



中山大學 软件工程学院
SUN YAT-SEN UNIVERSITY SCHOOL OF SOFTWARE ENGINEERING

SSE316: 云计算技术 Cloud Computing Technology

陈壮彬

软件工程学院

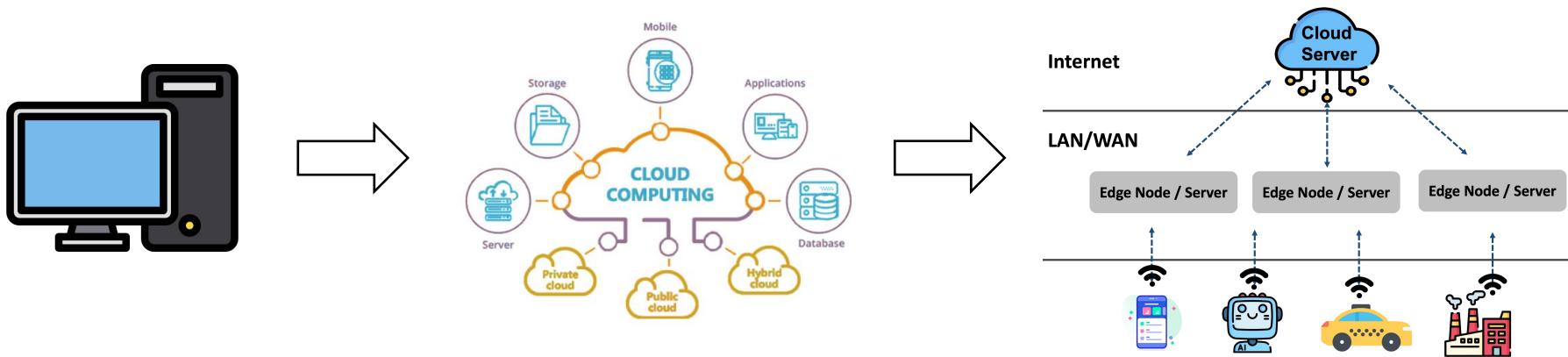
<https://zbchern.github.io/sse316.html>



边缘计算

- ❖ 边缘计算概述
- ❖ 边缘计算的典型应用
- ❖ 边缘计算的挑战
- ❖ 边缘计算中的 AI 部署

计算范式的发展



个人计算机

云计算

边缘计算

从 PC 时代到 Cloud 时代



- 云计算应用

- ✓ Software as a Service
- ✓ Personal life/classroom
- ✓ Scientific research



- 关键事件

- ✓ 2004 年 12 月份，IBM 把所有 PC 业务卖给联想

- Data center is the computer (2008)

Digital infrastructure 1.0 (05-15)

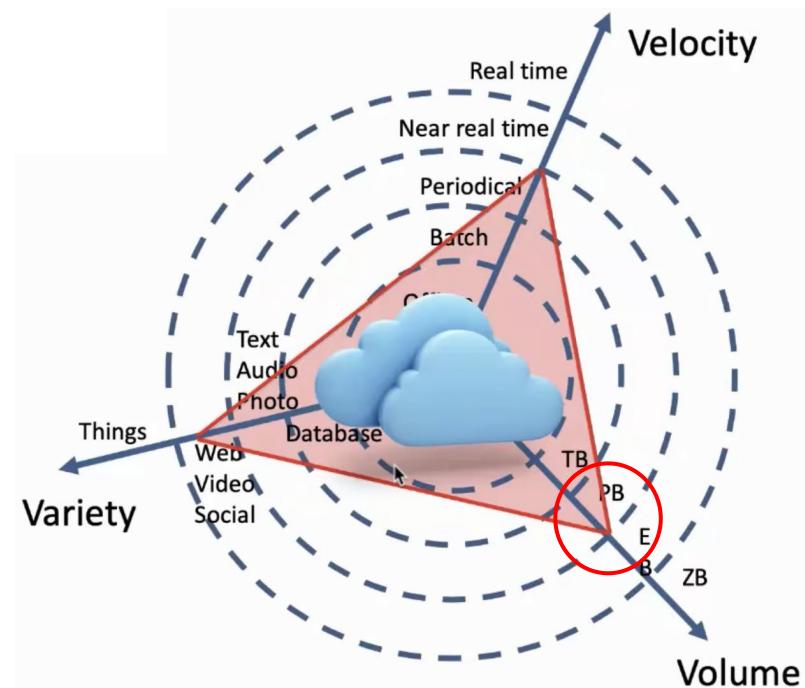


- 关键特征

- ✓ Push the data to the cloud
- ✓ Popular data processing model
MapReduce (OSDI' 04)
by Jeff Dean

- 商业/技术逻辑

- ✓ 资源集中管理/售卖带来的红利
- ✓ 网络传输能力大于数据的产生
- ✓ ...



Why edge computing?



- 概念不是最新的，CDN（Content Delivery Network）
- Push from cloud providers
 - Reduce latency, e.g., 10ms
 - Improve availability
 - Save bandwidth



2017 年 5 月 Satya:
Not cloud, not office
either, but intelligent
edge

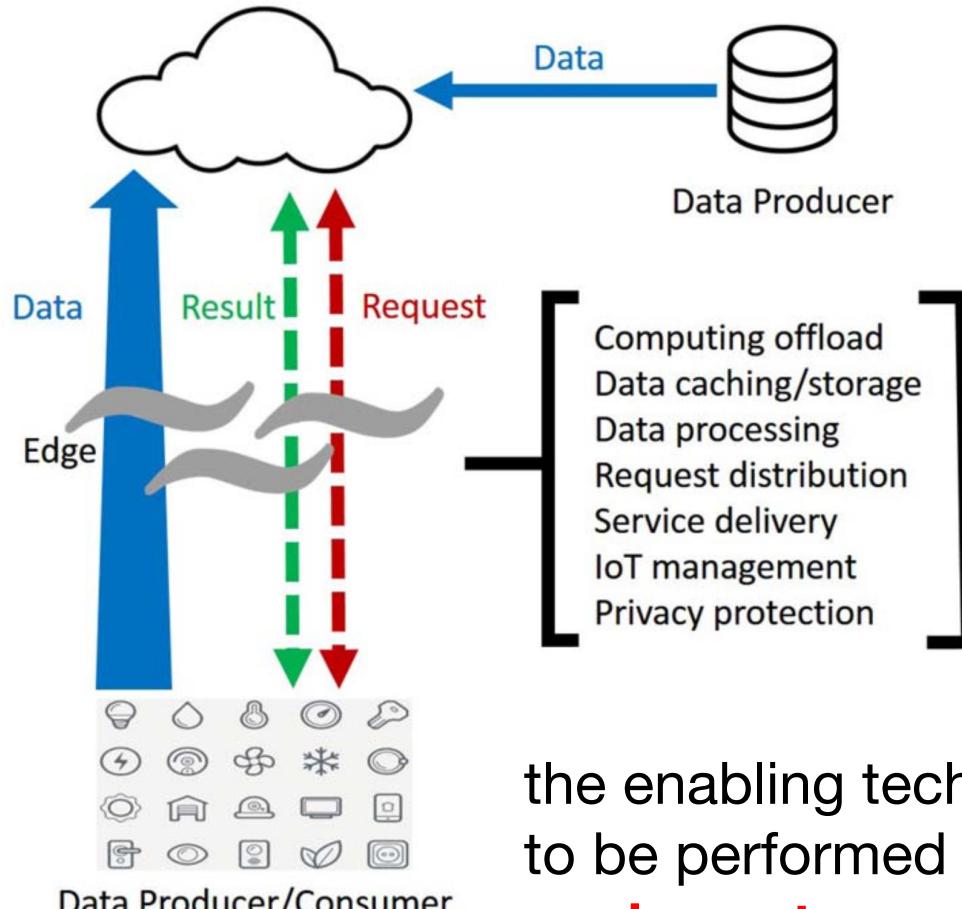
Why edge computing?



- **Push** from Internet of Things
 - ✓ Real-time context computing
 - ✓ Resource constraints
 - ✓ Security/privacy requirements



Edge computing in a nutshell



上行数据为物联网服务的数据
下行数据为云服务的数据

the enabling technologies allowing computation to be performed at the edge of the network, on **downstream** data on behalf of cloud services and **upstream** data on behalf of IoT services.

Digital infrastructure 2.0 (15-)

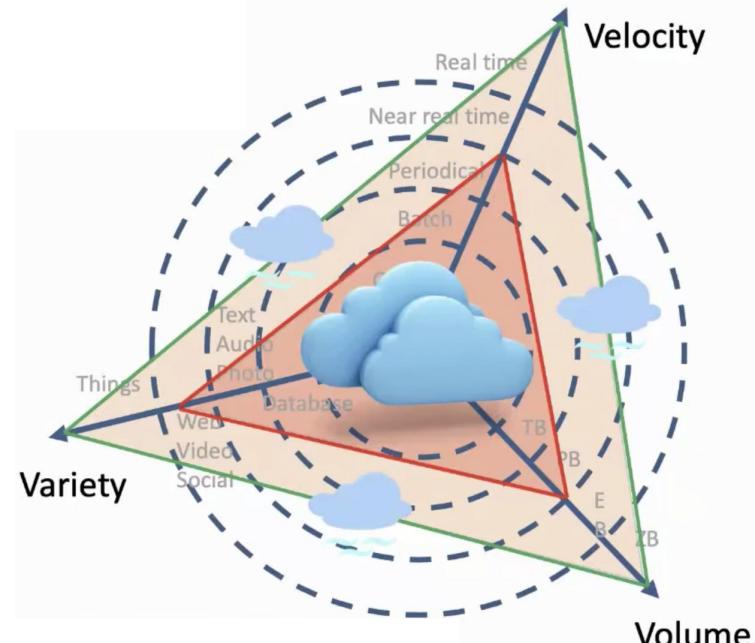


- 关键特征

- ✓ Data distributed at Cloud & Edges
- ✓ Possible processing models:
Cloudlet, Firework, FaaS

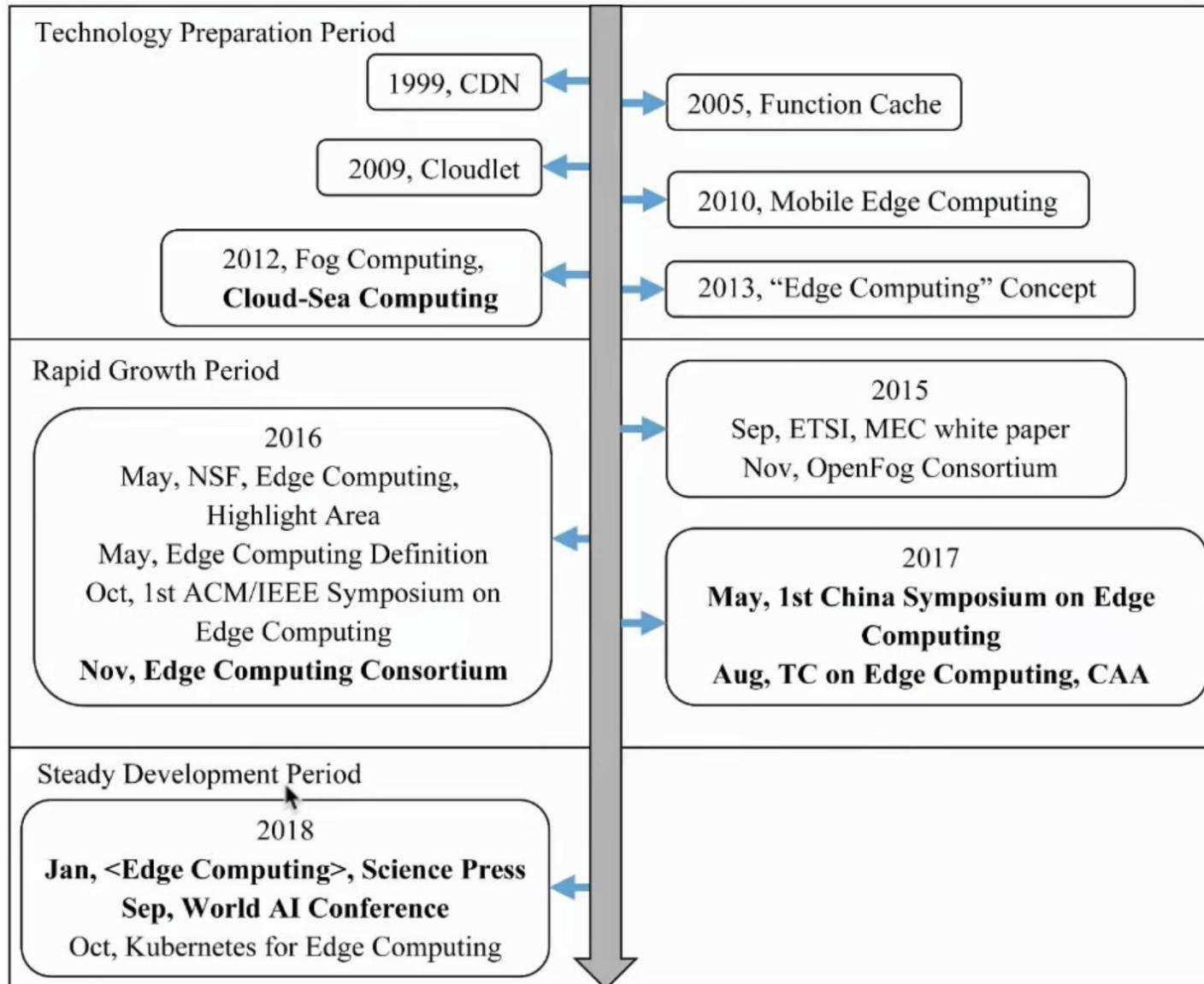
- 商业/技术逻辑

- ✓ IoT 设备种类/应用种类增多
- ✓ 数据量达到 ZB 级
- ✓ 网络的发展远远落后于计算的发展

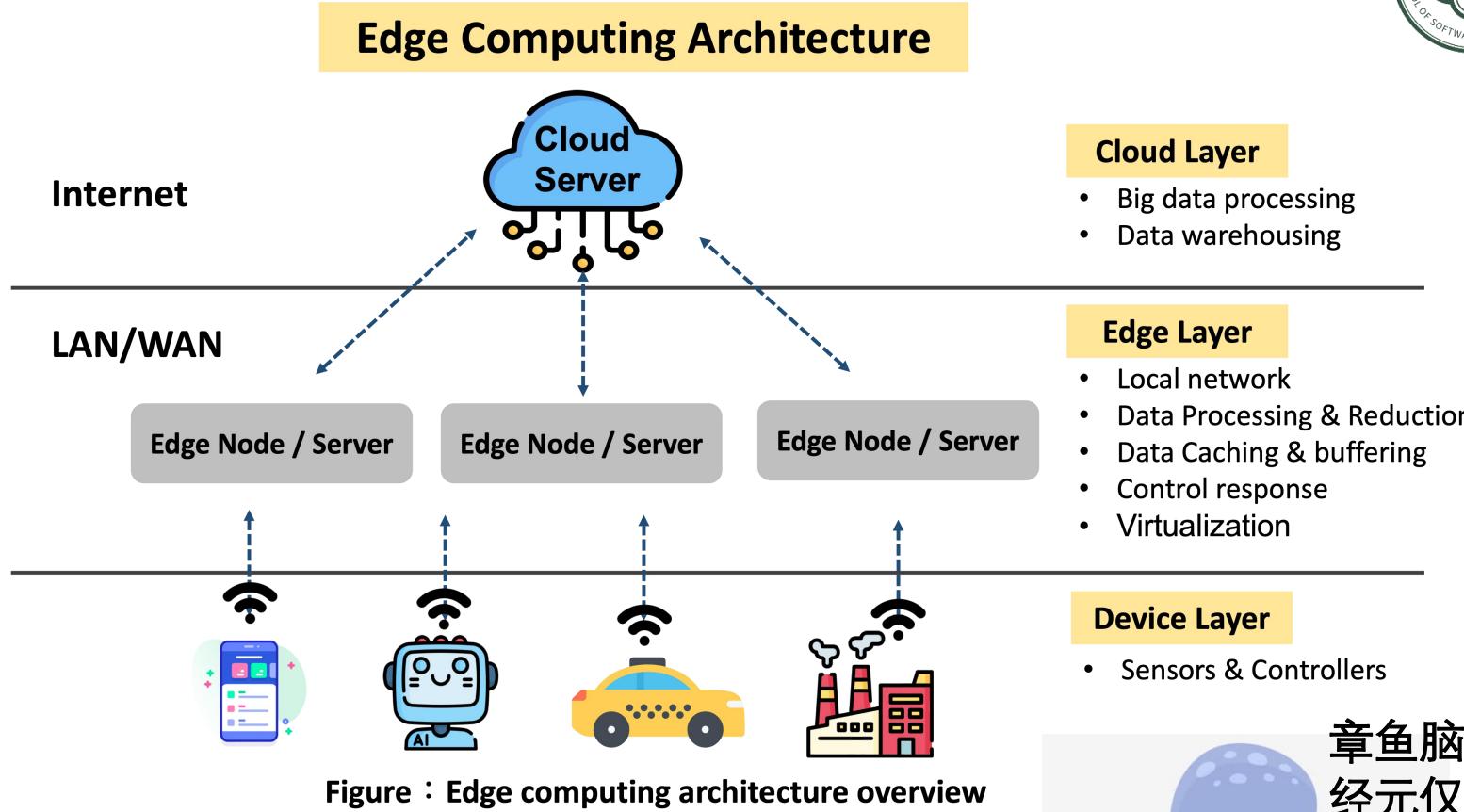


Gartner: Around 10% of enterprise-generated data is created and processed outside a traditional centralized data center or cloud. By 2025, Garter predicts this figure will reach 75%.

