

1.调研和说明并行计算、分布式计算、云计算的异同

并行计算：并行计算是相对于串行计算来说的。可分为时间上的并行和空间上的并行。时间上的并行就是指流水线技术，而空间上的并行则是指用多个处理器并发的执行计算。并行计算的目的是提供单处理器无法提供的性能（处理器能力或存储器），使用多处理器求解单个问题。

分布式计算：分布式计算研究如何把一个需要非常巨大的计算能力才能解决的问题分成许多小的部分，然后把这些部分分配给许多计算机进行处理，最后把这些计算结果综合起来得到最终的结果。最近的分布式计算项目已经被用于使用世界各地成千上万位志愿者的计算机的闲置计算能力，通过因特网，可以分析来自外太空的电讯号，寻找隐蔽的黑洞，并探索可能存在的外星智慧生命等。

云计算：云计算是最新开始的新概念，它不只是计算等计算机概念，还有运营服务等概念了。它是分布式计算、并行计算和网格计算的发展，或者说是这些概念的商业实现。云计算不但包括分布式计算还包括分布式存储和分布式缓存。分布式存储又包括分布式文件存储和分布式数据存储。

相似之处：

1. **处理大规模任务：**并行计算、分布式计算和云计算都旨在处理大规模的任务或问题，通过同时利用多个计算资源来提高计算效率和性能。
2. **利用多个计算资源：**这三种计算模式都利用了多个计算资源（如处理器、计算节点等）来完成计算任务，从而加速处理速度或增加处理容量。
3. **分解问题：**它们都涉及将大问题分解成小问题，然后分配给多个计算资源进行处理，最终将结果合并得到最终结果。

相异之处：

1. 并行计算：

- **同步性高：**并行计算通常是在同一台计算机上的多个处理器或核心上执行任务，并且这些处理器通常是紧密协调和同步的。
- **局限性：**并行计算的主要局限在于计算资源的有限性，受限于单个计算机的硬件资源，无法无限扩展。

2. 分布式计算：

- **分布性：**分布式计算涉及到多台计算机或计算节点之间的协作，这些计算资源可以位于不同的地理位置，并且通常是异步执行任务的。
- **灵活性：**分布式计算具有较高的灵活性和可伸缩性，可以根据需求动态添加或移除计算节点。

3. 云计算：

- **服务化：**云计算不仅仅是一种计算模式，还有一种服务化的概念，提供了通过网络按需获取计算资源的能力，通常以服务的形式提供，如IaaS、PaaS和SaaS等。
- **资源池化：**云计算通过资源池化的方式，将大量的计算资源整合在一起，并提供给用户按需使用，用户无需关心底层硬件或软件的细节。

2.课本第35页1.11题。并行计算应用调研

(1)计算机学科中transformer模型的并行计算，针对大模型中Encoder结构和Decoder结构可以进行并行化计算

Transformer模型有利于并行计算的主要原因是其自注意力机制。自注意力机制可以独立地处理每个输入序列中的元素，因此允许在计算时充分利用现代GPU的并行计算能力。与此相比，传统的循环神经网络（RNN）需要顺序地处理序列数据，这在计算过程中限制了并行性。

(2)应用效果我没有找到相应的实验

(3)共享内存计算平台，cuda

(4)没有找到相应的实验

(5)性能瓶颈:

- **内存限制**: Transformer模型的内存占用可能会成为性能瓶颈,特别是对于较大的模型和批次大小。在GPU上运行大型模型时,内存可能会成为限制因素,因为GPU的内存容量有限。这可能需要对模型进行分布式训练或者使用更大的GPU内存。
- **通信开销**: 在多GPU环境下, GPU之间的通信可能成为性能瓶颈。在分布式训练中,需要在GPU之间传输梯度和模型参数,这会引入通信开销。减少通信开销可以通过优化模型分布、减少参数同步频率等方式来实现。
- **计算不平衡**: 在某些情况下,模型中不同部分的计算负载可能不平衡,导致某些GPU的利用率较低,从而影响可扩展性。优化模型的计算图、调整批次大小等方法可以帮助解决这个问题。

3.根据调研的应用需求,按预算1000万人民币购置计算设备,请给出采购清单(组件型号、数量、单价)

型号: NVIDIA A100 PCIe显卡

单价: 78899元/台

数量: 126台