Hoang Nguyen

Data Engineer
Philadelphia, PA (open to remote) | (248) 971-5622 | hoangmnguyen.work@gmail.com
linkedin.com/in/hoangnguyen29497 | github.com/HoangMNguyen | www.hoangmnguyen.com

PROFESSIONAL EXPERIENCE

Penn Medicine, Center for Cellular Immunotherapies, Philadelphia, PA

Data Engineer Aug 2021 – Present

- Build ETL and reporting pipelines in Python (pandas) and SQL that turn clinical data-capture exports into standardized safety, enrollment, and study-status reports, cutting preparation time 90%.
- Developed Python desktop suite (PySide6, pandas), including a no-code report builder on a GUI-free core API; built serverless usage telemetry (AWS SAM, Lambda, DynamoDB) to track adoption.
- Design and maintain study databases and query workflows on Veeva Vault and PostgreSQL across dev/test/prod.
- Built data-cleaning and reconciliation tools (Python, VBA) for Veeva Vault CDMS, REDCap, and Velos exports, cutting data-cleaning time 60%.
- Owned database operations and the department tech stack, including access controls, auditing, backup/restore, upgrades, troubleshooting, documentation, and team training.

Data Coordinator (Co-op) Apr 2019 – Sep 2019

- Built a VBA and Python data-processing program for departmental reporting, saving an estimated 1,000+ working hours per year.
- Set up a shared data-processing system with automatic remote updates across workstations and new data-review protocols

Thomas Jefferson University Hospital, Philadelphia, PA

Research Assistant (Co-op) Apr 2018 – Sep 2018

- Built SQL databases for insomnia and traumatic brain injury research studies; maintained IRB, FDA, and sponsor documentation.

PROJECTS

PricePulse — github.com/HoangMNguyen/pricepulse | pricepulse.hoangmnguyen.com

- Built and operate a serverless price-history pipeline (Python, PostgreSQL, AWS Lambda/S3/EventBridge, Terraform) that ingests ~3.7k IKEA and UNIQLO products daily; ~220k price observations stored.
- Designed idempotent S3-to-PostgreSQL loads into monthly-partitioned tables with a 90-day materialized price baseline; pytest tests and GitHub Actions CI/CD with OIDC-authenticated Terraform deploys.

TECHNICAL SKILLS

Languages: Python (pandas, FastAPI), SQL (Oracle, SQL Server, PostgreSQL)
Cloud & DevOps: AWS (Lambda, S3, EventBridge, API Gateway, CloudFront, SES, CloudWatch), Terraform, Docker, GitHub Actions, Git, pytest
Data: Data modeling, table partitioning, data-quality checks
Clinical data systems: Veeva Vault CDMS/EDC, REDCap, Velos CTMS

EDUCATION

Georgia Institute of Technology, Atlanta, GA — M.S. Computer Science Expected Dec 2026
Coursework: Database System Implementation, Intro to Graduate Algorithms, High Performance Computer Architecture, Software Development Process, Knowledge-Based AI, Human-Computer Interaction, Software Analysis, Introduction to Operating Systems, Computer Networks, Machine Learning for Trading

Drexel University, Philadelphia, PA — B.S. Biomedical Engineering; Minors: Data Science, Economics 2021