诺亚鸿云

Hybrid-Computing Platform

致力于交付一套整体的人工智能基础建设方案
CPU通算＋GPU智算一体机、构建下一代智算私有云节点

方案优势

- **自主可控**
  从CPU到GPU全国产化体系结构，规避贸易摩擦，地缘政治等不稳定因素，构建持续的智算发展战略
- **独树一帜的设计体系**
  采用创新型的两段式设计体系，既：通算部分仍然使用高性能国产处理器体系结构，智算部分则采用自研的独立GPU基板。通过2颗GPU加速模组，促使GPUs在系统内部高速互联，以此来规避PCIe带宽瓶颈
- **开放式 & 易用性**
  始终遵循开放式的体系结构设计，可根据预算，喜好选择广泛的GPU品牌。通过自研的AIOS可在出厂便预制了GPU驱动程序，算子与框架适配服务，甚至包含友好的GPU调度与管理的统一可视化平台。
  解决国产GPU因为易用性推广的最后一公里
- **All in One 交付能力**
  HCP-48混合算力系统具备构建完整人工智能应用所需的配套环境。
  为复杂的AI应用程提供性能强大的虚拟化，容器等环境。而以GPU为核心的智算池也耦合在同一系统。
  支持混合算力的分布式部署，通过横向扩展，具备构建高可用性的大型GPU智算组网的能力

在当今人工智能产业快速的变革中，企业需要迅速的做出反应，通过拥抱人工智能带来的关键优势，降低运营成本，提高生产力，以此来提升企业在经济市场的竞争力。
然而，以通用算力构建的私有云向人工智能的战略性转型总是面临着各种问题，例如现有资产的利用率、设计人工智能基础设施的复杂度、如何确保国产化的自主可控、引入国产GPU与模型和框架的适配、以及智算架构的可扩展性为了应对企业不断发展的业务需求等等。

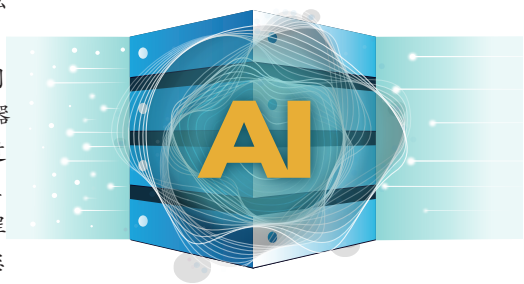
构建自主可控的人工智能基础设施
利用诺亚鸿云打造的HCP混合算力系统

Hybrid-Computing Platform 致力于打造一套完整的人工智能运行环境的解决方案，通过创新型混合算力的体系结构，降低企业迈入人工智能的部署成本。通算与智算两段化的系统设计标准，在安全方面，通算体系结构全部采用国产化自主可控设计，满足日益严苛的信息安全政策。而智算方面则适配了主流的国产GPU厂商，做到了从通用算力到智算的全国产化方案落地。

HCP方案突出高性能和无与伦比的可扩展性，这体现在：

- 我们的科研人员通过先进的电子电路设计，重新塑造了GPU的互联方式。
- 通过两段式的设计，重新设计高带宽的GPU基板，规避当前PCIe通道在智算产业的带宽瓶颈。
- 内置GPU加速模组，智算整机可通过模组直接 P2P互联，无需通过处理器与PCIe的中转。
- 通用基础算力配置自研的IPU加速卡，可有效的卸载国产化处理器与内存对于云组件的处理，如：云管平台、存储系统、磁盘IO、以及磁盘纠删 / 条带等、带来的性能开销。

