【工程与技术研究】

面向云计算并发访问的计算机大数据调度负载均衡方法

王艳兵

（徽商职业学院 电子信息系，安徽 合肥 230000）

摘　要：为有效解决计算机大数据调度能耗高和访问不安全问题，提出面向云计算并发访问的计算机大数据调度负载均衡方法。在云计算部署方案下，使用正负理想解大数据全面访问机制，保证访问数据的全面性。通过约束数据需求，使得授权合法使用者能够存取获许可的数据，实现多阶段访问的身份认证，保证访问安全；构建基于云计算的调度模型，通过数据分级调度策略达到多批大数据处理最大化收益目的。构建任务执行时间和能耗双重优化函数，结合数据传输路径迭代函数，实现对计算机大数据多批处理调度。访问结果显示总隐私权重最高为 16.4，调度结果显示能耗调度结果与理想结果拟合度较高，且资源利用率在 96%以上，表明负载能够达到最佳均衡状态。

关键词：云计算环境；计算机大数据；数据调度；调度能耗；负载均衡

中图分类号：TP 311.1　　**文献标识码**：A　　DOI：10.13486/j.cnki.1673-2618.2023.06.011

随着信息技术和网络技术飞速发展，计算机储存的数据种类及数量日益增多。由于网络规模不断扩大，将会有海量数据向计算机服务中心转移，从而导致数据中心能耗不断增大，运营商运行费用不断上升。计算机中心聚集了大量计算设备和平台，但由于无法合理地配置和利用这些资源，引起计算机服务中心能源消耗增大，从而导致了资源浪费和服务费用持续增长。文献[1]提出了一种基于神经网络的应用分析方法，该方法根据分块规模、分支执行步骤，使用针对神经网络规模化的应用方法，结合 Winograd 算法，实现对计算机数据存储的进一步优化；文献[2]提出了模拟退火法的应用分析方法，该方法通过开放排队网络对移动业务流时延进行优化建模，使用模拟退火求解模型，并在不同服务请求量和虚拟网络结构之间建立逻辑关联，实现对计算机虚拟网络功能部署。然而，这两种方法容易受到计算机存储内存影响，导致远程计算机无法根据目标需求发送计算机所需内容，也无法实现计算机数据有效反馈。针对该问题，本文提出了面向云计算并发访问的计算机大数据调度负载均衡方法。

1　计算机大数据云计算并发访问控制

1.1　基于正负理想解的大数据全面访问

在云计算环境中，多个用户同时访问和处理大数据是常见的情况，容易出现数据丢失导致数据访问不

收稿日期：2023-04-30

基金项目：安徽省高校自然科学研究重点项目（2023AH053112）；安徽省高等学校省级质量工程项目（2022cjrh044，2021zyjxzyk031）

作者简介：王艳兵（1981—），男，安徽安庆人，副教授，硕士，主要从事软件和大数据研究。

E-mail：2072237278@qq.com

全面的问题。大数据全面访问机制可以通过正负理想解的大数据全面访问机制，将不同用户或应用程序的数据隔离开来，防止不同用户之间的数据冲突和干扰，确保数据的安全性和完整性。正负理想解关系如图 1 所示。考虑到负理想解的关键参考作用，计算备选项与正理想解距离公式为

$$L_{A^+,A^-}=\{\sum_{i=1}^{n}[\omega_i(\eta_i^+-\eta_i)/(\eta_i^+-\eta_i^-)]^{\varepsilon}\}^{1/\varepsilon}。$$

式中，η_i^+、η_i、η_i^- 分别表示云计算部署方案指标正理想值、性能值和负理想值，ω_i 表示权重，ε 表示指标加权和最优值。在原有计算机大数据访问机制基础上，计算备选项与负理想解距离公式为

$$L_{A^+,A^-}=\{\sum_{i=1}^{n}[\omega_i(\eta_i-\eta_i^-)/(\eta_i^+-\eta_i^-)]\}^{1/\varepsilon}。$$

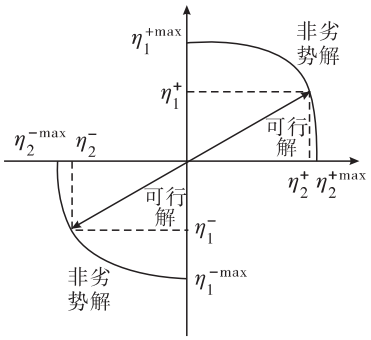


图 1　正负理想解关系图

ε 取值不同，可以反映决策者对指标偏离程度。当 $\varepsilon=1$ 时，强调云计算部署方案整体效用最大化；当 $\varepsilon\to\infty$ 时，强调云计算部署方案整体惩罚最大化。通过云计算部署方案的正负理想解大数据全面访问机制，可保证访问数据的全面性。

