

团 体 标 准

T/CESA 1228.2—2022

人工智能 计算设备调度与协同 第 2 部分：分布式计算技术要求

Artificial intelligence — Computation scheduling and cooperation—
Part 2: Distributed computing technical requirements

2022-09-29 发布

2022-09-29 实施



版权保护文件

版权所有归属于该标准的发布机构，除非有其他规定，否则未经许可，此发行物及其章节不得以其他形式或任何手段进行复制、再版或使用，包括电子版，影印件，或发布在互联网及内部网络等。使用许可可于发布机构获取。

目 次

前言.....	III
引言.....	IV
1 范围.....	1
2 规范性引用文件.....	1
3 术语和定义.....	1
4 缩略语.....	2
5 概述.....	2
6 计算设备技术要求.....	3
6.1 云侧设备.....	3
6.2 边缘设备.....	4
6.3 终端设备.....	4
7 分布式协同管理技术要求.....	5
7.1 通用技术要求.....	5
7.2 关键组件技术要求.....	6
7.3 云云协同.....	9
7.4 云边端协同.....	10
7.5 多端协同.....	13
参考文献.....	15

前 言

本文件按照GB/T 1.1-2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

本文件是T/CESA 1228—2022《人工智能 计算设备调度与协同》的第2部分。T/CESA 1228—2022已经发布了以下部分：

——第1部分：虚拟化与调度系统技术规范；

——第2部分：分布式计算技术要求。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由中国电子技术标准化研究院提出。

本文件由中国电子技术标准化研究院、中国电子工业标准化技术协会归口。

本文件起草单位：中国电子技术标准化研究院、华为技术有限公司、上海依图网络科技有限公司、上海天数智芯半导体有限公司、北京航空航天大学、中国医学科学院生物医学工程研究所、上海商汤阡誓科技有限公司、达闼机器人股份有限公司、昆仑芯（北京）科技有限公司、深圳市矽赫科技有限公司、深圳云天励飞技术股份有限公司、上海燧原科技有限公司、南通大学、上海计算机软件技术开发中心。

本文件主要起草人：董建、鲍薇、舒驰、赵春昊、徐洋、李小娟、马骋昊、余雪松、栾钟治、曹晓琦、许源、蒲江波、吴庚、张站朝、李艳华、洪宝璇、程冰、张明洁、曹志超、陈敏刚。

引 言

本文件是T/CESA 1228—2022《人工智能 计算设备调度与协同》的第2部分。T/CESA 1228—2022已经发布了以下部分：

——第1部分：虚拟化与调度系统技术规范。目的在于确立人工智能计算设备虚拟化与调度系统的架构，规定技术要求，描述了对应的测试方法。

——第2部分：分布式计算技术要求。目的在于确立人工智能计算设备分布式计算的参考架构，规定软硬件系统的功能和性能技术要求。

人工智能 计算设备调度与协同 第2部分：分布式计算技术要求

1 范围

本文件确立了人工智能计算设备分布式计算的参考架构，规定了软硬件系统的功能和性能技术要求。

本文件适用于用于分布式人工智能计算系统的设计、开发和测试。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

T/CESA 1228.1-2022 人工智能 计算设备调度与协同 第1部分：虚拟化与调度系统技术规范

3 术语和定义

下列术语和定义适用于本文件。

3.1

分布式机器学习 distributed machine learning

一种机器学习的方法或过程，利用互联网络作为首选通信载体，在不同子系统上完成同一个机器学习任务。

注：互联网络可包含局域互联网及广域互联网。

[来源：ISO/IEC 2382—2015，2178059，有修改]

3.2

联邦机器学习 federated machine learning

一种机器学习方法或过程，能让多个参加者协作构建及使用机器学习模型而不暴露参加者所拥有的原始数据或私有数据。

[来源：IEEE 3652.1，3.1]

3.3

增量学习 incremental learning

一种分为多个阶段的自适应学习方法，其中在前驱阶段学得的知识被转化为适当形式，以便为后继阶段的新知识提供基础。

注：基于机器学习实现的增量学习过程，称为“增量训练”。

[来源：ISO/IEC 2382—2015，2123002，有修改]

3.4

边缘计算 edge computing

一种分布式计算方法或过程，其中处理及存储过程发生在分布式系统的边缘。

注：计算发生靠近边缘的程度按系统的需求定义设定。

[来源：ISO/IEC TR 23188—2020, 3.1.3]

3.5

集合通信 collective communication

一种通信模式，其所传递数据在一组处理器单元（称为节点）上处理。

注：一次通信一般包含有多个接受者和发送者。

3.6

分布式训练 distributed training

一种人工智能学习过程，其训练任务的部分或者全部分配在多个计算节点完成。

3.7

分布式推理 distributed inference

一种人工智能推理过程，其中全部或者部分推理任务分配在多个计算节点完成。

4 缩略语

下列缩略语适用于本文件。

AI 人工智能 (Artificial Intelligence)

