



基于多中心集群协同计算的大数据平台技术研究与应用

阮宜龙, 徐雪灵, 法虎, 董丝纶, 姜磊, 杨磊, 燕媛媛
(中电信人工智能科技(北京)有限公司, 北京 100010)

摘要: 中国电信面向横跨多个地域、拥有众多集群的大型政企机构, 推出可以高效协同各类资源的广域大数据架构体系——云边智算大数据平台。该平台从集群维度对数据分区进行逻辑抽象, 将独立分散的数据集整合为一个“虚拟数据集”, 实现了一对多的数据集映射管理。同时, 该平台的计算负载数据集具有泛化特征, 能够灵活应对不同场景下的数据处理需求。另外, 该平台以关系表达式为中间表示, 支持多种计算引擎和调度系统, 能够在复杂的大型数据处理高容错场景中高效地完成批处理任务负载。目前, 云边智算大数据平台已在多种应用场景中落地, 平台在5G Core能力调度子系统(5GC)多中心大数据作业开发、运营方面提效17%, 且已实现8省前置大数据集群共计42 PB存储、84 TB内存、24 984 VCore计算资源的协同调度, 日均完成80 308次前置-核心两级任务调度。

关键词: 云边协同; 统一SQL; 任务优化; 大数据平台

中图分类号: TP391

文献标志码: A

doi: 10.11959/j.issn.1000-0801.2024152

Research and application of big data platform technology based on multi-centre collaborative computing

RUAN Yilong, XU Xueling, FA Hu, DONG Silun, JIANG Lei, YANG Lei, YAN Yuanyuan
China Telecom Artificial Intelligence Technology Co., Ltd., Beijing 100010, China

Abstract: China Telecom has launched a high-efficient and collaborative wide-area big data architecture system, the cloud edge computing big data platform, for large-scale governmental and enterprise organizations spanning multiple geographies and clusters. The platform logically abstracts data partitions through the cluster dimension, integrates multiple independent datasets into a “virtual dataset”, and achieves many-to-one dataset mapping management. At the same time, the computing load dataset of the platform has generalized characteristics, which can flexibly cope with the data processing requirements in different scenarios. In addition, the platform also supports a variety of computing engines and scheduling systems using relational expressions as intermediate representations to achieve batch tasks for large-scale, complex data processing in highly fault-tolerant scenarios. At present, the cloud edge computing big data platform has been applied in a variety of application scenes. The platform has improved efficiency by 17% in 5G Core capacity scheduling subsystem (5GC) multi-centre big data job development and operation, and has achieved the collaborative scheduling of a total of 42 PB of storage, 84 TB of memory, and 24 984 VCore computing resources, with a



daily average of 80 308 times of task scheduling between the front cluster and the core cluster.

Key words: cloud edge collaboration, uniform SQL, task optimization, big data platform

0 引言

传统的集中式大数据集群架构在不同集群内所使用的数据源类型不同,对应的实现技术和具体存储格式有较大差异,因此存在多集群间不易协同、数据难以共享的问题。与此同时,各种计算引擎升级迭代,与其耦合的计算任务带来了高昂的迁移成本,海量数据同步消耗巨量网络带宽,任务失败概率也大幅上升^[1]。并且各集群、机器节点、组件、云平台数量庞大且种类多样,缺乏统一元数据管理,用户在使用数据时不能快速获取数据具体的存储位置及存储形式,导致传统大数据集群存在安全管控困难、节点运营分散、组件管理不统一、配置管理烦琐和运营资源消耗大等问题。另外,传统大数据集群中数据生产成本高且自动化程度不足^[2],如跨 N 个地区的查询需要进行至少 N 次数据采集、查询等操作,且需要在前序操作完成后再手动创建后续任务。因此,以高效自动化的方式在多个集群间共享、分析数据已成为新技术的发展趋势。

当前国内外的数据虚拟化平台在元数据管理方面仅支持一对一数据集映射管理方式,且描述计算负载的数据集不具备泛化特征。在协同计算方面,这些平台仅支持中小数据规模下的分析型、探索型任务负载^[3-5]。而中国电信的云边智算大数据平台实现了多集群统一元数据管理,即从集群维度对数据分区进行逻辑抽象,将独立分散的数据集整合为一个“虚拟数据集”,实现了一对多的数据集映射管理;此外平台还实现了跨地域多集群间的资源协同调度,以关系表达式为中间表示,支持多种计算引擎和调度系统。该平台对开发者屏蔽了计算、存储、网络等资源差异和大数据基础设施的分散与异构特性,极大地简

化了多中心集群、多种数据源任务的开发、部署和运维流程,同时平台可以完成复杂的大型数据处理和高容错场景中的批处理任务,大幅提升广域大数据环境总体资源利用率和开发运营效率。目前该平台已将集团总部集群与各省分公司的集群结合,形成了 $1+31+N$ 的集群协同架构,实现了在各边缘集群任务一点发布后,由云上集群统一调度和协同计算。

