文章编号:1004-1478(2012)03-0028-03

海量并行处理系统的大数据读写接口优化研究

文必龙, 宗文栋

(东北石油大学 计算机与信息技术学院,黑龙江 大庆 163318)

摘要: 针对MPS系统中数据文件管理混乱、文件读写速率低的问题,采用道集流方式进行数据流程控制,并使用超时机制、数据库连接池和内存映射等技术对系统进行优化. 测试结果表明,优化后的系统数据管理有序、文件读写速度显著提高.

关键词: 数据库文件管理; 道集流; 超时机制; 连接池; 内存映射

中图分类号: TP311 **文献标志码:** A

Research on interface optimization for reading and writing large data of massive parallel processing system

WEN Bi-long, ZONG Wen-dong

(*School of Comp. and Infor. Tech.*, *Northeast Petroleum Univ.*, *Daqing* 163318, *China*)

Abstract: Aiming at the problems such as disorderly management of data file and low literacy rate in MPS, the system was optimized adopting the way collecting pattern for data flow control, and using the timeout mechanism, database connection pool and memory mapping technology. Test results showed that data managed orderly and file reading and writing speed increased in the optimized system.

Key words: database file management; way collecting; timeout mechanism; connection pool; memory mapping

0 引言

地震资料并行处理系统MPS(marine seismic data parallel process system)是中海石油研究中心与中科院地球物理所联合开发的,DHT是MPS的内部数据格式,它内容丰富、结构清晰、检索速度快[1]. 但在MPS中,地震数据的输入、输出和处理采用串行方式来完成,只有数据全部输入之后,处理才开始执行,因此输入、输出操作在系统处理过程中占用大量的时间; 系统采用暂存盘方式来存放临时数据,而地震数据庞大,因此需要大容量的磁盘来暂存数据,硬件配置要求非常高; 另外,由于磁盘读写方式较慢,采用暂存盘方式存放数据,会造成处理模块之间通信缓慢,运行效率低下[2]. 为了避免以上不足,本文对MPS系统进行改进,采用较为先进的道集流方式进行数据流程控制,并引入超时机制和数据库.

1 实现流程

通过与维护自定义打开文件表设置文件运行

收稿日期:2012-02-12

基金项目:国家863计划,国家重大专项(2006AA09A102—15)

作者简介:文必龙(1967—),男,湖北省仙桃市人,东北石油大学教授,主要研究方向为软件工程和集成技术.

超时机制来实现方案优化.

1.1　打开文件流程

文件描述表可以实时记录一个进程所有打开的文件列表，而且可以查询文件的超时状态；另外在性能方面，可以避免一个进程重复打开同一个文件，尤其是在网络环境中，频繁打开关闭文件会增加网络负载，降低程序运行效率. 维护一个打开文件描述符表，可以在一定程度上缓解网络负载、运行效率问题. 打开文件流程如图 1 所示.

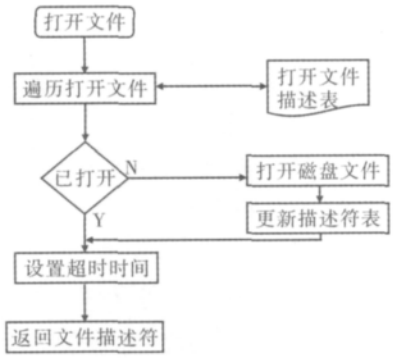

图 1　打开文件流程图

遍历打开文件表，是查询正要打开的文件是否已经被打开并尚未关闭，如果已经打开且尚未关闭，则直接将该文件描述符返回，并重新设置超时时间，置位 OPEN 位；如果正要打开的文件没有打开，则需要调用系统调用，如用 open( ) 去打开磁盘文件，并将成功返回的文件描述符追加到打开文件描述符表，同时设置超时时间，置位 OPEN 位.

1.2　设置文件超时流程

为了避免进程之间产生竞争条件，需要为打开文件描述符表中的每一项设置一个超时时间，当经历时间超过这个时间后将视该项超时，并可以根据情况关闭该文件. 运用超时机制可以防止其他进程因要打开的文件被占用而长时间阻塞，如果不设置超时，则文件一直处于打开状态，会影响系统内其他进程的正常运行. 设置超时时间流程如图 2 所示.

针对程序的运行类型(并行/串行) 的不同，超时时间的设置也是不同的. 在并行状态时，进程间产生竞争条件的几率要大于串行状态，所以并行时超时时间不应设置过大，否则会降低其他进程的运行效率；同理串行的超时时间可以适当调大一些，至于应该设置成多少，可能需要进一步确定. 图 2 中串行/并行超时时间比例为

$$T_s = 2T_p$$

其中，T_s 为串行超时时间，T_p 为并行超时时间.