边缘计算发展过程中的代表性事件



边缘计算架构

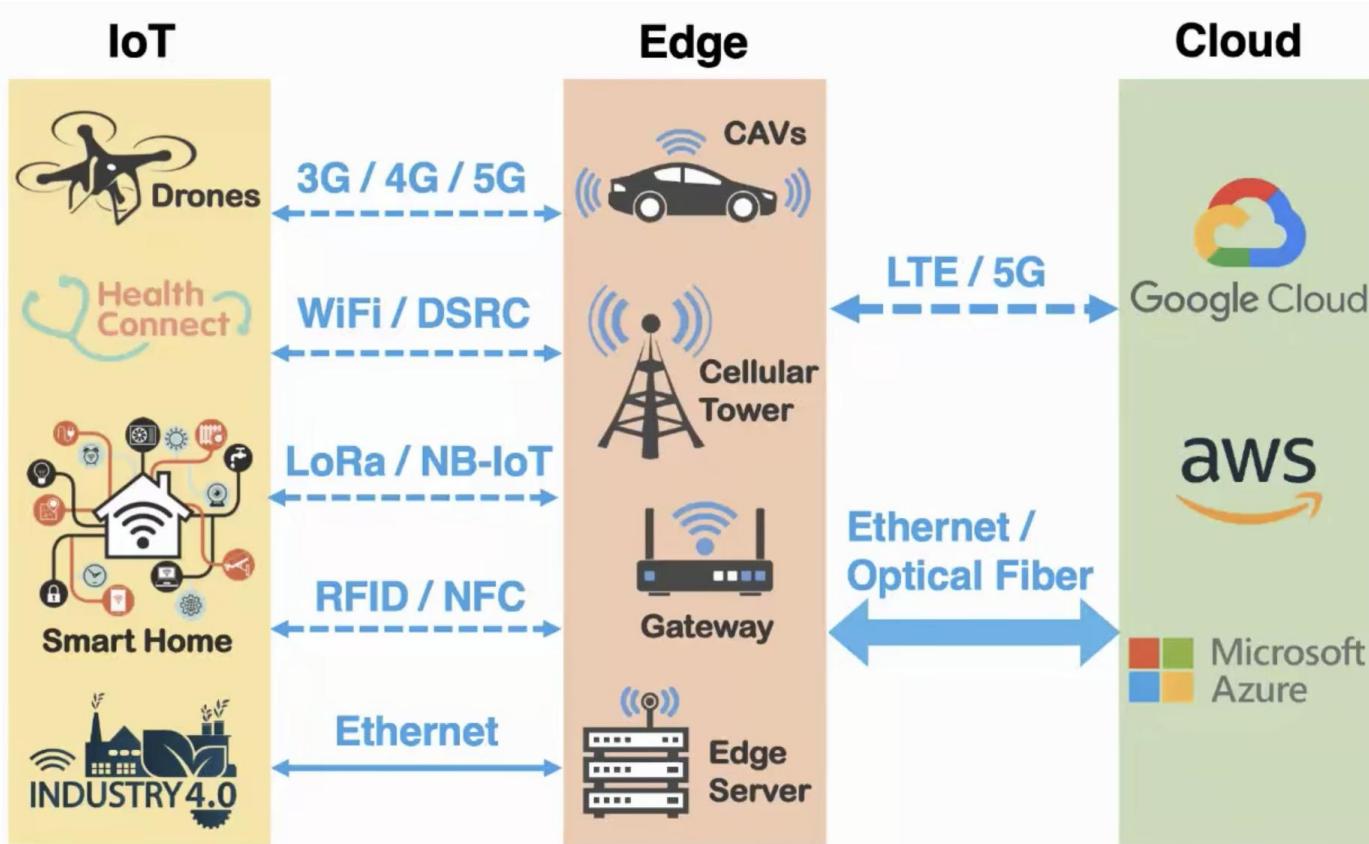


其余 60% 都分
布在八条触角上



章鱼脑部的神
经元仅占 40%

边缘计算的三层模型图

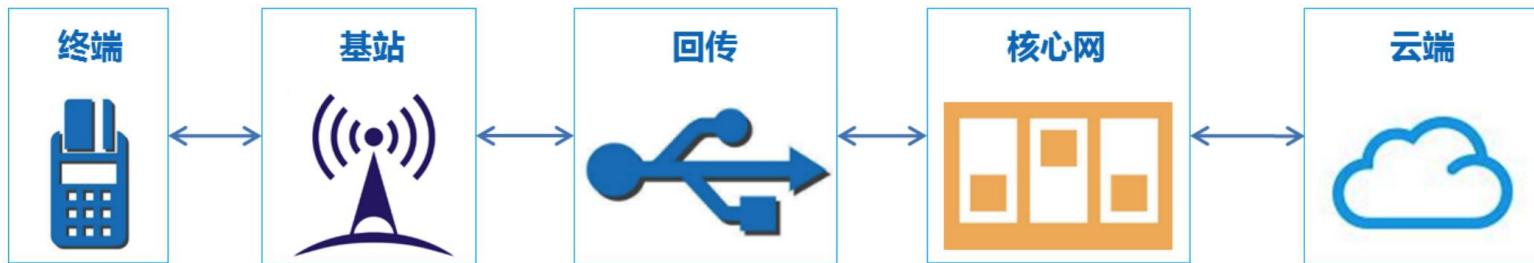


边缘计算的边缘是指从数据源到云计算中心路径之间的任意计算资源和网络资源

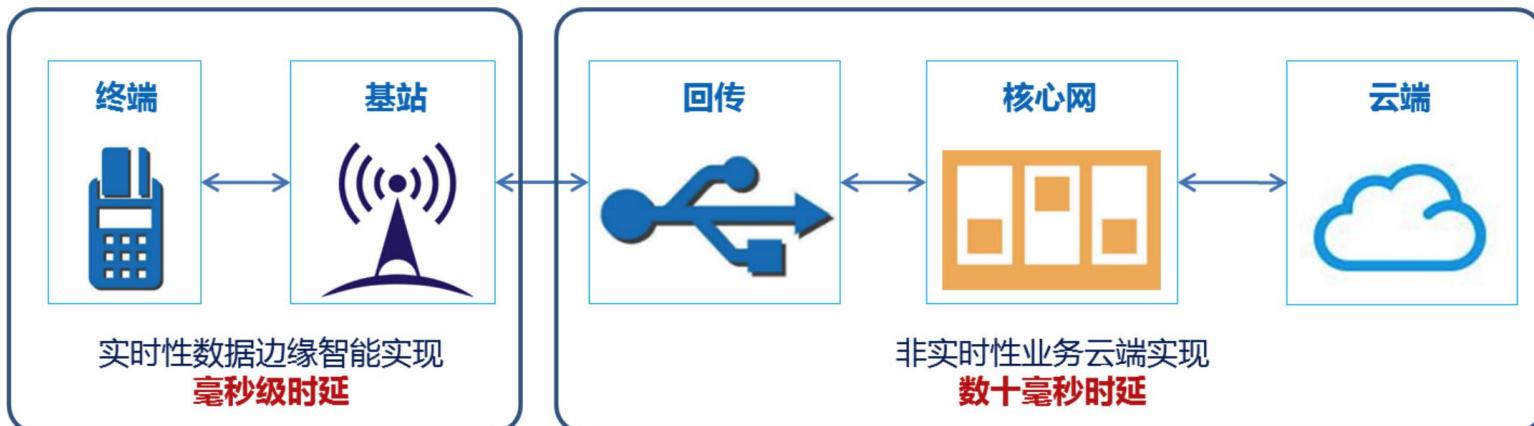
边缘计算的优势



- 低时延



从终端到核心网之间的时延为 20-30 ms



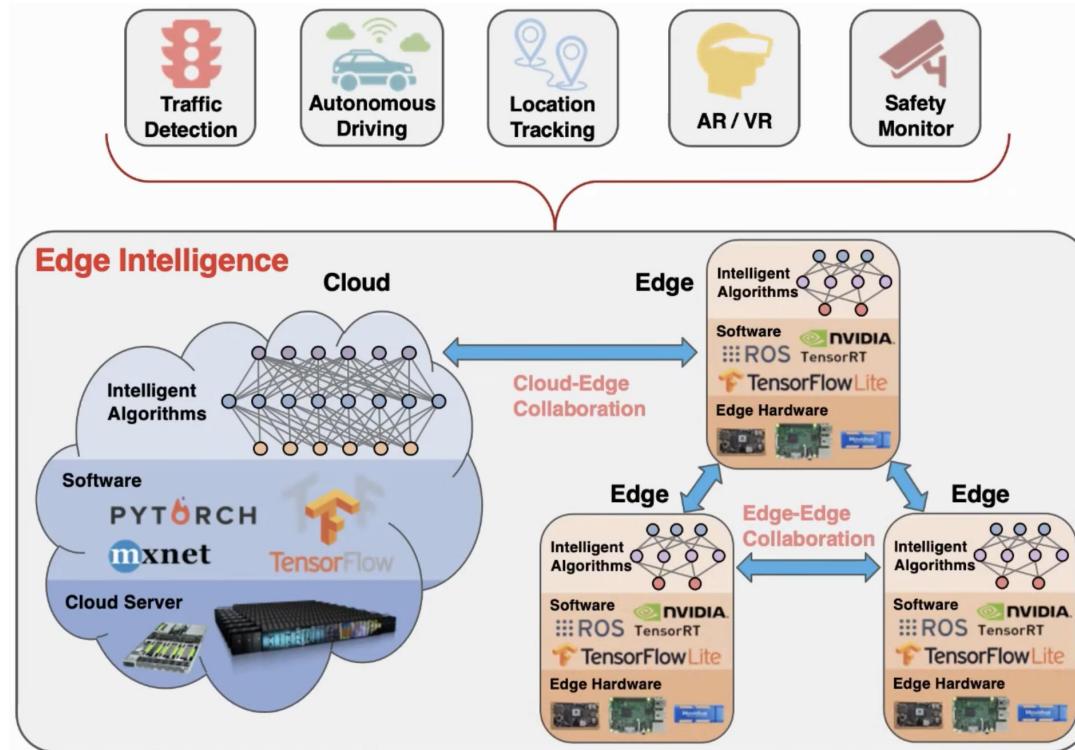
通过边缘计算实现**毫秒级**时延

边缘计算的优势



- 智能计算

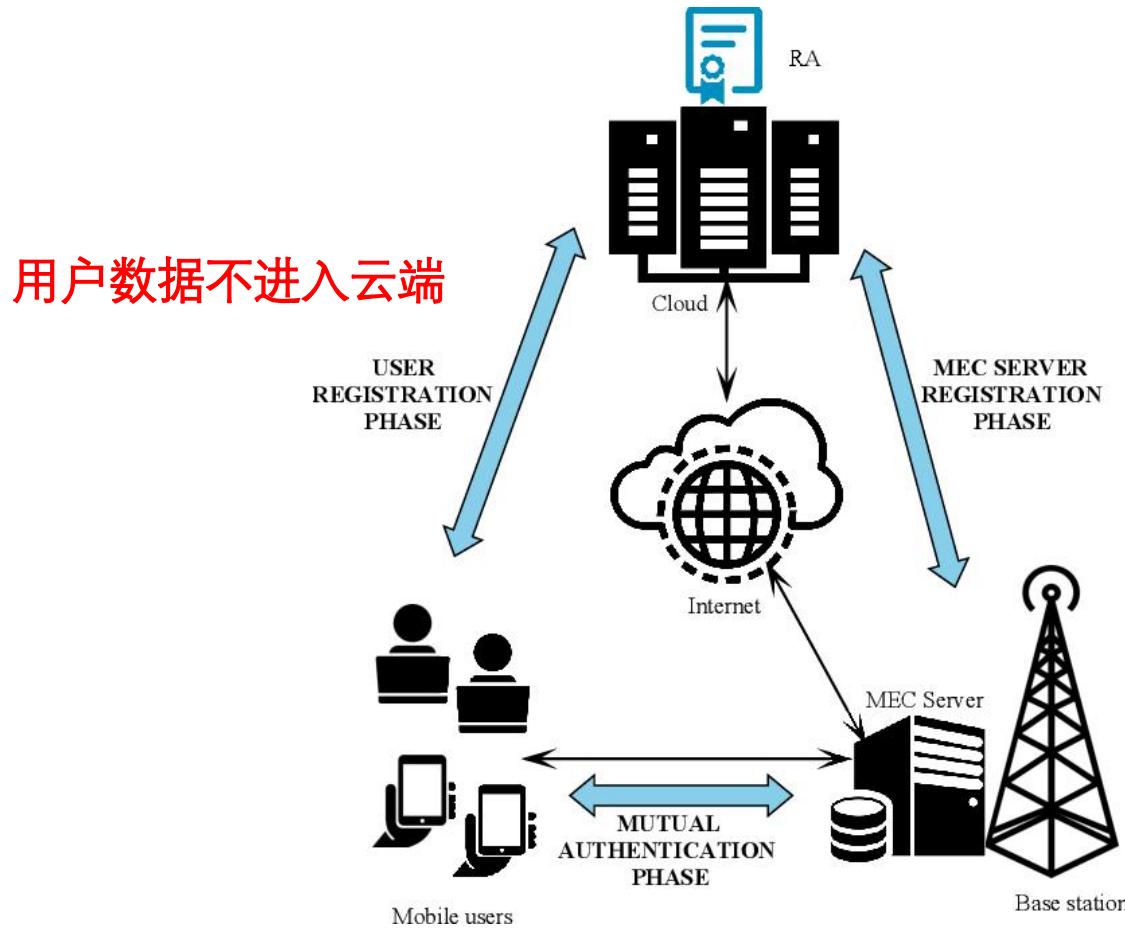
- ✓ 公共服务：人脸识别、入侵检测
- ✓ 车路协同：车辆碰撞预警和决策
- ✓ 工业智造：检测工业产品质量
- ✓ 虚拟现实（VR）或增强现实（AR）：边缘节点进行图像渲染



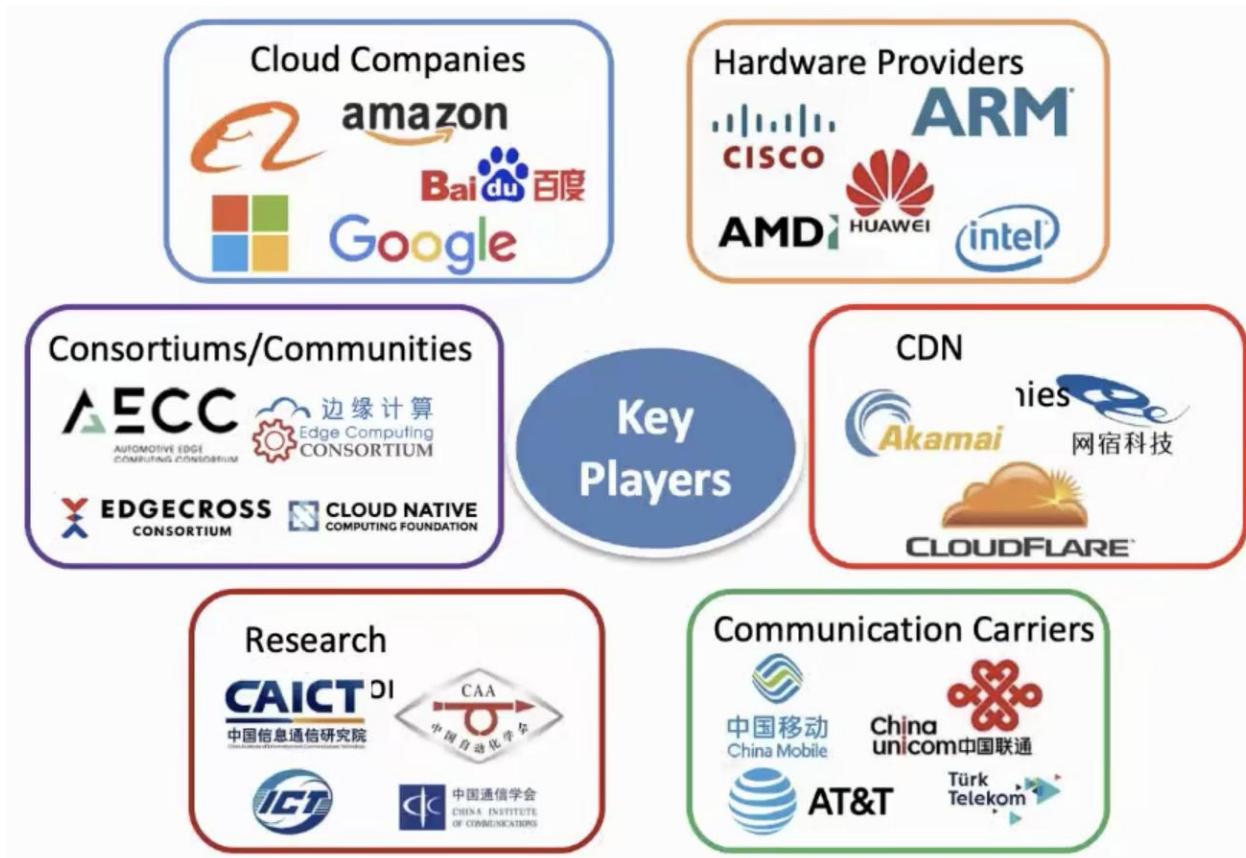
边缘计算的优势



- 数据安全与隐私



边缘计算参与者

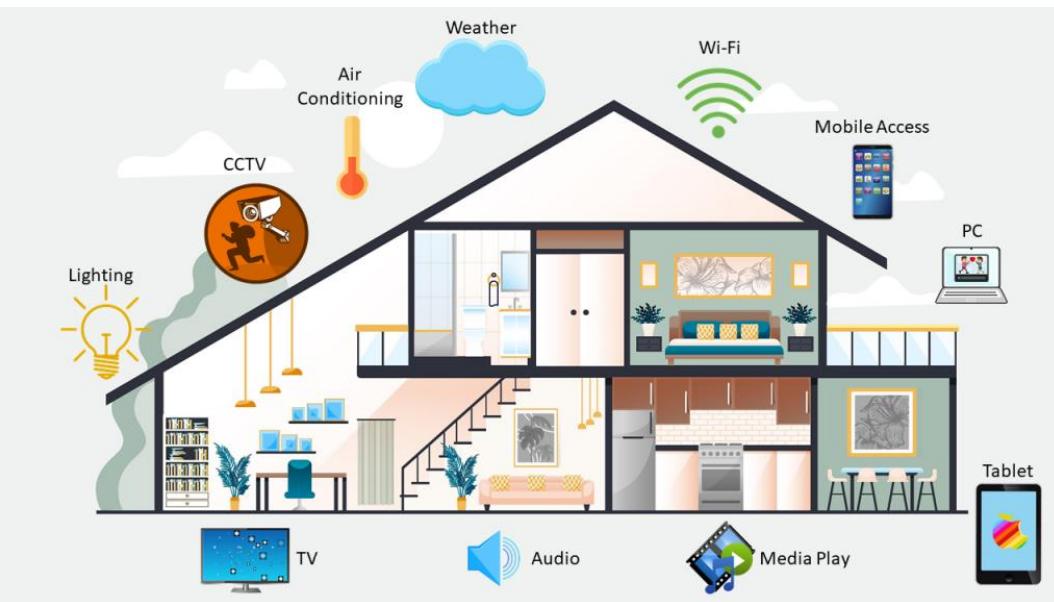


边缘计算的杀手级应用（Killer App）是什么？

智能家居



- 从简单的温度传感器到复杂的自动化系统，如智能照明、清洁和家庭娱乐系统



- 苹果 HomeKit
- 谷歌Home
- 亚马逊Alexa/Echo
- 三星 SmartThings
- 百度 DuerOS
- 阿里 AliGenie/天猫精灵 ×1
- 腾讯云小微/亲见 H2
- 京东 O2O 平台
- 小米盒子/中控路由器/台灯等

