

TC609

全国数据标准化技术委员会技术文件

TC609-XX-2025-XX

全国一体化算力网 算力资源管理与调度 技术要求

National Integrated Computing Network—Technical Specification for
Management and Scheduling of Computing Resources

（征求意见稿）

2025 年 6 月 7 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	II
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 缩略语	1
5 算力资源管理与调度总体功能架构	2
6 算力资源管理与调度功能要求	2
6.1 设计编排	3
6.2 智能调度	6
6.3 资源管理	8
7 算力资源管理与调度接口要求	10
附 录 A （资料性）资源订购使用流程	12
参 考 文 献	14

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。

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 算力资源管理与调度技术要求

1 范围

本文件规范了全国一体化算力网调度层中算力资源管理与调度的技术要求，包括设计编排、智能调度、资源管理等功能及接口要求。

本文件适用于全国一体化算力网监测调度平台中算力资源管理与调度功能建设、改造与优化。

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

调度引擎 Scheduling Engine

基于业务类型、资源状态、网络延迟、能耗等因素，选择最优策略，实现任务自动分发、部署与处理。

3.4

资源建模 Resource Modeling

对资源进行抽象与标准化描述的过程。

3.5

编排模板 Orchestration Template

可复用、结构化的配置定义文件，用于描述如何组合与调度各类基础资源（如计算、存储、网络）以满足特定业务或服务需求。

3.6

服务编排 Service Orchestration

组合资源能力和操作流程以满足业务需求的工作机制。

4 缩略语

SLA：服务等级协议（Service Level Agreement）

TOSCA：云应用的拓扑与编排规范（Topology and Orchestration Specification for Cloud Applications）

ML：机器学习（Machine Learning）
YANG：下一代建模语言（Yet Another Next Generation）

5 算力资源管理与调度总体功能架构

《全国一体化算力网 监测调度平台建设指南》中定义了全国一体化算力网监测调度平台总体架构，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。其中算力网调度层通过标准化接口对异属异构异地算力资源进行统一的算力资源管理，调用网络资源的服务能力实现按需连接传输，并根据算力网运营层下发的用户业务需求进行统筹调度，在满足业务需求的前提下提高全网资源利用率，功能架构见图1。