1.2　访问控制身份认证机制设计

在保证访问数据全面性的同时，为了保证数据的安全性和隐私保护，需要对访问身份进行验证，确保只有经过身份认证的用户才能访问数据，并且按照其权限进行合法的访问操作。访问控制机制利用预先定义的存取控制策略，对数据需求进行约束，使授权的合法使用者能够存取数据，而不能存取未获许可的数据。适当的访问控制机制，可以有效地限制使用者的存取权限，保证信息安全。利用云计算技术进行数据支持，将会给用户带来更好的数据处理体验效果。

访问控制身份认证机制主要包括密码、一次性令牌、条件属性验证，验证请求用户是否合法。当客户机以 ID 及口令登录时，若口令是有效的，则会产生一次令牌，将该令牌直接传送到客户端，结合一次性令牌提高云计算中客户认证效率。访问控制身份认证机制采用一次性令牌对客户端进行身份认证，而口令则是为了防止账户被非法访问而用来保护用户账户安全的密码。无论何时客户机登录，访问控制身份认证机制都会提供一种新的可使用标记。一次性令牌由系统自行根据下列公式自动产生，$ID\to OT$，$OT=\{p\}$。式中，OT 表示一次性令牌，p 表示素数。当客户端登录时，这个方法就会提供一个新的可用标签。这种一次性标记是通过系统按照

$$Q=\begin{cases}\text{if}\quad X_1\cap X_2\cap X_3\text{，被认证，}\\ \text{其他，未认证}\end{cases}$$

自动产生的。式中，X_1、X_2、X_3 分别表示计算机大数据密码、一次性令牌、条件属性。如果多阶段身份被认证，那么说明控制请求是有效性，具有一定可信性。反之，则不可信，无法通过访问控制身份认证。

2　计算机大数据多批处理调度

对计算机大数据云计算并发访问进行控制，可以提高数据访问的全面性和安全性，但是计算机大数据中的数据量过大和计算任务过多会造成计算机负载失衡，进而导致计算机系统性能降低，给计算中心带来更多能源消耗。因此，通过对计算机进行多批调度处理，在保证访问安全的条件下，可降低调度能耗，提高计算机的系统性能。

2.1　制定计算机大数据多批处理分级方案

利用 MapReduce 迭代法对基于云计算的多批次数据进行排序，每一个映射任务都会从 HDFS 上装载数据，而 Reduce 任务则会向 HDFS 发送一次迭代中间结果。在下一个循环中，同样装载和传送程序也会重复[3-4]。为防止因数据量过大和计算任务过多造成计算机负载失衡，将数据多批处理任务分配到固

定节点上。因此，构建基于云计算的任务调度模型(图 2)。

由于云计算具有大量计算机集群[5-7]，这些集群的结构复杂且异构，所以通过负载均衡处理，使各个计算机集群均衡地分配数据和计算任务，降低计算机集群的能耗，让网络资源得到更加平衡和充分的利用，避免某些计算机过载而导致系统性能下降，或某些计算机空闲而浪费网络资源。在服务端收到批量数据后，通过评估其价值函数，计算出任务收益，从而判定是否接受[8]。服务方接收任务时的增益

$$F=\begin{cases} r, t_1 \leqslant t_{\max}, \\ r-(t_0\times\delta), t_1 > t_{\max}。\end{cases}$$

式中，r 表示任务接收后所带来的收益，t_0、t_1、$t_{\max}$分别表示延迟时间、任务持续时间和任务完成时间，δ 表示收益变化值[9]。由于计算机处理大数据时，面对的是多批处理任务，所以该情况下的服务方接收任务增益可表示为

$$F_{\max}=\frac{\{\alpha F-(1-\beta)z\}}{t_B}。$$

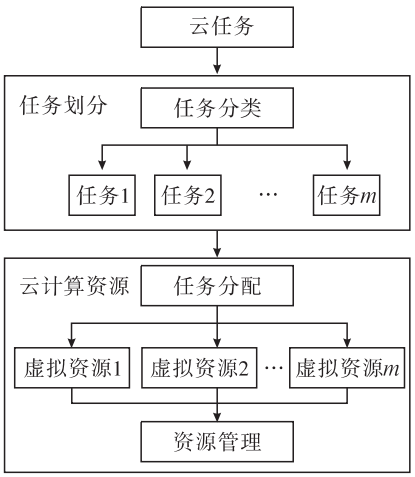


图 2 基于云计算的调度模型

式中，α、β 分别表示收益和未收益成本比例，t_B 表示任务 B 持续时间，z 表示任务 B 执行成本。将相同收益计算机大数据调度任务归一处理，通过数据分级调度策略达到多批大数据处理最大化收益目的[10]。

