采购招标项目参数要求

|  |  |  |  |
| --- | --- | --- | --- |
| 项目名称 | 人工智能超级计算平台 | 采购编号 | CGHW20190702-193-1 |
| 供货时间 | 2019.10.15 | 供货地点 | 吉林大学珠海学院网络中心 |
| 售后服务要求 | 按照合同书要求保障3年售后服务 | 安装调试要求 | 1.要求供应商在供货之前，安排技术人员对我校安装设备的机房进行实地考察并提出相关整改意见。2.要求供应商安排专业安装调试人员负责设备的安装调试以及设备验收工作。 |
| 项目概述：目前深度学习（DEEP LEARNING）是机器学习研究中的一个新的领域，其动机在建立、模拟人脑进行分析学习的神经网络，它模拟人脑的机制来解释数据，包括图像、声纹、文本、视频、数据等；深度学习是一个快速发展的人工智能领域，推动着众多计算机创新领域发展的引擎，涉及领域从高级医学和药物研究到自主控制、无人驾驶汽车等；深度学习也是当今世界最尖端的技术挑战之一。但深度学习也需要处理大量的并行计算。申请该大型计算系统主要服务于视频人工智能研究计算平台。此平台可为人工智能技术、类脑计算、大规模视频处理、先进视频编解码和高级计算机视觉机理提供高性能高并行性硬件支持。与所发展的算法相结合，并针对当前和未来的人工视觉智能与类脑计算的特点，发展具备高性能高通量的类人视觉和高级人工智能并行运算软硬件平台。该设备是目前用于深度学习并行计算的性价比最高的优质设备，有独立的专利权，短期没有其他设备可以替代，并且不会很快过时。该设备不仅可以用于科学研究，也可以用于本科生人工智能课程以及研究生部分课程的教学设备。该设备不仅用于计算机技术类的科学研究，也应用于大数据分析等研究。 |
| **重要技术指标（必填）** |
| 序号 | 指标名称 | 参数明细 |
| 1 | 设备名称 | 搭载Tesla V100 32GB的NVIDIA DGX-1系统 |
| 2 | 平台外型 | 3U机架式 |
| 3 | 处理器 | 2颗Intel Xeon E5-2698 V4 处理器 |
| 4 | 内存容量 | 512G RDIMM DDR4 |
| 5 | 硬盘 | 4块1.92TB SSD RAID0 高速企业级硬盘 |
| 6 | 接口及核心组件 | 8块Tesla V100 32G SXM接口，NVLINK架构，单卡配置为5120个CUDA核心，640个Tensor核心,7.8 FP64 TFLOPS|15.7 FP32 TFLOPS |125 Tensor TFlOPS20MB SM RF |16MB Cache32GB HBM2@900GB/s|300GB/s NVLink |
| 7 | 网络接口 | 双口10G网路 RJ45接口 和 4个100G IB接口 |
| 8 | 电源 | 1600W 3+1冗余电源 |
| 9 | 人工智能深度学习软件 | 提供深度学习框架应用软件images镜像caffe2, caffe , tensorflow, nvidia digits, mxnet, theano, cntk, pytorch, torch。这些深度学习框架以容器镜像的方式提供，需支持定期更新，可在云平台下载。 |
| 10 | 提供HPC软件 | 提供经优化过的HPC软件，包含gamess，gromacs，lammps，namd，relion，这些软件镜像需支持定期更新，可在云平台下载 |
| 11 | 集成式系统 | 提供第三方合作伙伴软件的容器镜像,包含chainer，h2oai-driveless，Mapd， paddlepaddle，这些软件镜像需支持定期更新。支持快速创建多种深度学习开发调试环境的容器，提供基于SSH的容器访问接口。支持容器的克隆、删除等操作。支持将创建的容器在线进行镜像打包，并支持将打包好的镜像上传镜像仓库，实现镜像版本的持续更新。 |
| 12 | 异构资源管理调度平台作业软件 | 提供动态资源调度管理功能，全面支持CPU、GPU等异构资源的混合调度，提供FIFO、作业优先级、资源配额等调度策略，支持节点配置、功能定位等原则对平台计算资源进行分组，调度器提供基于资源分区的训练任务调度策略，支持多用户登录。 |
| **一般技术指标（选填，不作为评标依据）** |
| 序号 | 指标名称 | 参数明细 |
| 1 |  |  |
| 2 |  |  |
| 3 |  |  |
| 4 |  |  |
| 申报人 |  | 单位负责人 |  |
| 实验室与资产管理处 |  |

**注： 1、参数不可与已获批学年采购预算有冲突，如有冲突以已批准采购文件为准2、不得含有排他性技术指标3、条目可根据具体情况增减4、本页不够可另起一页。**