智能家居

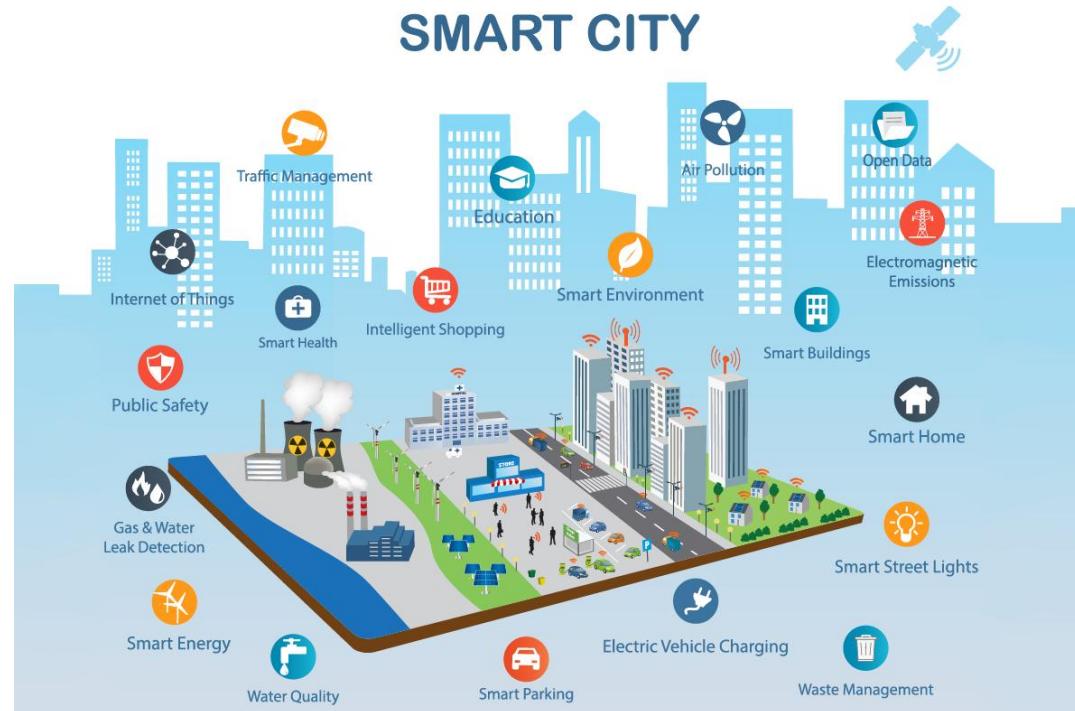


- 传统方式下，由于家庭里没有公网IP地址，要实现对家居设备的远程操控，必须通过云端的辅助来打通内网和外网的连接壁垒
- 而边缘计算能给智能家居提供很好的支撑
 - ✓ 一方面，边缘计算可以利用专用可靠的本地服务来处理家居应用产生的大量网络数据
 - ✓ 另一方面，边缘计算可以减少家居应用数据的泄露风险

智慧城市

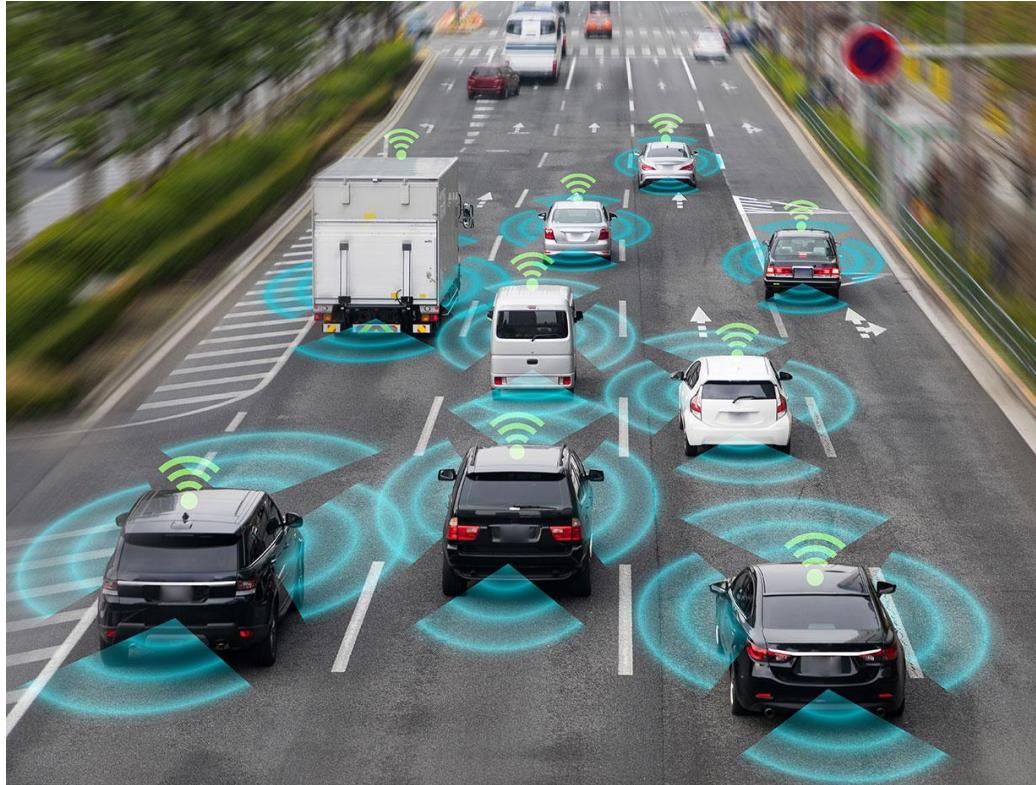


- IoT 技术已经从家庭发展到社区，甚至是城市规模的应用
- 智慧城市涵盖公共安全、医疗保健、旅游和运输等行业
 - 城市中的摄像头可以将视频流传送到边缘服务器进行实时处理和异常检测
 - 医疗保健需要与来自医院、药房、保险公司、物流公司和政府等多个实体协作

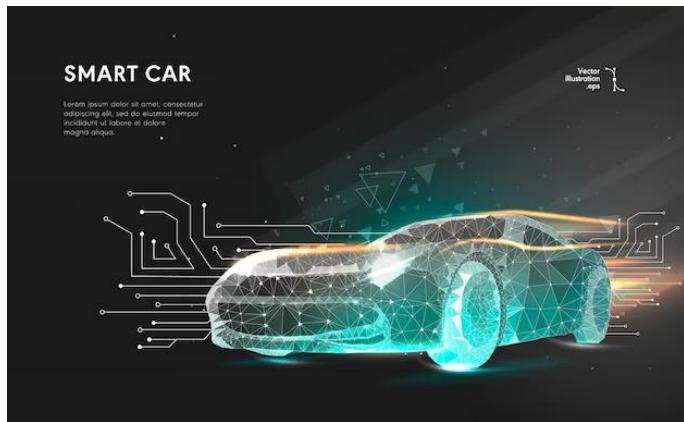




- 典型的车联网应用如自动驾驶、车载通信、车辆间的连接、车辆与道路基础设施的连接等，涉及多种类型大量数据（如视频、音频、信号数据）的处理



下一代个人智能设备？



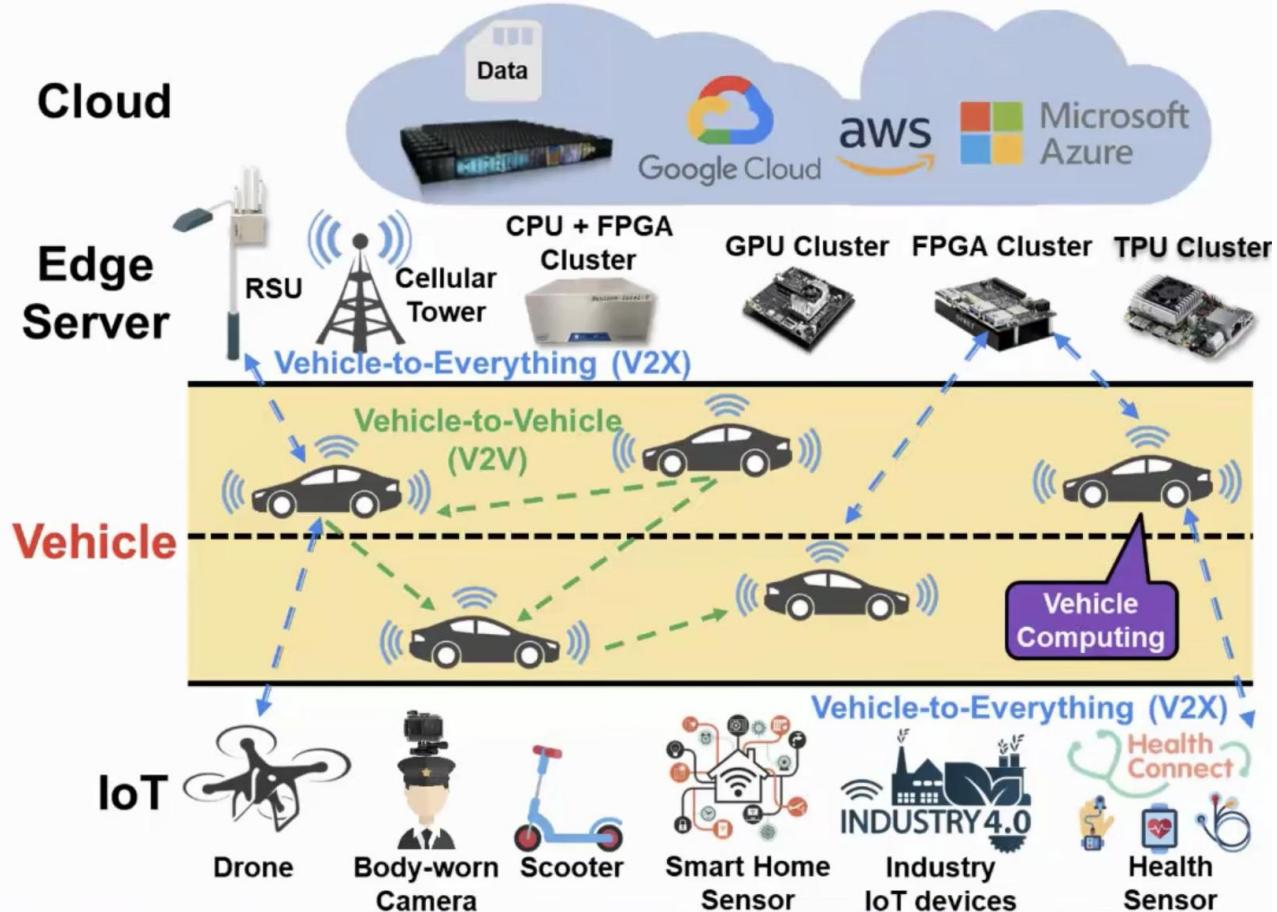
Vehicle computing



The 4-Tier Vehicle Computing Paradigm

Computing on CVs based on data from:

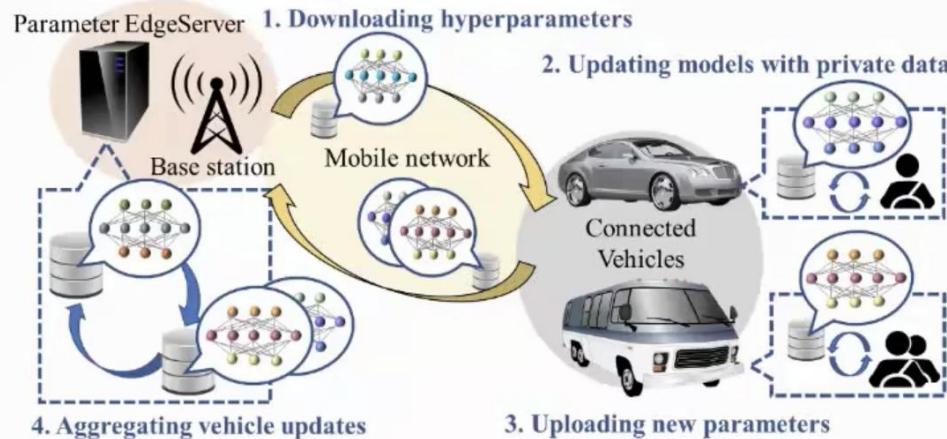
- In-vehicle sensors,
- Surrounding **connected** devices



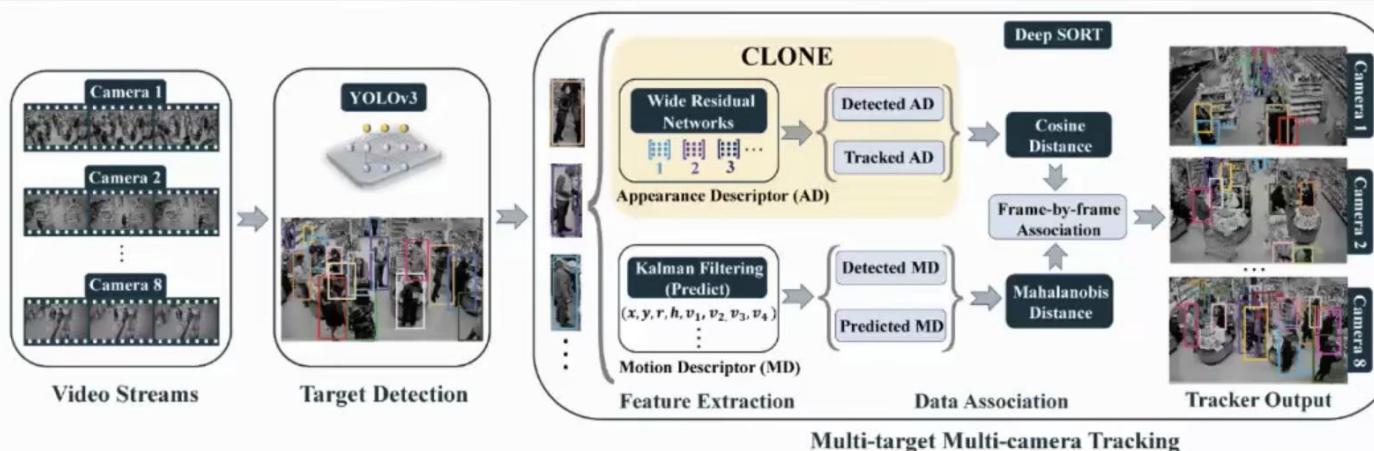
边缘计算在车联网中的应用



• 汽车电池寿命预测



- ❖ **CLONE**: a collaborative learning framework on the edges
- ❖ **CLONE_training**: EV battery failure prediction
 - Latency reduction
 - Privacy-preserving
- ❖ **Results**:
 - Reduce model training time significantly
 - Achieve equal or even higher accuracy



