**跨学科多模态人工智能计算平台(二期)服务器集群设备招标要求**

**一、设备要求和技术规格**

**1．总体要求**

**完整性要求：**本项目为交钥匙项目。中标人的工作范围不仅应包括“本技术规格”及所附的货物需求清单、技术条款和图纸上标明的要求，还应包括任何未明确标出的，但为保证全套设备安装后能够正常安全运行所不可缺少的配件及附件，投标时应自行计算和补充相应内容。

**报价要求：**投标人的报价不仅应包含招标文件提供的技术条款的要求，还应包括任何未明确标出的保证正常安全运行所不可缺少的配件及附件的全部费用。报价须包括但不限于货物费、材料费、设计和安装调试费、税金、运输和运输保险费、售后服务费以及其他的风险、责任等各项应有的费用。

**2. 设备配置清单**

**安装条件：**本项目所有服务器须为机架式服务器，可满足标准服务器机柜安装，所有服务器数量均匀安装在1个已有的冷通道内，冷通道有2排，每排8个机柜，单个机柜功耗不超过14kw。

**配置清单一览表：**

|  |  |  |  |
| --- | --- | --- | --- |
| **序号** | **货物名称** | **单位** | **数量** |
| 1 | GPU服务器（核心设备） | 台 | 33 |
| 2 | 存储服务器 | 套 | 1 |
| 3 | 管理服务器 | 台 | 3 |

