

AI 加速新材料研发 解决方案

AI accelerates new material research
and development solutions



场景分析

Scene Analysis

新材料作为新兴支柱产业，在新能源材料、生物医药、绿色建筑等领域有着举足轻重的作用，新材料应用有利于产业链结构调整和能力提升。新材料从研发到产业化周期十分漫长，研发周期缩短一半、研发成本降低一半是我国科学家们努力的目标。

- “材料基因工程”是当前材料领域公认的开发新材料的颠覆性前沿技术，通过计算技术、大数据技术和高通量自动化实验等关键技术，大幅提升新材料的研发效率，缩短研发周期，降低研发成本，促进工程化应用。
- AI4S正在成为科学研究的一种新范式，在材料领域也得到了广泛应用。通过深度学习算法构建模型，借助算力训练模型来预测材料性能、发现结构异化晶体等，进而达到发现新材料的目的，以深度学习为代表的人工智能技术改变了传统材料研究新范式。

在材料科学领域，高通量计算和深度学习已成为推动新材料发现与设计的关键技术。这些技术的核心在于能够获得、处理并分析海量的数据等。尽管这些技术极大地加速了材料研发过程，但同时也对算力资源提出了巨大的需求。

■ 材料数据生成、高通量计算算力需求大

材料数据生成是深度学习的首要步骤，涉及读取大量的输入文件等，随后执行计算以得到材料的物性数据。比如，仅仅是在电子云密度分布数据的生成上，构建具有简单立方对称性的10000种材料数据集，就需要超过10的七次方核时，动辄需要数年时间来完成。这一过程中，算力资源的短缺成为限制高通量材料计算任务的一个主要瓶颈。

随着高通量计算算力需求的增大，传统的物理实验和计算方法逐渐显得力不从心。以耐蚀材料的研发为例，传统方法需要进行成百上千次实验才能获得材料最佳的比例和浓度，而高通量计算方法虽然能够加速这一过程，但其本身的算力需求极大，尤其是在实现跨尺度的研究上，需求更是高得惊人。

■ 模型训练需要巨大的算力资源

模型训练作为深度学习在材料科学中的应用之一，尤其是卷积神经网络（CNN）的使用，能够构建覆盖原子特征、化学键特征和分子特征等多尺度的模型，但其逻辑复杂、参数量大，完成一次模型训练依然需要数百至数千张GPU卡协同工作。在单个GPU上，一个大型CN模型的训练可能需要耗费数周时间，而在一套计算集群系统上则可能仅需数天或数小时。

材料的研发，十分关注在现实场景中服役的表现。用于设计的材料完成筛选后，通过纳入成分-工艺-性能-环境等综合因素，对其进行仿真，获得材料在服役环境中的表现，完成仿真模拟与物理实验双重验证。材料仿真过程需要借助智算模型，缩短研发周期，以提升研发效率。

■ 计算平台建设周期长

“一代材料、一代技术、一代产业”，重大工程与装备“等米下锅”现象较突出，加快基础设施的建设，也是重中之重。计算平台建设面临周期长、缺乏定制化、以及运维手段过于单一等问题。据相关数据表明，传统数据中心建设平均周期为200天，业务上线90天，与客户要求数周业务上线时间相差甚远。



解决方案

Usage
Scenario

浪潮信息针对新材料研发领域存在的以上问题，结合材料基因工程研究所需的高通量计算和新材料预测模型开发的算力需求，提供一站式计算集群整体解决方案，全面协助客户完成材料基因计算集群快速建设和部署。













应用软件	 Shell连接		 远程桌面		 作业提交		 交互式作用		 文件管理		 资源申请	
	材料基因 ProCast	材料基因 ThermoCalc	材料表征 Pendant	材料评价 Quantum,ESPRESSO	分子动力学 LAMMPS, Amber	量子化学 GAUSSIAN, GAMESS	计算流体力学 Flunt	人工智能 Tensorflow, PyTorch				
平台管理	作业调度			集群管理				AI开发平台				
	· 作业管理 · 作业监控		· 应用管理 · 账单管理		· 资源监控 · 系统管理		· 告警管理 · 消费统计		· 数据管理 · 模型开发		· 统一管理 · 全流程支持	
硬件集群	计算节点					高速网络			并行存储			
	 NF5488 GPU加速		 i24 液冷		 TS860 大内存计算		 HDR 100 40Gbps IB 组网			 AS13000-H GPFS文件系统		
基础环境	模块化机房					 <ul style="list-style-type: none">· 机柜12台，整体功耗≤80.96KW· 封闭通道1套、精密配电柜1台、不间断电源1套、列间空调5台、配电站基站空调1台、机房环境动力监控系统1套						

