

云计算环境下海量数据高效调度模型仿真

方 晖

(惠州学院 教育技术中心, 广东 惠州 516007)

摘 要: 在云计算环境下海量数据调度的过程中, 数据节点的利用存在较强的随机性. 传统调度方法容易造成数据节点冲突, 从而导致调度效率降低. 提出基于数据节点冲突避免的云计算环境下海量数据高效调度方法. 将整个调度任务划分成多个子任务, 计算每个子任务任务量, 根据目标数据在数据节点的传输情况, 对数据节点的任务量进行计算, 根据计算结果, 建立优化的云计算环境下海量数据调度模型. 实验结果表明, 利用改进算法进行云计算环境下海量数据调度, 能够有效提高调度的效率, 避免了数据发生冲突的情况.

关键词: 云计算环境; 海量数据; 调度模型;

中图分类号: TP391.9

文献标识码: A

文章编号: 1000-7180(2014)09-0159-03

Huge Amounts of Data and Efficiently Under the Cloud Computing Environment Simulation Scheduling Model

FANG Hui

(Educational Technology Center of Huizhou University, Guangdong, Huizhou 516007, China)

Abstract: Huge amounts of data in a cloud computing environment scheduling process, the use of data nodes exist strong randomness. Traditional scheduling method is easy to cause data node conflict, leading to the scheduling efficiency lower. Based on data node to avoid conflict of huge amounts of data efficiently scheduling method in cloud computing environment. The whole scheduling tasks into subtasks, each subtask quantity calculation, according to the target data in data transmission of a node, the data node quota calculation, according to the calculation results, establishing optimized scheduling model huge amounts of data in cloud computing environment. The experimental results show that the improved data scheduling algorithm in cloud computing environment, can effectively improve the efficiency of the scheduling, to avoid the data conflict situations.

Key words: cloud computing environment; Huge amounts of data; Scheduling model;

1 引言

随着信息技术的快速发展, 数据管理已经步入云计算时代, 数据管理已经成为很多领域的核心部分, 发挥着越来越重要的作用^[1]. 在云计算环境下, 存在着海量数据, 利用合理的调度方法能够从海量数据中提取到目标数据, 因此, 如何针对云计算环境下海量数据进行高效调度, 已经成为数据管理领域需要研究的热门课题, 受到很多专家的重点关注^[2]. 当前阶段, 很多科研机构已经对云计算环境下的海

量数据调度任务进行了深入的研究, 并提出很多调度方法, 这些调度方法主要包括基于蚁群算法的云计算环境下海量数据调度方法^[3]、基于模拟退火算法的云计算环境下海量数据调度方法^[4]和基于数据节点关联系数的云计算环境下海量数据调度方法^[5]. 其中最常用的是基于数据节点关联性系数的调度方法^[6]. 由于云计算环境下海量数据调度方法在云计算环境下的数据管理领域发挥着无可替代的作用^[7], 因此, 拥有广阔的发展前景, 成为很多专家研究的重点课题^[8-10].

收稿日期: 2013-12-18; 修回日期: 2014-02-09

在实验的过程中,设置目标数据量是逐渐增加的.在此情况下,调度任务量越多,数据节点发生冲突的概率就越高.分别利用传统算法和改进算法进行云计算环境下海量数据调度实验.利用图2能够描述不同算法的调度时间.

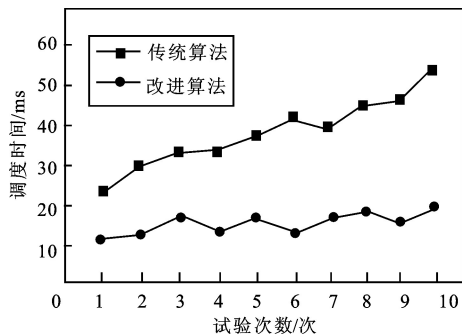


图2 不同算法调度时间对比

根据图2实验结果能够得知,随着云计算环境下的任务量逐渐增多,利用传统算法进行调度,数据节点发生冲突的概率增加增大,导致调度时间增加.而改进算法能够避免上述缺陷,从而保持了调度时间的平稳性.

将上述实验过程中的数据进行整理和分析,能够得到表1中实验结果.

