

# TC609

## 全国数据标准化技术委员会技术文件

TC609-6-2025-06

### 全国一体化算力网 算力多量纲计费 技术要求

National integrated computing power network—Technical requirements for  
computing power measurement and billing

2025-08-29 发布

2025-08-29 实施

全国数据标准化技术委员会 发布



# 目 次

前 言 .....	II
1 适用范围 .....	1
2 规范性引用文件 .....	1
3 术语和定义 .....	1
4 缩略语 .....	3
5 总体架构 .....	4
5.1 概述 .....	4
5.2 逻辑架构 .....	4
6 算力量纲 .....	5
6.1 面向资源节点 .....	5
6.2 面向任务节点 .....	9
7 算力计费 .....	11
7.1 面向资源节点 .....	11
7.2 面向任务节点 .....	11
8 功能管理 .....	12
8.1 采集接口 .....	12
8.2 权限管理 .....	12
8.3 计费规则配置 .....	12
8.4 数据采集 .....	13
8.5 统一记账 .....	13
8.6 数据校验 .....	13
8.7 可视分析 .....	13
8.8 数据存储 .....	14
8.9 历史追溯 .....	14
附 录 A（资料性）示例 .....	14
A.1 通用计算案例 .....	15
A.2 智能计算示例 .....	15
A.3 超级计算示例 .....	15
A.4 网络示例 .....	15
A.5 存储示例 .....	16
A.6 模型类任务示例 .....	16
A.7 渲染类服务示例 .....	17
A.8 数据类任务示例 .....	17
A.9 终端服务类任务示例 .....	18
参 考 文 献 .....	19

# 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：北京邮电大学、中国信息通信研究院、国家信息中心、中国移动通信有限公司研究院、中国电子技术标准化研究院、中移（苏州）软件技术有限公司、鹏城实验室、联通数字科技有限公司、中国联合通信网络有限公司、国家数据发展研究院、天津大学、天翼云科技有限公司、中国科学院计算技术研究所、京东科技信息技术有限公司、深圳市前海新型互联网交换中心有限公司、上海市新型互联网交换中心有限责任公司、阿里云计算有限公司、北京九章云极科技有限公司、中移（杭州）信息技术有限公司、腾讯云计算（北京）有限公司、杭州市新型互联网交换中心有限责任公司、中卫市新型互联网交换中心有限责任公司、中国农业银行股份有限公司、中电云计算技术有限公司、中科南京信息高铁研究院、紫金山实验室、江苏未来网络集团有限公司。

# 全国一体化算力网 算力多量纲计费技术要求

## 1 适用范围

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框架，规定了调度层算力服务，及与资源层、运营层、监测层进行交互的算力量纲、算力计费及功能管理基本技术要求。

本文件适用于国家级、区域级、城市级以及面向不同行业和企业开展全国一体化算力网算力服务交易、计费。

## 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

**算力** `computing power`

是指综合数据处理能力，从处理能力的分类可划分为通算算力、智算算力、超算算力、量子算力等。

注：综合处理能力包含硬件处理、软件处理及网络传输等综合能力。

### 3.2

**算力网** `computing power network`

是指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、服务智能按需。

### 3.3

**全国一体化算力网** `national integrated computing power network`

是指以信息网络技术为载体，促进全国范围内通用算力、智能算力、超级算力、量子算力等资源高比例、大规模一体化调度运营的数字基础设施。其核心目标是实现算力资源跨区域、跨行业高效协同，支撑数字经济高质量发展，服务国家“东数西算”战略布局。

### 3.4

**资源节点** `resource node`

也称算力资源，是指计算资源、存储资源以及节点内部网络资源等集合，通过该节点的管控系统/运营平台进行抽象并对外提供算力资源服务。

### 3.5

**任务节点** `task node`

是指接收用户任务请求并触发资源调用的逻辑接入点，通过该节点的管控系统/运营平台进行交互并对外提供算力任务接入服务。

### 3.6

**算力服务** `computing power service`

是指基于算力资源提供的可满足用户需求的计算、数据处理以及出具存储与传输等有价值的服务活动，包括但不限于算力资源本身、基于算力的特定工具以及具体计算和数据处理等行为。

