

编者按 关系型数据库在处理问题时将事物分解为二维表, 往往要进行多表联合查询, 不但增加成本, 而且影响系统性能。Caché数据库是一个用于高性能事务应用的后关系数据库管理系统, 具有面向对象的许多功能和一个事务型多维数据模型, 提供了比关系型技术更加高效的性能, 更大的扩展性, 更快速的编程能力和更加便捷的使用性能。笔者分析了Caché数据库的特性, 并对Caché数据库在HS中的应用优势进行了较为详尽的分析与解释, 对农业中植物保护、畜禽病害防治等多功能数据库的建立具参考意义。

基于 Caché 数据库的医院信息系统的研究

王建文 袁伟 (陕西科技大学计算机系, 陕西西安710021)

摘要 医院信息系统(HS)在医院管理及现代化建设中发挥越来越重要的作用。采用传统关系型数据库的HS产品日益显示出种种不足,如响应速度太慢,维护费用昂贵。Caché数据库是一种面向对象型数据库,具有许多卓越的性能,不但加快了软件开发的速度,还解决了“阻抗失配”问题。另外由于Caché采用多维数组的存储方式以及位图索引等技术,使其在与关系型数据库的HS对比中,展现出更大的优势,如响应速度加快,降低硬件成本,日常维护更加容易。

关键词 医院信息系统;面向对象;Caché

中图分类号 TP274 文献标识码 A 文章编号 0517-6611(2007)21-06685-03

Research on Hospital Information System Based on DataBase of Caché

WANG Jian wen et al (Department of Computer, Shaanxi University of Science & Technology, Xi'an, Shaanxi 710021)

Abstract Hospital Information System(HS) plays an important role in management of hospital and in modernization construction. Product of HS which adopts traditional Relational- DataBase increasingly reveals many shortages, for example: its response is too slow and cost of maintenance is expensive. Caché DataBase is an object-oriented DataBase, and it has many excellent performance. Not only it quickens speed of developing software, but also has solved problem of incompatibility of impedance. Besides, because Caché adopts mode of storage of multi-dimension and Bitmap Index, it makes Caché show even more superiority in contrast with Relational- DataBase, for example: response becomes quicker, cost of hardware is low and daily maintenance becomes easier.

Key words HS; Object-oriented; Caché

近年来,各个医院采用关系型数据库相继建立各自医院的信息系统(HS),由于医疗行业的特殊性和复杂性,运行不久就会逐渐显示出各种弊端。Caché数据库是美国InterSystems公司开发的世界独一无二的后关系型数据库。其优势在于加快开发时间(1/3),减少营运护理成本(30%~70%),提高读数据速度(7~12倍),具有处理大型复杂数据的优点。

1 关系与对象的矛盾:阻抗失配

目前很多软件在开发时,依然采取前台使用面向对象技术,后台选取关系型数据库的模式,这样就会产生一个“阻抗失配”问题。“阻抗失配”产生的原因是对象模型与关系模型之间缺乏固有的亲合力,所带来的问题包括:类的层次关系必须绑定为关系模式ID生成等。

1.1 类的层次绑定为关系模式 如果要将前台的面向对象协调地运行在关系型数据库上,一个重要的任务就是需要将类转换为1个或多个底层的表格。这个任务是将类中的每个属性对应成为表格中的列。如果类和表格能够一一对应,这种转换就十分简单,但是大多数的对象模型都需要把多个表格进行分解,并从中抽取一部分作为类的属性。还可能有很多类都会映射到相同的表中并根据自身类的需要对表的语义进行解释。

1.2 ID生成——生成对象ID并将其插入关系数据库 对象ID的生成过程十分重要,在关系数据库中通过Save()方法

生成不重复键时需要Object ID。在一一对应的类与表之间,最好的方法就是使用关系数据库引擎生成一个ID号,并将它作为对象的ID。但是一个类可能会跨越多个表,Save()方法可能需要把多个关系ID作为一组来调用。生成的ID必须是持久性的,这样才能在打开对象时获取表中的行来生成实例。因此,为了维护对象ID和关系ID的联系必须另外创建一个表。Caché是面向对象型数据库,这样前台与后台能很好地融合,无须再将类的层次绑定为关系模式,可以从根本上解决“阻抗失配”问题。

2 Caché数据库简介

关系型数据库把数据表示为简单的二维模型,而后关系型数据库采用的是独特的多维数据结构,这不仅能真实地反应和更好描述现实世界的复杂数据及其之间的联系,同时也使数据的存取能更快地实现。Caché就是一个用于高性能事务应用的后关系型数据库管理系统,该系统具有面向对象的许多功能和一个事务型多维数据模型。Caché整合了对象数据库访问、高性能的SQL访问和直接对多维数组访问。强大的多维数据访问,使这三种方法能够同时访问相同的数据,数据只要在单一的整合数据字典中描述1次,就可以被这三种方法访问。Caché提供了比关系型技术更加高效的性能、更大的扩展性、更快速的编程能力和更加便捷的使用性能。

