

Ha Khanh Nguyen

hknguyen.org [✉ ha@hknguyen.org](mailto:ha@hknguyen.org) [in haknguyen](https://www.linkedin.com/in/haknguyen)

Relevant Skills

- **Tools and Systems:**
 - Git, GitHub, GitLab CI/CD, Docker, Kubernetes, Bash, Terraform
 - Apache Airflow, GCP BigQuery, GCP Compute Engine, AWS S3, AWS EC2, AWS Athena, Snowflake, Apache Kafka
- **Languages (libraries, frameworks):**
 - Python (NumPy, pandas, scikit-learn, SciPy, seaborn, plotly, Dash, pyspark, pytest)
 - R (tidyverse, ggplot2, data.table, shiny), SQL (BigQuery, PostgreSQL), Java, C
- **Soft Skills:**
 - Agile workflow, non-technical stakeholder communication, public speaking, mentoring

Experience

Data Engineer II

Jan 2023 - Present

American Family Insurance

- Led data team's recover efforts as acting manager during major system outage. Prioritized pipelines based on customer needs and led ad-hoc data quality checks. Ensured business-essential dashboards and data were available, accurate, and met SLAs.
- Hosted training sessions for data engineers on deploying and maintaining infrastructure on GCP using GitLab and Terraform. Specific topics included GitLab CI/CD, Terraform, and GCP Cloud Composer (Airflow) environment deployment.
- Built and maintained pipelines for various BigQuery datasets to support analysts' reporting and dashboard needs, which allowed analysts to remove legacy dependencies and more efficiently transition to the cloud via GCP. Decreased time to deliver analysts' monthly reports from hours to minutes, improved data security, and increased data quality.

Data Engineer I

Jan 2022 - Jan 2023

American Family Insurance

- Led a legacy database retirement project which eliminated \$34,000 in yearly licensing fees. Deployed necessary infrastructure, built data pipeline, and developed the data validation process. Complete migration of business-essential data to GCP allowed teams across the organization easy access to data without compromising security. Data ingestion time reduced from one day to less than one hour.
- Architected and maintained infrastructure for forecasting in GCP. Deployed RStudio Server using Docker and GCP Compute Engine to support data analysts' modeling needs. Trained analysts to use GitLab, BigQuery, and RStudio Server. The provided infrastructure and training transitioned analysts from on-prem resources to cloud-based virtual machines and tooling, increasing data security and decreasing time-to-market of generated forecasts.
- Deployed, supported, and maintained an R Shiny web application on AWS EC2 instances to support planning and forecasting.
- Developed required infrastructure to establish a daily data pipeline from AWS to GCP using AWS CloudFormation and Terraform. Wrote documentation that has become the trusted source for engineers across the organization.
- Architected and maintained engineering infrastructure such as GCP Cloud Composer environments, Compute Engines, GitLab repositories, GitLab CI/CD pipelines, and GCP service accounts to support the data team's transition to the cloud.
- Nominated for the "All 4 OneAmFam Award" in recognition of outstanding achievement and performance.

Data Science Instructor

Oct 2019 - Dec 2021

Department of Statistics - University of Illinois Urbana-Champaign

- Duties included teaching three courses per semester and supervising graduate teaching assistants. Taught courses in mathematical statistics, statistical computing, and programming methods in both Python and R.
- Designed and developed the first Python programming course in the department of Statistics at Illinois. Introduced students to Git, GitHub, Python data types, data structures, and libraries including NumPy, pandas, scikit-learn, and Dash.
- Utilized Python, Docker, GitHub, and the open-source PrairieLearn framework to develop scalable and automated assignment grading in a wide range of topics including mathematical statistics and Python programming.
- Received "Teachers Ranked as Excellent" honors four times for several courses across multiple semesters.

Education

Master of Applied Statistics

2019

The Ohio State University

- Received the Distinguished University Fellowship.

BS Computer Science

2017

San Jose State University

- Minor: Mathematics. Received the Outstanding Graduating Senior Award and Summa Cum Laude Latin Honor.