

# 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 算力多量纲计费 技术要求

National Integrated Computing Network—Technical Requirements for  
Computing Measurement and Billing

(征求意见稿)

2025年6月7日

2025-XX-XX 发布

2025-XX-XX 实施

全国数据标准化技术委员会 发布



# 目 次

前言.....	III
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 总体架构 .....	3
5.1 概述 .....	3
5.2 逻辑架构 .....	3
6 算力量纲 .....	4
6.1 面向资源节点 .....	4
6.1.1 通用计算 .....	4
6.1.2 智能计算 .....	5
6.1.3 超级计算 .....	5
6.1.4 网络 .....	6
6.1.5 存储 .....	6
6.2 面向任务节点 .....	7
6.2.1 模型类任务 .....	7
6.2.2 渲染类任务 .....	7
6.2.3 数据类任务 .....	8
6.2.4 虚拟化任务 .....	8
7 算力计费 .....	9
7.1 面向资源节点 .....	9
7.1.1 按量计费 .....	9
7.1.2 周期计费 .....	9
7.1.3 卡包计费 .....	9
7.2 面向任务节点 .....	9
7.2.1 调用计费 .....	9
7.2.2 周期计费 .....	10
7.2.3 节点计费 .....	10
7.2.4 卡时计费 .....	10
8 功能管理 .....	10
8.1 采集接口 .....	10
8.2 数据采集 .....	10
8.3 数据校验 .....	11
8.4 计费规则配置 .....	11

8.5 统一记账 .....	11
8.6 历史追溯 .....	11
8.7 权限管理 .....	11
8.8 可视分析 .....	11
附 录 A (资料性) 示例 .....	12
A.1 通用计算案例 .....	12
A.2 智能计算示例 .....	12
A.3 超级计算示例 .....	12
A.4 网络示例 .....	13
A.5 存储示例 .....	13
A.6 算力卡示例 .....	14
A.7 模型类任务示例 .....	14
A.8 渲染类服务示例 .....	15
A.9 数据类任务示例 .....	16
A.10 虚拟化任务示例 .....	16
参 考 文 献 .....	18

# 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 算力多量纲计费技术要求

## 1 范围

本文件给出了全国一体化算力网算力多量纲计费的技术要求，提出了多量纲计费的功能框架，包含通用计算、智能计算、超级计算、存储资源、网络资源、模型类任务、渲染类任务、数据类任务和终端服务类任务等的多量纲定义及计费功能。

本文件适用于全国一体化算力网算力服务的多量纲计费实施。

## 2 规范性引用文件

本文件没有规范性引用文件。

## 3 术语和定义

下列术语和定义适用于本文件。

### 3.1

#### 资源节点 Resource Node

算力资源聚集地，提供包含计算、网络、存储等综合能力，根据聚集规模可分为卡资源节点、柜资源节点、集群资源节点等。

### 3.2

#### 任务节点 Task Node

算力任务服务使用者下达任务命令的逻辑节点，任务下达通常通过系统节点与资源调用相联。

### 3.3

#### 核时 Duration of Core Occupancy

一种算力服务的衡量单位，为给定环境下，处理器核心数和使用时长的乘积。

### 3.4

#### 卡时 Duration of Computing Card

单位时间内的标准化资源组合单位，为给定环境下（包含网络、存储等），GPU或计算卡的占用时间。

### 3.5

#### 标记单元 Token

一种算力服务的衡量单位，文本处理的最小单元。在模型处理过程中，数值大小与处理文本速度、能力成正比关系。

### 3.6

#### 资源类服务 Resoure-Based Service

将计算资源（如服务器、存储设备、网络带宽等）以一种可量化的资源形式提供给用户。一般分为通用计算、智能计算、超级计算。

### 3.7

### **任务类服务 Task-Based Service**

以完成特定任务为目标，为用户提供计算能力的服务模式。用户只需提交任务描述和相关数据，服务提供商负责调配所需的计算资源来完成任务。一般分为大模型服务、自然语言处理、图像视频识别、车联网训推一体化、视频分化渲染、调度服务和数据快递。