1 国内外发展现状

2021年上半年,我国大数据平台市场规模达54.2亿元,同比增长43.5%^[6]。2022年,中国大数据私有化部署市场规模达94.3亿元,大数据平台公有云服务市场规模达49.6亿元^[7]。随着数据应用的不断发展与演进,形成了多云、多集群、千湖万库的数据应用现状,如何高效协同、调度多个大数据集群间的数据资源与计算资源成为亟待解决的问题。国内外的学术界和产业界对上述问题进行了一些探索。Li等^[8]提出了一种利用云计算分散传统Hadoop MapReduce的优化框架以最小化计算和传输成本,该框架能够将原始的大规模问题分解为多个分别可解的子问题,为解决多集群难以同步的问题提供了思路。Fusco等^[9]提出了一种异构数据源语义集成方法数据集集成框架(data integration framework, DIF),在真实数据之上创建了虚拟视图,允许外部应用以透明方式访问集群数据,然而该方法应用的集群规模较小,未针对实际应用场景进行优化。Yang等^[10]建立了一个包括资源、应用、服务和安全的协同模型,但这个模型不能适配多集群异构。

2018年,谷歌公司提出了F1 Query^[11],可以覆盖企业级大数据处理和分析领域中的数据处理需求,既支持高效查询小规模 of 的联机事务处理

(online transaction processing, OLTP) 数据, 也支持对大批量异构数据进行快速即席查询, 同时可以对超大规模的数据进行可靠的数据抽取、转换、加载 (extract-transform-load, ETL) 处理, 但该方案不支持多集群、跨地域的协同计算, 在元数据管控方面仅支持一对一的数据集注册机制。2019年年末, 华为开源了基于 Presto 的数据虚拟化引擎 openLooKeng^[12], 它支持多集群、多种数据源的协同查询, 但其受限于 Presto 架构仅支持即席查询场景, 无法处理大规模数据的批量任务。Denodo^[13]公司的数据虚拟化产品则仅支持分析、探索型任务负载, 缺少对多中心跨源任务批处理的支持与调度管理, 也未能支持多集群多数据源的统一元数据管理机制。

综上所述, 目前国内外对去中心化大数据集群高效协同已有一些有益的探索, 但针对大型政企机构存在跨地域、多集群、数据源复杂的实际场景, 还需要一个能够统一纳管元数据、协同调度不同存储和计算资源且能够屏蔽不同集群底层异构性的大数据计算平台。

2 云边智算大数据平台

中国电信针对去中心化大数据集群协同问题, 设计开发了云边智算大数据平台, 其中的统

一元数据模块在元数据系统具备一对多的映射能力, 元数据模型的全局泛化能力和表征能力大幅增强, 且使用简单灵活; 而独特的任务优化引擎可以在多集群、跨地域的场景下, 利用任务优化算法同时支持即席查询与复杂大型批处理任务, 打破了计算引擎、集群和地域上的限制。

2.1 平台架构

中国电信的云边智算大数据平台实现了 Data Fabric 数据架构^[14], 云边智算大数据平台的多中心协同计算引擎架构如图1所示, 主要包含 Java 数据库连接 (Java database connect, JDBC) 驱动、统一元数据服务器、主节点 (Master)、任务管理器 (TaskManager)、省份工作节点 (Worker) 和对象存储缓存层, 第2.2节和第2.3节将对统一元数据和主节点中的任务优化引擎进行详细介绍。

如图1所示, 用户使用时首先通过 JDBC CLI 提交 SQL 任务; 在任务处理流程中, 平台的 Master 模块首先对提交的 SQL 进行风险评估, 拦截可能影响系统性能的高风险 SQL, 同时结合统一元数据服务对任务进行验证和权限控制, 并利用其提供的统计信息, 根据预定规则和成本对执行计划进行优化; 最后, 任务通过跨域任务卸载模块智能地根据数据所在集群和数据源类型进行卸

