



人工智能领域高性能计算 解决方案

人工智能是一种使用机器学习和深度学习等算法来模拟人类智能和实现自主决策的方法，通常用于自然语言处理、图像识别、智能机器人和智能交通等领域。

算力、算法、数据被称为人工智能发展的三大基本要素，随着人工智能（尤其是深度学习）模型和算法的发展，大型复杂模型、海量数据所需要的算力持续增长，高性能计算为机器学习提供了大规模并行、分布计算的基础设施。

从应用层面上来看，传统高性能计算应用和机器学习的方法相互借鉴，比如深度学习已被广泛使用在传统科学与工程计算中（如气候与天气研究），高性能计算中可扩展的优化方法也被用于大规模深度学习。人工智能对用 HPC 集群支撑深度学习任务，促进 AI 与 HPC 的融合提出了更高的要求。

■ 人工智能领域中HPC的重要性

1. 高性能计算为机器学习提供大规模并行、分布计算的基础设施

考虑到 AI 研究离不开庞大的数据、精确的数据标记和数十倍于一般科研项目的计算量，深度学习需要HPC听过较高的计算资源和计算能力。

2. 高性能计算和人工智能相互促进、相互融合

高性能计算和人工智能有着天然的联系，高性能计算不仅可以加速人工智能的快速发展，同时随着人工智能的不断成熟，又反过来促进高性能计算系统更好的部署和应用。

3. 高性能计算有助于商品化人工智能/机器学习的性能提升

高性能计算支持更多的硬件多样性、交互性支持以及针对性能优化的额外库/框架抽象，可以支持人工智能工作负载。

■ 人工智能领域建设高性能计算集群的难点

1. 计算能力不足

随着人工智能训练精度与规模要求的不断提升，人工智能基础设施正承担越来越高的负载水平，需要性能更加强大、算力充足、敏捷扩展的HPC。

2. 资源分散分布

人工智能用户操作只位于自己的工作站及存储之间，数据孤岛明显，计算资源相对分散，无法形成统一管理与统一分配。

3. 缺乏统一管理

人工智能应用业务流复杂，缺乏灵活有效的统一管理平台 and 全局统筹工具，没有统一的调度流程，不同任务无法形成统一业务流，响应缓慢。

■ 联智科技基础科研领域HPC解决方案概述

为了进一步提升高性能计算的性能表现，为人工智能等应用提供计算能力的支撑，联智科技推出高性能与人工智能融合解决方案。该方案通过 Web 界面融合 HPC 和 AI 两大应用，统一调度不同的任务到通用 CPU 集群及 CPU/GPU异构集群上计算，支持CPU/GPU混合调度，支持GPU单卡、多卡共享与独占，突破AI应用对算力要求高的局限。通过高效的分布式训练和推理，在基本不改变 HPC 用户使用习惯的情况下，解决了 AI 计算在 CPU 上的性能瓶颈，实现了统一高效的 HPC/AI 融合，为图像分类、语音识别、语言翻译和物体检测等深度学习应用案例提供更高效率的训练和推理加速。

■ 联智科技人工智能领域HPC解决方案价值

1. 联智科技提供HPC/AI集群建设全流程解决方案

提供一整套异构的、满足深度学习应用的HPC/AI整体解决方案，高效且经济地提供训练和推理加速，实现作业与训练任务的流程统一调度。

2. 解决深度学习神经网络更大规模、更复杂的问题

极高的计算性能、巨大的内存池和优化的节点间通信网络极大地扩展深度学习网络识别相关结构和模式的能力。

3. 降低传统 HPC 用户使用 AI 的门槛

集成AI 主流框架和Horovod 分布式系统，通过统一的集群资源调度平台，轻松实现基于常见主流框架（Tensorflow/Caffe）的分布式训练和推理，并在多节点分布式训练实例上取得极佳加速比。