HCP-48混合算力平台具备从GPU到CPU全国产化体系的整体交付方案。
企业可以通过本地化部署，即刻拥有集成私有云的人工智能完整基础设施。

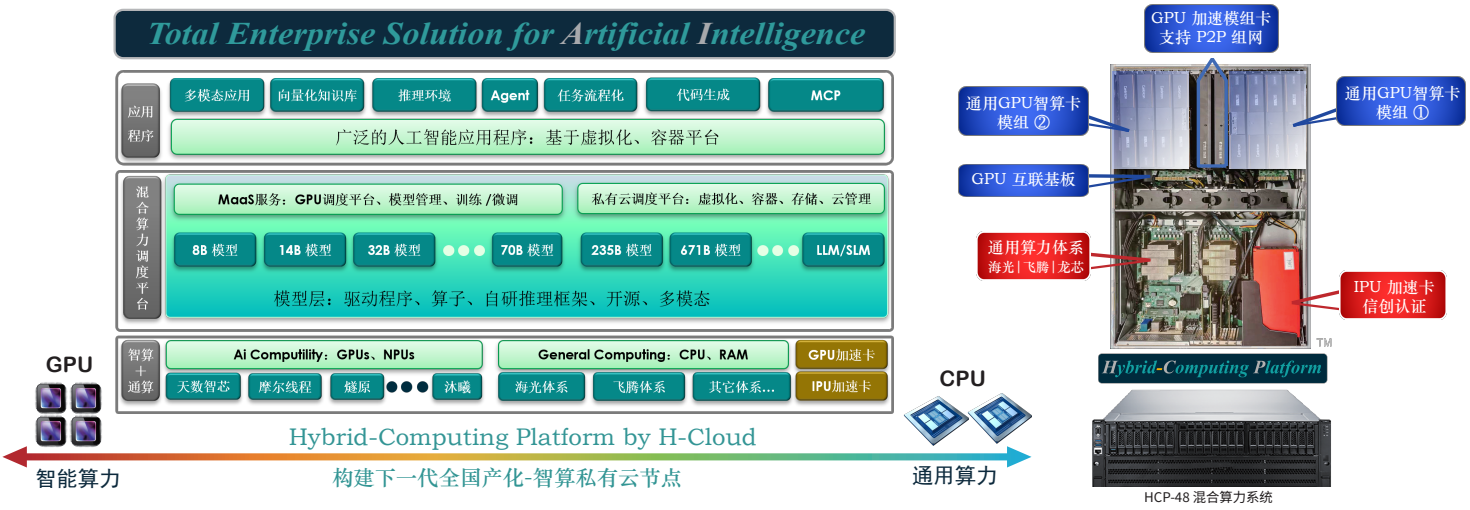


考虑到人工智能的落地需要配合大量的应用程序，HCP-48系统提供标准的虚拟化、容器环境，使用户无需为了兼容性而担忧。通过云管平台企业可以尽情的部署与Ai相关的应用程序，而分布式集群的能力可以使应用程序在额外的HCP节点进行故障转移，就像您在其它场景见到的私有云解决方案一样。
在私有云的建设中很多用户对“国产化处理器体系”诸多顾虑，是否能在融合架构中发挥更大性能与兼容性。我们的科学家为HCP通算层开发了具备自主可控的IPU加速中间件，通过有效的卸载与云组件相关的堆栈，促使处理器与内存等通用算力可以更加专注的服务应用程序。
IPU加速卡通过创新型的FPGA+ASIC嵌入式设计，预制云管平台、分布式存储系统、磁盘与IO的高级特性处理堆栈、以及一颗具备数百GB的RAID 分区芯片、服务于磁盘的High Caching。
IPU加速卡还具备自组网特性，使存储相关的流量和服务在跨节点交互时不经由额外的以太网适配器。综上，HPC的混合算力系统通过一系列先进的体系结构设计，可有效的加速依赖于人工智能的应用程序，包含与之相关的存储服务：向量化知识库、语料数据、模型载入与数据库。

Ai™
Hybrid-Computing Platform

HCP-48是诺亚鸿云基于创新理念“Hybrid-Computing Platform”打造企业级私有云智算节点。
通过HCP系统，耦合了以处理器为核心的通用算力来构建私有云平台，还包含了以GPU为核心的高性能智算体系，
符合全国产化设计，让智算基础设施高效运行的同时更加自主可控。

- 高性能--是在HCP体系结构设计之初一个重要的考量。我们工程师通过创新型的电子电路设计，开发了GPU的互联基板，通过交换芯片使GPU之间的互联带宽到达500Gbps。GPU加速模组卡则是把多个厂商GPU智算卡异构整合的又一个先进创新，还具备整机互联的直出端口，既:通过GPU模组卡400Gbps整机组网互联。



- 人工智能高度依赖的应用程序可运行在由HCP-48系统集成的云环境，通过虚拟化或容器环境屏蔽不相关性，而唯一被关注的则是国产化服务器体系带来的性能担忧。
- 我们工程师通过自研的“硬件中间件”方法，高效的卸载云组件相关的堆栈，让国产处理器与内存能够火力全开的服务应用程序，而在广泛的云组件中，存储与磁盘IO处理相关的的性能开销则是首当其冲。IPU加速卡具备完整的存储与磁盘IO处理堆栈，同时集成了通用的存储分布式技术，甚至还保留了10Gb/25Gbps存储组网端口，能够支持大规模的横向扩展需求。
- 系统结构的开放性和灵活性是我们一直追求的目标。HCP具备高度的灵活性，这不仅体现在您可以自由选择第三方的通用GPU算力卡。还可根据需求独立的扩展通用算力。GPU算力同样可以独立扩充，无需出厂既满配，可以根据应用需求、预算、喜好随需扩容。
- 易用性与维护性也是完整方案的一部分，HCP系统出厂既预装了主流国产厂商的GPU驱动程序，根据需求适配了广泛的框架和算子，甚至贴心的为用户下载大模型，这一切被高度集成在AIOS系统中。提供一套友好的可视化WebUI，随时供用户掌控集群内部的GPU调度、维护、监测、以及模型下载后的进行测试的工具。

方案效益：

•开放式 & 易用性

支持广泛的国产GPU智算卡与英伟达
自研的AIOS具备丰富的GPU可视化调度工具

•异构融合：

同一个GPU服务器内部可以支持不同厂家的国产GPU卡，
实现单服务器异构算力融合。

•提供一站式交付平台

提供国产 + 英伟达 等主流算力，用户选择更自由。

开“箱”即用，涵盖AI开发全流程，包含数据集、模型开发、训练、管理、部署功能，可灵活使用其中一个或多个功能。

企业从人工智能入门到扩展大型算力需求，横向扩展能力始终伴随人工智能业务的发展。

H-Cloud
Your Ai, Our Priority

探索诺亚鸿云更多的创新 Hybrid-Computing Platform

深圳市诺亚鸿云信息技术有限公司，成立于2017年7月，专注于超融合、桌面云、分布式存储、数据加速单元、信创AI工作站与模型一体机及企业数字化转型的产品集成、研发、咨询及维保业务。致力于提供最优质专业的IT解决方案，将产品、服务和用户体验做到极致。经过近10年发展，形成了诺亚鸿云性价比独一无二的产品，如超融合/双活/分布式存储/桌面云一体机、信创AI工作站、智算服务器、自主产权HCP混合智算系统等心竞争力产品与解决方案，获得大量用户认可！

Ai™ Hybrid-Computing Platform