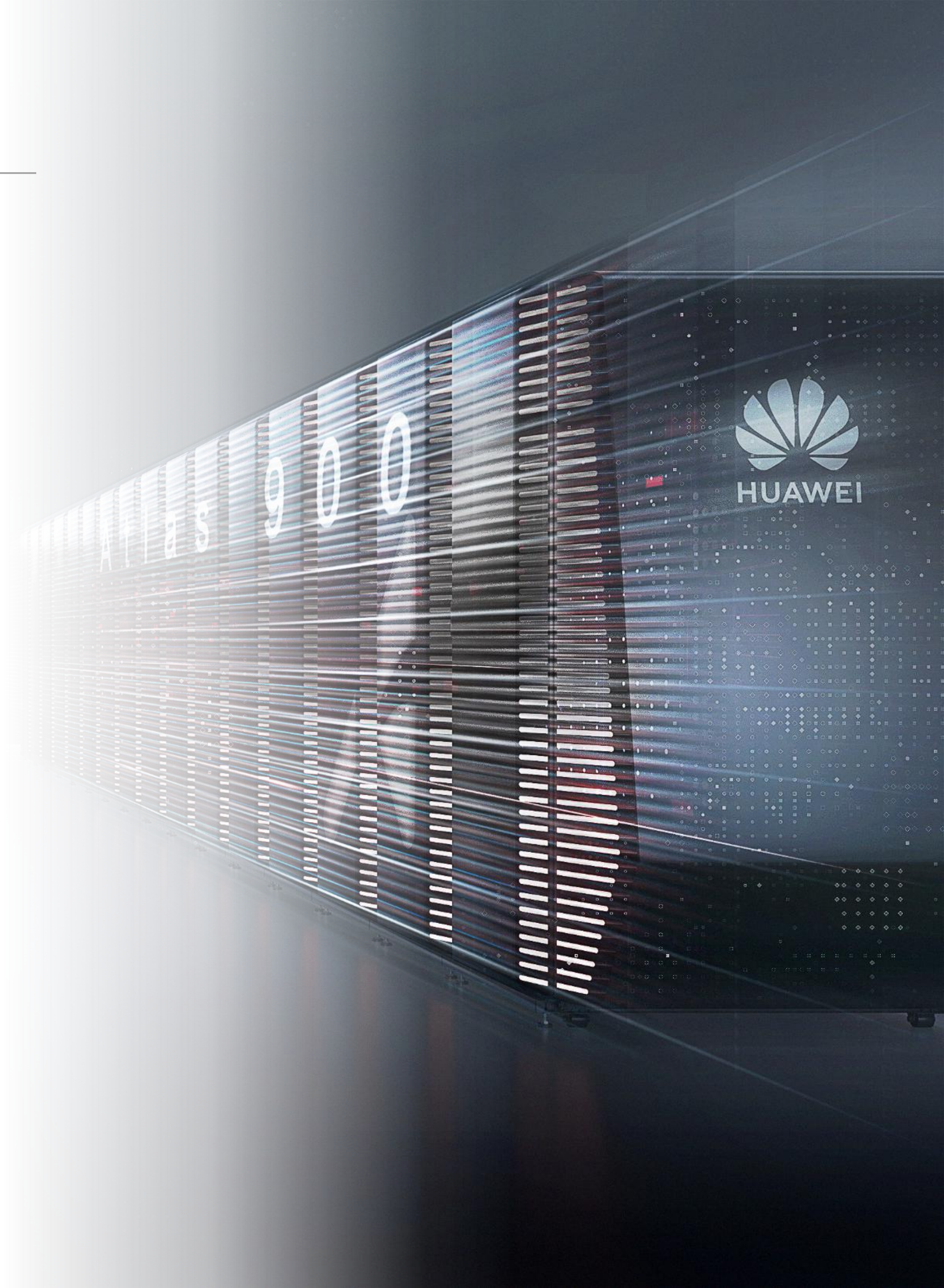
AscendCL单算子计算动态特性

HUAWEI

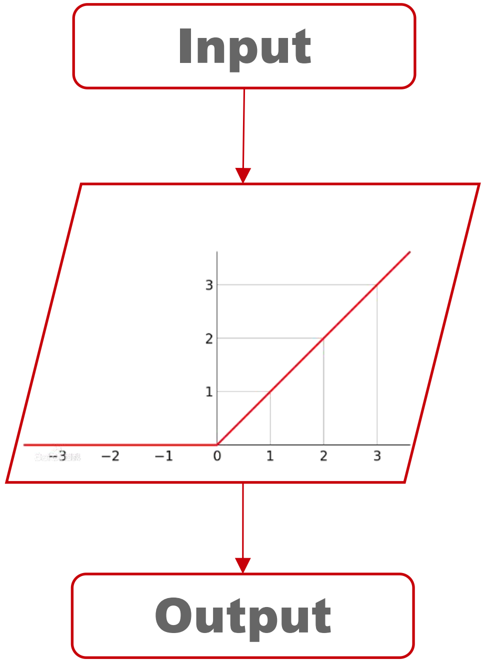
课程目标

- 学完本课程，您应该能够：
  - 掌握AscendCL单算子计算动态特性
- 为了达成上述目标，您应该具备如下知识：
  - 熟练的C/C++语言编程能力
  - ATC模型转换工具用法（单算子Json文件转换成离线模型）
  - AscendCL基础功能