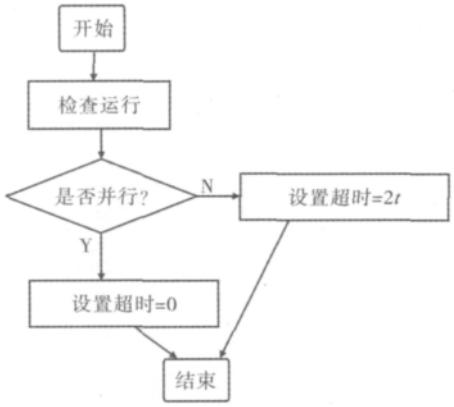

图 2　设置超时时间流程图

1.3　关闭文件流程

文件总是要关闭的，否则其他进程可能会受到影响. 在打开文件描述符表项中，uMask 中有几位是用来记录文件状态的，需要检查这几个状态后再决定是否关闭文件.

1) OPEN，CLOSED，TIMEOUT 均已置位，此时应该关闭文件.

2) OPEN，TIMEOUT 均已置位，但 CLOSED 尚未置位，此时说明进程可能仍在操作文件. 若时间过长导致文件超时，需要等待用户调用关闭文件接口，如调用 myClose( ) 关闭文件.

对应文件的状态位，可以定义几个符号常量，用以检查状态位：

```
#define FG_OPEN　0x0400
#define FG_CLOSED　0x0800
#define FG_TIMEOUT　0x1000
#define FG_OPEN_CLOSED　FG_OPEN | FG_CLOSED
#define FG_OPEN_TIMEOUT　FG_OPEN | FG_TIME-
OUT
#define FG_ALL　FG_OPEN | FG_CLOSED | FG_TIME-
OUT
```

2　关键技术

由于同一地震数据在不同节点上同时进行，也就是说，针对同一个地震数据需要形成多个作业. 而现有作业只描述处理，不描述输入、输出数，已经不能满足数据并行处理的需要，需要对作业的格式进行改进. 新系统中，数据的处理不再以文件为单位，而是以道集为单位进行，因此系统需要数据读

写接口能够以道集方式提供数据[3]. 为了有效地提高海量数据的读写速度,在新系统中引入了数据库,通过数据库连接池和内存映射技术将底层读写接口重新封装.

2.1 数据库连接池

在应用程序运行时,需要多次访问 MySQL 数据库,而在每次访问数据库之前首先要与数据库服务器建立连接. 然而在整个数据操作过程中,连接数据库的操作却要占用很大一部分时间,这势必导致服务器过载,运行效率低下等问题[4-5].

数据库连接池可提高多用户访问数据库的效率. 所谓数据库连接池就是将可用的连接存储于一个缓冲池中,当用户程序需要连接数据库时,先要查看缓冲池中是否有可用的连接,如果有则直接返回一个可用的连接,只有在没有可用的连接时才真正地连接数据库并返回数据库连接对象.

然而多数数据库管理系统只能提供服务器端的数据库连接池,即判断是否有可用连接的操作是在服务器端执行,虽然这样可以减少实际连接数据库的次数,但用户程序仍需要通过网络来与数据库建立联系. 因此,实际上很多应用程序都是采用客户端数据库连接池方法来实现这个功能,操作流程如图 3 所示.

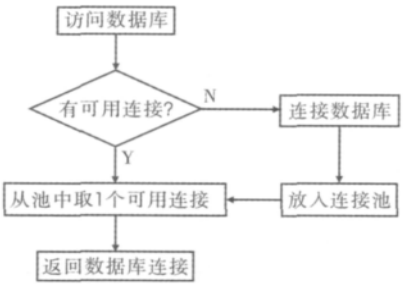

图 3 客户端连接池操作流程图

从图 3 可以看出,这个连接池模式与服务器端连接池的工作原理一样. 但是客户端的连接池是存放于客户端的,这样用户程序在连接数据库时,第一步是验证客户端的连接池,程序就避免了通过网络连接数据库服务器的操作,从而节省了网络资源和运行时间,提高了效率[6].

事实上,可以同时运用 2 种数据库连接池,以便最大程度上提高网络利用效率和缩短程序运行时间.

2.2 内存映射快速读写机制

数据存取接口从道号集计算道数据偏移量,然后从磁盘读取数据. 按一般文件读写操作,数据先读到内存缓冲区,再被程序读入. 当程序要读的数据不在内存缓冲区时,再读写磁盘. 这样频繁读写磁盘,会导致读写效率低下. 因此借鉴内存映射文件技术与虚拟存储管理思想,将基于虚拟映射的文件读取模型引入 MPS 中.

