



# Momenta Cloud

---

[www.momenta.ai](http://www.momenta.ai)

Sept 2018

# Agenda

- **Momenta 公司介绍**
- Momenta Cloud 介绍
- 标注分发云 MAC
- 弹性训练云 MTC



# 最顶级的人工智能科学家团队



**曹旭东**

**创始人&CEO**

微软亚洲研究院研究员  
商汤科技执行研发总监  
清华大学本科



**任少卿**

**合伙人，算法总监**

福布斯亚洲30岁以下30位精英  
Faster-RCNN 和 ResNet 作者  
CVPR 2016最佳论文奖得主  
中科大&微软亚研院联合培养  
计算机视觉专业博士



**夏炎**

**合伙人，算法总监**

深度学习和大数据领域专家  
中科大&微软亚研院联合培养  
模式识别和智能系统专业博士



**孙刚**

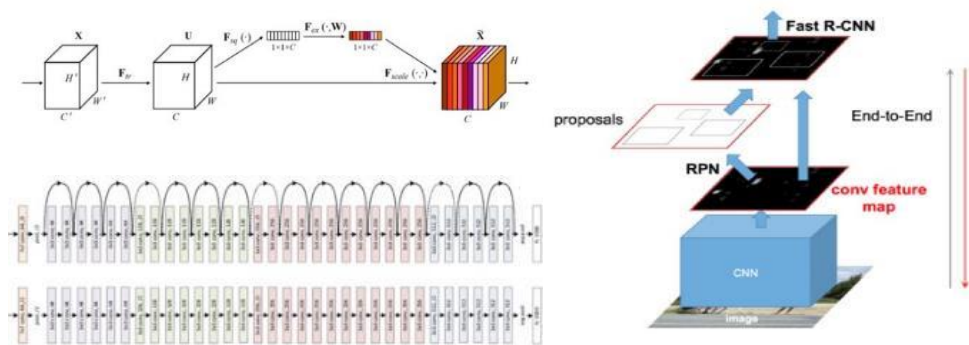
**合伙人，首席架构师**

高性能并行计算系统专家  
ImageNet 2017 图像分类冠军  
ImageNet 2016 场景分类亚军  
中科院大学计算机视觉专业博士

# 核心技术

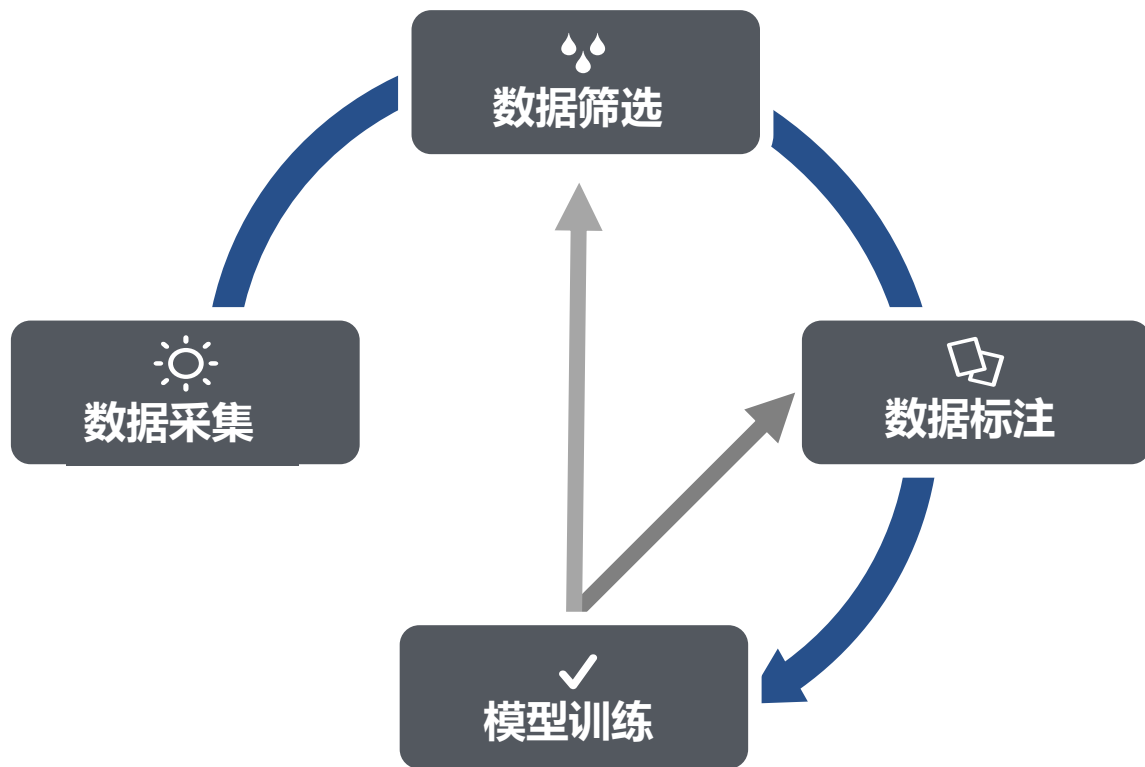
## 世界级人工智能研发能力

- ImageNet 2015物体检测，图像分类，场景识别**冠军**
- ImageNet 2017图像分类**冠军**
- MS COCO Challenge 2015多项比赛**冠军**
- **最具影响力**的深度学习网络结构设计作者
- **最具影响力**的物体检测框架设计作者



