Harsh P. Bajaj

Redmond, WA harshpbajaj@yahoo.co.in
(206) 825 0909 www.linkedin.com/in/harshbajaj543 https://github.com/harsh543

Senior Software Engineer

Software Engineer with experience building scalable, secure, and intelligent enterprise applications for Global Fortune 500 technology companies, ensuring high availability and performance. Proven expertise in full-stack development, AI/ML integration, cloud infrastructure, and microservices. Passionate about creating tools that enhance workforce productivity and collaboration and mentoring junior developers in ML and design patterns.

Skills

Full-stack development | Scalable distributed systems | SaaS platforms | Web application architecture
Cloud infrastructure | AI/ML | GenAI | Microservices | RESTful APIs | CI/CD | DevOps | BFF Patterns | OOD/OOP
Python | Java | C++ | C# | JavaScript | Go | Node.js | ReactJS | AngularJS
Spring Boot | .NET | TensorFlow | Keras | Synapse | MLFlow | Apache Spark | PyTorch | MCP
Azure (Service Fabric, ML, Functions, Compute) | AWS (Sagemaker, EC2, CloudFormation) | Docker | Kubernetes | Podman | App Insights | Geneva Monitoring
Kafka | Kusto Explorer | Git | Jira | Visual Studio | JetBrains
Windows | Linux | Android | iOS
SQL | NoSQL | DynamoDB | MongoDB | CosmosDB | OAuth | SSO | Authentication & Authorization
LLMs | Agents | Web Services | Container orchestration | Design patterns | Testing frameworks

Professional Experience

Microsoft - Redmond, WA July 2021 - present
Software Engineer II (Silicon, Cloud Hardware Infrastructure)

- Designed and implemented a scalable orchestration platform using Databricks, Apache Spark, and Azure Synapse, enabling seamless data processing across global enterprise datasets. Reducing the latency by upto 2 seconds.
- Developed a machine learning pipeline using Python and Synapse ML to predict GPU failures from telemetry data, deploying models as AI agents in Azure Foundry. Predicting proactively for about 13 days ahead.
- Designed secure telemetry ingestion pipelines for GPU hardware at rack level, integrating with Linux-based APIs and identity systems (certificates + Entra ID). Enabled encrypted data flow from edge to cloud observability tools like Kusto and Grafana. Reducing data ingestion latency by 5% and availability improvement to 99.99%.
- Collaborated cross-functionally with hardware and software teams to deliver high-performance, secure, and maintainable internal tools for cloud infrastructure monitoring.
- telemetry using DSTS and calling system level APIs in C++ and C# .NET to collect telemetry from Kafka and store in Kusto using certificates and managed entra-id.

Amazon Web Services - Seattle, WA July 2019 - July 2021
Software Development Engineer (AWS Identity)

- Implemented authentication services using Java, including TOTP-based login experiences for AWS Console, improving enterprise security adoption by 200%.
- Developed internal Python-based analytics scripts to monitor runtime infrastructure and backend metrics across EC2 fleets.
- Built scalable infrastructure for monitoring application health and usage patterns using SQL databases like PostgreSQL and Oracle within CI/CD environments.
- Maintained backend Linux services for internal tools, applying debugging techniques via shell scripting and log stream analysis.
- Developed and containerized services with Docker, integrated into CI/CD pipelines using AWS CloudFormation and CodeDeploy.
- Built internal observability infrastructure using Memcached for caching role-based access tokens, reducing user session fetch times by 40%.

FireEye - Milpitas, CA May 2018 - August 2018
Software Engineering Intern (Email & Cloud Security)

- Developed a custom K-Means clustering algorithm for logline analysis, integrated with Jira, automating issue tracking and improving internal support workflows.

- Made an `alert system` using SNS and Lambda application in Python to trigger the Email anomalies with the help of CloudWatch.

Yahavi-New Delhi, India Nov 2015 – Aug 2017 *Software Developer(Web & Android Application Development)*

- Built the mass mailer campaign with success. Using AWS SES and PHP and sent over 30,000 emails in a day.
- Monitored the mail server using Cloudwatch, SNS, and AWS lambda.
- Optimized the website code for a better load time of about 50% for all the webpages majorly optimizing using Javascript.

Motor & General Sales Pvt. Ltd-Lucknow, India Aug 2014 – Nov 2015 *Software Engineer*

- Maintained company's billing applications, inventories, and hire purchase using C++.
- Built an employee payroll software in wxDev-C++ which is a GUI based and offline desktop-based application for the company.

Education

Master of Science (M.S.), Computer Science
University of Illinois, Chicago
Relevant Coursework: Cloud Computing | Machine Learning | Distributed Systems | AI Safety | Big Data | Deep Learning

Bachelor of Technology (B.Tech), Computer Science and Engineering
Vellore Institute of Technology
Relevant Coursework: OOP | Data Mining | Linear Algebra | Graph Theory | Computer Networks

Projects

- Built a chatbot using Azure AI, LLMs, and KQL to help users optimize queries in Azure Data Explorer. Integrated document embeddings and SQL frameworks to enhance internal data accessibility.

- Developed a Network Anomaly Detection System leveraging autoencoder-based deep learning and multivariate statistical techniques to identify DDoS attacks in real-time. The system combined the Koalaverse anomalyDetection R package for log vectorization and factor analysis with a Keras-based autoencoder pipeline trained on normal traffic patterns. Latent representations were extracted and classified using a lightweight neural network to distinguish anomalous behavior. The solution was containerized for reproducibility and included a modular Python codebase with TSNE-based visualizations

- Built a Real-Time Fraud Detection System for e-commerce and banking transactions using advanced ML pipelines. The solution integrated geolocation-based feature engineering, SHAP/LIME explainability, and ensemble models (Random Forest, XGBoost) to identify anomalous behavior across diverse transaction streams. Deployed via Flask and Docker, the system enabled API-based fraud scoring and real-time monitoring. The project emphasized transparency and operational readiness with a modular codebase, interactive dashboards.

- Developed a Generative AI system to predict stock market based on the data from various datasources like Finance websites, News, and Federal reserve and deployed the LLM on Bedrock. Then used Generative to predict the direction of stock market and target ticker price.

- Designed a Investment Agent on Mosaic – Agent Bricks Framework, investment-focused agent using Databricks Agent Framework (MCP). The system orchestrated data retrieval, evaluation, and trade signal generation across real-time market feeds.