

Soumyendra Shrivastava

☎ (669) 900-3079 ✉ shrivastavasoumyendra@gmail.com in LinkedIn 🌐 GitHub 📍 San Jose, CA

Education

Master of Science in Software Engineering

GPA: 3.85

San Jose State University, San Jose, CA

May 2024

Relevant Coursework - Operating Systems, Distributed Systems, Advanced Parallel Processing, Machine Learning, Deep Learning, Advanced Data Mining

Bachelor of Technology in Information Technology

GPA: 3.76

Rajiv Gandhi Technical University, Bhopal, India

Jul 2021

Relevant Coursework - Operating Systems, Computer Networks, Cloud Computing, Data Structures, Design and Analysis of Algorithms

Skills

Languages: Java, Python, JavaScript, C#

Technologies/Tools: Amazon Web Services (ECS, EC2, ELB), Google Cloud Platform, Microsoft Azure, Kubernetes, Docker, Kafka, Tableau, MongoDB, MySQL, CI/CD, Jenkins, Git, Ansible, Agile, Elasticsearch, REST API, Linux

Frameworks: ReactJS, Angular, Node.js, Express.js, Flask, Django, Spark, HTML5, CSS3, PHP

AI/ML Stack: Scikit-learn, Tensorflow, PyTorch, LangChain, LLMs, OpenCV, D3.js, PyCaret, Pandas, NumPy, TensorRT

Experience

Software Engineer - Data

Aug 2024 — Present

Google(xWF)

Sunnyvale, CA

- Built an AI-powered platform with custom DAG-based pipelines for processing user-uploaded documents, improving data processing efficiency by 50% using Flask for the backend and Angular for the interface.
- Developed standalone microservices for document-specific workflows, training models from scratch and integrating role-based access control, increasing processing accuracy by 40% and ensuring secure, scalable operations.
- Enhanced user engagement by 35% by implementing interactive analytics and visualizations with Google App Engine and Gemini, while streamlining deployment pipelines to reduce system downtime by 40%.

Software Engineering Intern

Jun 2024 — Jul 2024

Etelic Inc.

San Jose, CA

- Developed an automated invoice processing system on a .NET SaaS platform using C# and Azure, reducing manual processing time by 50% and increasing dashboard accuracy by 40% with machine learning-driven insights.
- Streamlined Azure cloud deployment workflows, cutting downtime by 35% and ensuring high scalability, while enhancing user interaction and data visualization with Native JavaScript and HTML.

Software Engineering Intern - AIStore

Feb 2024 — May 2024

NVIDIA

Santa Clara, CA

- Developed and integrated performance metrics like throughput, latency, and storage operation counters for SDK, enhancing monitoring capabilities and user insights
- Implemented storage blob download API for SDK, while restructuring existing modules to improve code coverage to 90%. Utilized capacity planning while deploying AIStore production cluster
- Collaborated on integration of WebDataset format for deep learning applications into SDK to facilitate efficient reading and writing of data, optimizing dataset handling processes, particularly for PyTorch environments

Software Development Engineer

Mar 2021 — Jul 2022

Zetwerk

Bangalore, India

- Achieved a rapid 300ms response time for the Customer Dashboard through extensive data engineering optimizations and efficient real-time data stream handling and load balancing. Utilized insights from Splunk and Kibana for log analysis and performance monitoring, significantly improving system responsiveness
- Integrated Apache Kafka and Flink, along with Scala and Pyspark for real-time data processing. This integration, hosted on GCP, led to a remarkable 60% reduction in development time and leveraged Azure SQL and Cosmos DB for optimized data storage and retrieval.
- Improved code quality by 79% using SonarQube and reduced vulnerabilities using advanced scripting techniques. Streamlined the CI/CD pipeline using Azure, Docker, enabling seamless continuous integration and deployment processes for the team.