

David Freitag

New York, New York, US | +1 (262) 271-5365 | dkfreitag@gmail.com
github.com/dkfreitag | linkedin.com/in/davidkfreitag | davidfreitag.com

PROFESSIONAL SUMMARY

Senior Data Engineer skilled in building mission-critical, reliable, efficient, high-volume data pipelines. History of effective project leadership, stakeholder management, and mentorship of junior engineers. Former high school teacher.

WORK EXPERIENCE

Senior Data Engineer, American Family Insurance

Feb 2023 – Present

- Infrastructure Lead on a high-priority AI pilot initiative to improve underwriting risk review accuracy and speed. Scaled the MVP infrastructure to additional operating companies, lines of business, and sources of structured and unstructured data. Tech: Python/SQL, GitLab CI/CD, Terraform, GCP Pub/Sub, Docker/Cloud Run, BigQuery, Pydantic AI, LangGraph, Kubernetes.
- Architected and built multiple ETL pipelines to refresh terabyte-size datasets on demand for data science modeling. These datasets powered the models used to re-price AmFam's Auto and Property insurance products in all 50 states. Tech: BigQuery SQL, Airflow, Docker, Kubernetes.
- Improved, automated, and extended a vendor file transfer process for preparing and transferring data securely. Scaled the pipeline to prep and encrypt TB of data in parallel, improving runtimes from days to minutes. Tech: Google Cloud Storage, GCP Secret Manager, GNU Privacy Guard (PGP encryption), Google Batch.

Data Engineer, American Family Insurance

Aug 2021 – Feb 2023

- Built ETL pipelines to refresh monthly/quarterly datasets for data science modeling and product analytics. Tech: BigQuery SQL, Airflow, Docker, Kubernetes.

Data Science Intern, Pitney Bowes

Mar 2021 – Jul 2021

- Trained, tested, and tuned gradient-boosted machine learning models (XGBoost, LightGBM) to improve the accuracy of forecasting shipping container journey times within the Pitney Bowes network.

SKILLS

Languages: Python, SQL

Technologies: GCP, BigQuery, Pub/Sub, Google Batch, Docker, Pydantic AI, LangGraph, Kubernetes, Airflow, GitLab CI/CD, GitHub Actions, AWS Lambda, Glue, Kinesis, Athena, S3, Grafana, Terraform

EDUCATION

Master of Science, Information Systems, Data Analytics, Baruch College

May 2021

- GPA: 4.0
- Honors: Beta Gamma Sigma Honor Society, Beta Alpha Psi Honor Society

PROJECTS

Build Your First Serverless Data Engineering Project (maven.com/david-freitag)

May 2024 - Apr 2025

- Intro level course on serverless data engineering in AWS that I designed, marketed, and taught. Tech: REST API's, AWS Lambda, S3, Athena, Kinesis, Glue, Grafana.
- 4.8 stars, 100+ paid students

Citi Bike Data Platform (github.com/dkfreitag/citibike_data_platform)

Ongoing

- End-to-end data pipeline to ingest data from various Citibike API endpoints, do streaming and batch data processing, and make the data available in a serving layer for both analytics and data visualization. Tech (so far): Kafka, Docker, EC2, Lambda, Glue, Athena, S3, Spark, GitHub Actions, Terraform.