云环境下的数据库扩展策略的设计

周文琼[1],王乐球[2],郑述招[1]
(1. 广东科学技术职业学院 计算机工程技术学院,广东 珠海 519080;
2. 中山大学 资讯管理学院,广东 珠海 519000)

摘 要:针对云环境下的应用系统规模越来越庞大的问题,提出了一种扩展性较好的数据库服务器扩展模型。该模型架构分为三个层次:逻辑 SQL 处理层、DA 和 CP 层、物理数据库层。采用了读/写分离策略、数据库复制、负载均衡策略、服务器群集策略等技术,提出基于虚拟节点的加权一致性哈希负载均衡算法,根据物理节点的性能权值计算分配的虚拟节点数。通过仿真实验表明,该模型在负载均衡的性能上具有优势,在数据库层具有较好的扩展性。

关键词:云计算;数据库;一致性哈希;虚拟化;负载均衡
中图分类号:TP311.133 **文献标识码**:A **文章编号**:1673-629X(2014)09-0213-04
doi:10.3969/j.issn.1673-629X.2014.09.050

Design of Database Expansion Strategy under Cloud Computing

ZHOU Wen-qiong[1], WANG Le-qiu[2], ZHENG Shu-zhao[1]
(1. School of Computer Engineering Technology, Guangdong Institute of Science and Technology, Zhuhai 519080, China;
2. School of Information Management, Sun Yat-Sen University, Zhuhai 519000, China)

Abstract: For the problem of increasingly large scale application systems under cloud computing, propose a database server extensions model with better scalability. The architecture is divided into three levels, logical SQL processing layer, DA and CP layer, physical database layers. Use a read/write separation strategies, database replication, load balancing strategies, server cluster strategies and other technologies, and propose the weighted consistent hashing load balancing algorithm based on virtual node. According to the performance weight of physical node, compute the number of virtual nodes. The simulation experiments demonstrate that the proposed model improves system performance, and has good scalability in the database layer.

Key words: cloud computing; database; consistent hashing; virtualization; load balance

0 引 言

随着云计算的兴起,云计算环境下的应用系统变得越来越复杂,系统规模越来越庞大,迫切需要构建具有高可扩展性和高性能的数据库存储系统[1-3]。可扩展性是指当系统规模/容量增大、用户增加时,系统不需要调整系统架构,仅仅需要增加或增强相应的硬件设备,就可以适应新的应用规模。

实现可扩展性一般有两种方式:垂直扩展和水平扩展,前者是指增强硬件设备,后者是指增加硬件设备的数量。

相对于 Web 服务器或应用服务器的水平扩展,数据库层的水平扩展更难实现。Web 服务器或应用服务器实现可扩展性的主要技术手段是集群和负载均衡,因为 Web 服务器或应用服务器群集里的每台服务器几乎都不需要保存状态,集群内的每台服务器之间只需要进行少量的数据传输,所以 Web 服务器或应用服务器不需要与集群中的其他同类服务器进行交互就能较快地响应客户请求;但是,数据库服务器需要保存状态,每一次用户的数据增删改请求都会导致数据库服务器的数据变化,而且数据更改状态最终需要持久化保存到硬盘。这意味着每一次数据更新将引发集群内的数据库服务器的数据同步。

针对以上问题,文中对云计算环境下的数据库扩展技术进行了研究,提出了一种可扩展性较好的数据库服务器扩展模型。

收稿日期:2013-11-04 **修回日期**:2014-02-13 **网络出版时间**:2014-07-17
基金项目:国家自然科学基金资助项目(61003253);广东省中小科技型企业创新基金(2013B011201377)
作者简介:周文琼(1969-),女,重庆人,硕士,副教授,高级工程师,CCF 会员,研究方向为信息系统、Web 数据管理。
网络出版地址:http://www.cnki.net/kcms/detail/61.1450.TP.20140717.1233.047.html

1 云计算环境下的数据库水平扩展模型

云计算应用系统具有读多写少的特征,数据库层的负荷压力主要来源于读操作,为了有效管理海量的云环境数据存储,设计了一种云计算环境下的数据库扩展系统模型。系统模型图如图 1 所示,在 Web 服务器后,系统架构分为三个层次,由上到下依次为:逻辑 SQL 处理层、DA(Database Agent,数据库代理)和 CP(Cache Pool,缓存池)层、可扩展的物理数据库层。

