

数据库

流动计算

系统程序

(18)

67-69

流动计算中的数据库问题

范明 编译

TP311.13

一、引言

现在,“可移动计算”是指用户使用携带的便携式计算机(典型地是 IBM-PC 兼容机)运行最初为非移动环境设计的系统软件。当用户旅行时,这些计算机通过调制解调器与其它计算机连接;当用户在办公室时,通过直接网络连接。典型地,办公室计算机中存放数据的主拷贝,旅行前所需的数据卸到便携式计算机上,回来后重新加载到办公室计算机上。用户完全对确保文件的正确装卸负责。

近年来的技术进步已使得便携式计算机的能力和速度大大提高,物理尺寸减小。这种进步从系统软件的观点看是一种进化,而不是革命。然而,现在我们看到使用便携式计算机的方式正在发生根本性改变。便携式计算机的容量已变得足够大,足以容纳不久前还被我们称之为“超大型”的数据库;而且其速度已足够快,能够支持复杂的数据库操作。此外,无线网络技术使得这些计算机可以作为分布式系统的整体部分进行交互作用。最后,由于基于光笔的界面和更小的物理尺寸的计算机(包括“掌上计算机”)的引入,这些计算机的潜在用户群已大大增大。

这些计算机的能力增强后,用户开始以使用他们的台式工作站相同的方式使用这些系统。但在目前,由于可移动计算机可存放的数据卷和与其它计算机潜在的连接性能,用户还不能期望用手工装卸文件来管理他们的数据。另外,可移动计算机必须作为分布式系统的一部分操作,但要受到可移动计算机硬件本身、无线网络技术以及计算机和用户在物理上的移动等特定限制。

本文从数据库角度来考察流动计算。我们认为流动计算根本上改变了数据系统现存协议、模型和算法的一些假定。在这种环境下,数据库研究领域必须有所重新考察。

二、移动和无线技术

当前市场有大量设备都被广泛看作可移动计算机。我们认为诸如 HP95(带有袖珍无线电连接件)或光笔数据输入设备(如联合邮包服务机构携带的那种)等因其功能有限,不能看作通用的可移动计算

机。我们指的可移动计算机是任何 PC 兼容机,基于 Intel 386SX/486SX,具有4M 字节主存,100M 字节以上磁盘存储器。如果当前的技术发展趋势继续下去,则这些计算机的 CPU 速度、主存容量和磁盘容量大约8个月翻一番。然而,在可预见的将来,将仍然存在一些其它限制。

当前,可移动系统具有的电池容量相对有限,通常只能用两、三个小时,不能指望在今后五年内再提高30%以上。在不久的将来,当前单位重量的电能限制将继续有少许改善,在低能耗芯片方面的工作(例如,AT&T 的 Hobbit 处理器),以及在低电压结构方面的研究正在进行,然而,我们预料,由于用户将要求在他们的可移动机器上配备更快的处理器(以便他们运行应用程序之性能比得上在工作站上的运行),将导致对电能需要的增加。

可移动系统的屏幕尺寸近期不大可能增大很多(甚至可能减小!),体积小是这些机器的主要优点之一。除非有重大的技术突破(例如,虚实眼镜或可折叠的屏幕技术),否则,我们不能指望减轻当前的限制。

粗略地说,对于可移动的计算机有两种类型的网络结构。一是由诸如 NCR WaveLAN(基本上是“无线的以太网”)的局域网构成,它以相对低的能耗和花费提供小地域内(例如,在一幢大楼内)的连接。二是使用蜂窝式或袖珍式无线电调制解调器(例如,Ericsson GE Mobidem),提供大得多的地理区域上的连接,但花费相应增加。对于局域网,典型带宽为2M 比特/秒(WaveLAN),而对于远程无线电典型带宽是8K 比特/秒(Mobidem 袖珍无线电)。这些数字在今后几年内将会提高,可以肯定网络带宽仍然是近期系统的主要性能瓶颈。

最后,整个分布式计算环境的可靠性仍然是一个问号。现在还不知道如何实现可靠的网络连接,特别是当用户将这些系统置于电子噪音很大的地方时更是如此。难以估计一台脱网的计算机将导致多少次通讯中断(更不要说灾难性的故障了),然而,这些系统的运行条件肯定比典型的车间恶劣得多。因此,我们认为,网络分割的频度和持续时间至少比当前非移动的分布式系统高一个数量级。尽管上述论点

表明可移动环境比传统的环境更加故障四伏,但是某些特定的“故障”是可以预料的,一个用户可以预先通知由网络断开或关闭计算机;改变无线网络中的信号强度可以使系统预计到一个紧急拆连,这样,可移动计算环境的关键故障是“频繁的、可预见的拆连”。

