CHASE GOLDFELD

DATA ENGINEER

Experience

Lambda School — Remote
Associate Product Lead — Nov. 2020 to Current

- Founder, architecture designer, and lead engineer of COAG
- Codebase contributor and liaison between stakeholders and product developers to ensure timely product releases and MVP based on stakeholder needs
- Implemented data deduplication for Human Rights First via:
  - Stream processing by cross-referencing time stamps and locations
  - Batch processing by locality-sensitive hashing
  - Resulting in mitigated data expenses and an improved model accuracy by 26%
- Managed and contributed to the production of cross-functional applications, COAG, Human Rights First, Asylum, Story Squad, Village Book Builders, and Microfund, including product road maps, user stories, release calendar, and deployment plan for product migration to cloud via AWS
- Mentored and guided team leads and students through the Labs program to support the learning and professional growth of Lambda School students

Lambda School — Remote
Data Science Technical Project Lead — Sept. 2020 to Nov. 2020

- Codebase contributor and team lead that guided cross-functional teams to engineer products and features based on stakeholders' requirements
- Engineered Google's Tesseract OCR model for Asylum which allows users to upload PDF's into the model, predicts how a judge might rule on a specific asylum case, and identifies specific elements of an asylum case that will most impact a favorable or unfavorable ruling
- Led and communicated with teams consisting of up to ten students through a cross-functional project via Agile, Scrum, workflow environment
- Hosted daily stand-ups for students on the project and weekly 1:1's
- Improved students problem-solving skills by working through live code challenges

Projects

COAG — June 2021 to Current

COAG is an application for users who are searching for career opportunities and growth in the worlds fastest growing companies. Initially, this process would cost job seekers an exorbitant amount of time, up to 5-10 mins for 1 result. Now, COAG is the one-stop-shop for job seekers who are searching for careers with the world's fast growing companies. In a matter of seconds, COAG ingests data from multiple sources that identifies the fast growing companies in the world and then pinpoints jobs based on the user's criteria (resume, job title, experience, industry, company maturity, number of employees, etc) and displays results on COAG's interactive dashboard.

Tech Sack: Python | AWS | Cassandra | PostgreSQL | Docker | FastAPI | Airflow | Github Actions

- Constructed a data ingestion pipeline from multiple sources into a raw database
- Engineered a data pipeline that extracts data from the raw data base to the staging database where data is then transformed via mapping, deduplication, QA, and logs
- Built a custom interactive dashboard to present transformed data based on the user's criteria and outputting their results
- Designed and implemented COAG's ML model
  - User can upload their resume into COAG's ML model and then COAG will search for jobs based on user's resume
  - Output results based on COAG's text similarity algorithm matching feature

Data Engineering — Apr. 2021 to May 2021

A collection of Data Engineering projects related to data modeling, data pipelines, data lakes, infrastructure setup on the cloud, containerization, data warehousing, and automation with CI/CD.

Tech Stack: Python | PostgreSQL | AWS | Airflow | Cassandra | Docker | PySpark | CI/CD

- Engineered AWS Infrastructure as Code via launching EMR clusters from the CLI
- CI/CD pipeline with AWS MWAA
- Implemented ELT & ETL pipelines with Apache Airflow
- Containerization with Docker to standardize executable components that combine application source code with required libraries and dependencies to run the project in any environment
- Created data warehouse on the cloud via AWS Redshift
- Data Modeling with Apache Cassandra and PostgreSQL
- URL to repo: https://github.com/AuFeld/Data_Engineering_Projects

Human Rights First — Jan. 2021 to Mar. 2021

Human Rights First is an independent advocacy and action organization that challenges America to live up to its ideals. Engineered their data pipeline and an interactive 12-month timeline tracking police use of force in the United States.

Tech Stack: AWS Elastic Beanstalk | AWS RDS | PostgresSQL | Python

- Served as an Associate Product Lead that managed two cross-functional teams consisting of two project leads, eight data scientists, five back end developers, and ten front end developers.
- Improved the accuracy of tracking police use of force by 26% via implementing data deduplication, which reduced data duplication via locality-sensitive hashing.
- Constructed and implemented key features: a new form for reporting police use of force incidents | implemented tag system for cataloging incidents reported on Twitter and Reddit | categorized tags based on the National Institute of Justice's Use of Force Continuum
- Successfully constructed MVP on-time for stakeholder, Welton Chang, CTO of Human Rights First
- URL to repos: [DS] https://github.com/Lambda-School-Labs/human-rights-first-ds-f | [FE] https://github.com/Lambda-School-Labs/human-rights-first-fe-f

Contact

goldfeld.chase@gmail.com
aufeld.github.io/
+1 (561) 926-3903
Delray Beach, FL
chase-goldfeld/
AuFeld/

Education

Lambda School · Data Science — Aug. 2020

Lambda School is a 9+ month computer science & software engineering program that provides an immersive hands-on curriculum focused on Data Science.

- Served as a data scientist on a ready to deploy project, working alongside UX designers, web developers, and mobile developers to bring the project to fruition
- Completed a deep dive into Data Engineering, working with databases, productization, and big data
- Developed a solid foundation of descriptive and predictive statistics, including: linear algebra, linear regression, hypothesis testing, storytelling with data, and more
- Gained hands-on experience engaging with machine learning, being able to understand unsupervised learning, natural language, and neural networks

Skills

DATA ENGINEERING

SQL: MySQL, PostgreSQL, & SQLAlchemy
NoSQL: MongoDB
Frameworks: FastAPI, Flask, & Heroku
Containerization with Docker
Data Modeling with Cassandra
Python Libraries: Pandas & Numpy
ELT & ETL Pipelines with Airflow
CI/CD with Github Actions
Big Data Processing with PySpark

CLOUD SERVICES WITH AWS

Data Lakes with S3
Data Warehousing with Redshift
Clustering with Elastic MapReduce
ML Deployment with Elastic Beanstalk
Scaling with EC2
Elasticsearch
Automation with MWAA

COMPUTER PROGRAMMING

Languages: Python, SQL, & Scala
Jupyter Notebooks
OS: Linux, Mac, & Windows
Git
Bash

PROBLEM SOLVING WITH COMPUTER SCIENCE

Data Structures
Algorithms
Dynamic Programming

MACHINE LEARNING

Regression
Natural Language Processing
Neural Networks