

腾讯云 | 连接智能未来

# 极速训练：超大规模分布式机器学习

张文杰 腾讯云AI平台总经理



1. 智能钛机器学习 ( TI - ML )
2. TI Self-Learning
3. TI One
4. TI Accelerator
5. 极速训练



强大算力的一站式机器学习服务平台

多产品形态提供差异化机器学习能力

形成完整的开发者生态

图像识别

金融风控

智能投研

精准推荐

其他场景

Tencent Intelligence Machine Learning Platform

CPU / GPU / FPGA / ASIC/CVM





非算法人员



AI算法工程师



AI算法专家

Tencent Intelligence Machine Learning Platform

全自动、快迭代

**TI Self-Learning**

拖拽式、可视化

**TI One**

高性能、轻量级

**TI Accelerator**



**AI Project**

场景

算法

数据

技能 \ 成本	低	中	高
低	TI S		
中		TI One	
高			TI A

公有云服务

- 按需使用，按量付费
- 单机多卡、多机多卡
- 云端数据、定制算法

私有化一体机

- 数据存放限制及监管
- 多种硬件定制选型
- 私有云到公有云无缝迁移

# 智能钛机器学习





01

提供多个场景  
自动建模能力

02

提供深度定制功能  
优于业界效果

03

模型算法无需  
人工调参

04

模型结果更加简单，  
易于部署

实验名称: GBDT分类实验案例\_19537

配置管理

并行方式:  CPU  DATA

数据停止阈值:

最小split\_fraction:

用于构造桶的数据采样量:

类别数量:

sigmoid:

温馨提示: 右键点击组件可以查看运行结果和日志哦

实验名称: Ftr实验案例\_534518

配置管理

批量数量:

特征空间大小:

迭代次数:

lwr0001:

lwr0002:

alpha:

beta:

温馨提示: 右键点击组件可以查看运行结果和日志哦

# 智能钛一站式机器学习

- 难以快速入门机器学习
- 需要构建机器学习基础设施
- 缺乏模型构建能力



业界痛点

- 快速上手、教学容易
- 提供算法模型



行业需求



我们提供

- 操作可视化、拖拉拽式任务流
- 大量现成算法模型、算法结果具象化
- 灵活可自定义一站式机器学习平台



拖拽式任务流设计



运行模式灵活



支持10+机器学习框架



100+机器学习算法



效果可视化



强大的团队协作和分享

## TI One 值得选择的理由

自建

TI ONE

框架重复繁琐的步骤从头到尾搭建，并专人维护更新



框架



一键集成，开箱即用

大量样本数据集及开发人员进行算法设计和参数调优，记录每次训练结果



算法



拖拽式界面，合理的算法参数设置  
一键多种参数组合调优  
批量可视化评估，结果一目了然

模型训练和使用脱节；  
选取最优模型、使用模型都要手动开发



模型



模型训练与预测无缝衔接，可多流程复用  
模型自动更新，异常迅速告警

手工编写画图脚本



评估



实时动态可视化评估

监控脚本监控运行的任务，自行发现并处理异常；  
通过拷贝代码或者git分享算法



协作



多种监控配置，完善告警体系，多次自动重试；  
发布算法和模型，可控分享粒度和级别

The screenshot displays the TI One machine learning workflow interface. The main workspace shows a flowchart with the following steps:

- 训练集特征处理 (成功)
- 数据切分 (成功)
- 测试集特征处理 (成功)
- 流式率模型训练 (成功)
- 二分类评估 (成功)

On the right, the '算法配置' (Algorithm Configuration) panel is visible, showing the following parameters:

参数名称	配置值
输入数据	hdfs://PB-tn-tdw.tencent-c.com/...   HDFS URL
特征列	0 从1开始计数
目标列	1 模型训练时-1的列名, 从0开始计数
抽样率	1.0 范围: 0-1.0
并行数	20 Spark并行数

# 智能钛机器学习加速器



- 大规模训练时间长
- 计算集群成本高、太复杂
- 业务规模持续扩大、模型快速更新



业界痛点

- 轻便快捷的专业机器学习环境
- 整合常见的算法框架
- 分布式、高性能、易部署
- 大算力、高性价比



行业需求



我们提供

- 整合腾讯内部机器学习经验
- 大规模集群、资源调配、网络调优
- 框架优化、易用、高性价比



方便便捷



轻量级



多框架选择



高性能



成本低

```

→ /tmp
→ /tmp tictl job create \
--name=test01 \
--packagedir="cos://test-1255582819.cos.ap-shanghai.myqcloud.com/example:/data/mnist" \
--command-python --runtime-tia-1.7.0-gpu \
--args="mnist_saved_model.py ./mnist_model_01test" \
--wt-5U10G1P

create job success.
NAME      CREATED          END              STATE    MESSAGE
test01    2018-05-07 21:17:51.034349343 +0800 CST  a=+4384.402868380      creating
→ /tmp tictl job list
NAME      CREATED          END              STATE    MESSAGE
test01    2018-05-07 21:17:51 +0800 CST      Running    [MASTER/Running : Running/1]
→ /tmp
    
```

