

Xinjie (Edward) Hu -- Amazon Web Service (AWS) Certified Solution Architect Professional

+86 134-1751-4224 | xinjie.h@outlook.com | xinjie-h.github.io

EDUCATION

Southern University of Science and Technology (SUSTech)

南方科技大学

Shenzhen, China

Bachelor of Economics in Finance (GPA: 3.75/4.00)

Sep.2016-Jun.2020

- Computer Science Related Courses: *Computer Programming Fundamentals (Java), Discrete Mathematics, Data Structures and Algorithm Analysis*
- Math Fundamentals: *Mathematical Analysis (Calculus), Probability and Statistics, Linear Algebra.*
- Honors: University Scholarships for three consecutive years (Top 10%).

CERTIFICATION & SKILLS

- **Certification:** Amazon Web Service (AWS) Certified Solution Architect Professional;
- **Experience:** data warehouse construction, query & storage cost optimization, DBMS cloud migration, *AdaBoost, NLTK, OpenCV*
- **Skills:**
Familiar: Python, Spark, Hadoop core components, OLAP Data Warehouse (Doris, StarRocks)
Know Basic: Linux command, Redis, Java, C++;

PROFESSIONAL EXPERIENCE

Sohu (NASDAQ: SOHU) - Intelligent Media R&D Centre

搜狐 – 智能媒体研发中心

Beijing, China

Data Engineer & Junior Software Development Engineer (Backend)

Since Jul.2021

- Storage cost optimisation: Combined bucket policy and partitioning for table life-cycle planning, which has kept the storage cost of HDFS cluster hot data layer stable;
- Query acceleration: Applied various schemas, including columnar storage (Parquet, ORC), JSON SerDe, HyperLogLog based on analytical scenarios, which significantly reduced runtime for querying large tables;
- Data Middle Platform (DMP) construction: Consumed various data sources (RDBMS via Sqoop, Nginx streaming log via Kafka) and adopted MapReduce, Spark and Python for data cleaning and processing. Constructed data warehouse and built API services for downstream business;
- Ticket-Handling: Processed user's daily questions on query optimization. Analyzed query plan and tuned complex queries with adjusting memory and concurrency properties;
- Backend service: Operated xxx-job based task workflow system and built alerting system on monitoring instance status;
- Cloud migration: Migrated ~1 PB on-premise data asset to cloud Hadoop cluster and object storage system.

SAP - Digital Aged Care Team

Sydney, Australia

Internship - Data Scientist

Jan.-May. 2021

- Goal: Delivered machine learning based Parkinson disease prediction model;
- Literature Review: Read medical image literatures and researched on clinical quick methods of diagnosing Parkinson and dementia disease;
- Data & Feature Extraction: Used OpenCV to process hand-drawn spiral images collected from Australian hospitals. Used histograms, dimension reduction methods to extract a series of features for identifying asymmetry and smoothness patterns on spirals;
- Model: Adopted AdaBoost with hyper-parameter tuning and pipelines to predict Parkinson disease;
- Outcome: Accuracy: 92.35%, AUC 0.86, F1 score: 0.78.

Academic Research @ SUSTech Shuo YAN¹'s Innovation Economics Research Group

Beijing, China

Co-author & Research Assistant

Oct. 2018-Sep. 2020

- Discovery: Demonstrated managers' overseas work experience can significantly improve firm citing more overseas patents, and expedite innovations;
- NLP on patent: Used Python NLTK and Pandas packages to tokenize patent text and adopted KNN algorithm to classify samples into different industries. Managed over 2 million patent data by MongoDB;
- Econometrics: Regressed 10-year panel data in STATA. Adopted the endogenous growth model and instrumental variable method to construct the patent citation model;
- Publication: Our research *Executives' Foreign Work Experience and International Knowledge Spillovers: Evidence from China* has been published on *Emerging Market Finance and Trade* (JCR Q1, ABS2), and archived by the 12th Xiangzhong Economic Seminar, Chengdu (第12届香樟经济学论坛, 成都).

¹ Click [here](#) for Assistant Professor Shuo YAN's personal website.