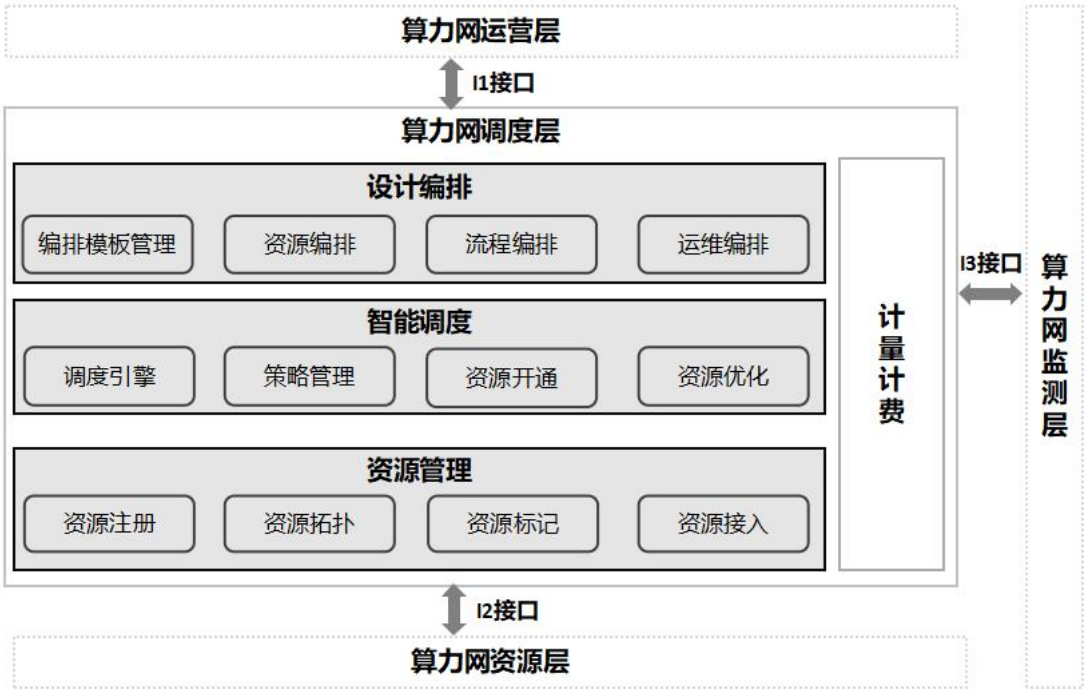


图 1 算力网调度层功能架构图

本技术文件遵从《全国一体化算力网 监测调度平台建设指南》要求，针对算力网资源管理及调度技术进行规范。算力网资源管理与调度技术相关模块包括：

- 1) 设计编排：负责算力网资源的抽象建模、服务模板构建与流程编排定义，实现服务的快速定义与复用。
- 2) 智能调度：负责调度决策核心功能，结合多维度信息进行策略评估与任务分配，实现任务分发与资源开通，提升业务服务质量和资源利用效率。
- 3) 资源管理：负责对算力、网络资源的状态、性能、拓扑等信息的管理。

本文件将规范上述模块的功能要求，并明确上述模块与周边协同关系。

注：权限与角色控制相关要求，参考《全国一体化算力网 算力计量计费技术要求》，本文件不做展开要求。

6 算力资源管理与调度功能要求

6.1 设计编排

6.1.1 编排模板管理

编排模板管理能力要求旨在提升模板的可复用性与个性化程度，满足多样化编排场景下的高效配置与灵活管理需求，具体要求如下：

- 应提供基础编排模板；
- 应支持查看编排模板的类别、名称、描述、资源组成；
- 应支持基于基础编排模板进行自定义配置，包括但不限于资源组成配置、资源规格配置；
- 应支持具有相应权限的角色基于任务数据、经验数据等信息进行模板创建、更新；
- 应支持具有相应权限的角色进行模板的创建、修改、删除、查看，以及版本控制（包括版本迭代与回滚）；
- 应支持模板市场，具有相应权限的角色可在模板市场进行模板发布、下线操作；
- 宜提供基于用户画像的推荐能力，基于用户历史行为数据进行编排模板定向推荐。

编排模板相关信息见表 1。

表 1 编排模板信息

分类	信息	说明
模板基本信息	模板 ID	模板唯一标识符
	名称	模板名称
	类别	模板分类（如通用类、存储类、网络类等）
	描述	对模板功能、适用场景的说明
	创建者	创建该模板的用户或角色标识
	创建时间	模板创建的时间戳
资源组成信息	资源列表	构成模板的资源项清单（如计算节点、存储资源、网络资源等）
	资源类型	每个资源项的类型（如 VM、容器、GPU 等）
	资源配置参数	每个资源项的规格参数（如 CPU、内存、存储大小等）
	拓扑关系	各资源间的依赖或连接关系（如网络连接、服务依赖等）
版本与变更管理	版本号	当前模板的版本标识
	版本说明	版本的简要说明
	历史版本列表	记录过往版本及其内容
	回滚标志	是否可回滚到该版本
	更新时间	当前版本更新时间戳
模板使用与推荐信息	使用频率	模板被使用的次数或频率统计
	用户行为标签	基于用户使用历史形成的行为标签
模板状态与市场信息	模板状态	当前状态（草稿、已发布、已下线等）
	发布者	执行发布操作的角色
	市场标签	发布至模板市场时的分类、标签信息