目 录 TABLE OF CONTENTS



复习/思考



开发前准备：

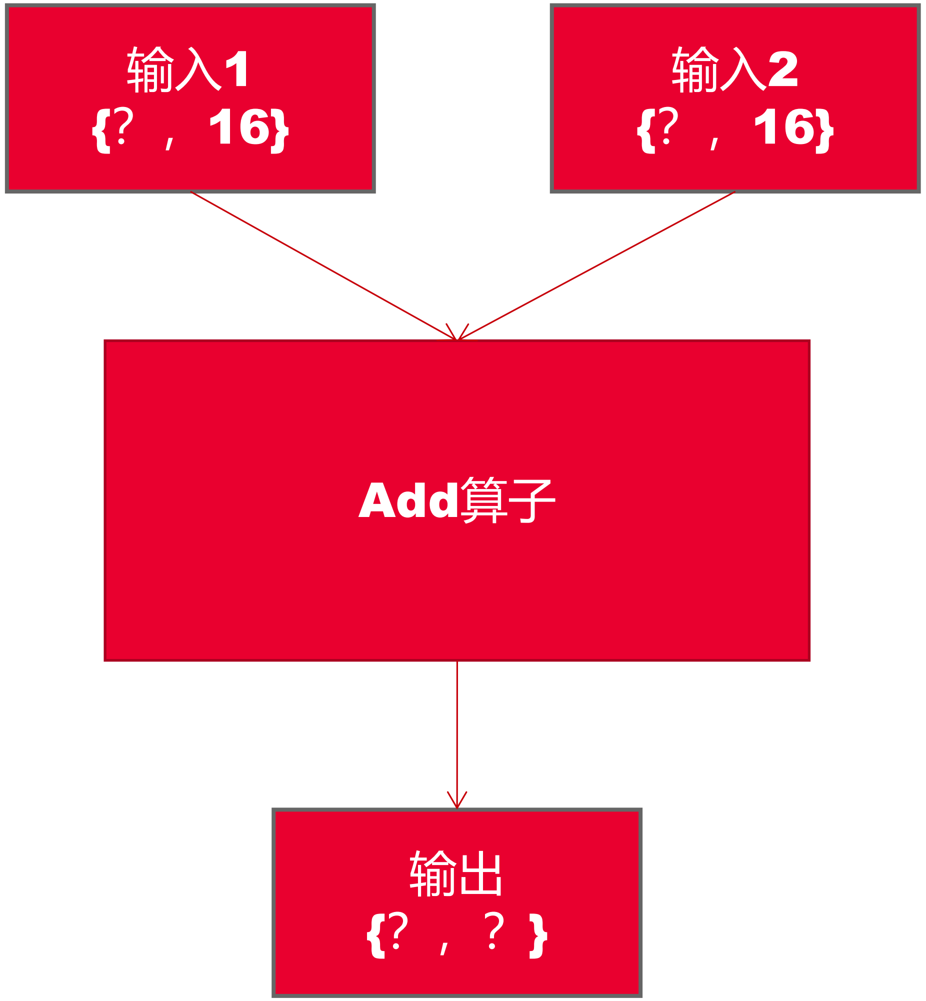
- 用atc工具将单算子Json配置文件转换成.om文件

开发中：

- 初始化
- 加载单算子om文件
- 创建运行资源