CPU 中央处理单元 (Central Processing Unit)

FPGA 现场可编程逻辑门阵列 (Field Programmable Gate Array)

GPU 图形处理单元 (Graphics Processing Unit)

NPU 神经网络处理单元 (Neural-network Processing Unit)

PCIE 外设组件互连 (Peripheral Component Interconnect Express)

RDMA 远程直接内存访问 (Remote Direct Memory Access)

5 概述

基于AI计算设备构建的分布式AI任务的运行环境，包含云侧、边侧和终端侧。环境集成通信、控制、设备接入、管理控制等多种功能，基于虚拟化技术手段，提供资源池化和分配调度能力，降低分布式应用的作业延时，提升作业吞吐率。系统实现提出的主要功能见图1，包含：分布式训练，云侧设备承担云、边缘和端侧模型的训练，并在部署或运行时分发到对应位置；在云侧、边缘、终端侧设备上部署模型，并执行推理任务；使能联邦学习、增量学习等方法。

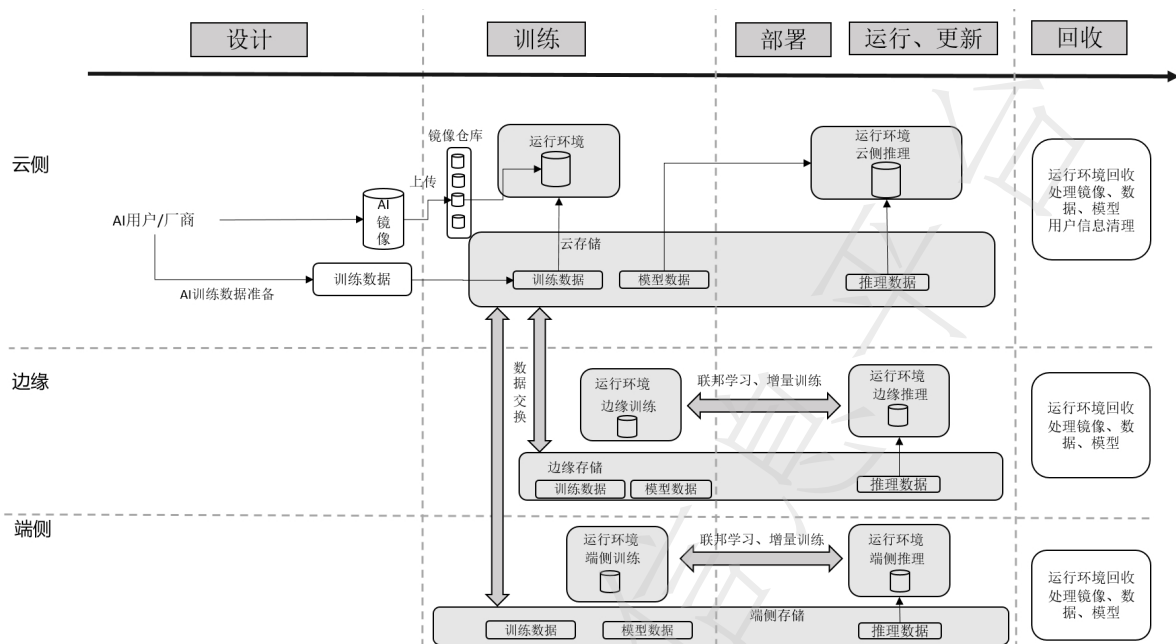


图 1 云边端分布式 AI 任务生命周期

6 计算设备技术要求

6.1 云侧设备

6.1.1 概述

云侧 AI 计算设备及其使用，符合以下要求：

- a) 支持以下应用类型：
 - 1) 云侧训练；
 - 2) 在云侧部署和运行 AI 推理任务，在推理过程中应用计算资源弹性伸缩策略；
 - 3) 增量训练；
 - 4) 联邦学习；
- b) 遵循以下使用原则：
 - 1) 训练过程与数据源分离；
 - 2) 监控 AI 任务全生命周期过程；
 - 3) 多租户共享计算、存储资源；
 - 4) 数据集中处理和分析；
 - 5) 使用虚拟化技术管理计算资源。

6.1.2 技术要求

云侧 AI 计算设备，符合以下要求：

- a) 应具备用于 AI 计算的加速处理器（如 CPU、GPU、NPU、FPGA 等）；
- b) 应支持模型训练所需的软件和模式，包含：
 - 1) 机器学习框架（如 MindSpore、TensorFlow、PyTorch）；
 - 2) 云侧分布式集群训练；

- 3) 模型可视化, 包含训练过程, 数据处理过程等;
- 4) 自动学习;
- c) 应提供与边缘设备、终端设备兼容的互操作协议及接口;
- d) 应支持训练或推理任务镜像管理;
- e) 宜实施针对异构 AI 加速处理器的虚拟化, 形成资源池统一调度 AI 计算。