3 关系型数据库的HS与Caché的HS的比较

3.1 基于Caché的数据库有更快的响应速度 关系型数据库在处理问题时将事物分解为二维表,然而现实世界约80%的事物是非结构化的。如果采用关系型数据库技术,一个简单的事物可能要被拆分为很多张表,在处理查询时,往往要

基金项目 陕西省教育厅资助项目。

作者简介 王建文(1947-),男,河南郑州人,教授,从事软件工程与数字图像处理研究。

收稿日期 2007-04-06

进行多表联合查询,需要很大的开销,从而影响系统的性能。由于 Caché 数据库面向对象的特性而无须进行多表的连接和在多个表之间进行来回跳转,在很大程度上降低了系统的开销,从而加快了系统的响应速度。

另外,一个成功医院的 HIS 系统应该具备满足联机事物处理(OLTP)高速响应的功能,即满足 OLTP 高速响应的要求。Caché 支持动态位图索引技术,位图索引技术是一种用一系列位串表示对于一个给定的索引值的对象 ID 集,位图索引适合 OLTP 的索引结构,是提高数据库性能的重要措施。由于位图索引的特性,当字段的值的分布非常不均匀时,即只有少数可能值时,这时可以采用位图索引。反之,建立常规的索引即可提高查询速度。所以将位图索引技术用于 HIS 系统时,用户能够快速搜索大型数据库,通常查询数百万条记录只需 1 s。

3.2 基于 Caché 的数据库能显著降低硬件的费用 目前基于关系型技术的 HS,一方面由于关系型数据库建表时,在没有数据的地方也要预留和占用空间,浪费了大量宝贵资源,增加了本来就不必要的硬件成本。另一方面,由于数据库的内部机制导致的系统处理速度慢,而不得不采用更高性能的硬件来弥补这一缺陷,无形中也增加了系统的硬件成本。相对而言,由于 Caché 数据本质上是可变的,并且存储在稀疏数组中,由于稀疏数组只存储非零元和必要的行号、列号,这样会节省大量的空间,大约只需关系型数据库 1/3 的空间,除了降低磁盘空间的要求,紧密的数据存储还提高了性能,所以基于 Caché 的应用程序可能比关系型的应用程序快 20 倍。此外,使用 Caché 数据库时不需要第二台计算机服务于数据仓库或者决策支持,也不需要每天把数据迁移到第二个系统,这样也会降低硬件的费用。与大小相同的系统相比,基于 Caché 的应用对硬件的要求很低,即使大量低成本的计算机也可以联合成一个非常强大的系统以支持大量的事物处理。

3.3 基于 Caché 的数据库日常管理更加容易 Caché 的日常管理非常方便,不仅可以自动化完成备份工作,还可以免掉一些对其他数据库而言一定要做的日常工作。传统的关系型数据库,数据库管理员(DBA)通常有很多任务要做,比如 DBA 要针对不同的应用制定不同的转储计划,以保证一旦发生故障能尽快将数据库恢复到某种一致的状态;关系型数据库运行一段时间后,由于记录不断增、删、改,会使数据库的物理存储情况变坏,降低数据存储效率,数据库性能下降,这些都需要 DBA 进行重新组织或部分组织。由于 Caché 采用多维数据模型,支持复杂数据的简单存储,并且不需要对数据的声明和定义,记录不断的增、删、改对其影响很小。Caché 还具备在线备份功能,通过簇等技术支持 24 h 连续运行或无缝地从故障中恢复。可见,Caché 只需便捷的系统管理,因而可减少故障产生和降低维护管理费用。

4 Caché 数据库在 HIS 中的应用

4.1 快速应用开发 对象技术是提高编程效率简单而有力的工具。由于 Caché 数据库采用面向对象的特性,开发者能够使用简单而实际的方法来思考和使用对象。HS 系统业务流程极为复杂,采用面向对象可以很大程度上加快应用程序

开发的进程。同时,对象本身的模块化和交互性能使得应用程序维护变得相对简单。此外,Caché 数据库使用对象技术,也可以从根本上解决“阻抗失配”问题。

Caché 应用的核心是非常快速的 Caché 虚拟机,为此 Caché 本身提供自带的脚本语言 Caché Object Script(COS),COS 对象脚本语言是一种以面向对象为基础的数据库设计语言。COS 对象的建立和函数的编写,同目前通用的面向对象语言的编程语言是类似的。所以使用 Caché 能很大程度上加快开发应用程序的速度。

4.2 独特的多维数据模型 由图 1 可见,在处理真实世界中的数据元素时,关系数据库以做表为主,将数据中的关系以二维的表格来描述。Caché 的多维数据存储方式是以倒树行的完全符合自然界的数据结构进行处理,保持其多维性。这种结构能存储非常丰富的数据,通过内置的 COS 语言能够直接访问多维数组结构,能够获得更高的性能和更好的存储利用率。

