# Dhanavikram Sekar

720-481-1463 | dhanavikram2000@gmail.com | github.com/dhanavikram | linkedin.com/in/dhanavikram

## EDUCATION

**M.S. in Data Science**, University of Colorado Boulder — May 2025
**Courses taken**: Datacenter Scale Computing, Database Systems, Machine Learning, Deep Learning

**B.E. in Electronics and Communication Engineering**, Anna University — May 2021
**Relevant Courses**: Data Structures and Algorithms, Communication Networks, Database Management Systems

## SKILLS

**Languages**: Python, Go, BASH, SQL (PostgreSQL, MySQL, T-SQL), R, Java, Javascript, HTML, CSS
**Libraries**: Pytest, Streamlit, Langchain, Scikit-Learn, PyTorch, Tensorflow, Hugging Face
**Data Engineering Tools**: Apache Spark (PySpark), Polars, Apache Airflow, Dask, Ray
**Tools**: Git, MLflow, Putty, Docker, Kubernetes, SonarQube, Postman, Power BI, Databricks, Openstack
**Operating Systems**: Unix/Linux (Systemd, Ping, Traceroute, Nmap, Administration, Shell scripting and Automation)
**Networking Concepts**: TCP/IP, UDP, DNS, DHCP, HTTP, ARP, NAT, PAT, Subnetting, Switching, VLAN
**Certifications**: AWS Certified Cloud Practitioner

## EXPERIENCE

**Research Assistant** — September 2024 – Present
*University of Colorado Boulder*

- Revamped existing **ETL data pipeline** using **polars** and vectorized logic, reducing runtime from **30 hours to 5-10 minutes**
- Optimized data storage by **75%** and improved I/O speed by **80%** by transitioning from CSV to Parquet format
- Deployed a **distributed** data platform using **Elasticsearch** and **Kibana** across a 10-node cluster

**Data Engineer Intern** — May 2024 – August 2024
*Navajo Transitional Energy Company*

- Built automated ELT pipelines, transferring **500,000+ records** from on-prem **SQL Server** to **Azure Fabric OneLake**
- Developed **10+** dashboards in **PowerBI** to convey information at a glance for decision making
- Automated web scraping of coal shipment data using Python, eliminating **5+ hours** of manual weekly tasks

**Software Engineer** — July 2021 – July 2023
*Tata Consultancy Services*

- Developed and tested feature engineering utilities for a proprietary data processing and machine learning framework using Apache Spark (PySpark) and Python (pandas and sklearn)
- Developed XGBoost, Prophet and ARIMA models as custom spark UDFs using **Pytorch** and statsmodel
- Integrated the custom UDFs as custom **SparkML Estimators** and **Transformers** for demand forecasting
- Implemented experiment tracking and model drift detection with **MLflow**, **Grafana**, and **Prometheus**
- Built **CI/CD** pipelines with **Jenkins**, integrating automated testing (unit and integration tests) using **pytest**
- Deployed the entire framework as a **Django RESTful API** and tested its functionality using **Postman**

## PROJECTS

**Micro-Demucs: Music Separation as a Service** [GitHub]

- Developed a **GCP**-based **microservices** backend for music separation: REST server as frontend API, **Redis** for queuing, **MinIO** for object storage, and Meta's demucs model for track separation
- Containerized components using **Docker** and deployed to separate **Kubernetes** clusters in GCP
- Implemented an **Nginx** ingress controller for public internet access within the GCP environment

**Multi Document LLM Chatbot** [GitHub]

- Built a Mistral-7B based **RAG system** to answer queries from uploaded documents
- Implemented MiniLM-L6-v2 embeddings from **Huggingface** to convert text to vector embeddings
- Utilized FAISS-CPU for local vector storage and retrieval of document embeddings