6.2 边缘设备

6.2.1 概述

边缘侧AI计算设备参与协同计算, 遵循以下使用原则:

- a) 与数据源头保持系统要求的耦合度;
- b) 考虑将小规模的计算与服务部署在边缘侧或终端侧;
- c) 考虑存储、传输、计算和安全任务;
- d) 在本地执行数据处理, 避免数据网络传输中的风险;
- e) 考虑业务场景要求的实时性。

6.2.2 技术要求

边缘侧AI计算设备, 符合以下要求:

- a) AI 边缘计算设备包含但不限于 AI 服务器、AI 加速卡、AI 加速模组;
- b) 应提供与终端设备、云侧设备的协同接口;
- c) 应支持至少 1 中机器学习框架 (如 TensorFlow、PyTorch、MindSpore 等);
- d) 应支持视频分析、文字识别、图像识别、声音处理等应用;
- e) 应具备数据隔离措施, 在受攻击时, 只影响本地设备, 而非其他参与协同的设备;
- f) 宜支持流数据的实时处理;
- g) 宜能在边缘存储实时信息;
- h) 宜能监控边缘节点的运行状态;
- i) 宜支持推理任务或训练任务。

6.3 终端设备

6.3.1 概述

终端侧AI设备参与协同计算, 在数据流转逻辑方面, 一般部署在随数据发生 (如数据源) 的位置附近。

6.3.2 技术要求

终端侧AI计算设备, 符合以下要求:

- a) 应支持基于轻量化机器学习框架的运行 (训练或推理);
- b) 应支持脱离机器学习框架的模型下沉, 在设备上直接运行;
- c) 应支持分布式训练, 包含:
 - 1) 能与其他终端设备、云侧设备、边缘设备协同, 完成训练;
 - 2) 兼容分布式训练框架, 包含但不限于 TensorFlow、PyTorch、MindSpore 等;
 - 3) 支持差分隐私机制;
- d) 宜支持基于全量框架的深度学习推理;
- e) 宜支持数据预处理 (如图像缩放、图像填充、图像旋转、声音去噪等);

- f) 宜使用终端设备电池电量等指标衡量、检测能耗水平；
- g) 宜支持与边缘设备或云侧设备的交互，完成推理；
- h) 宜支持多端协同，包括但不限于：
 - 1) 管理多个终端设备；
 - 2) 多个终端设备之间的数据协同与共享；
 - 3) 多个终端设备之间计算任务的调度与分发。

7 分布式协同管理技术要求

7.1 通用技术要求

分布式协同管理架构见图2，符合以下要求：

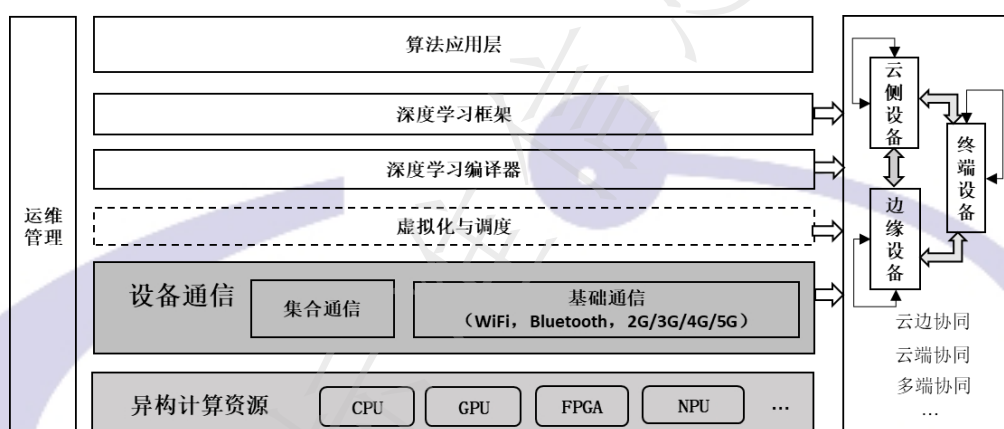


图 2 分布式管理架构图

- a) 应能管理机器学习算法的运行环境配置，包括支持云侧设备、边缘设备和终端设备；
- b) 应支持以下类型的应用在 a) 包含的计算设备上的运行：
 - 1) 云端模型裁剪、压缩和优化；
 - 2) 联邦学习；
 - 3) 协同推理；
 - 4) 基于云侧或边缘侧集群的分布式训练；
 - 5) 边-云协同的增量训练；
- c) 应支持以下支撑组件：
 - 1) 机器学习框架；
 - 2) 深度学习模型编译器；
 - 3) 集合通信库；
 - 4) 虚拟化与调度组件；
 - 5) 云侧、边缘侧支持容器化微服务架构；
- d) 分布式计算框架，应符合以下要求：
 - 1) 具备一致性，在多个计算设备共同训练一个模型时，不同计算设备上模型参数、计算过程应保持一致；
 - 2) 具备容错性，对分布式协同训练的集群，在训练过程中有节点或网络连接出现故障（如节点、交换或路由设备不工作或性能下降）时，能保证原有任务持续进行并完成；

