

# 中国东数西算节点万卡GPU集群算力负荷实时跟踪白皮书

最近和几位做高性能计算的朋友聊天，他们都在关注一个现象：那些部署在“东数西算”枢纽节点的大型数据中心，特别是动辄搭载上万张高性能GPU的AI训练集群，其能耗管理正成为一个既关键又棘手的课题。你知道吗，一个满载的万卡集群，其峰值功耗可能接近一个小型城镇的用电量。这不仅仅是电费账单的问题，更关乎整个算力基础设施的稳定性与可持续性。一份深入探讨其算力负荷实时跟踪机制的白皮书，其价值正在于此——它试图为这股澎湃的算力巨流，安装上精准的“流量监测仪”和“智能调度阀”。

## 中国东数西算节点万卡GPU集群算力负荷实时跟踪白皮书

最近和几位做高性能计算的朋友聊天，他们都在关注一个现象：那些部署在“东数西算”枢纽节点的大型数据中心，特别是动辄搭载上万张高性能GPU的AI训练集群，其能耗管理正成为一个既关键又棘手的课题。你知道吗，一个满载的万卡集群，其峰值功耗可能接近一个小型城镇的用电量。这不仅仅是电费账单的问题，更关乎整个算力基础设施的稳定性与可持续性。一份深入探讨其算力负荷实时跟踪机制的白皮书，其价值正在于此——它试图为这股澎湃的算力巨流，安装上精准的“流量监测仪”和“智能调度阀”。

现象背后是冰冷的数据。根据行业估算，一个典型的万卡GPU集群（以当前主流高性能卡计），其IT设备负载往往在5-8兆瓦之间，这还不包括为保障这些芯片稳定运行所必需的冷却系统能耗，后者通常能占到总能耗的30%-40%。这意味着，单个集群的总用电需求轻松突破10兆瓦。当数以百计这样的集群在西部能源富集区拔地而起时，它们对当地电网的瞬时负荷、电能质量以及长期能源规划的影响，是颠覆性的。负荷不再是平滑曲线，而是随着AI训练任务起落而剧烈波动的尖峰脉冲。传统的“以电定算”粗放模式，在“双碳”目标和经济效益的双重压力下，已经难以为继。算力与电力之间，需要更精细、更动态的协同。

那么，如何实现这种协同？关键在于“实时跟踪”与“智能响应”。这不仅仅是软件层面的监控大屏，更需要硬件层面的柔性支撑。这就引出了我们常说的“源-网-荷-储”一体化思路。在“荷”端，即算力集群本身，通过实时采集每一机柜、每一服务器乃至每一组GPU的功耗与性能数据，构建数字孪生模型，精准预测负荷变化趋势。更重要的是，需要有一个能够快速响应调度指令、平抑波动、甚至参与需求侧响应的“缓冲池”和“稳定器”——这正是储能系统可以大显身手的舞台。

这里可以看一个贴近的场景。想象在宁夏或甘肃的某个算力枢纽，一个专注于自动驾驶模型训练的万卡集群正在运行。某天下午，因训练任务集中爆发，集群功耗急速拉升至峰值，同时，当地电网恰逢检修时段供电能力受限。此时，如果数据中心配备了智能储能系统，它便能在毫秒级时间内，根据实时跟踪到的负荷飙升信号，无缝切入并释放预存的绿色电能，保障算力任务“零感知”连续运行，同时避免对公共电网造成冲击。等到夜间训练任务减少、电网负荷低谷且绿电充沛时，储能系统再从容充电，既消纳了过剩的可再生能源，也大幅降低了数据中心的用电成本。这套“算力负荷跟踪+智能储能响应”的组合拳，正是实现算力基础设施高效、绿色、弹性的核心。

在这个领域深耕，阿拉（上海话，意为我们）海集能有着近二十年的技术积累。从上海总部到南通、连云港的两大生产基地，我们构建了从核心部件到系统集成的全产业链能力。特别是在为通信基站、边缘计算站点等提供高可靠能源解决方案的过程中，我们深刻理解了“无电弱网”环境下保障关键负载

连续运行的极端重要性。这种经验，被我们融入到为大型数据中心和算力节点设计的储能解决方案中。我们的标准化与定制化并行的生产体系，能够针对不同枢纽节点的电网条件、气候环境，提供从磷酸铁锂电芯、高效PCS（功率转换系统）到一体化系统集成和智能运维的“交钥匙”服务。目标很明确：就是让算力不再受制于电力的波动，让每一份绿电都能被高效利用。

这份白皮书所揭示的，其实是一个更宏大趋势的缩影：数字世界与物理世界的能源系统正在深度耦合。未来的算力，必将是“可观测”、“可调控”、“可优化”的。它要求我们不仅要懂算力，更要懂电力；不仅要会建数据中心，更要会运营一个与电网友好互动的“虚拟电厂”。这其中，实时跟踪是感知的起点，而基于此的智能调度与储能缓冲，则是实现价值闭环的关键行动。

随着AI算力需求呈指数级增长，你认为，下一个十年，除了储能，还有哪些跨领域技术将成为支撑“东数西算”战略平稳落地的关键支柱？我们是否已经为应对算力功耗的下一轮爆发做好了足够的能源准备？

---

来源: <https://hjenergysolution.com>