### 3.7

#### **资源类服务 resource-based service**

是指一种算力服务模式，能将算力资源以一种可量化的资源形式提供给用户，包括计算资源、网络资源、存储资源等。

### 3.8

#### **任务类服务 task-based service**

是指一种算力服务模式，以完成用户业务需求为目标，按任务完成情况提供并计量计算、网络、存储等算力资源，包括大模型服务、车联网训推一体化、视频分化渲染和数据快递等。

### 3.9

#### **通用计算 general computing**

也称通算，是指由CPU（中央处理器）为主提供的计算能力，主要用于通用计算任务。

### 3.10

#### **智能计算 intelligent computing**

也称智算，是指由GPU（图形处理器）、FPGA（现场可编程门阵列）、ASIC（专用集成电路）等芯片为主提供的计算能力，主要用于人工智能相关的计算任务。

### 3.11

#### **超级计算 super computing**

也称超算，是指基于大规模并行架构的超高性能计算能力，主要用于科学计算、工程仿真、气象预测等领域的超大规模、高复杂度计算任务。

### 3.12

#### **网络 network**

是指由一系列基础设施、协议和服务组成的系统，用于连接计算节点（如服务器、虚拟机、容器、终端设备等），并实现它们之间数据（包括指令、请求、结果、文件等）的高效、可靠和安全传输。

### 3.13

#### **存储 storage**

是指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。

### 3.14

#### **模型类任务 model-related task**

是指依托算力资源开展的，人工智能模型构建、优化、运用等相关的一系列计算任务类型。运用深度学习、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、智能交互等服务。

### 3.15

#### **渲染类任务 rendering-related task**

是指依托算力资源开展的，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。

### 3.16

#### **数据类任务 data-related task**

是指依托算力资源开展的数据类操作任务类型，包含数据处理、数据分析、数据管理、数据应用、数据传输、数据存储等。

### 3.17

### **终端服务类任务 terminal-service-related task**

是指依托算力资源开展的，通过网络连接、云端虚拟化等形式，为终端用户提供对部署在云端的计算机系统、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。

#### **3.18**

### **核时 duration of core occupancy**

是衡量算力服务在单位时间内资源消耗量的标准单位，为给定环境下，处理器核心数和使用时长的乘积。

#### **3.19**

### **卡时 duration of computing card**

是衡量算力服务在单位时间内资源消耗量的标准单位，为给定环境下，计算任务占用计算卡的数量和占用时长的乘积。

#### **3.20**

### **标记单元 token**

是指一种算力服务的衡量单位，信息处理的最小单元。

#### **3.21**

### **有效请求量 billable request volume**

是指在任务类服务的调用中，客户端向服务端发起并成功到达服务端，且服务端能够识别、处理并最终向客户端返回了明确状态码（通常是HTTP状态码）响应的请求次数。

#### **3.22**

### **云端 cloud**

是指用户通过互联网远程访问、按需使用的共享式、可扩展的计算资源池（包括服务器、存储、网络等）。

#### **3.23**

### **卡包 card package**

是指一种预付费的计费形式，用户在特定周期内可消费规定量的指定算力资源或组合。

#### **3.24**

### **异构计算 heterogeneous computing**

是指基于CPU、GPU、NPU等不同架构计算单元协同工作的计算模式。

## **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）

vCPU：虚拟中央处理器（Virtual Central Processing Unit）

GPU：图像处理器（Graphics Processing Unit）

FPGA：现场可编程门阵列（Field-Programmable Gate Array）

ASIC：专用集成电路（Application-Specific Integrated Circuit）

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）

MB：兆字节（Megabyte）

GB: 吉字节 (Gigabyte)  
TB: 太字节 (Terabyte)  
BGP: 边界网关协议 (Border Gateway Protocol)  
Mbps: 兆比特每秒 (Megabits per second)  
FLOPs: 每秒浮点运算次数 (Floating-Point Operations Per Second)  
Mbps: 兆比特每秒 (Megabits per Second)  
IOPS: 每秒输入输出操作次数 (Input/Output Operations Per Second)  
ps: 图像像素 (pixel)  
FPS: 每秒帧数 (Frames Per Second)  
SLA: 服务等级协议 (Service Level Agreement)  
RESTful: 一种基于HTTP协议的API设计风格 (Representational State Transfer)  
JSON: 轻量级数据交换格式 (JavaScript Object Notation)  
XML: 可扩展标记语言 (eXtensible Markup Language)  
CSV: 逗号分隔值文件 (Comma-Separated Values)

