping and NLP techniques to process and analyze poetry data. It aims to create new poems by replacing words based on POS analysis, enhancing the understanding and generation of poetic content.
- [Flight Booking Application.](#)
Focused on resolving the challenge of creating a flight booking application that incorporates diverse design patterns and software engineering best practices. The goal is to establish a versatile and reusable solution for addressing the intricacies of the problem domain.