**主要技术参数：**

注: 表中“★”代表关键指标，不满足该指标项将导致响应被拒绝；“▲”代表重要指标；无标识则表示属一般指标项。

1.GPU服务器（核心设备）

|  |  |  |
| --- | --- | --- |
| **序号** | **指标项** | **指标要求** |
| 1 | 数量 | 33台 |
| 2 | CPU | 单节点CPU数量≥2块 |
| ▲每块CPU满足：核心数≥56，线程数≥112（提供证明材料） |
| ▲基础频率≥2.0GHz，最大睿频频率≥3.8GHz（提供证明材料） |
| CPU缓存≥105MB |
| CPU间高速互联总线链接数≥4 |
| 支持AVX-512指令集，支持AMX、DLB、IAA等多款加速器 |
| 3 | 内存 | ▲采用DDR5 ECC REG内存，单节点内存容量≥2TB（提供证明材料） |
| 4 | 系统盘 | 单节点包含2块系统盘 |
| 采用热插拔SSD |
| 单块硬盘容量≥960GB |
| 5 | 缓存盘 | 单节点包含2块支持热插拔缓存盘，单块缓存盘容量≥7.68TB且最高顺序读取速度≥2000MB/s |
| 6 | 主板 | 主板内部至少配置4个PCIe Switch，每个PCIe Switch提供2组16通道的PCIe 5.0连接到CPU |
| 7 | GPU | ★单节点总显存容量≥768GB，单节点单精度计算能力理论峰值≥350TFLOPS（提供证明材料） |
| ★GPU服务器配置专用互联芯片，用于GPU服务器内部GPU之间数据互联，每块GPU至GPU连接最大通道带宽≥900GB/s，（提供证明材料） |
| 单GPU显存带宽≥4TB/s |
| ▲所有GPU总算力满足单精度理论峰值≥11.5PFLOPS（提供证明材料） |
| 支持FP64、FP32、FP16、INT8计算精度 |
| 支持任意GPU间通过网络的RDMA计算 |
| 支持TensorFlow、PyTorch、MXNet、Theano等深度学习框架 |
| 单块GPU最大TDP功耗不超过400W |
| 8 | 单机通信带宽 | 基于p2p BandwidthLatencyTest实测的节点内任意两张GPU卡间通信带宽≥735GB/s，基于nccl-test实测的节点内Allreduce通信性能≥430GB/s，要求提供包括软硬件环境、测试结果界面截图的测试报告，并加盖原厂公章 |
| 9 | 单机算力效率 | 基于主流开源模型Llama2-7B（TP=8，序列长度=2048）的整机FP16训练实测算力效率与理论算力效率比例不低于70%，要求提供测试报告并加盖原厂公章 |
| 10 | 集群整体性能 | 在满足不少于30张所投GPU卡的环境下，实测13B参数量或以上模型的FP16精度训练下单卡token值不低于1500tokens/s, 70B参数量或以上模型的FP16精度训练下单卡token值不低于270tokens/s, 8\*7B MoE参数量或以上模型的FP16精度训练下单卡token值不低于1100tokens/s，要求提供包括并行训练策略、batchsize设置、通信性能分析等的测试报告，并加盖原厂公章 |
| 11 | 数据传输能力 | GPU节点间通信采用InfiniBand网络，单个物理端口通信带宽≥200Gbps |
| ▲至少12个GPU节点每张GPU卡可独占通信带宽≥100Gbps，其余GPU节点每张GPU卡可独占通信带宽≥50Gbps（提供证明材料） |
| 每个InfiniBand物理端口须对应不同的PCIe插槽 |
| 12 | 可用性 | 配置热插拔冗余电源及风扇，满足负载散热需求，至少配置1个万兆网口，并配备足量10G SPF+光模块 |
| 13 | 可管理性 | 支持IPMI、远程监控图形界面，可远程通过独立端口对服务器完全控制，包括远程的开关机、重启、更新 Firmware, 虚拟KVM, 虚拟光驱等 |
| 14 | 高速计算网络接入 | ▲配置足量的InfiniBand网络组网设备，满足所有GPU服务器的InfiniBand端口与其他服务器的InfiniBand端口接入同一个InfiniBand网络的需求，链路带宽≥200Gb/s（提供证明材料） |
| ★整个高速计算网络须为不高于7:3阻塞的全互联的二层胖树拓扑的InfiniBand网络（提供证明材料） |
| 计算网络中，单台网络组网设备延迟≤130ns |
| 配置足量线缆，满足GPU服务器所有InfiniBand端口接入高速计算网络的连接需要 |
| 额外预留至少3个InfiniBand网口，满足和现有集群对接要求 |
| 15 | 千兆管理网络接入 | 所有GPU服务器的IPMI管理网口与其他服务器的IPMI管理网口须接入同一个千兆网络 |
| 16 | 万兆管理网络接入 | 所有GPU服务器的万兆以太网口与其他服务器的万兆以太网口须接入同一个万兆以太网络 |
| 17 | 性能测试 | 针对GPU服务器涉及的性能参数指标，须出具符合要求参数的测试报告，包括但不限于Stream显存带宽测试、GPU多卡互联带宽测试、NCCL集合通信带宽测试、开源模型测试、内存带宽测试、IB网络带宽测试、IB网络时延测试、IMB MPI时延带宽测试等 |

