**AI医疗辅助大模型服务器需求**

为构建北流市人民医院AI医疗辅助大模型，现拟采购满足同时运行QwQ-32B、 Qwen2.5-Vision-Large-7B-Instruct和BAAI/bgereranker-large模型所需的GPU、 CPU、内存、存储等资源的服务器硬件（品牌机），核心配置参数如下：

|  |  |
| --- | --- |
| **组件** | **配置** |
| CPU | >2个处理器，单核>16，频率>3.0GHz， 总计48核 |
| 内存 | >256GB DDR5 4800MHz， 24插槽 |
| 硬盘 | 960GB M.2+2x1.92TB SSD，可扩展 |
| GPU | 4块，INT8性能>2400 TOPS，支持8块扩展 |
| 网络 | 4千兆+双万兆网卡，虚拟化6万并发 |
| 电源 | 4个1600W热插拔铂金，冗余设计 |
| 虚拟化 | 支持，集群3机， 6万并发 |