## 超大规模大数据平台

- 在线远程众包数据标注系统

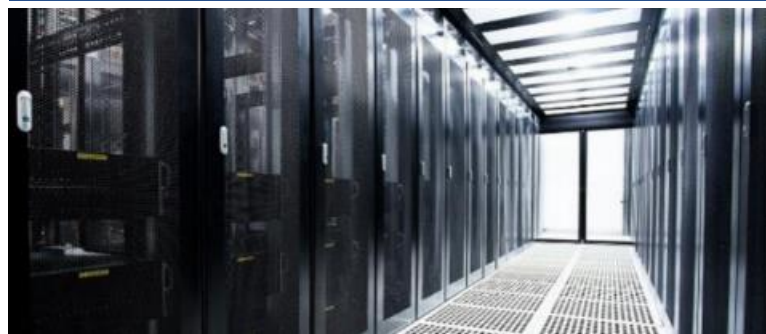




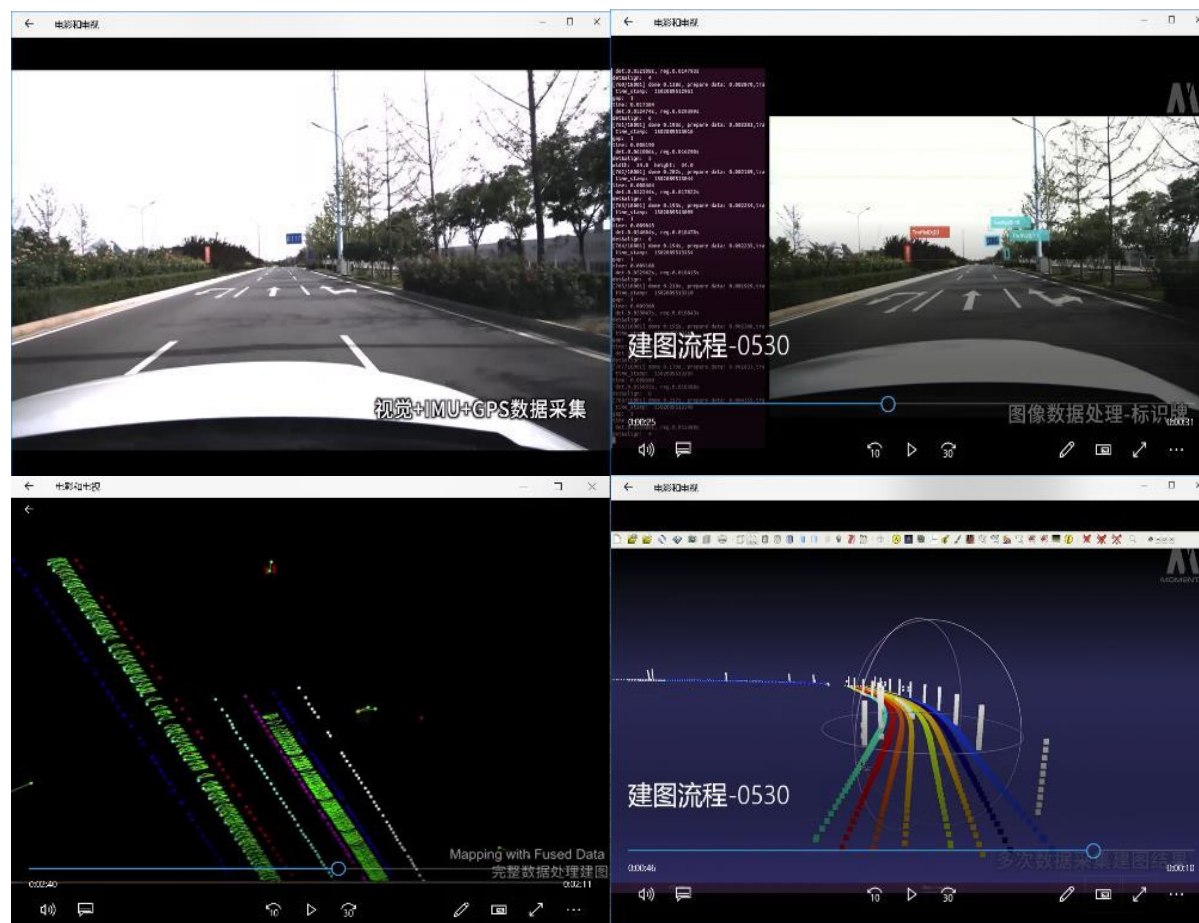
# 核心技术

## 最大的人工智能训练平台

- 拥有**全球最大规模**的深度学习训练平台  
(2000块GPU, 2018Q1)

