浙江工业大学采购项目竞价执行建议书（货物类）

|  |
| --- |
| **采购项目概况** |
| 项目名称 | 生物信息高性能计算设备扩容 |
| 确认书编号 | [2025]26867 | 预算金额（万元） | 540000 | 经费项目负责人 | 周晓根 |
| 采购联系单位 | 信息学院 | 具体项目负责人 | 周晓根 |
| 项目联系人 | 周晓根 | 项目联系人手机 | 15257118077 |
| 预计竣工时间 | 2025年6月10日 | 所在校区 | 屏峰 |
| 填表时间 | 2025年5月21日 | 填表人 | 周晓根 |
| **采购项目内容** |
| 序号 | 项目内容 | 规格型号 | 单位 | 数量 | 参考价（万元） | 品牌（推荐品牌数量3个及以上） |
| 最高单价 | 小计 |
| 1 | 生物信息高性能计算设备 | 见主要技术指标 | 套 | 1 | 54 | 54 | 浪潮、宝德、联想 |
| 2 |  |  |  |  |  |  |  |

（生物信息高性能计算设备扩容）

主要技术指标：

一、GPU计算节点（1台）：

|  |  |
| --- | --- |
| **指标项** | **技术要求** |
| 总体要求 | 机架式服务器，高度≥4U  |
| ▲处理器 | 配置两个Intel Xeon 8460Y+(40C, 2.0GHz) |
| 内存 | 最大支持32个内存插槽；支持高级内存纠错（ECC）、内存镜像（Ememory mirroring）、内存热备（rank sparing）等高级功能，最大支持4T内存容量，支持4800MT/s工作频率； |
| ▲硬盘 | 配置1张Raid卡（不低于RAID -SAS 8I 2G 12GB 支持0/1/5/6/10/50/60）最大支持12块3.5寸硬盘或24块2.5寸硬盘，最大支持8块NVME硬盘；本次配置2\*960G SSD硬盘，2\*3.84T NVME SSD； |
| ▲GPU卡 | 配置2块计算卡，单颗GPU参数要求：FP32≥59TFlops，FP16≥119TFlops，INT8≥238TFlops，显存大小≥48GB，显存带宽≥864GB/s，TDP≤350W；配置2块计算卡：单颗GPU参数要求：FP64≥9.7TFlops，FP32≥19TFlops，显存大小≥80G，显存带宽≥1900GB/s，NVLINK带宽大于等于400GB/s，CUDA核≥6900，Tensor核≥432，TDP≤350W |
| 拓扑切换 | 可支持一键拓扑切换，提供技术白皮书截图。 |
| ▲网卡 | 配置1张4GE千兆网卡配置1张双口万兆网卡，含光模块； |
| 电源 | 配置4个3000W热插拔电源，支持2+2冗余； |
| 管理 | 配置≥1个1Gb 管理端口，可实现与操作系统无关的远程对服务器的完全控制，包括远程的开机、关机、重启、更新Firmware、虚拟光驱、虚拟文件夹等操作，提供服务器健康日记、服务器控制台录屏/回放功能，能够提供电源监控，可支持动态功率封顶。 |
| 性能 | ★1、支持CPU和内存的压测功能，该功能需支持自定义CPU测试百分比、CPU测试时间、内存测试时间及采集频率,上述功能提供软件界面截图★2、AI 多机并行加速性能测试，测试至少两种深度学习框架（如 Caffe,TensorFlow, Torch 等），测试模型采用 AlexNet，或 Resnet 从 1个 GPU 卡到 16 个 GPU 卡的性能，加速比≥14x，提供对应的测试证明； |
| 集群管理 | ★1、配置联合管理功能，实现小规模集群的统一管理，简化了设备的运维管理，所管理设备类型包含存储、节点、交换机、路由器等，提供第三方测试报告；★2、所管理设备品牌包含但不限于华为、浪潮、联想、华三、MAIPU、ZTE等，所展示参数包含IP、健康状态、厂商、型号等，提供软件功能截图； |
| 功耗管理 | ★为防止机柜负载过高，节点管理平台支持基于数据中心机房的3D温度拓扑图，提供第三方测试报告；同时提供机柜进出风口温度参数，提供软件功能截图； |
| 功耗监控 | ★支持对机房进行制冷分析及提高制冷优化建议（支持ASHRAE推荐温度、ASHRAE一级许可和ASHERAE二级许可三种规范进行分析。列出高温热点设备，并给出建议和提供优化方案），提供软件功能截图 |