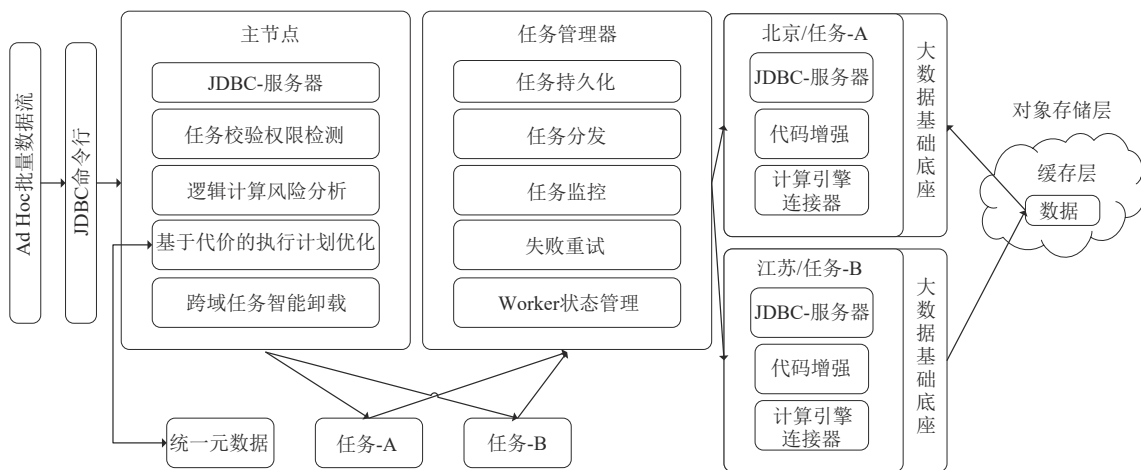


图1 云边智算大数据平台的多中心协同计算引擎架构



载^[15-16]。任务会根据数据依赖关系分层处理，同层任务并行执行，且每层任务完成后其结果会作为下一层任务的输入继续处理。在整个任务处理流程中，任务管理器记录并更新各个子任务的运行状态、数据源和所在集群等信息，同时全程监控任务运行情况。

实际场景中，本平台构建了广域大数据计算环境，支持统一声明多中心集群、多种数据源的任务，并可以进行多中心集群协同计算与协同调度。同时，平台屏蔽了多种资源差异和大数据基础设施的异构特性，逻辑上将数据统一挂载，极大地简化了多中心集群、多种数据源任务的开发、部署和运维流程。另外，本平台还可以纳管大型政企中已建设完成的公有云与私有云大数据集群环境。

2.2 统一元数据模块

传统元数据系统仅维护真实存在的数据库和表，缺乏对元数据的泛化能力。例如，F1 Query 和 Denodo 都只支持一对一的数据集注册机制，无法适配多中心集群的复杂需求。本平台创新地引入了“虚拟数据集”，实现了“虚拟数据集”和真实数据集的联动，提供了元数据方面的全局视角，使得元数据系统具备一对多的映射能力，元数据模型的全局泛化能力和表征能力大幅增强。

元数据是描述数据特征的数据，提供有关数据的内容、来源、格式等重要信息，为全景数据视图的构建，任务优化引擎的元数据查询、验证、绑定，执行计划的生成提供指引。统一元数据模块能够实现多中心集群库、表、列、权限等元数据信息的统一存储，为多集群场景中统一 SQL 的解析和优化提供支持，因此统一元数据模块对实现多中心协同计算至关重要。

为了保证数据的一致性，业务功能相似的多中心集群通常复用相同的数据仓库模型，即不同地域不同数据源的集群具有名称和结构相同的

库、表和列。传统元数据系统仅支持一对一的数据集映射管理，此类数据集在计算负载的描述方面不具备泛化特征，这导致任务描述存在大量的冗余语句，不利于复杂任务的开发与维护。为了解决上述问题，本平台从集群维度对数据分区进行逻辑抽象，将独立分散的数据集整合为一个“虚拟数据集”，只需通过简单的数据集索引即可操作多个集群的不同数据集，能够极大地简化复杂 SQL 描述语句，有效地提升开发效率。

统一元数据模块分为统一接口层、收集层和查询层，统一元数据模块结构如图2所示。统一接口层封装了不同数据源的连接方法，通过该模块可直连到底层引擎进行操作，因此平台支持多种类型的数据源，且具备按需扩展的能力。收集层调用统一接口层查询各个数据源的库、表、列、权限等元数据信息，当元数据有变更时，将元数据按照预定义的格式封装为消息推送到消息队列中，消息队列数据的高可用机制保证元数据不会丢失。消息队列监听器负责对元数据进行解析并持久化到数据库中。查询层提供查询检索接口，以便下游系统调用，且该层提供了数据源注册、数据源下线、数据源配置管理、元数据授权、鉴权、元数据目录树查询等功能。

虚拟数据集挂载实现了数据集的一对多映射关系。虚拟数据集挂载如图3所示，beijing、shanghai、guangzhou 这3个不同地域的真实数据集被挂载到一个名为 company_db.employee 的虚拟数据集上，其中，真实数据集采用五段式方法描述，即{地域，集群，数据源，数据库，表}，而虚拟数据集采用两段式方法描述，即{数据库，表}。

云边智算大数据平台采用虚拟数据集描述数据表从而实现任务声明的简化。任务声明对比如图4所示，针对同一个任务，传统方式的任务声明需要3个 select 子句并结合 union 操作，而采用虚拟数据集的任务声明只需一个 select 子句。其

