佳杰云星
CLOUD STAR

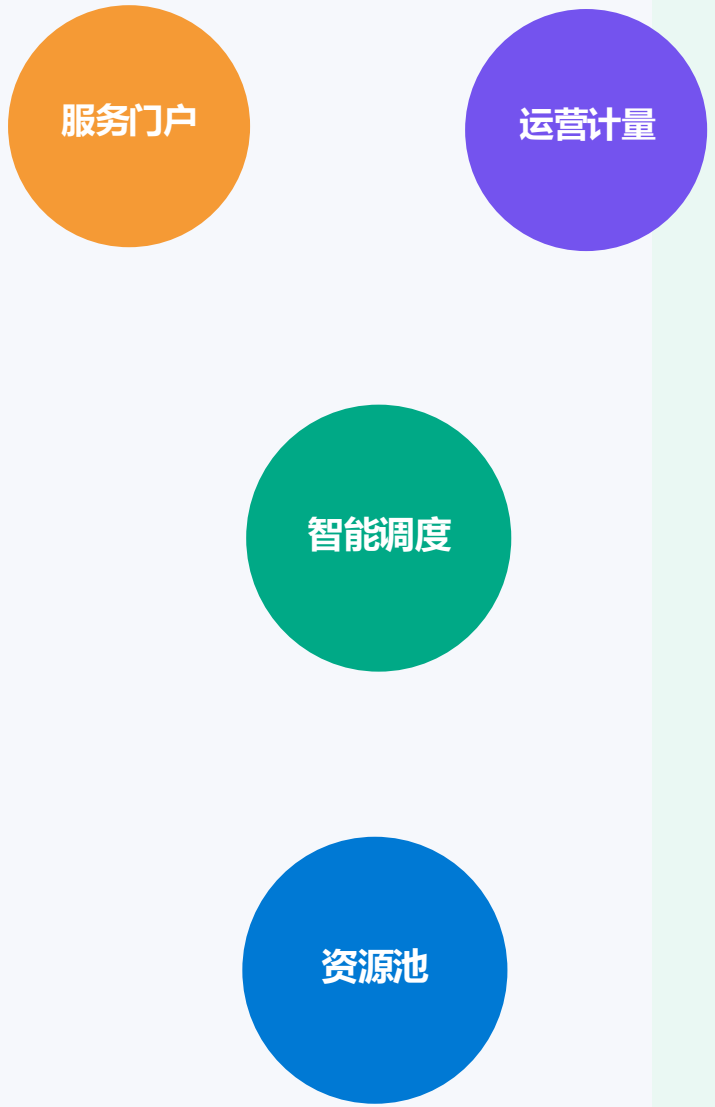
算力调度与管理平台

面向智算中心、区域算力枢纽和高校科研场景的异构算力统一纳管、智能调度与运营平台

GPU / NPU / DCU / MLU

多租户服务化

统一运营计量



用户挑战

佳杰云星 CLOUD STAR

算力能不能被高效使用和管理?

异构设备、分散资源和人工申请流程，让 AI 研发与算力运营之间形成断点。

01 资源分散

多品牌 GPU/NPU、裸金属、容器集群各自管理，资源视图割裂。

02 调度困难

训练、推理、Notebook 等负载特征不同，靠人工分配难以稳定匹配。

03 使用门槛高

开发者需要关心环境、镜像、卡型、队列和资源申请细节。

04 运营不可视

资源利用、租户配额、计量账单、成本收益难以形成闭环。

产品要解决的核心命题

把分散异构算力转化为可申请、可调度、可计量、可运营的统一算力服务。

佳杰云星 CLOUD STAR

一个平台，连接算力资源、AI 开发与运营服务

从底层异构设备到上层用户门户，建立完整的算力服务化链路。

运营管理	租户、目录、审批、计量、账单、分析
服务入口	Notebook、训练作业、模型部署、推理服务
智能调度	负载感知、异构适配、队列策略、跨池分发
资源池化	统一纳管、卡级切分、配额隔离、资源池视图
基础设施	GPU / NPU / DCU / MLU / CPU / 裸金属 / 容器集群