6.1.2 资源编排

- 资源编排模块负责各类资源建模与组合配置等，实现资源能力编排和模板定义。具体要求如下：
- 应支持对多类型资源（算力、存储、网络等）的基础属性（如 CPU 核数，内存容量，存储类型，带宽等）进行标准化建模（如 TOSCA、YANG）；
 - 应支持对异构资源基于统一逻辑进行抽象，并支持组合配置；
 - 宜支持通过图形化界面提供资源模板的可视化创建与复用，提升用户操作便捷性；
 - 宜支持多租户场景下的资源模板隔离与权限控制。

资源编排相关信息见表 2。

表 2 资源编排信息

分类	信息	说明
资源基本信息	资源 ID	资源的唯一标识符
	名称	资源名称
	类型	资源类型（如算力、存储、网络等）
	描述	对资源功能的简要说明
	所属模板 ID	资源关联的模板唯一标识
	所属租户	所属的租户或组织标识
资源属性建模信息	属性建模语言	用于描述资源属性的建模语言（如 TOSCA、YANG）
	资源属性集合	资源的基础配置属性（如 CPU 核数、内存大小、带宽等）
	弹性能力	是否支持伸缩、动态扩展等能力
	规格级别	高性能型、通用型等资源等级分类
	异构标识	是否为异构资源，及其抽象层次
资源组合与拓扑信息	组合 ID	资源组合所属的组合标识
	拓扑结构	资源之间的连接或依赖关系
	依赖关系	资源组件之间的先后部署或依赖逻辑
	调度策略	与资源组合相关的编排调度策略（如低延迟优先）
抽象与适配信息	适配器类型	资源适配所使用的对接插件类型（如 Kubernetes、裸金属 API）
	物理映射关系	抽象资源与底层资源的映射关系
可视化与复用信息	是否支持可视化建模	是否支持图形化界面呈现与建模
	是否可复用	是否支持作为模块在其他模板中复用
	版本控制	是否支持资源组件版本管理及记录

6.1.3 流程编排

- 流程编排面向业务场景定义资源操作流程，数据流转路径与操作间依赖关系。具体要求如下：
- 应支持基于业务场景定义服务操作流程（如状态转移、流程控制），包括服务流程的创建，修改、查询、删除；
 - 应支持资源调用顺序与依赖关系的建模，包括创建，修改，查询、删除；

- 应支持数据流转路径控制，包括配置、修改、删除、查询；
- 应支持多流程并发执行与编排状态监控；
- 应支持编排流程的异常处理（如暂停，回滚）与跳转处理（如事件钩子、事件回调）；
- 宜支持流程编排相关模板（服务流程、资源调用、资源依赖模板）的导入、导出、共享，便于跨场景迁移和快速上线。

流程编排相关信息见表 3。

表 3 流程编排信息

分类	信息	说明
基本信息	流程 ID	唯一标识该流程
	流程名称	流程的可读名称
	所属业务场景	流程对应的业务类型
操作流程定义	状态转移定义	定义流程中的各个状态及其转移逻辑
	控制逻辑	包括条件跳转、循环、并发等控制逻辑
资源依赖建模	调用顺序	各资源调用的时序依赖
	资源依赖关系	资源之间的依赖链（如 A 依赖 B 创建）
数据流控制	数据路径定义	各操作之间数据的输入输出关系
	数据格式/类型	流转的数据类型定义
异常处理	异常捕获策略	对异常流程如暂停、回滚的处理策略
	跳转机制	支持事件钩子、回调等跳转逻辑
模板管理	模板导入导出	是否支持流程模板的导入、导出、共享功能

6.1.4 运维编排

运维编排提供运维任务部署与运行，资源分配与回收等生命周期流程的图形化策略式编排。具体要求如下：

- 应支持运维服务的部署、重部署、运行、暂停、取消全过程控制；
- 应支持运维任务生命周期流程的策略化定义，如超出 CPU 使用率 90% 自动扩容；
- 应支持运维任务的自动触发，如定时健康检查，异常告警备份等；
- 宜支持异常节点的故障隔离与替代方案。

