

# PC 服务器实现海量数据存取的方法\*

## The Approaches of Massive Data Access Based on PC Server

郭 荣<sup>1</sup>, 杨 颖<sup>2</sup>

GUO Rong<sup>1</sup>, YANG Ying<sup>2</sup>

(1. 柳州市百货股份有限公司, 广西柳州 545001; 2. 广西大学计算机与电子信息学院, 广西南宁 530004)

(Liuzhou Department Store Ltd., Liuzhou, Guangxi, 545001, China; 2. School of Computer, Electronic Information, Guangxi University, Nanning, Guangxi, 530004, China)

**摘要:**在 HP LH6000 PC 服务器(双 PIII550 Intel 至强 CPU、1G 内存、Raid5 磁盘阵列)、WINNT4.0 网络操作系统和 Microsoft SQL Server 2000 数据库管理软件环境下,提出一种应用 PC 服务器实现海量数据存取的方法,该方法通过改进和优化数据分类、数据分割、数据加工和保存、数据访问等一系列措施解决海量数据存取的技术难题,较好地提高了数据库服务器系统的处理速度和效率,节约了投资和维护成本。

**关键词:**数据 数据库 数据表 系统

**中图分类号:**TP311.132 **文献标识码:**A **文章编号:**1002-7378(2009)01-0033-03

**Abstract:** By using the hardware of HP LH6000 PC Server, the network operation system of WINNT4.0 and the database management system of Microsoft SQL Server 2000, a method of massive data access is offered. It can solve the technical difficulty for massive data access through improving and optimizing data classification, partition, processing and store. At the same time, it can quickly improve the speed and efficiency of system process in database server, and greatly save the investing and maintenance cost.

**Key words:** data, database, table, system

管理信息系统(MIS)/企业资源计划(ERP)最大的特点和优点是数据积累。数据的积累不仅是对业务活动过程的记录,及时反映业务经营状况,还能从中归纳出有规律的信息,对以后的业务活动起到很好的指导作用,并且能够规范经营管理活动,提升管理水平。然而,随着数据的积累,数据表(Table)和数据库(DataBase)变得日渐庞大<sup>[1]</sup>。数据访问时,随着检索范围加大,查询效率会显著降低,因此系统处理和响应的速度越来越慢。当系统响应慢到一定程度时,数据处理的时效性就不能满足要求,甚至导致数据丢失。同时,用户也将失去等待的耐心而逐步放弃使用系统,使系统失去存在的意义<sup>[2]</sup>。因此,对信

息管理系统来说,随着使用时间的增长,如何保证系统响应速度保持在用户可以接受的范围内,是 MIS 系统设计开发中普遍面临的技术难题。

当前,很多管理信息系统开发者为了回避这个技术难题,普遍要求使用单位配置硬件层面上数据存取能力极强的小型机<sup>[3]</sup>。这样一来,使用单位将要增加几十万乃至几百万元的初始硬件投资,而且配套的软件平台投资和后继的培训、使用、维护成本也将大大增加。本文根据作者多年管理信息系统开发的研究和实践经验,提出一种应用 PC 服务器实现海量数据存取的方法,通过改进和优化数据的存取解决海量数据存取的技术难题,为信息系统使用单位节约投资和维护成本。

### 1 数据处理环境

数据处理环境为某大型商厦的企业资源计划管理系统,该系统有近 300 个使用站点,每年的业务数

收稿日期:2008-06-10

作者简介:郭 荣(1971-),男,高级工程师,主要从事电子商务和数据库技术研究。

\* 国家“863”高技术研究发展计划基金项目(2002AA4Z3430),广西自然科学基金项目(桂科青 200731023),广西教育厅桂教科研项目(200626)资助。

据量约 2G。该企业资源计划管理系统采用 HP LH6000 PC 服务器(双 PIII550 Intel 至强 CPU、1G 内存、Raid5 磁盘阵列),网络操作系统采用 WINNT4.0,数据库管理软件采用 Microsoft SQL Server2000。

