

算网大脑驱动的泛在算力调度架构设计与实践验证

Design and Practical Verification of a Ubiquitous Computing Power Scheduling Architecture Driven by Computing Network Brain

刘运奇¹, 邓玲¹, 冯铭能², 韦秀林²(1. 中国联通广东分公司, 广东 广州 510627; 2. 中讯邮电咨询设计院有限公司广东分公司, 广东 广州 510627)

Liu Yunqi¹, Deng Ling¹, Feng Mingneng², Wei Xiulin²(1. China Unicom Guangdong Branch, Guangzhou 510627, China; 2. China Information Technology Designing & Consulting Institute Co., Ltd. Guangdong Branch, Guangzhou 510627, China)

摘要:

随着算力需求的日益增长, 如何实现高效、灵活且智能的算力调度成为当前研究的热点。通过聚焦算网泛在调度核心概念, 探讨其在泛在算力调度中的应用, 提出一种创新的架构设计方案, 全面提升算力资源利用率和服务质量。

关键词:

算网大脑; 算力调度; 智能编排和调度; 计量中心

doi: 10.12045/j.issn.1007-3043.2026.03.005

文章编号: 1007-3043(2026)03-0024-05

中图分类号: TN915

文献标识码: A

开放科学(资源服务)标识码(OSID):



Abstract:

With the increasing demand for computing power, how to achieve efficient, flexible, and intelligent computing power scheduling has become a hot research topic. It focuses on the core concept of ubiquitous computing and network scheduling, explores the application of ubiquitous computing power scheduling, and proposes an innovative architecture design scheme to comprehensively improve the utilization rate of computing resources and service quality.

Keywords:

Computing network brain; Computing power scheduling; Intelligent orchestration and scheduling; Metering hub

引用格式: 刘运奇, 邓玲, 冯铭能, 等. 算网大脑驱动的泛在算力调度架构设计与实践验证[J]. 邮电设计技术, 2026(3): 24-28.

1 概述

为应对我国数字经济快速发展带来的海量算力需求, 我国启动了“东数西算”工程, 通过构建全国一体化算力网络, 将东部密集的算力需求有序引导至西部, 发挥西部在能源、土地等方面的优势, 促进绿色低碳发展, 优化全国算力资源配置。

在“东数西算”推进过程中仍面临带宽不足、网络时延高、算力调度能力弱等关键技术挑战, 亟需加强顶层设计, 统一标准体系, 建设大带宽、低时延、智能

高效的算力网络, 实现跨区域算力资源的协同调度与高效利用。

1.1 某省联通算网大脑现状分析

某省联通已经建设了一套算网大脑系统, 将中心算力、边缘算力和网络资源进行统一管理和调度, 通过智能化的方式提供算网一体化的服务能力, 为用户提供便捷、高效、可扩展的算力资源。

泛在计算时代对计算能力和网络资源提出了更高的要求。传统的静态资源分配模式已难以满足动态变化的需求, 目前系统存在以下问题。

a) 资源异构调度能力不足。计算资源呈现多架构(CANN、CUDA等)、跨地域分布特征, 在整合不同城

收稿日期: 2026-02-03

市数据中心的异构资源搭建跨区域 AI 训练平台时,需打破资源类型与地域隔阂,实现统一管理协同。

b) 网络结构复杂。现网存在带宽波动、时延不稳定、丢包等问题,给算力调度中数据的高效稳定传输带来挑战,需重点考量网络因素对调度效果的影响。

c) 调度任务复杂多样。计算任务涵盖批处理、实时处理、交互式任务等类型,不同任务对算力与网络需求差异显著,需精准匹配多样化任务需求,保障调度适配性。

因此需要对算网大脑进行重构,从根本上改变行业传统做法,实时感知网络状态,动态调整资源分配策略,从而优化整体性能。

1.2 建设需求分析

在数字化时代,构建高效、灵活、可靠的泛在算力调度至关重要。

a) 高效性。面对海量数据与高并发任务,算力调度需实现毫秒级响应与资源分配,提升任务处理效率与系统吞吐能力,保障关键业务高峰期的稳定运行,最大化资源利用效能,支撑大规模实时计算需求。

