浙江省科学技术奖公示信息表

提名奖项：科学技术进步奖

|  |  |
| --- | --- |
| 成果名称 | 大规模高效能智能计算关键技术及应用 |
| 提名等级 | 一等奖 |
| 提名书  相关内容 | 主要知识产权和标准规范目录、代表性论文专著目录。  1. 面向神经网络检查点数据的智能压缩存储方法和系统 ZL202110749499.6，证书号第6304907号  2. 基于内存池的智能计算系统架构和并行训练方法ZL202311072311.4，证书号第6416246号  3. 一种安全的调整确定性时延的方法ZL202111251690.4，证书号第 4938733号  4. 基于SDN与NDN的虚实结合动态流量调度方法及装置ZL202211483627.8，证书号第6330144号  5. 一种批量同质任务的处理系统，方法以及装置ZL202310676088.8，证书号第6348389号  6. NVAlloc: Rethinking Heap Metadata Management in Persistent Memory Allocators/Proceedings of the 27th ACM International Conference on Architectural Support for Programming Languages and Operating Systems  7. Accelerating Tensor Swapping in GPUs With Self-Tuning Compression/IEEE Transactions on Parallel and Distributed Systems  8. iCache: An Importance-Sampling-Informed Cache for Accelerating I/O-Bound DNN Model Training/Proceedings of the 2023 IEEE International Symposium on High-Performance Computer Architecture  9. HOME: A Holistic GPU Memory Management Framework for Deep Learning/IEEE Transactions on Computers  10. Exploring Post-Training Quantization of Protein Language Models/Proceedings of the 2023 IEEE International Conference on Bioinformatics and Biomedicine |
| 主要完成人 | 何水兵，排名1，教授，之江实验室  白冰，排名2，副研究员，之江实验室  杨非，排名3，副研究员，之江实验室  薛辉，排名4，副研究员，之江实验室  乔剡，排名5， 高级工程师，新华三技术有限公司  沈明，排名6，高级工程师，云尖信息技术有限公司  徐琪，排名7，副研究员，之江实验室  秦亦，排名8，工程师，之江实验室  朱俊，排名9，高级工程师，之江实验室  曹阳，排名10，无，之江实验室  杜锦，排名11，工程师，之江实验室  陈平，排名12，无，浙江大学  王健博，排名13，助理工程师，新华三技术有限公司 |
| 主要完成单位 | 1.单位名称：之江实验室  2.单位名称：新华三技术有限公司  3.单位名称：云尖信息技术有限公司  4.单位名称：浙江大学 |
| 提名单位 | 之江实验室 |
| 提名意见 | 浙江大学联合之江实验室等单位合作承担的“大规模高效能智能计算关键技术及应用”项目，针对大模型训练的算力卡脖子问题，依托一系列国家项目的支撑，在人工智能基础设施方面开展持续攻关，突破了一系列关键技术，成功研制大规模高效能智算系统，在面向大模型的计算优化、高效存储、网络通信和训练容错等方面开展了系统研究，形成了一套大规模智能计算系统的系统软件，支撑了国内最大算力最高密度单一万卡集群和国内最大规模混合万卡集群的构建和性能提升.主要创新包括：针对算力利用效率低的问题，提出了多种优化算法，实现了千亿参数大模型并行策略分钟级搜索；针对网络带宽利用率低的问题，提出了基于信元的通信负载均衡技术，提升了节点间网络有效带宽；针对存储系统效率低的问题，提出了基于数据重要性的高效缓存架构和自适应的显存-内存-磁盘多层存储间数据迁移方法，减小跨层数据搬运量；针对大规模训练容错开销高的问题，提出了并行细粒度增量检查点技术，缩短了检查点保存时间，提出基于机器学习的故障预测方法，大大缩短了硬件模块故障定位时间。  提名该成果为省科学技术进步奖一等奖 |