



2016 杭州·云栖大会  
THE COMPUTING CONFERENCE

云栖社区  
yq.aliyun.com

# PAI分布式机器学习平台

九丰  
阿里云数据事业部技术专家

2016  
The Computing Conference

主办单位:



战略合作伙伴:

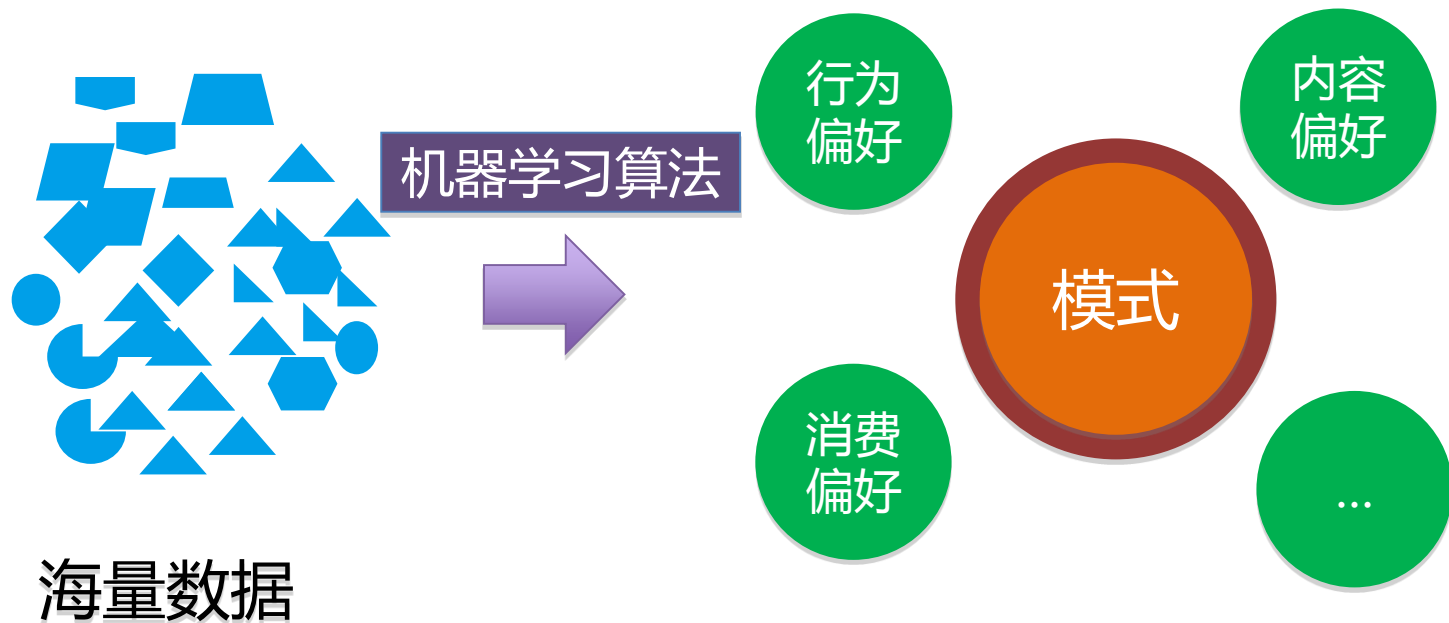


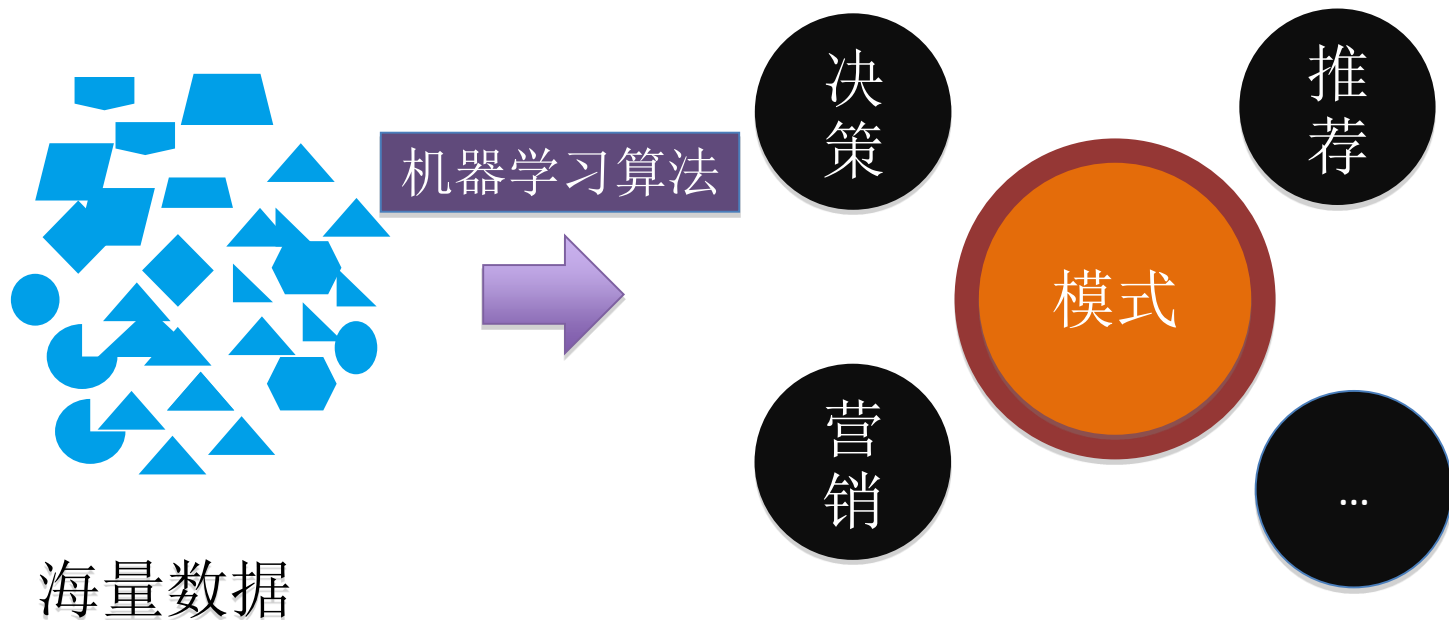
扫码观看大会视频



# 大数据时代 如何让我们的服务与众不同？







## PAI分布式机器学习平台

SQL

MR

.....

机器学习

## MaxCompute统一计算引擎

计算节点

计算节点

计算节点

计算节点

计算集群

计算节点

计算节点

计算节点

计算节点

计算集群

计算节点

计算节点

计算节点

计算节点

计算集群



扫码观看大会视频

## 产品特性

- 1 | 一站式机器学习服务，集成业内前沿算法
- 2 | 多种友好使用接口，提供拖拽式WEB操作界面
- 3 | 多种存储服务支持
- 4 | 模型管理
- 5 | 在线预测服务，模型一键发布



# 一站式机器学习服务

数据清洗

特征工程

离线训练

在线预测

- 加全采样
- 随机采样
- 分层采样
- 过滤和映射
- 拆分
- 归一化
- 标准化
- .....

- 主成分分析
- 特征离散
- 偏好计算
- 特征尺度变换
- 窗口变量统计
- .....

- 分类算法
- 聚类算法
- 回归算法
- 文本分析
- 网络分析
- .....

- 一键部署
- 负载均衡
- 灰度发布
- 轮转升级
- 自动流控
- .....



