

分布式数据库中维持数据副本一致性的方法

— SQL Server 7.0 数据复制的原理及应用

王靖 熊焰 (中国科学技术大学计算机系 合肥 2300027)

摘要 本文探讨在松散耦合的分布式数据库系统中,保持数据副本之间数据一致性的问题。介绍Microsoft SQL-Server 7.0中引入的基于事务日志、带有冲突解决的数据复制方法,并举例说明这种方法的应用。

Abstraction: This article discusses how to keep data consistency within multiple copies in a loosely coupled distributed database system, introduces the data replication method based on transaction logs in Microsoft SQL-Server 7.0 as well.

关键词 分布式数据库 数据一致性 2PC 数据复制

1 引言

数据库系统是计算机软件的一个重要分支。回顾数据库系统发展的历史,不难看出其发展的两个趋势:一个是由小型化向大型化发展;另一个是由集中式向分布式发展。地域的分布以及计算机系统体系结构的分布化必然导致数据库系统的分布化。

在一个分布式的数据库系统往往需要维护同一个数据库的多个副本,那么如何有效维护多副本之间数据一致性的问题就摆到了我们的面前。

也许我们首先会想到经典的两阶段提交算法(2-Phase Commit,简称2PC),2PC能够保持多副本间数据严格一致,但其开发期间工作量较大,运行期间要求副本间保持可靠的通信链路,运行中的通信成本不容忽视。另外2PC需要同步修改数据库的全部副本,当网络故障或某个副本机器故障发生时2PC即不能顺利进行,势必形成阻塞,降低系统的可用性。

其实在实际应用中,除了少量关键性的数据(我们称之为临界资源)之外,其余大量的非关键性数据,其副本只要维持近似的、基本地一致,系统即可正常运行,而数据库副本间的数据更新则等待稍异步地进行,必然能减少拥塞,提高系统可用性,降低运行成本。

维持多副本间数据近似一致的方法有很多,一种直观的方法是利用邮件系统将最新副本拷贝给其余副本,这种方法需要额外的软件支持或人工干预。本文旨在介绍MS SQL Server 7.0中内置的“数据复制”构件,利用此构件,用户可将精力集中于更新本地副本,而更新其余副本的工作则交给数据库系统去完成。这种方法既可靠又简便,不需要额外开销。

2 基本术语

SQL Server 7.0仍然同以前版本一样,借用了出版界的术语:

复制(replication) 将数据表视图及数据由源数据库复制到目标服务器的过程。源、目标服务器可以是同一个服务器,也可以不是。

文章(article) 被复制的基本单位。一篇文章中包含被复制的单个数据表的信息,可以是整个数据表,可以是数据表的部分行(称为表的水平划分),也可以是数据表的部分列(垂直划分),甚至是表的部分行中的部分列(水平划分+垂直划分)。

出版(publish) 使数据库能够被复制的过程。

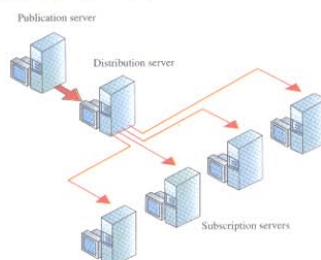
出版者(publisher) 被复制的服务器。又称出版服务器。

出版物(publication) 在一次复制中被复制的一组数据表。也就是文章的集合。**订阅(subscribe)** 请求或同意接受出版物的过程。

订阅者(subscriber) 接受出版物并更新本地副本的服务器。又称订阅服务器。

3 基本原理

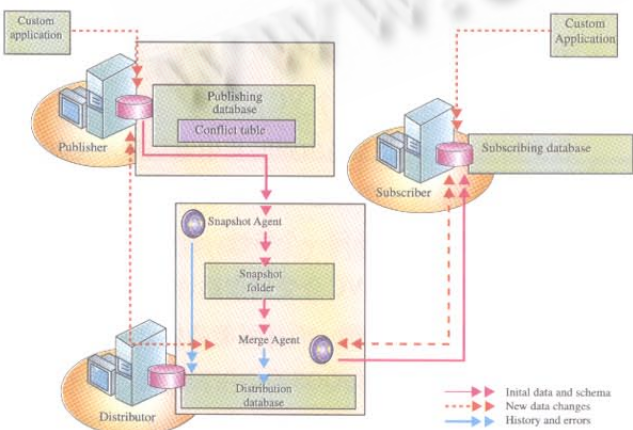
SQL Server的数据复制采用基于事务日志的方式(新增的快照复制除外)。其拓扑如下图所示:



日志阅读进程、分发进程是数据复制的主要组成部分。日志阅读进程定期阅读出版服务器中的系统事务日志库,将与数据复制相关的那部分事务日志拷贝到分发数据库中。分发数据库是一个存储转发数据库,保存所有等待发送到各订阅服务器中的事务日志。分发数据库驻留在分发服务器中,分发服务器既可是独立的服务器,也可有出版服务器兼任。分发进程则负责将分发数据库中的事务日志连续或定期地发往各个订阅数据库。订阅数据库根据接收到的事务日志更新本地副本,使之与出版服务器中的副本保持一致。

在 SQL Server 7.0 以前的版本中,对于出版物作了比较严格的限制:出版物只能在出版一方被修改,订阅一方不能主动修改。否则无法保证数据的一致性。这种限制必然降低了数据复制的可用性。在 SQL Server 7.0 中数据复制构件得到了较大的改进,尤其是新增的归并复制(Merge Replication)更是打破了这种限制。归并复制虽然仍然沿用出版者/订阅者术语,实际上不管是出版者还是订阅者均可修改出版物,任何一方的修改均被繁殖到其余副本。当出版者、订阅者同时修改出版物而发生冲突时,归并复制可以按照预先设定的优先权规则解决冲突,结果“胜出”一方的副本被复制给其余服务器。下面的插图描述了归并复制的过程。由插图可看出,归并复制第一次引入了一个归并代理(Merge Agent),其作用是当出版者、订阅者同时修改出版物而发生冲突时,根据预定的规则作出仲裁。

利用 SQL Server 7.0 的可视化管理工具 Enterprise Manager 可以完成数据复制的设置。首先选中相应的数据库,激活菜单 Tools --> Replication --> Create and Manger Publications...,在向导的提示下输入必要的信息即可完成设置。本文限于篇幅无法详细叙述。



注:在上图中,还有一个快照代理?Snapshot Agent?,其作用是定期产生出版物的快照。订阅方在初始化阶段,可以利用该快照生成本地的数据副本。

应用举例如下:

这是一个商品销售组织的商品销售 MIS 系统。该销售组织在城市范围内有若干个商店以及一个中心仓库,顾客购买大宗商品时,在商店付款后由商店出具提货单,然后顾客凭提货单到仓库自行提取货物或者约定时间送货。商店与仓库之间通过局域网或公用电话网相连。为了保证在通信链路中断的情况下,仍能正常工作,商店与仓库各自保存一个提货单的副本。提货单主要内容如下:

站点编号	货物种类	货物数量	已提货标志	...
------	------	------	-------	-----

其中,“站点编号”是商店、仓库的唯一标识;“已提货标志”为 true 表示顾客已提货,该标志初试值 false。

各商店保存自行销售的商品的提货单,仓库则保存各商店提货单的并集。仓库提取相应提货单子集与对应商店形成一对出版/订阅关系。由于 Merge Replication 中,出版者/订阅者可以彼此更新,是实际上的对等关系,所以原则上可任选一方作为出版者。由于出版者可以指定部分行作为一个出版物,所以以仓库为出版者更方便。在本例中,仓库以所有站点编号为 1 的提货单作为一个出版物,与商店(1)形成一对出版者/订阅者关系;而以所有站点编号为 2 的提货单作为一个出版物,与商店(2)形成一对出版者/订阅者关系。其拓扑如下:



当顾客在商店付款购买大宗商品时,商店 MIS 系统在本店提货单表中生成一条记录(已提货标志为 false),这条记录随后由数据库系统透明地复制到仓库提货单表中。顾客前往仓库提货后,仓库 MIS 系统将本地提货单表相应记录的已提货标志置为 true,售货商店提货单表中相应记录自动被更新,以便售货商店查询。

本系统投入运行数月来,系统运行稳定、正常。

4 结 论

Microsoft SQL Server 7.0 的数据复制,是一种基于事务的异步复制机制,运行高效、可靠,人工干预少,易于(下转第 23 页)

(上接第 21 页)

设置和维护。在松散耦合的分布式数据库中有较好的应用前景。

其不足之处在于:当多方更新同一数据发生冲突时,仅利用预先设定的优先权解决冲突,没有考虑其时间上的先后次序。笔者认为如能考虑时间的因素则冲突解决方案会更为完善。考虑到分布式系统中没有统一的绝对时钟,操作时间次序的确定可以借助于类似Lamport的逻辑时钟定序算法。■

参考文献??

1. *Report on the Workshop on Fundamental Issues in distributed Computing*, Fallbrook, CA, 1980
2. *Integrity in Distributed Databases*, C.Parent, Proc AICA 77, Pisa 1977
3. *Microsoft SQL Server 7.0 Books Online*, Microsoft Corporation
4. *Microsoft SQL Server 6.5 Books Online*, Microsoft Corporation
5. *MSDN*, Microsoft Corporation