VARUNKUMAR DADAJI SONAWANE

Bloomington, IN | vsonawa23@gmail.com| +1(812)929-3270| linkedin.com/in/varun-sonawane | github.com/varunsonawane

EDUCATION

Indiana University, Bloomington **Aug 2024 - May 2026**
Master of Science in Computer Science GPA: 3.8/4.0
(Relevant Courses: Engineering Cloud Computing, Applied Machine Learning, Advanced Database Concepts, Applied Algorithms)

Savitribai Phule Pune University **Aug 2019 - May 2023**
Bachelor of Engineering in Information Technology with Honours in Data Science GPA: 8.90/10.00

SKILLS

Programming & Scripting: Python, JavaScript, SQL, Bash
Databases & Cloud: PostgreSQL, MySQL, NoSQL, AWS (S3, Lambda, EC2, SageMaker, IAM), Azure, GCP, Snowflake
ML, LLMs & Analytics: Scikit-learn, TensorFlow, Pandas, NumPy, LangChain, Qwen, Prompt Engineering, ReAct, Feature Engineering
Big Data & ETL: Apache Airflow, PySpark, Spark, Hadoop, Data Pipelines, Data Modelling, Normalization, Batch Processing
BI Tools & Data Visualization: Power BI, Tableau, Excel, Git, Linux CLI, VS Code, System Design, Agile, Statistics

EXPERIENCE

Data Engineer Intern (The Commons XR, San Diego, USA) **May 2025 - Aug 2025**

- Achieved real-time, cross-system data syncing, as measured by a 100% reduction in manual handoffs and faster reporting cycles, by building data pipelines from **PostgreSQL** to both **PostgreSQL** and **BigQuery**, and implementing **CDC** using Google Cloud **Datastream**.
- Improved query maintainability and execution reliability, as measured by a 40% drop in SQL runtime errors and smoother dev handoffs, by refactoring legacy **SQL** into modular **SQLx** files and integrating them into Python-based BigQuery pipelines.
- Delivered ready-to-serve datasets for XR session analytics, as measured by improved dashboard and insight accuracy, by designing batch transformation workflows in BigQuery and developing interactive dashboards using **Plotly** and **Dash**.
- Enabled scalable **LLM**-driven activity scoring, as measured by faster prompt iteration and consistent response quality, by building a tuned **Retrieval-Augmented Generation (RAG)** model in **Vertex AI** and documenting the full prompt architecture.

Technical Lead (Unstop Igniters, India) **Mar 2022 - July 2023**

- Implemented **structured project management** for **4 hackathons** and **8+ coding competitions**, enhancing **team work** while showing **attention to detail** in **requirements gathering** across **technical challenges**.
- Coordinated **10+ member teams** for 5 industry speaker sessions and 7 technical workshops, defining **clear project scope** and **success metrics** while demonstrating **communication**, **organization**, and **problem-solving** abilities.
- Engineered systematic **technical evaluation rubrics** for multi-tier hackathon judging, implementing quantifiable metrics for **algorithmic efficiency**, code maintainability, **system architecture**, and API design patterns while delivering technical mentorship on RESTful services, **containerization**, and **cloud-native deployment** methodologies.

PROJECTS

IdeaGenie | LLM-Powered Innovation Ranking Engine [2nd Prize – Hackathon Winner] **Apr 2025**

- Architected an end-to-end idea evaluation engine using **Flask**, **Docker**, and **Qwen LLM via Ollama**, implementing the **ReAct (Reasoning + Action)** framework to simulate iterative reasoning steps for multi-criteria scoring.
- Designed a hybrid pipeline combining **vector similarity (cosine distance on sentence embeddings)** with LLM-based evaluations for idea relevance, originality, and feasibility; handled 100+ idea inputs with sub-**180ms** latency.
- Engineered a production-ready **REST API** with input schema validation, concurrency, and structured JSON response formatting with fallback logic for incomplete LLM outputs.
- Integrated advanced **prompt chaining**, dynamic context injection, and controlled generation using ReAct-style templates, improving LLM output consistency and ranking accuracy.
- **Engineered a modular backend architecture**, optimized for future containerization, and integrated seamlessly with a React frontend for real-time idea ranking and display.
- **Ranked 2nd out of 50+ teams** for AI-driven system design, LLM integration, and deployment readiness.

ShieldScraper | AWS-Based Automated Data Pipeline **Apr 2025**

- Developed a secure, fully automated web scraping pipeline leveraging Scrapy for data extraction, containerized with Docker, and deployed on **AWS ECS Fargate** for scalability and reliability.
- Implemented robust ETL processes using **AWS Glue** and **Lambda**, transforming scraped soccer data from ESPN into structured datasets stored in **AWS DynamoDB** and queried via **AWS Athena**.
- Designed real-time monitoring and alerting systems utilizing **AWS CloudWatch** and **AWS SNS**, enabling proactive management and rapid response to operational issues.
- Built interactive, insightful analytics dashboards with **AWS QuickSight**, providing stakeholders with actionable data visualizations to support strategic decisions.
- Automated daily scraping and data processing workflows using **AWS EventBridge**, significantly enhancing operational efficiency, reducing manual intervention, and ensuring seamless scalability.

ETL Pipeline for Weather Data using Airflow and Docker **Dec 2024**

- Developed an **automated ETL pipeline** to extract real-time weather data from the OpenWeather API, transform it using Python, and load it into a PostgreSQL database, processing data for over 10+ locations daily.
- Utilized **Apache Airflow** for task orchestration and **Docker** for containerized deployment, improving scalability and reducing manual intervention.
- Designed modular **DAGs** to automate data workflows, increasing data processing efficiency and enabling seamless analysis and visualization.