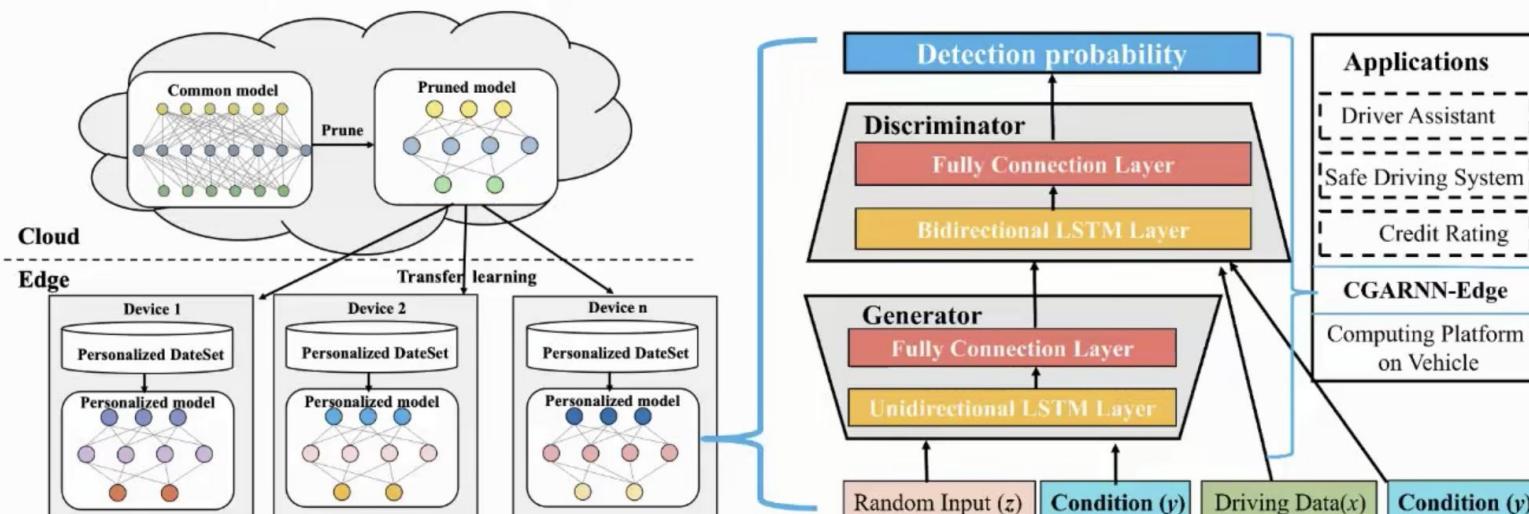
- ❖ **CLONE_inference**: Multi-camera multi-target tracking
- ❖ 8 cameras
- ❖ Comparable accuracy and speed as stand-alone learning

边缘计算在车联网中的应用



- 驾驶行为建模

- Collaborative Cloud-Edge Computation for personalized Driving BEhAvior Modeling
 - Models personalized driving behaviors based on driving data and contextual data.
 - The GARNN-based model adapts to the dynamic change of normal driving.
 - Developers build third-party applications on top of pBEAM.



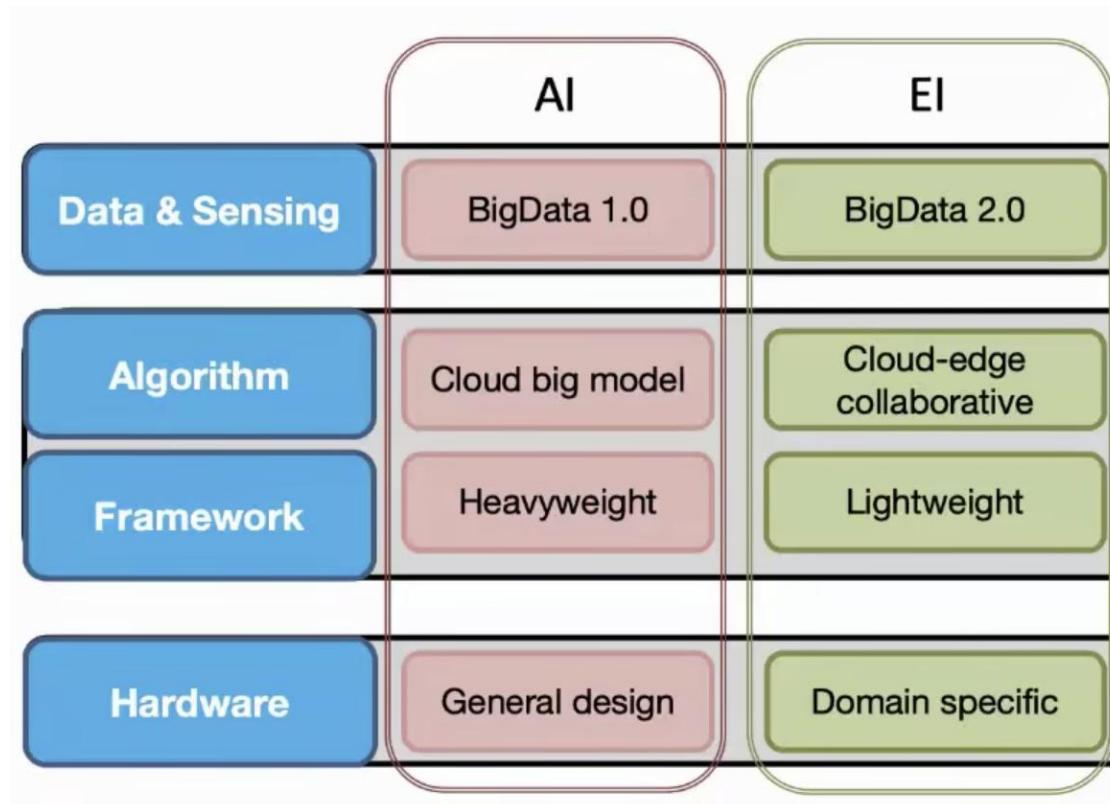
Overview of pBEAM

Model (LSTM + Conditional GAN)

从 CI 到 EI



- 从 Cloud intelligence 到 Edge intelligence

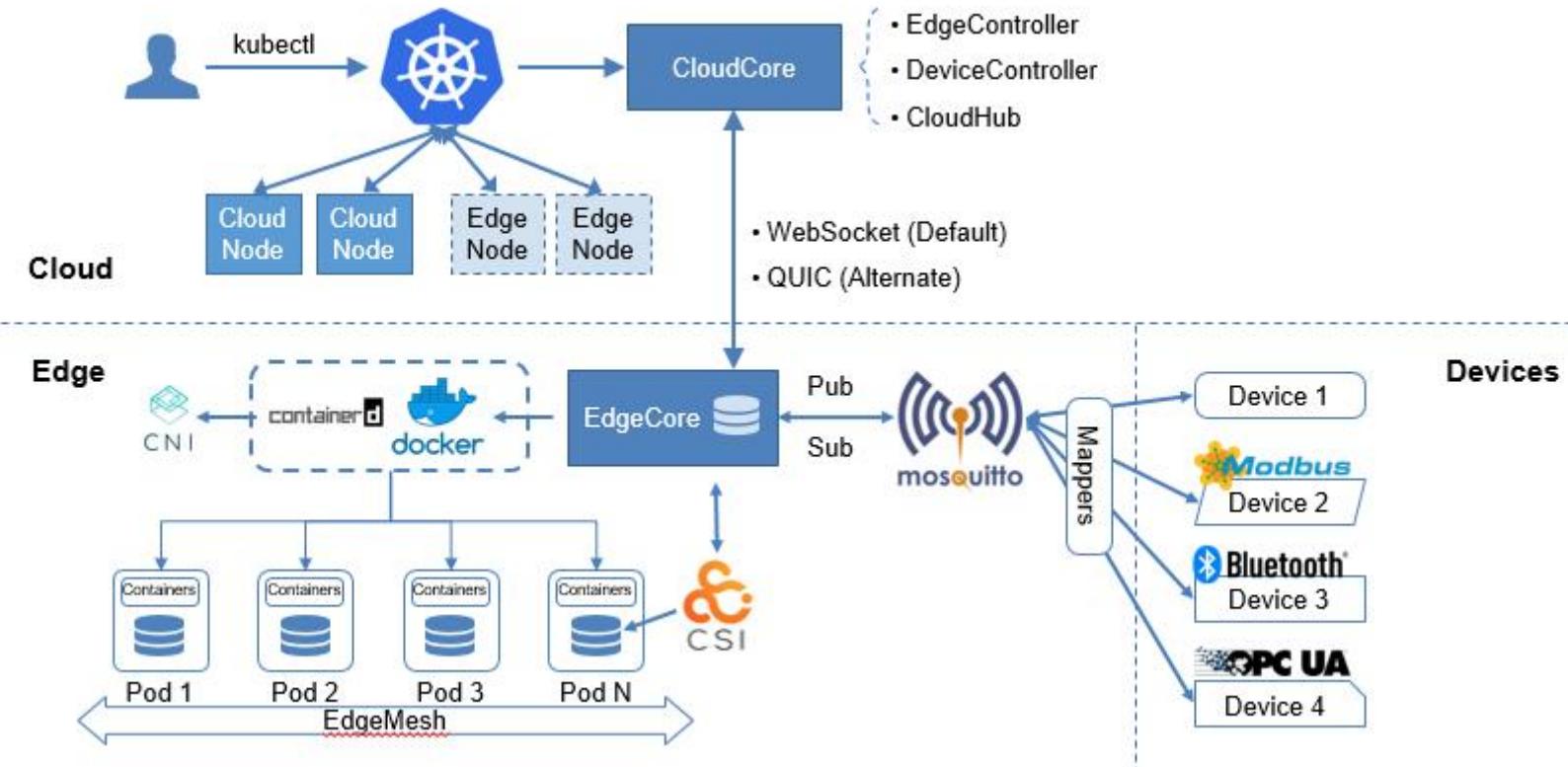


主流边缘计算平台



- 华为 KubeEdge

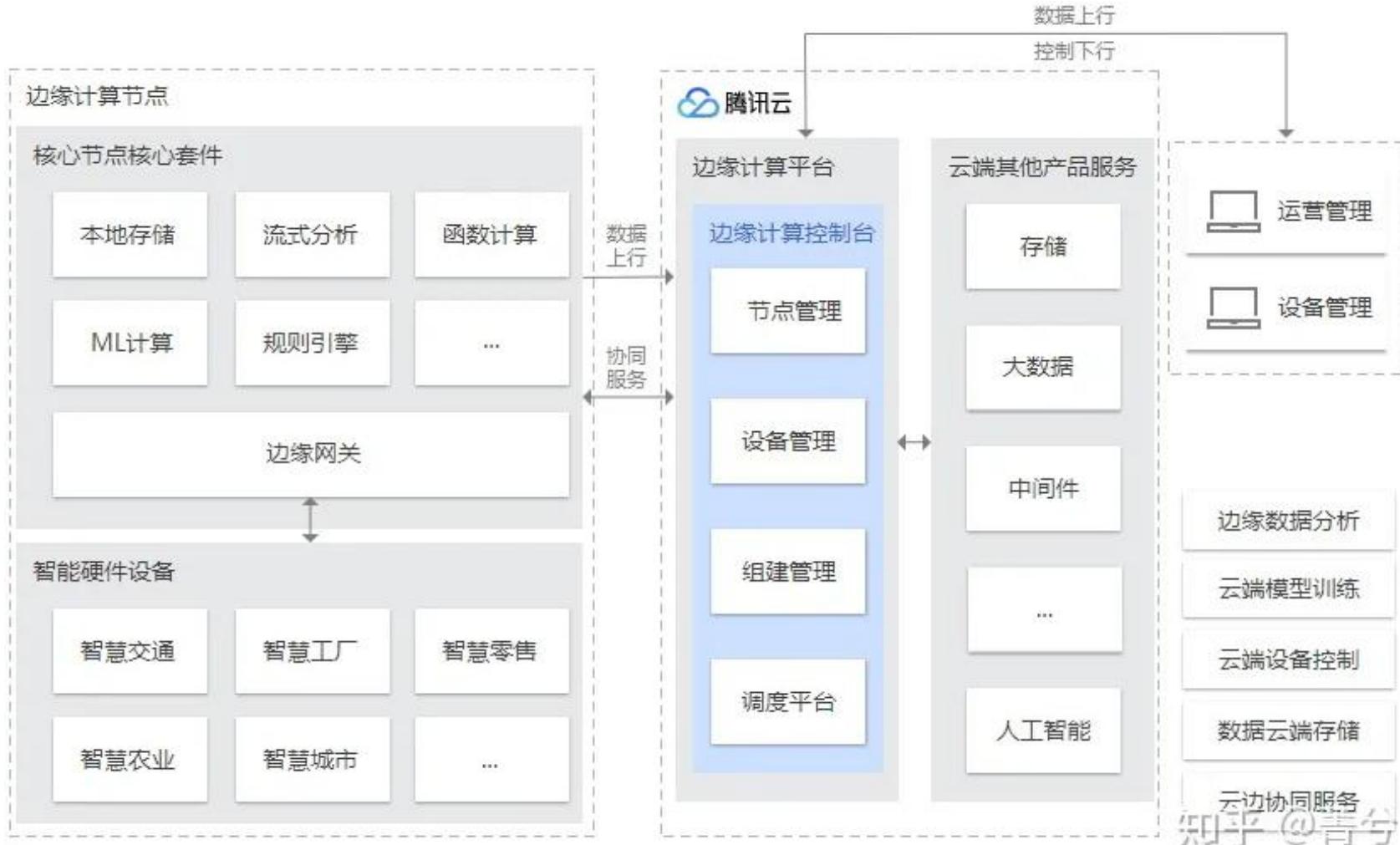
- ✓ 基于 Kubernetes，将本机容器化和设备管理扩展到边缘端的主机



主流边缘计算平台



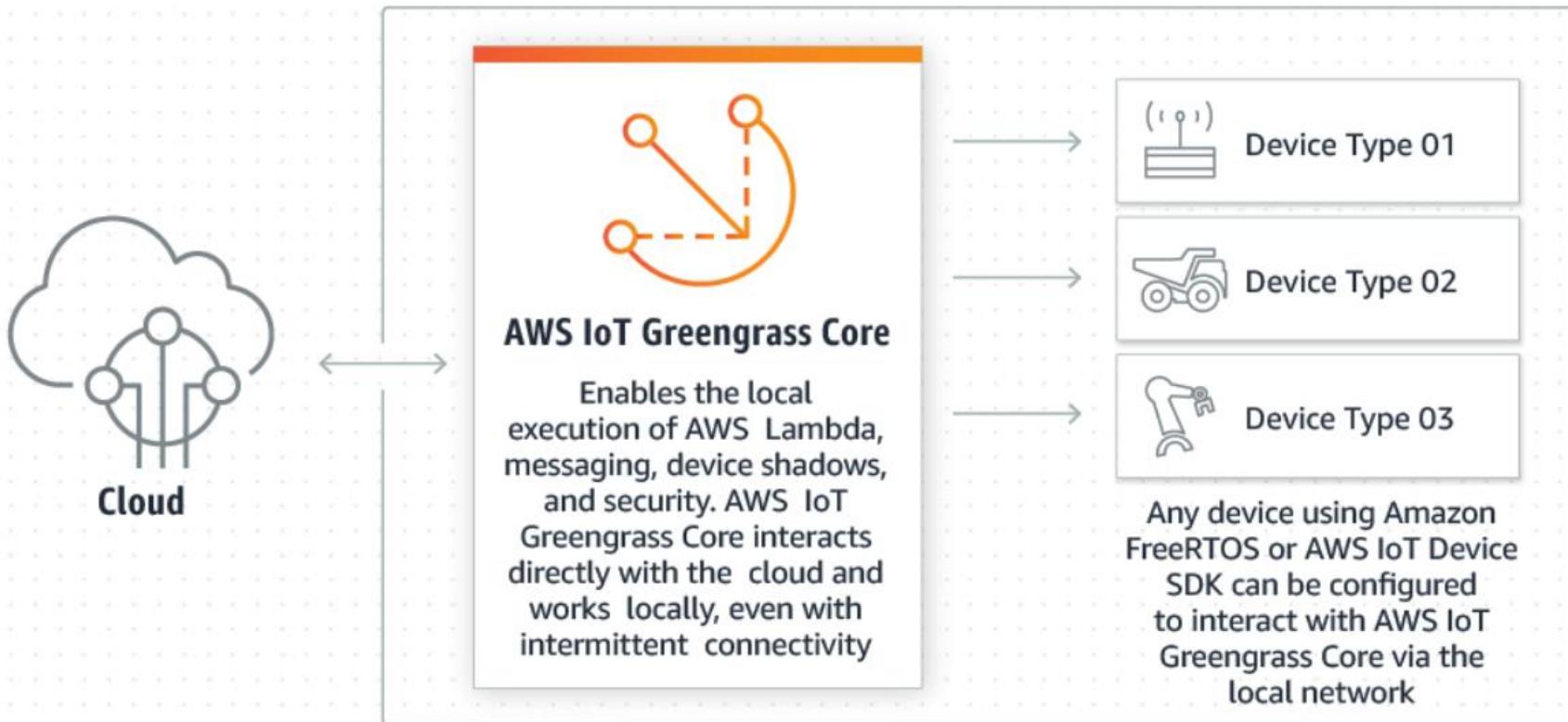
- 腾讯 IoT 边缘计算平台 IECP



主流边缘计算平台



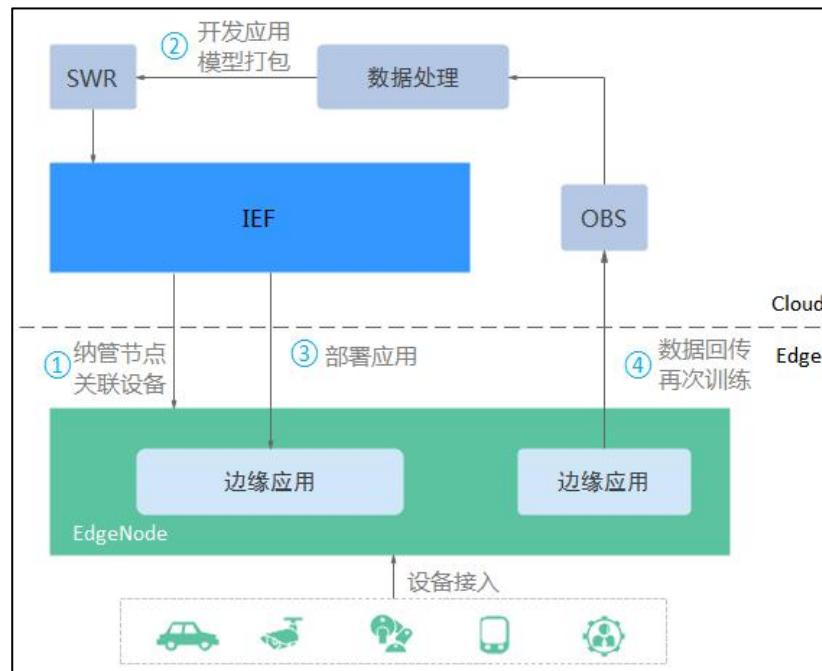
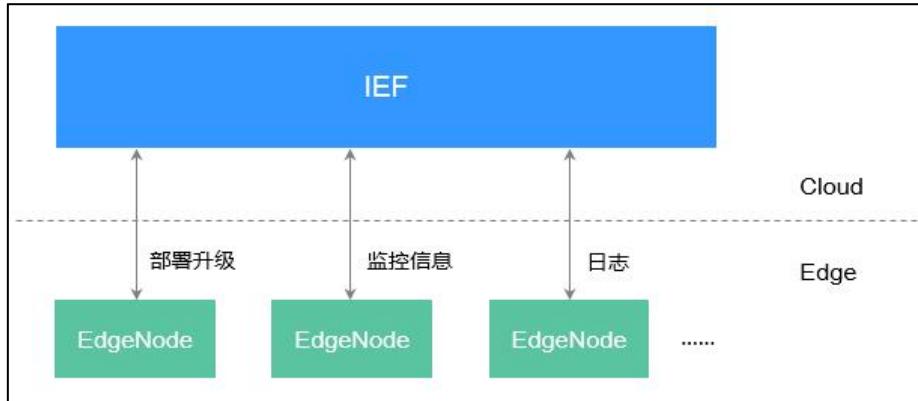
- AWS IoT Greengrass



