



香港中文大学（深圳）货物类资产购置论证报告

一、基本情况

项目名称	人工智能教学智算集群（1）
项目金额	353.3706 万
论证编号	LZ202603006

二、货物清单

序号	货物名称	数量	单位	是否接受进口
1	智算管理登录服务器	1.0	台	否
2	智算开发计算服务器	4.0	台	否
3	高性能分布式存储	3.0	台	否
4	智算管理平台	1.0	套	否
5	RDMA 25G 以太网交换机	2.0	台	否
6	RDMA 100G 核心交换机	1.0	台	否
7	智算 IPMI 管理网络交换机	1.0	台	否

三、产品技术要求

序号	货物名称	招标技术要求
1	智算管理登录服务器	<p>★1.1 CPU: 颗数≥ 2, X86 架构处理器, 每颗处理器主频$\geq 3.25\text{GHz}$, 每颗核心≥ 32核, 线程≥ 64, 缓存$\geq 256\text{MB}$;</p> <p>1.2 内存: ≥ 8块 32G DDR5 4800MHz 内存;</p> <p>1.3 系统盘: ≥ 2块 960G SSD 6G SATA 3.5in RI SSD, 顺序读取$\geq 550\text{MB/s}$;</p> <p>1.4 数据盘: ≥ 1块 3.84TB NVMe U.2 SSD, 顺序读取$\geq 6800\text{MB/s}$;</p> <p>1.5 阵列卡: ≥ 1块 8口 12Gb SAS, PCIe 3.0, 支持 RAID 0/1/10;</p> <p>1.6 网卡: ≥ 2块双端口 25G 网卡, ≥ 4个 25G 端口互联, SFP28 封装, LC 接口, 波长 850nm 光模块和配套光纤;</p> <p>1.7 标准机架服务器, 高度 2U;</p> <p>1.8 电源: 2 块$\geq 800\text{W}$ 交流电源模块 (需满足 80PLUS</p>



		<p>240V 铂金标准);</p> <p>1.9 主板支持≥ 6*PCIe 全高槽位、≥ 4*PCIe 半高槽位, ≥ 2*OCP 3.0 槽位。</p>
2	智算开发计算服务器	<p>▲2.1 机型拓扑: CPU-GPU 连接关系, Switch 机型 Balance 拓扑;</p> <p>★2.2 CPU : 颗数≥ 2, X86 架构处理器, 每颗处理器主频≥ 2.45GHz, 每颗核心≥ 64核, 线程≥ 128, 缓存≥ 256MB;</p> <p>2.3 内存 : ≥ 16块 DDR5 RDIMM-32GB 4800MHz;</p> <p>2.4 系统盘: ≥ 2块 480GB 6G SATA 3.5in RI SSD, 顺序读取≥ 550 MB/s;</p> <p>2.5 数据盘: ≥ 1块 3.84TB PCIe4.0 x4 NVMe U.2 SSD, 顺序读取≥ 6800 MB/s;</p> <p>2.6 网卡: ≥ 2块 25Gb 光口-双端口, ≥ 8个 25G 端口互连, SFP28 封装, LC 接口, 波长 850nm 光模块和配套光纤;</p> <p>2.7 阵列卡: ≥ 1块 8 口 12Gb SAS, PCIe 3.0, 支持 RAID 0/1/10;</p> <p>2.8 电源: 4 块 ≥ 2600W 交流电源模块 (需满足 80PLUS 240V 铂金标准);</p> <p>★2.9 GPU: 单台服务器配置 GPU 数量 8 块; 单块 GPU 显存≥ 32GB (提供制造商产品说明书或制造厂商公开发布的产品功能印刷资料) FP32 non-Tensor 计算性能≥ 104.8 TFLOPS; FP16 Tensor TFLOPS with FP16 Accumulate≥ 419 TFLOPS; 支持 CUDA 编程开发, CUDA 核心数≥ 21760;</p> <p>▲2.10 Switch 机型下, 设备内卡间 P2P NCCL All_Reduce 通信平均速度不低于 40GB/s, NCCL All_Gather 通信速度不低于 40GB/s, NCCL All_to_All 通信速度不低于 20GB/s (需提供 NCCL 通信速度截图及 P2PBandwidth 卡间带宽速度截图);</p> <p>2.11 支持 ≥ 8个 PCIe5.0 x16 双宽全高全长插槽; ≥ 5个 PCIe 5.0 x8 单宽全高全槽; OCP3.0 卡支持可选。</p>
3	高性能分布式存储	<p>3.1 ★配置≥ 3节点分布式混闪存存储集群, 3 节点满足数据盘总裸容量≥ 480T, ≥ 24个 25G 端口光模块, SFP28 封装, LC 接口, 波长 850nm 的光模块和配套光纤;</p> <p>单节点设备规格: 2U, 系统盘: ≥ 2个 240GB SATA SSD, 缓存盘: ≥ 2个 1.92T-U.2 NVME SSD(读密集型), 数据盘: ≥ 8个 20T 机械硬盘; ≥ 2张 25G 双端口网卡, 冗余电源;</p> <p>▲3.2 智能数据预读, 支持对数据预读功能, 将热点数据预读至缓存层, 从而提升读写性能;</p> <p>3.3 用户配额, 文件存储支持针对本地的用户/用户组和域用户/用户组设置默认配额, 限制用户的使用容量或文件数量, 防止个别用户/业务滥用存储空间, 挤占其它用户的生产资源;</p> <p>3.4 文件数据保护, 配置对任意目录层级打快照的功能, 并支持对目录以及该目录下的子目录同时打快照, 定时快照间隔最短支持 15 分钟, 支持快照数≥ 20000个。支持快</p>