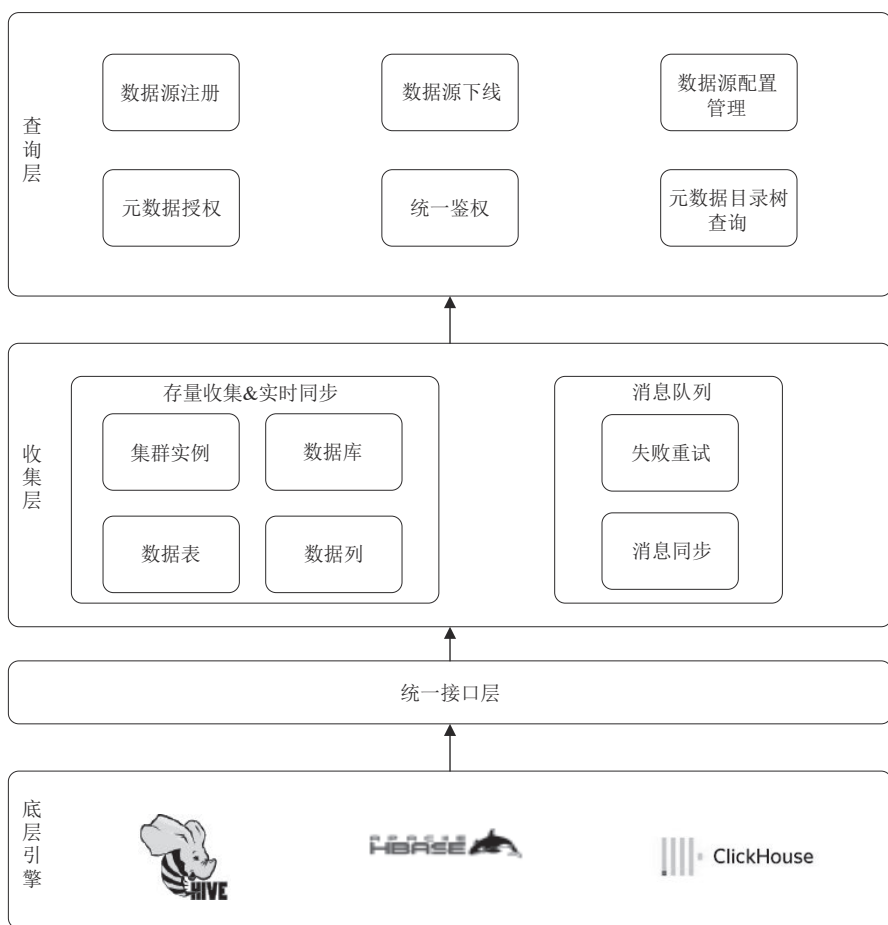


图2 统一元数据模块结构

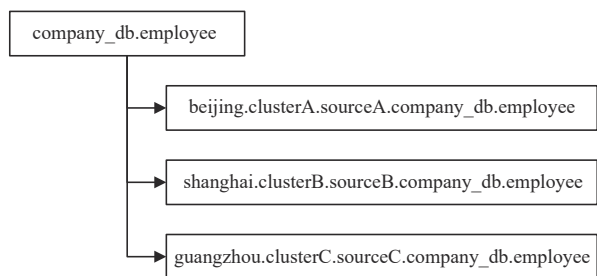


图3 虚拟数据集挂载

中，`ceclusterid` 是 SQL 解析过程中的保留字，它应用于从虚拟数据集映射到特定真实数据集的场景中，可以指定真实数据集的范围。从任务声明的对比可以看出，采用虚拟数据集的方式比采用传统方式更简单清晰，有效降低了开发难度。

SQL 解析过程对任务关系表达式中的各节点进行迭代处理，将原始 SQL 拓展为最终 SQL。若

遍历过程中发现了虚拟表，则获取查询条件；若查询条件中不存在预定义的 `ceclusterid` 保留字，则请求元数据，获取该虚拟表对应的所有真实表，将虚拟表改写为所有真实表 `union` 的形式；若查询条件中存在 `ceclusterid` 保留字，则需继续遍历该关系节点的条件树，对条件树进行分解，把对应的过滤条件映射到相应的真实表中，同时需要对拆分的条件进行改写，删除条件树中用于指定真实表的 `ceclusterid` 过滤条件，得到真实表、真实过滤条件的元组。最后对该虚拟表进行改写，将带虚拟表和虚拟过滤条件的关系节点改写为多个具有真实过滤条件的真实表的 `union` 形式。具有 `ceclusterid` 保留字的 SQL 解析示例如图5所示。



```
传统方式:  
select * from beijing.clusterA.sourceA.company_db.employee  
union all  
select * from shanghai.clusterB.sourceB.company_db.employee  
union all  
select * from guangzhou.clusterC.sourceC.company_db.employee
```



```
现有方式1:  
select * from company_db.employee  
  
现有方式2:  
select * from company_db.employee  
where cecclusterid in ('beijing.clusterA.sourceA', 'shanghai.clusterB.sourceB', 'guangzhou.clusterC.sourceC')
```

图4 任务声明对比

```
原始SQL:  
select * from company_db.employee  
where (cecclusterid in ('beijing.clusterA.sourceA', 'shanghai.clusterB.sourceB') and col_a > 10)  
or (cecclusterid in ('guangzhou.clusterC.sourceC') and col_b < 20)
```



```
最终解析转化成的SQL:  
select * from beijing.clusterA.sourceA.company_db.employee where col_a > 10  
union all  
select * from shanghai.clusterB.sourceB.company_db.employee where col_a > 10  
union all  
select * from guangzhou.clusterC.sourceC.company_db.employee where col_b < 20
```

