

# 微博大规模机器学习平台

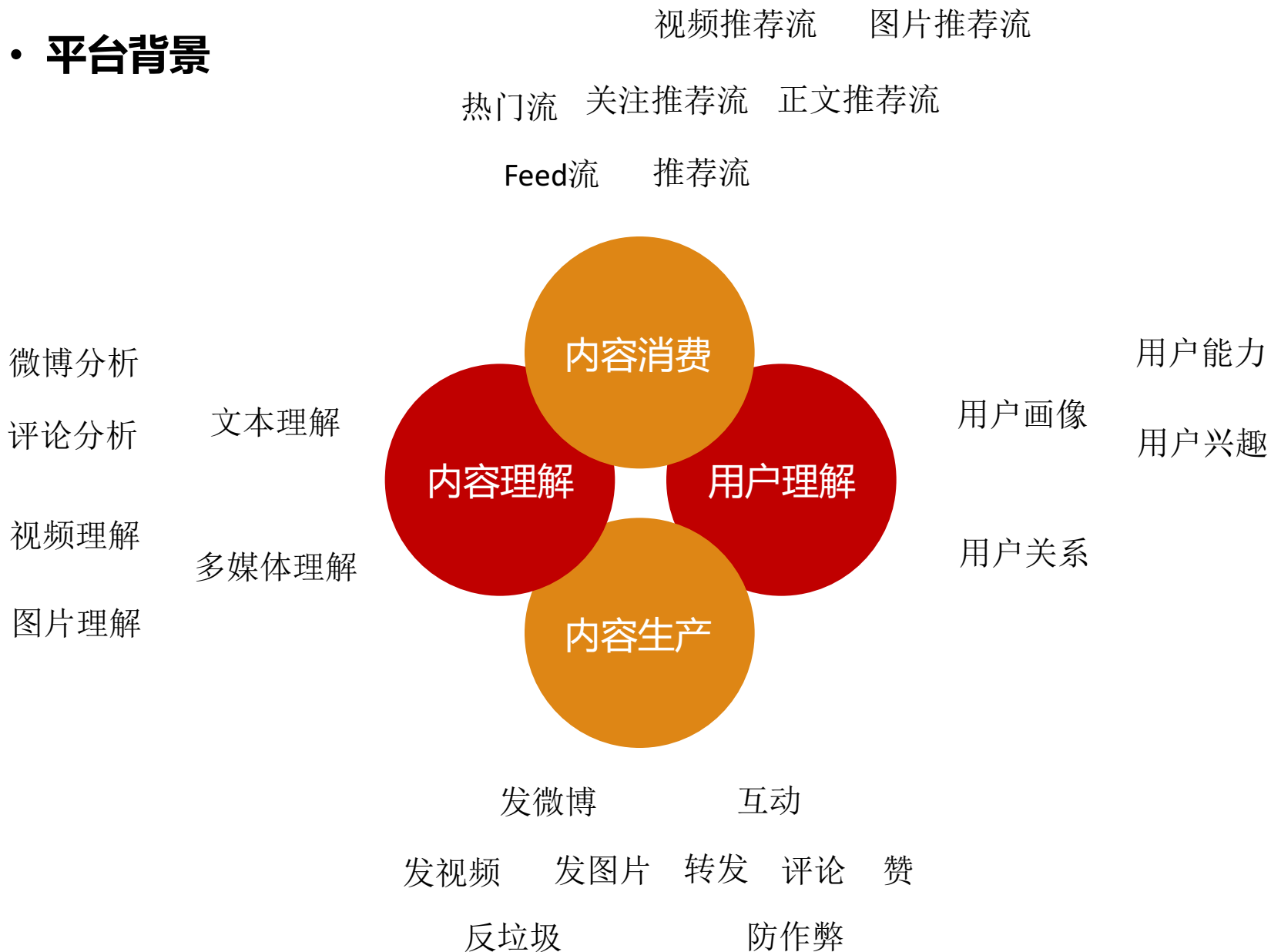
黄波  
资深架构师  
微博机器学习平台负责人  
2018.10



1. 平台背景
2. 平台架构
3. 平台架构-算法平台架构
4. 平台架构-数据平台架构
5. 平台应用
6. 平台效果

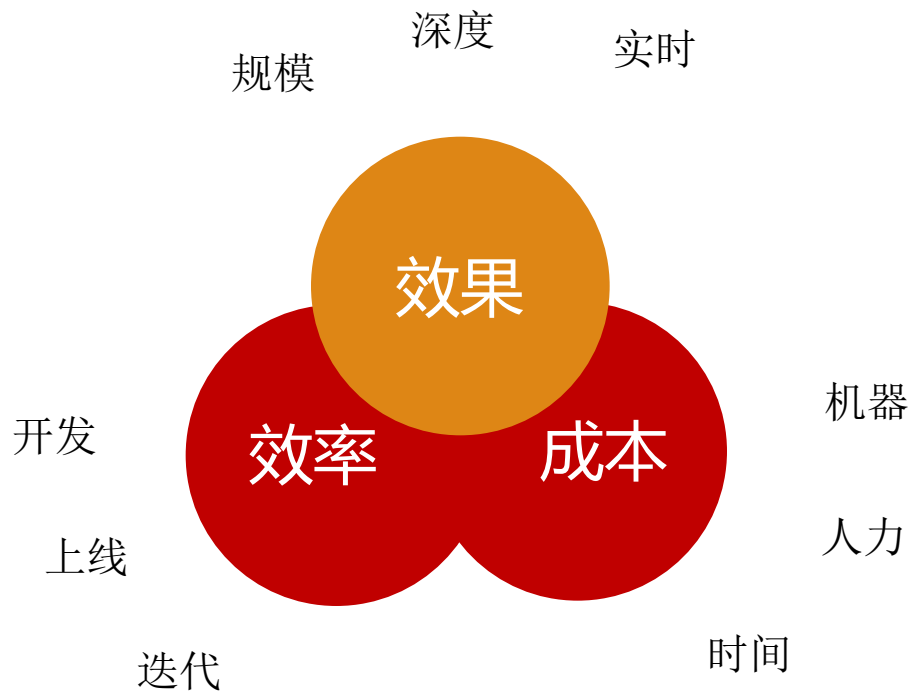
# 1 平台背景

## • 平台背景



# 1 平台背景

- 平台背景



## 2 平台架构



## 2 平台架构

### 业务

推荐/排序

文本分类/检测

图像/视频分类

反垃圾/反作弊

.....

### 机器学习平台

#### 数据 ( WeiData )

数据服务

数据存储

数据计算

#### 算法 ( WeiFlow )

模型服务

模型库

模型训练/评估

样本库

#### 引擎 ( WeiEngine )

推荐引擎

控制台

开发套件

#### 计算

Hadoop/Spark

Storm/Flink

TensorFlow /Caffe

.....

