AnushKrishna VenkataKrishnan

anush.venkatakrishnan@gmail.com | github.com/anushkrishnav | linkedin.com/in/anushkrishnav | Rochester, NY

EDUCATION

Rochester Institute of Technology   May 2025
Master of Science in Data Science   *Rochester, NY*
- Lead - Data Science & Engineering Club
- Backend Developer for RIT's Student Government, optimized and upgraded the petitions search system of **Pawprints**.

PSG College of Arts and Science   Aug 2019 - Aug 2022
Bachelor of Science in Computer Science with Data Analytics   *Coimbatore, TN*
- Recipient of the Student Achiever Award

EXPERIENCE

Research Assistant   Sep 2023 – Present
Rochester Institute of Technology   *Rochester, NY*
- Supporting **Prof. Mohamed Wiem Mkaouer** for a graduate-level class.
- Conducting **data mining experiments** to gain insights into developers' AI and LLM tool utilization.
- Conducting a study on how ChatGPT contributes to refactoring of Open Source projects.
- Building **Dask-based ETL Pipelines** to obtain and preprocess data to support the data mining experiments

Data Engineer   Dec 2021 – Jul 2023
Metabob   *Mountain View, CA*
- Led the automation of data collection pipelines, utilizing **Celery for job queue management** and optimizing processes for efficient data extraction.
- Developed a **Dask-based NLP pipeline** on **GKE and AKS**, achieving a significant **52% reduction in preprocessing and training time**. This parallelized various tasks for streamlined **batch training**.
- Oversaw the migration of the Dask ML pipeline and Postgres Database from Google Cloud Platform to Azure, ensuring a smooth transition and continued functionality.
- Deployed a Data Collection Pipeline on **Google Kubernetes Engine**, enabling efficient extraction and handling of over **2 million rows** daily for the data science team.

PROJECTS

MicroRides | *Python, Apache Kafka, Postgres, Redis, Streamlit*
- Engineered ETL pipelines in Python and Kafka, integrating real-time data from RTS stops, user locations, and destinations for dynamic route planning.
- Utilized Postgres for efficient storage and retrieval of transit data, optimizing the app's capabilities for thrifty transit planning.
- Implemented data streaming techniques to enhance dynamic routing, ensuring timely and cost-effective transit recommendations.
- Collaborated with the Uber API to seamlessly incorporate last-mile connections, providing users with a comprehensive and economical transit solution.

TECHNICAL SKILLS

Databases: MySQL, Cassandra, MongoDB, PostgreSQL, Redis
Big Data Frameworks: PySpark, Apache Airflow, Apache Kafka
Tools and Libraries: Python (multiprocessing, Spacy, Celery, Tensorflow, Numpy, Keras), Git, GitHub, Docker, Shell Scripting, Flask, Django, FastAPI
Cloud and Control Systems: Google Kubernetes Engine (GKE), Google Container Registry (GCR), Azure Kubernetes Service (AKS), Google Cloud, SQL, GraphQL

ACHIEVEMENTS

Winner of QCHacks 2021 (IBM Creative Challenge), hosted by Stanford x Yale.
Global Finalist (University Category) - Call for Code 2021, hosted by IBM.
Microsoft Cloud Challenge winner at Cal Hacks 8.0, hosted by UC Berkeley.
Winner of the Code What Matters award from JPMorgan, hosted by FIU.