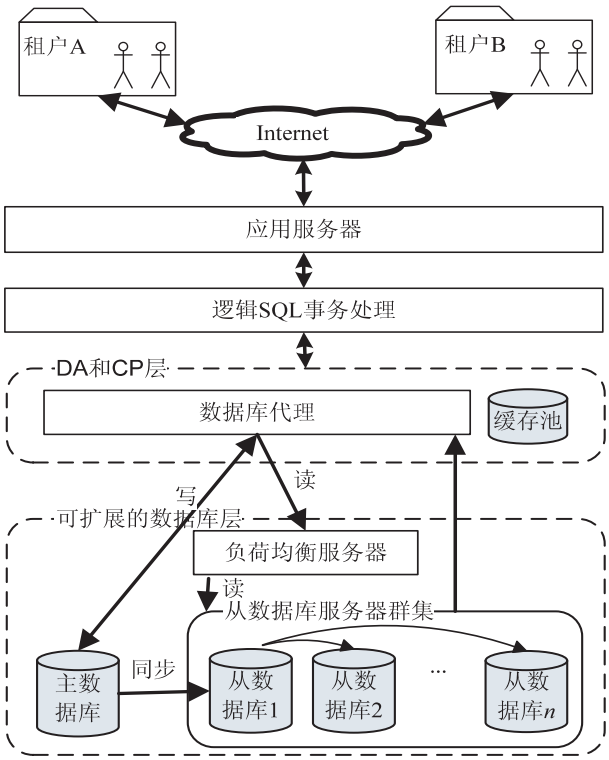

图 1 云计算环境下的数据库水平扩展模型

逻辑 SQL 处理层是系统的逻辑处理中心,对用户请求进行逻辑处理,根据接收到的包组装 SQL 语句,组装完成后将 SQL 语句包送给下一层“DA 和 CP 层”;DA 和 CP 层是本地数据代理和缓存池,数据库代理的主要功能是实现数据库连接控制和读写分离的策略,缓存池的主要功能是实现将访问数据 Cache 缓存到本地,从而缓解数据库层负荷,提高系统的数据访问性能;可扩展的数据库层采用读写分离和群集技术提升数据库层的可扩展性,该层实现的主要功能是数据库复制、数据库集群的负载均衡。

定义 1:云环境下数据库扩展模型是一个六元组,$M=(T,A,C,M,L,S)$,其中,T 表示逻辑 SQL 事务处理,A 表示数据库代理,C 表示数据缓存,M 表示主数据库,L 表示负载均衡服务器,S 表示从数据库。

1.1 实现读写分离策略

此模型中,通过数据库代理实现读写分离,数据写操作分配到主数据库上执行任务,数据读操作分配到从数据库群集上执行任务。对应用程序而言,读写分离是透明的。发生写操作后,使用数据库复制技术实现主从数据库同步问题。

从数据库群集内存储节点存有相同的数据,互为备份,这将提高系统的可靠性。

定理 1:设系统的可靠性为 R,主数据库节点设备的可靠性为 R_m,从数据库群集中的节点设备的可靠性为 R_s,节点数为 n,则系统的可靠性模型为:

$$R=R_m*(1-(1-R_s)^n)$$

1.2 数据库复制策略

读写分离会导致主从数据库不一致,主从数据库同步的主要方法是数据库复制技术,常用数据库复制方法主要有三种:快照法、触发器法、日志法。快照法将源数据库中某时点的存储对象采用快照技术保存起来,根据时点数据镜像同步目标数据库,需要对源数据库进行完全复制,效率较低;触发器法捕捉源数据库发生变化的数据,避免了完全复制,效率较高,但可能会导致主数据库不稳定和降低主数据库性能;日志法分析源数据库日志,获得源数据库的历史操作信息,通过日志归档与传递来实现数据同步,日志法是一种增量复制方式,效率较高。目前数据库数据复制技术大多采用日志复制技术,例如,Oracle 的流复制(Streams Replication)技术即日志法,被广泛应用于数据库复制。

设从数据库集群有节点 n 个,主从数据库同步的模式可以有:1 ∶ n 模式和 1 ∶ 1 ∶ (n-1)级联模式,这两种模式如图 2 和图 3 所示。前者是 n 个从数据库利用主数据库的日志完成同步,后者选择一个首从数据库,首先完成首从数据库的同步,然后,其余 n-1 个从数据库利用首从数据库的日志完成同步。

