

万卡GPU集群的能耗挑战与传统火电调频的液冷储能舱架构启示

最近和几位数据科学界的朋友聊天，他们的话题总绕不开训练大模型时，那些动辄上万张GPU卡组成的计算集群，耗电量实在有点“吓人”。这让我想起，在能源领域，我们其实早就面对过类似的、关于集中式高功率密度设施供能与散热的核心命题。比如，在传统电力系统中，为了给电网调频，我们常常会建设大型的火电厂。而为了让这些电厂的响应更敏捷、运行更高效，一种集成化的“液冷储能舱”架构应运而生。今天，我们就从这个看似遥远的角度，来聊聊它能为当下的算力能耗困局带来哪些启发。

【重要说明】本文/视频中所有关于节省金额、收益、回本周期、投资成本等数据，均为基于特定假设（如年用电量100万度、电价0.8元/度、光伏利用小时数等）的理论推演示例，不代表实际收益承诺，亦不构成购买或投资建议。实际收益受光照条件、电价波动、设备价格、安装费用、补贴政策等多种因素影响，可能存在显著差异。在做任何投资决策前，建议自行核实最新市场价格并咨询专业人士。

万卡GPU集群的能耗挑战与传统火电调频的液冷储能舱架构启示

最近和几位数据科学界的朋友聊天，他们的话题总绕不开训练大模型时，那些动辄上万张GPU卡组成的计算集群，耗电量实在有点“吓人”。这让我想起，在能源领域，我们其实早就面对过类似的、关于集中式高功率密度设施供能与散热的核心命题。比如，在传统电力系统中，为了给电网调频，我们常常会建设大型的火电厂。而为了让这些电厂的响应更敏捷、运行更高效，一种集成化的“液冷储能舱”架构应运而生。今天，我们就从这个看似遥远的角度，来聊聊它能为当下的算力能耗困局带来哪些启发。

现象是清晰的：无论是支撑AI革命的万卡GPU集群，还是保障电网稳定的火电调频系统，它们都代表着一种集中、高密度的能量消耗与处理单元。前者处理数据流，后者平衡电力流，但都面临“功率密度高、散热需求大、供电可靠性要求苛刻”的共性问题。根据一些行业分析，一个大规模AI训练集群的功耗可能接近一个小型城镇，其散热系统的能耗有时甚至能占到总能耗的40%。这不仅仅是电费账单的问题，更是整个系统设计逻辑的挑战。

数据会说话。在电力行业，为了提高火电机组的调频响应速度和精度，降低机组磨损，配套的储能系统——尤其是采用液冷技术的集装箱式储能舱——已经成为提升综合性能的关键。这种架构将电池、热管理系统（液冷）、能量转换系统（PCS）以及控制单元高度集成在一个标准化的舱体内。其优势在于：

高效热管理：液冷比传统风冷能更精准、更均匀地带走热量，确保电芯在最佳温度区间工作，这对于需要快速、频繁充放电的调频场景至关重要，能极大延长系统寿命。

功率与能量解耦设计：储能舱的功率输出（由PCS决定）和能量存储（由电池容量决定）可以相对独立地配置，以满足调频对快速功率支撑（高功率）和一定持续时间（适量能量）的双重要求。

模块化与可扩展性：像搭积木一样，可以根据调频容量需求灵活增加或减少储能舱的数量，部署快速，且便于后期运维。

这个思路，阿拉觉得完全可以迁移到超大规模计算中心的能源基础设施设计上。想象一下，将GPU集群的供电与散热视为一个整体来设计，而不是事后补救。一个高度集成、智能管理的“能源舱”概念

万卡GPU集群的能耗挑战与传统火电调频的液冷储能舱架构启示

，或许就是答案之一。在这方面，像我们海集能这样，在新能源储能和站点能源领域深耕近二十年的企业，其实已经积累了相当多的实践经验。从电芯选型、PCS设计，到整套系统的集成与智能运维，我们为全球客户提供一站式解决方案，尤其在应对极端环境和复杂电网条件方面，有着丰富的项目落地经验。

让我们看一个具体的案例，虽然它直接关联的是通信站点，但其底层逻辑相通。在东南亚某岛屿的离网通信基站项目中，当地气候高温高湿，电网脆弱。我们部署了光储柴一体化的站点能源柜。这个一体化柜集成了光伏控制器、锂电池组、智能液冷温控系统和柴油发电机控制器。通过智能能量管理系统，它优先使用光伏发电，并用储能电池调节波动，仅在必要时启动柴油机。结果是，在确保通信基站7x24小时不间断供电的前提下，柴油消耗降低了超过70%，并且系统在高温环境下运行稳定性远超预期。这个案例的核心，正是一体化集成、智能热管理（适应极端环境）和多种能源的协同优化。

那么，对于万卡GPU集群，我们能获得什么见解呢？我认为关键在于“系统级能效观”和“主动式热管理”。

对比维度

火电调频液冷储能舱

万卡GPU集群能源架构启示

核心目标

快速、精准响应电网频率波动，提升火电厂综合效益

稳定、高效支撑算力输出，降低PUE（电能使用效率），保障计算任务连续性

架构核心

预制化、模块化舱体，集成电池、液冷、PCS、控制

可考虑“算力能源一体舱”概念，将GPU服务器组、定制化配电、液冷散热模块、后备储能进行更高密度集成

热管理

液冷确保电池在最佳温区工作，提升寿命与性能

将液冷直接用于GPU芯片级散热，废热可进行更高品位回收利用，而非简单排散

智能控制

根据电网信号自动调节充放电，与火电机组协同

能源管理系统（EMS）与集群作业调度系统联动，在电价低谷或可再生能源充足时调度训练任务，实现“算力负载转移”

更进一步，我们可以思考，是否能为大型数据中心或算力中心配置专属的、模块化的大型储能系统（类似于储能舱的放大版），利用其快速调节能力，一方面平滑数据中心从电网取电的功率曲线，减少对电网的冲击，另一方面在电网需要时提供调频辅助服务，创造新的收益流。这实际上是将数据中心从

一个纯粹的能源消耗者，转变为具有一定灵活性的“产消者”。相关的技术探索，在一些前沿研究中已有讨论，例如美国劳伦斯伯克利国家实验室对数据中心柔性负荷的研究 (Lawrence Berkeley National Laboratory)。

海集能在江苏的南通和连云港布局的基地，一个擅长深度定制，一个专注标准化规模制造，这种“双轮驱动”的模式，恰恰能够应对从传统电力调频到新兴算力基础设施等不同场景的需求。无论是需要适应特殊气候、特殊电网条件的定制化方案，还是追求快速部署、成本优化的标准化产品，我们都能依托从电芯到系统的全产业链把控能力，为客户交付可靠的解决方案。我们相信，能源技术的进步，其价值最终体现在为各种前沿科技应用提供坚实、绿色、高效的底座。

所以，下一个值得探讨的问题是：当AI对算力的渴求持续增长，我们是否应该从设计之初，就将计算集群与智慧能源系统（包括高效储能和热管理）视为一个不可分割的整体来规划？这样的范式转变，会催生出怎样的新一代绿色算力基础设施形态？期待听到各位的思考。

来源: <https://www.hjenergysolution.com>