2.存储服务器

|  |  |  |
| --- | --- | --- |
| **序号** | **指标项** | **指标要求** |
| 1 | 数量 | 1套 |
| 2 | ▲文件系统 | 投标人应承诺配置商业版并行文件系统，该并行文件系统与所投存储产品兼容，且本次所投产品包含所有存储容量的并行文件系统的正版商业授权，所有存储服务器可通过该并行文件系统进行统一管理和使用（提供承诺函） |
| 3 | CPU | 单节点配置≥2块CPU |
| 单块CPU默认主频≥2.0GHz |
| 单块CPU核心数≥12个 |
| 4 | 内存 | 单节点配置≥128GB内存 |
| 5 | ★IO性能 | 整套存储子系统所有节点聚合读带宽≥136GB/s，聚合写带宽≥50GB/s，聚合读IOPS≥85.6万，提供产品证明（提供证明材料） |
| 6 | 系统盘 | 单节点配置2块热插拔SSD，单块硬盘容量≥480GB |
| 7 | ▲数据盘（提供证明材料） | 整套存储子系统数据盘可采用SSD或HDD  单块SSD数据盘容量≥7.68TB  单块SSD数据盘最高顺序读取速度≥2000MB/s |
| 8 | ★容量 | 整套存储子系统数据盘实际可用容量≥2.5PB（提供证明材料） |
| 9 | ▲可靠性 | 具备高可靠性，采用副本或纠删码方式保证数据安全冗余度，须至少保证任意1个节点或2块硬盘故障，业务不中断且数据不丢失（提供证明材料） |
| 10 | 数据传输能力 | 单节点须配置足量单端口InfiniBand网卡，网卡端口带宽≥200Gbps，满足存储通信需求 |
| 11 | GPU直连 | 支持GDS（GPU Direct Storage）功能，数据可不经过CPU、内存直接从存储放入到GPU应用程序中 |
| 12 | 权限认证管理 | 支持按用户/组/目录分别设置配额，支持AD/LDAP/NIS用户认证，支持ACL认证管理 |
| 13 | 配额 | 支持对用户、用户组、目录设置配额；配额类型支持统计配额、限制配额；统计配额仅监控存储的使用情况，不限制使用；限制配额监控存储使用情况，同时可以限制使用，超出阈值告警 |
| 14 | 客户端支持 | Linux私有协议访问，同时支持NFS、CIFS等标准NAS协议 |
| 15 | 统一命名空间 | 支持SSD 高速资源池和大容量资源池（如SAS或NL-SAS）合并成全局统一的命名空间 |
| 16 | ▲数据分级或分层功能 | 支持并配置数据分级或分层功能，支持配置数据写入策略和迁移策略，支持一次性或周期性数据迁移，支持按照根据数据缓存空间占比设置迁移策略，支持大小、用户、组、项目或文件扩展名等匹配设置分级或分层策略（提供证明材料） |
| 17 | 数据迁移 | 支持数据生命周期管理，可以根据策略自动迁移或备份数据 |
| 18 | 管理界面 | 提供GUI管理界面，支持运维可视化，无需第三方软件或插件，即可支持同一Web界面管理多套存储集群。 |
| 19 | ▲API接口 | 要求提供标准的API接口，可以供第三方系统调用存储运行信息的，包括但不限于：用户/用户组配额及用量，存储整体容量和使用量等（提供证明材料） |
| 20 | 兼容性 | 兼容POSIX标准接口，支持Linux、Windows、Vmware ESXi操作系统，KVM、vSphere等虚拟化平台 |
| 21 | 可用性 | 配置热插拔冗余电源及风扇，满足负载散热需求，至少配置1个万兆网口，并配备足量10G SPF+光模块 |
| 22 | 可管理性 | 支持 IPMI、远程监控图形界面，可远程通过独立端口对服务器完全控制，包括远程的开关机、重启、更新 Firmware, 虚拟 KVM, 虚拟光驱等 |
| 23 | 高速计算网络接入 | 配置足量的InfiniBand网络组网设备和线缆，满足所有存储服务器的InfiniBand网口与其他服务器的InfiniBand网口接入同一个InfiniBand网络，链路带宽≥200Gbps |
| 24 | 千兆管理网络接入 | 所有存储服务器的IPMI管理网口与其他服务器的IPMI管理网口须接入同一个千兆网络 |
| 25 | 万兆管理网络接入 | 所有存储服务器的万兆以太网口与其他服务器的万兆以太网口须接入同一个万兆以太网络 |
| 26 | 性能测试 | 针对存储服务器涉及的性能参数指标，须出具符合要求参数的测试报告，包括但不限于存储IO性能测试等 |

