高性能GPU显卡技术要求：

数量：2个

具体要求如下：

|  |  |
| --- | --- |
| 参数 | 技术规格要求 |
| 架构 | Volta |
| Tensor核心数量 | 640 |
| CUDA核心数量 | 5120 |
| 双精度浮点运算能力\* | 7TFLOPS |
| 单精度浮点运算能力\* | 14TFLOPS |
| Tensor性能 | 112 TFLOPS |
| GPU内存\* | 16GB HBM2 |
| 显存带宽 | 900GB/秒 |
| 系统接口 | Pcle Gen3 |
| 最大功耗 | 250W |
| 软件\* | 1. 可支持GPU的资源被分割（小到单片GPU的1/2），然后在运行环境中分配给用户。 2. 可以实现将多个系统聚合成一个独立且具备灵活性的系统，实现计算资源的扩展和管理，大幅度优化现有GPU解决方案，可将训练速度提升2-4倍，在同一网络条件下，提供多机多节点的计算资源（从单片到多片GPU）实时动态调配。 3. 可在共享存储区域创建私有文件区域，保证数据安全。可实现共享存储区域数据给不同workspace区域共同使用，避免重复数据上传。 4. 提供基于容器的管理，一键部署深度学习框架例如Caffe，Tensorflow。 5. 非开源软件或开源框架二次开发，需要提供原厂授权和指导，提供一定时间的试用时间和技术支持。 |
| 服务\* | 1. 3年产品原厂质保，具备远程机器检测能力，原厂400/800技术支持，可通过网络、电话、邮件等方式提供软硬件技术支持。 2. 提供安装服务及实施主机应用及数据库实施迁移，要求零时间切换，平滑迁移。 3. 原厂原包装，竞价结束后2天内到货，免费送货到指定地点并安装部署。 |