

# 昇腾产品形态说明

文档版本 01  
发布日期 2023-08-31



**版权所有 © 华为技术有限公司 2023。保留一切权利。**

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

## **商标声明**



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

## **注意**

您购买的产品、服务或特性等应受华为公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

## 目 录

1 昇腾产品形态说明.....	1
-----------------	---

# 1 昇腾产品形态说明

昇腾AI产品包含训练和推理产品，主要产品系列如表1-1所示，昇腾训练产品同时支持训练和推理业务。

表 1-1 昇腾产品系列

产品系列	产品型号
Atlas 200/300/500 推理产品	Atlas 200 AI加速模块、Atlas 300I 推理卡（型号：3000）、Atlas 300I 推理卡（型号：3010）Atlas 500 智能小站、Atlas 200 DK 开发者套件
Atlas 200/500 A2推理产品	Atlas 500 A2 智能小站、Atlas 200I DK A2 开发者套件、Atlas 200I A2 加速模块
Atlas 推理系列产品（配置 Ascend 310P AI处理器）	Atlas 300I Pro 推理卡、Atlas 300V 视频解析卡、Atlas 300V Pro 视频解析卡、Atlas 300I Duo 推理卡、Atlas 200I SoC A1核心板
Atlas 推理服务器系列产品	Atlas 500 Pro 智能边缘服务器、Atlas 800 推理服务器（型号：3000）、Atlas 800 推理服务器（型号：3010） Atlas 500 Pro 智能边缘服务器支持插入Atlas 300I 推理卡（型号：3000）、Atlas 300I Pro 推理卡、Atlas 300V 视频解析卡、Atlas 300V Pro 视频解析卡使用 Atlas 800 推理服务器（型号：3000）支持插入Atlas 300I 推理卡（型号：3000）、Atlas 300I 推理卡（型号：3010）、Atlas 300I Pro 推理卡、Atlas 300V 视频解析卡、Atlas 300V Pro 视频解析卡、Atlas 300I Duo 推理卡使用 Atlas 800 推理服务器（型号：3010）支持插入Atlas 300I 推理卡（型号：3010）、Atlas 300I Pro 推理卡、Atlas 300V 视频解析卡、Atlas 300V Pro 视频解析卡、Atlas 300I Duo 推理卡使用
Atlas 训练系列产品	Atlas 800 训练服务器（型号：9000）、Atlas 800 训练服务器（型号：9010）、Atlas 900 PoD（型号：9000）、Atlas 900T PoD Lite、Atlas 300T 训练卡（型号 9000）、Atlas 300T Pro 训练卡（型号：9000）

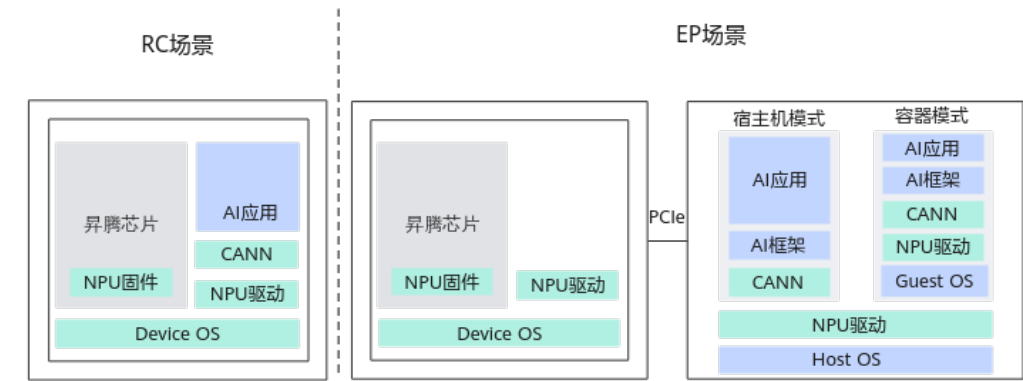
产品系列	产品型号
Atlas A2 训练系列产品	Atlas 800T A2 训练服务器、Atlas 900 A2 PoD 集群基础单元、Atlas 300T A2 训练卡

昇腾产品以昇腾AI处理器的PCIe（Peripheral Component Interconnect Express）的工作模式进行区分，分为RC（Root Complex）模式和EP（Endpoint）模式，如果PCIe工作在主模式，可以扩展外设，则称为RC模式；如果PCIe工作在从模式，则称为EP模式。

表 1-2 PCIe 的工作模式

PCIe工作模式	支持的昇腾产品	工作模式说明
RC模式	Atlas 200 AI加速模块、Atlas 200 DK 开发者套件、Atlas 200I SoC A1、Atlas 500 A2 智能小站、Atlas 200I DK A2 开发者套件、Atlas 200I A2 加速模块	产品的CPU直接运行用户指定的AI业务软件，接入网络摄像头、I <sup>2</sup> C传感器、SPI显示器等其他外挂设备作为从设备接入产品。
EP模式	推理产品：Atlas 500 智能小站、Atlas 200 AI加速模块、Atlas 推理系列产品（配置Ascend 310P AI处理器）	EP模式通常由Host侧作为主端，Device侧作为从端。客户的AI业务程序运行在Host系统中，昇腾产品作为Device系统以PCIe从设备接入Host系统，Host系统通过PCIe通道与Device系统交互，将AI任务加载到Device侧的昇腾AI处理器中运行。
	训练产品：Atlas 训练系列产品、Atlas A2 训练系列产品	

图 1-1 RC 和 EP 场景



在异构计算架构中, 昇腾AI处理器与服务器的CPU通过PCIe总线连接协同工作, 我们一般分别称为Device侧和Host侧, 具体说明如下:

- Host: CPU所在位置称为主机端 ( Host ), 是指与昇腾AI处理器所在硬件设备 ( 如Atlas 300I 推理卡 ) 相连接的X86服务器或ARM服务器, 利用昇腾AI处理器提供的NN ( Neural-Network ) 计算能力完成业务。
- Device: 是指安装了昇腾AI处理器的硬件设备, 利用PCIe接口与服务器连接, 为服务器提供NN计算能力。