Fei Huang

Amsterdam, Netherlands | +31 645 038 614 | huangf06@gmail.com

linkedin.com/in/huangf06 | github.com/huangf06 | feithink.substack.com

Data Engineer with expertise in building scalable data platforms and ETL pipelines. Skilled in Python, PySpark, Databricks, and cloud technologies. Hands-on experience designing data infrastructure and optimizing pipeline performance. M.Sc. in AI from VU Amsterdam. Seeking to contribute to Doghou...

EDUCATION

Vrije Universiteit Amsterdam, The Netherlands *Amsterdam*
M.Sc. in Artificial Intelligence (GPA: 8.2/10) *Sep 2023 – Aug 2025*
Thesis: Uncertainty Quantification in Deep Reinforcement Learning under Noisy Environments

Tsinghua University, China *Beijing*
B.Eng. in Industrial Engineering *Sep 2006 – Jul 2010*

Certification: *Databricks Certified Data Engineer Professional (Feb 2026)*

PROFESSIONAL EXPERIENCE

GLP Technology *Shanghai, China*
Data Engineer & Team Lead *Jul 2017 – Aug 2019*

- Engineered automated data pipelines using Python and SQL; integrated payment gateway APIs and coordinated collection operations.
- Joined as founding data team member at fintech startup; built credit scoring engine from scratch including data pipelines, feature engineering, and model deployment using logistic regression and scorecard methodology.
- Led PySpark workflow implementation to process large-scale datasets, reducing processing time and enabling daily risk reporting.

Baiquan Investment *Beijing, China*
Quantitative Developer *Jul 2015 – Jun 2017*

- Built automated ETL pipeline to ingest daily market data from vendors with data quality validation supporting downstream trading operations.
- Engineered high-performance factor computation engine using vectorized Pandas/NumPy operations; reduced calculation time for technical indicators across 3000+ stocks.
- Developed the underlying event-driven backtesting infrastructure that supported strategy simulation and performance reporting.

  Career Note: 2019–2023 included independent investing, language learning (English, German), and graduate preparation.

PROJECTS

Financial Data Lakehouse (Databricks/Spark) *Oct 2025 – Present*
- Architected resilient lakehouse using Auto Loader and Structured Streaming; pipelines handle schema evolution automatically for continuous crypto market data ingestion.
- Engineered data quality framework via fault injection testing; validated quarantine pattern to isolate malformed records without disrupting production.

TECHNICAL SKILLS

Languages: Python (pandas, NumPy, PyTorch), SQL, Bash

Data Infrastructure: Databricks, PySpark, Delta Lake, Hive, Apache Airflow, Docker

Cloud & Tools: AWS, GCP, Azure, Git, Linux, CI/CD

Certifications: Databricks Certified Data Engineer Professional

INTERESTS

Interests: Philosophy (Kant, existentialism), Dostoevsky, powerlifting, analytical writing

Blog: huangf06.github.io/FeiThink/en/