## 5 总体架构

### 5.1 概述

算力多量纲计费应面向用户侧，通过提取关键计量要素，综合反映算力资源使用情况，实现计费过程的标准化。该过程旨在解决因量纲不一带来的算力一体化难题，以标准化运行机制支撑全国一体化算力网体系化建设。

### 5.2 逻辑架构

本标准是算力网调度层的重要组成部分，提供国家级、区域级、城市级算力服务的统一量纲、计费及功能要求。该计费体系向下对接算力网资源层的资源输入，向上支撑算力网运营层的服务输出，同时同监测层数据同步，形成四维统一的计费体系，以保障全国一体化算力网高效运行和服务有序流通。

随着数字中国建设进入新发展阶段，应用侧对算力服务提出了新要求，算力服务呈现出从资源导向向任务导向转变的新特征。多量纲计费体系基于资源供给侧和用户需求侧，划分为面向资源节点和面向任务节点的计量计费，分别对应资源供给与任务执行场景下的计费需求。按照计量计费的关键要素，全国一体化算力网 算力多量纲计费技术要求逻辑架构分为算力量纲、算力计费与功能管理三个部分，通过量纲与计费机制的组合运用，实现多维度、可扩展的计费模式。见图1，其中：

——算力量纲，提供标准化通用计量项。包含面向资源节点的计量规范与面向任务的计量规范。

其中，面向资源节点的计量规范涵盖通用计算资源、智能计算资源、超级计算资源、网络资源和存储资源等资源类服务；面向任务节点的计量规范涵盖模型类任务、渲染类任务、数据类任务、终端服务类任务等现阶段典型任务类服务，并预留后续任务类型扩展空间，以支撑多样化服务场景。算力量纲的计量基础依赖于全国统一的算力标识体系，确保资源属性描述的一致性和可得性；

——算力计费，提供多维度通用计费模式。分为面向资源节点的计费和面向任务节点的计费。其中，面向资源节点的计费支持量级维度、周期维度、卡时维度等；面向任务节点的计费支持量级维度、周期维度、卡包维度等。量级维度用于反映资源使用规模的差异，支持按需、随用随计的方式进行计费；周期维度用于反映资源使用时间的差异，支持按月、按年等周期进



行计费；卡时维度用于反映任务执行时长的差异，支持按小时等精细粒度进行计费；卡包维度用于反映预定义资源组合的差异，支持以套餐或包的形式进行计费；

——功能管理，提供计费全生命周期基本功能，保障计费过程准确可信。对算力多量纲计费全生命周期内的基础功能项提出参考要求，主要包含采集接口、权限管理、计费规则配置、数据采集、统一记账、数据校验、可视分析、数据存储、历史追溯功能模块，确保计费过程的完整性与可信性。