#### **3.8**

### **通用计算 General Computing**

也称通算，由基于CPU（中央处理器）芯片的服务器提供，是基础的计算能力，主要用于运行通用计算任务。

#### **3.9**

### **智能计算 Intelligent Computing**

也称智算，侧重于人工智能相关的计算需求，由GPU（图形处理器）、FPGA（现场可编程门阵列）、ASIC（专用集成电路）等异构计算芯片提供。

#### **3.10**

### **超级计算 Supercomputer**

也称超算，超级计算机提供的计算能力，执行复杂且计算密集型的任务。

#### **3.11**

### **网络 Network**

是指基于计算机网络技术，通过分布式软件模块或平台为用户提供数据交换、资源共享、业务支持等功能的综合性服务。

#### **3.12**

### **存储 Storage**

通过技术手段将数据保存到特定介质中，并实现有效访问的业务模式。

#### **3.13**

### **模型类任务 Model-related Tasks**

基于超大规模预训练人工智能模型（如GPT等）构建的智能化服务，依托深度学习、自然语言处理等技术，通过API接口或云平台对外提供文本生成、语义理解、数据分析、智能交互等能力。包含模型训练、推理及垂直技术应用（语言/视觉）等。

#### **3.14**

### **渲染类任务 Rendering-related Tasks**

基于云计算技术的渲染服务模式，用户通过互联网将需要渲染的任务（如3D动画、影视特效、建筑可视化等）提交到云端服务器集群，由云端的高性能计算资源完成图像或视频的渲染处理，最终将结果返回给用户。

#### **3.15**

### **数据类任务 Data-related Tasks**

数据类任务是指数据处理、数据分析、数据管理、数据应用、数据传输、数据存储等数据类操作任务。

#### **3.16**

### **终端服务类任务 Terminal-service-related Tasks**

终端服务类任务是指通过网络连接、云端虚拟化等形式，为远程用户提供对计算机系统、应用程序或数据的访问与操作支持，实现跨地域的资源共享和协同工作。

## **4 缩略语**

下列缩略语适用于本文件。

AI: 人工智能 (Artificial Intelligence)

API: 应用程序编程接口 (Application Programming Interface)

CPU: 中央处理器 (Central Processing Unit)

GPU: 图像处理器 (Graphics Processing Unit)

NPU: 神经处理器 (Neural Processing Unit)

QPU: 量子处理器 (Quantum Processing Unit)

TPU: 张量处理器 (Tensor Processing Unit)

GB: 吉字节 (Gigabyte)

TB: 太字节 (Terabyte)

Mbps: 兆比特每秒 (Megabits per second)

FLOPs: 每秒浮点运算次数 (Floating-Point Operations Per Second)

## 5 总体架构

### 5.1 概述

算力多量纲计价, 应面向用户, 是提取关键计量要素, 综合算力服务能力进行标准化计费的过程, 以解决算力资源价值衡量与交易规则问题, 推动算力资源高效配置与市场化流通, 是推动全国一体化算力网形成的内生力量。

### 5.2 逻辑架构

全国一体化算力网 算力多量纲计费为算力网调度层的重要组成部分, 提供国家级、区域级、城市级算力网服务的统一计费规范及功能要求, 涵盖算网资源式服务、算网任务式服务的计费规范, 向下支持算力网资源层资源输入的标准化计费, 向上支撑算力网运营层服务输出的标准化计费, 是全国一体化算力网高效运行, 服务高质流转的核心要素。全国一体化算力网 算力多量纲计费总体分为算力量纲、算力计费、功能管理三个组成部分。