#### 存储

Hdfs/Odps

PS/WeiPS

#### 调度

Yarn/Mesos/K8s

Docker

#### 集群

数据仓库集群

实时计算集群

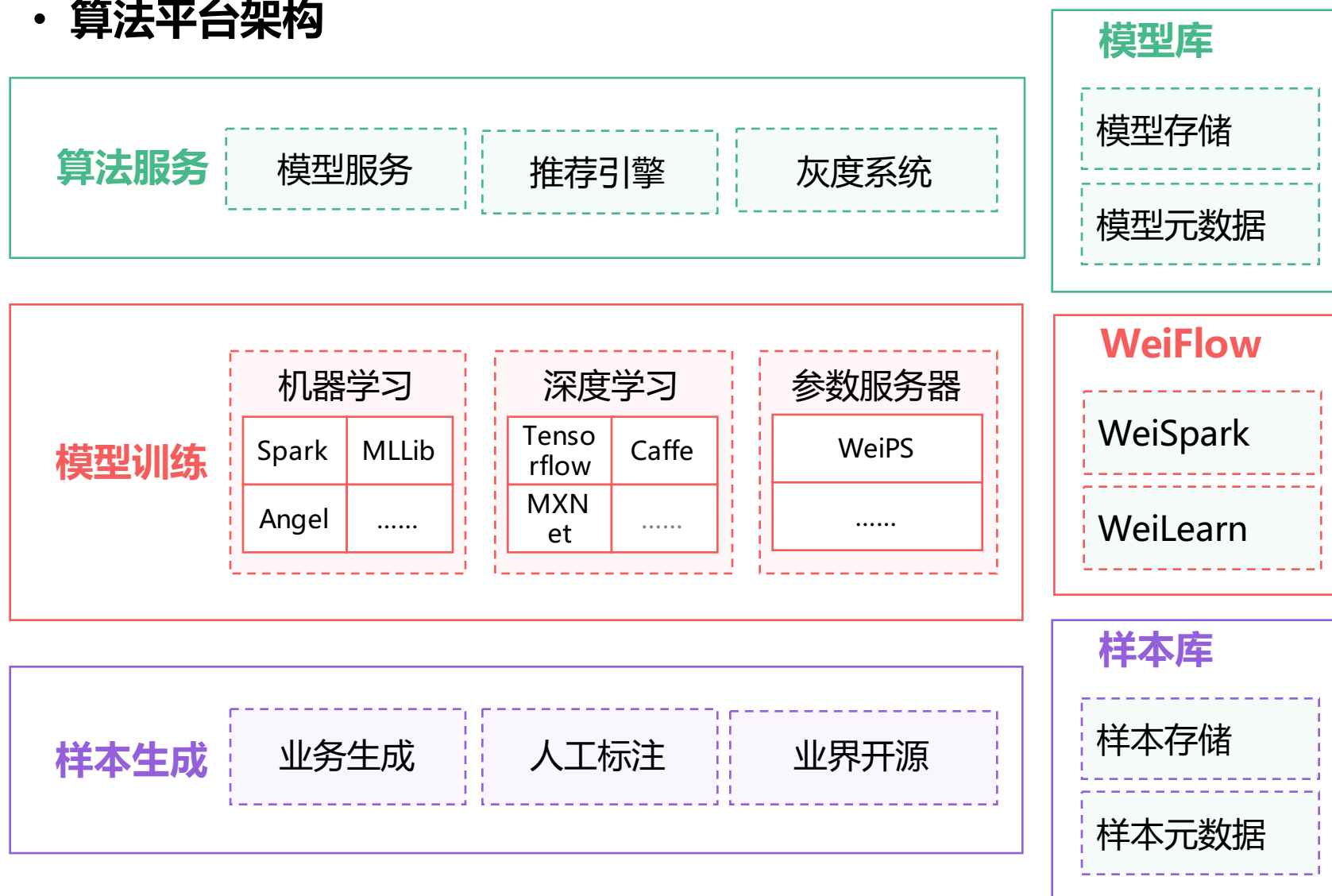
阿里云计算集群

高性能GPU集群

.....