| 安全性 | ★支持服务器厂商自研的主机安全系统加固软件（非PCIe形式的安全板卡、OTP双因素认证、黑匣子等其他服务器管理软件自带功能等），从操作系统内核实现对服务器的安全加固，与服务器同一品牌。该系统可实现内核级安全加固，增强型身份认证、服务完整性检测、注册表防篡改机制等功能。提供三级等保检验检测报告及移动存储介质注册、文件完整性检测、违规外联检测等软件功能截图 |
| 软件 | 人工智能开发平台（1套）： |
| 指标项 | 技术要求 |
| 功能 | 1、总体要求：支持通过Docker快速创建计算环境，基于Kubernetes进行可移植容器的编排管理，计算环境隔离互不影响，内置各种AI框架镜像，包括 Tensorflow、Caffe、pytorch、PaddlePaddle、MXNet主流框架版本，兼容web开源镜像和用户自定义镜像；2、支持集群资源统一调度，支持多用户，多作业同时运行，通过调度器来给作业动态分配资源，支持单节点单GPU, 单节点多GPU，多节点多GPU以及多节点GPU，以及CPU/GPU混合的多种调度方式；3、数据管理：提供基于Web的文件管理功能，支持文件和文件夹的创建、删除、重命名、在线编辑、权限设置，支持文件批量上传和打包下载；平台支持数据隔离与协同，个人数据可通过共享提供给平台其他人员使用；平台支持模型开发时的数据加速功能，用户可直接使用远端共享存储上的文件，也可以将远端存储的文件拉取到计算节点本地参加计算，从而提升计算性能，同时减轻高并发下存储Server端的I/O压力；提供软件功能截图；4、交互式开发：支持用户通过平台内置AI镜像或用户自定义镜像进行环境创建交互式开发环境，环境实例可以使用CPU资源也可以使用GPU资源；开发环境支持Jupyter、web shell的在线交互开发，支持对接第三方开发工具（如VSCode、PyCharm）；平台支持开发实例的持久化，在实例中安装的软件包在下次创建启动后可以继续使用；提供软件功能截图；5、模型训练：支持基于容器的模型训练功能。对Tensorflow、Caffe、pytorch、PaddlePaddle、MXNet框架，用户可自定义训练使用的框架版本，容器数量，GPU数量，内存，指定不同的GPU型号资源，并且可以实时查看训练日志，监控各容器内资源使用状况；支持AI模型的多机分布式训练，提供模板可以提交Tensorflow、Caffe、pytorch、PaddlePaddle、MXNet框架的分布式训练作业；提供软件功能截图；6、资源调度能力：使用SLURM调度，支持任务级别的资源调度方式功能，支持指定节点、GPU型号亲和调度、数据亲和调度策略，节点NUMA亲和调度，支持用户资源配额，用户组轮询调度等策略，高/中/低优先调度、支持紧急任务调度；★7、细粒度调度：提供基于web的GPU细粒度调度设置，允许多个任务指定GPU显存，调度到同一张GPU卡，GPU切分需同时支持：1）MIG模式：每张卡独立配置MIG方案；2）显存隔离：按GPU显存进行任意大小的切分，最小支持1GB显存；3）复用率：按GPU上并发任务个数进行配置提供软件功能截图；8、报表统计：支持从集群、资源组、节点三个维度进行计算资源的性能及使用统计，包括CPU利用率、GPU利用率、CPU使用核数、GPU使用卡数信息；支持对集群周期内的任务情况进行统计，按任务规模和任务时长进行数量统计；支持用户和用户组统计周期内的资源使用明细，包括任务总量、任务总机时、任务总CPU核时、任务总GPU卡时、数据存储 ；提供软件功能截图；★9、异构加速卡兼容性：支持异构加速卡统一调度，支持纳管国内外主流AI加速卡，包括但不限于英伟达、寒武纪、燧原、天数智芯、灵汐、华为昇腾等，支持统一管理调度，要求提供不少于3家AI国产芯片厂商与平台厂商的互认证证明并加盖双方公章；10、支持对接多种存储系统，包括NFS、BeeGFS、HDFS；支持通过界面配置同时对接多个存储系统；11、高可用：平台自动检测主备机状态及HA服务状态，当发现主机存在异常需要切换备机时，HA模块会自动切换管理服务到备机，并且在切换过程中不影响用户运行任务；12、提供自动报警功能，能够提供物理节点或容器的CPU、网络、GPU、磁盘使用的实时数据统计，报警项设置和报警策略，达到报警阈值时自动报警；13、支持集群横向扩展：当系统出现性能瓶颈，可以通过新增节点，添加到集群中实时扩展；14、安全管理：平台需支持内建身份认证机制接入，支持对接第三方账号系统；15、数据保密性：存储在数据库中的关键敏感数据以加密形式存储，支持AES等常见加密算法；16、系统安全：平台支持与其他系统隔离，防止其他系统未经授权访问本系统接口，提供API接口上增加签名校验或SSL安全证书；▲17、因此次为高性能设备扩容项目，故提供产品必须满足与一期设备集群兼容，如出现不兼容需退货处理。 |