全国一体化算力网算力多量纲计费的总体架构及功能模块见图1。

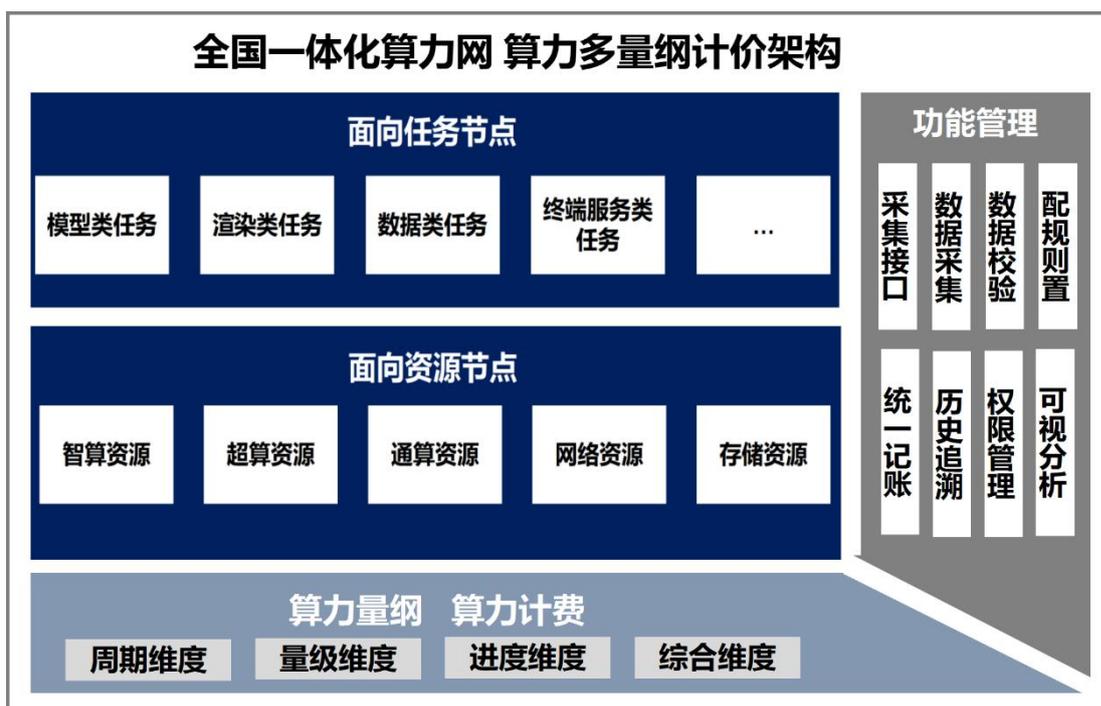


图 1 全国一体化算力网 算力多量纲计费参考框架

随着数字中国走向新发展阶段，应用需求侧对算网服务提出了新要求，呈现出新发展特征。一方面，算网服务逐渐从供给本位走向需求本位，从传统资源式供应走向任务式配给；另一方面，传统资源式本身也呈现出融合、多样的供应形式。任务式配给，以完成任务指令为目标，通过平台自动配给所需算力、网络、存储、数据、模型等资源，故而其计费量纲较之传统资源供给均发生较大变化，是故算力计费分为面向资源节点服务（即资源式服务）的计量计费，以及面向任务节点服务（即任务式服务）的计量计费。

算力量纲中，面向资源节点的计量对智算资源、超算资源、通算资源、网络资源、存储资源典型资源类服务分别进行计量规范，面向任务节点的计量对模型类任务、渲染类任务、数据类任务、终端服务类任务现阶段典型任务类服务分别进行规范，未来可扩充典型任务类服务。

算力计费中，分别对面向资源节点及面向任务节点的计价进行规范总结，分别适用于两类节点框架内服务。

功能要求部分对算力多量纲计费全周期基本功能项做出要求，包含采集接口、数据采集、数据校验、配置规则、统一记账、历史追溯、权限管理、可视分析。

## 6 算力量纲

### 6.1 面向资源节点

#### 6.1.1 通用计算

通过统计通用计算节点综合资源使用情况进行计量，量纲包含：

- 支持接入算力网监测/采集系统，对计算资源、网络资源、内存资源、存储资源进行采集；
- 支持对处理器数量、CPU 使用量、内存大小、硬盘存储空间大小、硬盘个数、网络带宽进行

计量；可支持能耗、运维成本、用算时间区间（闲时/忙时）、用量区间、服务类型计量。

量纲计量项见表 1。

表 1：通用计算量纲

计量项选择	单位	说明
处理器数量	个	处理器个数越多，性能越好，资源消耗越大。
CPU 使用量	核时	核时= CPU 核数 × 小时
内存	GB	内存容量一般指节点的随机存储器（RAM）的容量。
硬盘存储空间	TB	硬盘存储容量是指存储器可以容纳的二进制信息量。
网络带宽	Mbps	单位时间内能发送/接收的最大数据量。