主流边缘计算平台



- 华为智能边缘平台 IEF

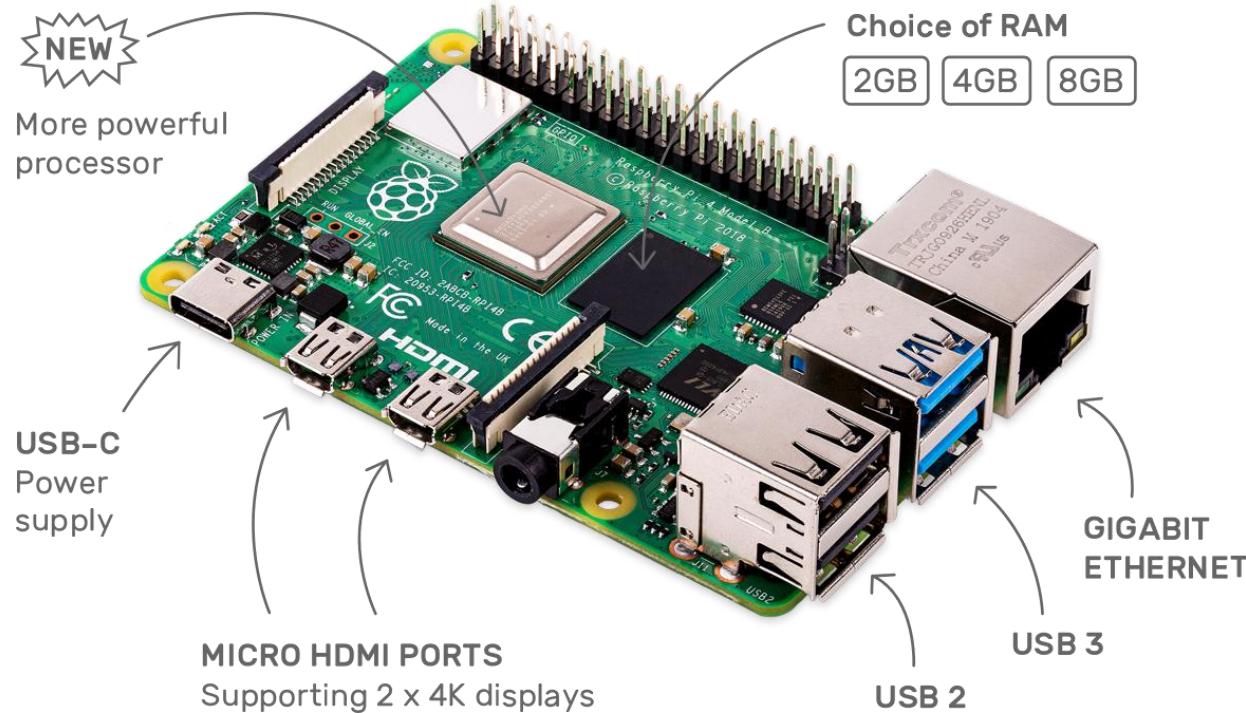


跨平台边缘设备



- 树莓派 (Raspberry Pi)

- ✓ 微型计算机，通常基于Linux，可以运行各种编程环境和工具



跨平台边缘设备



- Jetson Nano
 - ✓ Nvidia 开发的微型人工智能计算机，搭载了 Nvidia 的 GPU



跨平台边缘设备



- 华为 Atlas

- ✓ Atlas 500智能小站（型号：3000）是面向边缘应用的产品，具有超强计算性能、体积小、环境适应性强、易于维护和支持云边协同等特点

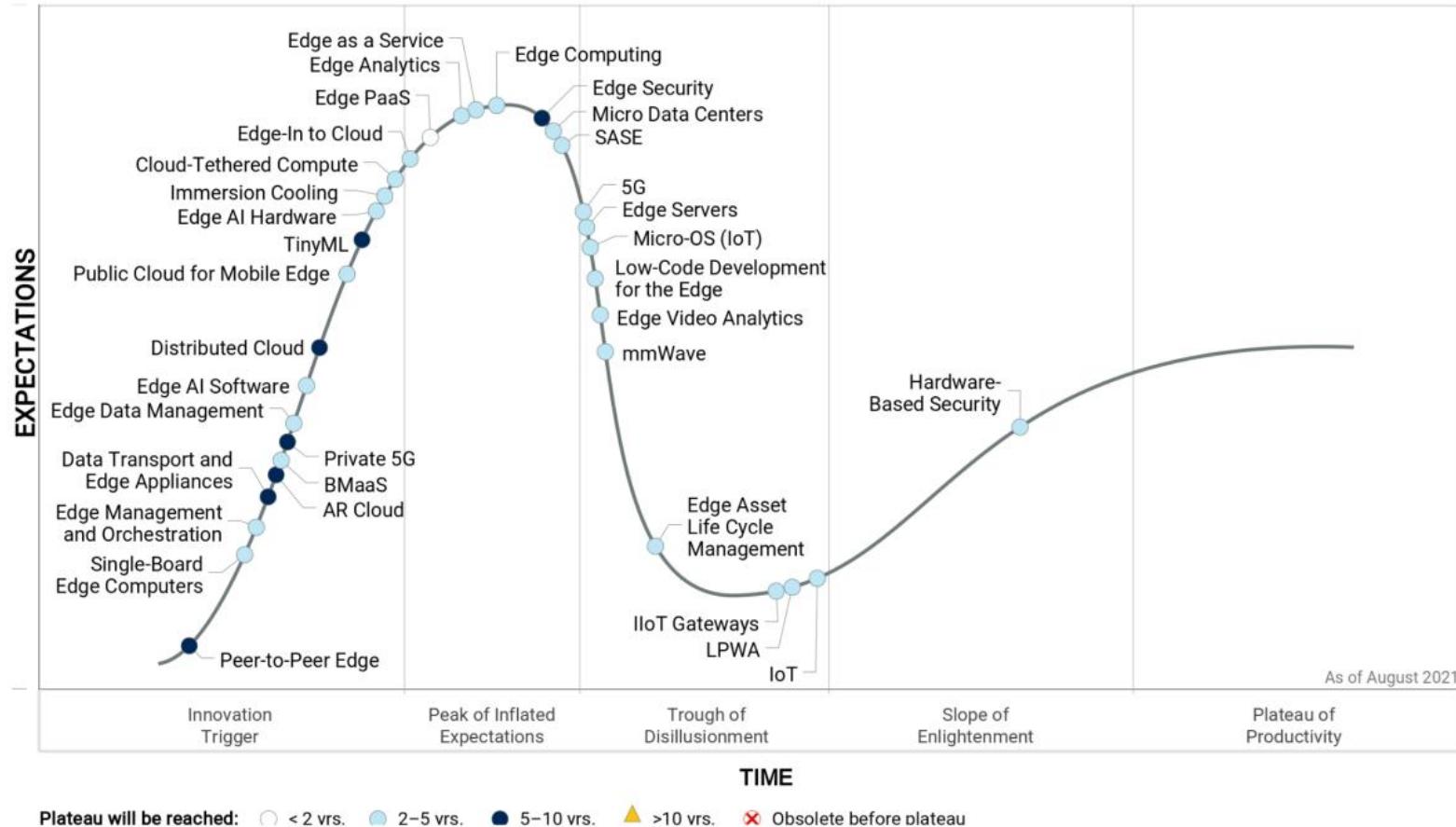


Atlas 500 智能小站

Gartner 技术成熟度分析



Hype Cycle for Edge Computing, 2021



Source: Gartner (August 2021)

747550

边缘计算的挑战



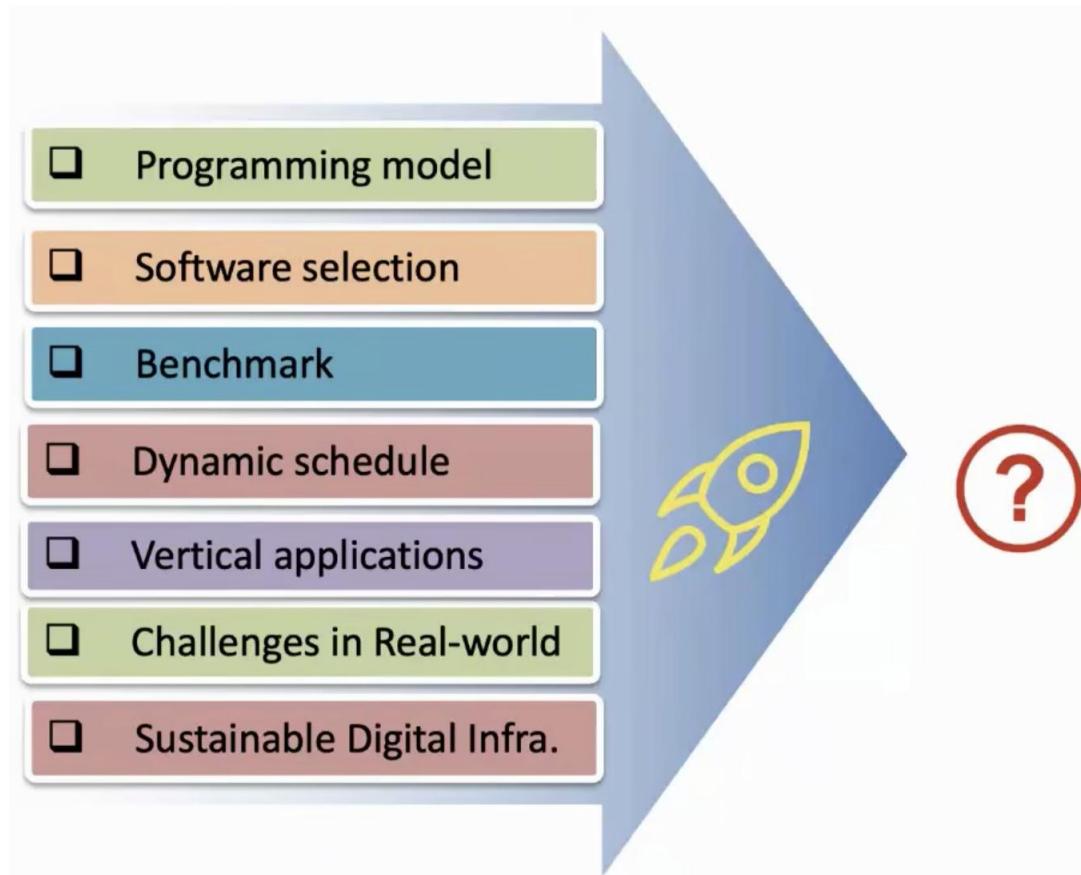
- 边缘节点的通用计算
 - ✓ 边缘设备数量庞大、分布广、异构
 - ✓ 需要形成统一接口规范、传输协议、时钟同步等问题
- 任务划分和分配
 - ✓ 传统云计算中，应用被建模成具有依赖关系的子任务，用 DAG (Directed Acyclic Graph) 图分析
 - ✓ 边缘计算面临任务类型复杂、数量庞大、到达时间不确定、设备可靠性等挑战
- 分布式存储和协同计算
 - ✓ 协同异构且数量庞大的边缘设备
 - ✓ 动态调整计算任务



边缘计算的挑战



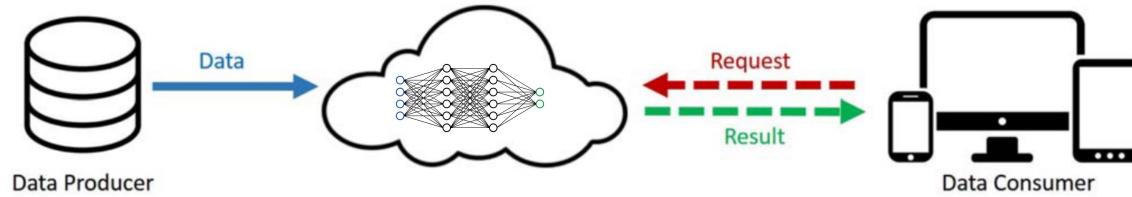
- 以一个具体的场景“车联网”为例，以下是常见的挑战



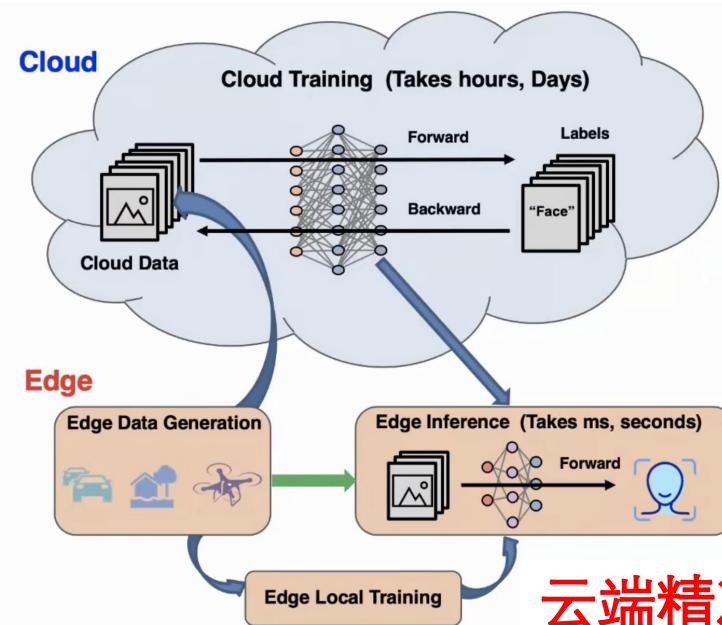
AI 应用构建流程



- 传统云计算中的 AI 应用构建流程



- 边缘计算中的 AI 应用构建流程



云端 AI 的挑战



- 高时延与高成本使得将大量边缘端数据传输到云端数据中心变得不切实际
- 大规模数据的采样带来知识获取的精度损失
- 部分项目数据的隐私性和实时性需求导致数据迁移到云端数据中心的方案不可选