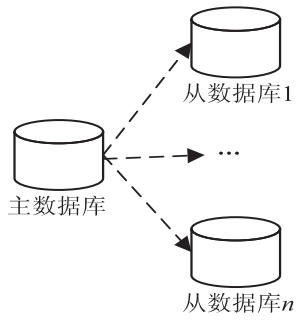

图 2 1 ∶ n 数据库同步模式

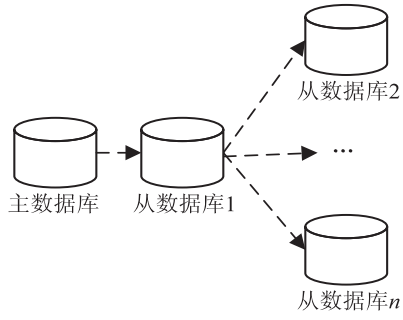

图 3 1 ∶ 1 ∶ (n-1)级联数据库同步模式

定义 2:设主从数据库同步的延时为 D,则 D 的主要影响因数是一个四元组 $D=(M,C_m,P,A_s)$,其中,M 表示同步模型,A 表示数据库代理,C_m 表示主数据库的捕获变更数据的速度,P 表示传播间隔和传输时间,A_s 表示从数据库更新数据的应用速度。

1.3　负载均衡策略

负载均衡系统性能的一个重要因素是算法[4-5],负载均衡算法划分为两类:静态算法和动态算法。静态算法易于实现,适用于同构并能预知负载量的集群系统;动态算法需要动态收集节点设备的负载情况和任务的执行特征,开销较大,适用范围更广泛。

综合考虑数据读操作的热点问题和命中率问题、节点性能差异问题,提出基于虚拟节点的加权一致性哈希负载均衡算法。

一致性哈希算法[6-8]修正了简单哈希算法,解决了热点问题,提出了在动态变化的环境中,哈希算法应该满足的四个适应条件:平衡性、单调性、分散性、负载。其原理如下:将存贮空间抽象为一个环,首先对存贮节点进行哈希计算,配置到环上节点;其次,对数据对象的 Key 进行哈希计算,按顺时针方向将其映射到最近的节点。如图 4 所示,采用哈希函数计算存贮节点 ID 的 Hash 值,将存贮节点 ID 映射到圆环形的地址空间上;计算每个数据对象的 ID 的哈希值;把数据对象映射到哈希空间:沿着圆环地址空间顺时针寻找存储节点,寻找到的首个节点确定为该数据对象的存放节点。

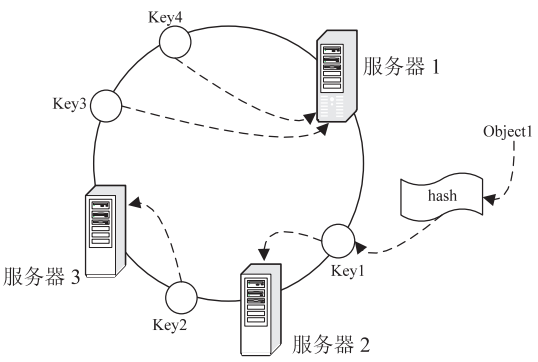

图 4　一致性哈希算法原理

但现实中数据库存贮节点不多,可能产生节点负荷不均匀。为了解决负荷倾斜问题,引入虚拟节点机制,即对每个物理节点分配多个虚拟节点,再增加虚拟节点到独立节点的映射信息,成为基于虚拟节点的一致性哈希算法。

另外,从数据库集群系统中的节点往往是异构的,在任务分配时应根据不同节点的性能区别对待,以达到能者多劳的效果,提出基于虚拟节点的加权一致性哈希负载均衡算法。

定义 3:设数据库集群有 n 个节点,$S=\{s_1,s_2,\cdots,s_n\}$,则节点 s_i 上的性能权值是一个五元组 $P=(C,F,M,D,N)$,其中,C 表示 CPU 数量,F 表示 CPU 频率,M 表示内存容量,D 表示磁盘 I/O 速率,N 表示网络吞吐量。