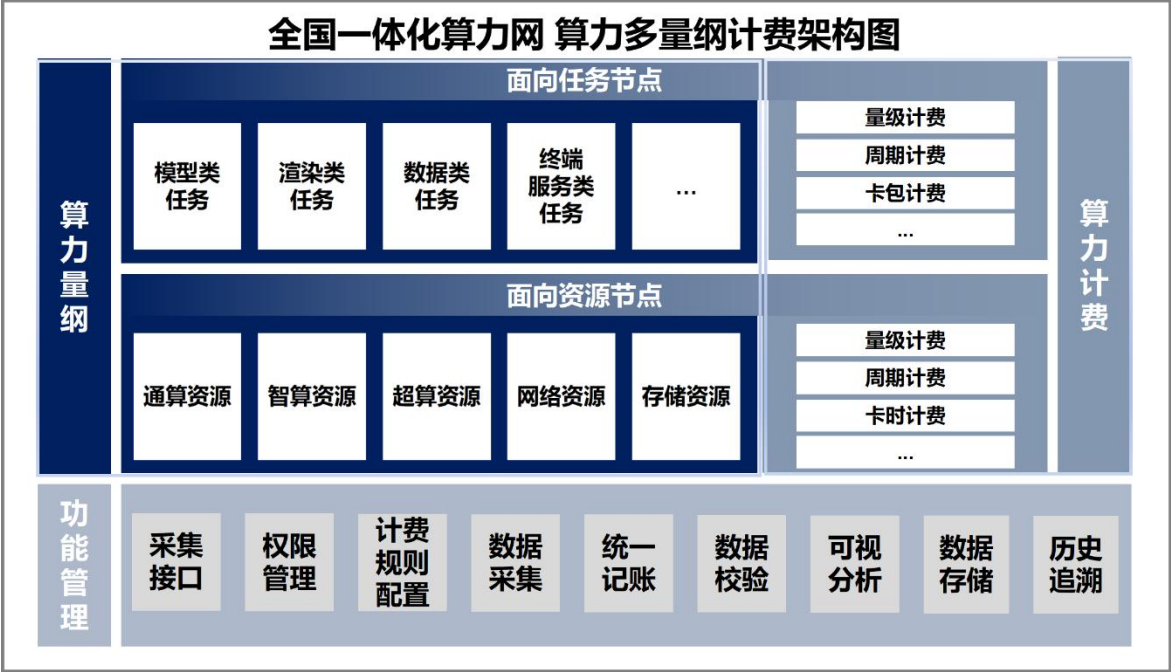


图 1 全国一体化算力网 算力多量纲计费参考框架

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、网络资源和存储资源进行采集；
- 支持对 CPU 型号、vCPU 核数、CPU 用量、内存用量、系统盘容量、公网带宽和内网带宽等进行计量；
- 可支持对数据盘容量、计算资源占用时段（闲时/忙时）和空闲资源占比等进行计量；
- 可支持按量级、周期和卡时等维度进行计费。

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
CPU 型号	/	指使用的 CPU 型号，用于反映 GPU 硬件的性能规格及适用计算场景
vCPU 核数	核	指分配的虚拟 CPU 核心数量，用于反映计算资源的配置规模
vCPU 用量	核时	核时 = vCPU 核数 × 小时数，用于反映 CPU 资源实际使用情况
内存用量	GB	指分配的内存容量大小，用于反映内存资源的配置规模
系统盘容量	GB	指分配的系统盘存储空间大小，用于反映持久化存储资源的配置规模
公网带宽	Mbps	指分配的公网出口带宽能力，用于反映数据外部传输能力
内网带宽	Mbps	指分配的内网带宽能力，用于反映数据中心或集群内部的网络传输能力

### 6.1.2 智能计算

通过统计智能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、网络资源和存储资源进行采集；
- 支持对 CPU 型号、vCPU 核数、内存用量、GPU 用量、GPU 型号、AI 加速卡型号/ASIC(如 NPU、TPU、DPU 等)、系统盘容量、公网带宽和内网带宽等进行计量；
- 可支持对数据盘容量、计算资源占用时段（闲时/忙时）和空闲资源占比等进行计量；
- 可支持按量级、周期和卡时等维度进行计费。

量纲计量见表2。

表 2 智能计算量纲

计量项选择	单位	说明
vCPU 核数	核	指分配的虚拟 CPU 核心数量，用于反映计算资源的配置规模
内存用量	GB	指分配的内存容量大小，用于反映内存资源的配置规模
GPU 用量	卡时	卡时 = GPU 卡数 × 小时数，用于反映 GPU 资源实际使用情况
GPU 型号	/	指使用的 GPU 型号，用于反映 GPU 硬件的性能规格及适用计算场景
AI 加速卡型号	/	指使用的 AI 加速卡的型号，用于标识其计算性能等级、架构类型及适用的人工智能计算场景
系统盘容量	GB	指分配的系统盘存储空间大小，用于反映持久化存储资源的配置规模
公网带宽	Mbps	指分配的公网出口带宽能力，用于反映数据外部传输能力

内网带宽	Mbps	指分配的内网带宽能力,用于反映数据中心或集群内部的网络传输能力
------	------	---------------------------------

### 6.1.3 超级计算

通过统计超级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、网络资源和存储资源进行采集;
- 支持对 CPU 型号、vCPU 核数、vCPU 用量、内存用量、GPU 用量、GPU 型号和系统盘容量等进行计量;
- 可支持对数据盘容量、计算资源占用时段(闲时/忙时)和空闲资源占比等进行计量;
- 可支持按量级、周期和卡时等维度进行计费。

