My Name

phone |@gmail.com |location | linkedin.com/in/

Experience

MOST RECENT JOB
Remote
Software Engineer Aug 2022 - January 2024

- Spearheaded the integration of MPNET vector generation into the division's top patent product improving prediction results by 55%. AWS SQS notifications trigger the processing of incoming data in Batch jobs that run a custom Docker image. The system then stores the results in Postgres and Snowflake, allowing front-end users to access and utilize this data. Infrastructure deployment is handled by Terraform and the use of text-field hashing eliminates redundant processing, optimizing compute resource spending. Thoroughly unit-tested using Pytest
- Designed and implemented a data parsing pipeline for new incoming data, processing 2.5TB of XML data from S3. A recursive SAX parsing algorithm was used in conjunction with Spark for Delta Lake writing hosted on Databricks with an S3 backend with data written to Snowflake
- Developed a raw data ingestion and processing pipeline for a new business direction, initialized with over 100GB XML data and having 0.5-1GB of updates daily stored in Databricks Delta Lake and Snowflake using Python and Spark and deployed via Terraform
- Increased accessibility of key machine learning predictions by writing an end-to-end pipeline on Databricks, leveraging a fine-tuned model on 15M+ records weekly and hosting the results in PostgreSQL, enabling consultants to use this data for client deals
- Refactored and improved upon a critical data pipeline using byte-seeking techniques to enable Spark to parse and process data on Databricks efficiently, increasing processing speed by over 12x, removing user downtime, and eliminating data staleness
- Improved consultant's turnaround time by 70% via automating processing using a Python pipeline allowing users to upload required files and quickly transform raw data into client-ready Excel reports using Pandas for data parsing and calculations

OTHER JOB
Location
Research Assistant Jul 2018 - Aug 2020

- Pioneered a program for juniors and seniors who lacked prior research experience to volunteer for one semester and develop transferable wet-lab skills to pursue other endeavors – mentored seven undergraduates
- Developed microRNA RT-PCR protocols specific to our equipment for use across all labs
- Increased data-logging efficiency by 40% and improved the safety of patients' personal information by converting the record-keeping system to a MySQL database stored with automated data input using Python from initial patient signup
- Analyzed medium-sized (10,000+ data points) experimental datasets using Python, MS Excel, and Tableau with the accepted normalization and statistical measures for advanced analysis and cross-validation with other labs

Projects

Reddit Viral Video Generator

- Used the Reddit API to automatically fetch the text of top posts in a particular subreddit, which is sent to Tiktok's Text-to-speech API to return the text read back in an automated voice, processed through a locally running Whisper AI model- transcribing the audio with timestamps, which are then parsed and overlaid on a snippet of background gameplay with TTS audio using PyDub and MoviePy and saved in 120 second intervals ready to be uploaded to TikTok and Youtube Shorts

Twitch Data Visualizer

- Utilized Celery Beat with Redis to asynchronously fetch livestreaming data from the Twitch API with validated OAuth key, organized using Pandas and hosted on a Django backend; styled with Tailwind CSS and containerized with Docker

Substack Clone – Blogging Website

- Engineered a Substack-style blogging platform using a Django and Bootstrap CSS styled-frontend. UI includes article genre filters on the homepage, dynamic text input fields with jQuery and ckeditor, and real-time username validation using HTMX

TradingView Public Scripts

- Created 7+ unique stock market and cryptocurrency trading indicators, focusing on statistical anomalies in price volatility, and integrating metrics like moving averages, volume, and VWAP. Saved by 750+ users on TradingView

Education

SCHOOL - Bachelor of Science

May 2018

- HONORS PROGRAM

Skills & Technologies

Python, JavaScript, Django, Flask, NodeJS, ReactJS, REST API, PyTest, Pandas, HTML, CSS, Tailwind, Docker, Git, Linux, Azure, Webpack, Docker, CICD, AWS, Batch, S3, SQS, RDS, PostgreSQL, Terraform, Spark, Databricks, ETL, Snowflake, Pytorch, Databricks