

# UCloud下一代AI as a Service平台

宋翔

UCloud 创新产品线

A decorative abstract graphic is positioned at the bottom of the slide. It features several overlapping, rounded shapes in shades of blue and purple, resembling stylized leaves or petals. Thin, light blue lines radiate from the center of the graphic, extending towards the corners of the slide.

### 强大的计算能力



### 更好的算法

- DNN
  - CNN
  - RNN
  - LSTM
  - GAN
  - ...
- Value network
  - Policy network
  - Sequence-to-sequence
  - CDNN
  - AlexNet
  - Inception
  - Fast RCNN
  - ...



### 更好的软件平台



### 丰富的应用场景

**自动驾驶**

- TESLA
- UBER
- WAYMO
- UISEE 驭势

**AlphaGo**

- AlphaGo

**智能推荐**

- Google
- 头条 今日头条
- Baidu 百度

**自然语言**

- Yi 追一科技
- 思必驰 AISPEECH
- 科大讯飞 IFLYTEK
- 云知声 Unisound
- 小i机器人 www.xisoi.com

**计算机视觉**

- TUPU 图普科技
- FACE++
- SENSETIME 商汤科技

**智能医疗**

- enlitic
- iCarbonX 医云智能

**智能家居**

- amazon echo
- Google HOME

# AI 产品开发周期

收集数据

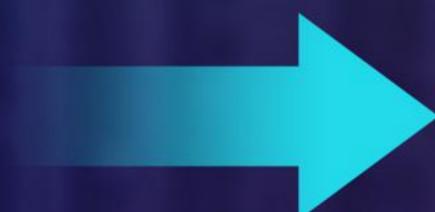
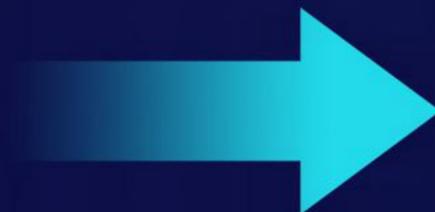
大数据收集