- 创建计算所需输入输出（固定Shape）
- 单算子计算
- 检查结果
- 销毁运行资源
- 去初始化

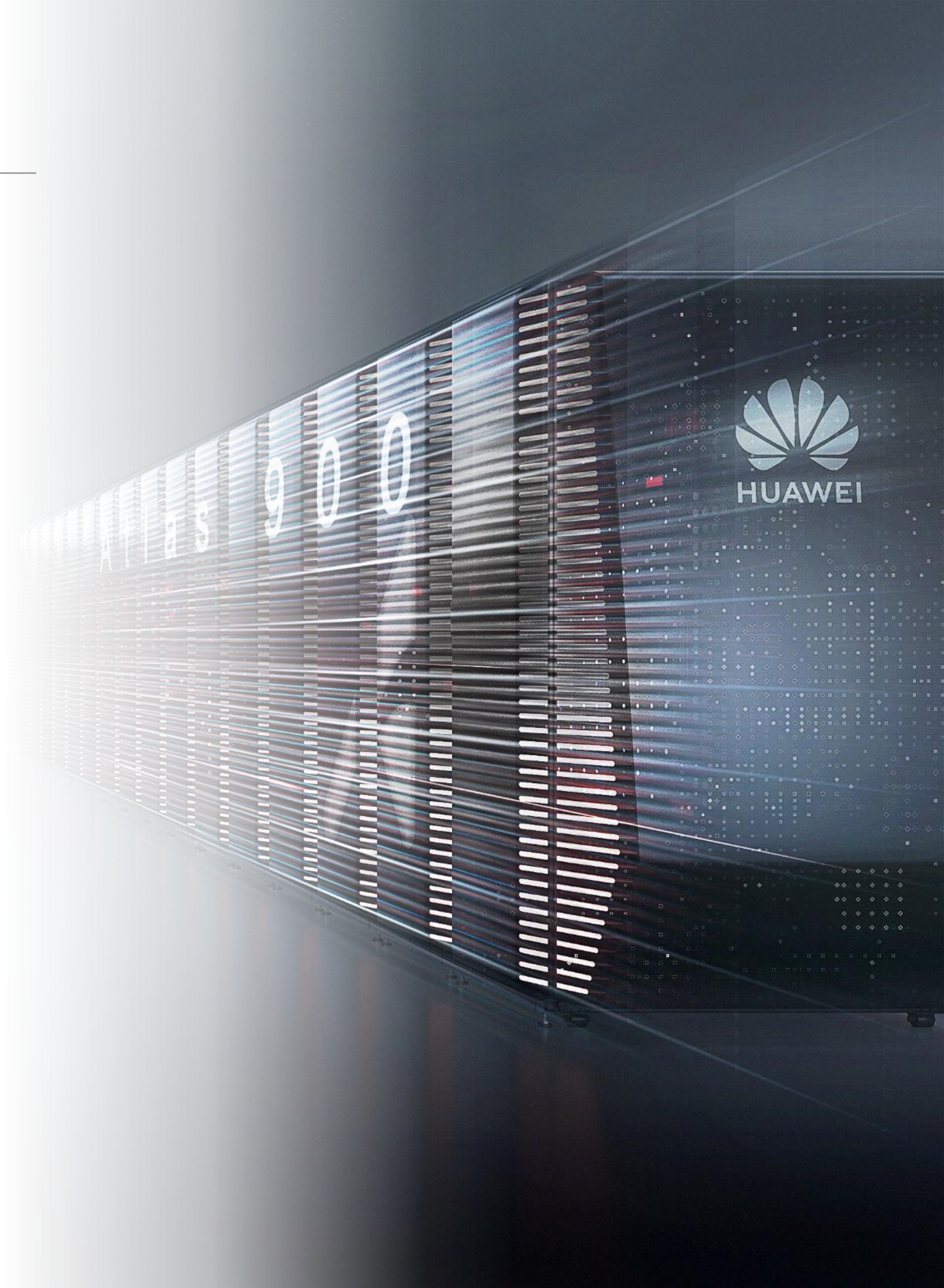
思考：
动态**Shape**场景下，
不知道输出张量形状，
无法为其创建内存，
怎么办？