运维编排相关信息见表 4。

表 4 运维编排信息

分类	信息	说明
基本信息	任务 ID	运维任务唯一标识
	任务名称	运维任务名称
	任务类型	运维任务类型，如部署、暂定、取消等
	生命周期状态	任务当前状态，如运行中，已暂停，已取消等
	创建时间	任务创建时间
	修改时间	任务修改时间

	结束时间	任务结束时间
策略定义	触发条件	如定时、事件、监控指标触发等
	策略动作	如扩容、缩容、重部署等
自动化处理	异常检测处理策略	如故障隔离、替代方案等
可视化配置	图形化描述	运维编排是否支持图形化方式定义

6.2 智能调度

6.2.1 调度引擎

调度引擎基于业务类型、资源状态、网络延迟、能耗等因素，选择最优策略，实现任务自动分发、部署与处理。具体要求如下：

- 应支持基于多因子（至少包括算力类型、性能、能耗、延迟、位置）的综合调度决策；
- 应支持对多调度因子设置权重排序；
- 应支持任务的扩缩容管理；
- 应支持基于任务优先级的调度排序；
- 宜支持基于 AI 算法进行调度优化，如多任务并发调度、依赖控制、动态优化。

调度引擎相关信息见表 5。

表 5 调度引擎信息

分类	信息	说明
基本信息	调度任务 ID	唯一标识一项调度任务
	调度任务类型	所调度的任务种类（算力、存储、网络等）
调度因子	算力类型	支持的算力类型（CPU/GPU 等）
	网络延迟	网络性能因子
	能耗	能效指标
	地理位置	资源部署的物理位置
策略控制	权重设置	支持对调度因子进行加权
	优先级控制	支持对任务设置调度优先级
AI 调度优化	AI 算法类型	是否使用 ML/AI 调度策略（如强化学习等）
	优化目标	吞吐量最大化、延迟最小化、多任务调度等

6.2.2 策略管理

策略管理支持调度策略配置、删除、修改、查询等，支持策略插件式扩展，支持针对调度策略进行仿真评估。具体要求如下：

- 应支持调度策略的创建、下发、查询、修改与删除；
- 应支持调度策略的多维度匹配与优先级控制；
- 应支持调度策略的组合与嵌套管理；
- 应支持调度策略生效范围的灵活配置（区域/业务/租户等）；
- 宜支持调度策略的调试、验证，具备策略仿真测试与回滚能力；
- 宜支持动态插件式策略扩展。

策略管理相关信息见表 6。

表 6 策略管理信息

分类	信息	说明
基本信息	调度策略 ID	唯一标识一条调度策略
	策略名称	可读策略名称
策略定义	匹配规则	多维匹配条件
	优先级	策略匹配优先级
	策略组合关系	是否为嵌套组合策略
	生效范围	策略作用的区域、业务、租户等
策略维护	验证方式	策略测试/验证机制
	扩展机制	是否支持插件式扩展

6.2.3 资源开通

资源开通能力要求如下：

- 应支持用户对编排方案进行选择确认；
- 宜支持多种编排方案对比能力，输出对比结果及选择建议。
- 应支持基于用户确认的编排方案生成对应资源调度指令集，为用户完成资源开通及部署；
- 应提供资源调度指令集管理能力，包括但不限于指令集中调度指令的查看、修改、删除等。

资源开通相关信息见表 7。

表 7 资源开通信息

分类	信息	说明
开通方案	方案 ID	编排方案标识
	方案名称	用户选择的开通方案名称
	方案类型	方案的类型，如预定义，自定义等
	方案描述	对方案的可读性描述
	方案状态	方案的执行状态，如待确认，已部署等
方案对比	方案列表	对比方案列表
	对比结果	方案的对比结果
	方案推荐	基于对比结果的优选方案
指令集管理	指令集 ID	对应开通方案生成的调度指令集 ID
	指令明细	包括创建资源、配置网络、挂载存储等动作
	指令编辑能力	是否支持查看、修改、删除指令集

