



构建万物互联的智能世界

# 数据新范式

## 释放AI新动能

# 华为

# 发布AI存储新品 助力释放AI新动能

2023年7月14日,华为发布大模型时代AI存储新品,为基础模型训练、行业模型训练、细分场景模型训练推理提供存储解决方案,助力释放AI新动能。

### 数据决定智能的高度,数据存储当作先锋

从互联网到移动化,从云计算到智能化,每一轮技术变革都为千行百业带来深远影响。而从通用大模型到行业大模型,AI大模型的持续迭代也将助推新一轮的科技革命。在从0到1构建AI大模型创新应用的同时,企业应该找到适配自身发展需求的数据底座,将数据存储建设作为AI建设的下一个加速点。

当前,企业在开发及实施大模型应用过程中,面临四大挑战:

第一,数据准备时间长,数据来源分散,归集慢,预处理百TB级数据需10天左右;第二,多模态大模型以海量文本、图片为训练集,当前海量小文件的加载速度不足100MB/s,训练集加载效率低;第三,大模型参数频繁调优,训练平台不稳定,平均约2天将出现一次训练中断,需要Checkpoints机制恢复训练,故障恢复耗时超过一天;第四,大模型实施门槛高,系统搭建繁杂,资源调度难,图形处理器(GPU)资源利用率通常不到40%。

### 构建数据新范式,释放AI新动能

作为数据的载体,数据存储成为AI大模型的关键基础设施。顺应AI发展趋势,针对不同行业、不同场景大模型应用,华为推出OceanStor A310深度学习数据湖存储与FusionCube A3000训/推超融合一体机,帮助解决企业开发及实施大模型应用难题。

OceanStor A310深度学习数据湖存储,面向基础/行业大模型数据湖场景,实现从数据归集、预处理到模型训练、推理应用的AI全流程海量数据管理。OceanStor A310单框5U支持400GB/s带宽以及1200万IOPS的性能,可线性扩展至4096节点,实现多协议无损互通。全局文件系统GFS实现跨地域智能数据编织,简化数据归集流程;通过近存计算实现近数据预处理,减少数据搬移,预处理效率可提升30%。

FusionCube A3000训/推超融合一体机,面向行业大模型训练/推理场景,针对百亿级模型应用,集成OceanStor A3000高性能存储节点、训/推节点、交换设备、AI平台软件与管理运维软件,为大模型伙伴提供全方位的部署体验,实现一站式交付。开箱即用,2小时内即可完成部署。训/推节点与存储节点均可独立水平扩展,以匹配不同规模的模型需求。同时,FusionCube A3000通过高性能容器实现多个模型训练推理任务共享GPU,将资源利用率从40%提升到70%以上。FusionCube A3000支持两种灵活的商业模式,包括华为昇腾一站式方案以及开放计算、网络、AI平台软件的第三方伙伴一站式方案。

### 合作开放,推动“产学研”同频共振

本次发布会上,华为数据存储产品线总裁周跃峰博士对话华为苏黎世研究所存储首席科学家张霁,就数据安全流转、企业数据接入大模型等话题展开探讨。张霁是华为众多新生代技术青年的代表,还有更多的华为员工在遍布全球的研究所以开展前沿研究,一起挑战难题、创造价值、推动科技发展。

截至目前,华为数据存储在全球拥有12个研发中心、4000多名研发人员、3000多项专利授权,并且与超过25所中国高校、30所海外高校建立合作关系,并成立8个联合实验室。华为自2019年起设置奥林帕斯奖,以激励全球科研工作者,突破存储领域关键技术难题。到2023年,已有多位来自全球顶尖高校与科研机构的专家获奖,他们在存储创新架构、创新算法等课题上取得了技术性突破,加速了科研成果产业化,实现产学研合作共赢。

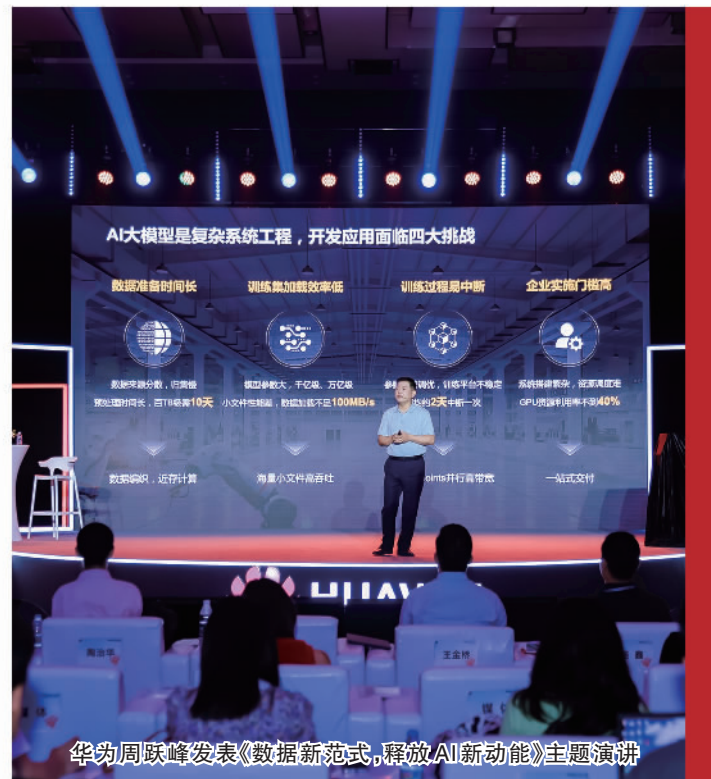
在本次发布会上,中国科学院自动化研究所紫东太初大模型中心、武汉人工智能研究院的专家及科大讯飞、智谱AI等企业相关负责人分别就AI大模型的应用实践以及基于华为AI存储的联合创新进行探讨,并对未来同华为在数据存储领域的合作进行展望。

数据来源:华为



构建万物互联的智能世界

选择华为 OceanStor 存储  
释放AI新动能  
数据决定智能新高度



华为周跃峰发表《数据新范式,释放AI新动能》主题演讲