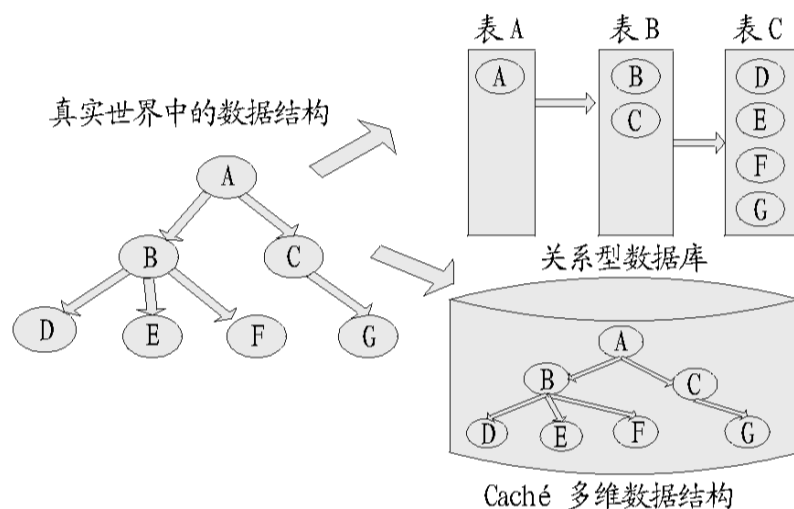


图1 关系型与对象型数据库的不同

Caché 的多维数组结构被称为“global”，带有许多下标的数据可以存储在 global 中。而且这种下标是没有类型限制的，能够存储任何类型的数据。

如图 2 所示,一个库存应用程序提供的信息有物品、大小、颜色、类型、数量。它的数据结构可能是这样(对 global 的操作要用到下标“^”)：

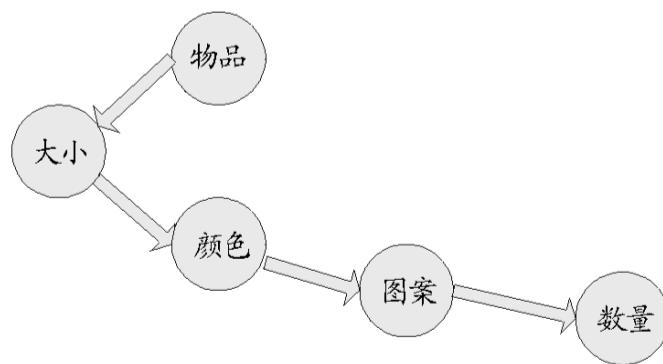


图2 仓库物品数状结构

\wedge stock(物品,大小,颜色,图案) = 数量

例如: \wedge stock(“大衣”,23,黑色,方格) = 12,这个等式的意思是库房有23号黑色四方格大衣12件。在这样的数据结构下,要确定是否有23号大小黑色方格的大衣,只需要访问一下这个数据节点就可以了。如果客户想要23号大小的大衣,但是不知道其颜色和图案,可以用多维数组列出这个数组: \wedge stock(“大衣”,4)。如果采用关系型数据库来完成上述行为,必然要建立多张表,然后通过外键进行连接,显然增加了复杂度和开销,也不能像多维数组那样更加形象和直观地对事物进行操作。Caché 数据库由于使用了 global 数据结构,从

而能够以一种简单的方式对丰富复杂的数据进行存储,也加快了软件开发的效率。

4.3 基于 Caché 服务器构建快速的 WEB 应用 为了适应越来越多 WEB 应用的需要,Caché 提供了自己的 Caché Server Page 技术,CSP 让用户能够建立和部署高性能的、高伸缩性的 WEB 应用。使用 CSP 开发 WEB 应用程序,CSP 在数据服务器上运行,并且和数据存放到一起,不但提高了效率而且降低了网络服务器的负荷,从而能处理更多的服务器请求。引人注目的是 Caché 最新版本已开始引进被称为 ZEN 的技术,这种技术将每个页面完全封装为类的形式,同时提供了许多丰富的控件,用户还可根据需要编写为自己软件服务的功能强大的控件。目前众多的 HIS 厂商开始建立基于 B/S 结构的 HIS 系统,使用 CSP 技术能高效地进行开发。

参考文献

- [1] InterSystems 公司.Caché 技术手册[EB/OL].[2007-01-12].<http://www.InterSystems.com>.
- [2] InterSystems 公司.Caché 初学者中文技术资料[EB/OL].[2007-01-14].<http://www.InterSystems.com>.
- [3] InterSystems 公司.对象与关系之间的矛盾[EB/OL].[2007-01-15].<http://www.InterSystems.com>.
- [4] O'NEILP, GRAEFEG. Multi-table joins through bit mapped join indices[J].SIGMOD Record,1995,2(3):8-11.
- [5] 萨师煊,王珊.数据库系统概论[M].3版.北京:高等教育出版社,2000:232-241.
- [6] 徐杰磐.面向对象数据库及其应用[M].北京:科学出版社,2003:1-4,35-39.
- [7] 雷光复.面向对象的新一代数据库系统[M].北京:国防工业出版社,2004:1-10,50-80.
- [8] 刘雄飞,陈金雄,王庆森.医院数字化建设的现状与发展思考[J].解放军医院管理,2003(4):310-360.
- [9] 安迎建.数据库原理与应用[M].北京:中国石化出版社,2004:150-170.
- [10] 张彤.平台化的 HIS- 医院信息系统的发展方向[J].医疗设备信息,2004(7):15-25.