

Vibhav Sharma

Linkedin | Github | vibhavsharmaha@gmail.com | Medium
+91 9380141825

WORK EXPERIENCE

Propheus by Evam Labs | Senior Data Engineer

Jan 2025 – Present

Bangalore, India

- Developed **UNet** segmentation models using conditionally masked Sentinel-1 and Sentinel-2 imagery aggregated into **GH5 geohash tiles** to derive spatial indices (e.g., greenness, waterbody coverage), contributing to **60% improvement** in house price valuation performance with **LLM-driven** explainability.
- Independently architected **Traffic Pulse**, processing 20+ CCTV feeds via **YOLO/OpenCV** for vehicle detection and graph-based anomaly analysis. Deployed on horizontally-scaled **Kubernetes** infrastructure with **Docker, Kafka**, and **CDN**.
- Addressed domain-specific challenges while developing a **Telecom Churn Prediction System** processing data from **10M users**, using ensemble models and **SHAP**-based local feature importance for explainability. Orchestrated daily **Spark** jobs using **Airflow**, with automated alerting for at-risk customers.
- Collaborated with data scientists to integrate **multimodal AI systems** and helped fine-tune models for image search optimization, improving accuracy by **30%**. Designed **distributed microservices** using **FastAPI, Kafka** for async processing, **Redis** caching for sub-millisecond response, and **Qdrant** vector search for multimodal embeddings.
- Engineered streaming data pipelines processing (approx **10K events/day**) from 20+ traffic cameras and vendor APIs (TomTom), using **Kafka** for real-time ingestion and **Apache Spark** for distributed processing with shuffle optimization, memory tuning, and data skew mitigation, storing results in **Hive** and **Delta Lake**.
- Led an Agile team of five** through full SDLC to build an end-to-end application that scraped all data (reviews, promos) of targeted retail stores and competitors, creating a dashboard with full frontend (**Next.js and Vue**) and backend (**Python, FastAPI, and MongoDB**), implementing **CI/CD pipelines** with automated testing and deployment using GitHub Actions.

Propheus by Evam Labs | Data Engineer

Nov 2023 – Dec 2024

Bangalore, India

- Designed and implemented a modular **Custom Agentic Framework** powering Alchemy's geospatial solution, featuring specialized **Retail Agents** (runtime event scrapers and footfall miners) and **Geo Agents** (runtime satellite analyzers and commute pattern miners), collectively processing data from **100+ sources**.
- Developed custom **Spark extensions** and **spatial join functions**, enhancing distributed processing efficiency bringing down the job compute time from **6 hrs to 30 min**.
- Engineered **kernel-level optimizations** for compute functions, achieving up to **60%** performance gains in data-intensive workloads.
- Created an **Experience Agent** to automate failure detection, analysis, and context extraction for **LLM** code generation.

Azira | Associate Software Engineer, R&D

Nov 2022 – Nov 2023

Bangalore, India

- Developed custom **Spark privacy join functions** capable of running on any machine via extension archives.
- Contributed to a **Clean Room framework**, enabling **cryptographic joins** for secure, privacy-compliant dataset integration.

CME Group | Software Engineer Intern

Apr 2022 – Sep 2022

Bangalore, India

- Developed workflow pipelines for business-critical tasks using **Java** and **Appian**.
- Assisted in the maintenance of **microservices** and performed deployment testing.

PROJECTS

Real-time Traffic Anomaly Detection | *OpenCV, YOLOv8, Spark, Kafka* - [Link](#)

- Built a real-time system processing 20+ CCTV feeds , used **graph-based algorithms** to identify traffic irregularities.
- Implemented intelligent **route deviation recommendations** for proactive congestion prevention.

Custom Spark Extensions | *Apache Spark, Scala, JTS, SBT* - [Link](#)

- Developed custom extensions for geospatial data processing, reducing processing time by 40% via **geohash based indexing**.

TECHNICAL SKILLS

Languages: **Adept:** Python, Scala, SQL, Java **Familiar:** Rust, C, C++

Big Data: Apache Spark, Custom Spark Extensions, Kernel Optimization, Real-time Stream Processing, ETL

ML & AI: Computer Vision, Multimodal ML, Fine-Tuning (LoRA), LLM Agents, Predictive Modeling

Geospatial: Apache Sedona, Satellite Imagery Analysis, Footfall Analysis, Graph-based Pattern Detection

Databases/Cloud: PostgreSQL, Elasticsearch, Vector DBs, Redis, Docker, Kubernetes, Kafka, AWS

Web: **Adept:** FastAPI, React, REST APIs, Spring Boot, Custom Web Scraping **Familiar:** NextJs , Typescript, Vue

EDUCATION

R V College of Engineering

B.Tech in Computer Science and Engineering - GPA 8.2/10.0

Aug 2018 – Aug 2022

Bangalore, India