6.2.4 资源优化

资源优化支持以保障用户业务服务质量为前提，基于业务运行情况、资源实时状态等信息进行动态资源优化。具体要求如下：

- 应提供基于 SLA 服务保障协议，明确在用户业务运行过程中可提供的稳定性保障；
- 应支持计算任务当前执行进度和计划进度的偏差检测能力；
- 应提供运行时自优化策略，支持以运行态保障要求为前提，基于用户业务运行情况、资源状态为进行动态资源优化（如闲时关停用户闲置资源）；

——宜支持资源热点检测与分流；
 ——宜支持基于滑动窗口或 ML 算法，进行资源负载预测。
 资源优化相关信息见表 8。

表 8 资源优化信息

分类	信息	说明
基本信息	调度策略 ID	唯一标识一条调度策略
	策略名称	可读策略名称
策略定义	匹配规则	多维匹配条件
	优先级	策略匹配优先级
	策略组合关系	是否为嵌套组合策略
	生效范围	策略作用的区域、业务、租户等
策略维护	验证方式	策略测试/验证机制
	扩展机制	是否支持插件式扩展

6.3 资源管理

6.3.1 资源注册

资源属性注册能力要求如下：

- 应支持枢纽/省/行业算力调度平台注册，支持配置包括但不限于下设纳管的不同类型算力类型；
- 应支持公有云、专有云、边缘云等多种云平台注册，支持配置包括但不限于云平台类型、所属服务商等信息；
- 应支持算力资源节点注册，支持配置包括但不限于算力资源节点所属地域、可用算力资源类型、可用算力资源规模等信息；
- 应支持算力资源注册，支持配置包括但不限于算力资源供给方式、算力资源搭载的 CPU、GPU 等芯片类型、芯片架构、芯片型号等信息；
- 应支持网络资源注册，包括但不限于私有网络、专线、专网等多种网络资源，支持配置网络资源规格、所属地域、所属服务商等信息；
- 应支持存储资源注册，如对象存储、文件存储、块存储等多种存储资源，支持配置存储资源规格、所属地域、所属服务商等信息；
- 应支持安全资源注册，如防火墙等多种安全资源。

资源注册相关信息见表 9。

表 9 资源注册信息

分类	信息	说明
平台信息	调度平台 ID	枢纽、省、行业调度平台标识
	云平台类型	公有/专有/边缘云
节点信息	节点 ID	资源节点唯一标识
	所属地域	地理区域信息
	可用资源类型/规模	节点可提供的 CPU/GPU 等种类与规模
资源明细	芯片类型/型号	资源搭载的计算芯片信息
	网络资源信息	专网/私网/专线等参数

	存储资源信息	对象存储、块存储、文件存储等类型及规格
	安全资源信息	如防火墙、WAF 等

6.3.2 资源拓扑

资源拓扑通过采集各区域资源分布、连接关系、网络性能与健康状态，构建动态资源拓扑视图。具体要求如下：

- 应支持各区域资源节点的识别与拓扑构建；
- 应支持实时性能指标与资源拓扑的动态关联；
- 应支持拓扑结构的动态更新；
- 宜支持拓扑图可视化展示、层级缩放与历史状态回溯。

资源拓扑相关信息见表 10。

表 10 资源拓扑信息

分类	信息	说明
节点标识	节点 ID	拓扑中识别的节点唯一标识
	节点类型	计算/网络/存储的节点类型
拓扑关系	拓扑连接关系	节点之间的连接结构
性能数据	性能关联	将性能指标（延迟、负载等）关联到拓扑结构上。
可视化信息	拓扑展示	拓扑的图形化、层级缩放及历史状态查看