面向角色

- 管理员
- 运营人员
- AI 开发者
- 租户用户

佳杰云星 CLOUD STAR

标准版覆盖算力调度平台的核心软件闭环

面向租户、运维管理员和系统管理员，提供从资源接入、配额调度到开发推理、监控审计的基础能力。

租户自服务

资源总览与配额查看
对象/块存储、访问密钥
自定义镜像、Notebook、推理服务

资源运维管理

集群注册、同步、监控、详情
资源池与节点管理
异构卡纳管与 GPU 虚拟化

调度与配额

租户级共享资源池配额
超配额自动排队
防碎片、公平、细粒度调度策略

监控告警

节点、集群、GPU/NPU、容器指标
告警规则、活动告警、历史告警
联系人、邮件/短信通知

AI 生命周期管理

文件/块存储与镜像管理
开发环境与推理服务统一查看
训练作业和模型资产管理基础能力

平台治理

租户、租户映射、角色和管理员
系统配置、卡型号、公共桶
站内消息、自监控、安全审计

标准版定位 **让平台先具备“可纳管、可申请、可调度、可监控、可审计”的基础生产能力。**

佳杰云星 CLOUD STAR

按业务成熟度扩展数据、模型、运营和网关能力

可选模块适合面向外部租户运营、AI 工程平台化、大模型服务化和数据治理场景逐步叠加。

数据治理

多源数据归集、资产管理、清洗和质量评估，提升训练数据可用性。

归集 | 资产 | 清洗 | 评估

训练与微调

通用训练作业、LLM 模型管理、LoRA/QLoRA 等微调流程。

作业 | 模型 | LoRA | QLoRA

运营支撑

资源申请审批、实名认证、订单、账单、收支明细、工单和公告。

审批 | 订单 | 账单 | 工单

模型网关

统一接入公有/私有大模型，支持路由、插件、监控和资产集市。

供应商 | 路由 | 插件 | 监控

运营级计量

按小时生成作业计量数据，支持租户、集群维度查询、导出和话单推送。

计量 | 导出 | 话单 | 推送

AI 平台扩展

面向企业级 AI 工程，扩展模型资产、推理 API、Token 用量和订阅机制。

推理 | API | Token | 订阅

佳杰云星 CLOUD STAR

从资源纳管到用户使用，形成闭环流程

用户只看到服务，平台负责资源、调度、计量与运营。

1 **纳管**
接入裸金属与异构计算卡

2 **池化**
形成统一资源池与规格

3 **发布**
上架算力服务目录

4 **申请**
租户自助申请与审批

5 **调度**
平台智能匹配资源

6 **运营**
计量、账单、分析、优化

结果 **管理员管理资源，运营人员管理服务，开发者专注模型与应用。**

佳杰云星 CLOUD STAR

底层硬件，中间池化适配，上层软件能力

平台把 GPU、NPU、DCU 等异构硬件抽象为统一资源池，再向上提供调度、开发和运营能力。

软件能力层

统一门户
申请算力服务
查看资源与任务

智能调度
负载感知
队列与策略匹配

AI 开发
Notebook
训练与模型部署

运营管理
租户配额
计量账单与分析

池化适配层

统一资源池与异构适配
设备接入 · 驱动/框架适配 · 资源切分 · 配额隔离 · 统一监控 · 故障与容量治理

硬件资源层

底层硬件与基础设施资源
NVIDIA GPU | 华为昇腾 NPU | 海光 DCU | 寒武纪 MLU | CPU / 裸金属 | 存储 / 网络

对用户的变化：只面对统一服务入口，不再直接处理具体卡型、驱动环境、队列位置和底层资源差异。

佳杰云星
CLOUD STAR

让 AI 开发者更快进入工作状态

围绕 Notebook、训练作业、模型和推理服务，提供一站式研发入口。

预置环境

减少框架和依赖配置成本

一键提交

训练任务统一调度

模型复用

模型管理到部署贯通

弹性推理

按业务负载调整服务资源

算力不仅要调度，还要能运营

面向多租户、服务目录、审批、计量和账单，支撑智算资源持续经营。

运营结果 **资源可定价、服务可上架、用量可追踪、成本收益可分析。**

佳杰云星 CLOUD STAR

四类场景优先适配

从单一智算中心到区域级算力枢纽，平台价值会随资源复杂度提升。

01 智算中心

统一纳管 GPU/NPU 资源，提供开发、训练、推理与运营能力。

02 区域算力枢纽

跨地域、跨资源池统一入口，支撑多主体算力服务化。

03 高校科研

面向课题组、课程教学和科研任务，提供配额、隔离与快速开发环境。

04 政企算力服务平台

结合审批、计量、账单和运营分析，支撑对内服务或对外运营。

选择逻辑 **资源越异构、租户越多、运营要求越强，越需要平台化治理。**

佳杰云星
CLOUD STAR

把算力投资转化为可持续服务能力

不仅提升资源利用，更降低 AI 应用落地门槛。

01 提升利用率

通过池化、调度和配额治理，减少资源闲置与碎片化。

02 降低门槛

开发者通过统一入口使用 Notebook、训练和推理服务。

03 统一运营

租户、服务、订单、计量和账单形成可管理闭环。

04 兼容国产化

支持多类型异构算力资源，适配国产化与多云环境。

一句话总结 **让算力像云服务一样被申请、调度、计量和运营。**

佳杰云星
CLOUD STAR

下一步：了解产品，申请演示

如果您正在建设智算中心、区域算力平台或高校科研算力环境，可以通过官网联系我们获取完整方案资料。

查看产品彩页

快速了解算力调度和运营平台能力。

预约产品演示

结合现有算力环境评估接入与调度方案。

申请完整方案

获取行业场景、部署架构和交付方案材料。

www.cloud-star.com.cn | 400-651-8860