边缘计算中的机器学习计算任务



- 云端训练，边缘端推理
- 边缘端训练，边缘端推理
- 云边协同的训练或推理

云边协同的机器学习任务优化

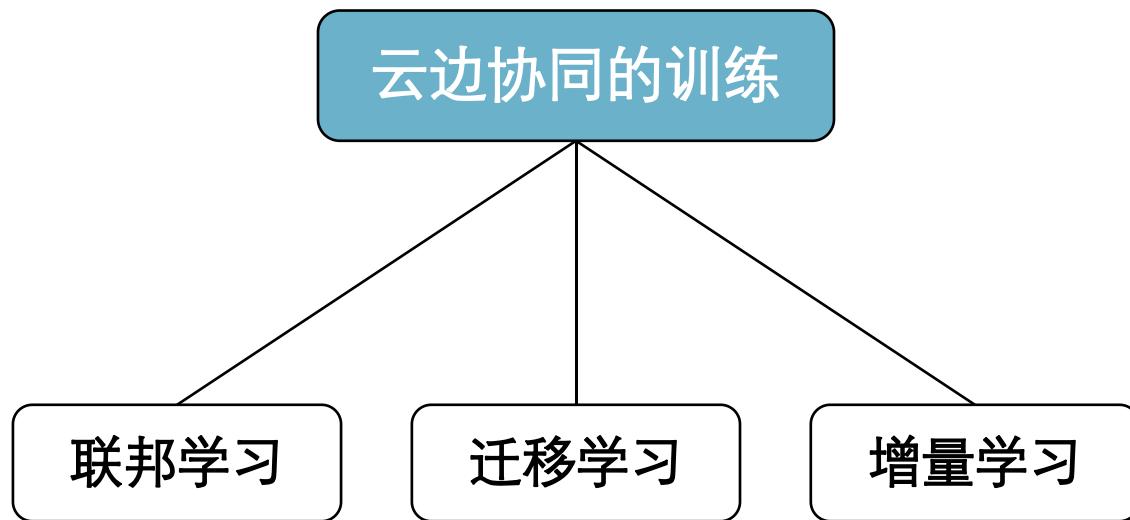


云边协同的训练

云边协同的推理

模型的训练和推理不仅发生在中心云上，也发生在网络边缘的设备上，如智能手机、IoT 设备等

云边协同的训练算法



云边协同的联邦学习



- 联邦学习 (Federated Learning)

- ✓ 一种机器学习的框架，它使得多个设备（例服务器）可以**共享模型的学习过程，而不需要直接共享数据**，保证了数据隐私和安全

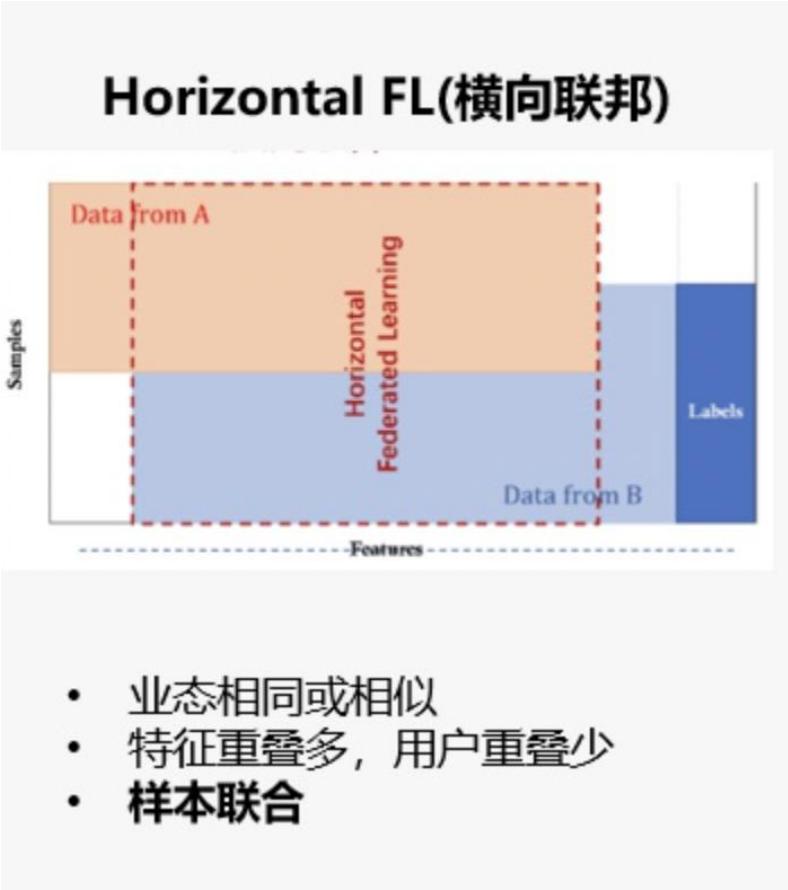
- 联邦学习的基本过程如下：

- ✓ 在中央服务器上初始化一个模型
 - ✓ 将模型发送到各个设备
 - ✓ 每个设备利用本地数据对模型进行训练并更新模型参数
 - ✓ 每个设备将**模型参数的更新**发送回中央服务器
 - ✓ 中央服务器汇总所有设备的模型更新信息，更新全局模型
 - ✓ 重复以上步骤直到模型收敛

横向联邦学习



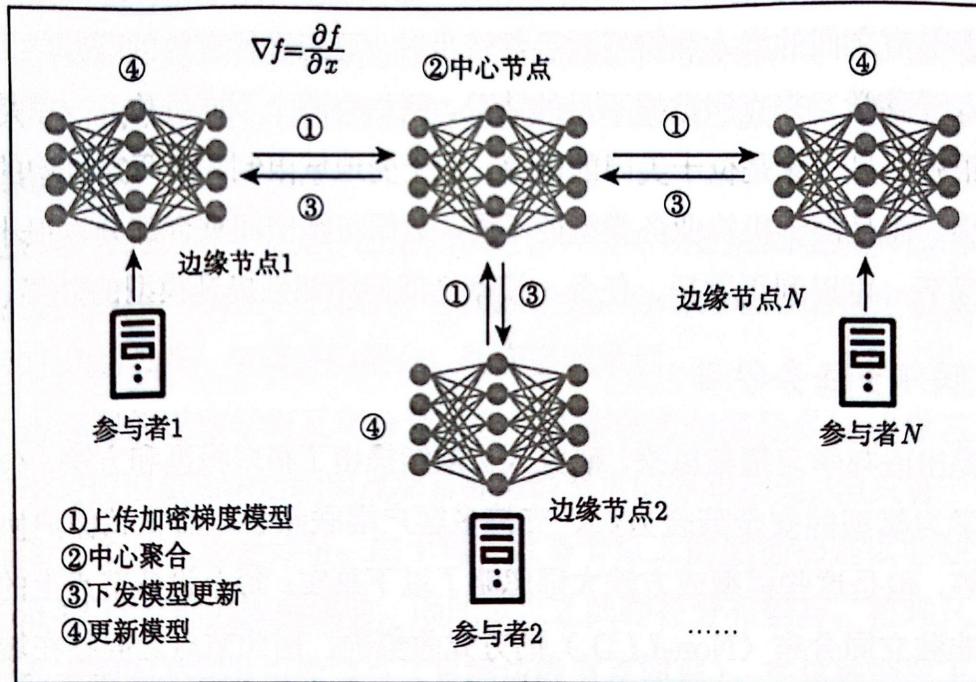
Horizontal FL(横向联邦)



一个区域性银行在两个地方运营，用户交集少，但业务非常相似，也就是特征空间是相同的。

可看成是基于样本的分布式模型训练，数据分布在不同的机器上。

横向联邦学习框架

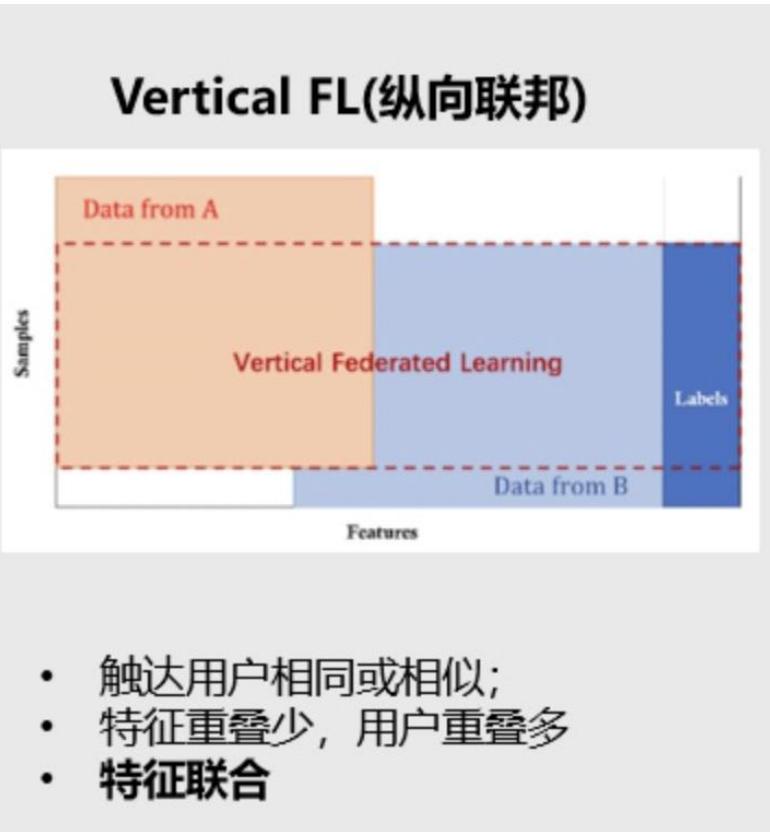


1. 边缘节点基于预处理数据和原始模型进行本地训练，生成梯度信息加密上传至中心节点
2. 中心节点收集边缘节点梯度信息后进行聚合
3. 中心节点下发聚合后的模型
4. 边缘节点更新自己的模型

纵向联邦学习



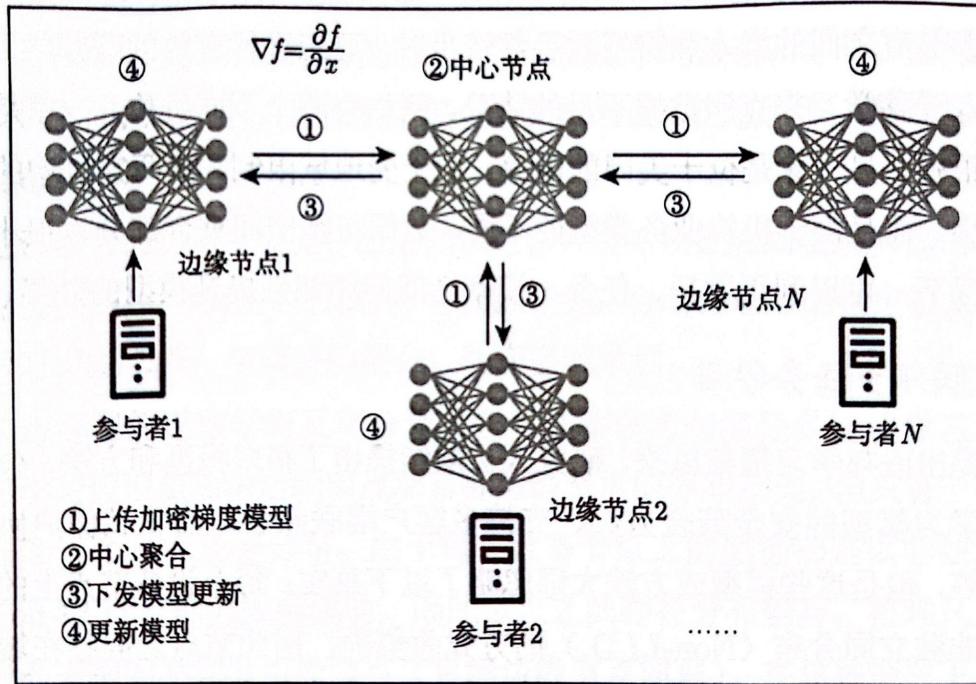
Vertical FL(纵向联邦)



同一个区域内的不同机构，一个是银行，一个是电子商务公司，用户空间交集大，但特征不同。

将不同的特征加以聚合，以增强模型的预测能力。

横向联邦学习框架

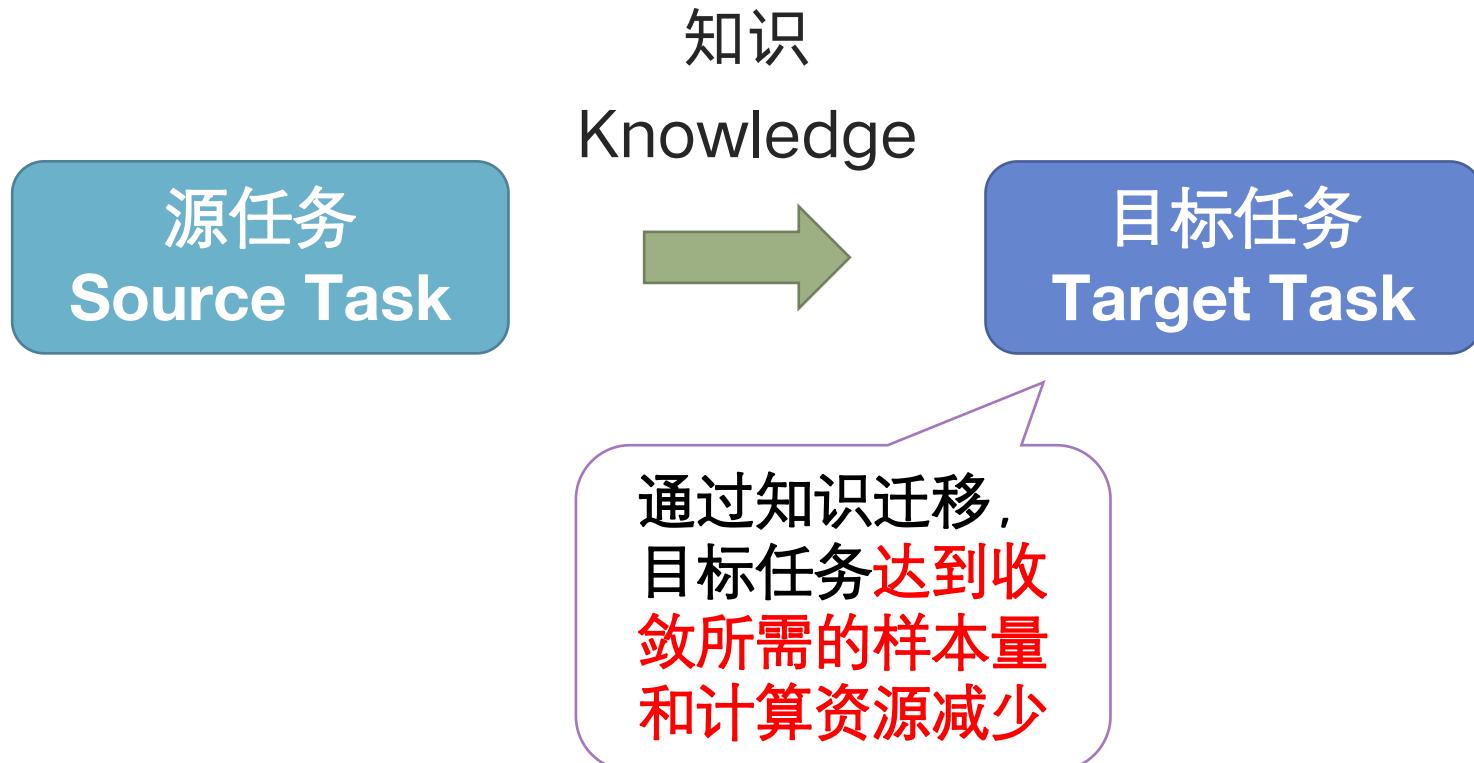


1. 在中心节点将加密样本对齐
2. 边缘节点基于对其样本分别计算和自己相关的特征空间，生成梯度信息加密上传至中心节点
3. 中心节点更新模型并分发给边缘节点
4. 边缘节点更新模型

云边协同的迁移学习



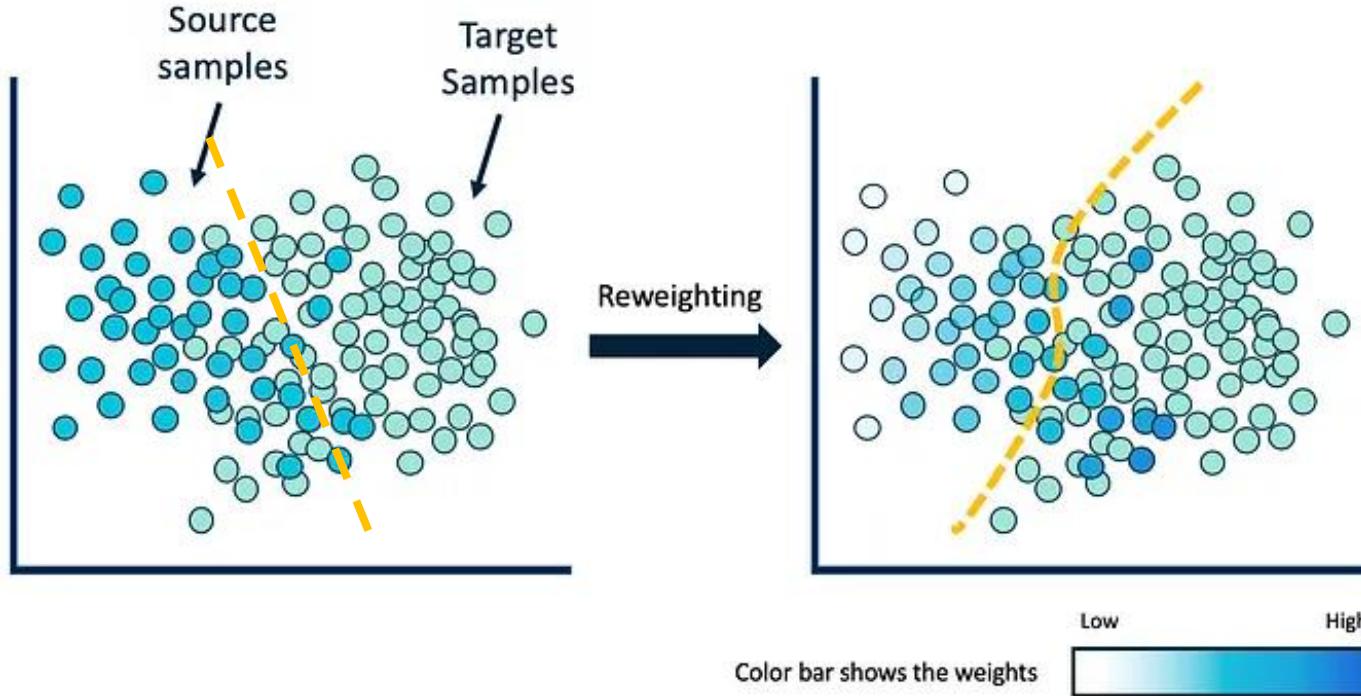
- 由于直接对每个边缘节点的任务从头开始训练成本太高，云边协同的迁移学习运用已有的相关知识来辅助尽快地学习目标边缘端的新知识