6.3.3 资源标记

资源标记根据资源能力、位置、类型、能耗等维度，生成用于调度的资源标记，支持动态更新与策略引用。具体要求如下：

- 应支持对资源设置标签（如地理位置、资源类型、功耗等级等）；
- 应支持对批量资源设置统一标签；
- 应支持动态更新与标记继承机制；
- 应支持资源标记与策略规则的绑定使用，如基于标签过滤匹配资源；
- 宜支持标签库管理，包括标签的创建，删除、修改、查询；
- 宜支持标签自动生成与推荐功能。

资源标记相关信息见表 11。

表 11 资源拓扑信息

分类	信息	说明
基本信息	资源 ID	被标记资源的唯一标识
	标签 ID	标签的唯一标识
	标签名称	可读的标签名称
	标签分组	如位置类，功耗类，资源标识类
标签管理	批量标识	是否批量标签标识
	标签组配置	用于批量配置标签

	标签继承	用于定义资源之间标签继承关系的配置策略
	创建时间	标签创建时间
	修改时间	标签最近一次修改时间
策略绑定	策略匹配	标签是否用于策略过滤/匹配
	自动生成与推荐	基于策略自动推荐或打标签
标签库	标签状态	标签是否可用
	操作标识	标签支持的操作（如修改，删除，共享等）

6.3.4 资源接入

资源接入负责对编排调度的资源进行接入管理。具体要求如下：

- 应支持多协议、多来源、多类型的算力资源接入格式转换；
- 应支持收集并处理算力资源性能信息，包括但不限于 CPU 利用率、内存利用率等指标；
- 应支持收集并处理网络资源性能信息，包括但不限于时延、带宽、抖动、丢包等指标；
- 应支持收集并处理存储资源性能信息，包括但不限于存储资源 IOPS、吞吐量等指标；
- 应支持对接入资源进行合规性检查；
- 应支持接入资源的健康监控与状态同步管理；
- 宜支持资源接入流程的可视化引导与日志审计。

资源接入相关信息见表 12。

表 12 资源接入信息

分类	信息	说明
基本信息	资源 ID	被接入资源的唯一标识符
	资源类型	标识接入资源的类型（如计算、网络、存储等）
	来源类型	资源来源渠道，如公有云、专有云、边缘云等
	协议类型	用于接入该资源的通信协议类型
性能信息	计算性能指标	如 CPU 利用率、内存使用率等指标
	网络性能指标	如时延、带宽、抖动、丢包等
	存储性能指标	如 IOPS、吞吐量等
合规与状态	合规性状态	资源是否符合接入标准的校验结果（合规/不合规）
	健康状态	接入资源的运行健康状态
	同步时间	最近一次状态同步的时间戳
管理与审计	格式转换配置	记录接入过程中的格式转换规则与配置
	可视化引导信息	与资源接入相关的可视化流程配置（如接入引导流程 ID）
	日志审计记录	记录资源接入过程中各步骤的操作日志

7 算力资源管理与调度接口要求

算力调度层与周边其他层之间的接口包括I1、I2、I3接口。

I1接口：算力资源管理与调度通过此接口，向算力网运营层提供可调用的原子功能及服务，及任务相关性能指标，告警信息。接口具体要求如下：

- 应支持调度原子能力（如资源申请、调度请求、任务部署等）的接口暴露；
- 应支持按任务维度提供关键性能指标（KPI）与运行时状态反馈；
- 应支持任务执行过程中的实时告警上报与异常信息推送；
- 宜支持 RESTful、gRPC 等主流接口协议，并保障接口可扩展性。

相关接口信息可参考《全国一体化算力网 算力运营服务技术要求》。

I2接口：算力资源管理与调度通过此接口，为任务分配算力网资源，获取算力网资源实时状态。接口具体要求如下：

- 应支持资源的接入的请求与响应；
- 应支持资源申请、释放、调整等资源操作请求的发起与响应；
- 应支持对特定资源进行占用、隔离、回收等控制操作；
- 应支持跨区域资源的感知与调度；
- 应支持告警事件的联动报告，包括调度失败、资源失效等；
- 宜支持对资源操作的事件追踪与结果回传。

