

AAKASH KHATU

🌐 aakashkhatu.me | 📞 585-604-9270 | ✉ ak7665@rit.edu | 🌐 linkedin.com/in/aakashkhatu/

EXPERIENCE

Data Engineer, Specialist Programmer

Infosys Limited

August 2021 – October 2022

Pune, India

- Designed and developed **ETL** pipelines on **Azure Data Factory** to extract data from SAP-ERP and CRM sources.
- Transformed the data using **PySpark** and **Azure Databricks** applying complex business logic, then loaded it into **Azure Data Warehouse**. Integrated the data into **Power BI** for Business Intelligence (BI) reporting.
- Streamlined data pipelines, reducing ETL completion time from 17 hours to 5 hours, making the process **over 3x faster**. Collaborated within an **Agile** Team, consistently delivering sprint targets.

Software Developer

Kamusi Project USA

August 2020 – July 2021

Remote

- Developed scripts in **Python** and **Cypher** for cleaning and importing data with varied formats and from various universities into a **Neo4j Graph Database**, managing a database of over 7,000 interconnected languages for translation.
- Redesigned database schema and optimized search queries to run under 15ms, making them over 65 times faster.
- Ensured smooth functioning of the **Linux** servers and automated system snapshot backups and recovery mechanisms using **Bash** Scripts.
- Transitioned Development environment from **Vagrant** to **Docker Containers**.

Deep Learning Research Intern

AjnaLens - Dimension NXG

June 2020 – July 2020

Mumbai, India

- Integrated **Tensorflow** models into Android apps using **Tensorflow Lite** for prototyping and benchmarking various **Deep Learning** Image Recognition models.
- Researched approaches to increase inference performance of **Machine Learning** models on embedded devices with **Java** and **C++** optimizations, while significantly reducing memory usage.

PROJECTS AND CERTIFICATIONS

SecureRAG : Self Hosted AI Agent for Secure Document Querying | LangChain, Python

July 2024 - Ongoing

- Developed a Completely Self Hosted **LLM** System using Llama 3.1 8B Model for Querying a Knowledge Base of Sensitive Documents utilizing **Retrieval Augmented Generation (RAG)**. Operates efficiently on a single GPU system, with accuracy comparable to OpenAI's GPT-3.5 Turbo.
- The primary goal is to enable Companies, Development Teams and Research Groups to interact with confidential documents, without relying on external third-party cloud based services and Implement a system that can be easily scaled and customized.

AWS Certified Cloud Practitioner | Certification

March, 2024

- Grasped the Fundamentals of **AWS Cloud** Services. Served as a stepping stone in translating my extensive knowledge of **Microsoft Azure Cloud** Services to their **AWS** equivalent technologies.

AWS-Powered Data Pipeline: Analyzing the Steam Dataset | Python, Pandas, AWS Services

January 2024

- Developed an ETL pipeline using **AWS Glue** to process a large Steam dataset, extracting key metrics and correlations.
- Utilized **AWS S3** for storing large dataset files and **DynamoDB** for managing cleaned and structured data, ensuring secure and scalable handling of both raw and processed data.
- Setup **AWS Lambda** Functions to query data in DynamoDB and return Data Visualizations over **API Gateway**.
- I learned to orchestrate various AWS Services and create an End-to-end Cloud Solution. Found out interesting insights from the Steam Dataset regarding the buying habits of users and created interactive visualizations with **Plotly Express**.

Grand Finalist in Smart India Hackathon | Python, NumPy, Pandas, Tensorflow, Tensorflow JS

June 2020

- Developed a Chrome extension using **TensorFlow.js** to gauge student engagement in online lectures. Followed strict user privacy guidelines by architecting a system where user's webcam data never leaves their system.
- Optimized machine learning model for real-time inference at 60 frames per second and achieved seamless performance even on old hardware without a major impact on system performance.

TECHNICAL SKILLS

Languages: Python (Proficient), C/C++ (Intermediate), SQL (Proficient), MongoDB, Neo4j - Cypher, JavaScript.

AI & ML Technologies: LangChain (Intermediate), CrewAI, ComfyUI, Chainlit, Supabase, Tensorflow, Statistics & Mathematics(Proficient), Jupyter.

Cloud Technologies: Docker, Linux, Azure Data Factory, Azure DataBricks, Azure Data Warehouse, , AWS S3, AWS EC2, Glue, Amazon DynamoDB.

Data Science: Pandas (Proficient), PySpark (Proficient), Apache Hadoop, Spark, Azure Data Factory, Azure DataBricks, Neo4j, Microsoft SQL Server.

Tools: Agile, Git, Microsoft Azure Cloud, Amazon Web Services, VS Code, Microsoft Excel, VS Code Debugger.

EDUCATION

Rochester Institute of Technology, Rochester, NY

Master of Science in Computer Science

Mumbai University

Bachelor of Technology in Computer Science

Current GPA: 4.0

Jan 2023 – May 2025

Cumulative GPA: 7.9/10

July 2017 – July 2021