TC609

全国数据标准化技术委员会技术文件

TC609-6-2025-XX

全国一体化算力网 智算中心算力池化技术要求

National integrated computing network——Technical specification for poolling of computing resource in artificial intelligence data center

（征求意见稿）

2025 年 8 月 1 日

202X-XX-XX 发布　　202X-XX-XX 实施

全国数据标准化技术委员会　发布

目　　次

前　言

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。

本文件由全国数据标准化技术委员会（SAC/TC609）提出并归口。

本文件起草单位：

全国一体化算力网 智算中心算力池化技术要求

1 范围

本文件规范了全国一体化算力网资源层中智算中心内的算力池化技术要求，包括算力资源抽象、资源池化管理、任务式资源申请与调度、业务编排等功能及接口要求。

本文件适用于全国一体化算力网监测调度平台中算力网资源层中智算中心异构算力资源相关池化能力的建设与实现。

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。

3 术语和定义

下列术语和定义适用于本文件。

3.1

算力 computing power；computility

图形处理器（GPU）、中央处理器（CPU）等设备执行计算密集型任务的计算能力。

3.2

算力网 computing network

支撑数字经济高质量发展的关键基础设施，可通过网络连接多源异构、海量泛在算力，实现资源高效调度、设施绿色低碳、算力灵活供给、服务智能随需。

3.3

算力池化 pooling of computing resource

通过算力虚拟化和应用容器化等关键技术，对各类异构、异地算力资源与设备进行统一注册和管理，实现对大规模集群内异构计算资源的按需申请与使用。

3.4

智算中心算力池化 pooling of computing resource in artificial intelligence data center

通过算力资源抽象、资源池化管理等关键技术，实现对智算中心内部异构算力资源的差异化屏蔽，面向上层调度、管理系统及应用程序呈现标准化算力设备及统一调用方法，支持计算任务在智算中心内任意算力资源上的灵活运行及迁移。

4 智算中心算力池化功能架构

针对算力网资源层中的各智算中心，在接入全国一体化算力网前，需对其提供的异构智能算力进行差异化屏蔽和池化，以便实现上层调度及管理系统对底层异构硬件的一致化灵活调度，以及应用程序在底层异构智算硬件上的灵活运行与迁移。

全国一体化算力网监测调度平台包括算力网运营层、算力网调度层、算力网资源层以及算力网监测层。其中算力网资源层通过资源并网将异属异构异地算力资源接入算力网，并依托于算力网服务用户业务需求。算力资源并网是通过网络连接实现算力资源的可达、可用，并通过API接口实现算力资源的管理、调度与计量。

结合智算中心算力池化定义及监测调度平台总体架构（见图1），智算中心算力池化位于算力网资源层中，基于多样化异构算力资源进行算力抽象及系统级资源池化管理，面向资源封装、调度及管理、应用等屏蔽异构算力差异，形成统一虚拟化算力及操作方法。此外，智算中心算力池化提供任务式资源申请与调度，支持上层调度及应用按计算任务进行资源调度及使用。智算中心算力池化相关模块包括：

a) 算力资源抽象：对异构算力资源进行统一抽象，形成包含设备、显存、计算在内的统一算力资源结构，面向上层调度及业务实现对异构算力资源硬件结构差异的屏蔽，并提供异构算力资源的标准化操作接口；

a) 资源池化管理：根据算力资源抽象对异构算力资源统一管理，面向上层调度及业务实现算力资源架构无感的池化能力，并提供标准化资源申请接口；

b) 任务式资源申请与调度：负责提供计算任务粒度的资源申请与调度，支持将智算应用中的多个计算任务拆分，按需为每个计算任务申请异构算力资源，并将计算任务调度至申请的算力资源实施计算；