样本迁移



- 加权源任务样本（使其与目标任务的分布接近），并与目标任务共享加权样本



样本迁移的步骤



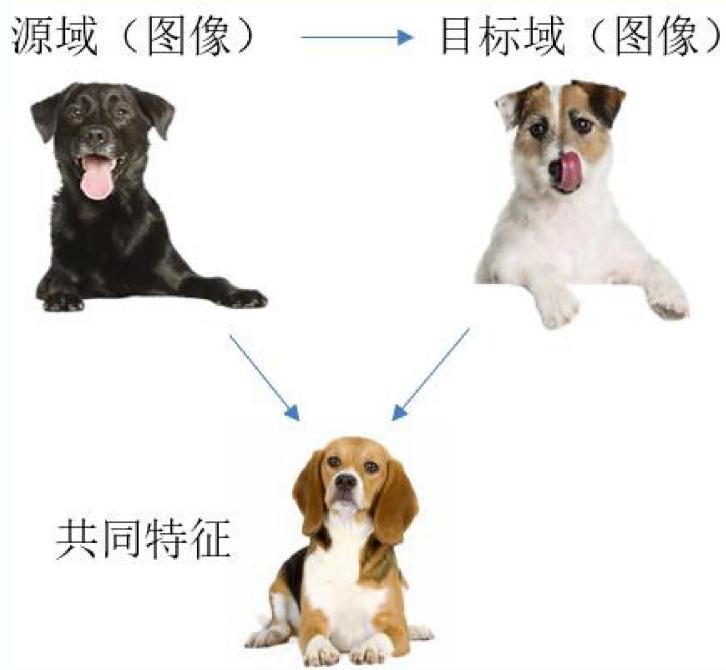
1. 提取源任务与目标任务的样本特征
2. 估计目标任务样本分布
3. 根据目标任务样本，调整源任务样本权重
4. 基于加权源任务样本与目标任务样本进行训练

特征迁移

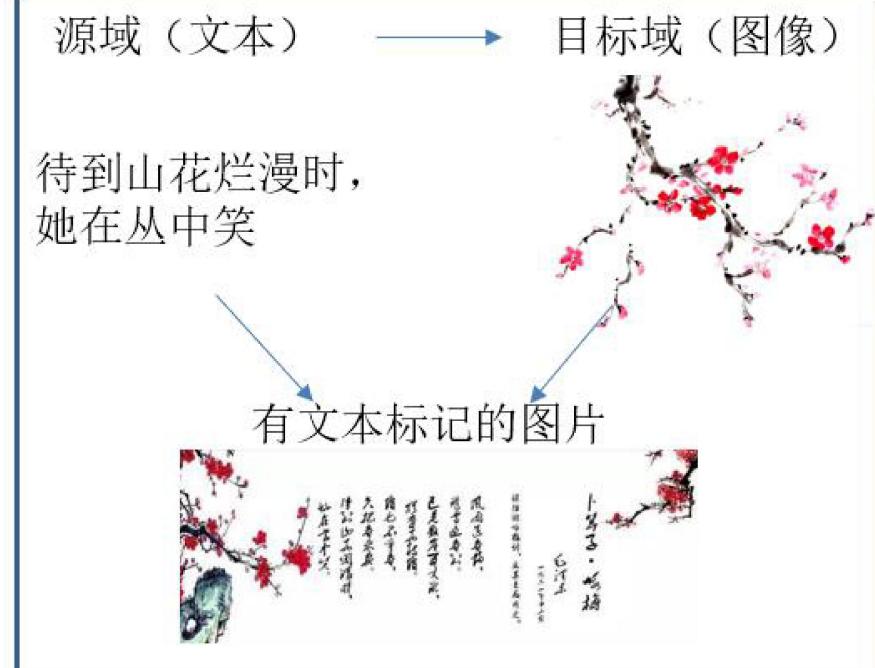


- 将源任务和目标任务的特征变换到相近空间后一起学习

源域和目标域特征空间一致



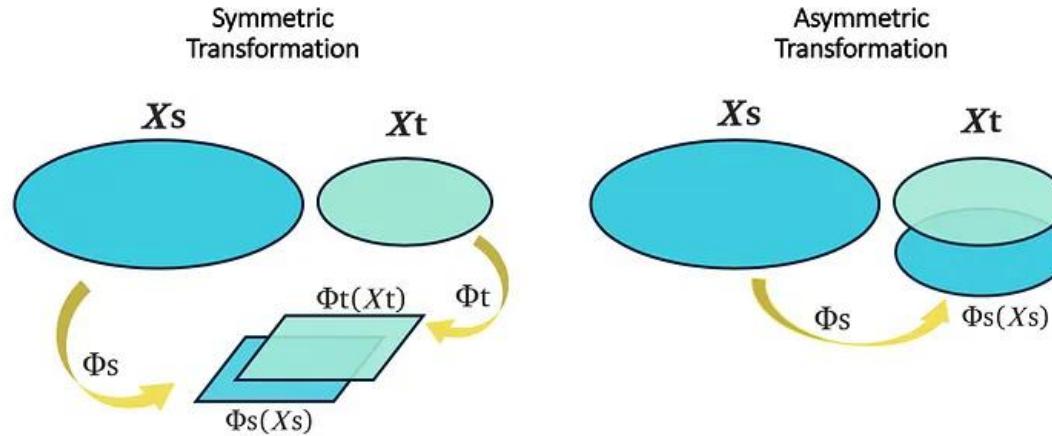
源域和目标域特征空间不一致



特征迁移的步骤



1. 分别在源任务和目标任务数据上提取特征
2. 将源和目标任务特征映射到同一空间，最小化分布距离
 - ✓ 将源和目标任务特征映射到同一空间
 - ✓ 将源任务特征映射到目标任务特征空间



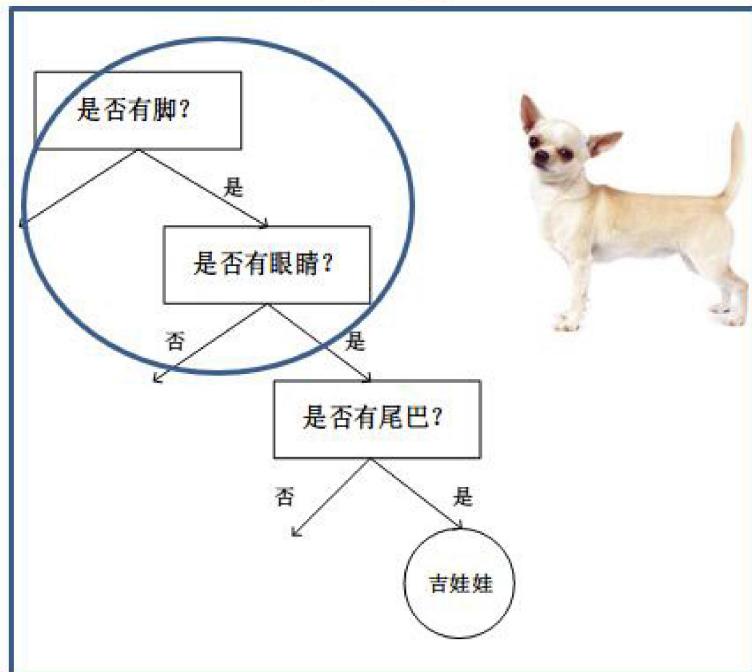
3. 基于映射后的源任务和目标任务特征训练模型

参数迁移

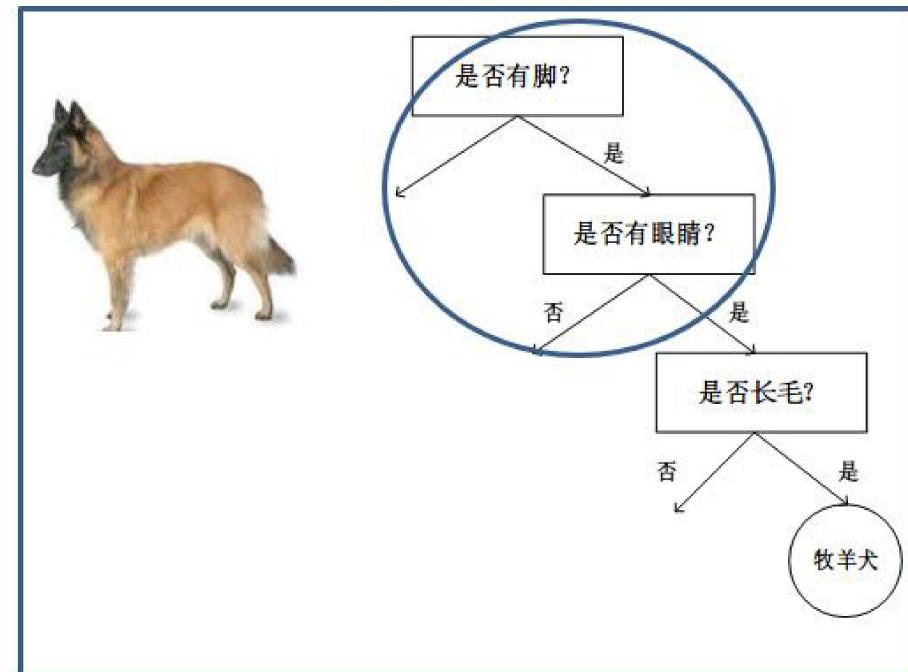


- 将源任务的参数与目标任务共享

源域（图像）



目标域（图像）



参数迁移的步骤



1. 在源任务数据上训练模型
2. 提取源任务模型的部分参数
 - ✓ 决策树节点特征
 - ✓ 神经网络前几层
3. 给定源任务模型的部分参数，基于目标任务数据继续优化模型

关系迁移



- 关系迁移关注源域和目标域的样本之间的关系

师生关系



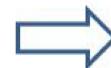
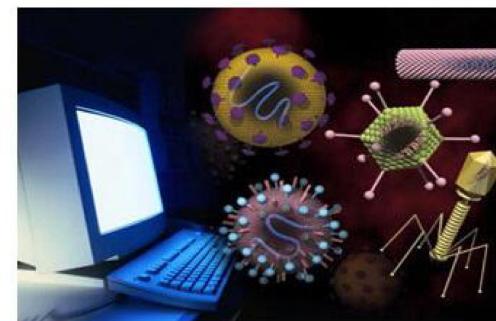
上下级关系



生物病毒



计算机病毒



关系迁移的步骤

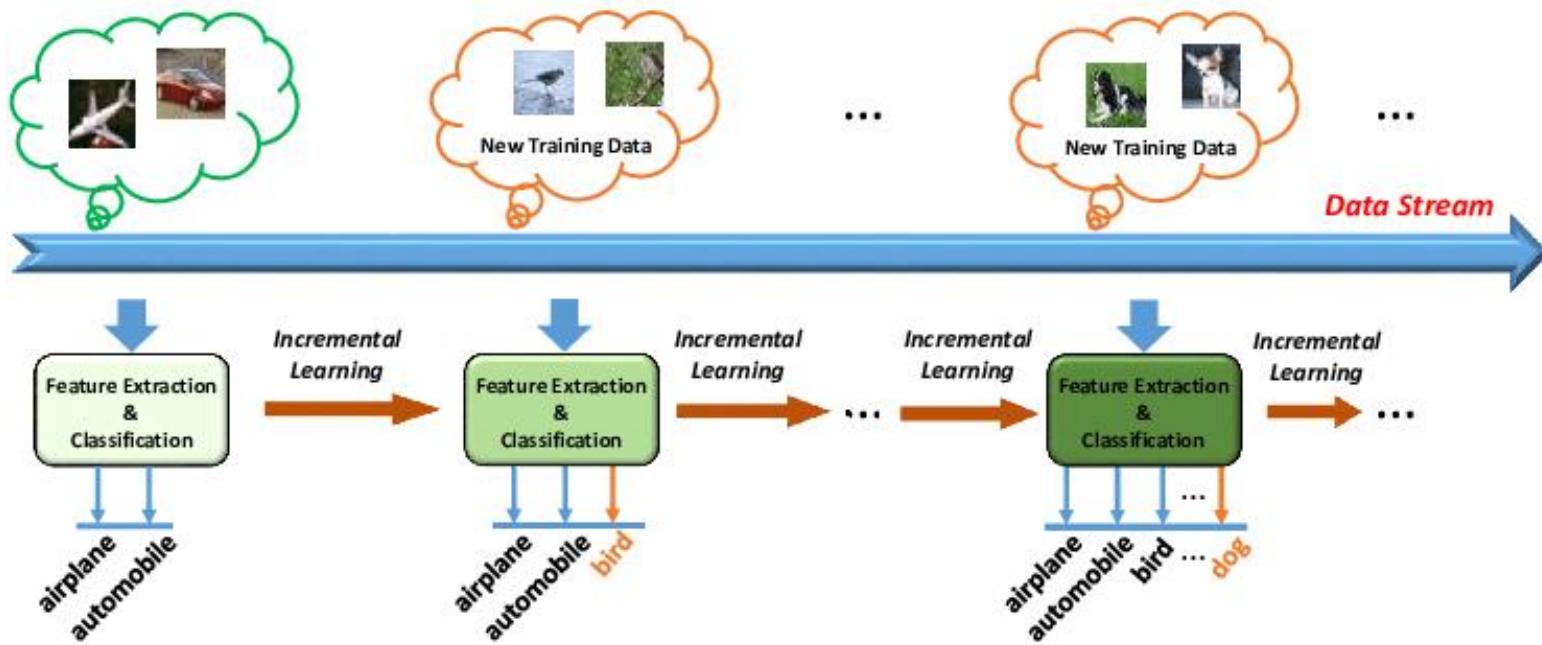


1. 在源任务数据上训练并提取出关系知识
2. 建立源任务与目标任务各关系实体之间的映射，将源任务关系知识映射到目标任务上
3. 给定映射后的源任务关系知识，基于目标任务数据继续优化关系知识

云边协同的增量学习



- 传统的机器学习采用批处理或脱机训练的方式学习，增量学习则让一个系统能不断地从新样本中学习新知识，并能保持大部分以前已经学习到的知识





- 在线学习 (Online learning)

- ✓ 一种可以随着新数据的不断接收，逐步调整模型的学习方式
- ✓ 主要针对动态变化的数据流
- ✓ 例如股票价格预测、在线推荐系统等
- ✓ 主要挑战是概念漂移(concept drift)

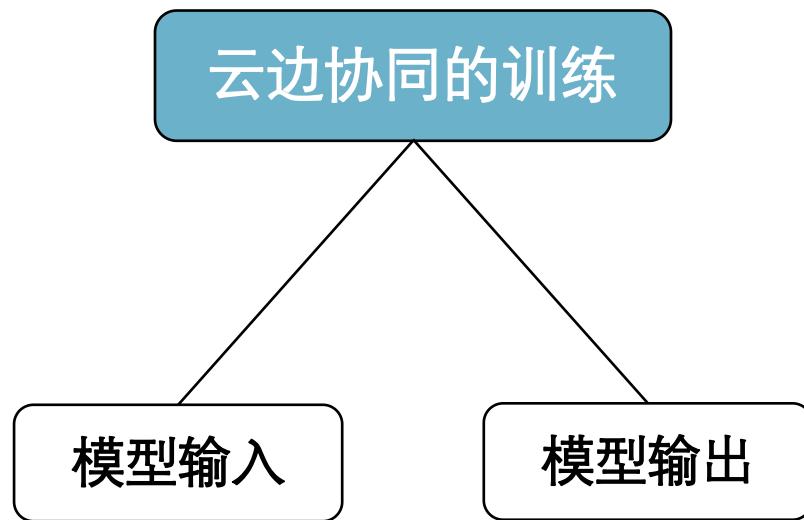
终生学习



- 终生学习 (Lifelong learning)

- ✓ 从一系列的任务中持续学习和适应新知识，而不是只从单一任务中学习
- ✓ 使机器学习模型像人一样能够通过经验积累知识，并在新的任务中利用这些知识
- ✓ 主要挑战是灾难性遗忘(catastrophic forgetting)，也就是在学习新任务时，模型会忘记旧任务的知识