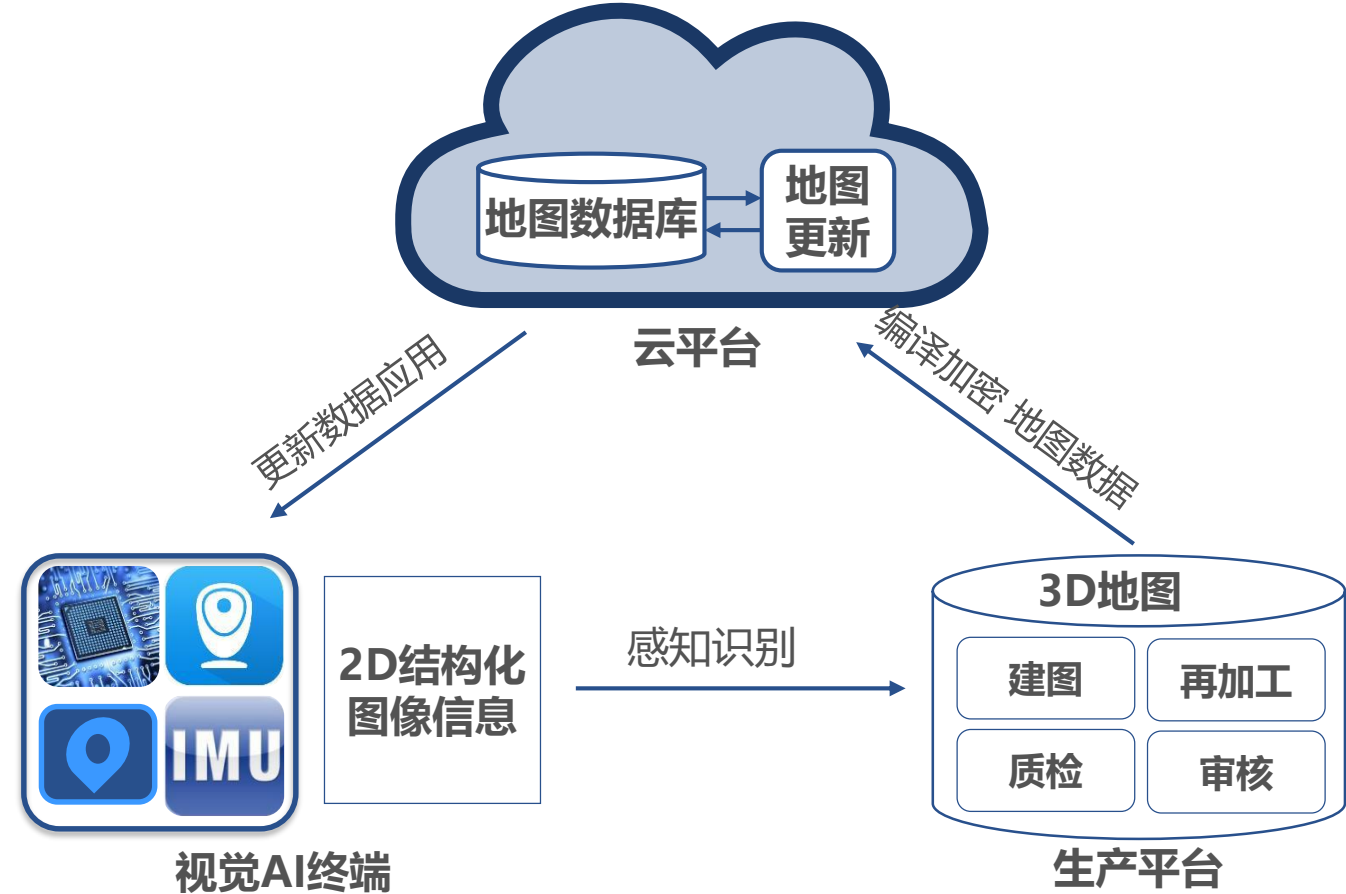


## 自动化地图编辑系统

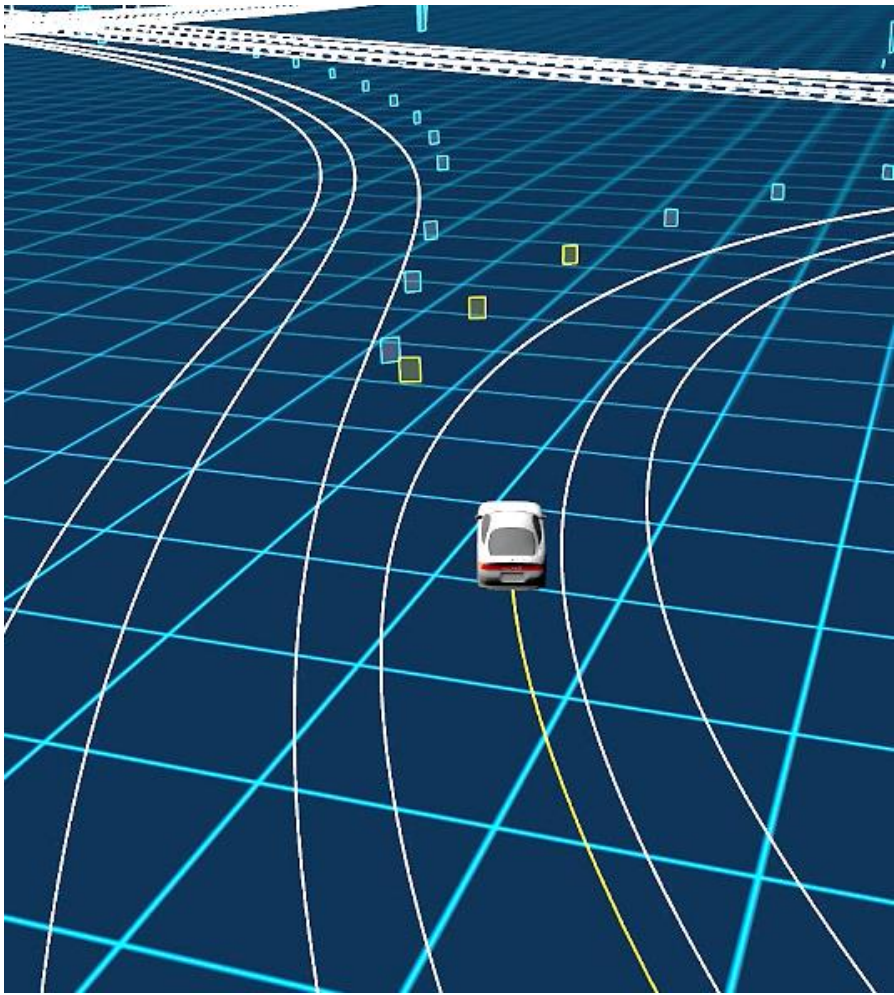


# 基于人工智能的地图技术路线

方案优势：成本低、更新快、精度高



技术路线示意图



自动驾驶导航地图示意图

# Agenda

- Momenta 公司介绍
- **Momenta Cloud 介绍**
- 标注分发云 MAC
- 弹性训练云 MTC





# Momenta Cloud

## 现状

- 车联网、数字化进行中
- 海量数据无处安放
- 大量的数据无法收集
- 数据价值等待被解放

## 发展

- 大数据是企业发展必经之路
- 落地数据应用到数据驱动
- 人工智能时代已经到来



## 机遇



车联网、数字化进行中  
海量数据无处安放  
大量的数据无法收集  
数据价值等待被解放

滴滴事件



技术发展



AlphaGo

AlphaGo



政策红利



主机优势



MOMENTA 为了无法计算的价值 | 阿里云





## 途径

化整为零，循序渐进

### 数据标准化

- 定义数据
- 统一存储
- 一致访问
- 共享原信息

### 数据管理

- 数据地图
- 血缘分析
- 权限体系
- workflow 设计

### 数据应用

- 数据处理
- 数据标注
- 模型训练
- 数据可视化

### 数据监管

- 数据审计
- 监控和告警
- 日志
- 计费

## 当前市场情况

首款支持深度学习产品  
机器学习产品没有得到普及  
直接进入人工智能时代

我们如何做

# Momenta Cloud - 一站式平台

## 车上设备

智能硬件及配套部署方案，赋能车辆资源，整合最广泛介质，通过车载设备实现分布式数据采集最大化数据池