创建任务及查询

```

→ /tmp
→ /tmp tictl model create \
--name=model01 \
--description=model01 \
--model="cos://test-1255582819.cos.ap-shanghai.myqcloud.com/example/mnist_model_01test:/data/model"

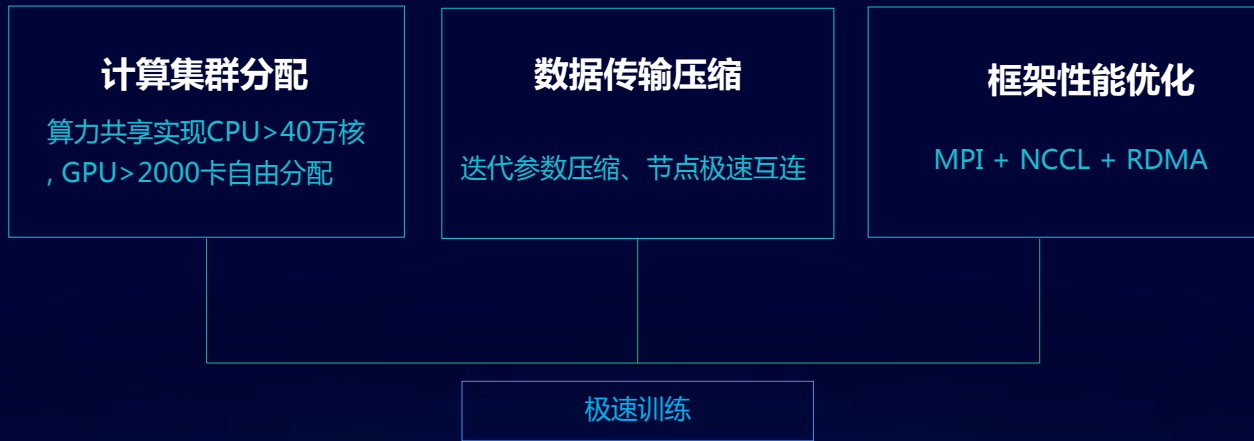
create model success.
NAME      CREATED          URL              STATE    MESSAGE
model01   2018-05-07 21:35:46 +0800 CST               creating
→ /tmp tictl model describe --n model01
NAME      CREATED          URL              STATE    MESSAGE
model01   2018-05-07 21:35:46 +0800 CST               Running   Deployment has minimum availability.
→ /tmp tictl model describe --n model01
NAME      CREATED          URL              STATE    MESSAGE
model01   2018-05-07 21:35:46 +0800 CST      100.143.51.48  Running   Deployment has minimum availability.
→ /tmp curl 100.143.51.48
Hello World!
→ /tmp
    
```

模型部署及验证

```

→ /tmp
→ /tmp tictl job list
NAME      CREATED          END              STATE    MESSAGE
test01    2018-05-07 21:17:51 +0800 CST      2018-05-07 21:23:05 +0800 CST  Succeeded
→ /tmp
→ /tmp tictl logs -n test01
    
```

查看任务及日志



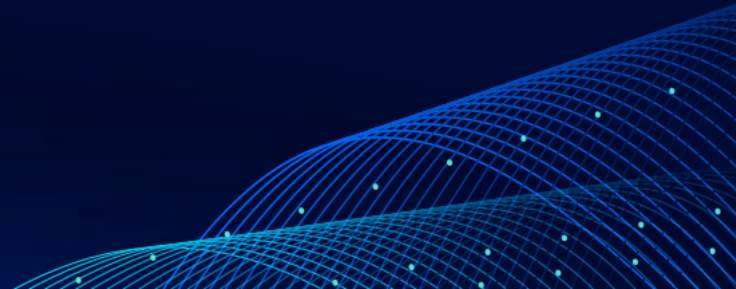
# TI A 训练加速对比

大规模GPU集群加速提升明显 (128块GPU卡时, 提升超过1倍)

- 节点扩展成本恒定
- 超大规模GPU集群 (>1000块GPU卡)



NAVINFO >  
四维图新



以我所能 为你而+

 腾讯云 | 连接智能未来