1. 管理交换机（1台）：

|  |  |
| --- | --- |
| 指标项 | 技术要求 |
| 功能 | 1、交换能力：交换容量≥2.56Tbps,三层包转发率≥1260Mpps，若产品官网或彩页中有X/Y参数，则以X参数为准；2、接口配置要求：固定接口≥16个10/100/1000Base-T电口，≥20个1G/10G SFP+ 端口，≥8个10G/25G SFP28端口；配置4个万兆多模模块插槽：可插拔模块化电源≥2，扩展插槽≥23、二层功能：最大MAC地址≥144K，支持4K VLAN，支持QinQ、支持STP/RSTP/MSTP，支持端口镜像，流镜像及远程端口镜像，支持Jumbo≥9K；4、三层功能：路由表≥48K，支持IPv4静态路由，RIPv1&v2，OSPF，BGP，ECMP，PBR;支持IPv6静态路由，OSPF V3，RIPng，ICMPv6，NDP，PMTU；支持VRF功能；5、组播协议：支持IGMP，PIM-DM，PIM-SM，PIM-SSM，Static Multicast，MLD v1/v2，MLD Snooping ，PIM-SMv6，MVR6；6、#安全：支持dot1X认证、AAA认证，支持Radius、TACACS+认证，支持IP SLA7、★电源：实配可热插拔的冗余电源，实现1:1冗余，支持220V双交流或高压直流240V供电，整机最大功耗≤100W；8、★Vxlan：支持VxLAN二层/三层网关，支持VXLAN tunnel，支持VXLAN 路由，支持BGP-EVPN。9、支持M-LAG技术，将多台物理机箱组成1台虚拟交换机，实现多台设备间的链路聚合 |

供应商响应附件要求：

1. 质保期三年，中标后5日内提供原厂授权和原厂售后服务承诺函，服务器SN码官网可查全部配件信息。

售后服务要求：

1. 质保期过后，厂商终身维护，并提供成本价零配件、免收人工差旅费。
2. 现场安装调试、并安排现场技术培训，终身提供技术服务、技术支持。
3. 中标后一周内签订合同 30 天内交货。
4. 提供数据整合迁移服务，提供完整的集群平台搭建和软件系统安装售后服务。人工智能开发平台搭建，集群调度软件的安装优化，为保障设备良好运行与售后保障。应在不超过2小时内做出响应，4小时内到场，不超过2个工作日内解决故障。
5. 成交后不能如期签订合同、不能如期供货的，将上报财政厅和政采云追究其相关责任。
6. 因此次为高性能设备扩容项目，故提供产品必须满足与一期设备集群兼容，如出现不兼容需供应商退货处理，并将上报财政厅和政采云追究其相关责任。