### 6.1.2 智能计算

通过统计智能计算节点综合资源的使用情况进行计量，量纲包含：

——支持接入算力网监测/采集系统，对计算资源、网络资源、内存资源、存储资源进行采集；

——支持对处理器数量、CPU使用量、内存、硬盘存储空间、GPU型号、GPU卡数、网络带宽、集群架构等进行计量；可支持对能耗、运维成本、用算时间区间（闲时/忙时）、用量区间进行计量。

量纲计量见表2。

表 2：智能计算量纲

计量项选择	单位	说明
处理器数量	个	处理器个数越多，性能越好，资源消耗越大。
CPU 使用量	核时	核时= CPU 核数 × 小时
GPU 使用量	卡时	卡时=GPU 卡数× 小时
GPU 使用量（或）	FLOPs	FLOPs = 核心数*每核心每周期 FLOP 数*频率 (Hz)
内存	GB	内存容量一般指节点的随机存储器（RAM）的容量
硬盘存储空间	TB	硬盘存储容量是指存储器可以容纳的二进制信息量
网络带宽	Mbps	单位时间内能发送/接收的最大数据量。

### 6.1.3 超级计算

通过统计超算节点综合资源使用情况进行计量，量纲包含：

——支持接入算力网监测/采集系统，对计算资源、网络资源、内存资源、存储资源进行采集；

——支持对处理器数量、CPU性能、内存、GPU型号、GPU卡数、网络延迟、集群架构进行计量；

可支持对能耗、运维成本、用算时间区间（闲时/忙时）、用量、用途进行计量。

量纲计量见表3

表 3：超级计算量纲

计量项选择	单位	说明
处理器数量	个	处理器个数越多，性能越好，资源消耗越大
CPU 使用量	核时	核时= CPU 核数 × 小时
GPU 使用量	FLOPs	FLOPs = 核心数*每核心每周期 FLOP 数*频率 (Hz)
内存	GB	内存容量一般指节点的随机存储器（RAM）的容量
硬盘存储空间	TB	硬盘存储容量是指存储器可以容纳的二进制信息量
网络带宽	Mbps	单位时间内能发送/接收的最大数据量

#### 6.1.4 网络

通过统计网络资源使用情况进行计量，量纲包含：

——支持接入算力网监测/采集系统，对网络资源进行采集；

——支持对网络带宽、节点连接密度、端到端时延、流量等计量；可支持网络扩展能力。

量纲计量见表4.

表 4：网络量纲

计量项选择	单位	说明
对内带宽	Mbps	决定了存储系统内部数据传输的速度
对外上行带宽	Mbps	存储系统向外部设备或网络传输数据的能力
对外下行带宽	Mbps	决定了存储系统从外部设备或网络接收数据的能力
流量	GB	反应用户真实使用的流量

#### 6.1.5 存储

通过统计存储资源使用情况进行计量，量纲包含：

——支持接入算力网监测/采集系统，对内存资源、存储资源进行采集；  
 ——支持对内存、磁盘容量、访问次数、访问流量进行计量。  
 量纲计量见表5

表 5：存储量纲

计量项选择	单位	说明
内存容量	GB	内存容量一般指节点的随机存储器（RAM）的容量
磁盘容量	TB	磁盘存储容量是指存储器可以容纳的二进制信息量
最高读写速率	GB/s	存储设备在数据传输时能达到的最快读取与写入速度
访问次数	万次	访问存储内文件的次数
访问流量	GB	访问存储内文件产生的对外下行流量

## 6.2 面向任务节点

### 6.2.1 模型类任务

面向模型类服务，以大模型服务、自然语言处理、图像视频识别等模型类任务所需综合能力为计量对象，量纲包含：

——支持接入算力网监测/采集系统，对并行上限、上下文窗口长度、多模态进行采集；  
 ——支持对模型参数量级、请求任务数量/大小/时间并发上限数量、窗口大小和多模态类型计量；  
 可支持对用算时间区间（闲时/忙时）、用量区间、用途计量。