图5 具有 cecclusterid 保留字的 SQL 解析示例

2.3 任务优化引擎

云边智算大数据平台的任务优化引擎在多集群、跨地域的复杂场景下，利用任务优化算法同时支持即席查询和复杂大型批处理任务，打破了计算引擎、集群和地域上的限制。平台通过插拔式优化策略，自顶向下地利用数据集的位置信息和统计信息，将大型任务拆分为小型子任务，并行处理不同的数据片段，加速整体计算过程。传统数据虚拟化平台仅支持中小数据规模下的分析型、探索型任务负载，本平台以关系表达式为中间表示，搭载多种计算引擎和调度系统，可以支持大型、复杂的数据处理、高容错场景的批任务

负载。与之相比，openLooKeng 受限于 Presto 架构，仅支持即席查询场景，无法处理大规模数据批量处理任务。

本平台使用的任务优化算法如图6所示，一个以SQL形式声明的批处理任务，涉及4个不同集群的数据集A、B、C、D，首先将该任务解析为一棵关系表达式树，根据数据集的位置特征从Scan节点自底向上开始对其染色。若树中的节点只经历过一次染色，则将其标记为纯色，若节点经历过两次及两次以上的染色，则将其标记为杂色，最后通过色彩信息逐级上溯，直至根节点。一组纯色节点将被封装为新的CecConverter节