- 3) 具备统一的编程模型，对于开发者和用户屏蔽机器学习框架和不同计算设备的差异；
- 4) 具备分布式存储能力，能根据不同的分布式计算场景（如分布式集群、端云，边云等）设置相应的存储方式；
- 5) 支持计算设备管理，对不同的 AI 加速卡或 AI 加速处理器，根据 AI 任务实施虚拟化和分配，增加计算设备利用率；

注：计算设备的管理和调度，涉及虚拟化和集群调度的部分，参考第1部分。

e) 分布式多终端协同，应符合以下要求：

- 1) 去中心化，即多个终端设备协同完成 AI 任务时，不存在常设的中心节点；
- 2) 实时维护与共享集群中节点状态，以实施调度决策；
- 3) 在调度时考虑网络传输的因素，具备实时性保障机制（如超时机制等）；

注：集群组网使用无线连接（如WIFI或蓝牙）时，会出现传输速率有限且不稳定的情况。

- 4) 在调度时考虑设备实际能力，并具备适应性协同策略（如按计算能力调度）；

注：多终端设备结构不同，能力不同，优势的外设和算力资源往往不在同一节点上。

- 5) 调度时考虑设备使用状态因素，计算调度过程不影响其他设备或应用的运行。如：正常的视频播放任务不应被打断，或导致播放卡顿等；
- 6) 考虑非常插电设备接受调度后计算能耗。

7.2 关键组件技术要求

7.2.1 机器学习框架

用于分布式机器学习任务的机器学习框架，应符合以下要求：

- a) 应能在云侧、边缘侧、终端侧等设备上运行；
- b) 提供模型部署功能，应符合以下要求，包括但不限于：
 - 1) 应能以云服务的形式在云侧部署，提供推理任务执行环境；
 - 2) 应支持能被裁剪的轻量化框架，在云侧训练的模型，能部署在端侧和边缘侧；
 - 3) 应支持脱离框架或只依赖轻量化框架的模型的运行；
 - 4) 应支持在内存受限的设备或环境中的运行；
 - 5) 宜支持模型量化，并在用户允许的误差范围内，使模型能从云侧迁移至端侧或边缘侧；
- c) 应支持云侧、端侧、边缘侧的计算协同模式，包括但不限于：
 - 1) 云侧训练一端侧推理；
 - 2) 云侧训练一端侧增量训练一端侧推理；
 - 3) 云侧一端侧的联邦学习；
 - 4) 云侧训练一边缘侧推理；
 - 5) 云侧训练一边缘侧增量训练一边缘侧推理；
 - 6) 云侧一边缘侧联邦学习；
 - 7) 云侧训练一边缘侧增量训练一端侧推理；
 - 8) 云侧一边缘侧一端侧联合推理；
- d) 宜支持跨云侧、边侧、终端侧的，一致的计算图中间表示，保证一次训练后的模型在端、边、云三种设备上的无（格式）转换部署。

7.2.2 虚拟化与调度

异构AI加速处理器或带有异构AI加速处理器的加速卡的虚拟化与调度组件，应符合以下技术要求：

- a) 应符合本文件第 1 部分的要求；

- b) 资源调度系统应能调用 AI 加速卡或 AI 加速处理器及其虚拟化实例，包含：
 - 1) 仅能用于训练任务的加速卡；
 - 2) 仅能用于推理任务的加速卡；
 - 3) 能用于训练及推理任务的加速卡；
- c) 分布管理中的异构资源调度系统，应支持以下调度模式：
 - 1) 云侧的训练卡调度，实施云侧分布式训练；
 - 2) 边缘侧的训练卡调度，支持边缘侧增量训练；
 - 3) 云侧或边缘侧的推理加速卡调度，支持大模型推理，克服单卡运行时内存限制；
 - 4) 云侧和边缘侧的训练加速卡的调度，能完成云侧训练—边缘侧增量训练的任务；
 - 5) 云侧和边缘推理加速卡的调度，能完成云侧训练—边缘推理的任务；
 - 6) 云侧和终端侧加速卡的调度，能完成云侧—端侧的联邦学习任务；
 - 7) 云侧训练加速卡和端侧推理加速卡的调度，能完成云侧训练—端侧推理任务；
 - 8) 边缘侧训练加速卡和端侧推理加速卡的调度，能完成边缘侧增量训练—端侧推理的任务；
 - 9) 边缘侧训练加速卡的调度，能完成边缘侧的分布式训练任务；
- d) 宜支持基于以下类型处理器的 AI 加速卡的调度，包括：
 - 1) CPU（边缘，终端）；
 - 2) GPU；
 - 3) FPGA；
 - 4) NPU 或其他定制化处理器。