定理 2:s_i 节点的性能权值计算公式如下:

$$P(s_i)=k_1\times C_i\times F_i+k_2\times M+k_3\times D+k_4\times N$$

其中,$i=1,2,\cdots,n$;$\sum k_j=1$,k_j 表示各指标的权值参数,反映不同类型的服务对各个指标的影响程度。

2　关键算法

2.1　基于虚拟节点的加权一致性哈希负载均衡算法

(1)计算节点性能权值。

根据定理 2 计算全部物理节点权值,获得 $P=\{p_1,p_2,\cdots,p_n\}$。考虑到数据库服务器的性能影响侧重于 CPU 运算速度和内存,例如 k_i 设置如下:$k=(0.5,0.3,0.1,0.1)$[9],在实际的应用过程中,这些参数可以根据系统运行情况进行调节,以期达到更佳的效果。

(2)定义虚拟节点数量。

要取得比较好的负载均衡效果,在服务器数量比较少的时候需要增加虚拟节点。

设实际物理节点数为 N,则虚拟节点总数[10]为:

$$V=N*\Omega(\log N)$$

每个物理节点分配到的虚拟节点数为:

$$V(i)=V*(p_i/C)$$

其中,p_i 表示物理节点 i 的性能权值;$C=\sum p_i$。

(3)将虚拟节点和物理节点进行映射。

产生 V 个虚拟节点:根据物理节点的 IP 和虚拟节点编号,计算哈希值[11-12],将虚拟节点散布在环上,哈希值计算方法为 MD5 算法;建立虚拟节点和物理节点的映射表。

(4)预负载放置策略。

负载均衡服务器接收到读操作的 SQL 时,设 $x=$ SQL,计算 Hash(x)。在环中按顺时针方向查找第一个可用的虚拟服务节点,依据虚拟节点和物理节点的映射表,找到可服务的物理节点,将该读 SQL 查询任务预分配到该物理节点。

2.2　实际读操作分配节点

因为主数据库服务器可能成为系统性能瓶颈,所以,主从数据库同步采用日志法和 1∶1∶(n-1)级联模式。1∶1∶(n-1)级联模式需要解决主从数据库短时间不同步的问题,传统方法有两种:方法一是当主从数据库数据不一致时,则将写和读请求都发送到主数据库上,当主从数据库数据一致时,再实行读/写分离;方法二是当主从数据库数据不一致时,则读请求等待,直到同步完成,再读取数据。为了不增加主数据库的

负载,兼顾 1∶1∶(n-1)级联模式,综合应用这二种方法,即当主从数据库不同步时,首先判断从数据库节点 1 是否完成同步,如完成,则将读操作分配给从数据库节点 1;如从数据库节点 1 未完成同步,则让读操作等待。同时,在数据库同步完成后,向等待读队列通知同步完成。算法描述如下:

//获得主数据库、从数据库 1、分配节点 i 的 SCN

(1) S_m = Master. get_system_change_number;

S_1 = $Slave_1$. get_system_change_number;

S_i = $Slave_i$. get_system_change_number;

(2) if($S_i = S_m$) then //如果 S_i 完成同步

return $node_i$; //则分配节点 S_i

(3) else if($S_1 = S_m$) then //如果 S_i 不同步、S_1 同步

return $node_1$; //则分配节点 S_1

(4) else //如果 S_i、S_1 都不同步

return $null_1$; //则让读操作等待

3 系统仿真与结果分析

系统仿真配置了 1 台电脑作为 DA 和 CP 层,1 台电脑作为负载均衡器,8 台电脑组成一个从数据库集群,1 台电脑作为主数据库服务器。电脑均为 DELL 微机服务器,配置为 2.27 GHz CPU,8 GB 内存,1 G 网络适配器。

数据库服务器的软件平台为:操作系统-Red Hat Linux 9.0;数据库-Oracle11g;主从数据库同步技术-Oracle 流复制。

主从数据库基于 Oracle 流复制的数据同步构建的主要步骤如下:

(1)将主从数据库设置为归档模式;

(2)配置主从数据库相互访问参数。编辑数据库的监听配置文件 listener. ora、客户端访问文件 tnsnames. ora,配置协议、主机 IP、端口号等参数;

(3)为每个数据库设置数据库全局名称,建公共的数据库链接;

(4)在主数据库上配置 Supplemental loging;

(5)配置数据复制的相关内容:表空间、用户、私有数据库链;

(6)在主备机上分别创建队列:在主数据库上创建 Master 流队列,在从数据库上创建 Backup 流队列;

(7)在主数据库上创建 Capture 捕获进程;在主数据库上创建 Stream propagation 流传播;

