

从需求开始，找到合适组合。

提供模型、预计用量、应用场景或算力规格，即可沟通资源组合与交付方式。



扫码咨询

电话：027-87805210

手机：18627001072

网址：xserver.cn

聚合服务带来的改变

01

统一入口

减少多平台切换与重复接口开发

02

灵活采购

适配试用、项目和长期使用需求

03

资源可控

用量可统计，方案可比较

04

持续支持

从咨询、接入到部署交付

模型、Token、算力，一站式对接。

资源如何进入业务

从 Token 调用、采购到算力部署，覆盖科研、开发、商业服务和项目交付。

01

Token 聚合

统一接入主流模型，按能力、速度和成本选择。

02

Token 代理

支持按量、按周期和项目需求对接。

03

算力租赁

提供 GPU 实例和专属资源。

04

设备销售

提供工作站、服务器和集群方案。

典型应用场景



科研与数据分析



软件研发与系统集成

智能应用、知识库、代码生成与自动化。



商业服务与业务运营

客服、内容、营销、办公与数据洞察。

案例覆盖

政府

政务智能化

高校

科研分析

企业

流程自动化

研究所

算力平台

AI 资源聚合服务

连接模型 • Token • 算力

一个入口，组合国内外主流大模型与高性能计算资源。



从资源选择到技术落地

灵活采购 / 统一接入 / 持续支持

获取匹配方案

电话：027-87805210 | 手机：18627001072

网址：xserver.cn

01

核心优势

Why choose aggregation

资源选择更简单

强调灵活性、透明度与持续服务，不局限于单一模型或单一资源形态。

01 多模型统一接入
减少重复接口开发

02 资源按需组合
模型与算力灵活搭配

03 采购方式灵活
适配试用与项目周期

04 用量清晰可查
便于预算核算与比较

05 按场景选模型
兼顾能力、速度与成本

06 算力弹性扩缩
应对峰值和阶段性任务

07 软硬件协同
覆盖接口、服务器与集群

08 持续技术支持
从咨询到部署交付

一套资源路径

需求梳理 → 模型与算力匹配 → 采购接入 → 部署支持

02

Token 与模型

Token and model service



统一接入，灵活采购

让不同模型在同一需求下可比较、可切换。

Token 服务路径



适用场景

- 智能问答与知识库
- 文本、图像与多模态调用
- 代码生成与研发辅助
- 批量内容与流程自动化

代表性模型品牌

- 字节豆包
中文与内容协作
- DeepSeek
推理与代码
- OpenAI GPT
通用智能与多模态
- Claude
长文本与代码
- Gemini
多模态与长上下文

具体型号、版本与供应清单以实际服务目录为准

通用模型 推理模型 多模态 智能体

03

算力与交付

Compute and delivery

算力按需到位

从云端实例到专属服务器，按任务规模组合资源。



资源形态

- 01 GPU 云端实例
快速开通，适合阶段性任务
- 02 专属算力资源
适合稳定负载与独立使用
- 03 GPU 工作站
面向研发与本地推理
- 04 AI 服务器与集群
面向规模化部署与交付

典型任务与交付支持

- 模型推理
- 训练微调
- 数据处理
- 项目扩容
- 私有部署
- 集群交付

获取匹配方案

电话：027-87805210 | 手机：18627001072 | 网址：xserver.cn