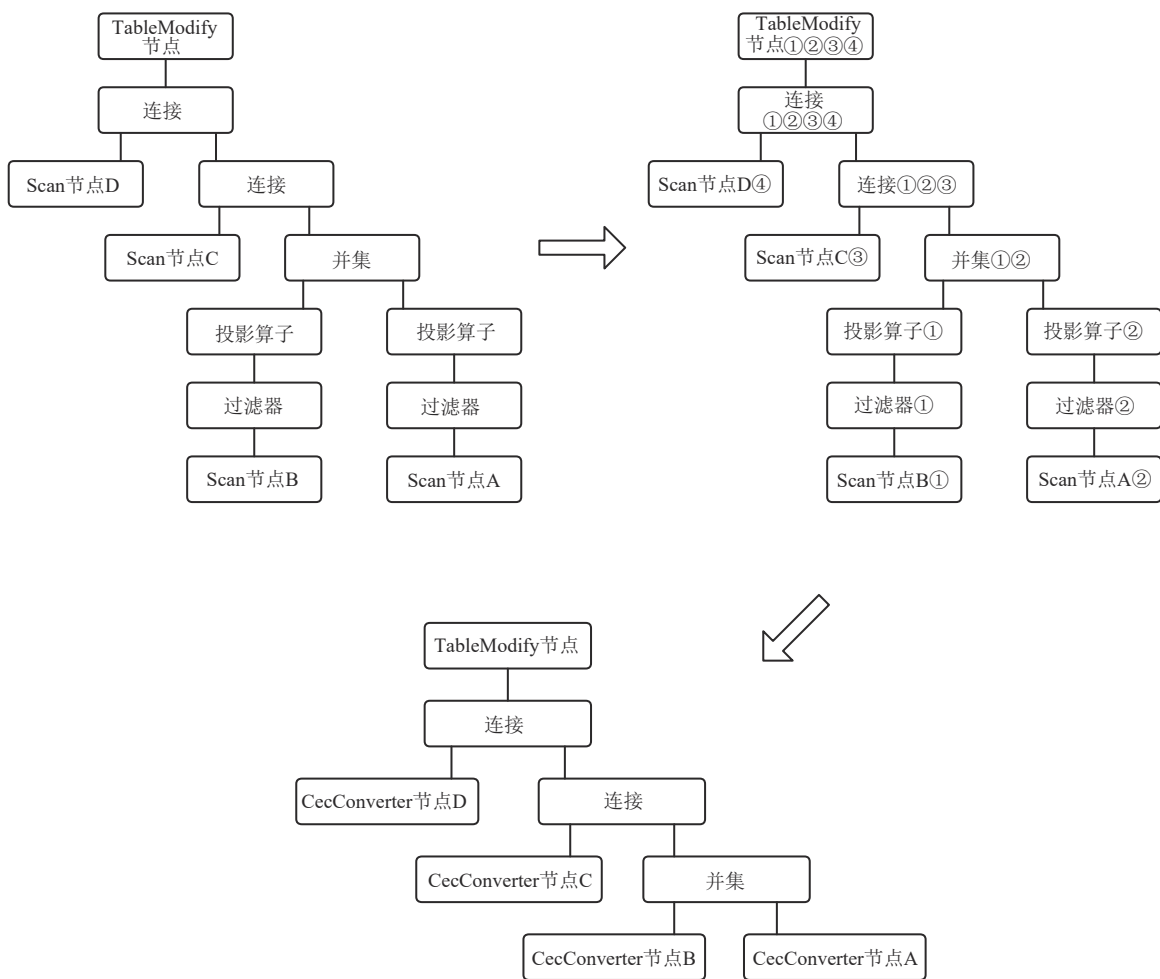


图6 本平台使用的任务优化算法

点，其余杂色节点成为一组，任务优化引擎自动会将每一组任务卸载到对应的集群中。

实现过程中，CecConverter 节点会根据任务配置标定到目标集群的调度系统，创建周期性工作流，并进行跨集群数据传输；TableModify 节点则在标定到目标集群的调度系统后，创建周期性工作流以完成数据的写入行为。批处理子任务固化为所在集群调度系统的工作流程如图7所示，与图6的染色结果对应，工作流A、B、C、D为CecConverter节点，工作流E为TableModify节点，其中，数据检查模块（DataChecker）负责判断数据是否到达；执行节点负责执行具体子任务工作负载，并完成最终的写入行为；数据传输模块（DataTransfer）负责跨集群数据传输。

云边智算大数据平台的任务优化引擎模块实现了多集群、跨地域的协同计算，通过任务优化算法将同逻辑的子任务卸载到不同集群中，突破了集群间、计算引擎间的限制，使平台可以支持大型、复杂的数据处理、高容错场景的批任务负载。

3 平台应用成效

目前，云边智算大数据平台已在中国电信集团总部和多个省分公司落地应用。平台具备统一元数据管理技术和多中心集群协同计算能力，已在中国电信集团的网络数据处理系统5GC信令详单查询业务场景中使用，可达到分钟级协同查询中国电信31省区市共4亿移动用户、万亿级明细

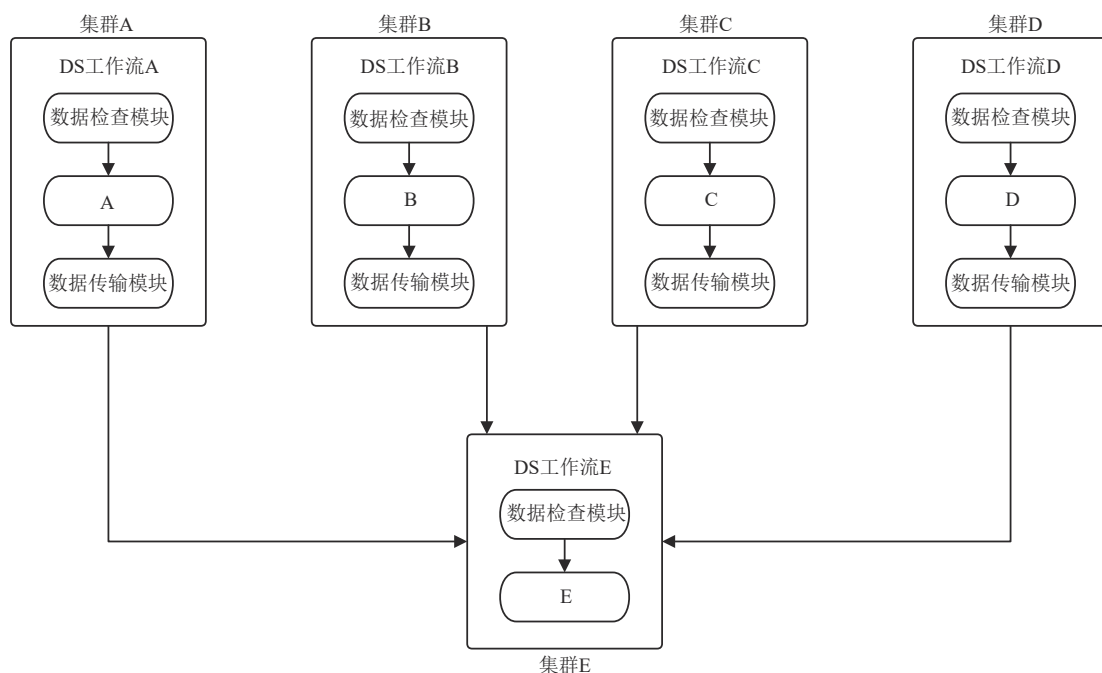


图7 批处理子任务固化为所在集群调度系统的工作流程

数据（5G深度包检测（deep packet inspection, DPI）、5G话单、接入和移动性管理功能（access and mobility management function, AMF）、会话管理功能（session management function, SMF））的需求。平台支持从核心集群一点调动广域31个省区市前置大数据环境的数据资源和计算资源，真正实现了批处理任务的多中心协同计算、协同调度和按需隐式数据共享。云边智算大数据平台已将各数据集群系统的子程序和主程序的接口开发、部署相关工作优化省略，据集团数据发展中心统计，在5GC多中心大数据作业场景中，本平台共提升开发、运营效率约17%，5GC系统应用提效见表1。

云边智算大数据平台具备多中心集群协同批处理能力，已支撑集团网络数据处理系统前置-核心两级作业调度场景。因为各前置集群中的批处理任务逻辑一致，所以可以通过本平台简化调度语句编写与分发，实现一点定义任务，多集群协同调度。目前，本平台已实现8个省区市前置大数据集群共计42 PB存储、84 TB内存、24 984