(8)在从数据库上创建 Apply 应用进程;

(9)在主数据库上启动 capture process;

(10)在从数据库上启动 Apply 应用进程。

测试环境的应用系统采用工程中心研发的 OA 平台,测试数据表记录数达 500 万条记录。测试的任务分别为 Select、Delete、Update、Insert。

在不同的并发请求数下,文中提出的数据库水平扩展模型与传统的单数据库层模型在系统响应时间上进行了比较,结果如图 5 所示。在系统并发请求数逐步增加的情况下,文中模型的系统响应时间更加快捷和稳定。

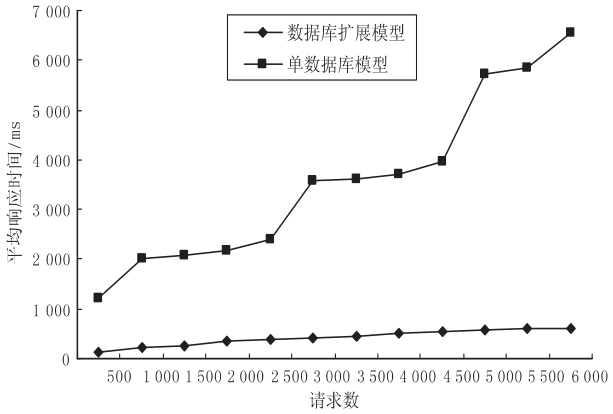

图 5 压力-响应时间比较图

4 结束语

文中对云计算环境下的数据库扩展策略进行了研究,提出了一种可扩展性较好的数据库服务器扩展模型。该模型架构分为三个层次:逻辑 SQL 处理层、DA 和 CP 层、物理数据库层,采用了读/写分离策略、数据库复制、负载均衡策略、服务器群集策略等技术,提出基于虚拟节点的加权一致性哈希负载均衡算法。通过仿真实验结果表明,该模型在负载均衡的性能上具有优势,在数据库层具有较好的扩展性[13-14]。

参考文献:

[1] 王意洁,孙伟东,周 松,等. 云计算环境下的分布存储关键技术[J]. 软件学报,2012,23(4):962-986.

[2] 史玉良,栾 帅,李庆忠,等. 基于 TLA 的 SaaS 业务流程定制及验证机制研究[J]. 计算机学报, 2010,33(11):2055-2067.

[3] 孔兰菊,李庆忠,李晓娜. 一种 SaaS 交付平台的多租户数据迁移策略[J]. 计算机应用与软件,2011,28(11):52-56.

[4] 熊安萍,刘进进,邹 洋. 基于对象存储的负载均衡存储策略[J]. 计算机工程与设计,2012,33(7):2678-2682.

[5] 陈 涛,肖 侬,刘 芳. 对象存储系统中自适应的元数据负载均衡机制[J]. 软件学报,2013,24(2):331-342.

[6] 周敬利,周正达. 改进的云存储系统数据分布策略[J]. 计算机应用,2012,32(2):309-312.

[7] 胡丽聪,徐雅静,徐惠民. 基于动态反馈的一致性哈希负载均衡算法[J]. 微电子学与计算机,2012,29(1):177-180.

[8] 周游弋,董道国,金 城. 高并发集群监控系统中内存数据

(下转第 221 页)

享。

学生信息流:学生信息的数据流转,主要以招生为权威数据源,招生的学生数据经教务进行学生分班编学号和迎新处理环节转化为在校学生,而在校学生的信息以教务为权威源进行在校信息的变更,同时通过数据中心将数据共享给其他业务系统。

4.2　数据应用分析

数据变更、使用。数据中心建立时,要充分考虑数据的按需变更和使用[13],考虑数据变化的频繁度,数据变化是阶段性的还是偶尔或者频繁、高度频繁的进行变化。如在高校应用系统之间,教职工信息的变化是偶尔的,教职工工资信息是阶段性的,新生数据是阶段性的,学生迎新是阶段性的,学生基本信息是偶尔的,学籍变动是偶尔的,学生选课是阶段性的,科研成果、项目、著作是频繁出现的,毕业论文选题、答辩是阶段性的等数据。

5　结束语

通过对高校数字化校园建设的现状分析,对数据中心建设中关键技术云计算的研究分析,提出了数据中心平台的建设内容及规划设计,进而结合高校各业务系统之间的数据流转,提出了高校数据共享的总体规划,以及数据共享中所需要的学校自编代码的实现和流转,各业务系统之间的数据共享矩阵。西安文理学院的数据中心平台已经基本形成,平台的建设及数据共享的实现,进一步验证了高校数据中心平台建设及数据共享的规划的有效性。

参考文献:

[1] Whiteaker J, Schneider F, Teixeira R. Explaining packet delays under virtualization[J]. Computer Communication Review, 2011, 41(1): 38-44.