量纲计量见表6

表 6：模型类任务量纲

计量项选择	单位	说明
并发上限	个	同时处理任务个数的上限
Token 消耗量	个	最小的文本处理单元
上下文长度	Mb	模型一次性处理的最大文本范围

### 6.2.2 渲染类任务

面向渲染类任务，以云渲染等渲染类任务所需综合算力能力为计量对象，量纲包含：  
 ——支持接入算力网监测/采集系统，应对计算能力、网络能力、存储能力等进行采集；  
 ——支持CPU核心数、GPU型号、GPU性能、内存与显存容量、数据传输速率计量；

——可支持包算时间区间（闲时/忙时）、用量区间、用途计量。  
量纲计量见表7。

表 7：渲染类量纲

计量项	单位	说明
处理器数量	个	处理器个数越多，性能越好，资源消耗越大。
CPU 使用量	核时	核时= CPU 核数 × 小时
GPU 使用量	FLOPs	FLOPs = 核心数*每核心每周期 FLOP 数*频率(Hz)
内存	GB	内存容量一般指节点的随机存储器（RAM）的容量
存储空间	TB	硬盘存储容量是指存储器可以容纳的二进制信息量
渲染分辨率	4K	分辨率 = 横向像素数 × 纵向像素数

### 6.2.3 数据类任务

面向数据快递任务，以数据快递等数据类任务为计量对象，量纲包含：

——支持接入算力网监测/采集系统，应对传输设备、网络能力、存储能力等进行采集；

——支持对传输方式、存储空间、数据导入存储设备时长进行计量。

量纲计量见表8。

表 8：数据类量纲

计量项选择	单位	说明
存储空间	TB	磁盘存储容量是指存储器可以容纳的二进制信息量
网络带宽	Mbps	单位时间内能发送/接收的最大数据量
输入数据大小	GB	待处理业务的数据大小
输出数据大小	GB	输出结果的数据大小

### 6.2.4 终端服务类任务

面向终端服务类任务，以云手机、云电脑、云服务等虚拟化类任务为计量对象，资源使用计量，量纲包含：

——支持接入算力网监测/采集系统，应对计算资源、网络资源、内存资源、存储资源进行采集；

——支持对处理器数量、CPU数量、内存大小、硬盘存储空间大小、硬盘个数、网络延迟大小计量；

——可支持对用算时间区间（闲时/忙时）、用途进行计量。  
量纲计量见表9。

表 9：终端服务类量纲

计量项选择	单位	说明
计算资源	FLOPs	FLOPs = 核心数*每核心每周期 FLOP 数*频率(Hz)
网络资源	Mbps	单位时间内能发送/接收的最大数据量
存储资源	TB	虚拟化系统可以容纳的二进制信息量