		<p>照重命名功能，支持快照点任意文件数据恢复；</p> <p>3.5 访问权限，支持存储本地用户权限管理，或支持对接标准 windowsAD 域控，对学校用户权限管理；</p> <p>▲3.6 文件压缩，为降低数据长期保存成本，文件存储应提供数据缩减的能力，支持以文件目录为单位配置数据压缩策略，可选择节省容量优先和性能优先两种策略，并支持查看计算压缩的数据量，提供产品功能页面截图；</p> <p>3.7 目录均衡打散，为使得所投产品在高并发场景下能发挥出更高性能，支持将某一目录下所有的元数据访问请求均衡打散到集群内多个节点的多个元数据服务 MDS 中；</p> <p>3.8 主机和硬盘故障容忍，支持数据副本和纠删码方式的冗余保护机制，支持 2 个或以上副本，纠删码最小 3 节点支持 4+2，5 节点支持 8+2，当主机或者磁盘故障，数据依旧可以正常访问；</p> <p>3.9 硬件亚健康检测处理，支持硬盘亚健康预测、检测和故障自动处理，坏道定时扫描，静默错误检测，能够提前预判硬盘故障风险，做好硬件更换准备，并在界面上做出提示；</p> <p>▲3.10 防勒索：基于业务安全考虑，为避免产生额外的经济损失，存储产品应具备勒索病毒的防护能力，内置免费病毒防护引擎，支持中毒前的勒索病毒防护、中毒时（或检测为疑似勒索病毒时）自动打快照、中毒后进行文件的安全恢复，能有效抵御勒索病毒，提供产品功能页面截图。</p>
4	智算管理平台	<p>4.1 整体要求：</p> <p>4.1.1 提供容器+镜像的管理，提供 GPU+CPU 集群的管理和调度；</p> <p>4.1.2 支持多数据中心管理，用户可以选择使用某一数据中心资源；</p> <p>4.1.3 具备 3 层组织架构管理；管理员、组织管理员、成员等，通过设定对组织及用户进行资源配额；支持对资源进行套餐分配及设置使用权限，可设置部分或全部使用；</p> <p>4.1.4 平台集成多种深度学习各种框架（如 TensorFlow, pytorch, caffe, keras 等）用户模型开发、模型训练等；支持用户自定义框架扩展；</p> <p>4.1.5 平台具备多种计费模式；提供完整的计量计费功能，管理可通过计费模块对 GPU、CPU、内存等资源进行费用设定，根据用户使用时长进行计费；</p> <p>▲4.1.6 系统支持管理员对下属课题组或用户付费的管理，支持用户通过支付宝等方式在线充值；也支持虚拟充值，在后台管理系统上由管理人员进行充值；也支持免费使用，提供产品功能页面截图，操作页面截屏作为证明材料；</p> <p>▲4.2 GPU 资源集中管理分配，提供产品功能页面截图，操作页面截屏作为证明材料；</p>



