

# 运维期末作业——作业指导书

---

## 作业目的

---

经过了一个学期在运维部的学习，部员们收获满满，现在是时候检验你的学习成果了。该作业旨在培养部员的知识检索能力、自主学习能力以及计算机核心素养。为下一届网协运维部部长选拔培养和选拔人才。该作业将作为选拔**运维下一届部长**的重要参考

## 作业指导

---

在摩尔定律逐渐失效的当下，传统的冯诺依曼架构逐渐不能适应 AI 等密集计算任务的需求了，各大公司相继推出特定领域计算芯片。计算集群或许将逐步进入异构计算时代。

针对 AI 所需要的巨大的数据带宽，英伟达公司开发了 NVlink 用于 GPU 之间的直接通信，迈络思公司开发了 RDMA 以绕开 CPU 实现各机柜之间的数据交换。Ktransformers 将稀疏矩阵计算放到 CPU 上计算，稠密矩阵放到 GPU 上计算，从而大幅降低显存需求。

你可以以异构计算在 AI 推理领域的应用和其原理为核心撰写报告。

AI 可以写代码，但不能运行代码。有时候，AI 生成的代码你运行会报错，这时候你必须手动将错误信息发给 AI，让 AI 帮你改错。如果 AI 能自己写代码，自主运行代码并查看运行效果就好了。或者你可以将 AI 接入 Blender 帮你建模？或者让 AI 自动帮你查资料？让 AI 操作数据库？

你可以从用户感知较强的人机交互角度出发，以 AI Agent 智能体为关键词，参考 MCP、OpenManus和RAG 相关资料，以 AI 交互为核心撰写报告。

价格太高？全量Deepseek R1可以用 CPU+内存 运行，只不过效率不是很理想，但是不可否认的是价格很低。

你可以从极致的成本角度出发，在2000元左右配置一台能运行全量Deepseek R1的服务器。1token/s 能玩，2token/s 流畅，3token/s 电竞。