相关接口信息可参考《全国一体化算力网 算力并网技术要求》。

I3接口：算力资源管理与调度通过此接口，从算力监测层定时获取全局资源信息及实时算力运行信息。接口具体要求如下：

- 应支持全局资源信息的定期同步；
- 应支持通过监测层实时查询算力资源运行信息；
- 应支持资源状态实时查询（包括空闲率、负载、位置、功耗等）；
- 宜支持以订阅-发布模式进行数据传输，提升监控系统响应效率；
- 宜支持多维度的监控指标定义（如资源利用率、延迟、能耗等）。

相关接口信息可参考《全国一体化算力网 算力监测技术要求》。

附录 A（资料性）

资源订购使用流程

A.1 用户明确资源需求的订购使用流程

用户可通过浏览运营层门户，明确并选择所需算力类型、规格、数量、算力资源供给方等信息后，资源订购及使用流程见图A.1：

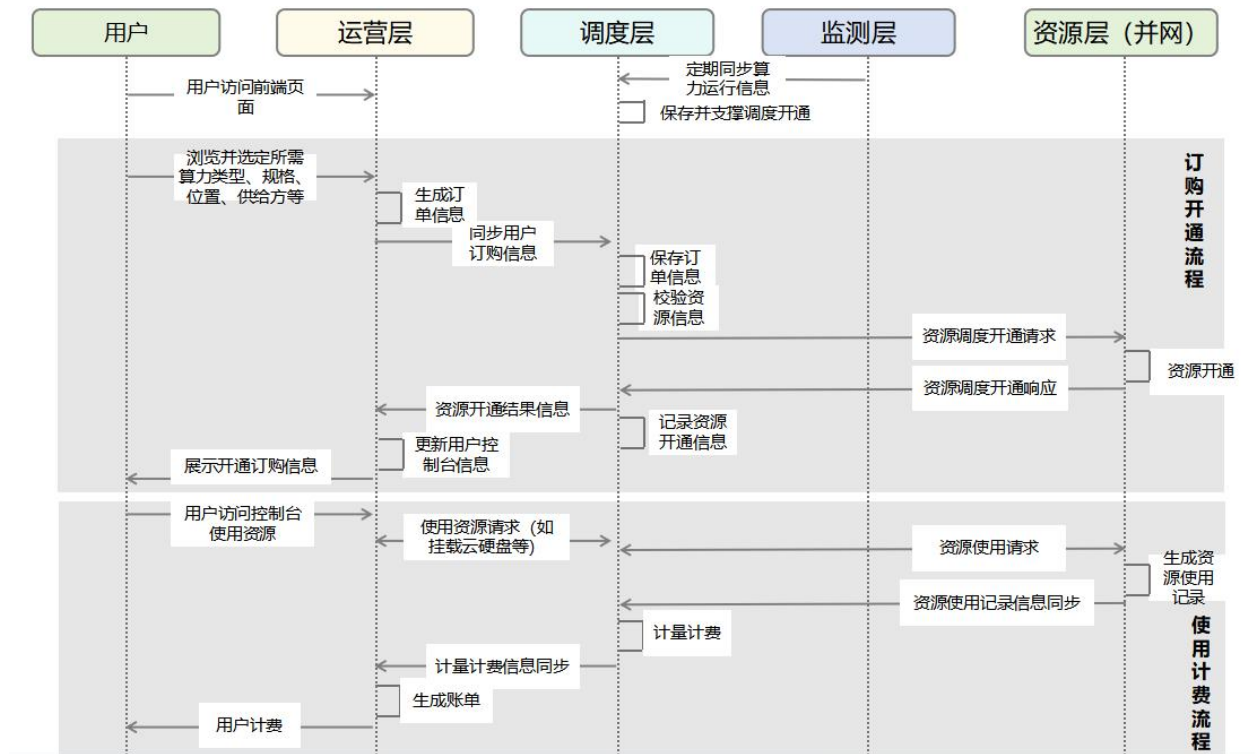


图 A.1 用户明确资源需求的订购使用流程

A.2 任务式服务订购使用流程

在用户仅提出业务需求或大致资源需求，而不明确或无需关注算力供给方或未知等信息时，以保障用户业务需求为主，可通过调度层、监测层、并网层等交互为用户提供合适的资源方案，用户确定方案并订购后，再进行资源使用、计量计费流程，见图A.2。

