**参数指标**

**1．** CPU核心数量：96核心；

**2．** CPU双精度浮点计算能力：6.45Tflops(即6.45万亿次/秒)；

**3．** GPU数量：12块RTX 4090；

**4．** GPU单精度浮点计算能力：876Tflops(即876万亿次/秒)；

**5．** 系统内存：1024GB；

**6．** 系统存储：64TB；

**7．** 网络带宽：10Gbps；

**8．** 系统总功耗：8000瓦。

|  |  |  |
| --- | --- | --- |
| **部件** | **技术规格** | **数量** |
| GPU节点 | 机箱：静音塔式机箱  CPU：1颗24核心 2.1GHz Intel Xeon Gold 5318Y  内存：8根32GB DDR4 3200MHz ECC REG  系统盘：1块960GB SATA SSD企业级固态硬盘  缓存盘：1块7.68TB U.2 NVMe SSD企业级固态硬盘  数据盘：1块16TB 256MB 7200转 SATA 企业级硬盘  GPU卡：3块NVIDIA Geforce RTX 4090一体式液冷版  网卡：1块双口万兆网卡 | 4 |
| 交换机 | 48口千兆4口万兆交换机，配套线缆 | 1 |
| 机柜系统 | 符合国家工业标准的42U、19寸专业服务器机柜  包括独立的布线系统、散热系统和供电系统  尺寸为2055mm\*600mm\*1000mm  承重要求800kg或以上 | 1 |
| 操作系统 | Linux 64位 企业版 | 1 |
| 集群管理软件 | 集群管理软件提供节点管理、用户管理、任务管理、状态监控等多种功能，任务管理不仅支持CPU调度，还支持GPU调度。网络方面不仅支持普通以太网的通信管理，还支持高带宽、低延迟的Infiniband网络。在集群状态监控上，系统提供一整套基于WEB的组件用于监控自身的活动及配置状况，如处理器、内存及交换分区使用情况、网络负载、节点状态和当前计算任务的进展等，内置故障报警邮件通知。系统为每个用户提供一个接口，用户通过这个接口提交自己的计算任务，计算完成后会反馈结果给用户，用户只用指定需要的资源，如CPU核心数、GPU数，而不用关心程序是在哪个节点，使用哪个CPU、GPU完成的。用户还可以对自己的队列进行管理，如：提交任务后可以查看自己任务的工作状态，改变多个任务的执行顺序，暂停或删除已提交任务。 | 1 |
| 应用软件 | 提供TensorFlow, Pytorch，Matlab，Mathematica等软件安装服务和技术支持（用户提供安装包）。 | 1 |
| 开发环境 | 集成OpenMPI, MPICH并行计算环境；  集成GNU C/C++编译器，GNU Fortran编译器；  集成LAPACK, BLAS, Atlas, FFTW等高性能计算函数库；  针对CUDA开发还将提供cuBlas, cuFFT, cuSparse, cuRand, npp, OpenCL等高性能计算函数库；  集成CUDA驱动、CUDA工具包、CUDA Samples。 | 1 |
| 安装调试 | 原厂工程师上门安装调试高性能计算集群，提供一些常见计算软件的安装和调试服务（如TensorFlow, Pytorch，Matlab，Mathematica等），并且提供3年的免费维护及相关问题的解决。 | 1 |
| 培训 | 提供完备的集群使用说明书；  原厂工程师上门提供集群系统使用培训，包括系统管理员培训和普通用户培训，培训方式为集中会议式培训和上机培训相结合。 | 1 |