量纲计量见表3。

表 3 超级计算量纲

计量项选择	单位	说明
CPU 型号	/	指使用的 CPU 型号,用于反映 GPU 硬件的性能规格及适用计算场景
vCPU 核数	核	指分配的虚拟 CPU 核心数量,用于反映计算资源的配置规模
vCPU 使用量	核时	核时 = vCPU 核数 × 小时数,用于反映 vCPU 资源实际使用情况
GPU 使用量	卡时	卡时 = GPU 卡数 × 小时数,用于反映 GPU 资源实际使用情况
GPU 型号	/	用于反映 GPU 硬件的性能规格及适用计算场景
内存用量	GB	指分配的内存容量大小,用于反映内存资源的配置规模
系统盘容量	GB	指分配的系统盘存储空间大小,用于反映持久化存储资源的配置规模
公网带宽	Mbps	指分配的公网出口带宽能力,用于反映数据外部传输能力
内网带宽	Mbps	指分配的内网带宽能力,用于反映数据中心或集群内部的网络传输能力

### 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;
- 支持对公网 IP、带宽类型、公网带宽、公网上行流量、公网下行流量时延、丢包率和抖动等进行计量;
- 可支持网络可靠性、安全加密和网络扩展等增值服务进行计量;
- 可支持按量级和周期等维度进行计费。

量纲计量见表4。

表 4 网络量纲

计量项选择	单位	说明
公网 IP	个	用于标识和访问外部网络的唯一地址
带宽类型	/	公网出口带宽的接入方式类型（BGP、普通专线等）
公网带宽	Mbps	公网出口的最大传输速率
公网上行流量	GB	从本地网络向公网发送的数据量
公网下行流量	GB	从公网接收至本地网络的数据量
时延	ms	数据在网络中传输的往返时间，影响实时处理性能
丢包率	%	数据在传输过程中丢失的比例，影响服务质量和数据完整性
抖动	ms	网络时延的波动幅度，影响语音/视频等实时服务体验

### 6.1.5 存储

存储资源按类型可分为块存储、文件存储和对象存储。

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，对存储资源进行采集；
- 支持对块存储的存储类型、存储容量、最大 IOPS 和数据传输量等进行计量；
- 支持对文件存储的存储类型、存储容量、最大 IOPS 和数据传输量等进行计量；
- 支持对对象存储的存储类型、存储容量、流量和请求操作等进行计量；
- 可支持对快照服务、管理功能、数据处理和数据取回等增值服务进行计量；
- 可支持按量级和周期等维度进行计费。

量纲计量见表5。

表 5 存储量纲

计量项选择	单位	说明
存储类型	/	指为满足不同数据可靠性、可用性与访问频率需求而划分的数据存储形式，主要包括标准存储、低频存储、归档存储等。
IOPS	次/秒	每秒输入输出操作次数，用于反映存储设备处理请求的能力
数据传输量	GB/s	指单位时间内数据传输的总量，用于反映存储或计算系统的传输能力
存储容量	GB	用户数据所占用的存储空间
流量	GB	数据传输实际产生的流量，主要包括流入流量、流出流量、回源流量和跨区域复制流量

请求操作	万次	用户对存储资源的所有访问和管理操作，主要包括读/写请求等
------	----	------------------------------

## 6.2 面向任务节点

### 6.2.1 模型类任务

面向模型类任务，以大模型、自然语言处理、图像视频识别等典型模型类任务为计量对象，量纲包含：

- 支持接入算力网监测/采集系统；
- 支持对模型参数量级、最大上下文长度、Token 输入/输出量和多模态类型（文本、图像、音频等）等进行计量；
- 可支持对并发上限等增值服务进行计量；
- 可支持按量级、周期和卡包等维度进行计费。

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
参数量	B	模型可训练参数的总规模
输入/出量	Token	单次请求中输入/输出的 Token 总量
上下文长度	Token	单次请求中模型可处理的 Token 上限
多模态类型	/	支持的模态类型（如文本、图像、音频等），用于识别多模态任务能力
并发上限	个	同时处理的会话数量

### 6.2.2 渲染类任务

面向渲染类任务，以云渲染等典型渲染类任务为计量对象，量纲包含：

- 支持接入算力网监测/采集系统；
- 支持 vCPU 核数、vCPU 用量、内存用量、GPU 用量、带宽、并发上限、磁盘容量、渲染分辨率和渲染帧数等进行计量；
- 可支持按量级、周期和卡包等维度进行计费。

