

# 超强A系列异构算力产品主打胶片



# 软通计算 超强 A系列异构算力产品，打造异构算力产品标杆



## 数据中心

### 中心训练型服务器

900 AI集群



数据中心机房

超强AI算力、极佳AI拓展  
更优AI能效

超强A900P 47U



数据中心机房

超强AI算力、极佳AI拓展  
更优AI能效

超强A800T A2 4U4路



数据中心机房

更高算力密度、极致能效比  
高速网络带宽

超强A800I A2 4U2路



数据中心机房

高算力密度，高能效比  
大模型推理

### 中心推理型服务器

超强A801I 2U2路



数据中心机房

灵活配置，适配多项负载、  
智能视频分析

超强A800I 2U2路



数据中心机房

超强算力、超高能效

## 边缘计算

### 边缘型

超强A500P 2U1路



大型园区、商场、电力、交通等

超强算力、超高能效

### 开发者套件

超强A200I A2 DK开发者套件



学校、算法公司、智能制造企业

高集成度、丰富扩展接口

### 加速模块

Atlas200I A2



摄像头、无人机、机器人等

极致性能、超低功耗

# 推理硬件平台

## 边缘模块形态

超强200I A2 加速模块

高度集成 超强解析



20T

OPS INT8 AI算力

超强200I DK A2 开发者套件

开箱即用 参考丰富



8T

OPS INT8 AI算力

## 标卡形态

Atlas 300V Pro视频解析卡

高清解析 业界第一



128路

1080P 30FPS

48GB

LPDDR4X

Atlas 300V 视频解析卡

高清解析 高性价比



100路

1080P 25FPS

24GB

LPDDR4X

Atlas 300I Pro 推理卡

智能加速 性能领先



140T

OPS INT8 AI算力

24GB

LPDDR4X

Atlas 300I Duo 推理卡

双芯配置 性能翻倍



280T

OPS INT8 AI算力

48/96GiB

LPDDR4X

## 服务器形态

超强A500P 智能边缘服务器

2U短机箱 适用短机柜



3张

Atlas 300V, 300I/V Pro

超强A800I 推理服务器

2U标准 灵活扩展



8张

Atlas 300V, 300I/V Pro

超强A801I 推理服务器

自研平台 极致扩展



8张

Atlas 300V, 300I/V Pro

超强A800I A2推理服务器

4U8模组 高性能配置



8模组

高性能NPU模组

# Atlas 200I A2-超强边缘视频分析



Atlas 200I A2 加速模块

关键特性	规格描述
形态	采用MXM连接器: 82 mm (长) * 60 mm (宽) * 7 mm (高)
内存	LPDDR4X, <b>12 GB</b> , 总带宽 51.2 GB/s 或LPDDR4X, 4 GB, 总带宽 34.1 GB/s
AI算力	整数精度 (INT8) : <b>20</b> /8 TOPS 半精度 (FP16) : <b>10</b> /4 TFLOPS
编解码能力	内置DVPP预处理单元 图片JPEG/PNG编解码能力 视频 <b>40路</b> 1080P 30FPS
典型功耗	20 TOPS: 25 W / 8 TOPS: 21 W
ISP	4KP45 (4K画质每秒45帧) / 8KP15 / 16KP4
图形显示	2路HDMI图片输出, 满足现场结果直接显示