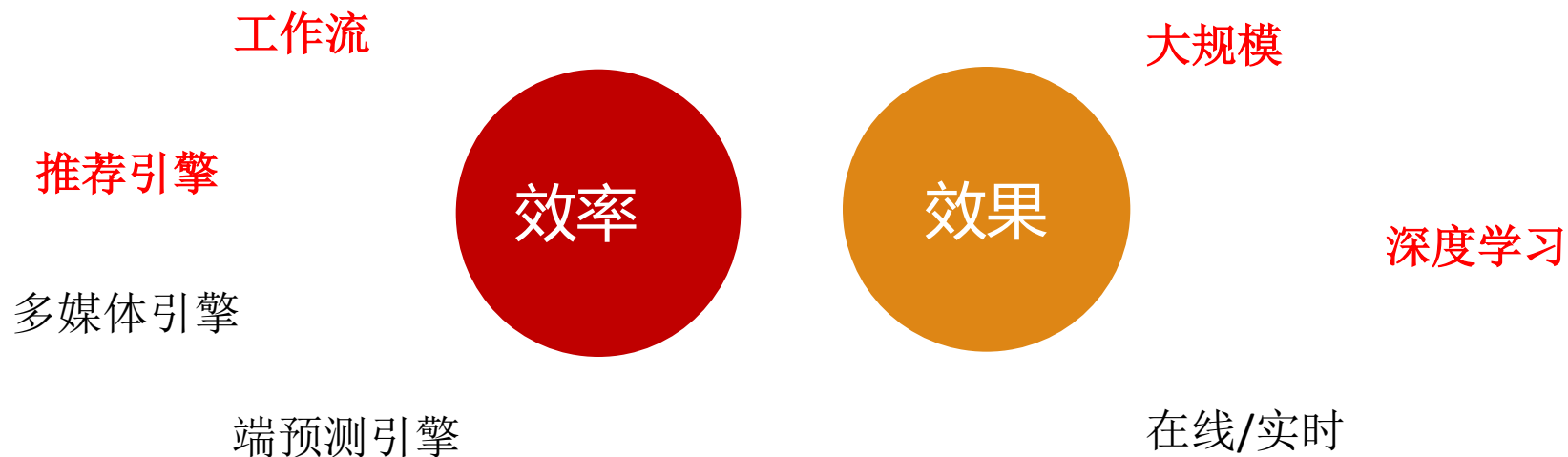
### 3 平台架构-算法平台架构

#### • 算法平台架构



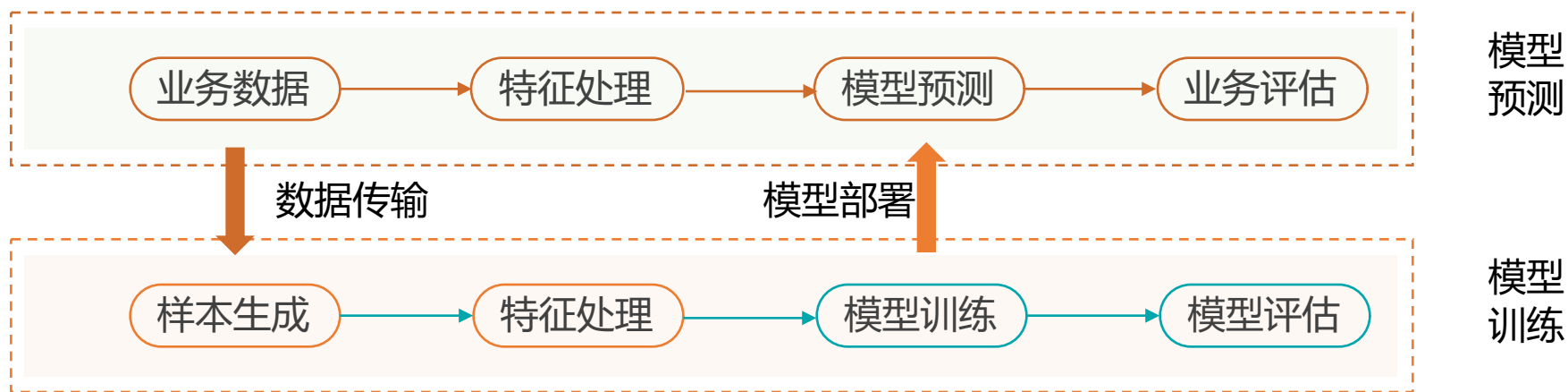
### 3 平台架构-算法平台架构

- 算法平台架构





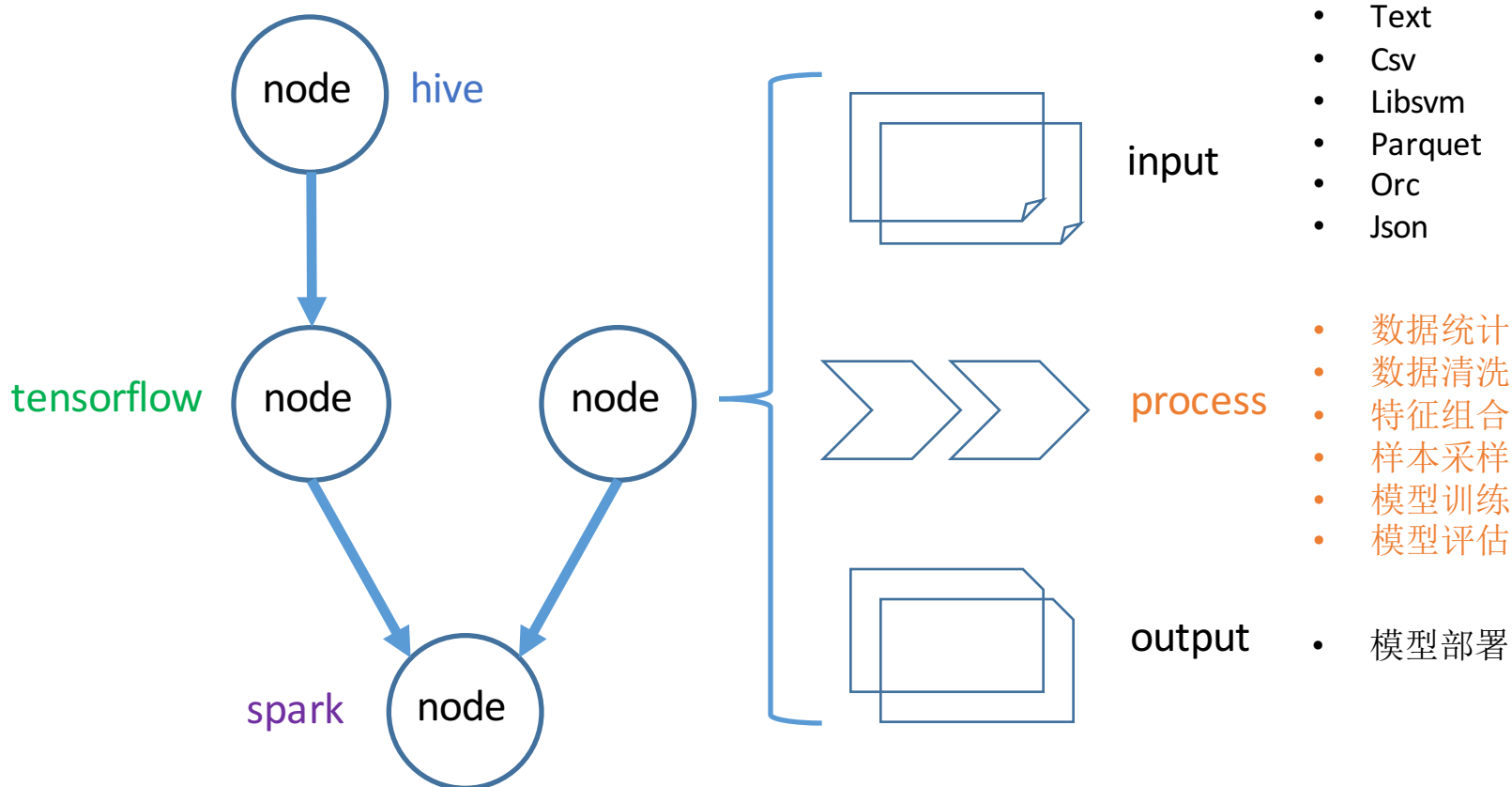
## 3.1 平台架构-算法平台架构-机器学习 workflows



## 3.1 平台架构-算法平台架构-机器学习 workflows

### ➤ 机器学习工作流框架WeiFlow

- 统一数据计算框架
- 双层DAG，兼容异构环境



## 3.2 平台架构-算法平台架构-推荐引擎

### • 推荐引擎

#### 业务逻辑系统

输出逻辑

策略逻辑

模型逻辑

物料逻辑

#### 灰度实验系统

控制台

参数配置

流量分配

审批上线

#### 数据分析系统

报表数据

