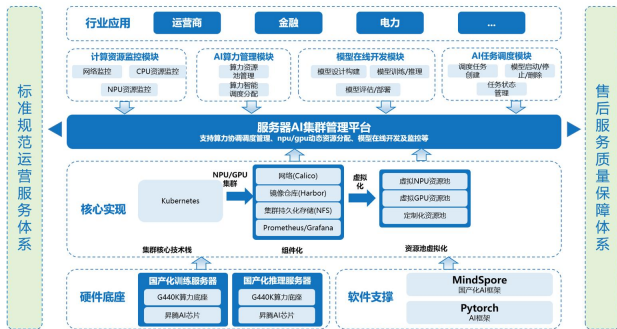


服务器集群管理平台



服务器集群管理平台是烽火通信自主研发的一款高性能集群算力管理分配一体解决方案，该平台旨在帮助企业提升AI应用效率，降低成本，并简化AI算力管理复杂性，并支持国产AI硬件，增强我国在AI技术领域的自主可控能力，可广泛应用于电力、金融行业、运营商、智能制造等行业。通过灵活应用于多个领域，该算力一体化解决方案可助力各行各业智能化转型和效率提升。

功能全面

- **计算资源监控**：实时监控NPU、GPU的利用率、温度和功耗；
- **AI任务调度管理**：支持任务创建、调度、交付及状态管理；
- **智能调度算法**：根据资源使用情况和任务优先级动态调度算力；
- **模型推理/训练日志**：支持问题追踪及分析；
- **AI算力管理**：采用NPU/GPU算力虚拟化技术，兼容国产算力卡，提升资源的灵活性与效率，增强数据安全和自主控制。

参数指标

计算性能	可支持8台以上服务器、20p以上算力的中大型集群管理及调度，满足大规模模型训练需求。
灵活扩展	支持横向扩展，每增加一个计算节点，整体性能提升至少20%；支持纵向扩展，单节点性能提升可达30%。
生态兼容性	支持多种NPU和GPU硬件，兼容主流深度学习框架。支持超过100种不同的金融工具和软件包；API调用成功率不低于99.9%，平均API响应时间小于200毫秒。
智能调度	资源利用率平均提高30%以上；任务响应时间缩短至少50%。
用户友好	新用户首次登录后，可在15分钟内完成基本设置；用户界面评分 (UI Score) 达到4.8/5 (满分5分)。

应用场景

服务器集群管理平台

可广泛应用于运营商、金融行业、电力及智能制造等行业



方案价值



开箱即用

- **一键部署**：通过脚本化管理，用户只需一条命令即可完成平台的部署，快速上线并开始调度管理，减少复杂的系统配置和运维压力。
- **灵活集成**：服务器集群管理平台兼容主流硬件和AI框架，能够适应企业在多场景下的应用需求，帮助用户无缝集成已有系统，实现计算资源的全面整合。

服务模式

服务器集群管理平台解决方案可私有化部署，可根据业务需要，定制化训练/推理模型，编码在线开发与指令，并提供部署应用相关的技术支持和咨询，以及后续更新升级等专业服务。