**集成度高 以一当四, 不需单配CPU和ISP, 硬件性价比更高**

Atlas 200I A2

等于

CPU

4核LX900  
1.6GHz

NPU

20/8 TOPS  
INT8

编解码

最大40路1080P@30fps  
2路8K@30fps  
**业界最强**

图形显示

单核Mali-G52  
2路HDMI

**应用场景** 嵌入边缘设备, 如机工控机、智能网关等室内、桌面级部署的边缘AI场景, 大型机器人等供电充足应用场景, 安防场景边缘后端 (非太阳能供电)

智能边缘设备



智能机器人



视频分析后端



**1** 视频性能强, 单路性价比高

Atlas 200I A2

INT8 算力

**持平**

VS

NV Jetson Orin Nano

硬件视频解析

**3x** ↑

SoC内存

**1.5x** ↑

**2** 接口丰富, 行业场景更容易适配

Atlas 200I A2

**MXM接口**  
(314pin)

硬件实现更多的网络/摄像头/USB/串口/音频等

VS

NV Jetson Orin Nano

**DDR4 SO-DIMM接口**  
(260pin)

# 超强A200I DK A2：开箱即用、参考丰富



Atlas 200I DK A2  
开发者套件

Atlas 200 DK 演进一代产品

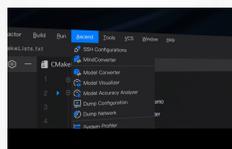
关键特性	规格描述
形态	135 mm (长) * 120mm (宽) * 44 mm (高)
AI算力	整数精度 (INT8) : <b>8</b> TOPS 半精度 (FP16) : <b>4</b> TFLOPS
摄像头接口	2*MIPI-CSI 支持两个树莓派摄像头
USB接口	1*USB Type-C 仅支持从模式; 2*USB3.0 Type-A
以太网接口	2*RJ45千兆网口
编解码能力	内置DVPP预处理单元 图片JPEG/PNG编解码能力 视频 <b>20路</b> 1080P 30FPS
典型功耗	24 W
存储设备	1*NVMe M.2 SSD; 1*Micro SD卡
图形显示	2路HDMI

**应用场景** 昇腾AI开发者上手学习，实践创新场景，提供配套软硬件

**AI应用创新**  
机械臂/智能小车...



**高校教学**  
AI教学实训



**行业算法验证**  
ISV等行业算法



## 1 丰富示例易学易用

- 3大端到端场景化Demo, 100+场景样例 (视图分析、OCR、动作/手势识别等)

智能车

机械臂

语音交互

## 2 预训练模型简单适配

- ModelZoo**: 900+高性能预训练模型, CV/NLP/语音等

## 3 推荐配件方便组装

推荐配件指导, 降低选型兼容难度

连接扩展



接口扩展



组件扩展



场景扩展



## 4 高效工具极简易用

- 一键制卡工具**: 一键安装镜像, 30min环境搭建
- 模型适配工具**: Windows下的端侧模型适配工具, 支持用户完成端侧模型适配开发全流程

# Atlas 300I Pro推理卡



## 超强算力

- 单卡最大提供140 TOPS INT8算力，为数据中心推理提供更强大支持
- 支持8 core \* 1.9 GHz CPU计算能力

## 超高能效

- 提供 2 TOPS / W 超高能效比，达到业界2.1倍，节省机房部署成本

## 安全启动

- 设备启动链完整，启动状态确定，防止后端植入

## 关键特性 规格描述

形态	单槽位半高半长PCIe卡
AI算力	<b>140</b> TOPS INT8 / 70 TFLOPS FP16
CPU算力	8 core * 1.9 GHz
内存规格	24GB LPDDR4X，总带宽204.8 GB/s，支持ECC
硬件编解码	H.264、H265视频编解码，JPEG图片编解码
最大功耗	72 W

## 应用场景



OCR 识别



搜索推荐



语音分析



内容审核

# Atlas 300V 视频解析卡



## 超强算力

- 单卡最大提供100 TOPS INT8算力，为数据中心推理提供更强大支持
- 支持8 core \* 1.9 GHz CPU计算能力

## 超强视频解析

- 支持100路 1080P 25FPS/80路 1080P 30FPS 视频实时分析能力
- 支持JPEG和视频硬件编解码，提升图片和视频应用分析性能

## 安全启动

- 设备启动链完整，启动状态确定，防止后端植入

## 关键特性 规格描述

形态	单槽位半高半长PCIe卡
AI算力	100 TOPS INT8 / 50 TFLOPS FP16
CPU算力	8 core * 1.9 GHz
内存规格	24GB LPDDR4X，总带宽204.8 GB/s，支持ECC
硬件编解码	H.264、H265视频编解码，JPEG图片编解码 支持100路 1080P 25FPS/80路 1080P 30FPS 视频解码
最大功耗	72 W