业务曝光、互动  
长期趋势

实验数据

实验效果数据对比  
天级更新

#### 物料服务

物料管理

物料统一规范  
统一管理 统一监控

物料服务

统一入库  
线上服务

#### 特征服务

特征数据

微博、用户  
关系、环境

分发控制

较feed更严格  
质量控制

#### 模型服务

统一框架

统一输入输出规范  
统一上线流程

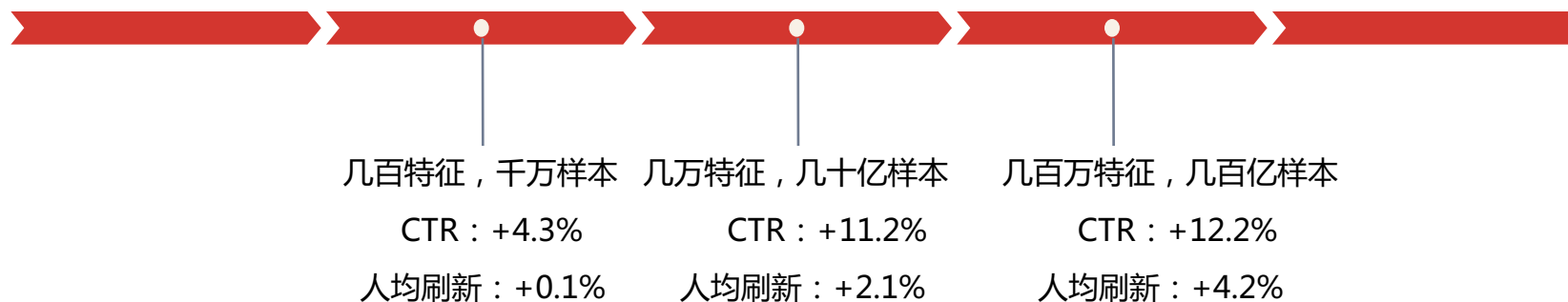
多模型

多样化模型支持

## 3.3 平台架构-算法平台架构-大规模机器学习

### ➤ 大规模机器学习

- 实践证明：机器学习规模越大，效果越好



### ➤ 规模：几千亿样本、几十亿参数

- 参数服务器WeiPS：解决样本和特征规模化问题

## 3.3 平台架构-算法平台架构-深度学习

### • 深度学习

业务

用户向量

内容向量

文本分类

图像/视频分类

.....

