AI大模型GPU算力集群采购需求

**一、采购标的需实现的功能或者目标，以及为落实政府采购政策需满足的要求：**

**（一）采购标的需实现的功能或者目标**

本项目采购AI大模型GPU算力集群1套，配置为GPU≥8块，单卡141GB HBM3e显存，支持第四代NVLink（900GB/s带宽），用于支撑大规模多模态模型训练与推理研究，满足智能感知、语言理解、大规模仿真等任务对高算力和高带宽的要求。要求实现大模型的高效并行训练、低延迟推理服务，具有高显存、高吞吐、高并行和良好的能耗控制能力，支持主流深度学习框架，满足科研环境下复杂任务的长期稳定运行需求。

**（二）为落实政府采购政策需满足的要求**

1.根据《政府采购促进中小企业发展管理办法》（财库【2020】46号）规定，本项目采购标的为中小型企业制造、承建或承接的，投标人应提供办法规定的《中小企业声明函》，否则不得享受相关中小企业扶持政策。投标人应对提交的中小企业声明函的真实性负责，提交的中小企业声明函不真实的，应承担相应的法律责任。

本项目采购标的对应的《中小企业划型标准规定》所属行业为：工业。

2. **□ 本采购项目允许进口产品参加。**

**（说明：请项目单位根据采购实际情况在“□”中打勾（☑）。未进行勾选的，视为只接受本国产品参加）**

**二、采购标的需执行的国家相关标准、行业标准、地方标准或者其他标准、规范：**

采购项目中所含的投标产品及制造商应符合国家有关部门规定的相应技术、计量、节能、安全和环保法规及标准，如国家有关部门对投标产品或其制造商有强制性规定或要求的，投标产品或其制造商必须符合相应规定或要求，投标人须提供相关证明文件的复印件。

**三、采购标的概况**

（一）采购项目名称： AI大模型GPU算力集群

（二）采购数量及计量单位： 1台

（三）最高限价：人民币 2,680,000 元。

（四）交付时间：合同签订后 7 天内。

（五）交付地点： 西安交通大学兴庆校区彭康楼 。

（六）付款进度安排：验收合格后第一年支付合同总额的50%，验收合格后第二年支付合同总额的50%。

**四、采购标的需满足的质量、安全、技术规格、物理特性等要求：**

1. CPU≥2\*Intel Xeon Platinum 8558（48核/96线程，主频2.1-4.0GHz）；

2. 内存≥32根64GB DDR5 RDIMM（5600MHz）；

3. 系统盘≥1块960G U.2 NVMe PCIe4 SSD；

4. 存储≥2块3.84TB U.2 NVMe PCIe4 SSD；

5. GPU≥8块，单卡141GB HBM3e显存，显存带宽：4.8TB/s，支持FP8/FP16/TF32/INT8精度计算,FP16半精度1.98P（稀疏模式）,最高TDP 700W，互联技术：第四代NVLink（900GB/s带宽），支持多卡全互联拓扑；

6. 可靠性：要求厂商和配件GPU计算卡生产商具备长期合作关系，提供相关证明;（1）生产厂家具备GPU计算卡生产商解决方案合作伙伴授权证书，（2）生产工厂通过ISO9001质量管理体系认证和3C强制性产品认证；（3）出厂提供ATA功能测试报告和ATA压力测试报告；

7. 管理软件：

7.1. 平台支持免费安装，持续免费升级与补丁发布，通过Web页面或命令行自动获取最新版本，无需额外授权。确保在Ubuntu操作系统（20.04或更高版本）上运行稳定；自主研发平台,提供国家颁发的软著，须和服务器同品牌，需要提供证明材料/截图

7.2. 提供统一的中文web管理界面，支持web界面中英文切换, 同时提供命令行客户端连接集群进行操作，客户可通过pip等安装命令更新深度学习组件,系统所有功能均对外提供Restful API，方便用户进行二次开发；

7.3.平台提供集群管理员、分区管理员、项目管理员、普通用户多级多角色管理；不同的角色具有不同的访问权限, 系统对不同的角色有不同的访问授权，只有特定角色的用户才能访问系统特定的功能另外系统提供安全审计员角色，安全审计员可以对系统操作记录进行审计，需要提供证明材料/截图

7.4. 平台提供公共镜像、分区镜像、项目镜像、私有镜像多层级镜像管理,允许用户之间分享镜像,同时提供云端下载功能,如Nvidia NGC/Docker hub等,此外允许用户上传镜像包, 平台解压并加载镜像到平台镜像仓库,平台允许将运行中容器保存为镜像；