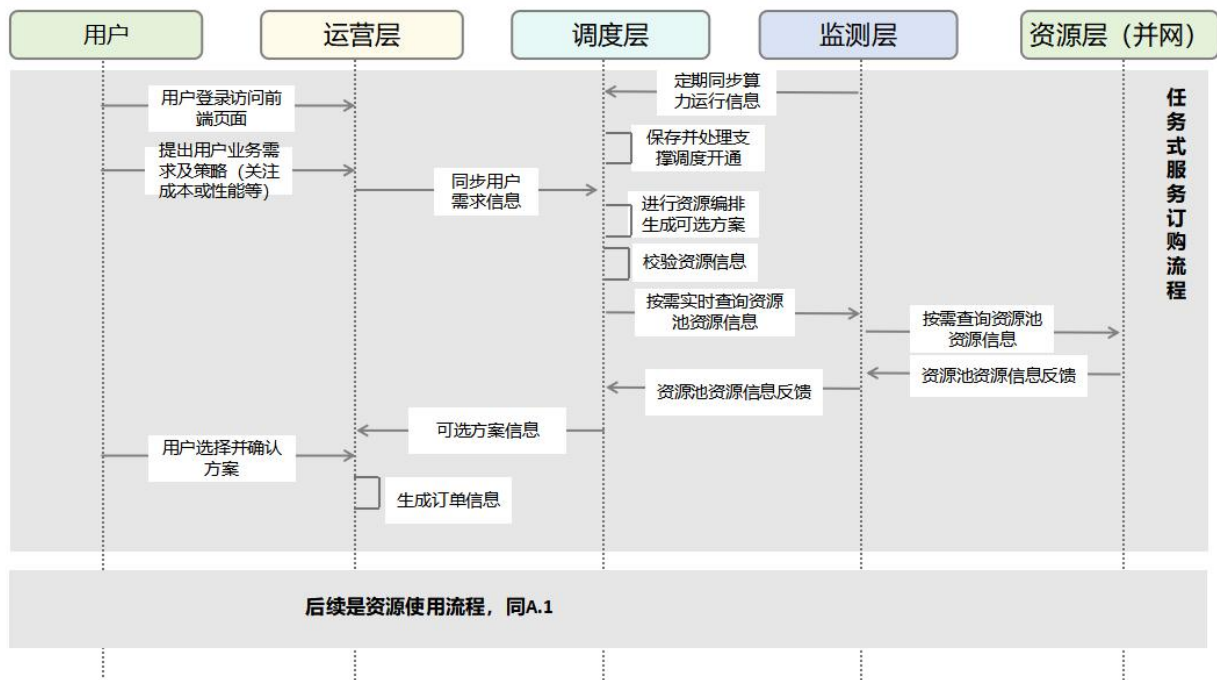


图 A.2 任务式服务的订购使用流程

参 考 文 献

- [1] 国家数据基础设施建设指引（发改数据〔2024〕1853号）
 - [2] 《全国一体化算力网 监测调度平台建设指南》（数标委〔20250425发布版〕）
 - [3] YD/T 4255-2023 《算力网络 总体技术要求》
 - [4] YD/T 6046-2024 《算力网络 算网编排管理技术要求》
 - [5] YD/T 6116.1-2024 《算力互联互通能力要求 第1部分：总体框架》
 - [6] 2024-0579T-YD 《算力互联互通能力要求 第2部分：算力调度》
 - [7] 2022-1553T-YD 《泛在算力调度管理技术要求》
-