7.2.3 深度学习编译器

深度学习编译器应具有兼容性，能将不同框架下模型或计算图编译到指定的AI加速卡上运行，实现模型跨平台移植。

7.2.4 安全组件

7.2.4.1 安全原则

分布式AI计算协同管理框架的使用，应按图1规定的生命周期和流程，识别以下风险，包括但不限于：

- a) 易受窃取等攻击的资产及相关攻击模式，包含：
 - 1) 训练脚本或代码文件；
 - 2) 训练数据集文件；
 - 3) 模型文件（包含预训练模型文件、训练结果文件和在计算节点中的模型文件）；
- b) 模型的文件，在云、边、端侧的完整性被破坏（如被修改）；
 - a) 模型及相关文件（含训练脚本、预训练模型、训练数据集、模型参数文件）的机密性被破坏；
 - b) 被其他应用恶意攻击，导致输出结果错误，或无法正常输出结果；
- c) 推理输入数据不被截获或修改；
- d) 云侧、边缘侧的容器构建、部署、运行安全威胁；
- e) 边缘侧接口的非授权接入和访问；
- f) 在无中心（或无云侧）的部署架构中忽略使用热备、冗余机制，造成数据丢失或业务中断。

7.2.4.2 安全要求

7.2.4.2.1 训练阶段，应实现以下安全防护机制：

- a) 云侧：
 - 1) 对训练文件的签名，能加密 AI 模型和训练脚本并上传至云侧；
 - 2) 对上传数据的加密；
 - 3) 运行训练脚本前，对训练镜像文件的身份校验；
 - 4) 对预训练模型的解密；
 - 5) 对训练数据的解密；
- b) 边缘侧：
 - 1) 对云侧下发镜像的身份校验和模型的加解密；
 - 2) 对联邦学习梯度信息的加解密；
 - 3) 对增量训练数据的签名和加密；
 - 4) 对端侧设备上传的数据的签名和加密；
- c) 终端侧：
 - 1) 对云侧、边缘侧下发模型的身份校验和加解密；
 - 2) 对端侧联邦学习梯度数据的加密。

7.2.4.2.2 推理阶段，应实现以下防护机制：

- a) 云侧：
 - 1) 在云侧环境中配置含有 AI 模型的容器运行的最小访问权限；
 - 2) 对 AI 模型身份的合法解密；
 - 3) 在加载镜像到云侧运行环境时，检查身份合法性；
- b) 边缘侧：
 - 1) 密钥管理，对云侧下发的镜像或模型实施身份校验；
 - 2) 对推理数据的加解密；
 - 3) 入侵检测；
 - 4) 边缘用户身份校验，应用授权访问；
- c) 端侧：
 - 1) 密钥管理，对云侧下发到端侧模型实施身份合法性检查；
 - 2) 推理数据在本地的加解密；
 - 3) 在模型上传至云侧或边缘侧进行推理前，加密模型；
 - 4) 终端设备使用虚拟机或容器时，在多终端设备间实现基于虚拟机或容器的安全隔离机制。

7.2.5 集合通信

7.2.5.1 提供统一的通信协议、数据结构定义和元语，机器学习框架利用集合通信库，能操作不同的计算设备。

7.2.5.2 实现 AI 加速卡间通信，包含云侧内部分布式训练和端边云协同训练所需通信。

7.2.5.3 能在下列集合通信拓扑结构上工作，包含：

- a) 基础类：
 - 1) 全网状结构(Full Mesh)；
 - 2) 星形拓扑结构 (Star)；
 - 3) 环形拓扑结构 (Ring)；
 - 4) 树拓扑结构 (Tree)；

- b) 扩展类：
- 1) 3D Torus 拓扑结构；
 - 2) Dragonfly 拓扑结构；
 - 3) 由以上两种或以上网络拓扑结构混合而成的混合拓扑结构。

7.2.5.4 支持集合通信步调，实现同步功能，满足协同节点异步使用要求，并具备超时管控机制，避免通信无法完成的问题。

7.2.5.5 支持通信协议，包含但不限于：PCIE，RDMA 和 TCP/IP 等。

7.2.5.6 支持集合通信能力表征指标，包含：

- a) 集合通信耗时均值；
- b) 集合通信耗时方差。

7.2.5.7 支持集合通信能力表征指标的测量方法，包含：

- a) 单 AI 加速卡或者计算设备上集合通信耗时测量；
- b) 全部 AI 加速卡或者计算设备上的集合通信耗时的均值和方差的计算；
- c) 时间的统计，符合以下要求：
 - 1) 单 AI 加速卡或计算设备的时间测量的起始时刻为：集合通信的使用者启动集合通信的时刻；
 - 2) 时间测量的结束时刻为：集合通信的使用者感知到集合通信完成的时刻；
 - 3) 集合通信的启动和结束如果存在固定开销，应被计入到通信耗时里。