## 7 算力计费

### 7.1 面向资源节点

#### 7.1.1 量级计费

按量计费模式是指根据实际使用的资源规格和使用时长按标准单价计费；

——计费对象为通算、智算、超算、网络、存储资源本身；

——计费方式为单价×实际资源用量×使用时长，资源用量包括vCPU核数，GPU数量，存储容量，网络流量等。

#### 7.1.2 周期计费

按周期计费模式是指按月、季度或年度周期预付资源费用，按资源规格进行计费；

——计费对象是通算、智算、超算、网络、存储资源的使用周期；

——计费方式为资源单价×资源配额×周期时长。

#### 7.1.3 卡包计费

卡包计费模式是指在一定价格内规定指定使用给定资源或资源组合的计费形式，包含算力卡、算力包等；

——计费对象是给定通算、智算、超算、网络、存储资源的使用权；

——计费方式为：单位时间的固定部分+卡时单价×卡时数量（卡数量×小时数），其中卡时为单位时间内的标准化资源组合单位。

### 7.2 面向任务节点

面向任务节点的服务以调用资源组合为特征，计费总体围绕调用进行，计费形式与计费单位在资源类服务上有所创新。

#### 7.2.1 量级计费

调用计费模式是指根据有效请求量的频率、数据处理复杂度及服务质量计费；

——计费对象为任务类服务的全部调用次数，包括模型类任务、渲染类任务、数据类任务、虚拟化任务等；

——计费方式为单价×有效请求量，其中，单价的影响因素包括单次调用消耗Token量、服务响应延迟、结构化数据输入输出总量等；

### 7.2.2 周期计费

周期计费模式是按照一定周期单位进行计费（如自然月、季度等），每个计费周期包含一定的额度；

——计费对象为任务类服务的调用周期，包括模型类任务、渲染类任务、数据类任务、虚拟化任务。超出额度的范围按照超额计费规则进行计费；

——计费方式为单位周期价格×周期数量。

### 7.2.3 进度计费

节点计费模式是指按任务进度进行计费；

——计费对象为任务类服务的全部调用节点进度，包括模型类任务、计算类任务、数据类任务、虚拟化任务；

——计费方式为资源单价×资源配额×任务进度。

### 7.2.4 卡时计费

卡时计费模式是指消耗一揽子标准化资源服务的计费形式；

——计费对象为任务类服务调用的标准化资源组合，包括模型类任务、计算类任务、数据类任务、虚拟化任务；

——计费方式为单位时间的卡时单价×卡时数量（卡数量×小时数），其中卡时为单位时间内的标准化资源组合单位。

## 8 功能管理

功能管理面向算力网络计量计费提供功能，包含下述8项功能。

### 8.1 采集接口

——支持对接外部系统，实现数据互通；

——支持同步结算数据，生成凭证或对接支付平台（如支付宝、银行接口）；

——提供 API 接口，支持监测调度各系统在权限范围内调用计量结算结果。

### 8.2 数据采集

——支持实时或周期性采集各类计量对象的数据；

——支持多种数据传输方式，如有线（RS485、以太网）、无线（4G/5G、LoRa、NB-IoT）等，兼容不同协议（Modbus、OPC UA、MQTT 等）。

### 8.3 数据校验

- 支持对采集数据进行合法校验，剔除异常值（如负值、跳变数据、重复数据、无主数据）；
- 支持补录缺失数据（基于历史趋势或算法估算）；
- 支持数据格式化转换；
- 支持多种数据采集方式（如接口、异步消息队列、文件等）。

### 8.4 计费规则配置

- 支持灵活配置计费策略，如：
  - 单一计价：无关用量等，按照单一价格进行计费；
  - 阶梯计费：分阶段设置不同单价；
  - 分时计费：按峰谷平时段设定差异化费率；
  - 峰值计费：按峰值或95峰值；
  - 复合计费：组合用量、时间、设备类型等多维度计费；
  - 资源包抵扣：设置资源包内可以抵扣的免费使用量；
  - 支持自定义公式，如损耗分摊、折扣计算、附加费（如税费、手续费）等。

### 8.5 统一记账

- 支持由运营方、资源提供方和资源使用方共同参与的协同计费管理模式；
- 支持通过实时采集资源使用数据，由系统自动生成临时账单并推送至各参与方进行确认，在各方达成一致后形成最终账单，进入分账结算环节。

### 8.6 历史追溯

- 支持版本管理和变更追溯。

### 8.7 权限管理

- 支持分配不同角色（管理员、操作员、用户）的系统访问权限；
- 支持用户自助查询（如历史结算记录、实时用量数据）。

### 8.8 可视分析

- 支持趋势分析：展示用量、费用随时间的变化趋势；
- 支持对比分析：不同用户、区域、设备间的横向对比；
- 支持异常预警：识别计量设备故障或异常用能（如突然激增 / 骤降）；
- 支持图表（折线图、柱状图、饼图）、Excel/CSV 文件导出。

附录 A  
(资料性)  
示例

A.1 通用计算案例

表 A.1 通用计算服务基础模板

规格示例	按核时	进度节点	按周期	按量
1台处理器，2CPU，4GB内存，0.25T存储空间，1块硬盘、网络延迟小于10ms	元/核时	元/50%任务进度	元/月	元/p
1台处理器，4CPU，8GB内存，0.5T存储空间，1块硬盘、网络延迟小于10ms	元/核时	元/50%任务进度	元/月	元/p
2台处理器，6CPU，48GB内存，1T存储空间，3块硬盘、网络延迟小于10ms	元/核时	元/50%任务进度	元/月	元/p
2台处理器，8CPU，256GB内存，4T存储空间，12块硬盘、网络延迟小于10ms	元/核时	元/50%任务进度	元/月	元/p
.....			元/月	元/p

