

当前选择 DBMS 时应考虑的若干问题

王红 卓越 (清华大学自动化系 100084)

数据库管理系统(DBMS)是绝大多数客户/服务器应用的核心。但是为某个特定应用选择一个最合适的DBMS却没有一个十分简单明了的方法。你必须考虑支持DBMS软件的平台的能力、SQL标准、预存例程和触发机制、数据库管理和监视工具等等很多问题。本文简要介绍DBMS的一些特点和功能以及选择DBMS应考虑的问题。

1. 数据库模型

虽然目前存在很多数据库模型,但关系数据库管理系统(RDBMS)仍牢牢统治着客户/服务器世界。关系数据库的成功在于关系模型虽然非常简单,但能解决绝大多数领域的问题。但是在某些应用领域,新出现的面向对象的DBMS(OODBMS)和多维DBMS(MDDBMS)发展很快。

对于需要保存复杂的数据结构或进行联机分析处理的系统,这些新型数据库工作得很好。OODBMS(例如GemStone系统公司的GemStone)能够以完整对象的方式保存数据,因此非常适合于使用面向对象的模型的开发工具(例如,基于IBM的SmallTalk的开发工具)。多维DBMS则为使用数据挖掘工具(Data Mining)搜索数据的用户提供了一种更自然的工作方式。不过除非你有这些特殊需求,RDBMS通常是客户/服务器系统的最佳选择。使用RDBMS开发应用,技术上面临的风险最小。例如,绝大多数流行的客户/服务器开发工具是通过数据库中间件或ODBC与大多数RDBMS通信。可是这些流行的开发工具一般不能与非关系数据库通信。

目前7个主要的DBMS垄断了关系数据库管理系统市场,它们分别是:Informix软件公司的Informix - Online Dynamic Server、Sybase公司的Sybase SQL Server、计算机联盟公司的CA - OpenIngres、IBM的DB/2系列、Oracle公司的Oracle Server、微软的SQL Server和Borland公司的Interbase等。

2. 软、硬件平台

从很多方面看,DBMS运行平台决定了数据库服务器的总体性能。除了硬件外,DBMS还依赖操作系统的

功能高效地提供硬盘、缓存和网络服务。因此,为DBMS选择合适的平台和选择DBMS一样重要。当基于X86、RISC和Sparc的服务器占据了市场的主要份额,操作系统一般为UNIX和Windows NT。UNIX有很多版本适合于不同的处理器,包括在X86处理器上运行的SCO Unix、UnixWare和Solaris,在Sparc处理器上运行的Solaris以及在RISC处理器上运行的HP - UX、AIX、IRIS等等。UNIX能够提供真正的抢先的多任务和多线程服务、具有优异的I/O性能、内存管理和任务管理性能,但是管理UNIX操作系统比较复杂。高端的UNIX服务器(如IBM的RS/6000、HP的HP - 9000)价格昂贵,一般超过\$50,000。尽管目前UNIX操作系统仍然是主要的选择,但是Windows NT正在迅速的流行起来,它不仅能够提供UNIX的绝大多数功能(如果不是全部功能的话),而且人机界面友好、易于管理。

3. 绕过操作系统工作

为了提高性能,一些DBMS绕过操作系统,直接控制系统资源。例如Informix、Sybase和Oracle都能够绕过操作系统本身的文件系统,直接访问被称为原始硬盘分区的物理上的硬盘分区。这样就没有了通过操作系统调用访问硬盘的负担,更快的硬盘I/O操作意味着更快的DBMS。虽然使用原始硬盘分区通常能够提供最佳的性能,但备份数据库需要利用特定的DBMS开发商提供的特殊工具。不过,DBMS使用原始硬盘分区还是使用操作系统本身的文件系统一般由用户自己决定。

DBMS提高性能的另一个措施是将DBMS服务程序作为操作系统内核一级的进程运行。一般基于UNIX操作系统的数据库服务器都采用这种方式。不过最好的例子是在NetWare文件服务器上作为NetWare可装载模块(NLM)运行的DBMS。因为NetWare没有零级(Ring 0)保护,所以DBMS就可以在Ring 0上运行。大多数基于NLM的DBMS,包括Oracle公司和Sybase公司的产品,都在Netware上运行的很好,速度很快。这种做法虽然能提高DBMS性能,但也是非常危险的,因为DBMS数据库服务器软件的一个错误操作就可能导致

整个服务器崩溃。

4. 管理客户连接

DBMS 服务器建立一个客户连接需要很多内存和 CPU 资源,因此受有限资源的限制,它只能处理有限的客户连接。例如,如果 DBMS 服务器建立一个客户连接需要 5% 的资源,则它最多只能同时建立 20 个客户连接,当然每个客户连接需要多少资源取决于客户实际在干什么。DBMS 一般以下述 3 种方式处理客户连接:每个客户单独一个进程、多线程或综合以上两种方式。

第一种方法为每个客户连接单独产生一个进程,这样 10 个客户连接就需要执行 20 个单独的进程。这种方法的优点是在各自的进程地址空间内建立的连接不受其他进程的影响,尤其是出现错误的进程的影响。而且多进程操作系统能够比较容易地将这些客户进程分配给若干个处理器以平衡负载。缺点是耗费资源,因为每个客户都要求一个重负载的进程。此外,还要增加进程之间通信等额外负担。采用这种方法的 DBMS 有 DB/2、Oracle Version 6 和 Informix。