图1：材料基因计算集群解决方案

在硬件集群层，为满足材料预测模型开发需求，通过部署人工智能开发平台，调度GPU服务器集群完成模型一站式开发、训练和部署。针对高通量计算大算力要求，分别部署多节点服务器和大内存计算服务器，实现新材料成分、结构、性能分析和高效筛选。

在基础环境层，提供模块化机房，采用一体化集成设计理念，将机柜、供配电、制冷、综合布线、消防、防雷、监控等功能于一体，通过列间空调近热源直接送风，避免热岛效应，相比传统机房，微模块机房制冷效率提升12%以上。



方案价值

Program
Value

- 基于深度学习建模，实现数据驱动多尺度材料计算与预测，使“算”出新材料成为可能
 - 利用深度学习技术对材料失效建模，对3000多万条腐蚀数据筛选计算，实现低合金钢大气腐蚀实时速率与动力学演化规律的高效精准预测，预测精度从50%提高至94%。
 - 通过不同腐蚀预测模型分析及仿真测试，可以快速得到耐高温高湿、耐腐蚀，具有更长寿命的新材料，有利于解决我国每年因材料腐蚀造成的数万亿元损失。

■ 高通量计算“创造”大量数据，以强大算力基础设施，提升新材料研发效率

- 采用至强8358 CPU服务器构建数十乃至数百P算力 的高性能计算集群，使得材料电子云密度数据集生成时间，由数年压缩至数月甚至更短的时间即可获得。
- 利用材料高通量计算和大数据分析、筛选技术，在近 4×10^4 种材料中发现了8000余种拓扑材料，超出历史上发现拓扑材料数量的10倍。

■ 提供一站式“交钥匙”式服务，加快数据中心建设周期，实现集群快速上线

- 各系统采用模块化工厂预制，现场积木式快速搭建，建设周期相比传统方案加快50%以上。采用行间精密空调实现近端制冷、水平送风，较常规空调节能率超过30%。
- 通过部署智能型动力与环境集中监控系统，达到数据中心动力、环境、能耗的动态管理，具备现场触控屏运维、WEB远程运维和手机APP运维的能力，实现无人值守。



相关产品

Related
Products



■ AI服务器 NF5488A5

- 全新架构，采用PCIe 4.0和NVlink3.0互联，实现AI性能、通信带宽、显存容量全翻倍；
- 8颗高性能加速卡和2颗64核CPU，NVlink3.0全互联，PCIe链路极致优化。



■ 液冷计算节点 i24LM6

- 2U空间支持4个双路计算节点，实现业界最高密度最多支持8颗Intel Ice Lake处理器，支持液冷升频，计算性能提升5%以上；
- 采用温水冷却方式，水冷散热效率高达80%，水冷覆盖CPU、内存和VR，PUE ≤ 1.1 ，减少数据中心制冷设备部署，显著降低制冷成。



■ 八路大内存计算节点 TS860M5

- 持8颗全新一代处理器，支持高达36T内存。计算性能提升超过20%，内存带宽最大提升20%，适合大内存计算、异构计算等计算密集型场景；
- 整机RAS特性80余项，实现全模块化容错设计，超高的性价比，超强的计算性能，满足高性能计算等关键应用。



■ 人工智能平台AIStation

- 提供从开发到部署的AI全流程数据、算法和开发环境流转共享，简化开发和部署工作，支持根据模型训练要求自动匹配最佳算力分配策略；
- 在训练和推理的不同维度，允许一张卡执行多个任务，提升集群资源的整体利用率，减少重复工作，让模型自动持续迭代。