A.2 智能计算示例

表 A.2 智能计算服务基础模板

规格示例	周期	卡时	按量	按节点
1台处理器，6CPU，48GB内存，1T存储空间，1块硬盘，M2050，1卡，网络延迟小于10ms，集群采用 rdma架构	元/月	元/卡时	元/p	元/50%任务进度
2台处理器，12CPU，512GB内存，4T存储空间，1块硬盘，P4，12卡，网络延迟小于10ms，集群采用 rdma架构	元/月	元/卡时	元/p	元/50%任务进度
2台处理器，8CPU，512GB内存，1.2T存储空间，1块硬盘，P100，4卡，网络延迟小于10ms，集群采用 rdma架构	元/月	元/卡时	元/p	元/50%任务进度
4台处理器，8CPU，40GB内存，4T存储空间，6块硬盘，V100，1卡，网络延迟小于10ms，集群采用 rdma架构	元/月	元/卡时	元/p	元/50%任务进度
8台处理器，16CPU，96GB内存，8T存储空间，10块硬盘，A100，1卡，网络延迟小于10ms，集群采用 rdma架构	元/月	元/卡时	元/p	元/50%任务进度
.....				

A.3 超级计算示例

表 A.3 超级计算服务基础模板

规格示例	周期	卡时	算力资源使用量	按节点
2台处理器，6CPU，48GB内存，M2050，1卡，网络延迟小于10ms，集群采用 rdma架构	元/月	元/卡时	元/p	元/50%任务进度
2台处理器，12CPU，512GB内存，V100，2卡，网络延迟小于10ms，集群采用 rdma架构	元/月	元/卡时	元/p	元/50%任务进度
.....				

A.4 网络示例

表 A.4 网络服务基础-入口模板

规格示例	按周期
入口流量 每月 0GB 到 10,000GB	元/GB
入口带宽 每月 10,000GB 到 50,000GB	元/月
入口带宽 每月 50,000GB 到 150,000GB	元/月
入口带宽 每月 超过 150,000GB	元/月
入口带宽 大块数据 (Bulk )	元/月
入口带宽 每月 5Mbps 到 10Mbps	元/月
入口带宽 每月 10 到 100	元/月
入口带宽 每月 100 到 1GB	元/月
1G之上	元/月
出口带宽 每月 0GB 到 10,000GB	元/月
出口带宽 每月 10,000GB 到 50,000GB	元/月
出口带宽 每月 50,000GB 到 150,000GB	元/月
出口带宽 每月 超过 150,000GB	元/月
出口带宽 大块数据 (Bulk )	元/月

公有IP地址：公有IP地址应按照购买的IP地址数量和使用的时间进行定价，单位：元/个/月。

A.5 存储示例

扩展的磁盘空间：扩展的磁盘空间按照容量和计量周期进行定价

表 A.5 存储服务基础模板

规格示例	按周期
1G-10G	元/GB/月
10G-100G	元/GB/月

表A.5 存储服务基础模板（续）

规格示例	按周期
100G-500G	元/GB/月
500G-1T	元/GB/月
1T之上	元/GB/月

A.6 算力卡示例

表 A.6 算力卡服务基础模板

规格示例	按周期	按卡时
20G网络流量，1Gbps下行峰值速率，150Mbps上行峰值速率，2T云存储空间，2000卡时	元/月	元/卡时
30G网络流量，1Gbps下行峰值速率，150Mbps上行峰值速率，2T云存储空间，2000卡时	元/月	元/卡时
50G网络流量，1Gbps下行峰值速率，150Mbps上行峰值速率，2T云存储空间，2000卡时	元/月	元/卡时
50G网络流量，1Gbps下行峰值速率，150Mbps上行峰值速率，4T云存储空间，5000卡时	元/月	元/卡时
50G网络流量，1Gbps下行峰值速率，150Mbps上行峰值速率，10T云存储空间，20000卡时	元/月	元/卡时
.....	元/月	元/卡时

A.7 模型类任务示例

A.7.1 大模型服务示例

表 A.7 大模型服务基础模板

规格示例	按token(输入1k tokens/ 输出 1000 tokens)	按周期	按API调用
GPU-3.5, 4K Token/1 API, 小于10ms	元/元	元/月	元/次
GPT-4 Turbo, 128K Token/10 API, 小于10ms	元/元	元/月	元/次
DeepSeek-V2, 128K Token/10 API, 小于10ms	元/元	元/月	元/次
.....			