数据建模+训练

AI模型训练

AI模型部署

AI在线服务



## 大数据收集

计算

UHost  
UHadoop  
UDDP

存储

UFile  
UDisk  
UFS

网络

VPC  
VDPN

## AI模型训练

GPU云主机、物理机

Nvidia  
K80

Nvidia  
P40

## AI在线服务

UCloud AI as a Service

UCloud 基础架构平台

# UCloud AI as a Service

why

what

how

# Why UCloud AI as a Service?

# AI在线服务任务的挑战



# UCloud AI as a Service平台

按需资源分配  
自动负载均衡

成本

UCloud  
AI as a  
Service

可靠性

全系统分布式化  
计算节点自动容灾

资源自动扩容

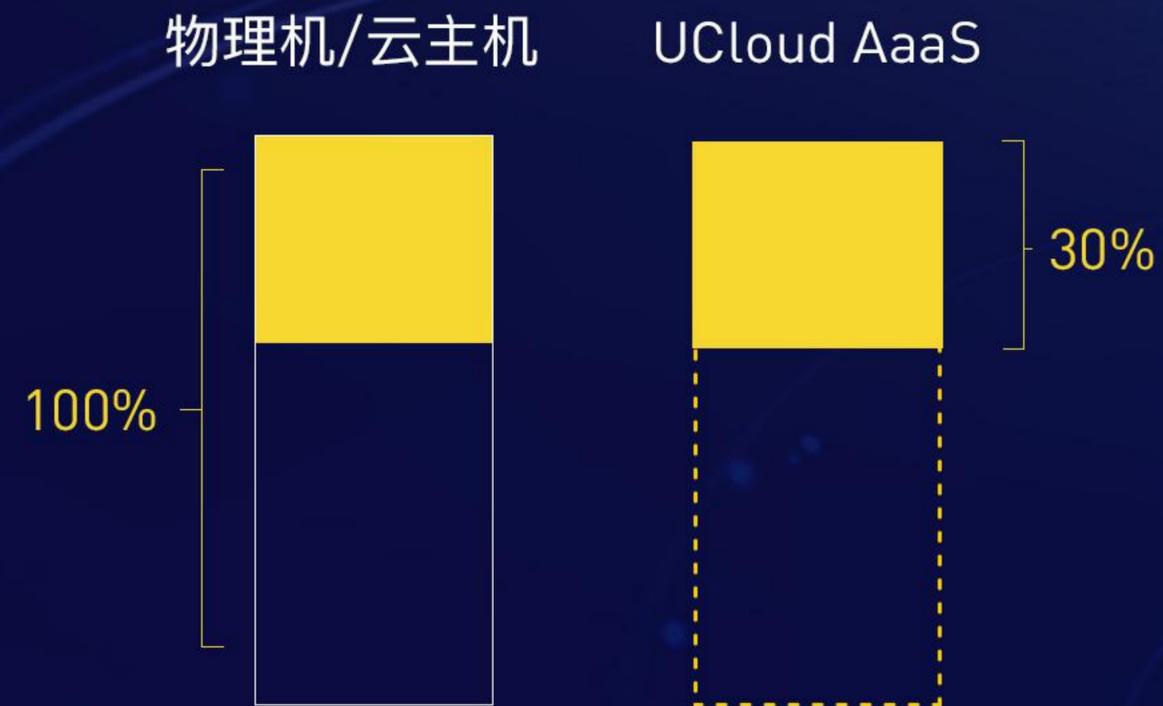
弹性

易用性

兼容各类AI框架  
便捷的任务管理

海量计算资源

# 成本

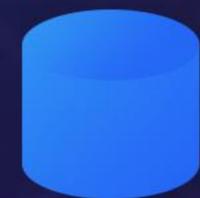


# 运维管理

	自建集群	UCloud AaaS
负载均衡	×	○
自动扩容	×	○
分布式容灾	×	○

# 资源

自建集群



UCloud AaaS



# What is UCloud AI as a Service?

# AI 产品周期中的UCloud AI as a Service平台





# UCloud AI as a Service 多AI框架支持

TensorFlow SDK

Inference 代码

模型文件

Keras SDK

Inference 代码

模型文件

Caffe SDK

Inference 代码

模型文件

MXNet SDK

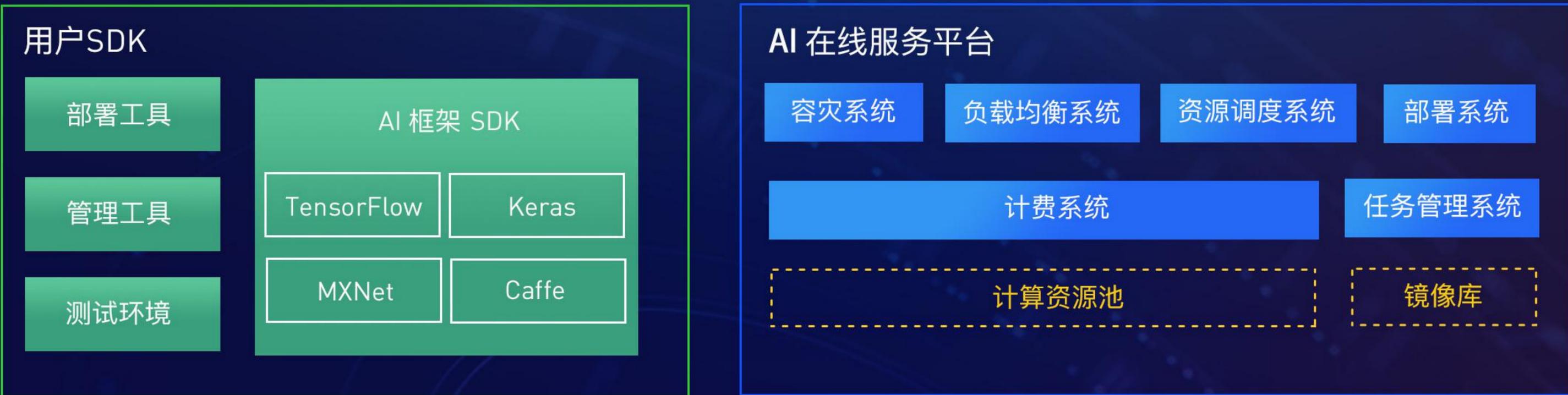
Inference 代码

模型文件

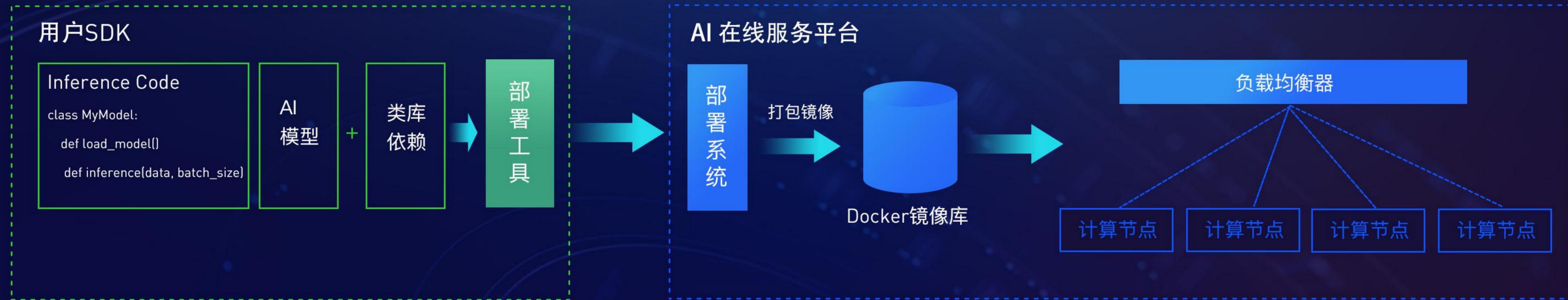
UCloud AI as a Service 平台

How UCloud AI as a Service work?

# AI as a Service 系统架构



# AI as a Service任务部署流程



# AI在线服务平台负载均衡



# AI在线服务平台资源调度



# AI在线服务平台自动容灾



# AI在线服务平台任务管理

## 任务控制

任务创建

任务删除

模型更新

## 任务状态管理

任务A

任务当前状态

任务执行时间

