

查询昇腾 AI 处理器型号

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 查询昇腾 AI 处理器型号.....1

1 查询昇腾 AI 处理器型号

昇腾产品通过昇腾AI处理器实现AI能力，用户使用CANN软件的API接口进行模型训练、推理应用开发、算子开发、应用和模型调优时，不同型号的昇腾AI处理器对应的API接口输入或输出参数会有差异，因此本节内容为用户提供查询所用昇腾AI处理器型号的方法，方便用户进行编程或迁移业务时，根据昇腾AI处理器型号配置API参数。

昇腾 AI 处理器型号介绍

昇腾AI处理器包含Ascend 910系列AI处理器、Ascend 310系列AI处理器。

查询方法

使用命令查询昇腾AI处理器型号前，请确保已安装NPU驱动和固件。

在昇腾设备执行**npu-smi info**命令，系统会回显AI处理器型号值，查看“Name”字段对应的AI处理器型号值。

Ascend 910系列AI处理器回显类似如下信息（回显的AI处理器数量以现场昇腾设备实际配置为准）：

```
+-----+
| npu-smi 23.0.rc2                Version: 23.0.rc2                |
+-----+
| NPU  Name      | Health  | Power(W)  Temp(C)  | Hugepages-Usage(page)|
| Chip          | Bus-Id  | AICore(%) | Memory-Usage(MB) | HBM-Usage(MB)      |
+=====+
+-----+
| 2   910A      | OK      | 69.4      44      | 0 / 0      |
| 0      | 0000:B2:00.0 | 0      860 / 15137 | 0 / 32768  |
+=====+
+-----+
| 5   910A      | OK      | 65.2      38      | 0 / 0      |
| 0      | 0000:DA:00.0 | 0      867 / 15137 | 0 / 32768  |
+=====+
+-----+
```

Ascend 310系列AI处理器回显类似如下信息（回显的AI处理器数量以现场昇腾设备实际配置为准）：

```
+-----+
| npu-smi 23.0.rc2                Version: 23.0.rc2                |
+-----+
| NPU  Name      | Health  | Power(W)  Temp(C)  | Hugepages-Usage(page)|
| Chip Device    | Bus-Id  | AICore(%) | Memory-Usage(MB)      |
+=====+
```

```
+-----+
| 0  310          | OK      | 12.8   57      969 / 969      |
| 0  0            | 0000:03:00.0 | 0      5669 / 7759      |
+-----+
```

如表1-1所示，同一系列的AI处理器根据硬件规格的差异对应不同的AI处理器型号值，如Ascend 910*和Ascend 910B*都为Ascend 910 系列AI处理器，但是硬件规格有差异，因此对应的CANN软件API接口输入或输出参数会有差异。为方便用户区分API参数的设置，后续文档描述API参数值时以AI处理器规格区分不同的参数值设置，如图1-1所示。

表 1-1 芯片规格分类

AI处理器型号值分类	AI处理器规格分类	AI处理器系列
910后带字母：910A、910B、910ProA、910ProB、910PremiumA	Ascend 910（后续文档中以昇腾910 AI处理器表示）	Ascend 910 系列AI处理器
910B后带数字：910B1、910B2、910B3、910B4	Ascend 910B（后续文档中以昇腾910B AI处理器表示）	
310	Ascend 310（后续文档中以昇腾310 AI处理器表示）	Ascend 310 系列AI处理器
310B后带数字：310B1、310B4	Ascend 310B（后续文档中以昇腾310B AI处理器表示）	
310P后带数字：310P1、310P3	Ascend 310P（后续文档中以昇腾310P AI处理器表示）	

图 1-1 不同规格 AI 处理器对应的参数配置示例

参数名	输入/输出	描述
gm	输入	A矩阵在Global Memory上的首地址。 昇腾910 AI处理器，支持的数据类型为：float 昇腾310P AI处理器AI Core，支持的数据类型为：float 昇腾910B AI处理器，支持的数据类型为：half/float