3. 管理服务器

|  |  |  |
| --- | --- | --- |
| **序号** | **指标项** | **指标要求** |
| 1 | 数量 | 3台 |
| 2 | CPU | 单节点配置≥2块CPU |
| 单块CPU默认主频≥2.0GHz |
| 单块CPU核心数≥24个 |
| 3 | 内存 | 单节点内存≥256GB |
| 4 | 系统盘 | 单节点配置2块热插拔SSD，单块硬盘容量≥480GB |
| 5 | 数据盘 | 单节点配置2块热插拔SSD，单块硬盘容量≥7.68TB |
| 6 | 数据传输能力 | 单节点配置≥1块单端口InfiniBand网卡，网卡端口带宽≥200Gbps，及≥1个SFP+万兆以太网口 |
| 7 | 可用性 | 配置热插拔冗余电源及风扇，满足负载散热需求，至少配置2个万兆网口，并配备足量10G SPF+光模块 |
| 8 | 可管理性 | 支持 IPMI、远程监控图形界面，可远程通过独立端口对服务器完全控制，包括远程的开关机、重启、更新 Firmware, 虚拟 KVM, 虚拟光驱等 |
| 9 | 管理软件 | ▲整体要求：投标人应承诺提供1套管理软件，实现对本项目涉及计算资源进行管理，所有软件指标均能提供对应的软件功能页面截图或产品白皮书对应功能简介等证明材料，提供的管理软件须为成熟的商业软件，提供商业版授权，满足本项目所有服务器使用需求（提供承诺函） |
| 节点和用户管理：支持通过软件对集群进行管理，实现节点上下线、新增或删除节点，新增、修改或删除用户、用户资源分配等功能 |
| 集群监控：支持对节点状态、节点使用率、作业状态等指标进行监控 |
| 资源调度：支持集群资源统一调度，支持多用户，多作业同时运行，通过调度器给作业分配资源，支持单节点单GPU,单节点多GPU，多节点多GPU等多种调度方式 |
| 容器管理：包括可移植容器的编排管理，计算环境隔离互不影响，内置常用AI框架镜像，兼容web开源镜像和用户自定义镜像，用户可进行镜像上传、分享，以便提供其他用户使用 |
| 报表：支持以用户、资源组等维度对集群资源使用情况进行统计并生成报表 |
| 大模型支持：可支持大模型训练的快速开展，支持通过软件平台，对大模型训练状态进行实时监测，包括但不限于节点使用率、节点状态等，同时具备一定的训练故障自动修复能力 |
| ▲任务及作业管理兼容性：管理软件提供标准的API接口及文档或手册，并支持与学校平台对接实现任务/作业管理功能；可通过提供的接口进行任务/作业的资源需求情况、作业状态等；能通过接口实现作业的基本维护操作，如提交、取消等（提供证明材料） |
| ▲作业及用户信息兼容性：承诺管理软件提供标准的API接口及文档或手册，支持与学校平台对接实现调取作业/用户用量信息的功能；可通过提供的接口获取具体调用节点/调用CPU资源数/调用GPU资源数等资源申请/或分配情况等；若系统采用队列、分区等方式对计算资源进行资源划分管理时，应能提供具体作业的用户名/用户组、分区名/队列名等分区信息等标识，以实现通过接口获取作业和资源使用信息（提供承诺函） |
| 数据管理兼容性：管理软件提供标准的API接口及文档或手册，并支持与学校平台对接实现数据管理功能：提供数据管理接口，支持包括但不限于安全获取用户数据列表、传输数据、修改数据、复制数据等 |
| ▲数据传输兼容性：承诺可以完成与学校平台的对接实现数据管理功能，管理软件提供标准的API接口及文档或手册，通过提供的数据传输接口，能够利用代理中转节点实现与其他集群同步数据到本集群，同时提供数据完整性校验（提供承诺函） |
| ▲统一认证对接：支持CAS，LDAP，OAuth等标准协议，支持NIS或LDAP等标准协议，实现用户/用户组的配置及验证接口（提供证明材料） |
| ▲统一监控接入：承诺配合学校部署学校统一监控，本项目中所有设备，包括：计算节点、存储服务器、管理服务器等，应支持基本的监控通信协议如SNMP或根据需要部署监控代理来实现与统一监控平台的对接，帮助学校实时获取设备运行情况，包括但不限于：CPU负载、内存负载、GPU卡负载、运行功耗、机箱温度、故障告警等。相关信息如若无法准确获取，应配合部署需要提供必要的技术支持和文档，如：技术文档、API说明或MIB库等（提供承诺函） |
| ▲交互式开发：支持网页端在线交互开发，可在线使用如Jupyter、Webshell或其他在线编辑器，或可对接其他第三方开发工具（如VS Code等）（提供证明材料） |
| 显存隔离：支持容器间GPU显存隔离、GPU复用或基于MIG的容器创建 |
| ▲可靠性：软件支持高可用部署，至少1个管理节点故障时，业务不受影响（提供证明材料） |
| 10 | 高速计算网络接入 | 配置足量的InfiniBand网络组网设备和线缆，满足所有管理服务器的InfiniBand网口与其他服务器的InfiniBand网口接入同一个InfiniBand网络，链路带宽≥200Gbps |
| 11 | 千兆管理网络接入 | 配置足量的千兆网络组网设备，满足所有管理服务器的IPMI管理网口与其他服务器的IPMI管理网口接入同一个千兆网络的需求，并至少配置2个万兆端口以及对应SFP+万兆模块用于上连汇聚交换机，同时须配置足量线缆 |
| 12 | 万兆管理网络接入 | 配置至少2台48口万兆网络组网设备，满足本项目所有服务器的万兆以太网接入需求，所有网口须配置足量线缆和SFP+万兆模块，交换机需支持100G堆叠，并提供2根100G QSFP28 100G堆叠线2根。每台交换机须配备至少2个40G上连光接口，同时须配置足量线缆及至少8个40G多模光模块 |
| 13 | 安装部署 | 投标须提供详细的集群安装部署相关文档，包括但不限于：网络拓扑图、软件系统规划、安装实施方案等 |
| 14 | 安装调试总体要求 | 设备到货后15天内完成平台的安装调试，性能测试，学校平台对接等工作，确保计算平台可以进入多模态大模型训练的测试阶段 |
| 15 | 安装调试 | 硬件设备安装上架、走线、服务器操作系统及驱动程序安装和本项目所涉及软件的安装部署 |
| 16 | ▲学校平台对接 | 承诺可以完成与统一身份认证和统一监控系统的功能对接，满足通过学校E算平台可集中访问和使用的需要（提供证明材料） |
| 17 | 与一期平台兼容性 | 已建成一期平台为独立集群，由32台8块高端GPU卡服务器、20台高速全闪存储服务器、3台管理服务器组成，整个高速计算网络为14台InfiniBand HDR 200Gbps交换机组成的全线速二层胖树网络。通过平台间联动调试，实现与本项目一期平台间的数据高速传输，同时提供一期、二期集群合并的技术方案 |