## 多种使用形式

Web界面

PyODPS

Console



```
from odps.ml.classifiers import RandomForest
```

```
df = iris.roles(lable="name")  
train, test = df.split(0.6)
```

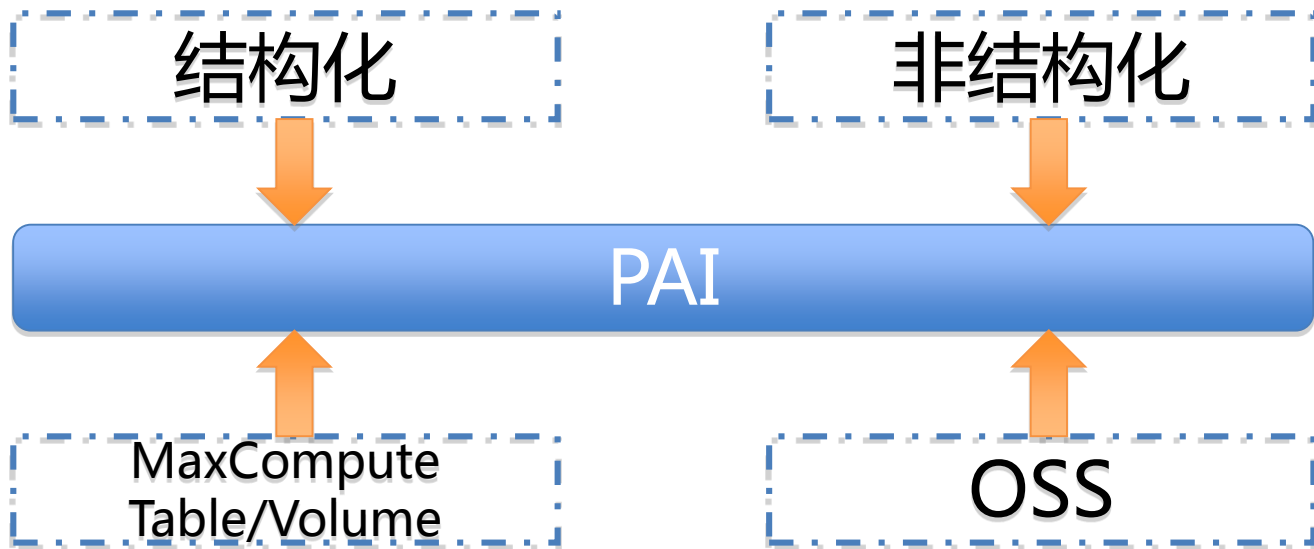
```
model = RandomForest(tree_num=1)  
train  
predic
```

1. pai -name randomforests
2. -DmodelName=rf\_model
3. -DinputTableName=wpbc
4. -DlabelColName=label
5. -DtreeNum=1000;





