

Atlas 300 AI 加速卡

2020.10

一、Atlas 300 AI 加速卡介绍

Atlas 300 卡作为加速卡，通过 PCIe 接口与主机连接。每个 Atlas 300 卡包含 4 个 Ascend 310 处理器，每个 Ascend 310 处理器包含 2 个 Davinci AI Core 和 8 个 A55 ARM Core。即每个 Atlas 300 包含 8 个 Davinci AI Core。

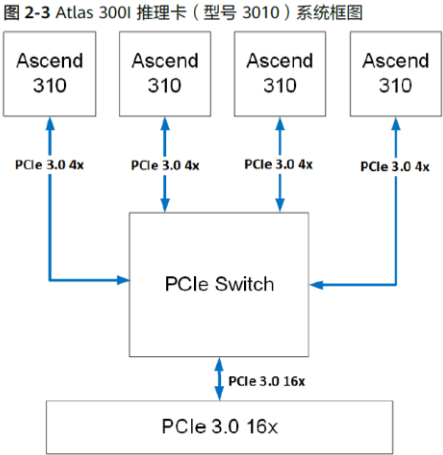


表 4-1 Atlas 300I 推理卡（型号 3010）规格

特征	规格
形态	Low Profile标卡，支持全高和半高两种拉手条
AI处理器	昇腾310 AI处理器，四个 <ul style="list-style-type: none">2个DaVinci AI Core8个A55 Arm Core（最大主频1.6GHz）
AI算力	<ul style="list-style-type: none">半精度（FP16）：44 TFLOPS整数精度（INT8）：88 TOPS
内存	<ul style="list-style-type: none">LPDDR4X容量：32GB（4x8G）位宽：512bit（4x128bit）速率：3200Mbps总带宽：204.8GByte/s（4x51.2GByte/s）支持ECC

Atlas 300 3000 是配合 ARM 服务器使用的，Atlas 300 3010 是配合 X86 服务器使用的。

二、2288H V5 服务器 (X86) + Atlas 300 加速卡 服务器信息：

- IP: 8.44.145.27
(小网地址，需要通过 vpn 接入；没有大网，不能直接连外网，不能直接 apt-get、

pip 等更新软件)

- 登录用户名: deepai, 密码: deepai123
- root 用户密码: deepai123
- HwHiAiUser 用户密码: deepai123 (HwHiAiUser 通过 useradd 创建, 非 adduser, 不建议直接登录系统, 通过 su HwHiAiUser 切换)
- 系统: Ubuntu 16.04.6 Server 版

三、系统驱动和开发环境

1. D 芯片的 npu 驱动: root 用户安装, HwHiAiUser 用户使用, NPU 20.0.0 以下版本

D 芯片信息查看命令 npu-smi info

npu-smi 1.2.3		Version: 1.3.8.B899			
NPU Chip	Name	Health Bus-Id	Power(W) AICore(%)	Temp(C) Memory-Usage(MB)	
175 0	310	OK 0000:B1:00.0	12.8 0	58 2375 / 8192	
175 1	310	OK 0000:B2:00.0	12.8 0	60 2375 / 8192	
175 2	310	OK 0000:B3:00.0	12.8 0	58 2375 / 8192	
175 3	310	OK 0000:B4:00.0	12.8 0	59 2375 / 8192	

详细 npu-smi 工具可参考文档: “Atlas 300I 推理卡用户指南 (型号 3010)”

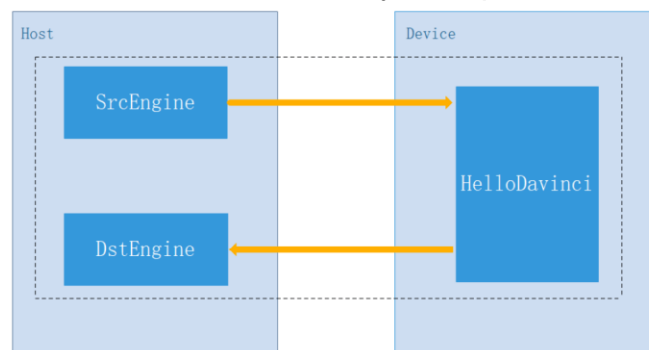
2. D 芯片的 MindSpore DDK (Device Development Kit) 环境

- a) 路径: /home/HwHiAiUser/tools/che/ddk/ddk

```
HwHiAiUser@deepai:~/tools/che/ddk/ddk$ cat ddk_info
{
  "VERSION": "1.3.8.B899",
  "NAME": "DDK",
  "TARGET": "ASIC"
}HwHiAiUser@deepai:~/tools/che/ddk/ddk$
```

3. 运行 demo

- a) 路径: /home/HwHiAiUser/hyb/samples



```
HwHiAiUser@deepai:~/hyb/samples/Samples/HelloDavinci$ ./out/main
Hello Davinci!
The sample end!!
```

HelloDavinci 生成的信息:

```
HwHiAiUser@deepai:~/hyb/samples/Samples/HelloDavinci$ cat out/davinci_log_info.txt  
This message is from HelloDavinci
```

四、模型转换，算子开发

- a) 请参考文档 “Atlas 300 AI 加速卡应用软件开发指南（型号 3010）”

五、参考文档

- [1] Atlas 300I 推理卡用户指南（型号 3010）
- [2] Atlas 300 AI 加速卡应用软件开发指南（型号 3010）