## 数据传输

打通外部数据与数据平台的连接，在原始数据与数据平台间构建高速通道，高效、便捷、自动化的实现数据同步



## 数据管理

混沌未凿是数据管理面对的共同问题，海量数据、多种类型，非机构化数据与日俱增，数据治理迫在眉睫



## 数据监控

数据生产、传输、加工、处理、应用等流程全面监控，为数据价值保驾护航，监控核心数据，掌握实时动态



## 数据处理

犹如淘金，原始数据如同淤泥，需要经过洗涤、处理、抽丝剥茧后得到其中的金沙，真正有价值的数据往往是其中的金子



## 数据标注

人工智能依赖标准结果输入，需要监督学习的方式，使用已有数据集作为基础，为大脑提供标准



## 模型训练

模型、网络是人工智能的广义产出，不同模型应对不同场景，最好的模型、智能需要结合实际场景不断挖掘



## 模型产出

模型需要应用、需要落地，智能需要与工业结合只有将可靠、准确的模型应用到设备或系统端才真正运转起来



## 模型评估

繁多、复杂的模型拥挤在系统前，除了自有模型越来越多的开源、第三方模型可选如何选出最适合的一个



# Momenta Cloud - 一站式平台

## 车上设备

- 通用采集硬件
- 多种采集模式
- 支持多种设备
- 灵活装备方案
- 可扩展存储设备

## 数据传输

- 高速硬盘
- SD卡
- 4G

## 数据管理

- 统一数据存储
- 弹性数据存储
- 非结构化数据无差别存储
- 灵活取用

## 数据监控

- IOT
- 电子围栏
- 路测管理

## 数据处理

- 非结构数据加工处理
- 图片、视频多类型支持
- GPU加速
- 灵活的处理方式
- 模型闭环复用

## 数据标注

- 一键拥有私有标注平台
- 强大的权限管理
- 支持引入外部团队
- 拥有全球最高效的标注资源, 满足标注需求
- 标注流程定制化
- 数据驱动标注, 精细化运营
- 标注结果一致性管理