b) 灵活性。调度系统需具备动态感知任务特征与资源状态的能力,支持多策略自适应调度,灵活适配批处理、实时计算、AI训练等多样化负载,实现资源按需分配与弹性伸缩。

c) 可靠性。在关键业务场景中,需保障算力服务的高可用性与容错能力,支持故障快速恢复与任务迁移,确保计算过程连续稳定,满足业务连续性与服务质量的严格要求。

根据以上分析及业务需求,有必要对现有系统架构进行升级和迭代开发。

2 算网大脑驱动的泛在算力调度架构

2.1 整体架构

如图1所示,某省联通新型算网大脑采用感知、决策与执行三层协同的泛在调度架构,实现对算力与网络资源的端到端智能管控^[1]。它集成算网雷达的资源发现机制与泛在调度中心的分布式任务分配算法,实现跨算力和网络的统一业务编排服务。该算网大脑向上提供算网业务服务能力,向下协同光网/算力管控系统,支持端到端业务编排、智能决策、跨域资源分配和统一协同调度等功能。

2.2 分层分域管理的算网架构

整体架构围绕资源融合、智能调度与开放服务三大核心理念进行设计,支撑多场景、多租户的算力需求,推动算力资源的集约化运营与高效利用。

通过按地理或管理边界划分区域单元,构建以省、市为核心的分布式算网大脑,结合专用高速网络连接,形成以集中决策为核心、分布式执行为支撑的协同控制机制。该控制机制能够实现算力资源的分域管理、属地化运营与全域高效协同,打造一体化分布式算网体系^[2]。

2.3 算力评估标准体系

2.3.1 评估指标

围绕算力因子、成本因素、服务指标3类核心指标构建灵活、可扩展的算力评价体系。

a) 算力因子是衡量资源性能的核心技术指标,包括以 FLOPS 为单位的计算能力、存储容量与读写速度、网络带宽与传输延迟。这些指标反映算力供给的技术水平与响应能力。