## 2 PC 服务器实现海量数据存取的方法

PC 服务器实现海量数据存取的方法主要包括数据分类、数据分割、数据处理算法和数据访问 4 个部分。

### 2.1 数据分类

为了减少数据的冗余,防止存储异常,提高处理速度,在开发过程中,针对各类管理对象都进行编号处理,并保存在不同的“对象定义表”中。同类对象放在同一个定义表中,编号不重复,并用不同的字段定义管理对象的属性。记录和加工业务数据时,只处理对象编号和业务数据,把结果放在不同的“业务数据表”中。在进行数据检索和存取时,通过数据视图(VIEW)或直接进行“对象定义表”和“业务数据表”的联接。根据业务数据被加工汇总的程度,又进一步把“业务数据表”划分为不同的“原始业务数据表”、“按日汇总业务数据表”、“按月汇总业务数据表”。经过记录和加工的业务数据,有部分是用来指导和规范业务活动的,把这部分数据分别存放在不同的“业务管理表”中。数据分类结构如图 1 所示。

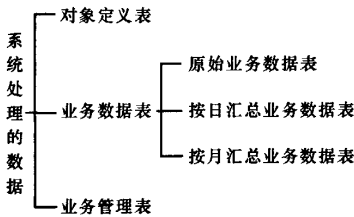


图 1 数据分类结构

### 2.2 数据分割

在使用系统过程中,用户最关心的数据是当月的业务数据,其次是当年的业务数据,再次是往年的业务数据。查询业务数据的频率从高到低排列,依次为当月、当年、往年,而且查询的时间段极少跨年度。根据这个规律,在数据库设计中,把“业务数据表”按年度分割存放,每年的“业务数据表”放在一个数据库(“年度数据库”)中,同时对当月的“业务数据表”另外放在一个数据库(“当前数据库”)中,这就保证了每个数据库和每个表的内容不会无限增长。

“对象定义表”中的各个字段值具有相对稳定的特点,因此也把它们按年度来保存。每个“年度数据

库”中保存一份对应的“对象定义表”,当年的“对象定义表”保存在“当前数据库”中。

“业务管理表”存放的是一些规则和规律,数据量不大,但是要求被查询时能迅速返回结果,而且这些规则和规律只对正在发生和将来发生的业务起作用,因此把它们放在“当前数据库”中。数据分割存放结构如图 2 所示。

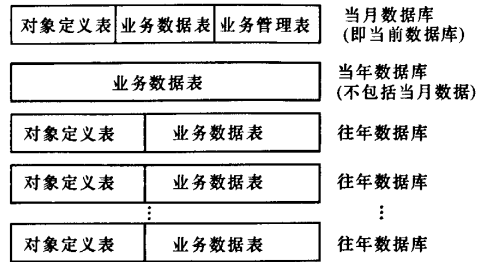


图 2 数据分割结构

### 2.3 数据加工和保存

在实际的数据访问中,作业层工作人员最关注原始数据和流水清单数据,而管理层和决策层人员最关注按日或按月汇总的业务数据报表。大部分用于管理和指导的数据也是要求按日或按月汇总。因此,很多数据不是需要使用时才从原始数据来直接汇总加工,而是可以采用“以空间换时间”的原则,在适当的时间,事先进行数据的按日或按月汇总,生成对应的汇总数据表。利用这种规律,在受理作业层工作人员对原始数据和流水清单的要求时,可以直接从“原始业务数据表”中访问检索;同时,在受理管理层和决策层人员的业务数据报表要求时,避免了对大量原始数据的直接访问检索和加工汇总,可以针对性地从“按日汇总业务数据表”或“按月汇总业务数据表”中的少量数据中检索汇总,这样大大地提高了系统响应的速度和效率。

具体的数据处理算法如下:(1)每天新增的管理对象和对管理对象的修改直接保存在“当前数据库”的各个“对象定义表”中。(2)在每日规定的结算时间或业务活动结束后,进行每日数据结算。对原始业务数据进行加工整理,生成各类按日汇总业务数据,保存在“当前数据库”的各个“按日汇总业务数据表”中,同时更新“业务管理表”。(3)在每月月底的月报截止日,进行月份数据结算和转存。把当月的“业务数据表”分类汇总,生成按月汇总业务数据,保存在“当年数据库”的各个“按月汇总业务数据表”中。(4)把“当前数据库”中的各个“按日汇总业务数据表”转存到“当年数据库”的对应“按日汇总业务数据表”中,并且清空“当前数据库”中的各个“按日汇总