## 模型训练

- 基于Caffe的深度优化框架
- 图像领域最好用框架
- 自助式训练管理平台
- 可视化操作
- 即时修改即时调试, 效率第一
- 自由使用多种标注结果

## 模型产出

- 生成可运行文件
- 可应用于车载设备或其他环境
- 更新至数据处理, 完善闭环

## 模型评估

- 模型对比运算
- 基于相同评测集产出评估报告
- 智能提示潜在问题

## 方法与实践

- 设计思考
- 敏捷& DevOps
- 精益实践
- 质量工程
- 自动化
- 可靠性工程(SRE)
- 行为驱动开发
- 数据创新

HOW

WHAT

Momenta  
Cloud

敏捷  
创新

人工智能时代  
最强产品

WHO

## 组织运作模式

- 人才/技能
- 产品导向
- 精益企业
- 加速变更
- 众包
- 敏捷管理
- 开放创新
- 企业文化

# 深度学习数据流





# Agenda

- Momenta 公司介绍
- Momenta Cloud 介绍
- **标注分发云 MAC**
- 弹性训练云 MTC



# 标注云 MAC:



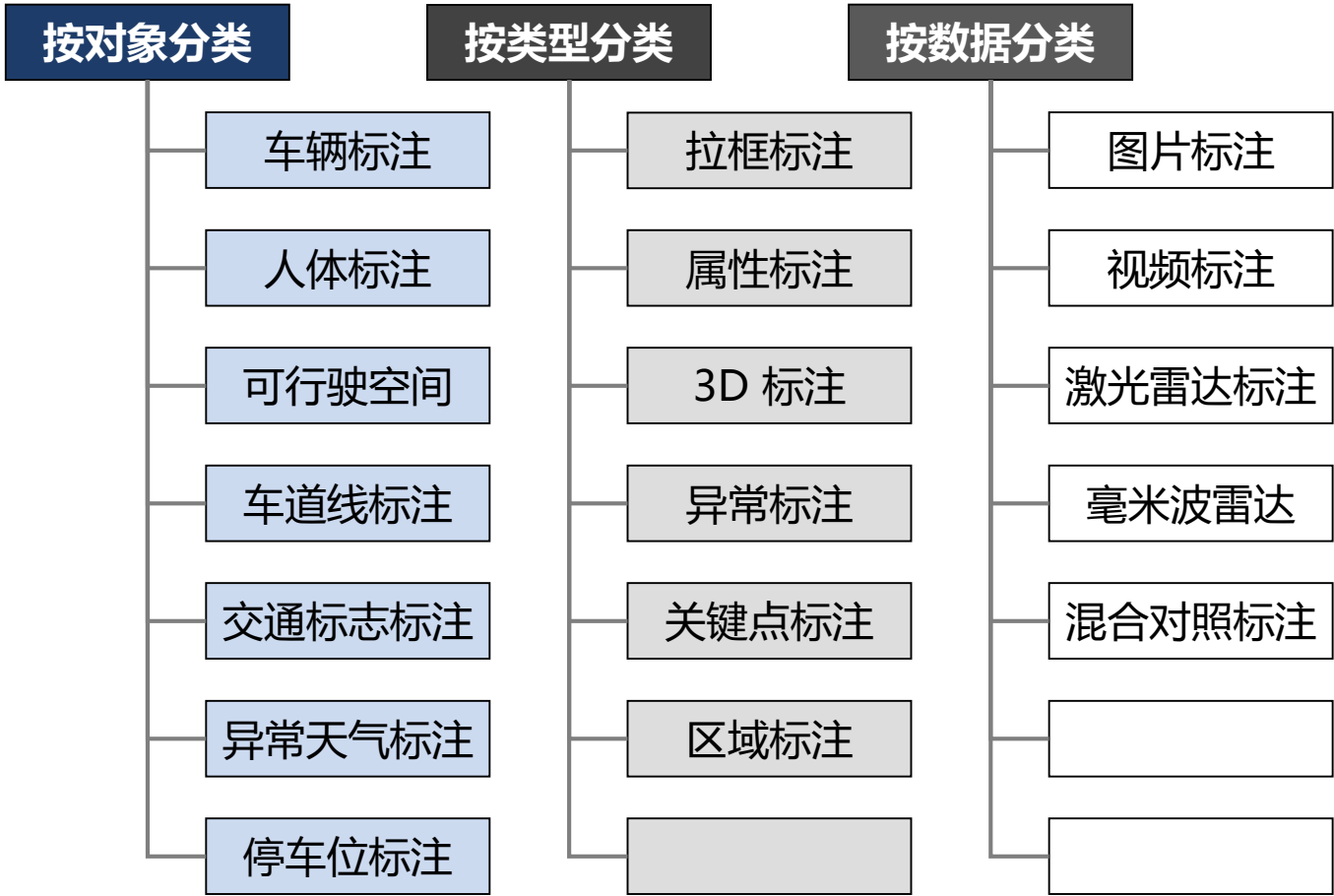
Momenta分发标注云 (Momenta Annotation Cloud)是 Momenta 独立研发用于自动驾驶标注的云平台，系统包括标注客户端、团队管理、运营管理等多个模块。采用众包模式，实现了低成本、高质量、快速的海量数据标注。