表1 不同算法调度时间数据表

实验次数/次	目标数据量/个	调度时间/ms	
		传统算法	改进算法
1	1 000	23	11
2	2 000	30	12
3	3 000	35	18
4	4 000	34	13
5	5 000	38	17
6	6 000	42	13
7	7 000	39	17
8	8 000	43	18
9	9 000	44	16
10	10 000	52	19

在上述实验的过程中,系统的资源利用率能够用图3进行描述.

根据上述实验结果能够得知,随着调度任务量逐渐增大,利用改进算法进行调度的资源利用率远高于传统算法.

将实验过程中的资源利用率实验数据进行整理和分析,能够得到表2中的实验结果:

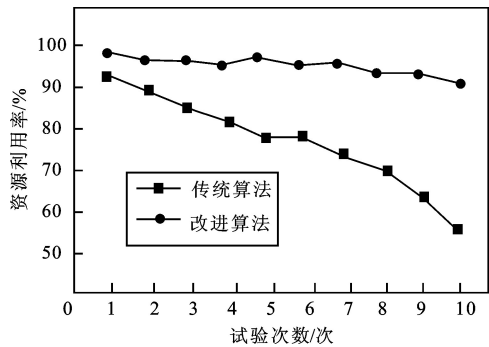


图3 不同算法资源利用率对比

表2 不同算法资源利用率数据表

实验次数/次	目标数据量/个	调度时间/ms	
		传统算法	改进算法
1	1 000	92	99
2	2 000	90	97
3	3 000	84	97
4	4 000	81	96
5	5 000	78	97
6	6 000	79	96
7	7 000	75	96
8	8 000	71	95
9	9 000	65	94
10	10 000	57	93

系统平衡度能够描述数据节点平衡性.假设系统平衡度越高,则证明数据节点的负载性越高,反之,则表明数据节点存在大量的冲突.利用图4能够描述实验过程中的系统平衡度.

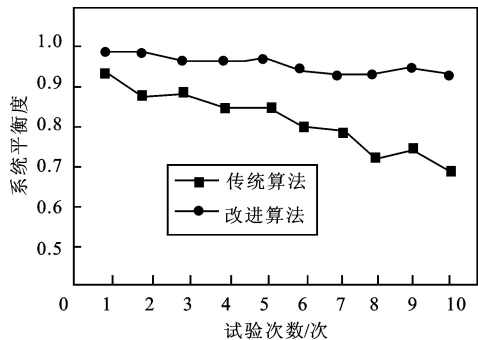


图4 不同算法系统平衡度比较

根据上述实验结果能够得知,利用改进算法进行云计算环境下海量数据调度,系统的平衡度接近1.而传统算法的系统平衡度较低.充分表明了改进算法的优越性.

4 结束语

针对传统调度方法进行云计算环境下海量数据

(下转第166页)

ID	EXPERTNAME	PROFITLE	OFFICE	DIALOGUETIME
1	张古田	副主任医师	泌尿外科	星期一上午
2	燕翔	副主任医师	泌尿外科	星期一上午
3	连惠波	副主任医师	泌尿外科	星期一下午
4	甘卫东	主任医师	泌尿外科	星期二上午、星期二下午
5	张士伟	副主任医师	泌尿外科	星期二上午、星期二下午
6	王斌	主任医师	骨科	星期一上午、星期二下午
7	邱勇	主任医师	骨科	星期一下午
8	骆东山	主任医师	骨科	星期一下午
9	王骏飞	副主任医师	骨科	星期一下午
10	朱锋	副主任医师	骨科	星期一下午
11	朱丽华	主任医师	骨科	星期二上午
12	朱泽章	主任医师	骨科	星期二上午
13	任杰	副主任医师	骨科	星期二上午、星期二下午
14	蒋青	主任医师	骨科	星期二上午
15	袁毅路	主任医师	普外科	星期一上午
16	尹卫民	副主任医师	普外科	星期一上午
17	包善华	副主任医师	普外科	星期一上午
18	徐庆祥	主任医师	普外科	星期一下午
19	桑剑锋	副主任医师	普外科	星期一下午
20	陈蓉	副主任医师	普外科	星期一下午
21	管文贤	主任医师	普外科	星期二上午
22	周铁	副主任医师	普外科	星期二上午、星期二下午
23	陈刚	主任医师	普外科	星期二下午
24	张双玲	副主任医师	普外科	星期二下午

图2 数据库关系表截图

用,而且灵活性很强及具有可扩展性,具有很强的实用价值.不足之处:一是由于一张 Excel 表格对应一个配置文件导入到一个数据库关系表中,存在数据的冗余问题,如图2的科室,坐诊时间;二是没有考虑如何抽取多张相关联的 Excel 表数据.