c) 业务编排：根据应用的算力资源需求分配并编排资源，拉起实例。

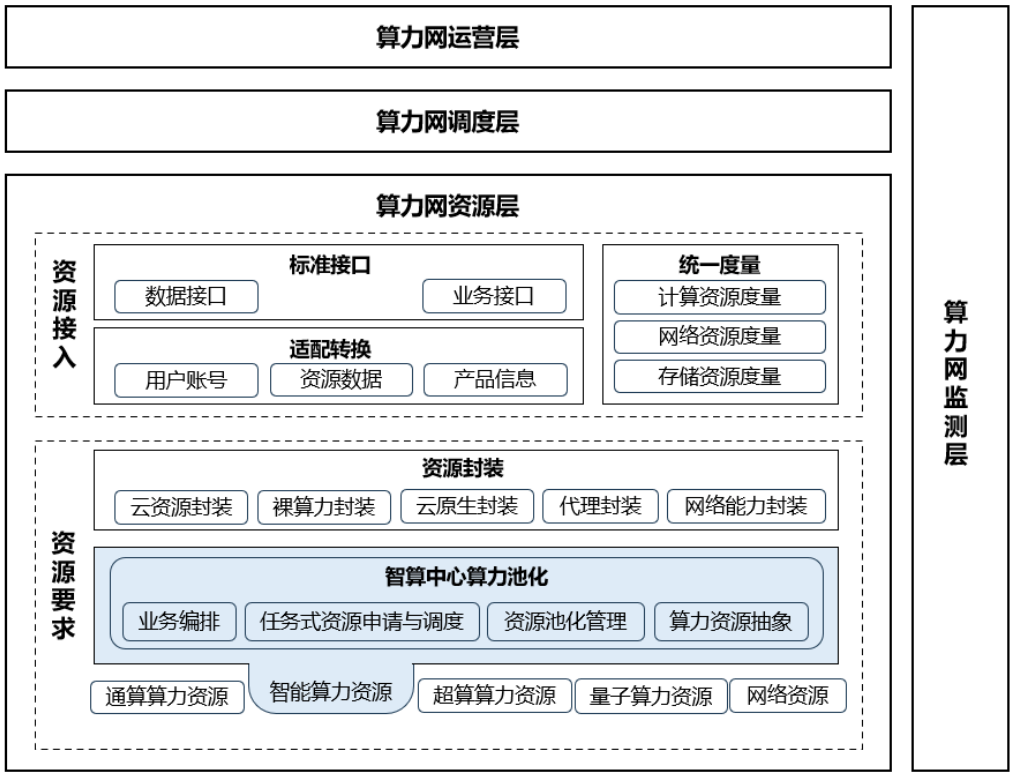

图1　智算中心算力池化功能架构

5　智算中心算力池化功能要求

5.1　算力资源抽象

5.1.1　设备抽象

设备抽象旨在定义统一的算力设备结构及信息，以面向多样异构设备形成统一的设备形态，具体要求如下：

a) 每台算力抽象设备应具备一个主机端和一个或多个设备端。其中主机端应至少具备 CPU 及其内存，设备端应至少具备一张加速卡，每张加速卡具备计算内核及显存。

5.1.2 计算抽象

计算抽象旨在对异构算力资源加速卡中的计算内核进行统一抽象，具体要求如下：

a) 计算内核抽象应具备四级粒度不同的计算单元，具体包括标量计算单元、向量计算单元、张量计算单元、函数计算单元；
b) 标量计算单元为计算内核抽象中的一级计算单元，应支持一维标量数据的 add、mul、div、sin、cos、sqrt、log 等基础运算；
c) 向量计算单元为计算内核抽象中的二级计算单元，应支持二维向量数据的 add、mul、div、sin、cos、sqrt、log、max、min 等基础运算；
d) 张量计算单元为计算内核抽象中的三级计算单元，应支持多维张量数据的 add、mul、div、sin、cos、sqrt、log、max、min 等基础运算；
e) 函数计算单元为计算内核抽象中的四级计算单元，应支持多维张量数据的 add、mul、div、sin、cos、sqrt、log、max、min 等基础运算，以及卷积等复杂函数运算；
f) 计算内核抽象应支持对 INT4、INT8、BF16、FP16、FP32 等精度的数据计算；
g) 计算内核抽象应支持与异构算力资源加速卡中的至少一种计算内核进行分级映射。

5.1.3 内存抽象

内存抽象旨在对存储、主机内存及异构算力资源加速卡中的显存进行统一抽象，具体要求如下：

a) 内存抽象应具备五种存储介质抽象，具体包括存储系统、主机端内存、设备端全局内存、设备端本地内存、设备端私有内存，其中存储系统可选；
b) 存储系统为设备之外的独立存储系统，应支持缓存数据、与主机端内存数据交换；
c) 主机端内存应支持缓存数据、从存储中读写数据、与设备端内存数据交换；
d) 设备端全局内存应支持数据缓存、与主机端内存数据交换、全部计算内核数据读写访问；
e) 设备端本地内存应支持数据缓存、一组计算内核共享数据读写访问；
f) 设备端私有内存应支持数据缓存、一个计算内核数据读写访问。

