

SUMMARY

Senior Python Developer & Data Engineer with 5+ years of experience building scalable, cloud-native data platforms. Expert in **Python**, **FastAPI**, and **Large Language Models** (LLMs). Proven track record designing microservices, leading small teams, optimizing ETL pipelines (reducing runtime by up to 60%), and delivering actionable insights through data-driven and agentic AI solutions.

EMPLOYMENT

Lead Python Developer Virtusa (Consulting for Citi Bank) Jun'24 – Present

- Developed LLM-powered agentic workflows within Maphub using **LangGraph**, **LangChain**, and **LLMs** to dynamically generate data models, configure UI elements, and enable conversational data interactions.
- Designed and implemented microservices with REST APIs using **FastAPI** to extract, validate, transform, and load data into MongoDB and Oracle databases, supporting inputs from XLSX file, and database sources.
- Built Text2SQL benchmark framework and script to **evaluate** and **optimize** LLM performance on complex enterprise schemas using metrics like **EX** (Execution Accuracy), and **QAS** (Query Affinity Score).
- Optimized schema dump token consumption via intelligent pruning, compression, and selective retrieval, significantly reducing context length and **inference costs by 50%** while maintaining query accuracy.

Specialist Programmer Infosys May'19 – Dec'21

- Engineered a scalable streaming data pipeline on **GCP** using **Pub/Sub** and **Cloud Run** for automated data ingestion and processing, enabling **real-time store analytics** and contributing to a 15% increase in sales.
- Built and managed independently four complex **ETL batch data pipelines** using Apache PySpark, Delta Lake, Databricks, Apache Airflow, and AWS services, serving **three clients**: Levi's, Kraft Heinz, and HCSC.
- Optimized data processing performance and resource utilization by strategically implementing window functions, leading to a **30% reduction in Spark job runtime** and 15% decrease in cluster memory usage.

EDUCATION

- **Northeastern University** - Master's in Computer Science (Jan'22 – Dec'23)
 - **Teaching Assistant** - Database Management Systems; Human-Computer Interaction
 - **Research Assistant** - List Curator (Summer'22 - Fall'23)
- **Amrita School of Engineering** - Bachelor of Technology in Computer Science (Aug'15 – May'19)

PROJECTS

- **Text2SQL Benchmark** — Built an open-source framework with a composite scoring system and intent-aware column selection to improve LLM evaluation on enterprise schemas. **Python, LLMs, SQLite**
- **Wander Finds (Gemini AI Hackathon '24)** — Developed an AI app with personalized travel recommendations using Gemini and Google Maps for location-aware suggestions. **Firebase, LangChain**
- **Shark Tank Analysis (Fall '23)** — Created an interactive dashboard analyzing entrepreneur trends and valuations across industries and states. **D3.js, JavaScript, Tableau**
- **Hate Speech Detection (Summer '22)** — Built a real-time in-browser text classification tool for detecting hate, offensive, and neutral content using ML models. **XGBoost, TF-IDF, Pyodide**

SKILLS

- **Languages:** Python; JavaScript; Java; Bash; C
- **Databases:** MongoDB; MySQL; Postgres; DuckDB; SQLite
- **Certifications:** [PCAP™ – Certified Associate Python Programmer](#); Smart Analytics, ML, and AI on GCP
- **Workflow & Platforms:** Delta Lake; Apache Spark; Airflow; Databricks
- **ML/AI Tools:** LangChain, LangGraph, Large Language Models (LLMs), AWS AutoGluon, XGBoost, Pyodide
- **Visualization:** Tableau, D3.js

PUBLICATION

- Vandanapu, K. — Hadoop and Natural Language Processing Based Analysis on Kisan Call Center (KCC) Data; 2018 International Conference on Advances in Computing, Communication, and Informatics.