## 存储服务支持



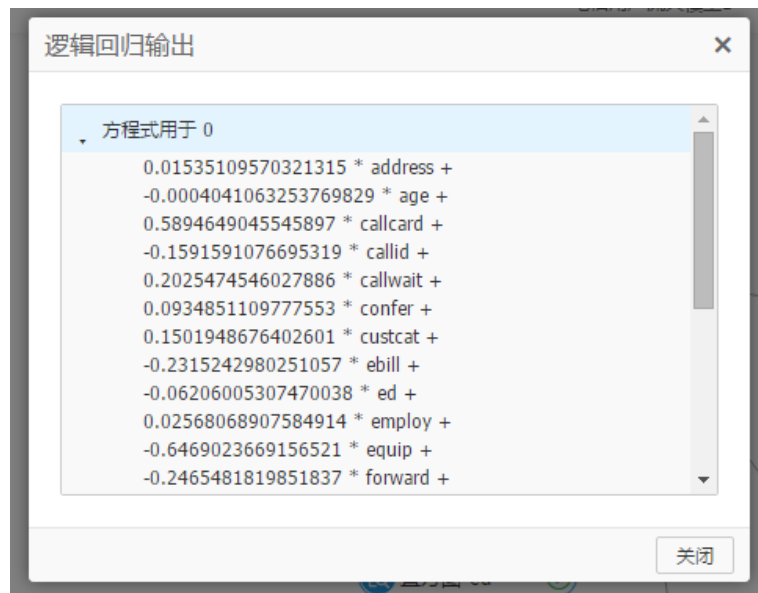
# 模型管理

模型可视化

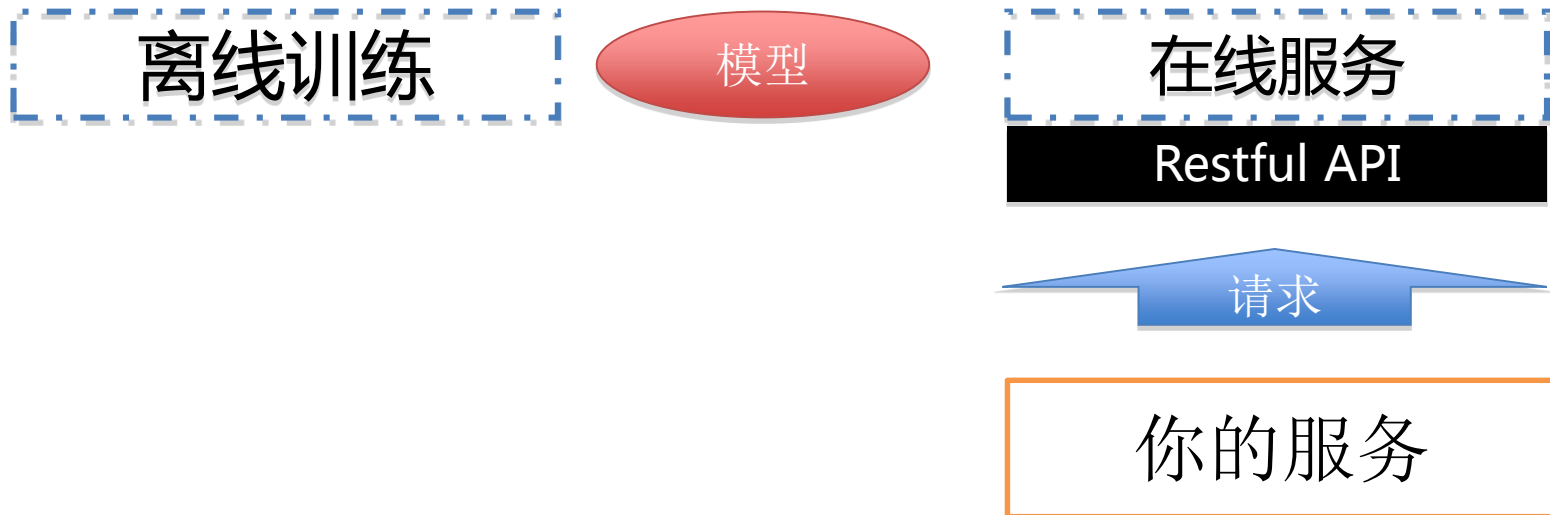
离线模型

在线模型

PMML



## 在线预测服务



# 在阿里的应用



千人千面

芝麻信用分

智能客服



扫码观看大会视频

## 在阿里的应用



商家反作弊

退运费险

舆情监控

证件智能审核

千人千面

芝麻信

更多应用场景.....





# 接下来我们会推出更多服务



深度学习 人工智能 自然语言理解  
异构计算服务 大数据  
图片识别 语言翻译  
高吞吐 语音识别 非结构化  
分布式存储



# GPU异构计算集群

MaxCompute

全局计算调度

计算节点

计算节点

计算节点

计算节点

CPU计算集群

GPU

计算节点

GPU

计算节点

GPU

计算节点

GPU

计算节点

GPU计算集群





## 多种分布式深度学习框架

### 分布式Caffe

- 多机多卡支持
- 高性能通信框架
- 支持Sequential Learning
- 性能优化
- 异步更新
- 模型分片

### 分布式Kaldi

- 分布式语音深度学习框
- 线性加速比
- 分布式调度
- Pipeline
- 多数据源接入

### TensorFlow

- 强大的分布式深度学习框架
- 支持多种深度学习网络
- 分布式调度
- 资源动态管理
- GPU卡虚拟化管理
- 抽象API封装
- 多数据源接入





# PAI，让你的服务更加**智能**



2016 The  
Computing  
Conference  
**THANKS**