[2] Kangarlou A, Gamage S, Kompella R R, et al. vSnoop: improving TCP throughput in virtualized environments via acknowledgement offload[C]//Proc of the 2010 ACM/IEEE international conference for high performance computing, networking, storage and analysis. New Orleans: IEEE Computer Society, 2010.

[3] 单　斌. 浅谈云计算在网络体系构建中的应用[J]. 中国新技术新产品, 2012(16): 20-20.

[4] 孟凡立, 徐　明, 张　慰. 基于云计算的高校数据中心设计与实现[J]. 现代教育技术, 2012, 22(3): 99-103.

[5] 罗军舟, 金嘉晖, 宋爱波, 等. 云计算:体系架构与关键技术[J]. 通信学报, 2011, 32(7): 3-21.

[6] 王晓虹, 刘　莹, 张艳凤. 高校数据集成系统的 ETL 设计与实现[J]. 计算机技术与发展, 2011, 21(7): 186-189.

[7] Alkis S, Dimitrios S, Malu C. Representation of conceptual ETL designs in natural language using semantic Web technology[J]. Data & Knowledge Engineering, 2010, 69(1): 96-115.

[8] 吝春妮. 数字化校园中数据中心建设与数据库安全[J]. 软件导刊, 2011, 10(4): 163-165.

[9] 杨　健, 陈春玲, 欧阳志友. 数据中心虚拟化技术的 TCP 性能分析与研究[J]. 计算机技术与发展, 2012, 22(12): 115-118.

[10] 王钱永. 数字化校园中数据中心平台的研究[J]. 中国教育信息化, 2007(11): 14-17.

[11] Wang Guohui, Eugene Ng T S E. The impact of virtualization on network performance of Amazon EC2 data center[C]// Proc of INFOCOM2010. [s. l.]: [s. n.], 2010.

[12] 孙洪睿. 高校共享数据平台的研究与实现[D]. 哈尔滨:哈尔滨工程大学, 2011.

[13] 秦剑波. 高校数字化校园建设中数据整合的探讨[J]. 科技资讯, 2007(19): 91-91.

(上接第 216 页)

库的设计与应用[J]. 计算机应用与软件, 2011, 28(6): 128-130.

[9] 周松泉. 一种新的服务器集群负载均衡算法[D]. 南昌:南昌航空大学, 2012.

[10] Karger D, Sherman A, Berkheiemer A, et al. Web caching with consistent hashing[J]. Computer Networks, 1999, 31(11-16): 1203-1213.

[11] Zeng Wenying, Zhao Yuelong, Ou Kairi, et al. Research on cloud storage architecture and key technologies[C]//Proceedings of the 2nd international conference on interaction sciences: information technology, culture and human. New York: ACM, 2009: 1044-1048.

[12] Dougherty B, White J, Schmidt D C. Model-driven auto scaling of green cloud computing infrastructure[J]. Future Generation Computer Systems, 2012, 28(2): 371-378.

[13] 周文琼, 徐　猛, 尚　敏, 等. 基于工作日历的工作流时间管理的研究与实现[J]. 测控技术, 2013, 32(7): 91-94.

[14] 周文琼, 李庆忠, 范路桥, 等. SaaS 模式多租户数据存贮模型的研究与实现[J]. 计算机科学, 2013, 40(10): 194-197.

云环境下的数据库扩展策略的设计

作者： 周文琼， 王乐球， 郑述招， ZHOU Wen-qiong， WANG Le-qiu， ZHENG Shu-zhao

作者单位： 周文琼,郑述招,ZHOU Wen-qiong,ZHENG Shu-zhao(广东科学技术职业学院 计算机工程技术学院,广东 珠海,519080)， 王乐球,WANG Le-qiu(中山大学 资讯管理学院,广东 珠海,519000)

刊名： 计算机技术与发展 ISTIC

英文刊名： Computer Technology and Development

年，卷(期)： 2014(9)