	<p>4.2.1 管理集群中 GPU 服务器及 GPU 卡的运行状况，资源使用情况，包含 GPU 总数/使用数，GPU 核心平均利用率，GPU 显存平均利用率等；</p> <p>4.2.2 支持多台服务器的 GPU 卡资源统一纳管，统一分配，支持 GPU 按照型号进行分组，当用户需要某一种类型的 GPU 卡时，就会从资源池中调度这一类型型号的卡给用户；</p> <p>4.2.3 支持单机单卡，单机多卡，多机多卡分配资源；提供单 GPU 卡多任务共享、GPU 切割（vGPU）方式分配资源，提升 GPU 利用率；</p> <p>4.2.4 组织和用户优先级设定，管理员可对用户添加优先级（1-5），优先级高的优先调度资源；</p> <p>▲4.3 大模型训练和推理功能（本项全部子功能均需提供软件功能截图和操作页面截屏作为证明材料，任意一条未提供截图证明则视作 4.3 条款不满足）</p> <p>4.3.1 平台管理员可自主新建开发、训练框架类型或版本；用户按需申请深度学习环境，包括深度学习框架，网络模型，GPU 及 CPU 等资源，支持用户对环境框架、资源配置进行动态调整；</p> <p>4.3.2 支持对环境做定时快照和备份，用于环境的快速回滚；支持把当前环境一键打包至镜像仓库，用户可以基于该镜像创建新的环境；</p> <p>4.3.3 提供 Jupyterlab 和 Jupyter notebook 的集成，支持 AI 相关数据集集成到 Jupyter 中，支持一键添加代码和数据集，提供在线编程及部署；支持 ssh 访问创建的开发环境，支持通过 Pycharm、VScode 等工具进行远程调试；</p> <p>4.3.4 支持 web 及 shell 提交训练任务，支持用户可以实时查看自己任务的运行结果；</p> <p>4.3.5 支持可视化作业管理，基于常用的 AI 算法框架及准备好的算法可以进行多版本任务参数调优，优化并增强机器学习全生命周期管理流程，提升模型训练任务的迭代效率；</p> <p>4.3.6 提供本地数据集如：计算机视觉 Images（超过 1810 万张图片）、自然语言处理 Audio（超过 1900 小时音频数据）、Multi-ModalImage-textparis（超过 630 万的数据对），可以在上述数据上进行训练与推理功能测试；</p> <p>4.3.7 提供预训练模型库包含以下模块：Deepseek 32B, Qwen3-30B-A3B 、 YOLO 、 FasterRCNN 、 SSD 、 MaskRCNN、 MMDetection、 SwinTransformer、 StyleGAN、 WasserteinGAN、 DenseNet、 CSPNet、 SimCLR、 U-Net、 MaskR-CNN 、 MobileNet 、 FLAME 、 YouTube-8M 、 AttentionOCR、 Transformer、 LSTM、 BERT、 RoBERTa 模块；</p> <p>4.4 镜像管理： 具备本地镜像仓库，支持镜像分组管理及分享，允许用户上传自定义镜像，支持将用户环境一键打包镜像仓库；镜像权限管理，可设置为公有或者私有；</p>
--	--