图片来源：Momenta

# 面向自动驾驶，全面支持多种标注类型

MAC 的独立标注工具是目前业界最先进、支持任务种类最多的标注工具之一，并以其简学、易用、高效等特点获得国内主流标注团队的一致好评；

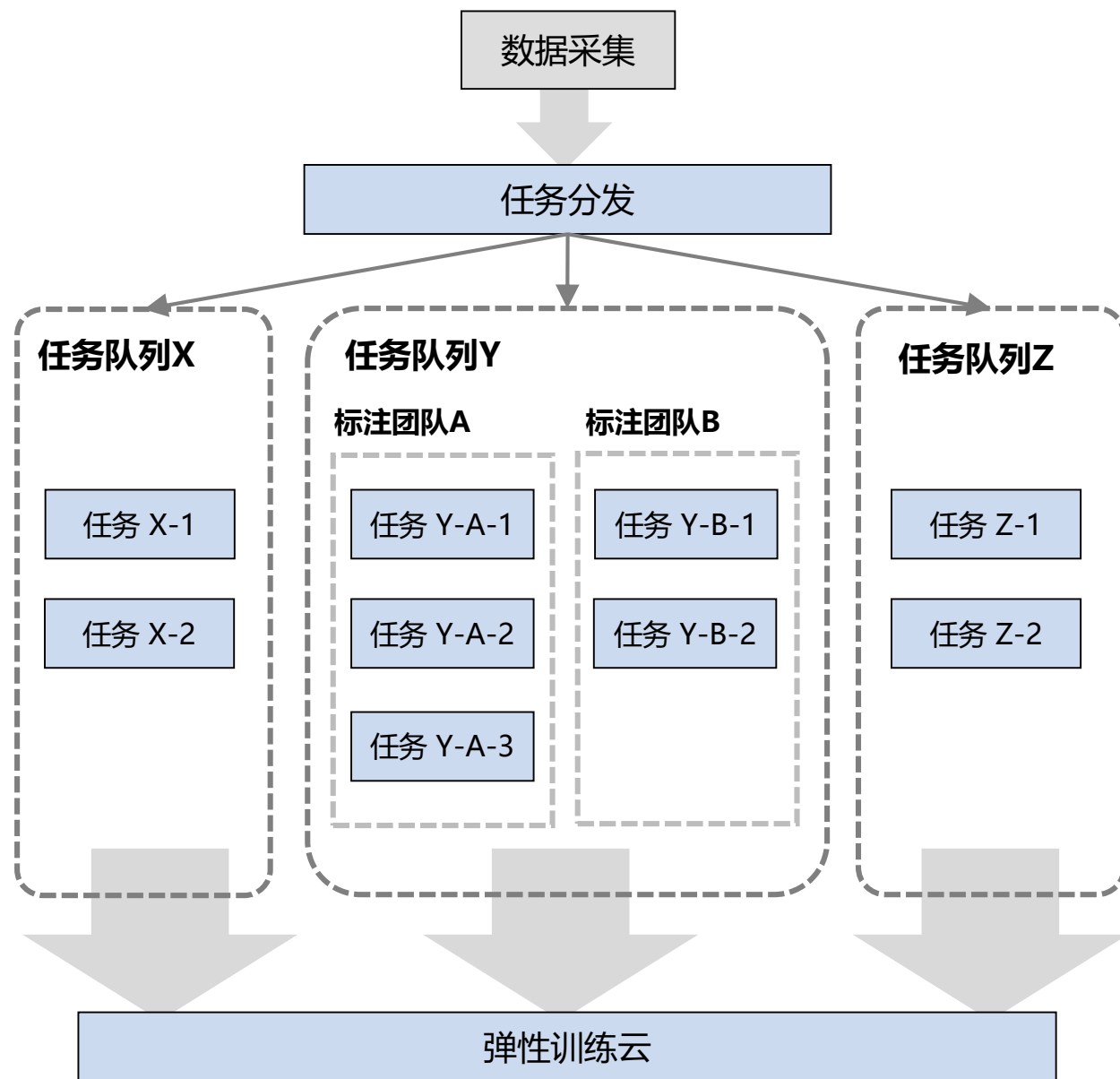
- **支持多种标注任务：**支持多种驾驶路面对象、标注方式、输入数据类型；
- **跨平台：**支持 Windows / Mac / Linux 等多种平台，即将支持 Web 标注
- **完善的培训课程、文档：**针对不同的标注任务，均具备完善的培训材料，新的标注团队可以迅速掌握



# 采用众包任务、降低成本

MAC 采用众包模式，可将标注任务分发至可以在地理隔离的多个标注团队，从而实现了低成本、高质量的标注；

- **团队资质评定：**标注团队经过对应培训后具备相应的资质后可以领取对应的任务；并根据标注结果进行自动更新；
- **竞价模式任务池：**同一个任务进入任务池后，可通过竞价模式发布给多个团队，最终由最低价者获得；
- **任务调度：**可以按照任务的大小、紧急程度、质量要求高低分配给不同的团队，实现资源的优化利用；





# 严格的审核机制，控制标注质量

MAC 通过严格的流程、规则把控，在低成本标注的同时保证了标注结果的质量；

- **严格的审核流程**：完整的审核流程、对错误图片的处理流程，有效防止错误数据、不合格数据流入训练环节；
- **专业的奖惩规则**：专业运营团队积累多年经验制定的一系列奖惩机制，有效激励高效标注团队、淘汰不合格团队；
- **先进的自动审核机制**：引入模型审核、混合审核等多种自动审核机制，自动化审核流程提高效率；同时降低了人工影响的不可控因素，进一步提高了质量；