5.2 资源池化管理

资源池化管理支持对智算中心异构算力资源进行统一管理，面向上层调度及业务实现算力资源架构无感的池化能力，具体要求如下：

a) 资源纳管：应支持纳管异构算力资源，提供资源注册、发现、管理等功能，并支持获取设备及异构算力基本信息，包括但不限于设备型号、设备供应商、设备状态、设备物理加速卡类型、设备物理加速卡数量、设备物理加速卡标识、设备物理加速卡状态、多设备拓扑等；
b) 异构调度：应支持以异构算力统一抽象中的资源粒度进行异构资源调度，最小调度单位为一个 XPU，其中 XPU 是一个抽象的加速卡，具备特定规格的抽象计算内核、抽象内存；
c) 资源映射：应支持 XPU 与实际物理加速卡资源量的映射，支持根据物理计算内核与抽象计算内核、物理内存与抽象内存之间的比例进行映射，一般一个物理加速卡可映射为 1 个或多个 XPU，当物理加速卡资源不足够映射为整数个 XPU 时，向下取整；资源映射时可根据计算单元、内存、算力等多维度映射；

d) 服务质量：需支持自动 QoS 管理，按分配的 XPU 资源数量限制上层应用可以使用的资源而避免争抢；
e) 弹性扩缩：需支持 XPU 算力及显存根据负载自动在线 Scale Up/Down 而业务不感知；
f) 分布式调用：需支持在上层应用无感知的情况下调用 XPU 资源；
g) 迁移能力：需支持 XPU 上计算任务的热迁移能力，将 XPU 上任务热迁移到本地其它服务器上的剩余资源充足的卡上，任务能够继续运行而上层业务无感知，热迁移过程中资源量仍使用 XPU 最小调度单位；
h) 资源生命周期管理：需支持以 XPU 为对象管理资源生命周期，包括但不限于实例化、升级、查询、终止等。

5.3 任务式资源申请与调度

任务式资源申请与调度支持按计算任务粒度实现资源调度，具体要求如下：
a) 计算任务拆分：需支持将应用中的多个计算任务分为一个或多个计算任务组，每个计算任务组中包含一个或多个需在同构加速卡上执行的计算任务；
b) 调度队列：需支持计算任务调度队列，计算任务可加入队列等待分发与计算执行、按优先级分配，需支持多计算任务优先级设置；
c) 计算任务资源建议：需支持根据 XPU 规格对拆分后的一组计算任务提出规格建议；
d) 任务迁移：需支持在异构加速卡间执行计算任务热迁移，迁移后计算任务能够继续运行而上层业务无感知；
e) 任务管理：需支持任务的 GUI 展现，包括对当前任务计算状态、在运行期间的 XPU 算力和显存的分配及实际使用情况进行展现，对以上信息支持任务日志记录及任务报表导出。

5.4 业务编排

业务编排应支持根据应用的算力资源需求分配并编排资源，拉起实例，具体要求如下：
a) 配置图编排：需支持编排和管理配置图，配置图相关配置需包含名称、租户、标签列表、注解、键值对信息；
b) 租户网络编排：需支持编排租户网络，租户网络相关配置需包含名称、租户、标签列表、注解、CNI 插件、物理网络、接口名称前缀、MTU、CIDR 等信息；
c) 存储卷编排：需支持编排存储卷，存储卷相关配置需包含名称、租户、标签列表、注解、存储类别、数据源、卷模式、访问模式、资源需求等信息；
d) POD 编排：需支持编排 POD，一个 POD 下可以定义多个容器，同一个 POD 下的容器共享网络命名空间，POD 相关配置需包含名称、租户、标签列表、注解、优雅关闭时间、容器、卷、重启策略、节点选择规则、镜像抽取密钥、主机名、租户网络列表、DNS 策略、DNS 配置、子域名等信息；
e) 部署编排：需支持编排部署，部署相关配置需包含名称、租户、标签列表、注解、POD 副本数及 POD 模板等信息；
f) 服务编排：需支持编排服务，服务相关配置需包含名称、租户、标签列表、注解、POD 选择规则、服务类型、会话保持策略、暴露的 TCP/UDP 端口列表等信息；
g) 状态集编排：需支编排状态集，状态集相关配置需包含名称、租户、标签列表、注解、关联的服务名、POD 副本数及 POD 模板、PVC 模板等信息；
h) 自动伸缩器编排：需支持编排自动伸缩器，自动伸缩器相关配置需包含名称、租户、标签列表、注解、最大副本数、最小副本数、扩缩目标、扩缩依据等信息。