		<p>4.5 分布式训练功能：支持分布式并行训练功能，支持 TensorFlow、pytorch 分布式训练，同时集成 horovod、ray 分布式训练框架；用户按需动态申请分布式训练所需要的 GPU 数量和节点数量，平台可以实时监控每个节点的运行情况；</p> <p>4.6 模型管理：支持统一的模型管理功能，可以集中管理所有在模型训练中得到的模型，导入并管理在本地开发的模型，满足模型需要不断迭代和调试的管理需求；支持对模型进行多版本管理，可以将模型发布为服务；</p> <p>4.7 模型服务：支持将训练好的 AI 模型部署成在线服务 (WebService)，提供在线 UI 预测推理能力，为用户提供可调用 API 接口。支持设置模型服务自动停止时间，避免不必要的资源浪费；</p> <p>4.8 数据管理：采用分布式文件存储，把本地硬盘组成分布式存储；用户按需申请存储空间，提供 scp 访问和 Web 两种访问方式，支持用户存储扩容、重命名、修改密码、共享存储、文件编辑、压缩解压缩等功能；数据标注工具服务化，通过平台集成的数据标注工具，可以在线完成数据标注。</p>
5	RDMA 25G 以太网交换机	<p>5.1 交换容量$\geq 8\text{Tbps}$；</p> <p>5.2 包转发率$\geq 2400\text{Mpps}$；</p> <p>★5.3 支持≥ 48个 25G SFP28 端口，≥ 8个 100G QSFP28 端口；</p> <p>5.4 支持数据中心 RDMA、Puppet、Chef 特性，可以实现数据中心的自动化运维管理；</p> <p>▲5.5 支持 SPAN/RSPAN/ERSPAN 镜像和多个镜像观察端口，可以对网络流量进行分析；提供官网截图证明。</p>
6	RDMA 100G 核心交换机	<p>6.1 交换容量$\geq 6.4\text{Tbps}$；</p> <p>6.2 包转发率$\geq 2030\text{Mpps}$；</p> <p>★6.3 支持≥ 32个 100G QSFP28 端口，≥ 2个 1GE SFP 端口，≥ 12个 100G 端口互联，QSFP28 封装，MPO 接口，波长 850nm 光模块和配套光纤；</p> <p>6.4 支持数据中心 RDMA/Puppet、Chef 特性，可以实现数据中心的自动化运维管理；</p> <p>6.5 支持 SPAN/RSPAN/ERSPAN 镜像和多个镜像观察端口，可以对网络流量进行分析；</p>
7	智算 IPMI 管理网络交换机	<p>7.1 交换容量$\geq 672\text{Gbps}$；</p> <p>7.2 包转发率$\geq 207\text{Mpps}$；</p> <p>7.3 支持≥ 48个 10/100/1000BASE-T 端口，6 个 1/10GE SFP+端口，2 个可插拔电源；</p> <p>7.4 支持 STP/RSTP/MSTP/PVST 协议；</p> <p>7.5 支持 STP Root Protection。</p>



四、售后服务和验收要求

序号	目录	售后需求
(一) 免费保修期内售后服务要求		
1	免费保修期	原厂保修，货物免费保修期__3__年，自最终验收合格之日起计算
2	维修响应及故障解决时间	在保修期内，一旦发生质量问题，中标人保证在接到通知后__2__小时内响应，__4__小时内赶到现场进行修理或更换。
3	培训方案	设备安装完毕后，提供免费的产品使用培训计划，根据用户人员变动情况提供持续的培训和技术支持，主要内容为产品的基本结构、性能、主要部件的构造及原理，日常使用操作、保养与管理，常见故障的排除，紧急情况的处理等，培训地点主要在产品安装现场或按甲乙双方协商安排。
4	售后	3年软硬件维保服务，软件维保期内免费升级，具有中国计算机软件著作权登记证书的软件。
(二) 免费保修期外售后服务要求		
1	维保期外	中标人保证继续为采购人提供货物的维修服务，中标人以市场零售价格__八__折的配件价格向采购人提供备品备件。
(三) 其他交付要求		
1	关于交货	<ol style="list-style-type: none"> 1. 交货地点：香港中文大学（深圳） 2. 交货义务：中标人承担的货物运输、安装调试、验收检测和提供货物操作说明书、图纸等其他类似的义务。 3. 交货期限：签订合同后__30__天（日历日）内交货。
2	关于验收	<ol style="list-style-type: none"> 1. 采购人和中标人应在交付时对货物进行开箱验货，以确认货物的数量、型号、规格等是否符合合同要求。 2. 如货物经安装、调试、试运行后验收的，中标人应在货物到货并经开箱验货合格后__15__日内完成货物安装、调试的所有工作。 3. 采购人验收合格前，除货物已由采购人实际使



