

SURYA RAMAN

SOFTWARE ENGINEER. DOG PARENT. MOTORCYCLE ENTHUSIAST.

Toronto, Canada | suryaramansuresh@gmail.com | [Website](#) | [LinkedIn](#) | [Github](#)

Skills

Node.js	TypeScript	MySQL	PostgreSQL	Elasticsearch	Kubernetes	Docker
Python	Golang	Redis	MongoDB	Kafka	OpenTelemetry	Git

Work Experience

Engineering Manager | yellow.ai JAN 2024 - PRESENT, TORONTO, CANADA

- Owned and scaled the complete messaging infrastructure (channels & integrations) for the yellow.ai GenAI platform, enabling the **reliable processing of over 8 billion messages** per year.
- Managed the team that architected and delivered a self-serve ETL pipeline for RAG (using agentic chunking, vector databases), achieving a **100% reduction in customer data ingestion wait times**.
- Managed a team of 14 engineers** across three verticals, resulting in improved team capabilities and project velocity.

SDE - 3 | yellow.ai JAN 2023 - DEC 2023, TORONTO, CANADA

- Increased yellow.ai platform's uptime to 99.9%** by eliminating Single Points of Failure across the system through **Geographical Database replication**, Database Proxies, and **geo-steered load balancing**
- Implemented a **blue-green data migration strategy** for chatbots, resulting in **zero downtime and a 170% increase in migration success rate**

SDE - 2 | yellow.ai JAN 2021 - DEC 2022, BANGALORE, INDIA

- Spearheaded the engineering efforts for building and maintaining WhatsApp Business **Kubernetes** infra
 - Collaborated with the WhatsApp/Meta** team to run load tests with various Kubernetes workload configurations and understand the optimal resource configuration
 - Revamped services to provision new workloads in Kubernetes with this configuration, resulting in a **400% increase in messaging throughput** and **reduced the cost of the infrastructure by 88% per month**
 - Built a proxy to port customers using legacy APIs to newer APIs with **Node.js, MongoDB, and Redis**
 - Setup **Prometheus** dashboard and introduced alerts for all aspects of the system (queue size, IOPS, remaining disk size) that **increased observability to 100%**
- Used **New Relic** to identify problem areas in a core authentication service; Introduced cache and rewrote legacy code to **reduce latency by 124%**
- Introduced **artifact management** with **Nexus**, converted an internal monolith into smaller libraries
 - Integrated this with a **Jenkins** pipeline to do automatic **semantic versioning**

SDE - 1 | yellow.ai MAR 2018 - DEC 2020, BANGALORE, INDIA

- Introduced admin panel for managing the yellow.ai platform using **React, Ant Design, and Typescript**
- Implemented OAuth 2.0 with OpenID Connect (OIDC), Proof Key for Code Exchange (PKCE)** which allowed clients to access the yellow.ai platform via any OAuth 2.0 provider
- Implemented a service for converting HTML/URL to pdf/image with **Typescript, Puppeteer, and browser pool**
- Improved chatbot load time by optimizing code, lazy loading assets, and client-side caching which **reduced page load time from 6 seconds to 1.2 seconds**

BI Developer | Visual BI Solutions (an Atos Company) MAY 2017 - MAR 2018, CHENNAI, INDIA

- Integrated Business Intelligence tools for the product **BI Hub** such as Microsoft Power BI, SAP Business Objects, SQL Server Reporting Services, and TIBCO Spotfire using **Python, MySQL, and Java**

Education

Computer Science and Engineering | Anna University AUG 2013 - MAY 2017, CHENNAI, INDIA

- Implemented Capstone Project on "BLE Tracking System" using Arduino, NRF24L01, Android, and Firebase