7.2.5.8 提供面向机器学习框架的统一接口，屏蔽异构硬件的差异。

7.3 云云协同

云云协同的框架见图3，协同过程符合以下要求：

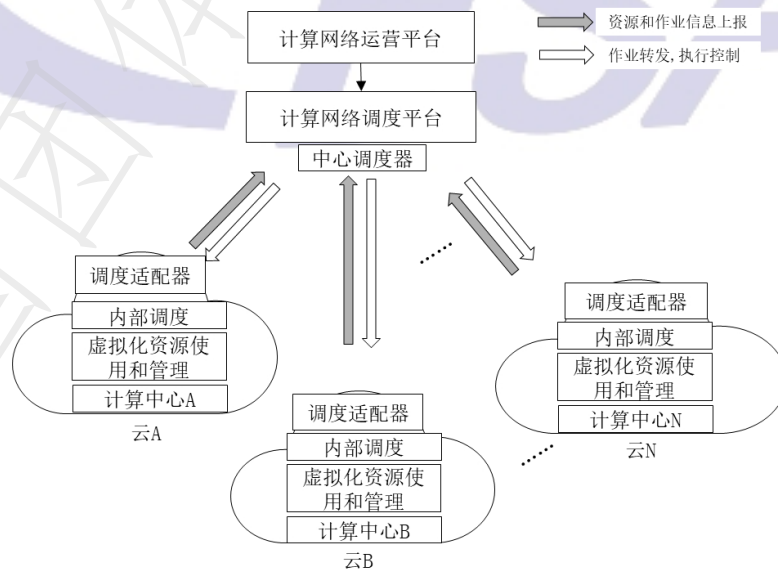


图 3 云云协同框架

- a) 云云协同框架，应完成中心调度模式或无中心调度模式（如基于云联邦的调度）；

- b) 使用中心调度模式时，中心调度器：
 - 1) 应能收集和监视全网计算资源的状态；
 - 2) 应能跨云（或计算中心）调度作业，转发和传输作业运行数据；
 - 3) 应实现负载感知调度、价格感知调度、网络感知调度或数据感知调度等策略；
- c) 使用中心调度模式时，调度适配器符合以下要求：
 - 1) 应适配不同云中，异构调度器的实现；
 - 2) 应能收集调度器资源和作业执行状态信息，并上报至中心调度器；
 - 3) 宜能向用户提供统一视图，使用户能使用、全网计算资源执行计算任务；
- d) 宜实现基于统一账户的认证、确权；
- e) 宜实现统一计算消耗计量和计费。

7.4 云边端协同

7.4.1 云边端协同管理框架

云边端协同的框架见图4，包含边云管理系统，边云通道和边端通道。边缘节点的管理，应实现以下功能：

- a) 应用管理，如双机备份、容器故障检测、模型管理、AI 加速卡插件管理；
- b) 边缘中间件（如消息总线、规则引擎等）管理；
- c) 端设备服务；
- d) 安全可信模块；
- e) 设备管理。

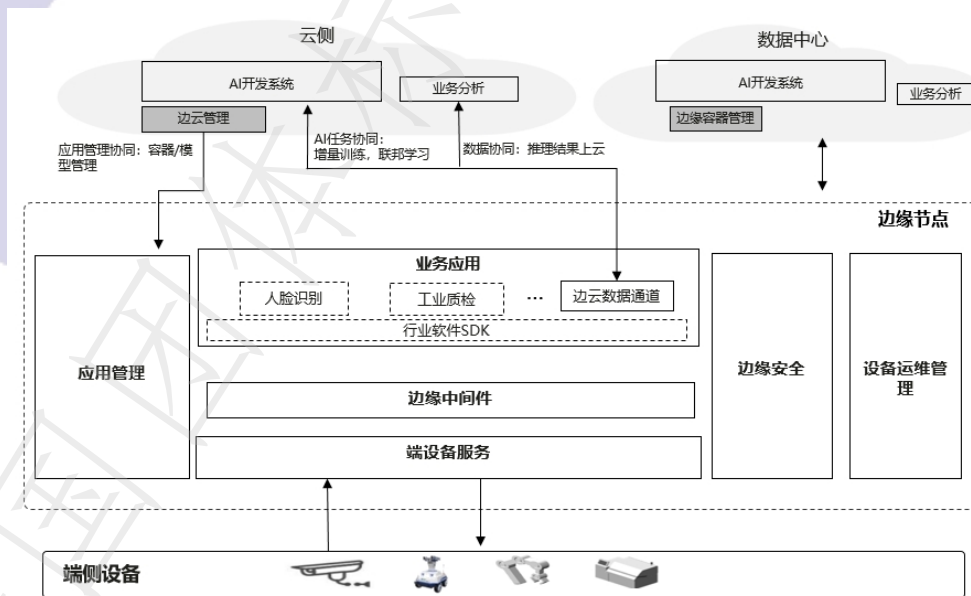


图 4 云边端协同框架

7.4.2 云边模型部署

云边模型部署流程见图 5，符合以下要求：

- a) 将云侧模型部署到边缘侧时，边云管理系统应符合以下要求：
 - 1) 纳管边缘节点并关联边缘节点的设备；
 - 2) 云侧能将 AI 应用或模型以容器镜像的形式下发，部署到边缘节点；