6　智算中心算力池化接口要求

6.1　概述

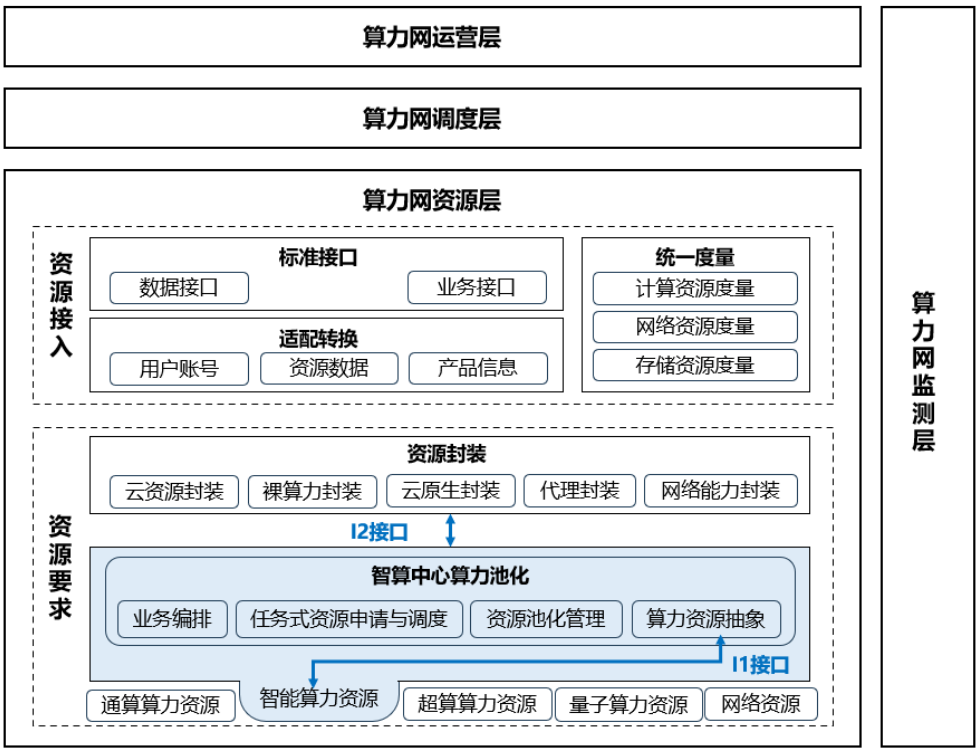

图2　智算中心算力池化接口

智算中心算力池化与算力网资源层其他功能模块之间的接口包括I1、I2接口，见图2。

6.2　I1 接口要求

异构算力资源通过此接口与算力资源抽象对接，以实现设备、计算、内存等资源的上报与映射，以及异构资源个性化操作接口的差异化屏蔽。接口具体要求如下：

a) 应支持向资源封装呈报设备、计算、内存等接口，并支持调度层、监测层通过该接口获取底层异构资源信息以及 XPU 资源信息；
b) 应支持设备管理系列接口，至少支持对异构加速卡的设备注册、设备获取、设备信息获取、设备占用、设备释放；
c) 应支持内存管理系列接口，至少支持对异构加速卡的内存空间创建、内存空间释放；
d) 应支持计算任务核函数管理系列接口，至少支持计算任务创建、计算任务下发、计算任务释放、计算任务参数设置、计算任务使用计算核信息获取；
e) 应支持上下文管理系列接口，至少支持上下文创建、上下文保持、上下文释放；
f) 应支持事件管理接口，至少支持事件等待、事件释放；
g) 应支持根据获取的设备信息将异构加速卡与标准规格 XPU 设备映射，并提供映射比例查询接口。

6.3　I2 接口要求

异构算力资源通过此接口与资源封装对接，以实现算力资源的差异化屏蔽及统一封装接口调用。接口具体要求如下：

a) 应支持资源生命周期管理系列接口，至少支持使用算力的应用的实例化、升级、查询、终止、扩缩容、迁移；
b) 应支持标准 XPU 规格的资源申请接口，并支持将标准规格 XPU 资源申请与物理加速卡资源映射；
c) 应支持算力资源封装类型、属地信息等基础信息注册接口；
d) 应支持以 XPU 为标准规格的算力资源注册接口，该接口应至少支持 XPU 规格、XPU 数量、XPU 算力等属性；
e) 应支持计算任务队列接口，支持业务按照一组计算任务或一个计算任务粒度向下提交计算任务；
f) 应支持计算任务资源查询接口，向业务提供标准 XPU 规格的资源申请建议；
g) 应支持计算任务组合配置接口，支持用户及调度按需描述多个计算任务之间的组合关系；
h) 应支持计算任务迁移接口；
i) 应支持计算任务状态查询接口。

参 考 文 献

[1] 《全国一体化算力网 监测调度平台建设指南》标准草案
[2] 《数据领域常用名字解释（第二批）》