业务数据表”。(5)在每年年底,进行年度数据转存。把“当前数据库”中的各个“对象定义表”复制一份到“当年数据库”中,生成新一年的数据库表结构。(6)将“当前数据库”作为新一年的“当月数据库”,删除“对象定义表”中已经淘汰的管理对象,整理“业务管理表”。

#### 2.4 数据访问

因数据是按年度存放的,在访问数据时,为使处理简单,约定访问的时间段范围不能跨年度。这样在访问数据前,首先判断要访问哪类数据。若为管理数据,则直接从当月数据库中的“对象定义表”和“业务管理表”中查找。若为定义数据或业务数据,则调用一个公用存储过程,输入参数为访问时间段的起始日期和截止日期,返回“对象定义表”和“业务数据表”所在的数据库名(若访问时间正好跨当年当月和当年往月时,则要另外返回一标记说明业务数据跨数据库)。而后,在指定的数据库中访问对应的数据表。

### 3 结束语

本文通过改进和优化数据分类、数据分割、数据加工和保存、数据访问等一系列措施,仅利用常用的 HP LH6000 PC 服务器就成功地实现了某大型商厦

企业资源计划管理系统海量数据的存取。目前该系统已经运行五年多,使用该系统汇总商厦一级部门一年的业务数据只须几十秒钟,绝大多数报表能在十几秒内完成,单据录入也没有停顿现象,较好地解决了使用 PC 服务器实现海量数据存取的技术难题,既避免了困扰用户的系统越用越慢现象,又能够将数据方便的按年度备份归档,提高了系统的性能和可用性,使系统处理和响应速度大大加快,保障了系统的良好性能。同时还节省了大量的硬件投资和日后的维护费用。这些改进的方法和措施,可以为其他管理信息系统的设计提供较好的借鉴作用。

#### 参考文献:

- [1] 戴子良. 海量数据处理分析[J]. Windows IT Pro Magazine: 国际中文版, 2007(1): 37-39.
- [2] 张金乙, 姜文志, 蒋伟俊, 等. 高速海量数据的接收和存储系统的设计与实现[J]. 计算机时代, 2007(12): 67-68.
- [3] 孟清, 杨树强, 蔡建宇, 等. 面向海量数据统计分析查询的中间层语义缓存[J]. 计算机工程, 2005(20): 67-69.

(责任编辑:韦廷宗)

## 我国成功开发出氟碳铈矿清洁冶金集成技术

稀土矿矿藏中除含有大量宝贵的稀土资源外,还分别含有 0.2%~0.3% 的放射性元素钍和 8%~10% 的氟。长期以来,该矿主要采用氧化焙烧——盐酸优溶工艺。该工艺存在流程长、稀土收率低、钍和氟资源大量流失的缺点,同时对环境也造成严重污染。

中国科学院长春应化所的研究人员经过多年的研究,先后突破了萃取体系和萃取剂的遴选、工艺流程的开拓、稀土萃余液中钍和稀土分离等系列技术关键,发明了一种从氟碳铈矿浸出液中萃取分离钍、钍的新工艺,并应用这种具有我国自主知识产权的核心工艺,成功开拓出氟碳铈矿清洁冶金与分离制备一体化集成技术。与现有的氧化焙烧——盐酸优溶流程相比,该集成技术在国际上首次从生产源头出发解决了放射性废渣和废气排放带来的环境污染;生产过程不排放氟气;原流程中作为污染源排放的氟、钍得到了有效回收,回收率超过 90%;稀土收率由 80% 提高到 90% 以上,铈产品纯度由 98% 提高到 99.999% 以上,有效实现了氟、钍、稀土的高效回收和稀土资源的清洁冶金及高值化。

(据科学网)