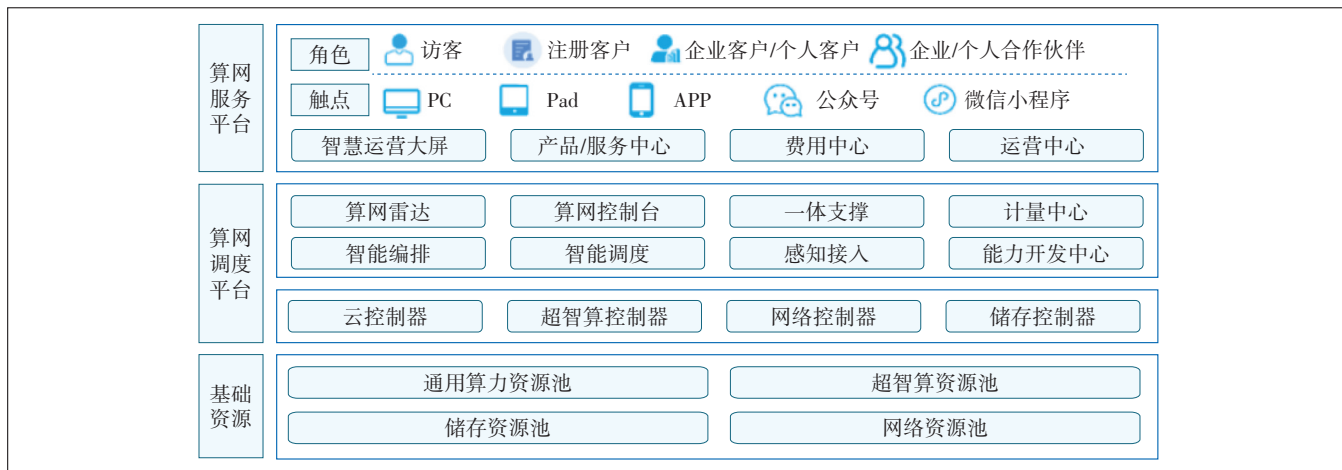


图1 算网服务及调度平台详细架构

b) 成本因素涵盖设备采购、租赁、电力消耗及运维管理等全生命周期成本,可用于评估算力使用的经济性与可持续性,为性价比最优的资源配置决策提供支持^[3]。

c) 服务指标聚焦业务保障能力,包括安全防护、容灾备份和服务质量,以确保服务稳定可靠。

2.3.2 评估模型

该评估模型采用分阶段评估与场景化适配相结合的设计思路,根据任务复杂度和应用环境动态调整策略,实现算力资源的精准匹配与高效利用。

a) 分阶段评估。聚焦用户需求分析,综合评估业务类型、数据规模、计算强度等关键参数,明确所需算力的性能基线。引入服务化能力,构建统一的服务镜像仓库与服务目录,实现软件环境的标准化封装与快速交付,从而提升算力服务的可用性与部署效率^[4]。

b) 场景适配。面向不同应用特征,建立差异化评估权重。通过按场景动态调整评估维度,实现算力配置的最优适配,提升资源利用率与业务支撑能力。

2.4 智能编排和调度

2.4.1 编排中心

编排中心通过资源抽象与模板化、 workflow 设计、脚本生成实现整合资源、优化流程、保障任务高效执行的目标。

a) 资源抽象与模板化。将异构计算资源统一抽象为标准化对象,并通过预定义模板实现资源的快速配置与批量管理。

b) workflow 设计。支持用户以图形化方式定义任务依赖关系和执行流程,实现复杂业务逻辑的可视化编排。

c) 脚本生成。根据用户配置和资源状态自动生成自动化部署脚本,驱动算力资源按需调度与任务执行^[5]。

2.4.2 调度中心

调度中心通过任务分配、负载均衡和优先级管理功能,实现算网资源的灵活调度。

a) 任务分配。根据任务需求与资源实时状态,智能匹配并调度任务至最优计算节点,提升执行效率。

b) 负载均衡。动态监控节点负载,合理分发任务流量,避免资源拥塞或闲置,保障系统整体性能稳定。

c) 优先级管理。支持按任务紧急程度设置优先级,确保高优先级任务优先获得资源并及时执行^[6]。

2.5 资源管理

资源管理包括资源发现、资源注册和资源更新。

a) 资源发现。通过网络扫描与探测技术,自动识别网络中新增的服务器、存储等计算资源,并将其纳入统一管理范围。

b) 资源注册。支持手动录入与自动接口方式,将资源的类型、配置及状态等信息登记至算网大脑,完成资源纳管。

c) 资源更新。持续监测资源运行状态,动态同步负载、故障等变化信息,确保算网大脑掌握实时资源视图^[7]。

2.6 计量中心和能力开放中心

2.6.1 计量中心

计量中心作为算网资源管理的关键环节,承担着用量统计与用量预测的重要职能。为计费结算和成本管理提供精准依据。计量中心记录计算时长、存储容量、网络流量等资源使用数据,为计费结算和成本管理提供精准依据^[8]。

2.6.2 能力开放中心

能力开放中心通过提供北向 API,允许第三方系统便捷地集成和调用算网大脑的功能。它通过 API 实现对算网资源的查询、任务提交、资源调度等操作,拓展算网大脑的应用场景和服务范围^[9]。

