Jun Hu

hu-jun.com | jun.hu@columbia.edu | (646) 271-1419

EDUCATION

COLUMBIA UNIVERSITY
M.S. in Computer Science
Jan 2017 - May 2018
New York, NY

TSINGHUA UNIVERSITY
Ph.D. in Biomedicine
Sep 2006 - Jan 2012
Beijing, China

WUHAN UNIVERSITY
B.S. in Biomedicine
Sep 2002 - Jul 2006
Wuhan, China

STATUS

Permanent Resident
Relocating Acceptable

COURSEWORK

Databases
Advanced Database Systems
Big Data Analytics
Advanced Big Data Analytics
Machine Learning
Artificial Intelligence
Analysis Of Algorithms
Computational Graphics
Microservice/Cloud Apps
Natural Language Processing

PUBLICATIONS

Research Papers: 49
Citations: $>$ 700 times
Full list: bit.ly/2xVZYic

SKILLS

Java • Python • Shell • R
Docker • Elasticsearch • Kafka
Graph Database • SQL • Hive
HTML • CSS • Sass • Javascript
TensorFlow • Scikit-learn
Git • LaTeX

LINKS

Portfolio: hu-jun.com
Linkedin: /jun-hu-664639a2
Scholar: bit.ly/2xVZYic
Github: /vibrioh
Facebook: /hjun1
Twitter: /vibrioh1

WORK EXPERIENCE

GRAPHEN, INC — New York, NY
Principal Research Scientist, Big Data — Sep 2018 — Current
Research Scientist — Jan 2018 – Sep 2018
Delivered 1 FinTech project ($10 M) as a DevOps engineer on-site.
Delivering 1 Cybersecurity project ($0.8 M) as a data scientist.
Delivered 1 HealthCare project (non-profit with Harvard).
Implemented Client Module of the Graph Database.
Implemented Integration and Automation of Ardi Platform
Accomplished Business Logic and Educated Customers.

ICAHN SCHOOL OF MEDICINE AT MOUNT SINAI — New York, NY
Postdoctoral Fellow — Apr 2012 - Dec 2016
Automated Optical-Mapping Data Transformation.
Developed Pipeline of Large-Scale Sequencing Analysis.

PROJECTS

Learn More: hu-jun.com/projects.html

ICBC NON-PERFORMING LOANS PREDICTION
Industrial and Commercial Bank of China (ICBC), the largest bank in the world, requests IBM an AI solutions ($10 M) to predict non-performing loans, simulate loan default cascades by user modifying relations/properties, find hidden relationships and display dynamic network evaluation. Graphen, inc provides high-performance distributed graph database solutions ($2 M).
$\Longrightarrow$ I am in charge of the system integration (Shell/Python scripts), the REST/Client and Dynamic Graph backend (Java/Spring). The containers of Grest/Gasync (Flask/uWSGI/Nginx/Ubuntu, Docker) are deployed with clustered Kafka and running on openSUSE servers. Upstream ETL processes data from ICBC Data Warehouse daily, triggering High Availability Module, which is monitoring server health status, switching between available servers and executing requests. After loading through Async/Kafka, the predefined analytics scripts (business logic) compute all the features needed for UI and ML module, such as guarantee cycles, closeness, betweenness, page rank and so on. The results are both stored locally in graph database and ported to Oracle Database.

BOC ADAPTIVE CYBERSECURITY MONITORING
Bank of China New York Branch Smart-i ($0.8 M). The system shall gather all relevant and available information about users, devices, applications, and network events, detect anomalies at various levels, and provide an aggregate risk assessment to predict cybersecurity risks.
$\Longrightarrow$ I am in charge of feature selection and graph database.