## 应用场景



智慧城市



智慧金融



智慧交通



智慧园区

# Atlas 300V Pro 视频解析卡



## 超强视频解析

- 支持128路1080P 30FPS 视频实时分析能力
- 支持JPEG和视频硬件编解码，提升图片和视频应用分析性能

## 超大内存容量

- 单卡支持LPDDR4X 48GB超大内存，在视频解析、检索聚类、OCR等场景中提升处理性能

## 安全启动

- 设备启动链完整，启动状态确定，防止后端植入

## 关键特性 规格描述

形态	单槽位半高半长PCIe卡
AI算力	<b>140</b> TOPS INT8 / 70 TFLOPS FP16
CPU算力	8 core * 1.9 GHz
内存规格	<b>48GB</b> LPDDR4X，总带宽204.8 GB/s，支持ECC
硬件编解码	H.264、H265视频编解码，JPEG图片编解码 128路 1080P 30FPS 视频解码
最大功耗	72 W

## 应用场景



智慧城市



智慧金融



智慧交通



智慧园区

# Atlas 300I Duo 推理卡



## 超强算力

- 单卡最大提供280TOPS INT8算力，为中心推理提供更强大算力支持
- 最大支持96GB内存，为大模型推理并发提供支持
- 支持16 core \* 1.9 GHz CPU计算能力

## 超大内存容量

- 提供1.86 TOPS/W 超高能效比，业界领先

## 高性能视频分析

- 支持256路1080P 30FPS 视频实时分析
- 支持JPEG和视频硬件编解码，提升图片和视频应用分析性能

## 关键特性 规格描述

形态	单槽位全高全长PCIe卡
AI算力	<b>280</b> TOPS INT8 / 140 TFLOPS FP16
CPU算力	16 core * 1.9 GHz
内存规格	<b>48GB</b> 或 <b>96GB</b> LPDDR4X，总带宽408 GB/s，支持ECC
硬件编解码	H.264、H265视频编解码，JPEG图片编解码 256路 1080P 30FPS 视频解码
最大功耗	<b>150 W</b>

## 应用场景



# 超强A500P智能边缘服务器：满足“短机柜”部署

支持 **1 ~ 3** 张卡

“短机柜”  
室外电信设备机柜 (III、IV)  
典型尺寸深度**650mm**

插卡式



Atlas 300I Pro  
Atlas 300V Pro  
Atlas 300V

关键特性	规格描述
形态	2U服务器, 短机箱 (86.1mm x 447mm x <b>475mm</b> )
AI加速卡	最大支持 <b>3</b> 张 Atlas 300I/V Pro推理卡
CPU	1x 鲲鹏920 (单路24核)
内存	4个DDR4内存插槽, 最高3200MT/s
AI算力	最大 <b>420</b> TOPS INT8 或 <b>384</b> 路 1080P 30FPS视频解析 (硬件解码能力)