		<p>用的情况外，货物的一切风险（包括但不限于货物的损毁、灭失及可能的侵权等），均由中标人承担。</p> <p>4. 验收时，双方按照合同约定及时对交付的货物进行验收，中标人应配合采购人的履约检查及验收。</p> <p>5. 中标人货物经过大学组织的验收后，中标人需提供产品保修文件。</p>
3	检测验证	<p>如采购人发现中标人提供的货物与投标资料明显不相符且中标人不能提供证据，采购人有权直接通过第三方检测机构对于中标人提供的本项目全部或部分货物，依据投标技术响应情况逐一测试验证，其检测结果作为验证中标人提供货物与其投标资料是否相符的认定标准。如检测结果符合合同要求，其检测费用由采购人承担；如检测结果不符合合同要求，其检测费用由中标人承担。</p>

五、配套条件落实情况

主要配套条件落实情况：（明确具体的设备物资安装和使用场地、配套设施落实情况、特殊的使用环境要求，水、电、防磁、防震、机房等其他的配套要求，是否有承重问题等。原则上使用大学经费采购的 100 万元以上的大型仪器设备，均应列入 UCRF 统一管理，应提前明确存放地址。）

场地已配备了电源，网络，相关的配套设备都已具备。运维人员、运行机制、安全保障等方面均已建成成熟的管理机制。

设备物资管理和维修维护落实情况：（符合要求的设备物资管理人员或操作人员的落实情况，应明确设备物资具体的管理人员或团队，以及后续维修维护经费的支出渠道等。）

学校具备完善的 IT 数据中心管理/运维团队。

设备管理或操作人员资格证、设备物资购置和使用许可证等的落实情况：（涉及安全风险的填写。若是特种设备需取得《中华人民共和国特种设备作业人员证》或《中华人民共和国特种设备安全管理人员》，特种设备的使用许可证；放射源或射线装置所需的辐射安全许可证等。）



非特种设备

安全风险防护措施落实情况：（涉及安全风险的填写，涉及辐射安全、生物安全的按规定做环境安全风险评价；如涉及污染物、废弃物排放、危险品和易燃易爆等危险因素，则应提出计划的处理方式。）

不涉及安全风险

六、购置合规性

（配置是否符合国家及学校规定的配置标准，对属于国家或地方控制采购的设备物资，特别审批或许可产品是否已取得购置许可等。是否符合国家安全、卫生、环保等强制性规定）

配置符合国家及学校规定的标准，不属于国家或地方控制采购的设备物资，符合国家安全、卫生、环保等强制性规定。

七、共享方案（含校内外）

（根据国家和地方的相关要求，所有设备均应向全校无条件开放共享，单台件 ≥ 50 万的设备应按规定向社会开放共享）

按大学的相关制度执行。

八、专家论证意见

本次购置的设备用于教学，购置理由充分，能满足教学的实际需求，整体购置方案基本可行。

根据需求方提供的证明材料，市场上有多家供应商满足该项目技术要求和供货要求。

专家组一致同意通过“人工智能教学智算集群（1）”项目的论证。