Jun **Hu**

Software Engineer · Research Scientist

92-02 50th Avenue, Elmhurst, NY

(646) 271-1419 | jun.hu@columbia.edu | www.hu-jun.com | vibrioh | hu-jun | Jun Hu

Lawful Permanent Resident (Alien of Extraordinary Ability)

Summary

In the AI startup company, my work involves both product development and onsite project consulting. I have implemented the application module of the company's graph database and contributed to the delivery of two AI solution projects at a total price of $11 million. I also collaborated with the Harvard research lab on a cancer classification study.

Skills

Programming	JAVA, Python, R, Shell
System	Git, Docker, Kafka, Redis, High Availability, Asynchronous Operation, Cron, Logstash, AWS, Google Cloud
Database	Graph Database, ElasticSearch/Kibana, PostgreSOL, DynamoDB, Active Directory
Web	RESTful API, HAProxy, Microservise/Serverless, Flask, JavaScript, HTML, SASS
Documenting	Sphinx, LaTeX, FreeMarker, Markdown, Jira

Work Experience

Graphen, Inc — *New York, NY*

Software Development Engineer & Research Scientist — *Jan 2018 –– Current*

- **Python Application Module:** Designed and implemented a module from scratch to facilitate users calling the graph database functions. Co-authored RESTful layer to call C++ code via SWIG. Finished and maintained over 8000 lines code and generated over 100 pages standard document through Sphinx.
- **DevOps:** Deployed a non-performing loan prediction AI system to the bank client. Dockerized repositories, integrated servers, organized pipelines, automated daily batch processing, and created high availability monitoring, switching and recovery system. Worked closely with the clients to translate business requirements to technical details. Drafted SIT/UAT test scripts and documentation.
- **Data Engineering:** Explored on the bank client's data, and designed schema on production. Understood the data shape and risk patterns gained from the bank and built a data generator that can automatically export complicated, meaningful and realistic data for offsite product developing.
- **Java Back End:** Supported a dynamic graph UI. Implemented .doc/.xls report generating system with FreeMarker. Created a user, ip, department auto updating system from Active Directory to Redis weekly.
- **Machine Learning:** Researched on advanced persistent threat and the bank client's McAfee SIEM, Forcepoint and MS Exchange logs. Proposed features for the behavior-based AI solution. Delivered a B-cell receptor based cancer classification solution to Harvard X. Shirley Liu Lab, and introduced the natural language processing concept of N-grams into protein fragments. Used neural network model and GPU acceleration.

Icahn School of Medicine at Mount Sinai — *New York, NY*

Postdoctoral Fellow — *Apr 2012 - Dec 2016*

- Built a Bio-Linux server and established a pipeline of large scale sequencing analysis in the laboratory.
- Created the automatic reporting script on analyzing temporal spatial metrics data in python.
- Analyzed and visualized on experimental data and and published 19 research articles..
- Oversaw research progress of the laboratory graduate students and lead 8 research projects.

Education

Columbia University — *New York, NY*

M.S. in Computer Science — *Jan 2017 - May 2018*

- Courses Highlighted: Advanced Database Systems, Advanced Big Data Analytics, Machine Learning, Artificial Intelligence, Analysis Of Algorithms, Computational Graphics, Microservice/Cloud Apps & Natural Language Processing

Tsinghua University — *Beijing, China*

Ph.D. in Biology — *Sep 2006 - Jan 2012*

- Courses Highlighted: Bioinformatics & Biostatistics

Wuhan University — *Wuhan, China*

B.S. in Biology — *Sep 2002 - Jul 2006*

- Courses Highlighted: Fundamentals of Operating System & C Language Introduction