★**指标和▲指标如需提供技术支持资料，以制造商公开发布的印刷资料（产品彩页、产品说明书、产品白皮书、官网截图）、界面截图、第三方检测机构出具的检测报告或在有效期内的证书扫描件等为准，投标人可以只提供上述材料关键页的复印件，关键页需体现投标产品的品牌、规格型号、相关性能指标，相关性能指标需以醒目的方式标明；如需提供承诺函需标明页码并盖章，凡不符合上述要求的视为无效技术支持资料。**

**二、其他项目要求**

**1．项目团队要求**

须提供1名工程师进行1年的5x8小时驻场服务，主要负责本计算平台设备日常巡检、硬件故障处置、用户使用问题处理、协助大模型训练过程中性能调优分析、管理平台运维工作。驻场工程师需要有二年以上HPC或AI领域从业经验，具有工程师及以上职称优先考虑；并承诺除非该工程师离职或招标人提出，否则不得另换他人。驻场期间由招标方安排日常工作。要求提供该工程师的姓名、社保证明及工作简历（含现职位说明、工作经历等）。

设备原厂商能够提供完备的技术服务团队，在人工智能领域面向多模态大模型训练过程中的参数调试、性能优化、故障排查等方面提供技术支持。团队有能力提供大模型训练成熟服务的优先考虑，需提供相关服务的发布链接、界面截图、用户手册、相关案例。团队能提供相关中文巨量模型的学习优化方法的优先考虑。

**2．质保要求**

仪器最终验收合格后提供整机60个月质保，所有软件、硬件需提供7x24小时原厂维保，10分钟电话响应，4小时上门，硬件故障承诺24小时内更换配件完成修复，软件保修期内提供免费升级。每季度提供系统巡检服务并出具巡检报告。

**3．验收标准**

在用户和设备技术人员双方确认设备的各项技术指标均达到要求且运行正常后方可对设备进行验收。