三、数据库系统内部

3.1 查询处理

计算环境的变化引起数据库系统查询处理部分的修改并非罕见。例如,考虑由严格的集中式数据库系统到早期的基于相对低带宽的广域网的分布式系统的进化。这一改变迫使查询优化程序在选择最优查询策略时大多集中在网络传输花费上,这导致了半连接算法的研究。随着局域网的流行,通讯费用变得不那么重要,优化程序又回过头来重新考虑各种资源的花费。流动计算的许多特点又一次迫使我们重新考虑以前的选择。本节我们集中考察以下两方面的影响:网络环境和电池供电式可移动系统的可靠性。

可移动数据库系统的可能网络环境将是高度异种的并具有广泛、混合的带宽和结构。可以预料,带宽规格将促使优化算法的改变,但仍然有一些新问题需要考虑。例如,可以预料在无线网络上发送信息的费用将很高(对于 RAM Mobile Data's Mobitex 月费用在50到80美元之间,并且每1000个字符信息大约需0.25美元。)这种费用可能导致查询优化程序的某些设计者只注重使事务处理的经费开支最少。作为第二个例子,考虑蜂窝式调制解调器用于通讯。相对于连续通话,通讯开始的附加费用将促使考虑长事务的查询处理策略,这些长事务不依赖于频繁的短通讯,而是需要较少(尽管较长)的对话。

正如我们在第二节指出的,看来当前的电池技术限制用户通话时间为两小时左右,一个必须考虑的问题是如何降低数据库的使用对这种资源的依赖。一种可能的办法是研制基于电能消耗选取策略的查询优化算法。这样,使每秒处理的事务数量最多的通常标准可能代之以使平均每个事务消耗的电能最少的标准。

流动计算机环境的限制将使得我们更可能接受查询的近似回答,因为除近似回答外可能根本没有回答。对于流动计算环境,可能接受的结果包括准拷贝(quasi-copy)。

流动计算查询处理的另一特点与查询本身有关。例如,查询的参数可能与提出查询的可移动计算机的位置有关,也可能依赖于计算机的移动方向和

速度,或者在查询处理期间,依赖于位置的数据可能发生变化。

流动查询的这些特点不仅影响查询优化,而且影响查询语言 and 用户界面。

3.2 事务处理

流动计算的应用将需要各种事务和类事务的服务,其范围包括如远程订单输入这样的数据处理任务到更复杂的交互的任务,如日程管理、类 CAD 活动、和类 saga。由于可移动计算机具有大量的局部存储,事务处理至少要像传统的分布式数据库系统一样复杂。由于用户需要在拆连期间能够有效地工作,因此可移动计算机将需要在事务管理方面在很大程度上是自治的,这就导致了多数据库的问题。

这样,流动计算包含了事务处理研究界当前感兴趣的许多研究领域的一些特点。然而,正如我们下面将看到的,流动计算使这些问题更富变化性。

3.2.1 故障 频繁的、可预测的可移动计算机的拆连表明,在预测到拆连时系统一定能够对活动的事务采取特定的行动。

(1)如果没有进一步的交互行动,事务处理可以迁移到一台非移动的计算机上。

(2)远程数据可以在预期的拆连之前卸下,以支持拆连之后需要继续在可移动的计算机上执行的交互事务。

(3)日志记录可以由可移动计算机传送到非移动的计算机上。

(4)可移动计算机可以采取行动,“申明脱网”。分布式协议可以处理拆连,其开销低于当前拆连发生后才发现拆连的模式。

对于可移动环境,另一种独特的容错是稳定的存储器的不稳定性,在实践中,磁盘存储器被看成一种非移动系统存放日志的稳定存储器。高可靠的系统使用重复日志,然而,唯有可移动计算机更易于受到由于用户将计算机掉到地下、机场安全系统导致数据破坏、或者甚至整机丢失或被盗而引起的灾难性故障的伤害。因此,上面提及的在预期的拆连之前将日志记录由可移动计算机传送到非移动计算机上可能确实是一项至关重要的活动。

当一台可移动计算机重新连接时,多半有紧迫的通讯要求。可移动计算机上的局部事务可能需要远程数据,远程事务可能需要局部数据,而系统可能有重要的内务操作(例如,由可移动计算机传输日志记录)。由于重新连接只有有限的时间,因此这些通讯任务需要按照系统整体性问题、对用户请求的响应时间和资源的有效利用等优先权来进行排序。

3.2.2 分布式计算问题 尽管流动计算环境只不过是一种分布式系统,但这种系统的费用和性能不同于传统的分布式系统。

可移动的计算机通讯的费用将比非移动计算机之间的通讯费用更高。这种费用不仅是使用网络的费用,而且包括可移动计算机的电池费用。由于接收消耗的能源比发送少,通讯费用函数是非对称的。数据成批由/向可移动计算机发送将降低整体花费。