7.5. 平台支持对资源进行横向扩展，无缝添加新的节点, 支持对平台资源进行逻辑分区，每个分区设置CPU、Mem和GPU使用配额，在分区、项目和个人配额中，可以针对不同GPU类型设置不用的GPU配额, 用户在Web GUI中查看节点资源配置，查看节点CPU、GPU、内存、网络、硬盘等使用情况和历史曲线；

7.6. 为防止资源浪费,管理员可以开启交互式任务运行时长限制，当交互式任务运行时间超过限制，平台删除任务并释放资源,平台支持提供Jupyter lab和jupyter notebook的集成，支持AI相关数据集成到Jupyter中；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；

7.7. 平台支持第三方存储, 例如gluster、lustre等, 分布式存储支持IB高速网络和RDMA，数据读写效率高,支持NFS数据支持本地数据缓存，提高IO性能,用户可以在Web GUI上进行数据上传、下载、删除、压缩、解压、复制、移动和内容浏览, 用户可以通过ftp进行数据的上传、下载等操作；

7.8. 支持GPU按照份数切分，单张GPU可按照1/2/4/8份数进行切分，实现GPU高利用率；需要提供证明材料/截图

**五、采购标的需满足的服务标准、期限、效率等要求**

1. 质保期： ≥ 3 年，质保期内免费维保≥2次/年，免人工服务费。质保期满后，仍需提供专业维修服务，投标人在投标文件中需注明维修服务单项报价。
2. 服务响应时间：接到维修电话后4小时内给予明确答复，8小时内到达现场维修。维修人员到现场后若问题特殊无法现场修复的，供货方需在24小时内给出合理解决方案。
3. 培训要求：提供培训电子资料及视频；供方免费为用户培训至少 1 名操作人员进行为期至少3天的现场操作培训以及应用培训，保证用户掌握有关设备的使用、维护、管理和应用等工作要求。不定期的免费提供相关设备应用方面的技术咨询等。
4. 驻场：提供原厂工程师至少三年驻场服务。

**六、采购标的的履约验收标准**

|  |  |  |  |
| --- | --- | --- | --- |
| 现场的检验指标及方法 | | | |
| 序号 | 功能或指标 | 验收或测试方法 | |
| **项目建设单位验收要求：** | | | |
| 1 | 货物外包装与外观无损伤 | 现场核查 | |
| 2 | 货物配置、包括备品备件、耗品耗材等提供齐全，货物实物品牌、规格、型号、配置数量与采购结果、合同约定相符。 | 依据《合同》及其附件（包括但不限于《采购需求》《供应商投标（响应）文件》《投标澄清函》《技术协议》等）约定，现场核查。 | |
| 3 | 所有功能和指标参数（包括边界极限值）达到采购结果合同约定要求。 | 依据《合同》及其附件（包括但不限于《采购需求》《供应商投标（响应）文件》《投标澄清函》《技术协议》等）约定，现场测试，供应商应提供《产品出厂检测报告》《产品合格证书》和根据合同约定提供《第三方检测报告》。 | |
| 4 | 提供《培训视频》影像资料 | 现场核查 | |
| 5 | 验证测试设备的运行稳定性 | 试运行验证测试设备运行稳定达标 | |
| 6 | 《供应商货物类项目完工报告》《项目建设单位货物类项目完工自验收报告》《项目建设单位货物类项目完工自验收报告》《第三方检测报告》等与验收相关的材料由项目建设单位妥善保管存档。 | | |
| **学校验收复核要求：** | | | |
| 1 | 项目建设单位填写《学校采购货物类项目验收复核申请表》 | | |
| 2 | 提供《供应商货物类项目完工报告》 | | |
| 3 | 提供《项目建设单位货物类项目完工自验收报告》 | | |
| 4 | 学校组织验收专家组现场复核供应商与项目建设单位货物到货完工验收完成情况 | | |
| 验收时是否需要供应商提供样品 | | 是□ | 否☑ |
| 验收时是否需供应商提供必要的其他设备 | | 是□ | 否☑ |
| 除现场验收外，需提供的其他验收要求 | | | |
| 除现场验收外，是□否☑需提供第三方检测报告 | | 对于检测机构的要求：国家正规检测机构，出具的检测报告由验收复核专家认可之后作为验收复核通过的主要依据。  对于检测执行标准的要求：各项检测项目标准以检测机构按照行业相关要求最新适用并执行的标准为准。 | |