# Agenda

- Momenta 公司介绍
- Momenta Cloud 介绍
- 标注分发云 MAC
- **弹性训练云 MTC**



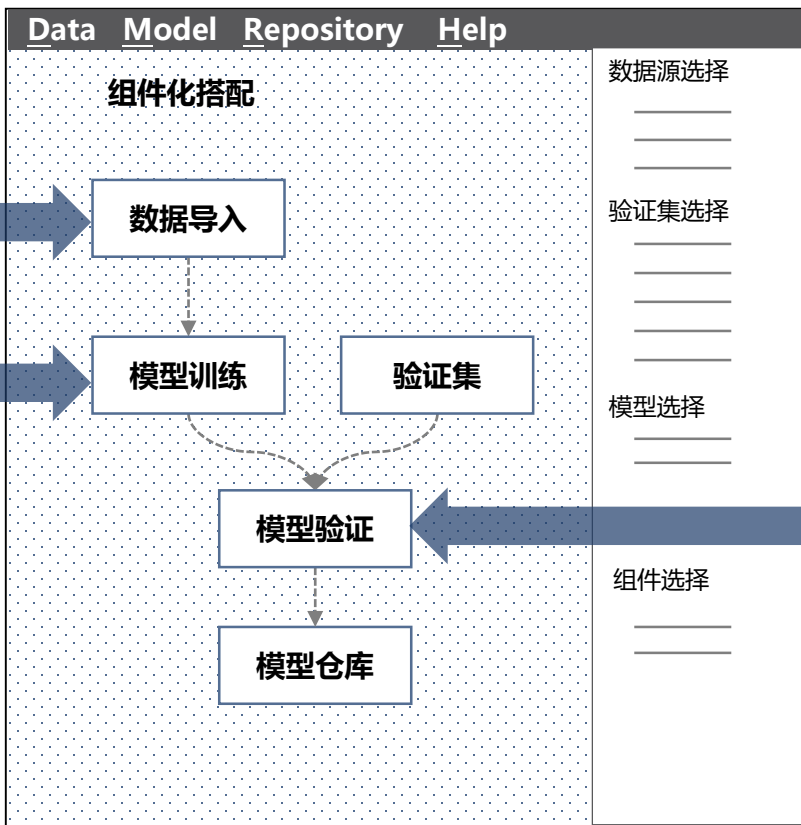
# 弹性训练云 MTC

## 数据导入

- 支持多种数据类型;
- 支持多批次数据合并;
- 支持多种组合规则;

## 模型训练

- 多机多卡并发训练;
- 共享式集群, 支持多人多任务同时进行;
- 支持多种主流训练框架;
- 所有任务由 Kubernetes 自动调度完成;



## 模型验证

- 训练所得模型自动在验证集上进行验证;
- 通过验证的模型将进入模型仓库, 供后续流程使用;

弹性训练云 (Momenta Training Cloud) 是 Momenta 研发团队用于模型训练的分布式共享深度学习训练系统。单个集群最多可部署至不超过 16 个节点上, 提供了接近线性加速比的高性能加速, 并实现多人、多任务共享资源池, 提高了 GPU 使用效率。