量纲计量见表7。

表 7 渲染类量纲

计量项选择	单位	说明
-------	----	----

vCPU 核数	核	指分配的虚拟 CPU 核心数量，用于反映计算资源的配置规模
vCPU 用量	核时	核时 = vCPU 核数 × 小时数，用于反映 CPU 资源实际使用情况
内存用量	GB	指分配的内存容量大小，用于反映内存资源的配置规模
GPU 用量	卡时	卡时 = GPU 卡数 × 小时数，用于反映 GPU 资源实际使用情况
磁盘容量	GB	用户数据所占用的磁盘空间
带宽	Mbps	数据传输速率，用于反映服务提供方与用户之间网络连接传输数据的快慢
并发上限	个	同时处理的会话数量
渲染分辨率	px	分辨率 = 横向像素数 × 纵向像素数
渲染帧数	FPS	每秒渲染的图像帧数

### 6.2.3 数据类任务

面向数据类任务，以数据快递等典型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；
- 支持对链路规格、数据迁移量、增量数据同步量、跨地域流量和时延等进行计量；
- 可支持按量级、周期和卡包等维度进行计费。

量纲计量见表8。

表 8 数据类量纲

计量项选择	单位	说明
迁移链路规格	/	根据链路配置的性能等级（如：small、medium、large）收费
数据迁移量	GB	全量迁移阶段实际传输的数据总量
增量数据同步量	GB/小时	增量同步阶段按小时统计传输的数据量
跨地域流量	GB	数据从源地域传输到目标地域的总量
时延	ms	数据在网络中传输的往返时间，影响实时处理性能

### 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、云主机等典型终端服务类任务为计量对象，量纲包含：

- 支持接入算力网监测/采集系统；

- 支持对计算资源和存储资源等进行计量；
  - 可支持对网络带宽等增值服务进行计量；
  - 可支持按量级、周期和卡包等维度进行计费。
- 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	/	包括 vCPU、内存、GPU 等
存储资源	/	包括系统盘和数据盘等
网络带宽	Mbps	网络出口的最大传输速率

7 算力计费

7.1 面向资源节点

7.1.1 量级计费

- 量级计费模式是指根据资源实际使用情况，按资源用量进行计费；
- 适用资源类型：计算资源、网络资源、存储资源等基础资源；
  - 计费方式：资源规格单价 × 计费时长。按秒计算计费时长，使用时长不足最低阈值时按最短时长计算；资源规格包括 vCPU 核数、GPU 卡数、存储容量、网络带宽等。

7.1.2 周期计费

- 周期计费模式是指按月、季或年度周期，预付资源使用费用，按资源用量进行计费；
- 适用资源类型：计算资源、网络资源、存储资源等周期性资源使用场景；
  - 计费方式：资源规格单价 × 购买时长。

7.1.3 卡时计费

- 卡时计费模式是指消耗一系列标准化资源服务的计费形式；
- 适用资源类型：算力资源类型；
  - 计费方式为：卡时单价 × 卡数量 × 小时数。

7.2 面向任务节点

面向任务节点的服务以任务对资源的组合占用为特征，计费形式与资源节点一致，但计费单位更加精细化。

7.2.1 量级计费

量级计费模式是指依据任务在计费周期内的实际使用量进行计费，该使用量可按调用次数或Token数量等方式进行计费：

- 适用任务类型：适用于模型类任务、数据处理类任务等按调用次数或Token数量计费的服务场景；
- 调用次数计费方式：调用次数单价  $\times$  有效请求量；
- Token 计费方式：输入Token 单价  $\times$  输入Token 数 + 输出Token 单价  $\times$  输出Token 数。

### 7.2.2 周期计费

周期计费模式是指按照预设的固定周期（如月、季、年度等）收取费用，每个计费周期包含约定的资源使用额度：

- 适用任务类型：适用于模型类任务、渲染类任务、数据处理类任务、虚拟化任务等需要周期性提供和保障资源的服务场景；
- 计费方式：套餐价格  $\times$  周期数 + 超额单价  $\times$  超额用量。