AI模型版本

QPS

## A/B 测试管理

任务A URL

80%

模型 A.alpha 服务

20%

模型 A.beta 服务

# AI as a Service 平台案例Demo

手写数字识别MNIST



MNIST Data

<http://yann.lecun.com/exdb/mnist/>

MNIST TensorFlow

[https://www.tensorflow.org/get\\_started/mnist/beginners](https://www.tensorflow.org/get_started/mnist/beginners)

MNIST Inference

MNIST 模型: `mnist.mod`  
MNIST代码: `MnistService:`  
`load_model()`  
`inference()`



AI 在线  
服务部署

### 训练

Input数据: <http://yann.lecun.com/exdb/mnist/>

模型输出网络定义:

```

输入x, y_
y = x * W + b
y_ = softmax(y)
    
```

MNIST 模型: `mnist.mod`

### 服务代码编写

```

def load_mode(self)
    y = x * W + b
    y_ = softmax(y)
    saver.restore(sess, mnist.mod)
    
```

加载模型

```

def execute(self, data, batch_size)
    input = convert(data, batch_size)
    predict = sess.run(y_, x = images)
    return predict
    
```

数据预处理

模型计算

### 部署

`python deploy_tool.py`

`--model_dir=./checkpoint_dir \`

模型

`--code_files=mnist.py \`  
`--main_file=mnist \`

代码

`--docker_dependency_apt= xxx \`  
`--docker_dependency_pip=xxx \`  
`--docker_files=xxx \`

依赖

`--request_url=http://api.ucloud.cn:80`



# 如何获取UCloud AI as a Service 服务信息

联系

您的销售经理  
或架构师

访问

我们的产品页面



THANKS!