内存映射文件技术是操作系统提供的一种新的文件数据存取机制. 利用内存映射文件技术,系统可以在 4 GB 的地址空间中为文件保留一部分空间,并将文件映射到这块保留空间. 一旦文件被映射之后,操作系统将执行管理页映射、缓冲以及高速缓冲等任务,而不需要调用分配、释放内存块和文件输入/输出的 API 函数,也不需要自己提供任何缓冲算法[7]. 另外当遇到大数据量文件时,内存映射文件技术能够分配一块足够大的内存来满足请求.

在 MPS 系统中,通常地震数据都比较大,有的已达 T 级,所以一般不宜一次将整个文件全部映射到地址空间中去,因此采用虚拟内存管理的方式,根据文件访问具有局部性原理,一次只将文件的一部分进行映射,如需要访问的数据已被映射则直接使用,否则重新映射. 由于减少了磁头定位参数计算次数和磁头定位的次数,文件访问的效率会得到很大提高.

3 优化结果

采用内存映射访问模型对程序进行测试,并与 FOCUS 进行比较,结果见表 1. 实验采用 Red Hat Enterprise Linux 4.0 的操作系统,MySQL 数据库,数据大小为 4.17 G,处理道集 240 个,其中每道集含 1 159道,数据道长 6 s,样点数 3 000,采样率 2 ms. 从表 1 中可以看出采用内存映射的 MPS 的文件访问效率提高了很多.

表 1 对比结果 min

服务器	MPS1.0 (传统方式)	MPS2.0 (内存映射)	FOCUS
SERVER1	23.37	8.50	9.32
SERVER2	43.25	43.25	43.02

4 结论

本文对 MPS 系统进行了优化. 采用道集流方式

(下转第 48 页)

4 结语

数字图书馆作为图书馆未来的发展趋势,将会拥有海量的数字资源. 图书馆的目标就是要充分发挥这些数字资源的作用,而避免信息过量. 本文基于数据挖掘技术,通过数据清理、数据整合,并加上规约算法,对图书馆信息管理的数据进行挖掘和预测,实验结果表明,数据挖掘的结果不仅能为图书馆的业务管理提供数据参考,而且能指导传统图书馆管理员的日常工作. 从而有利于图书馆调整图书管理策略,进一步提高服务质量和管理水平.

参考文献:

[1] Littman M L, Dumais S T, Landauer T K. Automatic Cross-language Information Retrieval Using Latent Semantic Indexing [M]. Belongia: Kluwer Academic Publishers, 1998: 15 - 21.

[2] Vinokourov A, Shawe-Taylor J, Cristianini N. Inferring a Semantic Representation of Text Via Cross-Language Correlation Analysis [M]. Cambridge: MIT Press, 2002: 91 - 98.

[3] 孙权森,曾生根,王平安,等. 典型相关分析的理论及其在特征融合中的应用 [J]. 计算机学报,2005,28(9):1524.

[4] Agrawal R, Imielinski T, Swami A. Mining association rules between sets of items in large databases [C] // Proc of the 1993 ACM SIGMOD Int Conf on Mana of Data, Washington DC: ACM Press, 1993: 207 - 216.

[5] 郭力平,雷东升,冷永杰,等. 数据库技术与应用 [M]. 北京: 人民邮电出版社,2007.

[6] 李静. 数据仓库中的数据粒度确定原则 [J]. 计算机与现代化,2007(2):57.

(上接第 30 页)

进行数据流程的控制,超时机制和文件描述符表的引入,提高了文件打开的速度; 数据库连接池技术的使用,提高了网络利用效率,缩短了程序运行时间; 内存映射机制的加入,提高了文件访问的效率.

参考文献:

[1] 常玉连,朱保国,任福深,等. 面向油田地面工程系统的数据库接口软件设计 [J]. 油气田地面工程,2003,22(8):62.

[2] 凡哲元,邢绍献,苏映宏,等. 油田开发规划优化决策系统研究 [J]. 油气地质与采收率,2003,10(6):34.

[3] 梁达平. 试析大型制造业 ERP 软件数据库性能优化技巧 [J]. 甘肃科技,2008(10):24.

[4] 陈悦,白杰,王林. 软件项目开发的性能优化 [J]. 微处理机,2009,30(3):99.

[5] 关晓晶,魏立新,杨建军. 基于混合遗传算法的油田注水系统运行方案优化模型 [J]. 石油学报,2005,26(3):114.

[6] 赵改善,孔祥宁,王于静,等. 64 位集群计算平台波动方程叠前深度偏移的性能优化 [J]. 勘探地球物理进展,2005,28(1):57.

[7] 杨存祥,张晓辉,石军. 基于 DSP 和 FPGA 的油田测井系统总线通信接口设计 [J]. 仪表技术与传感器,2010(4):67.