

Yu-Wen (Sonic) Lai

ysl273@nyu.edu | (929) 401-8800 | New York, NY | [linkedin.com/in/sonicywlai](https://www.linkedin.com/in/sonicywlai) | github.com/hsnusonic

PROFILE

Innovative software development engineer with extensive exposure to distributed systems and back-end development. Background in physics and information systems. Passionate about big data, software development, and distributed programming. Proficient in problem solving and collaborating with cross-disciplinary people.

EDUCATION

New York University

Master of Science in Information Systems

New York, NY

May 2021 (expected)

National Taiwan University

Master of Science in Physics, CS GPA: 3.86/4.3

Taipei, TW

Jun 2016

Bachelor of Science in Physics

Jun 2012

Relevant Coursework: Fundamental Object Oriented Programming, Data Structures and Algorithms, Algorithms Design and Analysis, GPU Programming, Machine Learning, Systems Programming, Web Retrieval and Mining, Computer Architecture

WORK EXPERIENCE

Acer Inc.

Software Engineer, Value Lab

New Taipei, TW

Sep 2016 - Jul 2019

- **Cancer Immunotherapy ► Distributed Computing | Python | Bash**
 - Organized the bare metal cluster by Ansible and Sun Grid Engine queuing system, lessening 66% effort to manage the cluster
 - Implemented Whole Exome Sequencing data analysis pipeline by Workflow Description Language, enhancing efficiency up to 10x by taking advantage of parallel processing and better memory allocation
 - Standardized package management on servers by Conda, decreasing colleagues' 85% time for building a clean environment
- **Big Data Platform Building for FarEasTone Telecommunications ► Spark | Kafka | HBase**
 - Collaborated with IBM to design real-time analysis of location-based user data, gaining \$645K contract for Acer
 - Achieved processing 1.5M records/min by PySpark program which extracts data from Kafka and loads data into HBase
- **Smart Taxi Operating Platform for Taiwan Taxi ► Spark | Web Services | Hadoop**
 - Developed web crawlers for collecting information about concerts and exhibitions, improving 5% accuracy of the prediction model
 - Analyzed street hails in real-time by Spark Streaming, acquiring about 100K records/min
 - Orchestrated data collection API by Flask and Nginx for unveiling user behaviors
 - Designed the data pipeline from MySQL and AWS S3 to HDFS by Fluentd, Sqoop, Hive, and Impala, importing 35 GB data daily

SAS Institute Taiwan Ltd.

Consultant Trainee (Intern), Professional Services Division

Taipei, TW

Apr 2015 - Feb 2016

- **Pattern Recognition ► SAS | Python**
 - Implemented Gabor filter as texture features and HSV color model as a color feature for defects detection on semiconductor wafers
- **Data Mining ► SAS Enterprise Miner**
 - Advised customers on building pipeline for demand forecasting with SAS Enterprise Miner

PROJECTS

- **Water Quality Anomaly Detection, HackNTU ► Azure | Data Visualization** Aug 2015
 - Integrated governmental open data and Google Maps API into a website monitoring the water quality of Taipei
 - Established one-class SVM model for detecting outliers by Azure Machine Learning
 - Visualized time series data of water quality including turbidity, chlorides, and pH by D3.js
- **Chinese Characters Recognition, National Taiwan University ► Python | Machine Learning** Jan 2015
 - Achieved 70% precision of 32 possible labels of handwritten Chinese characters by KNN, random forest, and SVM
 - Implemented data pre-processing, training, and validation by Numpy, scikit-learn, and LIBSVM

ACHIEVEMENTS

- Produced mandarin technical translation of Romeo Kienzler's book, *Mastering Apache Spark 2.x*, in 2018
- Won 2nd Place out of 200+ teams with \$3,200 grant at **Big Data Analytics for Semiconductor Manufacturing 2015**
- Won Bronze Medal (placed 4th out of 2000+ teams from all over the world) with \$20,000 grant at **Acer Incredible Green Contest 2012**

SKILLS

- **Languages:** Python, C/C++, Java, SQL, Bash, R, SAS, JavaScript
- **Libraries:** PySpark, Scrapy, Flask, Numpy, Pandas, scikit-learn, TensorFlow, CUDA, LIBSVM, D3.js
- **Tools:** Conda, MySQL, Impala, Hive, Sqoop, Kafka, Nginx, Git, Docker, Ansible, Hadoop, HBase, Fluentd, Prometheus, Grafana, AWS, Azure