ZHEJIAN JIN

(917) 609-4100 | zj2324@columbia.edu | www.linkedin.com/in/zhejian-jin

EDUCATION

Columbia University, New York, NY
M.S. in Electrical Engineering Expected Dec 2022

- Coursework: Cloud Computing, Distributed Storage Systems, Large-scale Stream Processing, Heterogeneous Computing, DB System Implementation, Big Data Analysis, Blockchain, Practical Deep Learning Systems Performance

Shanghai Jiao Tong University, Shanghai, CN
B.S. in Electrical and Computer Engineering Aug 2021

- Coursework: Data Structures and Algorithms, Operating System, Computer Networks, Artificial Intelligence, etc.

SKILLS & INTERESTS

Languages: C, C++, Golang, Python, MySQL, CUDA, MATLAB, Verilog
Tools: AWS, GCP, Docker, Kafka, Hadoop, Spark, Airflow, ZeroMQ

EXPERIENCE

Jupiter Research Capital, New York, NY Jun 2022 - Sep 2022
Quantitative Developer Intern

- Built high-speed exchange data parsers and data searching engines with self defined index builder for large binary sequential data from China A-shares stock exchange, Nasdaq ITCH and NYSE Pillar Integrated Feed
- Finished low-latency trades&orders FIX protocol message asynchronous parsing module and high-throughput database module supporting 30k+ inserts/second for live trading messages and historical log files
- Wrote data feeder protocols using REST API, websocket API and FIX protocol for 9 crypto markets (FTX, KuCoin, etc.)

Columbia University, New York, NY Jan 2022 - May 2022
CSEE4121 Computer Systems for Data Science Course Assistant

- Designed programming projects using Apache Spark, Spark Streaming, HDFS in Google Cloud Dataproc
- Held office hour weekly and answer questions on Piazza daily to answer students' questions from homework and class content, took teaching-team meeting and graded homework with other 9 CAs for the class with 300+ students

East Money Information Co., Ltd., Shanghai, CN Dec 2019 - Mar 2020
Back-End Software Developer Intern

- Implemented network framework with Python to develop websites for large-scale stock data in MongoDB visualization
- Utilized Apache Kafka to maintain the distributed publish-subscribe messaging system
- Developed the Data Quality Detection System. Wrote TCP private protocols for specific market messages in Golang to handle 5 problems including packet loss, delay, corruption, duplication and reordering
- Completed the Data Quality Detection System with WeChat alarming function. Improved the network latency through multithreading TCP technique and optimized data-processing time 3 times faster than original design

RESEARCH EXPERIENCE

Columbia University, New York, NY Oct 2021 - Dec 2021
Evaluation of Disaggregated Persistent Memory System

- Benchmarked latency and throughput for both local and remote devices including Hard Disk, Ramdisk, DRAM, PM (Persistent Memory), remote DRAM and remote PM, where remote devices are accessed through RDMA over InfiniBand network adapter, PM is emulated by pmem.io and LXC containers are used to restrict the memory
- Designed disaggregated integral test framework, generated workload by memaslap and measured performance, stated insights and seven guidelines on how to design a disaggregated memory system designed for PM

Shanghai Jiao Tong University, UM-SJTU Joint Institute, Shanghai, CN Sep 2020 - Apr 2021
Deep Reinforcement Learning-based TSP Solvers

- Proposed the MAGIC2 model and Stochastic Curriculum Learning method. Optimized the RL model is by Policy Gradient with a dynamicfurthest-insertion+2-opt baseline during training (Accepted by IEEE SSCI 2021)
- Achieved state-of-the-art results as a Deep Reinforcement Learning solver for TSP, taking both generalizability and computational time into account. E.g.: 8.79% Gap from Concorde Method on TSP200 problem while saving 75% time

ENGINEERING PROJECTS

Columbia University, Intro to Blockchain, New York, NY Mar 2022 - May 2022

Dark: A decentralized social network

- Built Dark, a decentralized social network allowing users to post and comment through the Dapp on the CCN network
- Composed 2 smart contracts in Solidity: a contract of the decentralized social network class and a contract of the migration. applied Truffle and Ganache to deploy the contracts, develop the application, and run tests

Columbia University, Heterogeneous Computing, New York, NY Oct 2021 - Dec 2021

GPU Acceleration of K-Means Clustering

- Proposed multiple GPU parallelization ways to speed up the naive K-means, including using shared memory to calculate distance for each data point and using shared memory and parallel scan to sum up each centroid
- Implemented the algorithms on NVIDIA Tesla T4 in GCP. Beaten the Sklearn K-means algorithm by speeding up 5 times when the number of data points comes to 10^6 with dimension of 8 and the cluster number is set to 5

Columbia University, Cloud Computing, New York, NY Sep 2021 - Dec 2021

Cloud Restaurant Scheduling Platform Using AWS

- Formulated REST APIs of a restaurant scheduling platform and deployed 3 atomic microservices handling user and address, restaurant and schedule information on AWS EC2, docker and AWS Elastic Beanstalk
- Performed multiple Lambda Functions to provide different functionalities with different components with API Gateway
- Integrated third party Oauth and encapsulated external cloud service to verify an address

Shanghai Jiao Tong University, Capstone, Shanghai, CN May 2021 - Aug 2021

Cross Camera Video Analysing System at Edge

- Accomplished a video analysing system dynamically assigning workloads across 5 cameras and a NVIDIA edge cluster
- Applied ZeroMQ for data transmissions among 5 cameras and between cameras and the edge cluster
- Transferred the YOLO PyTorch model to RKNN type to be compatible with the NPU API. Pre-compiled the model and deployed the model to cameras by Docker achieving 20+frames per second

Shanghai Jiao Tong University, Operating System, Shanghai, CN Oct 2020 - Nov 2020

Development of Multi-thread Database

- Classified queries with table, maintained a dependency table and constructed a thread pool to run queries in parallel
- Divided large tables into small parts and executed operations such as searching and calculation simultaneously
- Utilized the random-access property of std::vector and kept a keymap to store indices for given keys in order to improve the efficiency of conditional queries