2.2 大数据多批处理调度方案的实现

在云计算环境下，用户提交的任务调度为待执行 B 任务，将调度分配到适当计算资源节点，以满足用户要求。调度分为两个阶段：第一阶段的调度是根据用户运行时间来安排任务到虚拟机[11]；第二级段调度是根据任务特性和负载状况，对虚拟机进行合理分配，以保证系统资源负载平衡，同时降低系统整体功耗[12]。为了快速完成任务，需要大量的运算节点，这导致计算中心的能耗和运营商的成本上升。为了使任务能在最短时间内完成，必须调动更多计算节点，这就给计算中心带来更多能源消耗[13]。为了在低成本、短时间内调度多批大数据，构建了任务执行时间和能耗双重优化函数

$$T_{总}=\max\left(\sum_{a=1}^{j}T_t(a,b)\right)。$$

式中，$T_t(a,b)$ 表示任务从节点 a 到节点 b 预计完成的时间，j 表示任务完成轮次。

任务调度能耗成本为

$$C_{总}=\sum_{a=1}^{j}\sum_{b=1}^{j}[C_1(j)+C_2(j)+C_3(j)]。$$

式中，$C_1(j)$、$C_2(j)$、$C_3(j)$分别表示单位时间计算、传输和存储所消耗的能量。云环境中，当任务分配时，适应性高的个体会被更多地遗传到下一代，而适应性低的个体会在每次的竞争中被淘汰[14]。

采用双目标优化方法，以减少任务完成时间，降低计算服务中心能量消耗。构建基于时间－能耗双重任务调度的适应度函数

$$H(I)=\omega_{时间}h_{时间}(I)+\omega_{能耗}h_{能耗}(I)。$$

式中，$\omega_{时间}$、$h_{时间}(I)$分别表示时间权重和适应度函数，$\omega_{能耗}$、$h_{能耗}(I)$分别表示能耗权重和适应度函数[15]。假设计算机大数据从节点 a 到节点 b 的最短传输路径为 d，那么数据传输路径迭代函数可表示为

$$d^m(b)=d^{m-1}(a)+\omega'(a,b)。$$

式中，m 表示迭代次数，$\omega'(a,b)$表示节点 a 到节点 b 的权值。充分考虑收益要素，通过计算获取调度任务和计算机空闲状态下相匹配的动态调度方案，公式为

$$T(B,G)=u(B)-t(B,G)+\omega_B v(B,G)。$$

式中，$u(B)$表示任务执行优先等级，$t(B,G)$表示计算机空闲状态 G 下任务执行初始时间，ω_B 表示任务权重，$v(B,G)$表示任务处理速度，通过该式实现对计算机大数据多批处理调度，降低调度能耗，提高计算机的系统性能。

3　方法测试

为了检验面向云计算并发访问的计算机大数据调度负载均衡方法合理性，采用某企业的云计算管理系统平台进行测试。云计算管理系统平台如图 3 所示。根据图 3 的云计算管理系统平台，基于网络流量数据集，从任务调度和数据访问两方面对计算机的能耗调度、资源利用率和总隐私权重展开测试，其结构如图 4 所示。

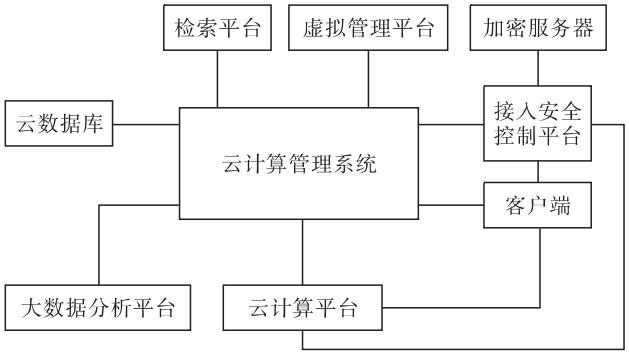


图 3　云计算管理系统平台

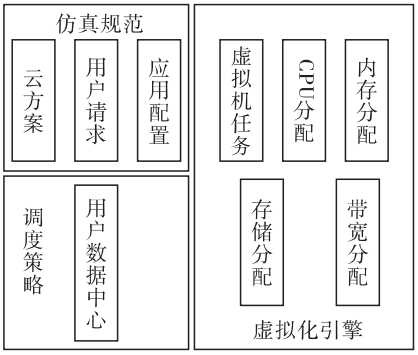


图 4　CloudSim 分层结构

计算机相关配置及负载参数如表 1 所示。

表 1　计算机相关配置及负载参数

参数名称		参数值	参数名称		参数值
计算机的配置	型号	Inteli5 3.1 GHz	负载值	CPU 使用率/%	50
	内存	4G		内存使用率/%	80
	操作系统	Windows 10		磁盘读取速度/(MB·s^{-1})	100
				网络传输速度/(MB·s^{-1})	10