第二种方法则是将所有用户连接和数据库服务器程序都以线程的方式在同一地址空间内运行。线程是轻载的“进程”,它比真正的进程运行得快、要求的资源少。而且这种方法具有自己的内部调度,不依赖于操作系统的进程保护机制。因为它对操作系统自身服务的要求比较少,所以跨平台移植性较好。缺陷是如果其中某一个线程运行出现错误,都有可能使整个 DBMS 崩溃。而且这种方法不能象我们期待的那样均匀地分配操作系统资源,任何一个线程都有可能使处理器饱和,使得其他线程都不得不等待执行。采用这种方法的 DBMS 有 Sybase System 11 和 Microsoft SQL Server。

还有一些 DBMS 综合以上两种方法的优点,例如,Oracle 7,它使用多线程的网络监听程序来建立初始的客户连接,将客户请求交给称为调度员的进程处理。调度员依次将从客户传来的消息放进一个内部消息队列。由所有客户共享的服务进程从消息队列中提取消息,在数据库服务器引擎内部执行客户请求后,再次通过内部消息队列将响应结果返回给客户。这种方法的优点是不需要为每个客户连接单独分配一个进程就能够为每个客户连接保留一个处于保护状态的进程环境。但是采用队列有延时问题。

5. 并行查询处理

大多数硬件和服务器操作系统平台(包括 X86、

RISC 和 CISC)都支持多处理器并行处理。并行查询处理有两种不同的方法。一是由操作系统自动将数据库查询操作分到各个处理器(但共享内存);二是由 DBMS 分配数据库查询操作(什么都不共享)。如果你选择后者的话,就必须专门购买一种特殊版本的、专门设计来进行并行查询操作的 DBMS。在这种 DBMS 服务器中有一个客户查询管理器,它能够将一个数据库查询分解成若干个小的查询,然后将这些小的查询分配给在几个不同的处理器上的数据库引擎并行执行。通常,每个处理器上只有一个数据库引擎,其功能等价于一个独立的 DBMS 服务器。查询操作完成后,查询管理器将所有结果组装起来形成一个总的结果返回给客户。以这种方式调度查询操作代价昂贵,只是在几年前刚刚开始实验。目前有几个 DBMS 开发商提供这类产品,Oracle 的 Parallel Query Option(并行查询选项)、Sybase 的 Sybase MPP(以前称 Navigation Server)以及 Informix 的并行 DBMS 服务器。它们主要用于数据仓库或对数据库响应时间要求很高的应用。

6. 预定义的例程、触发器和规则

这些功能一般用来将部分客户/服务器应用放在 DBMS 上执行,以平衡负载,还能够减少网络流量。简单地说,就是保存在 DBMS 上的一组 SQL 语句和逻辑程序,开发人员将它编译后作为数据库对象保存在 DBMS 上。客户程序向 DBMS 发送一条消息就能执行那些预定义的例程。触发器则是由数据库服务器进程在正常的数据库业务处理(如插入、删除或更新记录等)过程中执行的操作自动激活并执行的预定义的例程。规则是一种特殊的触发器,其作用是在将数据插入数据库之前检查它是否合法,例如考试成绩不应该是负数。

预定义的例程、触发器和规则为客户/服务器应用开发人员提供了很多便利。但这些功能一般由某种特定的 DBMS 专有,不同版本的 DBMS 提供的功能有很大差别。例如,微软的 SQL 服务器对于每个插入操作只能激活一个触发器,而 Computer Associates 的 CA - Open-Ingres 能够激活多个触发器。此外,预定义的例程、触发器和规则一般不能在不同的 DBMS 之间移植。

7. 数据库管理和监视工具

DBMS 一般都捆绑有数据库管理和监视工具。利用数据库管理工具能够备份和恢复数据库、进行用户管理、安全管理和性能调整。利用数据库监视工具能够观察 DBMS 运行的各个方面,包括高速缓存器、磁盘、处理

器等的使用情况等。如果需要附加的功能,例如定制的报告机制,一般能够在第三方找到所需要的产品。

8. 安全性

我们一般要求 DBMS 提供数据库、表、列(即字段)级的安全措施,有时候甚至是行一级的安全措施,来确保只有授权的用户才能够查看或更新数据。虽然绝大多数 DBMS 都提供了最基本的安全措施——数据库或表一级的安全措施,但它们一般都不能彻底防止非法入侵,需要开发人员在数据库自身的安全措施之外采取更严密的防范措施。

目前市场上有一些安全版本的 DBMS 能够达到政府安全标准,如 C2 级标准或 B1 级标准。这些 DBMS 能够严密的监视用户活动、检测非法入侵者,扮演“数据库警察”的角色。例如,Informix - Online Secure 7.x 就能提供这些服务。Oracle 和 Sybase 也都有安全版本的 DBMS。采用这些安全版本的 DBMS 虽然能够大大增强安全性,但也必须付出一定的代价:这些 DBMS 一般价

格较贵,附加了许多安全处理,运行性能稍差一些。

9. SQL 标准

在选择一种 DBMS 之前有必要了解 ANSI/ISO SQL 标准。目前有 3 个 SQL 标准:SQL - 89、SQL92 和 SQL3。SQL - 89 是其中最老的一个标准,现在一般不再使用,而是参照 SQL92。同前者相比,SQL92 为 DBMS 增添了许多新功能,例如:

- * 支持 SQL 代理
- * 支持 SQL 客户/服务器连接
- * 为高级编程语言提供嵌入的 SQL 功能
- * 动态 SQL 查询
- * 支持高级的数据类型,如 BLOB

虽然,大多数 DBMS 还未能完全支持 SQL92,SQL3 已经问世了。SQL3 增加了更多的功能,例如面向对象的 SQL 功能等。

(来稿时间:1996 年 11 月)