### 深度学习平台

样本库

样本库

样本标注

模型训练

模型评估

模型训练

模型预测

模型预测

模型库

端预测

端预测

端优化

多媒体流

多媒体流

媒体特征库

计算

Hadoop/Spark

TensorFlow /Caffe

存储

Hdfs/Odps

PS/WeiPS

调度

Yarn/K8s

Docker

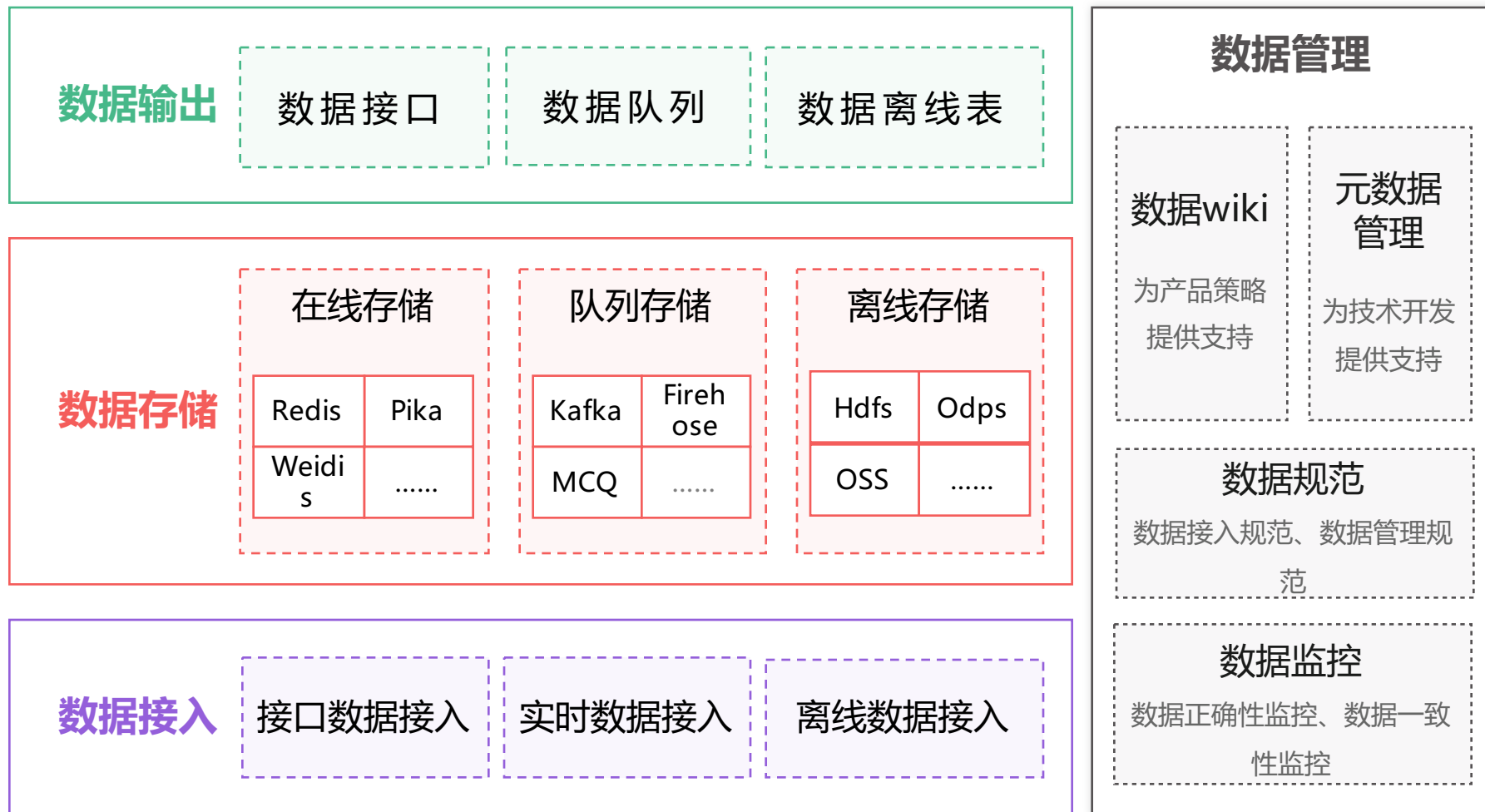
集群

阿里云计算集群

高性能GPU集群

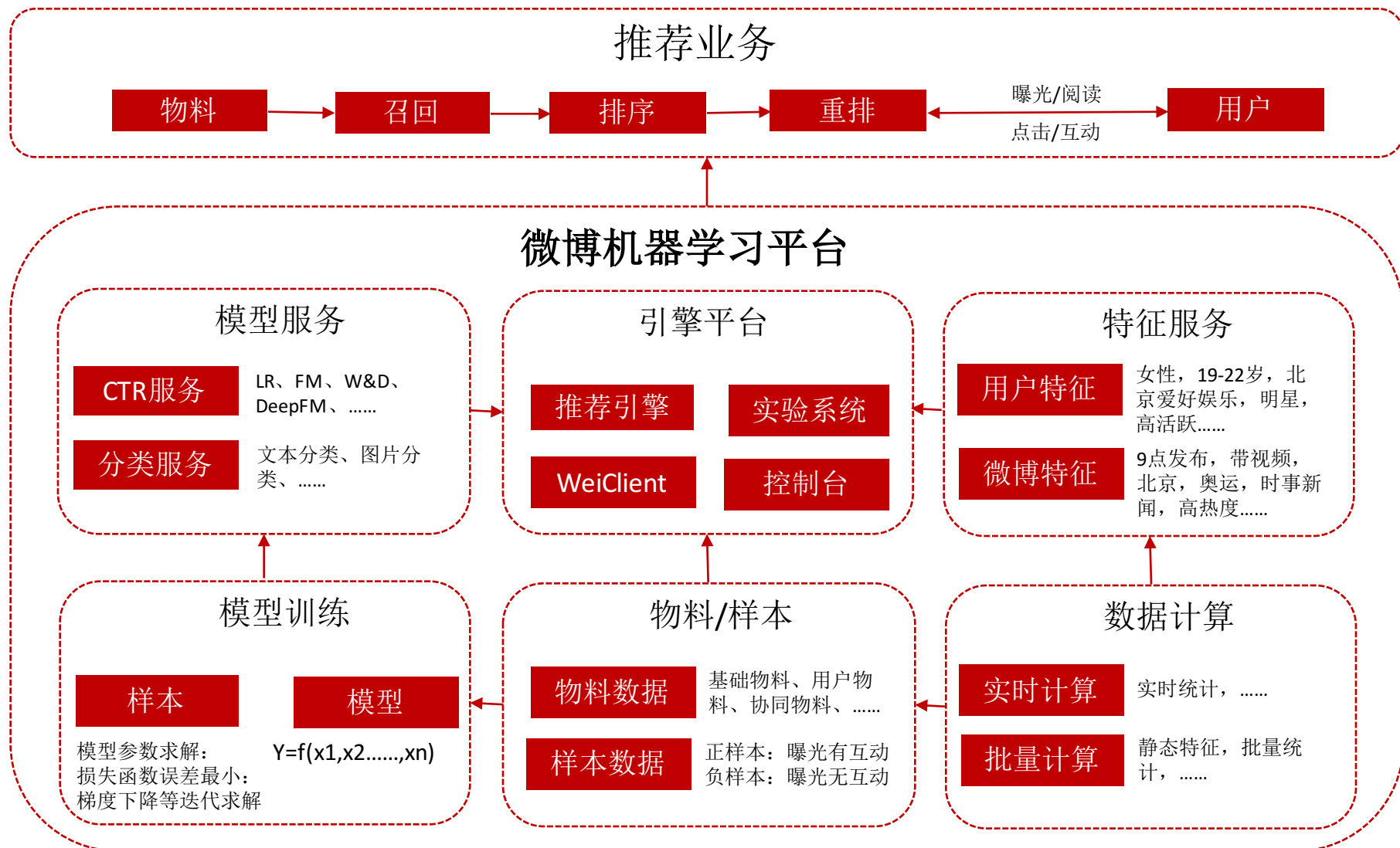
## 4 平台架构-数据平台架构

### • 数据平台架构