VCore计算资源的协同调度，完成日均80 308次前置-核心两级任务调度需求，各省区市协同批处理任务调度量统计如图8所示。

在平台的可视化界面中用户可以查看所纳管集群、数据源的统一元数据虚拟表，在界面左侧以树状结构展示了各集群数据源与所在地域；在界面中部的查询空间，平台支持即席查询和离线批处理两种查询方式，并将查询任务逻辑以右侧的有向无环图（directed acyclic graph, DAG）形式展示；此外，用户可以通过平台对提交任务的执行时间、任务使用资源、队列和用户的资源情况进行可视化的观测和审计操作。

4 结束语

云边智算大数据平台针对具有多地域、多集群的大型政企机构构建，能够实现任务一点发布、多集群协同计算、计算和存储资源统一调度。平台的统一元数据技术将已纳管的不同地域集群逻辑上抽象为一个“虚拟数据集”，对开发者和运营人员屏蔽计算、存储、网络等资源和大

表1 5GC系统应用提效

工作项	工作内容	旧架构工作量/ 人天	新架构工作量/ 人天	备注
模型物化	数据模型设计	5.0	5.0	
模型物化	指标算法设计	5.0	5.0	
数据存储	存储策略配置	5.0	5.0	
批量计算	数据算法模型开发与配置	5.0	5.0	新增1张表
任务配置	任务运行时配置	5.0	5.0	
作业管理	数据依赖关系设计与配置	5.0	5.0	
接口开发	5GC子程序新增接口，安装功能需求完成接口研发， 查询单集群数据集并输出查询结果	1.5	0.0	
接口开发	5GC主程序新增接口，通过查询31个省区市子程序新增 接口完成数据的汇总计算	1.5	0.0	
接口测试	5GC本地测试	1.5	1.5	基于已建设报表进行接口开发 (被优化的工作项工作量为0)
接口部署	5GC程序部署单省子程序以及核心节点的主程序	1.5	0.0	
接口注册	5GC程序新增接口在Dcoos平台注册	3.0	3.0	
接口联调	开展下游厂商联调工作	2.0	2.0	
接口部署	联调完毕后，在31个省区市完成5GC子程序部署、割接	3.0	0.0	
合计		44.0	36.5	

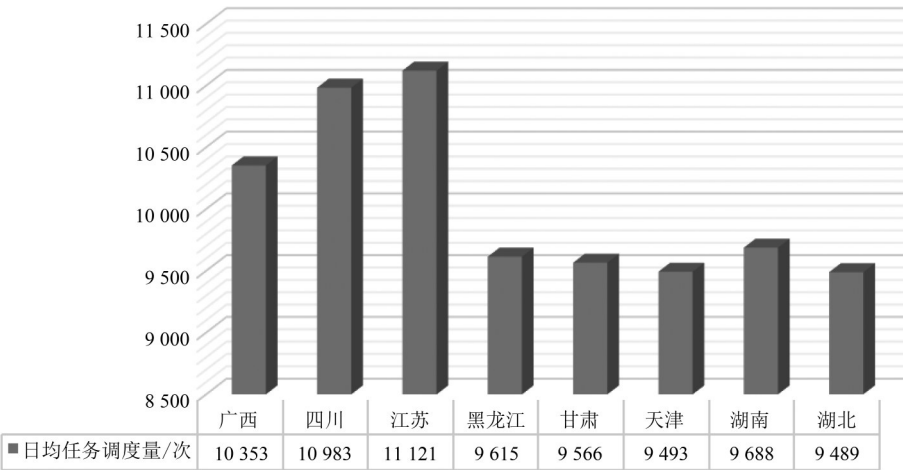


图8 各省区市协同批处理任务调度量统计

数据基础设施的分散与异构特性；任务优化引擎可以完成分钟级协同查询中国电信31个省区市、4亿移动用户万亿级明细数据，并且支持任务的多中心协同计算、协同调度、按需隐式数据共享，使得5GC多中心大数据作业开发、运营方面提效17%。此外，该平台以关系表达式为中间表示，支持多种计算引擎和调度系统，实现了大型、复杂的数据处理高容错场景的批处理任务，目前已实现8个省区市前置大数据集群存储与计

算资源的协同调度，日均完成80 308次前置-核心两级任务调度。在多中心集群协同计算、多中心作业协同调度等方面，云边智算大数据平台已具备众多优秀特性，达到了国际领先水平。

未来，云边智算大数据平台也将继续专注于不同场景的协同融合应用：在批处理场景中，将提升多中心异构数据源任务联合计算能力，优化多中心多表关联计算的能力；在实时处理场景中，将支持多中心异构数据源的批流一体化，使