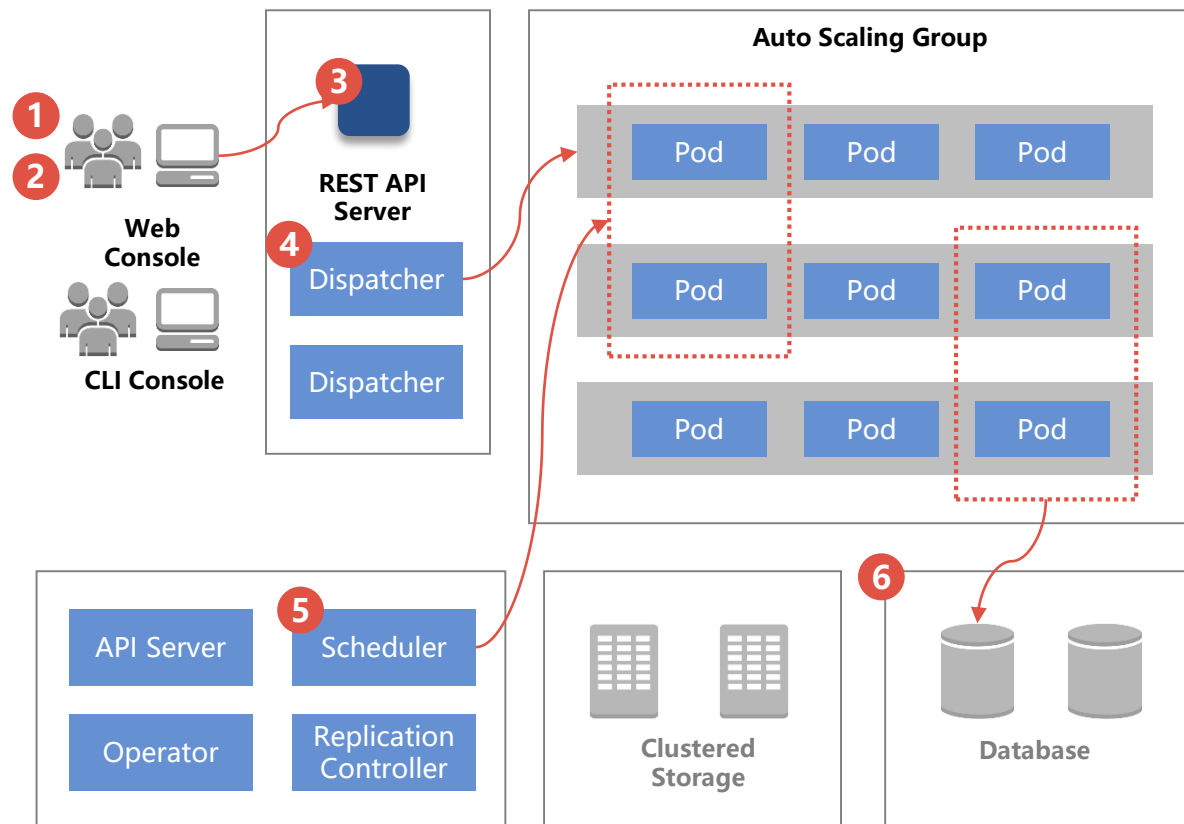
数据来源: Momenta

# 开放式系统架构，更易于集成

系统采用开发式架构，组件与市场主流方案兼容，更易于集成和二次开发；

- **开放式平台**：基于 Google 成熟框架 Kubernetes 一站式解决方案，所有组件均采用维护成本低、易于迁移；
- **多训练框架**：支持 Caffe, Caffe2, PyTorch, TensorFlow 等多种训练框架；
- **多种方式部署**：支持私有云、混合云、公有云多种方式部署；
- **多云部署**：与阿里云、AWS 等云服务厂商实现无缝对接；

MTC 系统架构示意





# 分布式多卡训练、高效加速

支持多机多卡训练，提供高加速比，训练速度更快，

MTC 训练加速比<sup>2</sup>

- **高效加速**：在 8x8 节点上实现接近线性训练加速比，相对传统单机训练速度提高超过 10 倍以上<sup>1</sup>；
- **资源共享**：多个团队之间共享同一个集群，并支持团队之间的配额分享；
- **自动任务调度**：通过排队自动调度系统，任务可以无间歇自动调度至文件系统上，极大提高GPU利用率；

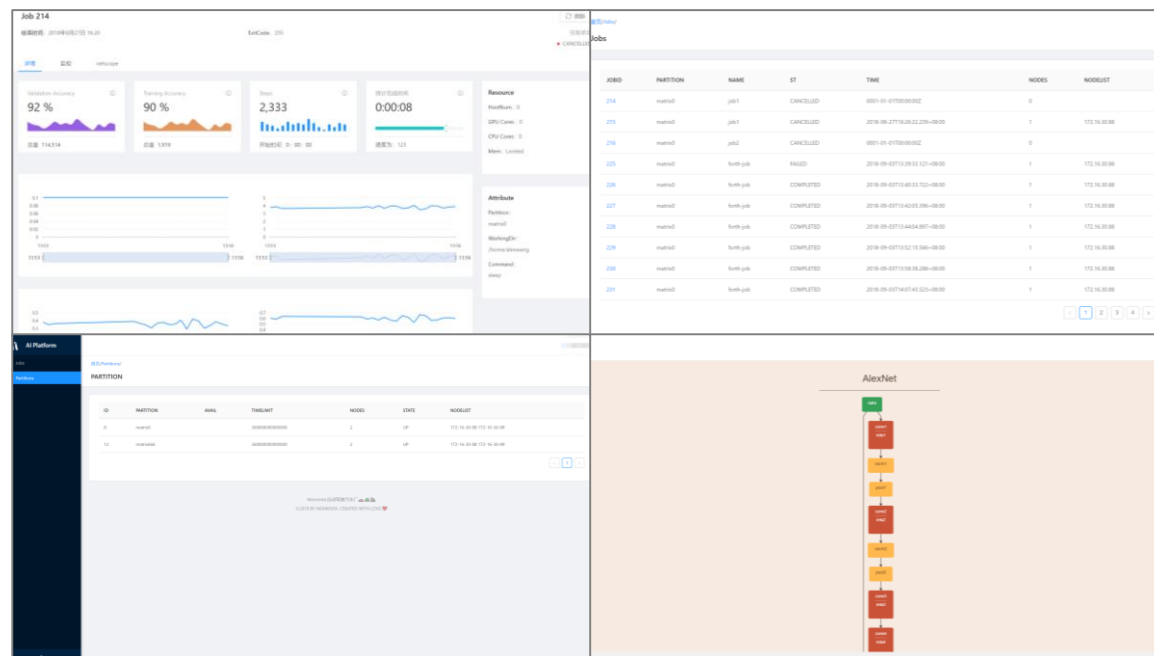
数据来源：Momenta

# 可视化界面，实时动态监控训练结果

支持多机多卡训练，提供高效加速比，

- **可视化界面**：通过网页可视化界面，同时提供命令行界面及API接口，便于系统集成和二次开发；
- **实时监控**：多种类型数据探头，同时支持自定义探头，可以实时监控不同节点、不同任务的状态、模型参数变化情况、数据使用情况等；
- **统一管理接口**：可统一管理数据、集群节点、模型结果，支持不同用户类型的权限定义、管理；

MAC 动态监控任务、训练状态，展示模型结构，运行结果



图片来源：Momenta

# 基于成功经验、快速部署上线

2

Momenta 2 年时间打造，吸取人工智能行业  
业界成功经验

50

训练集群已经成功部署  
在超过 50 个节点

100

数百只合作团队、上千  
个标注任务



# DEMO

---



为了无法计算的价值 |  阿里云  MOMENTA