(上接第161页)

调度的过程中,无法避免由于调度任务量过大造成数据节点冲突,从而导致调度效率降低的弊端.为此,提出一种基于数据节点冲突避免的云计算环境下海量数据高效调度方法.将整个调度任务划分成多个子任务,对全部调度任务中的子任务量进行计算,为模型的建立提供准确的依据.根据目标数据在数据节点的传输情况,对数据节点的任务量进行计算,根据计算结果,建立云计算环境下海量数据调度模型.实验结果表明,利用改进算法进行云计算环境下海量数据调度,能够有效提高调度的效率,避免了数据节点发生冲突的情况,取得了令人满意的效果.

参考文献:

- [1] 丁警,陈国良,顾钧. 计算网格环境下一个统一的资源映射策略[J]. 软件学报,2002,13(7):1303-1308.
- [2] 吴惠中等. 基于遗传算法的网格资源调度算法[J]. 计算机研究与发展. 2004,41(12): 2195-2199.
- [3] 汪德帅,张一川,张斌. 支持多租约 SaaS 应用按需服务的负载均衡策略[J]. 沈阳:东北大学学报:自然科学

参考文献:

- [1] 王晓刚,杨春金. Excel 到 SQL Server 数据库的数据导入导出技术研究[J]. 计算机与数字工程,2007,35(3):58-61.
- [2] 孙雪梅,丁军航,原明亭. 基于 C#.NET 的 Excel 表格数据导入数据库技术研究[J]. 信息技术与信息化,2009(6):109-110.
- [3] 王喆,潘懋,郭艳军. 基于 XML 元数据和 Schema 的 Excel 信息提取研究[J]. 计算机工程与应用,2008,(33):135-137.
- [4] 王湘波,王家华. Excel 与数据库间基于语法规则的数据交换技术[J]. 微电子学与计算机,2008,25(1):37-41.
- [5] 潘兆勇. 在 NET 中自定义配置文件[J]. 中文信息,2003(8):43-50.
- [6] 谭涛,陈汶滨,陈亚丽,等. NET 远程控制 Excel 在储量管理系统的应用研究[J]. 计算机应用与软件,2007,(02):43-44+71.
- [7] 曲英杰,刘卫东,战嘉瑾. 可重构密码协处理器指令系统的设计方法[J]. 计算机工程与应用,2004,(2):56-59.

作者简介:

- 王允富 男,(1986-),硕士研究生. 研究方向为智能信息处理.
武 尚 男,(1989-),硕士研究生. 研究方向为智能信息处理.
张晓如 女,(1963-),教授. 研究方向为智能信息处理.

学版,2011,32(3):352-355.

- [4] 余永红,柏文阳. 强制数据隐私和用户隐私的外包数据库服务研究[J]. 计算机应用研究,2011,28(1):304-308.
- [5] 尚明生. 网格计算中的任务调度模型研究[J]. 计算机工程,2006,32(2):7-9.
- [6] 钟求喜,陈火旺. 任务分配与调度中遗传算子的设计[J]. 国防科技大学学报,2000,22(3):34-38.
- [7] 周文俊,曹健. 基于预测及蚁群算法的云计算资源调度策略[J]. 计算机仿真,2012,09(29):239-242.
- [8] 赵春燕. 云环境下作业调度算法研究与实现[D]. 北京:北京交通大学,2009:5-31.
- [8] 宋远骏,杨孝宗,李德毅,崔东华. 多机多任务实时系统云调度策略[J]. 计算机学报,2000,23(10):1107-1110.
- [10] 刘燕. 基于云计算信息处理系统体系结构设计[J]. 科技通报,2012,28(18):100-102.

作者简介:

- 方 晖 男,(1976-),硕士,工程师. 研究方向为计算机软件技术、教育技术、实验室管理.