### 7.2.3 卡包计费

卡包计费模式是指根据标准化资源服务的实际消耗量，按卡时、调用次数或Token数量进行计费：

- 适用任务类型：适用于模型类任务、渲染类任务、数据处理类任务、终端服务类任务等服务场景；
- 计费方式：卡包单价  $\times$  购买时长。

## 8 功能管理

功能管理面向算力网多量纲计费提供功能，包含下述主要功能。

### 8.1 采集接口

- 支持接口安全认证；
- 支持对接外部系统，实现数据连接；
- 支持接口访问频率限制和流量控制；
- 支持同步结算数据，生成凭证和对接支付平台（如第三方支付、银行接口）；
- 支持双向身份认证（如 TLS/SSL 加密）及数据签名机制，确保数据传输安全；
- 提供 RESTful API 接口，支持 JSON/XML 数据格式；
- 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 权限管理

- 支持角色管理；
- 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）。

### 8.3 计费规则配置



- 支持计费单位标准化定义；
- 支持灵活配置计费策略，如：
  - 单一计价：按照固定价格进行计费，不随用量变化；
  - 阶梯计费：依据购买时长分阶段设置差异化单价；
  - 分时计费：按照峰、谷、平等时段设定不同费率，其中，峰时段指资源需求高、利用率集中的时段；谷时段指资源需求低、利用率相对空闲的时段；平时段则介于峰、谷之间，资源利用率保持常态；
  - 峰值计费：根据峰值或 95 分位峰值进行计费；
  - 复合计费：基于用量、时间、设备类型等多维度进行组合计费；
  - 资源包抵扣：通过预先配置的资源包额度抵扣相应使用量；
  - SLA 计费规则：支持基于服务等级协议（SLA）的计费机制；
  - 支持预付折扣、满减规则、资源包叠加抵扣等商业策略；
  - 支持基于地理位置的差异化费率配置；
- 自定义公式计费：当既有计费方式无法满足需求时，可通过输入数学表达式实现灵活计费，支持折扣计算及附加费用（如税费、手续费）等场景。自定义公式计费示例：基础定义费用 ± 自定义费用（如附加费用、折扣费用等）。

#### 8.4 数据采集

- 支持实时或周期性采集各类计量对象的数据；
- 支持断点续传和数据压缩机制；
- 支持高并发设备数据实时采集与存储；
- 支持多种数据传输规范，如有线、无线等，兼容不同协议。

#### 8.5 统一记账

- 支持由运营方、资源提供方和资源使用方三方共同参与的协同计费管理模式；
- 支持实时采集资源用量数据，自动生成临时账单并同步推送至三方进行核对，在各方达成一致后形成最终账单，进入分账结算环节；
- 临时账单应明确规定使用期限，如需在 24 小时内各方确认，否则形成最终账单；
- 支持账单争议解决模式，当任何一方对临时账单提出异议时，系统应支持争议提报、多方协商及仲裁确认流程，并根据最终确认结果形成最终账单。

#### 8.6 数据校验

- 支持数据采集源头（设备/系统）的元信息绑定，确保可追溯；
- 支持对采集数据进行合法校验，剔除异常值；
- 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）；
- 支持校验规则配置与更新。

#### 8.7 可视分析

- 支持费用趋势分析：用于展示用量与费用随时间变化的趋势；
- 支持成本构成分析；
- 支持对比分析：用于对不同用户、区域、设备等对象进行横向对比；
- 支持费用异常提示：用于识别计量设备故障或用量异常（如激增 / 骤降）；
- 支持图表（折线图、柱状图、饼图）、Excel/CSV 文件导出；
- 支持实时仪表盘展示算力资源使用热力图与计费趋势。

## 8.8 数据存储

- 支持数据安全：计费数据存储需符合 GB/T 35273《个人信息安全规范》加密要求；
- 支持审计合规：系统需满足等保三级日志留存要求，支持监管接口对接；
- 支持完整性保护：对计费数据实施实时校验，发现篡改或损坏立即告警并启动自修复；
- 支持数据追溯：为每条计费数据自动生成全局唯一标识并记录链路级信息，确保“数据一规则一账单”全链路可追溯。

## 8.9 历史追溯

- 支持日志管理；
- 支持版本管理和变更追溯；
- 支持审计功能，关键操作（如规则修改、账单调整）需留存完整操作日志；
- 支持历史结算记录自动导出。

## 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	按量	按周期	按卡时
1台CPU处理器，2核vCPU，2GB内存，1.5Gbps内网带宽，20GB系统盘，20GB数据盘(可选)，100Mbps公网带宽	元/秒	元/月	元/卡时
1台CPU处理器，2核vCPU，4GB内存，1.5Gbps内网带宽，20GB系统盘，20GB数据盘(可选)，100Mbps公网带宽	元/秒	元/月	元/卡时
1台CPU处理器，4核vCPU，8GB内存，2Gbps内网带宽，20GB系统盘，20GB数据盘(可选)，400Mbps公网带宽	元/秒	元/月	元/卡时
.....			