2.7 业务流程

详细业务流程如图 2 所示。算网大脑通过资源需

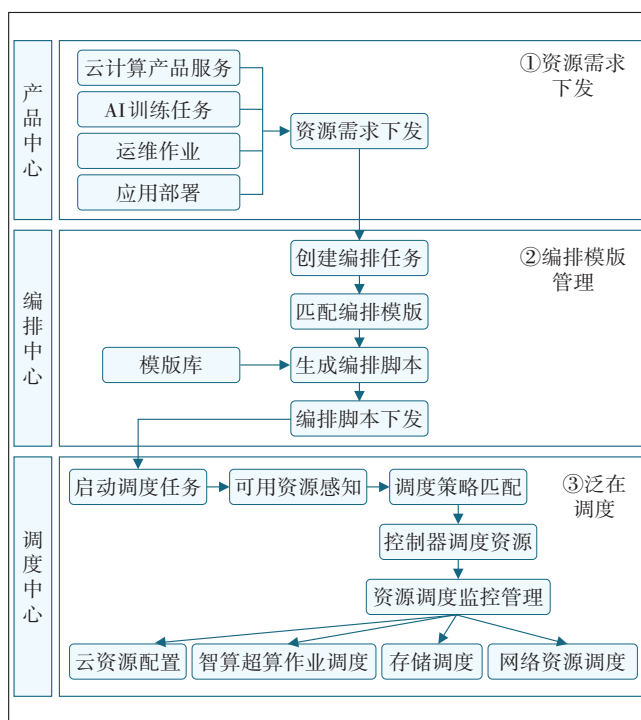


图 2 业务流程

求下发、编排模版管理、泛在调度流程实现算网业务的灵活开通和调度。

a) 资源需求下发。在算网资源调度分配场景(云计算、AI训练等)中,通过API向算网大脑下发需求,算网大脑解构算力需求至编排中心。经解构算力与匹配模板后,自动生成脚本并下发至调度中心完成编排。

b) 编排模版管理。可为云服务器、裸金属服务器等新增业务创建全新模板,也可基于现有模板修改创建新模板。所创建的模板形成模板库,用于在资源编排流程中生成脚本。算网大脑解构算力需求并匹配编排模板,生成脚本至调度中心完成任务。

c) 泛在调度流程。泛在调度模块在接收编排中心下发的脚本后启动调度任务,先感知可用资源平台及区域的资源信息,再基于调度策略评估确定最优配置方案,南向对接控制器下发指令并监控执行结果,指令成功后,向上游反馈任务完成。

3 应用部署

3.1 网络调度

网络调度流程如图3所示。算网大脑贯通智能专线从售前、售中到售后的全流程服务,实现云网资源可查、业务开通可视、SLA状态实时掌控的一体化智慧

服务。

3.2 灵活入算

灵活入算示意如图4所示。算网大脑基于多维度算路因子与星图算路算法,实现定源多宿的意图驱动选路与GIS可视化评估,并通过地址标准化、结果缓存及控制器协同,支持灵活入多算与业务流量的快拆快建调度^[10]。

3.3 运力管控

算网大脑通过一张全局运力图直观展示算力机房、服务节点及其资源和性能信息,结合地理化视图呈现链路时延、带宽利用率等关键指标,同时利用矩阵化查询提供任意节点间的详细链路数据,实现对网络问题及时发现和处理,助力运营商优化算网布局和服务范围。

4 应用价值分析

针对传统模式难以应对业务需求的动态变化与资源异构性的问题,本系统创新性地构建了以感知为基础、决策为核心、执行为落点的闭环运行理念。通过全局资源感知与动态状态反馈机制,实现资源发现、智能编排、分域协同与全生命周期管理。支持基于实时负载、SLA需求与能耗指标的自适应调度决策,