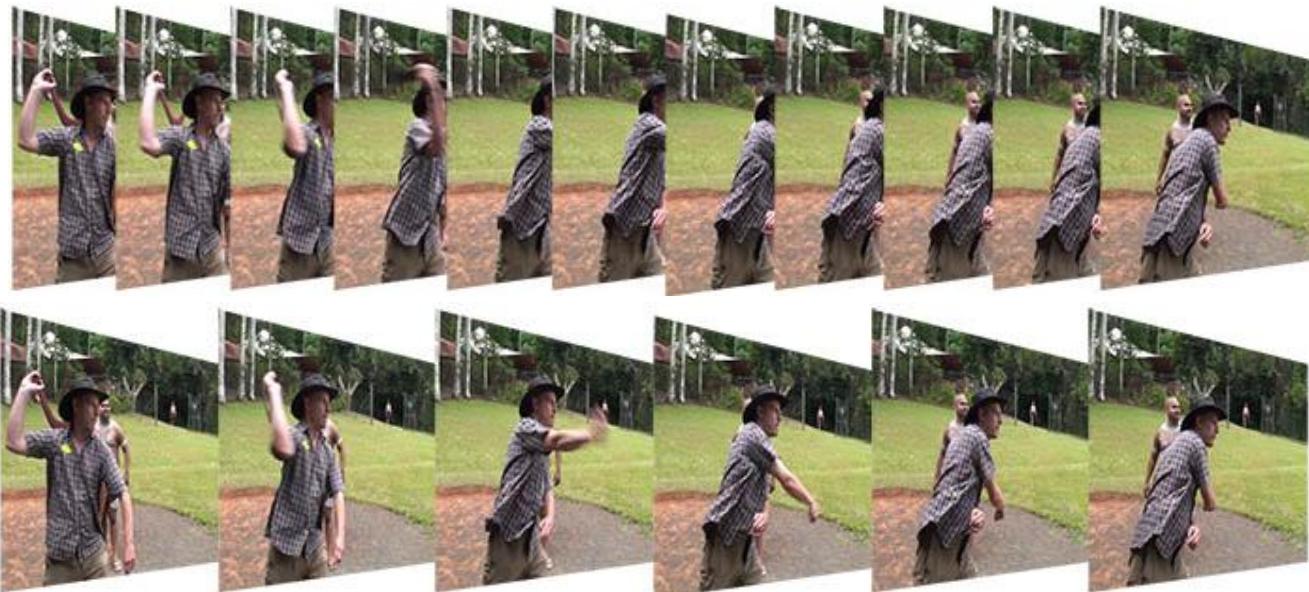
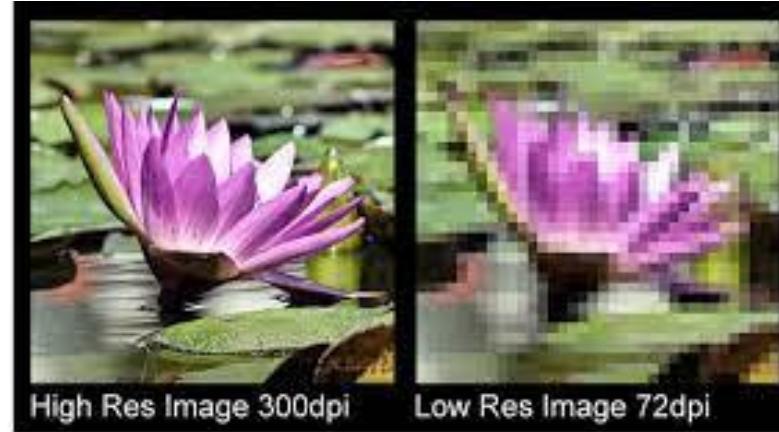
云边协同推理的优化



模型输入优化



- 1. 降低输入数据的配置



模型输入优化



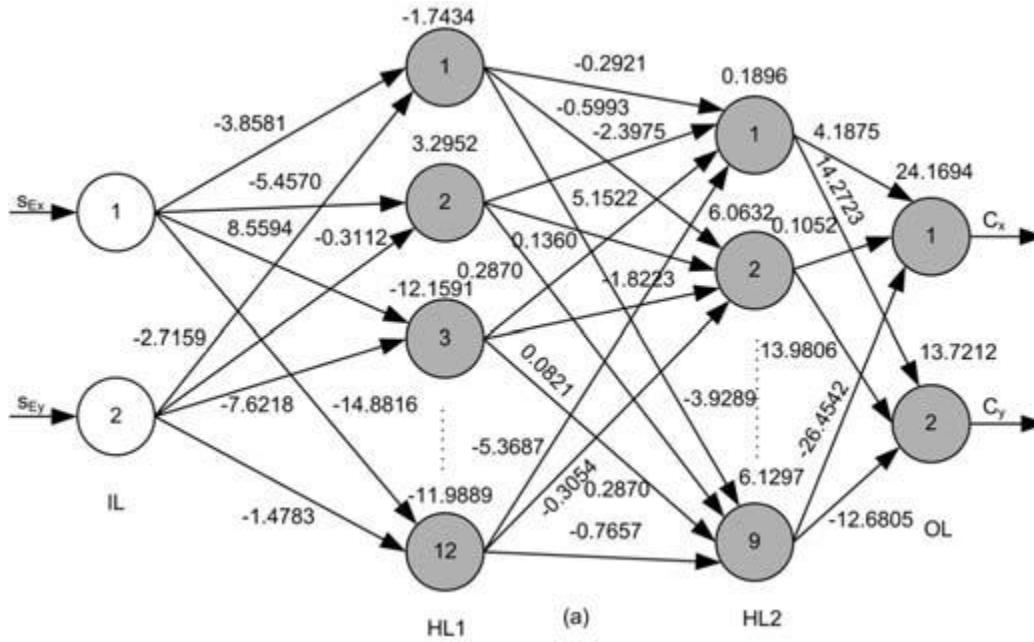
- 2. 缩小搜索空间/锁定感兴趣域



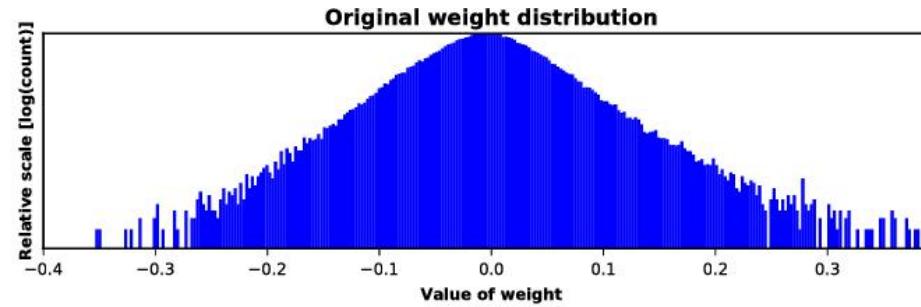
模型结构优化



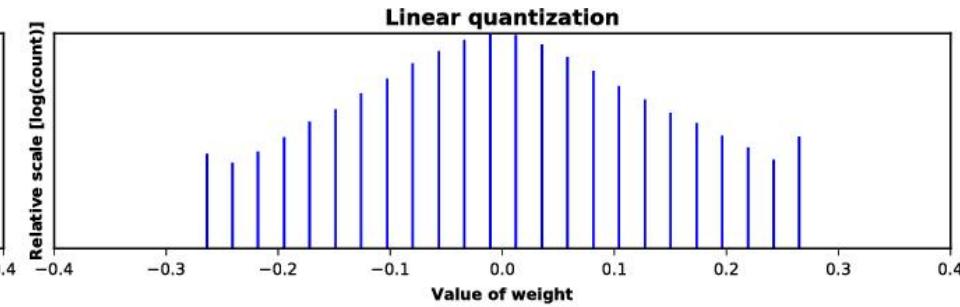
• 1. 权重量化 (weight quantization)



Original weight distribution



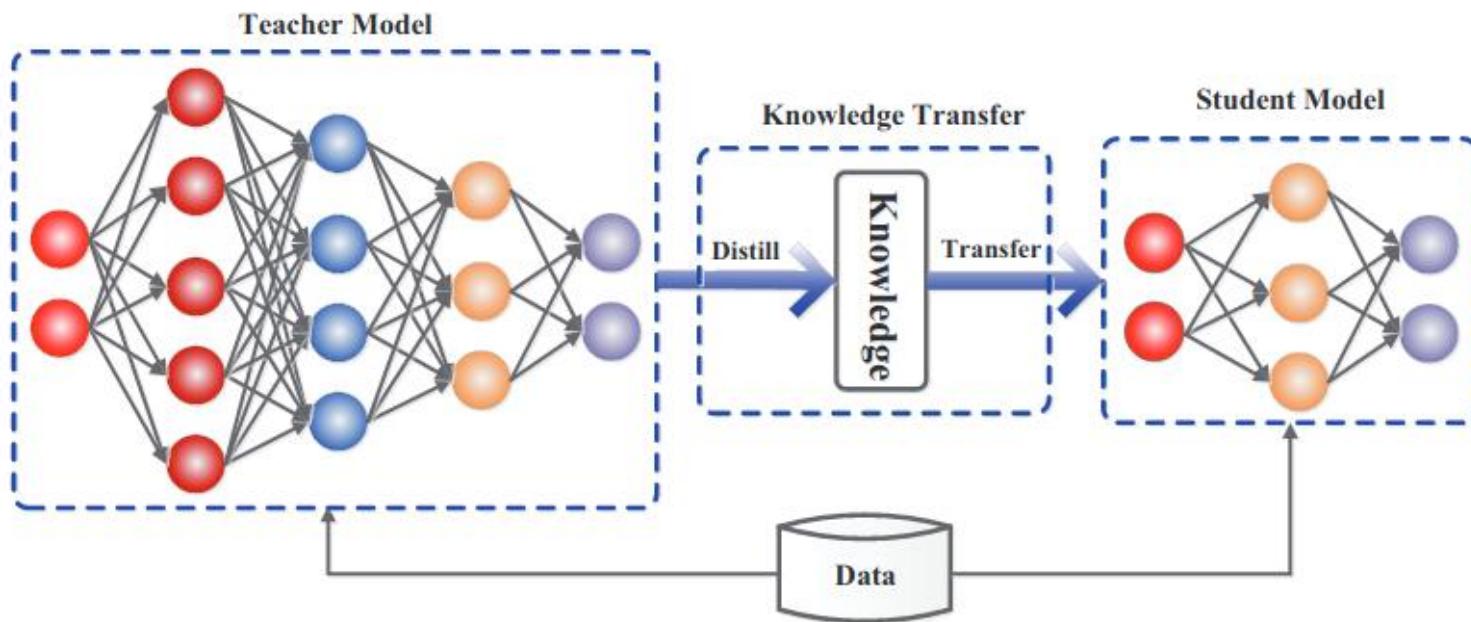
Linear quantization



模型结构优化



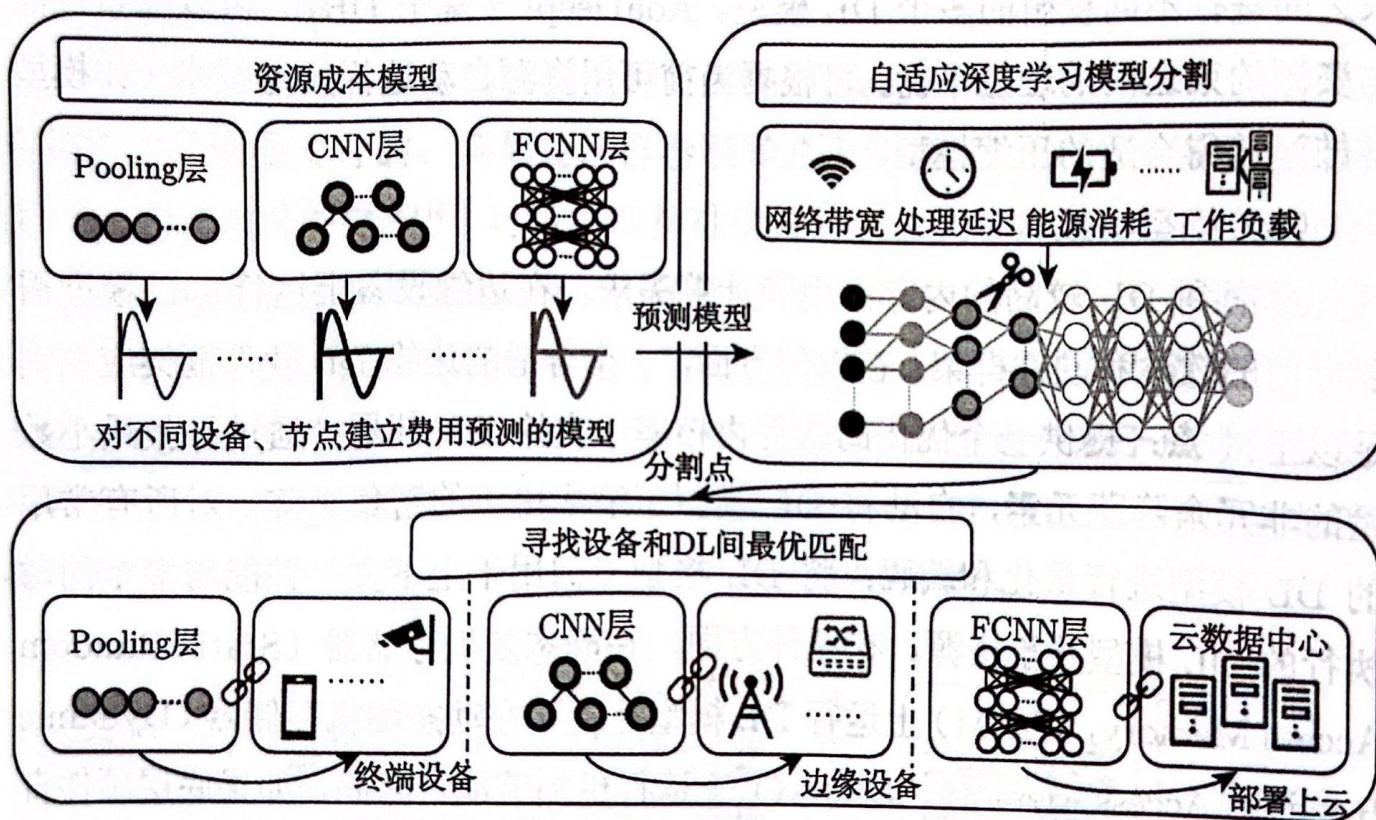
- 2. 知识蒸馏 (Knowledge distillation)



模型结构优化



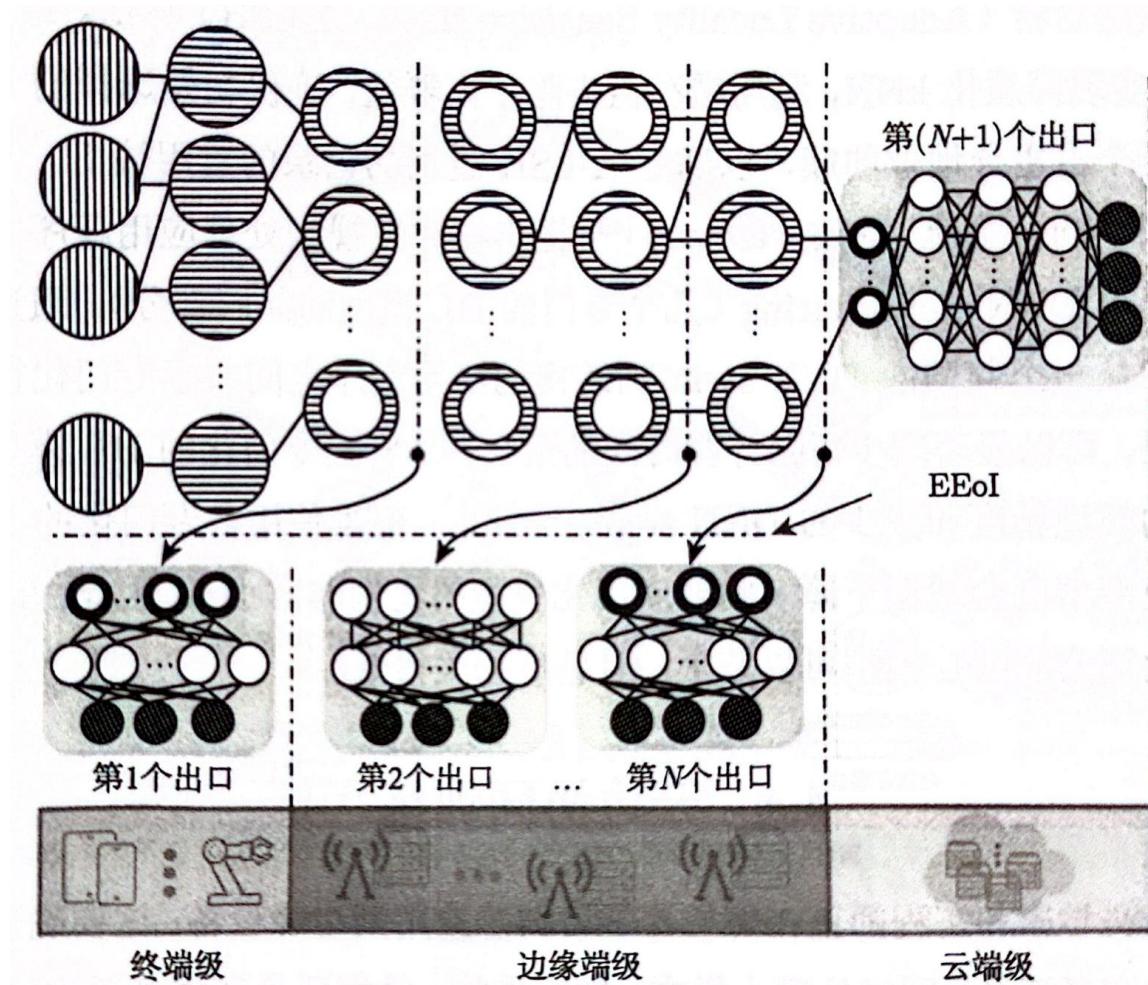
• 3. 网络切割 (Segmentation of DL Models)



模型结构优化



• 4. 早退网络 (Early Exit of Inference)





中山大學 软件工程学院
SUN YAT-SEN UNIVERSITY SCHOOL OF SOFTWARE ENGINEERING

谢谢

陈壮彬

软件工程学院

<https://zbchern.github.io/sse316.html>