## 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	按量	按周期	按卡时
1台GPU处理器，2核vCPU，2GB内存，1.5Gbps内网带宽，20GB系统盘，20GB数据盘(可选)，100Mbps公网带宽	元/秒	元/月	元/卡时
1台GPU处理器，2核vCPU，4GB内存，1.5Gbps内网带宽，20GB系统盘，20GB数据盘(可选)，100Mbps公网带宽	元/秒	元/月	元/卡时
1台NPU处理器，28核vCPU，116GB内存，13Gbps内网带宽，200GB系统盘，200GB数据盘(可选)，400Mbps公网带宽	元/秒	元/月	元/卡时
.....			

## 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	按量	按周期	按卡时
1台GPU处理器，96核vCPU，384GB内存，25Gbps内网带宽，100GB系统盘，100GB数据盘(可选)，400Mbps公网带宽	元/秒	元/月	元/卡时
1台CPU处理器，4核CPU，8GB内存，5Gbps内网带宽，50GB系统盘，50GB数据盘(可选)，100Mbps公网带宽	元/秒	元/月	元/卡时
.....			

## A.4 网络示例

表 A.4 网络服务基础模板

规格示例	按量	按周期
入口带宽	元/GB	元/月
出口带宽	元/GB	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	按量	按卡包
标准存储	元/GB/月	元/GB
低频存储	元/GB/月	元/GB
归档存储	元/GB/月	元/GB
.....		

## A.6 模型类任务示例

### A.6.1 大模型服务示例

规格示例	按Token	按调用次数	按周期
Qwen-plus 最大上下文长度1,000,000Token	元/Token	元/万次	元/月
DeepSeek-V2 最大上下文长度1,000,000Token	元/Token	元/万次	元/月
Hunyuan-T1 最大上下文长度1,000,000Token	元/Token	元/万次	元/月
.....			

表 A.6 大模型服务基础模板

### A.6.2 自然语言处理示例

表 A.7 自然语言处理基础模板

规格示例	按周期	按调用次数
语言识别	元/年	元/万次
语言处理	元/年	元/万次
语言生成	元/年	元/万次
文本翻译	元/年	元/万次
.....		

### A.6.3 图像识别示例

表 A.8 图像识别服务基础模板

规格示例	按周期	按调用次数
图像识别	元/年	元/千次
图像搜索	元/年	元/千次
图像标签	元/年	元/千次
.....		

### A.7 渲染类服务示例

#### A.7.1 云渲染示例

表 A.9 云渲染服务基础模板

规格示例	按周期	按量	按卡时
1台GPU处理器，4核vCPU，8GB内存，4G显存，6Mbps网络带宽	元/月	元/并发用户数/秒	元/卡时
1台GPU处理器，8核vCPU，16GB内存，6G显存，6Mbps网络带宽	元/月	元/并发用户数/秒	元/卡时
1台CPU处理器，8核vCPU，16GB内存，8Mbps网络带宽	元/月	元/并发用户数/秒	元/卡时
.....			

### A.8 数据类任务示例

#### A.8.1 数据快递示例

表A.10 数据快递服务基础模板

规格示例	按周期	按量
------	-----	----

数据迁移	元/月	元/GB
数据同步	元/月	元/GB
.....		

## A.9 终端服务类任务示例

### A.9.1 云电脑示例

表 A.11 云电脑服务基础模板

规格示例	按周期	按量
1台CPU处理器，2vCPU，4GB内存，32GB云硬盘	元/月	元/小时/台
1台GPU处理器，16vCPU，64GB内存，32GB云硬盘	元/月	元/小时/台
2台GPU处理器，22vCPU，80GB内存，32GB云硬盘	元/月	元/小时/台
.....		

## 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》标准草案
-