

Shuyuan Cao

24 years old | Female | Hunan Province, China
13298692458 | 1084255088@qq.com/224712221@csu.edu.cn



Education

Central South University (985, Recommended student) - Artificial Intelligence - Master Sep 2022 - Present

GPA: 3.78/4.0 **Research Area:** Recommend System、User Portrait

Major honors: **Province-level First Prize** of Hunan Graduate Student Artificial Intelligence Innovation Competition;
University-level **Outstanding Student** (2023); University-level **First Prize Scholarship** (2023)

Participating project: **National Social Science Foundation project** "Research on Intelligent Recommendation of Multimodal Resources for Children's Graded Reading for Smart Library" (Project number: 22BTQ033)

Research achievement:

Published paper: "Graph Convolutional Network-Based Repository Recommendation System" JCR Q2

Accepted paper: "Research on Children's Reading Portraits Based on Knowledge Graph" Chinese C, **Top Journal of Information Science**

Central South University (985) - Software Engineering - Bachelor Sep 2018 - Jun 2022

GPA: 86.47/100

Competition award: **National-level First Prize** of The 12th China College Student Service Outsourcing Innovation and Entrepreneurship Competition(Group leader); National-level College Student Innovation Project (Group leader)

Major honors: **Outstanding Graduate** of Hunan Province, **Outstanding Member** of Central South University (2022),
University-level **First Prize Scholarship** (2019), University-level **Outstanding Student** (2019)

Internship experience

PerfXLab - AI acceleration engineer Feb 2022 - Aug 2022

The internship concentrates on the operator optimization part of the **AI compiler**. Using the hardware characteristics, the optimization development of various operators on **dtu** is carried out, and the ultimate goal is to realize **the acceleration of AI training**. This project mainly uses C++ for development, Python for testing, to ensure the testing performance of four types of data fp32, fp16, int8, bool. I was responsible for the optimization and development of **concatenate operator**, experienced several rounds of iterative development, and finally realized the performance optimization of more than **500%** in main cases.

[Memory access optimization] : Using the hardware memory access feature (the higher the level of memory performance is higher), the continuous data is moved to the top hardware for **complex calculations**, and the **continuous memory access** is done on the bottom end.

[Time optimization] : Obtain the vpd graph through performance test, then determine the **operator performance bottleneck** (most operators are **IObound**, which is **intensive in accessing storage**). By adopting the **prefetching approach**, data is divided into multiple segments and transferred to different levels of memory access devices. This enables asynchronous concurrency, allowing memory access times to overlap and effectively **hide latency**, thereby significantly improving operator performance.

Alibaba Cloud - Basic R&D engineer Jun 2024 - Sep 2024

The work revolves around **cloud server and cloud computing**. I was responsible for the performance testing, targeted optimization of main sale instances, and development of ECS performance platform.

[Comprehensive Competitiveness Evaluation of Main Sale Instances]: Conducted comprehensive performance evaluations of up to 30 instance types (e.g., Yitian series, 7th, and 8th generation) across up to 20 key performance scenarios (such as Redis, MySQL, Hadoop, etc.). Provided data-driven insights into the advantages, cost-performance ratios, and optimal use cases for each instance based on actual performance and current pricing, supporting instance pricing strategies and user group targeting.

[Performance Testing and Optimization Methods]: Conducted performance testing and provided **optimization** recommendations across three primary application scenarios and other use cases, addressing performance at the **NIC queue** layer, **server** layer, and **ECS product** layer. Key achievements include:

- Achieving up to a **210% performance improvement** in MySQL and resolving issues of jagged performance fluctuations.
- Enhancing linear database **performance by up to 39%**.

[ECS Performance Platform Development]: Integrated **two new scenarios** into the ECS performance platform, enabling automatic performance monitoring during tests and automated result visualization.

Project experience

Research on recommendation of multimodal resources for children's graded reading for Children
- **Project Manager, organizer, developer**

Conducted data collection and user profiling in the early stages and built a children's reading platform based on Spring Boot, Spring MVC, and MyBatis-Plus in the later stages. The platform is expected to attract over 10,000 users after launch. The project involves data mining, recommendation systems, statistical analysis, and back-end development.

[User Profiling]: Collected data through questionnaires, processed and analyzed the data using Python, and integrated data crawled from various websites to build a book resource library. Connected users' book preferences, modality preferences, and type preferences with the resource library to create user profiles represented in the form of a knowledge graph.

[Reading Ability Assessment]: Applied the Group Decision-Making Analytic Hierarchy Process (GDM-AHP) and Entropy Weight Method to calculate the weights of multi-level labels for children's reading abilities, ultimately constructing a unified evaluation model.

[Platform Development]: Developed a children's reading platform based on the SSM framework, including modules for user management, reading ability evaluation, book recommendation, profile generation, book reading, and book reviews.

Professional skills

[Programming language] Familiar with **Python, C++, Java, Shell** programming.

[Operating system] Familiar with the basic knowledge of the **operating system**, including hardware structure, memory management, process scheduling, etc.

[Algorithm framework] Familiar with **tensorflow** common operators, such as bitcast, concat, reduce, slice, gather, transpose, etc., part familiar with the underlying implementation.

[Transformer architecture] Understand **transformer** Architecture, able to fine-tune the **LLaMA7B**.

Other information

[Certificate] : English CET-6 (543 points), CCF-CSP certification