

# Nikhil Baskar

8667877361 | vishalnikhil0307@gmail.com | [github.com/Nikhil0307](https://github.com/Nikhil0307) | [linkedin](#)

## Skills

---

**Technical Skills:** Python, Java, Scala, SQL, HTML, CSS, JavaScript, Bash.

**Frameworks & Tools:** CherryPy, Django, Servlets, Spring, ONNX, gRPC, Docker, Kubernetes, Redis, MongoDB, MySQL, PostgreSQL, CI/CD, MLOps, Git.

**Core Competencies:** Agile methodology, Waterfall methodology, Performance optimization, Mentorship

## Experience

---

**Software Developer, Zoho Corporation – Chennai** 2023 – Present.

- Enhanced Zoho CRM AI/ML model serving times through the resolution of high latency concerns during model loading from a storage medium.
- Implemented an in-memory cache for models that are used frequently and a distributed cache system using gRPC and protobuf was designed to bridge the gap between the storage medium.
- gRPC and protobuf were chosen after extensive research and benchmarking as the most appropriate options due to their speed and efficiency over the current HTTP1-based REST API.
- Enhanced the speed and efficiency of ML workflows, demonstrating a strong commitment to optimizing inference performance.
- 50% less latency was seen when providing the model, increasing customer satisfaction, product availability, and market success.
- Mentored interns, evaluating their performance and guiding them to expand their knowledge

**Software Developer Intern, Zoho Corporation – Chennai** 2022 – 2023.

- Managing diverse ML models built by separate teams on a server created dependency conflicts and hindered upgrades. Containerizing models in individual Kubernetes pods offered versioning flexibility but resulted in resource under/over utilization due to pod proliferation.
- Introduced the ONNX framework to streamline model serving by tweaking the existing flow. Converted supported models to ONNX format, categorized models as ONNX supported and unsupported, and optimized pod usage with ONNX runtime for efficient resource management by serving all ONNX supported models in a single pod with dynamic scaling.
- Reduced model sizes and enabled in-browser inference with ONNX for smaller models, significantly removing load times for models loaded in-browser.
- Reduced resource usage, resolved dependency conflicts, enabling seamless integration of diverse models without impacting latency.

## University Projects

---

### Home Food Application (*freelancing*)

- Developed a real-time web application using Python, boosting customer base by 20% for a home-based food company.
- Enabled pre-booking of meals through the website, enhancing order management and customer convenience.
- Automated data handling with Excel, ensuring efficient tracking and periodic data purging for privacy.

### Cook Book - Recipe Web Application

- Developed a Recipe web application using Struts, Servlet, and Spring in the backend to gain experience with legacy systems and modern Spring technology.
- Designed, utilized and implemented a cache system to reduce backend calls and leveraged Apache Lucene for optimizing search functionality and the ELK stack to make search results faster and more efficient.

### Taxi Booking Application

- Developed a taxi booking app as a way to understand and incorporate various technologies.
- Using PHP and MySQL, I built the backend, which improved my knowledge of server-side development with PHP.
- Enhanced user experience with integrated MapBox API for location selection and driver navigation.

## Education

---

**Veltech Multitech Dr Rangarajan Dr Sakunathula Engineering College – Be in Computer Science**

2023