3.1　访问测试

使用隐私保护方法来评估测试效果，隐私权保护率是指仅由被授权客户机能够正常存取云计算数据和整个云计算数据比例。隐私保护率计算公式为 $PR=\dfrac{N}{M}\times 100\%$。式中，$N$ 表示获取正确数据数量，M 表示数据总数。随着隐私权保护程度的提高，非法使用者在云端中获得数据服务数量也会降低，该算法验证能力也会随之提高。基于此，对比分析神经网络法、模拟退火法和本研究方法总隐私权重(图 5)。由图 5 可知，在保证所有数据均传输完成的前提下，使用神经网络法、模拟退火法的总隐私权重不如本研究方法总隐私权重大，所以使用本方法传输过程中隐私保护效率更高。

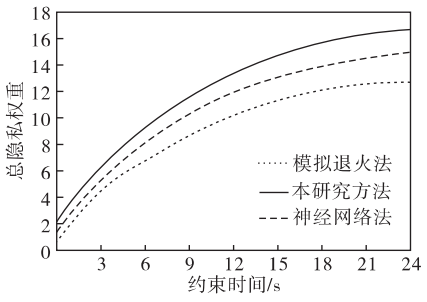
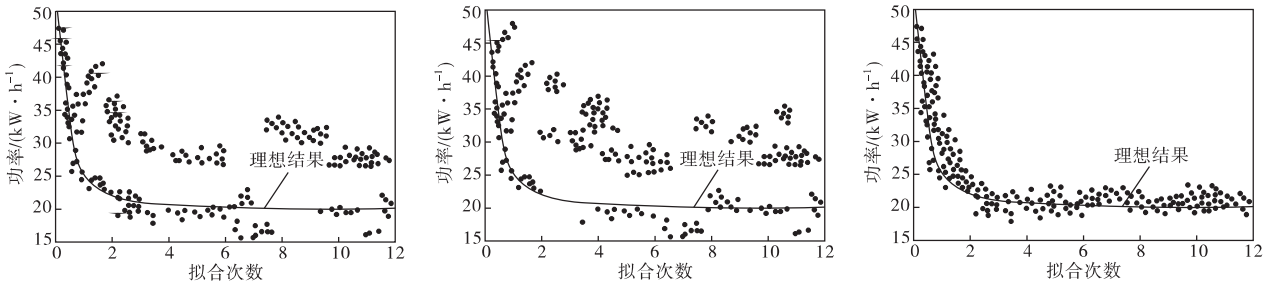


图 5　三种方法总隐私权重对比分析

3.2　调度测试

CloudSim 提供了专门的虚拟机、内存、带宽等接口，可以模拟虚拟环境中云计算技术，并能在这种环境中进行多批次数据调度。计算机运行时间和能耗是调度的关键因素，以调度能耗为例，使用神经网络

法、模拟退火法和本研究方法对比分析能耗调度情况，对比结果如图 6 所示。



(a)神经网络法 (b)模拟退火法 (c)本研究方法

图 6 三种方法能耗调度结果对比分析

由图 6 可知，使用神经网络法、模拟退火法能耗调度结果与理想结果相差较大，无法全部拟合在同一条曲线上，由此说明使用传统两种方法拟合效果较差，即能耗调度结果不理想；使用本研究方法能耗调度结果与理想结果接近，大部分数据拟合在同一条曲线上，由此说明使用本研究方法拟合效果较好，即能耗调度结果理想。

以调度时间为例，使用神经网络法、模拟退火法和本研究方法对比分析能耗调度情况，对比结果如图 7 所示。由图 7 可知，本研究方法在进行负载均衡调度时，网络资源利用率高于神经网络法和模拟退火法，说明使用本研究方法可以提高网络资源的利用率，负载均衡性能良好，能够达到最佳均衡状态。