- b) 边云管理系统应具备以下功能：
- 1) 将云侧服务能力部署到边缘，可支持视频识别、文字识别、图像识别等大数据任务的处理；
 - 2) 能将模型以容器和函数两种方式运行，满足用户轻量化部署的要求；
 - 3) 支持边缘节点以注册的方式，接入云侧；
 - 4) 支持云侧的业务以容器的形式，部署到边缘侧；
 - 5) 支持对边缘侧 AI 加速卡利用率的监控；
 - 6) 能管理的边缘节点数量宜不少于 128。

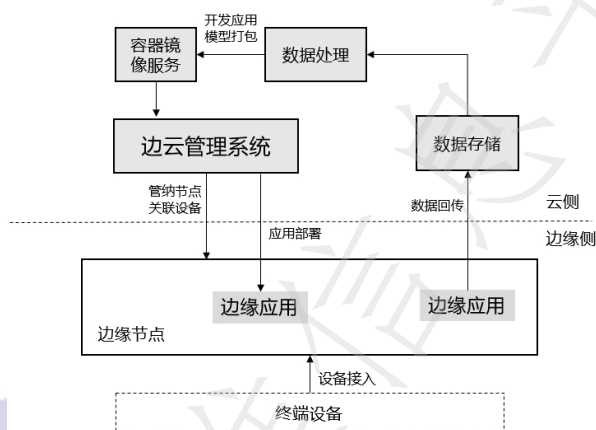


图 5 边云模型部署流程

7.4.3 云边 AI 任务协同

云边协同框架见图 6，边云管理系统，符合以下要求：

- a) 应支持增量训练；
- b) 应实现并提供以下训练样本筛选功能：
 - 1) 根据推理结果的置信度筛选增量训练的数据；
 - 2) 数据去重；
 - 3) 数据聚合；
- c) 应实现梯度数据和增量样本数据的压缩和传输；
- d) 应实现数据加密，满足 7.2.4 中边缘侧安全要求；
- e) 应提供边缘存储功能，能按 7.1 d) 4) 的要求存储终端侧的推理数据和云侧的模型；
- f) 宜支持基于边缘节点设备和云侧设备的联邦学习，集合通信满足 7.2.5 的要求，机器学习框架满足 7.2.1 的要求。

