SI CHEN

4703344900 ⋄ sichen.emory@gmail.com ⋄ https://github.com/meditates/

SUMMARY

Machine Learning Engineer and Software Developer with 7+ years of research and industry experience. Skilled in building scalable data and HPC solutions, containerizing complex systems, and leveraging machine learning to optimize performance. Passionate about bridging academic research with real-world data engineering and software development to deliver efficient and high-impact results.

SKILLS

Programming Language: Python, C, C++, Java, SQL, Shell scripting
Machine Learning & Data: PyTorch, Tensorflow, scikit-learn
DevOps & Containers: Docker, Singularity, GitHub Actions, Spack
Systems & Tools: Linux, Gem5, Perf, CUDA, HPC clusters, AWS, GCP

EXPERIENCE

National Center for Atmospheric Research (NCAR) | Research Intern   5/2023 - 12/2023
- Containerized HPC simulation applications using Docker and Singularity, reducing build time by 60%.
- Implemented CI/CD workflows (GitHub Actions) to automate validation processes, decreasing deployment errors by 40%.
- Deployed containers with diverse MPI/compiler versions across CPU and GPU nodes on supercomputers infrastructures, using Spack for management of software dependencies.

Bytedance | Research Intern   5/2021 - 8/2021
- Developed an AI-powered chatbot using natural language processing (NLP) models to automate error log diagnosis and root cause analysis, reducing support staff workload by 30%.
- Conducted API testing using Postman in the internal deep learning infrastructure platforms.

Netapp | Research Intern   5/2020 - 8/2020
- Optimized performance headroom predictive metrics for ONTAP data management software using queue theory and the half-latency rule, improving CPU utilization efficiency by 20%.
- Implemented workload merge processes for high availability (HA) failover scenarios, using performance indicators such as IOPS and service time to construct latency—utilization curves.
- Developed workload characterization by statistical analysis of service time distribution (mean and std) in Jupyter NoteBook to improve workload identification precision and curve fitting accuracy.

Emory University | Graduate Research Assistant   8/2017 - 12/2024
Optimized application performance across heterogeneous hardware using HPC simulations and ML
- Achieved a 127× speedup in model training time by developing a cross-architecture HPC prediction system with a Meta-learning model and Gem5 simulator.
- Enhanced the SimPoint architecture simulation acceleration framework using advanced clustering, achieving a 5× speed up in simulation time while maintaining accuracy.
- Utilized feature selection analysis to identify critical hardware performance events with Perf in HPC workloads, reducing data collection time by 95%.
- Developed a gradient-boosting classification model for storage provisioning using time-series-based feature extraction to identify concurrent I/O workloads.

EDUCATION

Ph.D. in Computer Science, Emory University, GPA: 3.98/4.00   *2017 - 2024.12*
M.S. in Electrical Engineering, Huazhong University of Science and Technology, GPA: 3.6/4.00   *2004 - 2006*
B.S. in Electrical Engineering, Huazhong University of Science and Technology, GPA: 3.84/4.00   *2000 - 2004*

SELECTED PUBLICATIONS

S Chen, S Gonzalo, A Wildani, *SimPoint++: Less Simulation Points*, SC24 Women in HPC Workshop
S Chen, S Gonzalo, A Wildani, *Few-shot HPC application runtime prediction*, Cluster 2023
S Chen, J Liu, Avani Wildani, CENSUS: *Counting Interleaved Functional Tenants on Shared Storage*, MSST 2020