得一份代码同时支持批量计算和实时计算；另外，平台将结合数据湖组件能力，实现数据的增量入湖和快速查询；在交互式查询场景中，将专注于多中心集群即席查询方面的优化，实现分钟级返回多中心集群明细数据，持续改进任务优化算法的性能；在多中心集群分布式调度方面，将开展联邦能力研发，推进分布式调度系统与现网业务的融合，并探索实现AI多级算力的协同。

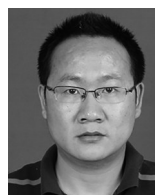
参考文献:

- [1] 百度智能云. 深入理解数据存储架构设计: 从集中式到分布式存储[EB]. (2024-02-18).
Baidu Intelligent Cloud. Deeper understanding of data storage architecture design: from centralized to distributed storage[EB]. (2024-02-18).
- [2] YI X, LIU F, LIU J, et al. Building a network highway for big data: architecture and challenges[J]. IEEE Network, 2014, 28(4): 5-13.
- [3] BRISBANE S. Decentralising Big Data Processing[D]. Sydney: The University of New South Wales, 2016.
- [4] JOHN F, GOPINATH S, SHERLY E. A decentralised framework for efficient storage and processing of big data using HDFS and IPFS[J]. International Journal of Humanitarian Technology, 2020, 1(2): 131-143.
- [5] 朱姝. 边缘云网下的新业务探索与应用[J]. 电信科学, 2023, 39(Z1): 55-61.
ZHU S. Exploration and application of new business under edge cloud network[J]. Telecommunications Science, 2023, 39 (Z1): 55-61.
- [6] IDC 中国. 2021 H1 大数据平台市场份额报告[EB]. (2021-12-27).
IDC China. 2021 H1 Big data platform market share report[EB]. (2021-12-27).
- [7] IDC 中国. IDC: 把握机遇, 穿越周期——2022 年中国大数据平台市场报告发布[EB]. (2023-07-25).
IDC China. IDC: seize the opportunity, traverse the cycle—China big data platform market report 2022 released[EB]. (2023-07-25).
- [8] LI P, GUO S, YU S, et al. Cross-cloud MapReduce for big data[J]. IEEE Transactions on Cloud Computing, 2020, 8(2): 375-386.
- [9] FUSCO G, AVERSANO L. An approach for semantic integration of heterogeneous data sources[J]. PeerJ Computer Science,

2020, 6: e254.

- [10] YANG J, LEE T Y, LEE W T, et al. A design and application of municipal service platform based on cloud-edge collaboration for smart cities[J]. Sensors, 2022, 22(22): 8784.
- [11] SAMWEL B, CIESLEWICZ J, HANDY B, et al. F1 query: declarative querying at scale[J]. Proceedings of the VLDB Endowment, 2018, 11(12): 1835-1848.
- [12] LIU X, YU Y, YU M, et al. An edge-cloud collaborative computing system for real-time internet-of-things applications[C]// Proceedings of International Conference on Computer Science and Education. Singapore: Springer Nature Singapore, 2022: 652-663.
- [13] PAN A, RAPOSO J, ÁLVAREZ M, et al. The Denodo data integration platform[C]// Proceedings of the 28th International Conference on Very Large Data Bases. New York: ACM Press, 2002: 986-989.
- [14] HECHLER E, WEIHRAUCH M, WU Y. Data fabric architecture patterns[M]// Data Fabric and Data Mesh Approaches with AI: A Guide to AI-based Data Cataloging, Governance, Integration, Orchestration, and Consumption. Berkeley, CA: Apress, 2023: 231-255.
- [15] LYU S, DAI X, MA Z, et al. A heterogeneous cloud-edge collaborative computing architecture with affinity-based workflow scheduling and resource allocation for internet-of-things applications[J]. Mobile Networks and Applications, 2023: 1-17.
- [16] ISLAM A, DEBNATH A, GHOSE M, et al. A survey on task offloading in multi-access edge computing[J]. Journal of Systems Architecture, 2021, 118: 102225.

[作者简介]



阮宜龙 (1977-), 男, 中电信人工智能科技(北京)有限公司大数据研发中心总经理、高级工程师, 主要研究方向为大数据平台架构规划与建设运营。



徐雪灵 (1997-), 女, 中电信人工智能科技(北京)有限公司工程师, 主要研究方向为大数据 PaaS 平台研究、存算分离、分布式缓存。



法虎 (1998-), 男, 中电信人工智能科技(北京)有限公司工程师, 主要研究方向为多中心协同计算、基因大数据压缩算法。



董丝纶（1997-），男，中电信人工智能科技（北京）有限公司工程师，主要研究方向为通用查询优化器。



杨磊（1986-），男，中电信人工智能科技（北京）有限公司高级工程师、PaaS 研发总监，主要研究方向为多模态融合 PaaS 平台和多中心跨域协同技术。



姜磊（1989-），男，中电信人工智能科技（北京）有限公司高级工程师，主要研究方向为数据虚拟化。



燕媛媛（1990-），女，中电信人工智能科技（北京）有限公司高级工程师，主要研究方向为大数据 PaaS 平台规划研究与管理。