

Vaibhav Vinay Ranashoor

Data Platform Engineer | +1 716-295-3648 | vaibhavvinayranashoor@gmail.com | [LinkedIn](#) | NYC Metro Area

PROFESSIONAL SUMMARY

Data Platform Engineer and AWS Certified Data Engineer – Associate experienced in building data pipelines and platforms with **Python and SQL on AWS and GCP**. Skilled in **ETL/ELT, data modeling, data-quality automation, and query optimization**, with containerized deployments and **CI/CD**. Built pipelines processing **10M+ records**, services sustaining **99.9% uptime**, and API performance tuning that cut latency by **60%**.

EDUCATION

Master of Science in Engineering Science (Data Science), University at Buffalo (SUNY), GPA: 3.96/4.00 February 2026
B.Tech in I.T., V.J.T.I. (University of Mumbai) May 2021

CERTIFICATIONS

AWS Certified Data Engineer - Associate ([Credential Link](#)) Issue date: 28th July 2025
High Performance Computing Programmer - Artificial Intelligence, C-DAC Issue date: 2nd September 2024

EXPERIENCE

Volunteer Data Engineer, Saayam For All (501(c)(3) non-profit), Remote: August 2026 to Present

- Built and maintained **Python ETL pipelines** that extract raw data, apply transformations, and load it into structured storage on **Amazon S3**; wrote **SQL** queries to retrieve, filter, and aggregate data for business intelligence and reporting.
- Implemented **data-quality checks** for missing, duplicate, and inconsistent records, monitored daily pipeline runs and escalated failures, built **Power BI** summary dashboards, and contributed to **Python backend services on AWS**.

Volunteer Software Engineer (AI/Data Platforms), University at Buffalo, NY: July 2026 to Present

- Designed and maintained a **PostgreSQL data platform (SQL, Django ORM)** for an AI-assisted Engineering Mechanics learning platform, modeling six related tables for engineering assets, AI chat sessions, and student Canvas events; implemented **relational schema design**, constraints, **versioned schema migrations**, and **query performance tuning**.
- Built **Python and Django REST Framework data-processing services** for **ingestion, validation, and complex transformations** of XML and JSON data; persisted problem schemas, graph definitions, grading context, model metadata, chat histories, and student-event timelines to produce reliable, traceable datasets for AI and analytics workflows.
- Containerized the platform with Docker and deployed it to **GCP Cloud Run with PostgreSQL on Cloud SQL**; implemented **deployment automation and GitHub Actions CI/CD** for container releases, API testing, and database migrations, with structured logging and production debugging supporting reliable, repeatable deployments.

AI Intern (Part-time), Vaspian, Buffalo, NY: September 2025 to November 2025

- Built an automated **Python and pandas data-ingestion pipeline** to ingest, clean, validate, standardize, and transform **10,000+ structured and unstructured records** from REST and JSON sources; implemented **data-quality and consistency checks** that produced reliable, analysis-ready datasets for retrieval and model-evaluation workflows.
- Containerized** the **Python** service with **Docker**, deployed it on **AWS**, and implemented **unit and integration testing, health checks, structured logging, and monitoring**, sustaining **99.9% uptime** and reducing production regressions.

Student Assistant (Part-time), University at Buffalo, NY: February 2025 to August 2025

- Built the early-stage **full-stack web application (TypeScript and React frontend, Python and Django backend)** for an **Engineering Mechanics e-learning platform** with faculty in Civil, Structural and Environmental Engineering.

Trainee, Capacity Building Programme (HPC Programmer - AI), C-DAC, Bengaluru: March 2024 to July 2024

- Built **Python and Bash scientific-data pipelines** on the **PARAM Utkarsh** supercomputer to ingest and validate molecular-simulation inputs and collect run configurations, evaluation metrics, and prediction outputs for analysis.
- Ran and tuned HPC workloads with **SLURM, OpenMP, and CPU/GPU nodes**, resolving scheduler and resource-allocation issues to improve end-to-end workflow execution time by **30%**; used **Git-based version control** and wrote technical documentation.

Software Developer (Deputy Manager), Jio Platforms Limited, Mumbai: June 2021 to August 2022

- Developed full-stack features for an agritech analytics platform serving **10,000+ farmers: Python backend services and REST APIs with Django and FastAPI**, JSON-based data-integration workflows, and **TypeScript and React** user interfaces; performed application and data-access **performance tuning**, reducing API latency by **60%, from 500 ms to 200 ms**.
- Built automated **data-quality, schema-validation, and error-handling controls using Python and SQL**, ensuring consistent data across customer-facing analytics and reporting pipelines and maintaining **99.9% data reliability**.
- Expanded **unit and integration testing**, diagnosed production failures through **structured logging and observability**, and delivered reliable releases through **Git-based CI/CD**, reducing production incidents by **60%** and accelerating deployment cycles.

PROJECTS

NYC Taxi ride demand prediction: Built a serverless **Python/PySpark** pipeline with Lambda, S3, Glue, Athena, and SQL to ingest, validate, and transform **10M+ trip records** into curated datasets for forecasting and a **Power BI dashboard**.

DiaBuddy, diabetes healthcare assistant (team project): Fine-tuned **LLaMA-3 8B Instruct with LoRA** on ~4.3K curated diabetes Q&A pairs and built a **RAG pipeline (LangChain, sentence-transformer embeddings, FAISS)** over a 2,032-page clinical guideline; evaluated with BLEU, ROUGE-L, BERTScore, and BART-MNLI.

TECHNICAL SKILLS

AWS: S3, Glue, Lambda, Athena, Redshift, Kinesis

Data Engineering: PySpark, Kafka, Airflow, dbt, Snowflake, PostgreSQL, ETL/ELT, data modeling, data-quality checks

Languages & Frameworks: Python, SQL, Bash, TypeScript, pandas, Django, FastAPI, React, LangChain, FAISS

DevOps, Cloud & BI: Docker, Git, GitHub Actions (CI/CD), GCP (Cloud Run, Cloud SQL), Power BI