**应用场景** 大型园区、电力、交通、商场等场景



智慧加油站



智慧营业厅



智慧交通



智能制造

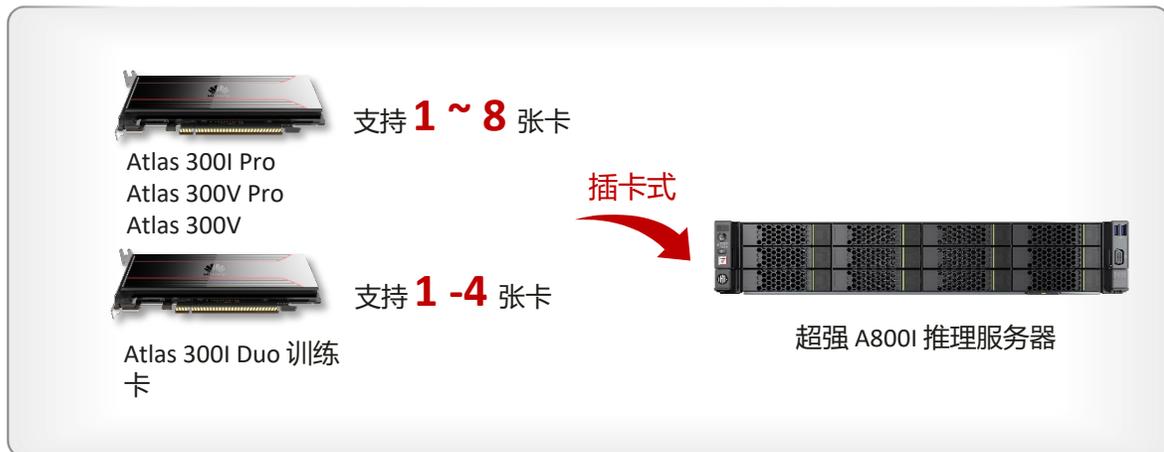


智慧社区



智能楼宇

# 超强A800I 推理服务器



关键特性	规格描述
形态	2U服务器
AI加速卡	最大支持 <b>4</b> 张Atlas 300I Duo推理卡, 或 <b>8</b> 张300I/V Pro加速卡
CPU	2x 鲲鹏920
内存	32个DDR4内存插槽, 最高3200MT/s
AI算力	最大 <b>1120</b> TOPS INT8 或 <b>1024</b> 路 1080P 30FPS 视频解析 (硬件解码能力)

**应用场景** 部署在数据中心, 使能中心推理



## 超强算力

支持8张Atlas 300I/V Pro 推理卡, 满足多场景推理需求;  
整机可提供1024路高清视频实时分析 (1080P 30FPS)  
搭载超强算力的鲲鹏920处理器, 高效加速应用

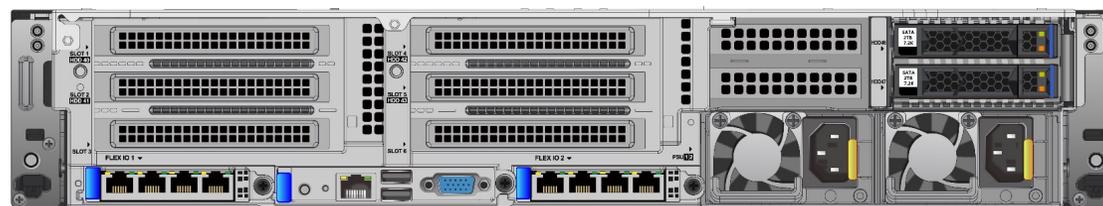
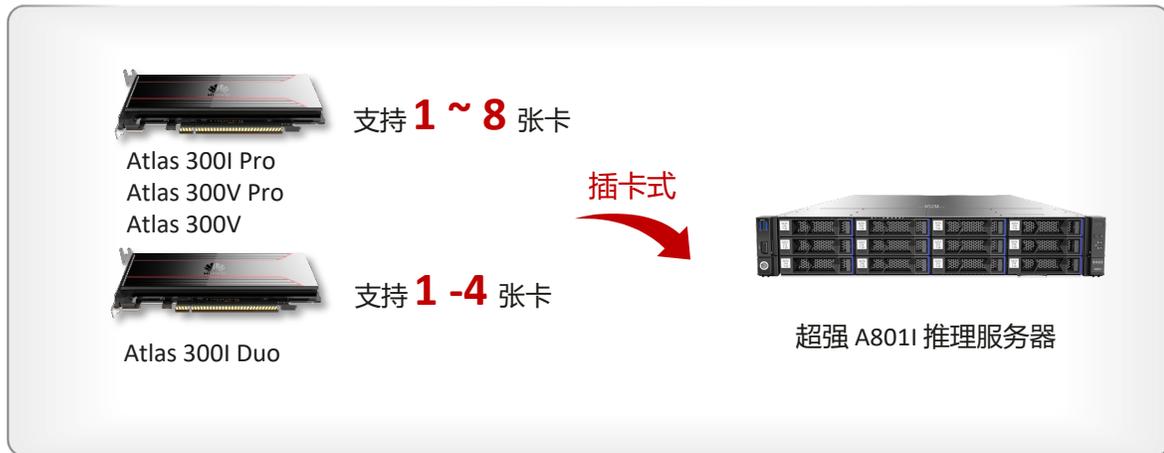