算子动态shape概述



对于支持动态Shape的算子，无法明确算子的输出Shape时，需要用户调用配合调用**aclopInferShape**接口、**aclGetTensorDescNumDims**接口、**aclGetTensorDescDimV2**接口、**aclGetTensorDescDimRange**等接口，推导或预估算子的输出Shape，作为算子执行接口**aclopExecuteV2**的输入。

目 录 TABLE OF CONTENTS



动态shape情形1

```
[
  {
    "op": "Add",
    "input_desc": [
      {
        "format": "ND",
        "shape": [-1,16],
        "shape_range": [[0, 32]],
        "type": "int64"
      },
      {
        "format": "ND",
        "shape": [-1,16],
        "shape_range": [[0, 32]],
        "type": "int64"
      }
    ],
    "output_desc": [
      {
        "format": "ND",
        "shape": [-1,16],
        "shape_range": [[0, 32]],
        "type": "int64"
      }
    ]
  }
]
```

模型编译时不指定shape，模型执行时根据输入固定shape，能推导出具体输出shape。

此场景下，可调用**aclopInferShape**接口来计算输出数据的形状。

函数原型

aclError **aclopInferShape(const char *opType,** 算子类型

int numInputs, 输入个数

aclTensorDesc ***inputDesc[],** 输入Tensor描述

aclDataBuffer ***inputs[],** 输入数据的DataBuffer（注意要用Host内存来创建）

int numOutputs, 输出个数

aclTensorDesc ***outputDesc[],** 输出Tensor描述

aclopAttr ***attr)** 算子属性

HUAWEI

动态shape情形2


```
[
  {
    "op": "TopK",
    "input_desc": [
      {
        "format": "ND",
        "shape": [-1],
        "shape_range": [[1,-1]],
        "type": "int32"
      },
      {
        "format": "ND",
        "shape": [],
        "type": "int32"
      }
    ],
    "output_desc": [
      {
        "format": "ND",
        "shape": [-1],
        "shape_range": [[1,-1]],
        "type": "int32"
      },
        {
        "format": "ND",
        "shape": [-1],
        "shape_range": [[1,-1]],
        "type": "int32"
      }],
      "attr": [
      {
        "name": "sorted",
        "type": "bool",
        "value": true
      }
    ]
  }
]
```


模型编译时不指定shape，模型执行时根据输入固定shape和常量，能推导出具体输出shape。

此的场景下，调用**aclopInferShape**接口、**aclGetTensorDescNumDims**接口（获取Tensor描述的维度数量）、**aclGetTensorDescDimV2**接口（获取Tensor描述中，指定维的长度）、**aclGetTensorDescDimRange**接口（获取Tensor描述中，指定维的长度范围）等，推导或预估算子的输出Shape，作为算子执行接口aclopExecuteV2的输入。

动态shape情形3

```
[
  {
    "op": "Where",
    "input_desc": [
      {
        "format": "ND",
        "shape": [-1],
        "shape_range": [[1,-1]],
        "type": "int32"
      }
    ],
    "output_desc": [
      {
        "format": "ND",
        "shape": [-1, 1],
        "shape_range": [[1,-1]],
        "type": "int64"
      }
    ]
  }
]
```

模型编译时不指定shape，模型执行时根据输入固定Shape，无法得到算子的准确输出Shape，但可以得到输出Shape的范围。

此的场景下，调用**aclopInferShape**接口、**aclGetTensorDescNumDims**接口（获取Tensor描述的维度数量）、**aclGetTensorDescDimV2**接口（获取Tensor描述中，指定维的长度）、**aclGetTensorDescDimRange**接口（获取Tensor描述中，指定维的长度范围）等，推导或预估算子的输出Shape，作为算子执行接口aclopExecuteV2的输入。

代码示例

acl_exec
ute_add_dynamiczip

Thank you.

昇腾开发者社区

http://ascend.huawei.com

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.