Zeyuan (Faradawn) Yang

https://faradawn.github.io/ | faradawn@uchicago.edu | (773)-231-3964

EDUCATION

The University of Chicago — Chicago, IL

Bachelor of Science in Computer Science and Mathematics (GPA: 3.66/4.0) — Sept. 2020 - Jun. 2024

RESEARCH EXPERIENCE

Optimizing Cache Performance for Multi-Process Servers — Nov. 2022 – Present

Traditional cache prefetchers struggle with the non-sequential access of modern multi-process applications. We develop a hybrid approach combining machine learning to identify access streams with classical algorithms for processing. This method outperforms both ML and huristic methods by up to 10% on certain traces.

Byte-VAE: A Novel, Memory-Efficient Image Generation Model (pdf) — Sept. 2023 – Dec. 2023

Vector-Quantize Variational Auto-Encoder (VQ-VAE) suffers from high memory usage due to need of a codebook. We designed an innovative rounding method that eliminates the in-memory codebook, achieving near-zero memory consumption while maintaining identical image generation quality.

Reducing I/O Latency in SSD Clusters via Neural Network — Nov. 2022 – Dec. 2023

Disk read and write (I/O) operations incur the highest latency in cloud environments. To avoid admitting an I/O to a "busy" solid-state drive (SSD), we designed a lightweight neural network for each SSD to estimate the completion time. Our model successfully reduced average I/O latencies by up to 79.6% with 10 nanoseconds overhead.

Performance Study of Seagate's Distributed Storage — Nov. 2021 – Nov. 2022

Seagate Technology's dsitributed storage, CORTX, faced a latency issue. I used a system, Analytic and Diagnostic Database, to collect metrics like queue depths and operation latencies. Compiling this data, I created a detailed graph of each component's time consumption, leading to a system upgrade that improved throughput by 20%.

PROFESSIONAL EXPERIENCE

Ovative Group — Minneapolis, MN

Data Scientist — May. 2023 – Aug. 2023

- Enhanced data ingestion from **13.31s to 1.17s** by introducing Hive partition to a Hadoop storage.
- Reduced **4.7x** processing time by converting CSV to Parquet, creating the first **Delta Lake** for ETL data analytics.

Seagate Technology — Chicago, IL

Research Scientist — Dec. 2021 – Jan. 2023

- Participated in writing a **C/C++** client for CORTX distributed cloud storage to improve performance.
- Created a large-scale **Kubernetes** deployment that achieved linear scalability in throughput.

TikTok — Beijing, China

Software Engineer — Jun. 2021 – Sept. 2021

- Enhanced the runtime of a **JavaScript** Sudoku from **140s to 0.02s** by optimizing the backtracking algorithm.
- Created a full-stack **HTML-Django** scheduling website that saved visitors **10 minutes** of registration time.

PROJECTS

IOS App: *Latin Garden* — Jan. 2021 – June. 2021

- Built a vocabulary app for Latin plant names. Received a **4.9 rating** with **122** reviews on App Store.

Enterprise Website: *US-Korea Conference 2022* — Jan. 2022 – Oct. 2022

- Developed an enterprise-level website streamlining paper submissions for **123** researchers.

YouTube: *Productivity and Tutorials* — Jan. 2020 – Present

- Created videos on how to install CUDA, Kubernetes, etc. Another channel on **Bilibili** has 5,000 subscribers.

SKILLS & INTERESTS

ACM-ICPC Programming Contest: 2023 Regional Silver Medal, 2022 Silver Medal.
Skills: C, C#, C++, Python, Java, JavaScript, R, SQL, PHP, Pytorch, GCP/AWS, data structure, database, Docker.
Courses taken: Operating Systems, Networks, Architecture, Machine Learning, Computer Vision, Statistics.