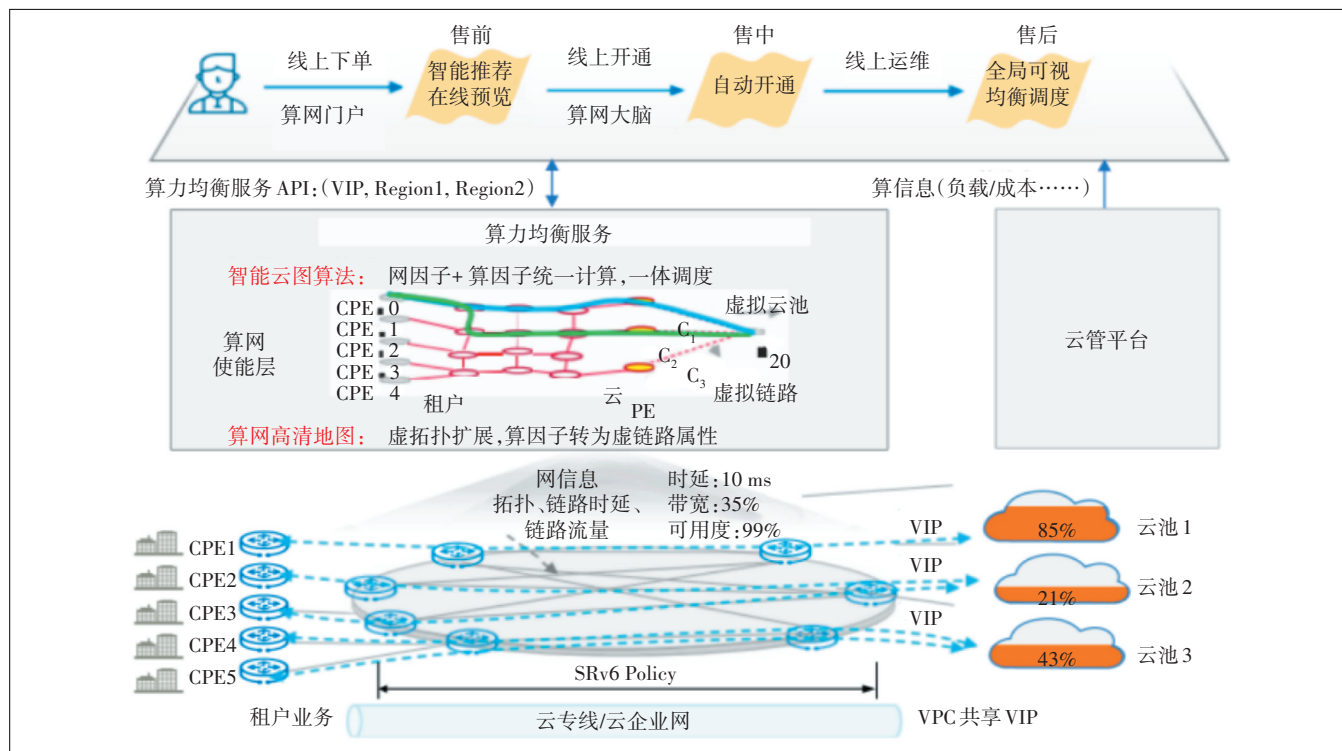


图3 网络调度流程

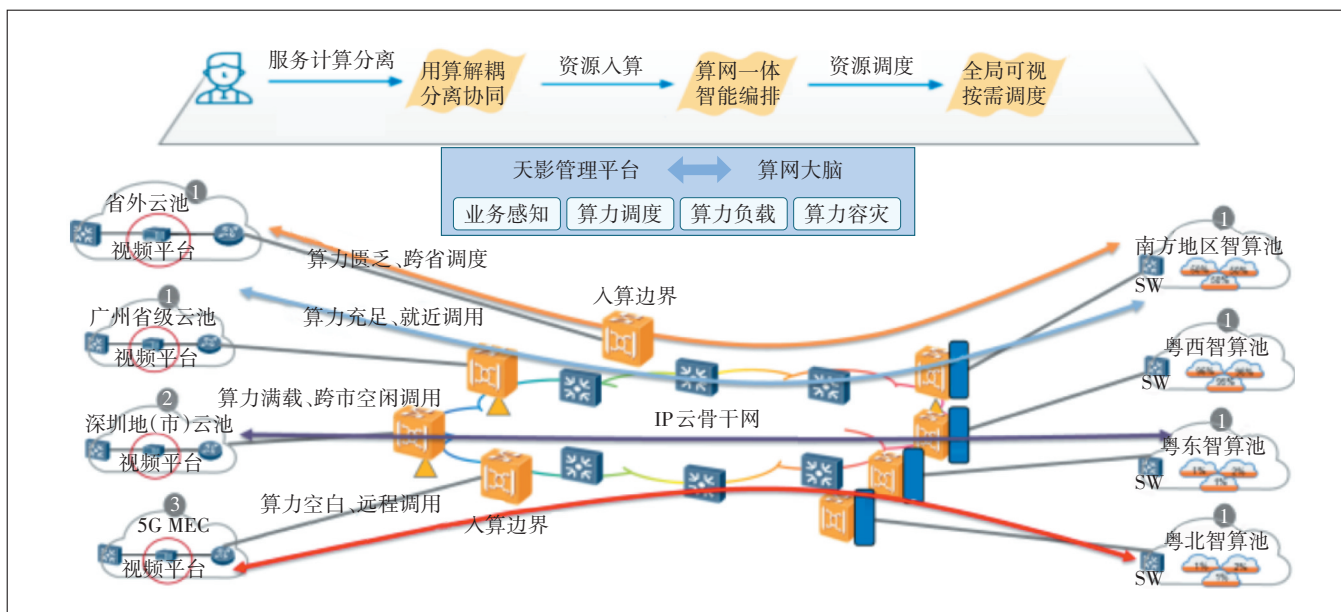


图4 灵活入算示意

显著提升资源利用率与调度响应速度。

现网应用表明,该系统在任务调度效率、跨域协同能力与能效比方面均优于传统方案,推动泛在算力调度向自动化、智能化持续演进。

5 结束语

本文通过构建由感知、决策与执行三层组成的调度架构,对算网资源进行闭环智能管控,实现跨区域、多层级算力资源的统一纳管与智能调度。系统创新性地融合算力、网络与安全资源,支持资源发现、智能编排、分域调度与全生命周期管理,推动泛在算力调度向自动化、智能化演进。

综上所述,随着数字时代的到来,面对日益增长的数据处理需求和复杂的业务场景,构建一个高效、灵活、可靠的算网大脑系统不仅是提升资源利用效率的关键,也是保障数字经济健康发展的基石。某省联通将通过不断探索和实践,致力于打造一个智能化、可扩展的智能算网调度平台,不仅为用户提供便捷高效的算力资源和服务,也为全国一体化算力网络建设贡献了自己的力量。随着技术的不断发展和完善,该平台将在更多领域发挥重要作用,助力各行各业实现数字化转型。

参考文献:

[1] 李明,王强.算网大脑驱动的泛在算力调度技术研究[J].通信技

术,2024,47(9):234-241.

