

CHASE GOLDFELD

DATA ENGINEER

Experience

Lambda School

Associate Product Lead

Remote

Nov. 2020 to Apr. 2021

- 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, 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

Data Science Technical Project Lead

Remote

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

Data Engineering

Apr. 2021 to Current

A collection of Data Engineering projects related to data modeling, data pipelines, data lakes, infrastructure setup on the cloud, containerization, data warehousing, and a dashboard to track and monitor data pipeline KPIs with alerts

Tech Stack: Python | PostgreSQL | AWS | Airflow | Cassandra | Docker | Grafana | Prometheus | PySpark

- Engineered AWS Infrastructure as Code via launching EMR clusters from the CLI
- Constructed a data pipeline analytics dashboard with Grafana to monitor KPIs and Prometheus for event alerting
- 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 | PostgreSQL | 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>

Groa

July 2020 to Aug. 2020

Groa uses machine learning to power an innovative recommendation engine with a feedback loop that generates tailored movie suggestions to users based on their unique taste.

Tech Stack: Word2Vec | AWS | PostgreSQL | Google Analytics | FastAPI

- Engineered the Google Analytics API to collect, measure, and report key metrics to stakeholders with data visualizations
- Built the Groa API using the FastAPI web framework and serves recommendations using a Word2Vec Model
- Implemented the Groa database with PostgreSQL on the AWS Relational Database Service
- Utilized ElastiCache (running Redis) for caching and Elasticsearch service from AWS for searching
- Successfully improved Groa: 66% Decreased Bounce Rate | Improved Model Response Time of 300% | Increased Average Session Duration Time of 665%
- Architecture Diagram: <https://www.notion.so/Architecture-Details-a8ef24c79c554a0fbf6bfab4163e588c>
- URL: <https://www.groa.us>
- URL to repo: <https://github.com/Lambda-School-Labs/Groa-ds>

Contact

✉ goldfeld.chase@gmail.com

🌐 aufeld.github.io/

📞 +1 (561) 926-3903

📍 Delray Beach, FL
in 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

Pipeline Monitoring with Grafana & Prometheus

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

COMPUTER PROGRAMMING

Languages: Python, SQL, & Scala

Jupyter

OS: Linux, Mac, & Windows

Git

Bash

PROBLEM SOLVING WITH COMPUTER SCIENCE

Data Structures

Algorithms

Dynamic Programming

MACHINE LEARNING

Regression

Natural Language Processing

Neural Networks