A.7.2 自然语言处理示例

表 A.8 自然语言处理基础模板

规格示例	按周期	按量	按卡时（规定卡类型， 增减部分模板参数）
语言识别，10万	元/月	元/次	元/卡时

表A.8 自然语言处理基础模板（续）

规格示例	按周期	按量	按卡时（规定卡类型，增减部分模板参数）
语言识别，100万	元/年	元/次	元/卡时
语言处理，10万	元/年	元/次	元/卡时
语言处理，100万	元/年	元/次	元/卡时
语言生成，300万	元/年	元/次	元/卡时
文本翻译，10万	元/年	元/次	元/卡时
.....	元/月	元/千次	元/卡时

A.7.3 图像识别示例

表 A.9 图像识别服务基础模板

图像标签/媒资图像标签调用API次数，可用区	按周期	按量	按卡时（规定卡类型，增减部分模板参数）
1000次调用，北京	元/月	元/千次	元/卡时
100万次调用，北京	元/月	元/千次	元/卡时
300万次调用，北京	元/月	元/千次	元/卡时
100万次调用，北京	元/年	元/千次	元/卡时
300万次调用，上海	元/2年	元/千次	元/卡时
.....	元/月	元/千次	元/卡时

A.8 渲染类服务示例

A.8.1 云渲染示例

表 A.10 云渲染服务基础模板

规格示例	按周期	按量	按卡时（规定卡类型，增减部分模板参数）
4vCPU, 8GB内存, 2 TF SP / 30T INTGPU性能, 4G显存, 6Mbps 网络带宽	元/天	元/并发用户数/小时	元/卡时
4vCPU, 16GB内存, 4 TF SP / 30T INTGPU性能, 6G显存, 8Mbps 网络带宽	元/月	元/并发用户数/小时	元/卡时

表A. 10 云渲染服务基础模板（续）

规格示例	按周期	按量	按卡时 (规定卡类型, 增减部分模板参数)
10vCPU, 30GB内存, 16 TF SP / 30T INTGPU性能, 4G显存, 8Mbps 网络带宽	元/月	元/并发用户数 /小时	元/卡时
20vCPU, 60GB内存, 16 TF SP / 30T INTGPU性能, 12G显存, 10Mbps 网络带宽	元/月	元/并发用户数 /小时	元/卡时
32vCPU, 60GB内存, 16 TF SP / 30T INTGPU性能, 24G显存, 10Mbps 网络带宽	元/月	元/并发用户数 /小时	元/卡时
32vCPU, 60GB内存, 16 TF SP / 30T INTGPU性能, 24G显存, 10Mbps 网络带宽	元/月	元/并发用户数 /小时	元/卡时
.....	元/月	元/并发用户数 /小时	元/卡时

A. 9 数据类任务示例

A. 9.1 数据快递示例

表 A. 11 数据快递服务基础模板

规格示例	按周期	按量
硬盘传输, 500GB, 1天	元/天	元/块
硬盘传输, 2TB, 1天	元/月	元/块
高性能存储设备, 60TB, 1天	元/月	元/块
高性能存储设备, 120TB, 1天	元/月	元/块
高性能存储设备, 120TB, 2天	元/月	元/块
.....	元/月	元/块

A. 10 虚拟化任务示例

A. 10.1 云手机示例

表 A. 12 云手机服务基础模板

规格示例	按周期	按量	按卡时 (规定卡类型, 增减部分模板参数)
2vCPU, 4GB内存, 32GB云手机存储空间	元/月	元/小时	元/小时
4vCPU, 4GB内存, 32GB云手机存储空间	元/月	元/小时	元/小时

表A. 12 云手机服务基础模板（续）

规格示例	按周期	按量	按卡时 (规定卡类型, 增减部分模板参数)
8vCPU, 8GB内存, 32GB云手机存储空间	元/月	元/小时	元/小时
8vCPU, 8GB内存, 32GB云手机存储空间	元/月	元/小时	元/小时
8vCPU, 16GB内存, 32GB云手机存储空间	元/月	元/小时	元/小时
.....	元/月	元/小时	元/小时

## 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〕）
-