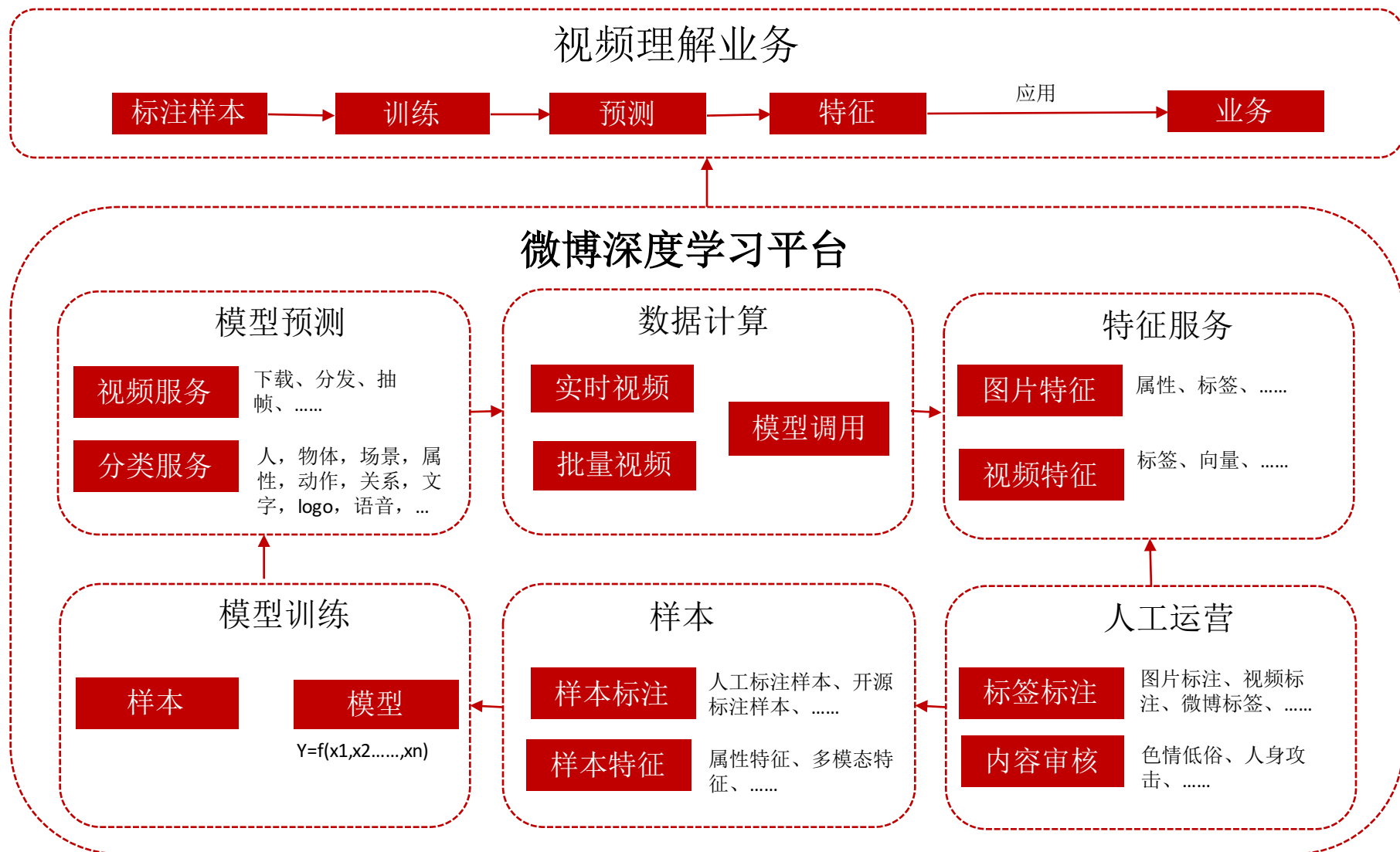
- 推荐应用
- 视频理解应用

## 5.1 平台应用-推荐应用



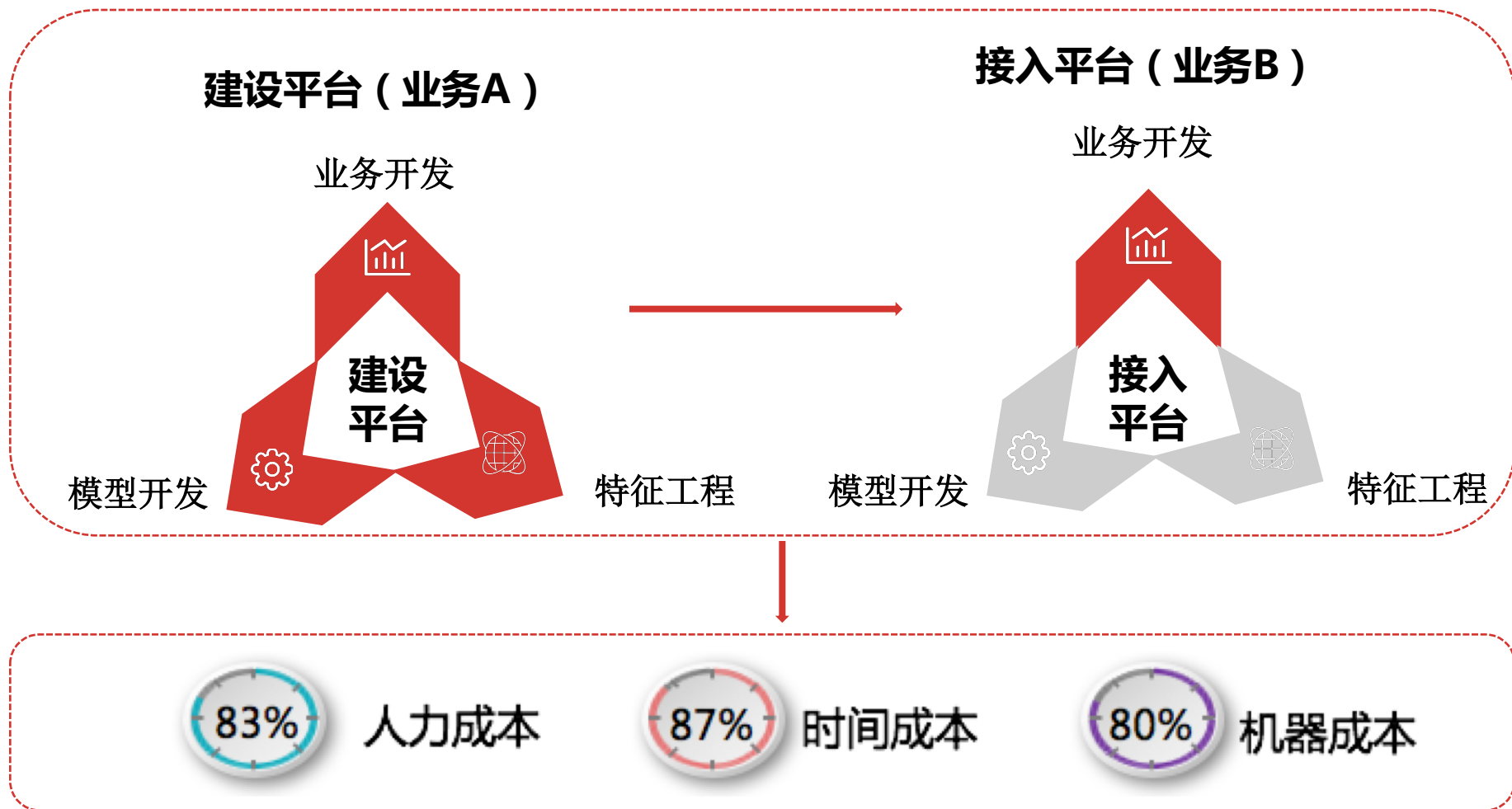


## 5.2 平台应用-视频理解应用

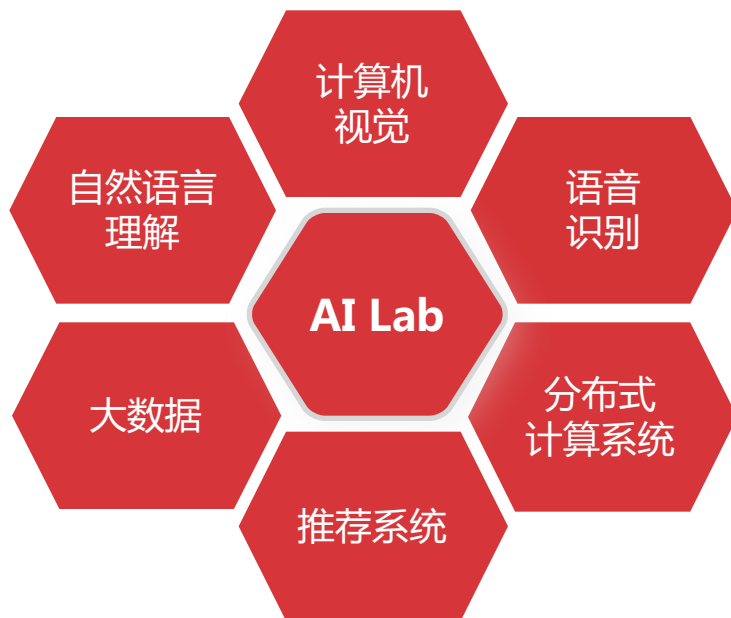


## 6 平台效果

### • 成本/效率



# 人工智能在社交媒体领域大有可为！



该二维码7天内(11月9日前)有效，重新进入将更新

地点：北美-硅谷、北京、杭州

yuqian8@staff.weibo.com

**以微博之力，让世界更美**

**感谢聆听！**