## 超高能效

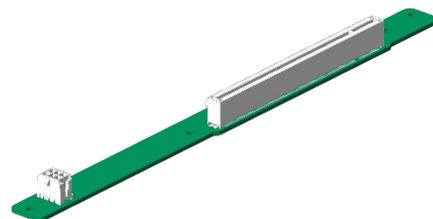
发挥鲲鹏架构多核、低功耗优势, 推理场景构建高效能、低功耗的AI计算平台  
Atlas 300I Pro单卡功耗仅72W, 为AI服务器算力加速同时提供更优的能效比



# 超强A801I 推理服务器：与超强A800I产品差异点

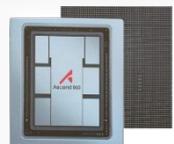


→ 8\*PCIeX8+2\*2.5后置



→ 中置X8 PCIe插槽，额外支持1张标准Raid卡

# 超强A800I A2 推理服务器



超强 A800I A2 推理服务器

## 关键特性 规格描述

形态 4U 昇腾模组服务器

AI加速卡 最大支持 **8** 昇腾AI**推理**加速模组

CPU 4x 鲲鹏920

内存 32个DDR4内存插槽, 最高3200MT/s

AI算力 **高效推理算力**

## 应用场景 部署在数据中心, 使能中心推理



### 更高算力密度

4U 高度最高提供业界领先算力

搭载8颗超强算力的昇腾 AI 模组, 高效加速应用



### 极致能效比

提供超高能效比, 节省运营费用



### 高速网络带宽

8\*200G RoCE v2高速接口

8路 NPU 通过 HCCS 或 PCIe 高速互联



内容生成



智能问答



内容审核



搜索NLP



互联网推荐



智能客服

# 训练硬件平台

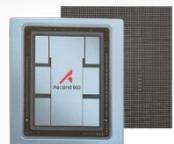


超强 A900P A2  
超强 A900 AI集群



超强A800T A2  
训练服务器

# 超强A800T A2 训练服务器



超强 A800T A2 训练服务器

## 关键特性 规格描述

形态 4U 昇腾模组服务器

AI加速卡 最大支持 **8** 昇腾AI**训练**加速模组

CPU 4x 鲲鹏920

内存 32个DDR4内存插槽，最高3200MT/s

AI算力 **业界领先AI算力**

**应用场景** 部署在数据中心，使能中心训练



### 更高算力密度

4U高度最高提供超强算力

搭载8颗超强算力的昇腾910处理器，高效加速应用



### 极致能效比

提供超高能效比，节省运营费用，降低训练成本



### 高速网络带宽

8\*200G RoCE v2高速接口

8路NPU通过HCCS Fullmesh互联，双向互联带宽达392GB/s



智慧城市



运营商



智慧金融



教育科研

# 超强 A900P/A900Pc A2 : 集群基础单元



超强 A900P A2 集群基础单元, 具有极致算力密度、极高AI能效、极简交付部署等特点。该基础单元广泛应用于深度学习模型开发和训练, 适用于人工智能计算中心、智慧城市、科研探索、运营商等需要重算力的场景。

## 极致算力密度

- 47U 高度可提供高密度AI 算力
- 支持机柜单元扩展, 全节点200G网络互联

## 极高AI能效

- 能效比最高可达 0.5 PFLOPS / kW
- 全液冷散热, 免冷机、免空调, PUE $\leq$ 1.15

## 应用场景



## 极简交付部署

- 计算节点高可靠全盲插
- 机柜工厂预集成, 整柜运输管理, 交换机一键配置

## 产品规格

形态	47U 机柜
NPU	64 * 昇腾910
CPU	32 * 鲲鹏920
内存	256个DDR4内存插槽, 最高3200 MT/s 单根内存条容量支持32 GB / 64 GB
HBM	总容量 4096GB, 带宽 1600 GB/s
AI算力	高密度AI算力
本地存储	最大支持 80 * 2.5英寸硬盘
供电	交流: 8路4+4, 电源: 380V/32A
散热方式	液冷
工作环境温度	工作温度: 5°C ~ 40°C (41°F ~ 104°F) (符合ASHRAE Class A1/A2/A3/A4)
结构尺寸	2250mm (高) × 600mm (宽) × 1050mm (深) (不含前后造型门) 2250mm (高) × 600mm (宽) × 1500mm (深) (含前后造型门)



感谢聆听