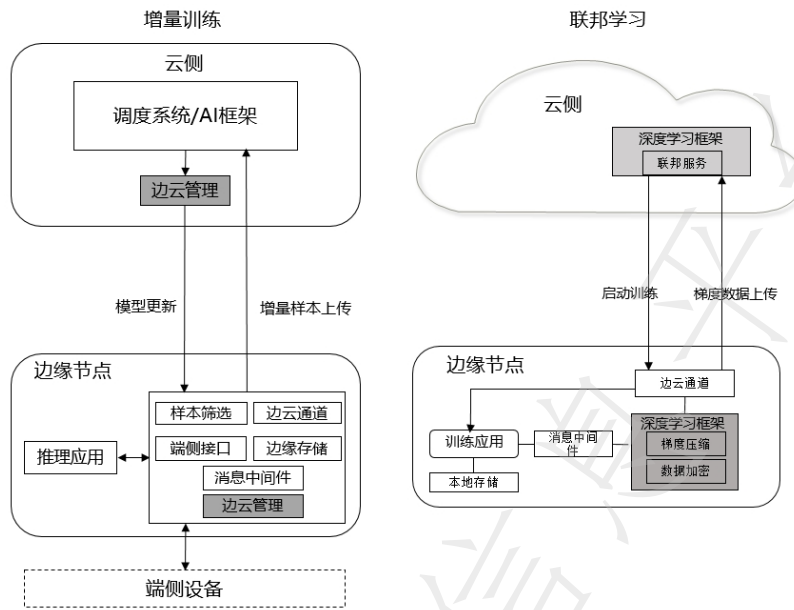


图 6 边云 AI 任务协同框架

7.4.4 端边云数据协同

边云数据协同框架见图 7。

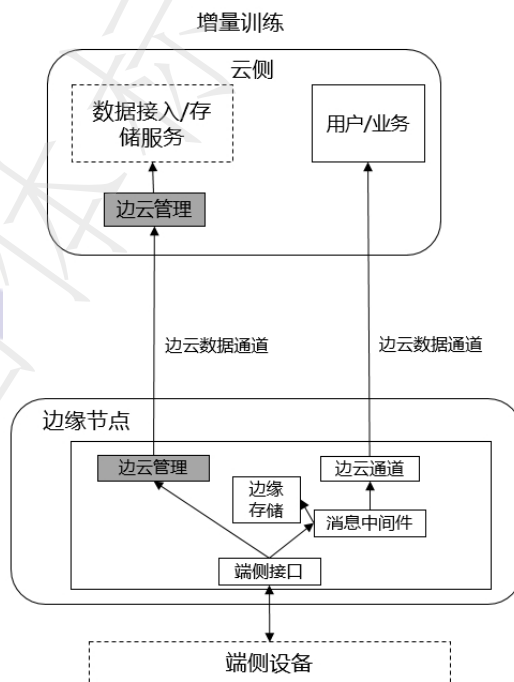


图 7 边云数据协同框架

7.4.4.1 终端侧设备与边缘侧的协同（见图 7），符合以下要求：

- a) 应支持端侧设备多协议接入；
- b) 宜支持下列终端设备接入，包含：

- 1) 摄像机;
- 2) 智能机器人;
- 3) 物联网传感器;
- 4) 激光雷达、毫米波雷达;
- 5) 信号机;
- 6) 气象感知;
- 7) 道路单元。

7.4.4.2 边缘侧设备提供面向 AI 应用的多媒体服务（见图 7），符合以下要求：

- a) 应能转发多媒体数据（到云侧、边缘侧或其它终端）；
- b) 应支持多媒体数据的回放；
- c) 宜支持多媒体数据的转码、点播、直播、录像和检索；
- d) 宜支持地图数据的共享分发。

7.4.4.3 边缘侧的数据存储，符合以下要求：

- a) 应能将推理结果的缓存在本地；
- b) 能实现以下类型数据的存储：
 - 1) 媒体数据；
 - 2) 关系型数据库；
 - 3) 时序数据库。

7.4.4.4 应实现或能使用边缘多容器通信中间件。

7.4.4.5 应实现实时数据传递（如监视器或摄像机采集的视频数据）。

7.5 多端协同

7.5.1 多端协同管理

多端协同的框架（见图8）应实现以下协同方式：

- a) 外设协同；
- b) 数据协同；
- c) 计算协同，包含：
 - 1) 统计计算；
 - 2) 推理计算；
 - 3) 训练。

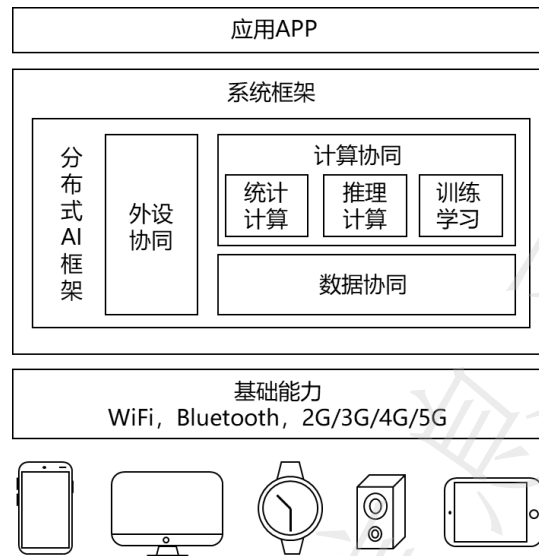


图 8 多端协同管理框架

7.5.2 多端数据协同

多端协同系统，符合以下数据协同要求：

- a) 应能在设备间传输数据；
- b) 应支持多端、多模数据融合；
- c) 应能在设备间共享读写数据；
- d) 离线时，宜实现数据在端侧设备的自动保存，在网络恢复后，自动同步至云或边缘侧。

7.5.3 多端外设协同

多端协同系统，应支持跨外设的输入、输出（如摄像机捕捉的影响，直接在大屏上显示，并不需要在主控端如智能通信终端上显示）。

7.5.4 多端计算协同

多端协同系统，符合以下计算协同要求：

- a) 应拆分任务，并分配到多个设备上，分别完成后整合、返回结果；
- b) 应支持以下类型推理计算任务，包含但不限于：
 - 1) 发起、拆分、分发和管理推理任务；
 - 2) 小任务从低端到高端的转发；
 - 3) 小任务计算靠近转发；
 - 4) 端侧节点设备能按自身能力动态加载对应模型；
 - 5) 端侧节点设备能按节点状态、网络状态、计算功耗等因素判断并选择协同模式；
 - 6) 端侧节点实现容错机制，包含但不限于：异常检测、任务重试、失败重发；
- c) 应支持以下类型训练任务：
 - 1) 模型全量训练；
 - 2) 深度学习模型微调。

参 考 文 献

- [1]ISO/IEC 2382:2015 信息技术 词汇 (Information technology — Vocabulary)
- [2]ISO/IEC TR 23188:2020 信息技术 云计算 边缘计算概览 (Information technology — Cloud computing — Edge computing landscape)
- [3]IEEE 3652.1 联邦机器学习体系结构框架和应用指南 (Guide for Architectural Framework and Application of Federated Machine Learning)
-