通讯费用结构的这些改变以及上面讨论的故障模型的考虑将改变已有的分布式协议,也可能导致新的协议。我们猜想,对于主要由一台特定的可移动计算机使用的数据库,主拷贝算法将能很好地工作。

3.3 安全性

流动计算中的许多安全性问题不是数据库系统特有的,而另一些是数据库特有的。

由于可移动计算机在各种网络中的出现和消失,防止一台机器伪装另一台成了大问题,当一台可移动计算机脱离局部环境时,它发送和接收的数据容易被窃取和未经授权被拷贝。由于有些外部包将是发往访问计算机的合法包,因此允许可移动计算机连接的网络不能进行现在作为安全机制的包过滤。

可移动的计算机访问一个“外部”环境将消耗该环境的某些资源,这包括网络带宽。对于数据库应用,前面讨论的电源考虑表明还包括磁盘存储器的使用,或许还有 CPU 周期。由于数据库应用对环境的需求不容忽视,记帐(和传统的分时系统非常类似)对于收费和限制访问计算机对外部环境的影响都是需要的。

也许,在流动计算环境下的数据库安全性的最有趣的方面是系统成分之间的正确性。传统地,数据库系统将它们的操作和安全性建立在所使用操作系统之上。显然,对于可移动计算机和外部环境为成功地工作而共享资源的情况,这种正确性的程度的某些是必需的。然而,现在还不清楚这种正确性程度的准确特性。理想地,正确性应当限于对可移动计算机可以认证自己的标准系统软件成分。

上面讨论的是可移动计算机用户关心的安全性问题。然而,外部环境的管理员同样也关心安全性问题,这种关注比当前的“可移动计算”模式更大。在当前的模式下,外部环境的用户登录在一个局部的宾客帐号上,由此用户通过无线网连接到他自己的内部环境。在流动计算环境下,一台主机可能伤害宿主机——或者无意地,或者有意地。这种伤害的可能性比通过宾客帐号大得多。允许流动计算机访问它所需的局部资源,能够对伤害加以限制吗?

四、人机界面

传统的数据库查询语言,特别是 SQL,不易被基于光笔的计算机使用。当前的手写体识别不适合于像典型的 SQL 查询那样长的输入,这样图形用户

界面是需要。在图形用户界面下,通过指向模式的元素或数据项本身来表示查询。图形用户界面方面的早期工作将键盘和鼠标的混用,并且还开发了典型工作站的大屏幕的用户界面。对于流动计算环境下的数据库访问,必须绝对尽可能少地输入文本,以支持无键盘的、基于光笔的计算机。此外,由于可移动计算机的尺寸小,界面也不能依赖大屏幕的使用。

这两种限制正在被 MITL 的 Neon 项目考察。为了减少表示数据库模式所需的屏幕空间,他们使用了嵌套关系模型的概念,用单个属性表示模式的部分。一种基于泛关系的方法与嵌套关系结合使用,使得查询的指定更加简便。由于泛关系的使用而导致的歧义性通过基于光笔的选择解决。

尽管当前原型仍需要键盘,但是他们正在减少对键盘的依赖,最终目标是完全依赖光笔。帮助减少对键盘的依赖的特色是使用光笔从模式中选择属性,使用光笔从数据库选择数据值插入查询中,使用光笔从属性域中选择数据插入查询中。特别是后两项对界面系统和人两方面都提出了严格的性能要求。他们正继续考察这些问题。

他们正在设计系统,设计时考虑到异种数据库环境。这不仅包括访问多关系数据库的能力,而且包括访问其它类型数据库,甚至诸如文件系统的非数据库信息源的能力(这需要考察诸如将一个关系与文件连接这种看来奇怪的想法)。

五、结束语

在过去的几年里,数据库研究界遇到的大部分挑战涉及数据库的形式概念如何适应存储设备、网络、处理速度、内存容量等方面的实际限制。尽管这些限制不断改变,但这些变化通常只是循序渐进的。只有一些不常有的革命性变革要求数据库的研究者重新考虑旧方法以产生新技术。按照我们的观点,流动计算就是这些革命性变革之一。

当前,对流动计算的系统研究主要在网络和操作领域,然而,远超出这些领域的问题直接影响数据库系统。流动计算引起了对电能管理、稳定的数据存储、无线通讯和用户界面的关注。我们已经注意到,研究界才开始重视这些问题对查询处理、事务处理、分布式计算、安全性和数据库查询语言的影响。新的硬件引入可移动计算机市场具有重要的实际影响,为数据库研究界打开了一个机遇之窗。

主要参考文献: Alonso R. and Korth H. F., Database System Issues in Nomadic Computing, In Proc. of the 1993 ACM SIGMOD Intl. Conf. on Management of Data, Washington, DC, 1993, pp388-392.