[2] 杨帆,徐志强.算网大脑驱动下的智能调度系统设计与实现[J].通信工程学报,2024,48(10):345-352.
[3] 李华,王丽.算网大脑:泛在算力调度的理论与实践[M].北京:人民邮电出版社,2023.
[4] 孙健,刘洋.泛在算力调度中的智能资源分配策略[J].现代通信,2023,36(11):105-112.
[5] 刘洋,赵亮.智能化算力调度技术及其应用[M].北京:机械工业出版社,2021.
[6] 张伟,赵丽.基于人工智能的泛在算力调度优化算法[J].计算机应用,2024,44(6):876-883.
[7] 张磊,高峰.基于AI的泛在算力调度管理[M].北京:清华大学出版社,2022.
[8] 陈晓峰,周凯.通信网络智能化调度与优化[M].北京:电子工业出版社,2022.
[9] 徐志强,杨帆.AI技术在通信网络调度中的创新应用[M].北京:清华大学出版社,2020.
[10] 陈晓峰,黄敏.深度学习在泛在算力调度中的应用探索[J].数据通信,2023,46(7):67-74.

作者简介:

刘运奇,工程师,学士,主要从事云计算、算力网络等领域研究和IT系统规划建设;邓玲,正高级工程师,硕士,主要从事5G网络、算力网络规划建设和科技创新工作;冯铭能,高级工程师,博士,主要从事数据网、核心网、算力网络的相关咨询设计工作;韦秀林,工程师,学士,主要从事数据网、网络安全、算力网络的相关系统研发工作。