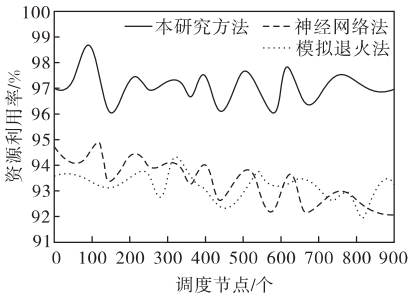


图 7 三种方法负载均衡调度结果对比分析

4 结束语

为了解决计算机大数据调度能耗高、访问不安全的问题，提出了一种面向云计算并发访问的计算机大数据调度负载均衡方法：给出了一种基于多级身份验证的云计算隐私保护算法，采用口令、一次性标识、条件属性等多种方式进行身份验证，以防止在计算复杂度较低情况下进行非法访问，从而提高了云数据隐私权保护。通过对多批数据的排序处理，保证了在以后数据调度中仍然能够很好实现负载平衡，防止了因负载不平衡而造成业务崩溃。

参考文献：

[1] 李茂文，曲国远，魏大洲，等. 面向 GPU 计算平台的神经网络卷积性能优化[J]. 计算机研究与发展，2022，59(6)：1181-1191.

[2] 陈卓，冯钢，刘怡静，等. MEC 中基于改进遗传模拟退火算法的虚拟网络功能部署策略[J]. 通信学报，2020，41(4)：70-80.

[3] 吕佳玉，竺智荣，姚志强. 云计算环境下的双通道数据动态加密策略[J]. 计算机应用，2020，40(8)：2268-2273.

[4] 李晓会，陈潮阳，伊华伟，等. 基于云计算和大数据分析的大规模网络流量预测[J]. 吉林大学学报(工学版)，2021，51(3)：1034-1039.

[5] 刘洋，赵瑞锋，李波，等. 基于 Docker 技术的静态安全分析云计算应用[J]. 电力科学与技术学报，2021，36(4)：181-187.

[6] 王若龙. 大数据分析技术在通信网络系统优化中的应用研究[J]. 电视技术，2021，45(8)：4-6.

[7] 武兰芬，姜军. 基于双源数据的云计算创新合作网络多维分析[J]. 科研管理，2020，41(2)：142-151.

[8] 刘炎培,朱淇,赵进超.边缘环境下计算密集型应用的卸载技术研究[J].计算机工程与应用,2020,56(15):1-14.
[9] 黄冬晴,俞黎阳,陈珏,等.面向移动边缘计算的联合计算卸载和资源分配策略研究[J].华东师范大学学报(自然科学版),2021(6):88-99.
[10] 董若楠,张光杰,刘渊,等.天地一体化信息网络动态重构技术与仿真方法[J].小型微型计算机系统,2020,41(5):1065-1070.
[11] 郁宁,王高才.基于可信期望的跨域访问安全性研究[J].计算机应用研究,2020(11):3406-3410.
[12] 屠袁飞,杨庚,张成真.一种面向云端辅助工业控制系统的安全机制[J].自动化学报,2021,47(2):432-441.
[13] 屠要峰,陈正华,韩银俊,等.基于持久性内存和 SSD 的后端存储 MixStore[J].计算机研究与发展,2021,58(2):406-417.
[14] 彭定洪,黄子航,王铁旦,等.面向云计算部署方案评价的区间犹豫模糊双重妥协评价方法[J].计算机集成制造系统,2021,27(6):1768-1779.
[15] 林学聪,董征,刘友武.用于云计算数据访问的多阶段身份认证[J].计算机工程与设计,2021,42(12):3396-3400.

Load Balancing Method of Computer Big Data Scheduling for Cloud Computing Concurrent Access

WANG Yanbing
(*Department of Electronic Information, Huishang Vocational College, Hefei* 230000, *China*)

Abstract: In order to effectively solve the problems of high energy consumption and to insecure access of computer big data scheduling, a load balancing method for computer big data scheduling for cloud computing concurrent access is proposed. Under the cloud computing deployment scheme, the positive and negative ideal solution big data comprehensive access mechanism is used to ensure the comprehensiveness of access data. By constraining data requirements, authorized legitimate users can access licensed data, achieve multi-stage access identity authentication, and ensure access security. Then, a scheduling model based on cloud computing is constructed to maximize the benefits of multi batch big data processing through data hierarchical scheduling strategy. A dual optimization function for task execution time and energy consumption is built, and the iterated function of data transmission path to achieve multi batch processing scheduling of computer big data is combined. According to the test results, the access results of this method show that the total privacy weight is the highest at 16.4. The scheduling results show that the energy consumption scheduling results have a high fit with the ideal results, and the resource utilization rate is above 96%, indicating that the load can reach the optimal balance state.

Keywords: cloud computing environment; computer big data; data scheduling; dispatching energy consumption; load balancing

(责任编辑:王新亮)