

PCs 数据库机群系统高可用性的研究与设计^{*}

祝永志 孔令鑫 侯秀杰

(曲阜师范大学计算机科学学院 日照 276826)

摘要 随着用户对数据库系统的可用性、可伸缩性以及性能指标的要求越来越高,使用机群技术进行容错和并行处理已成为一种有效可行的解决方案。本文就数据库机群及其可用性理论作了探讨与研究,并给出了一个由廉价 PCs 微机组成的具有较高可用性的数据库机群实施方案。

关键词 数据库系统, PCs 机群, 高可用性, 容错处理

Design and Realization of High Available PC Clusters

ZHU Yong-zhi KONG Ling-xin HOU Xiu-jie

(College of Computer Science, Qufu Normal University, Rizhao 276826, China)

Abstract As users have a more and more high demand for the availability, scalability and performance of the database system, using the cluster technology to carry on fault-tolerant and parallel processing has become one kind of effective feasible solution. This paper has made the discussion and the research on the database cluster and its high availability, and presented a scheme composed by the inexpensive PC computers that has the high availability.

Keywords Database system, PC clusters, High availability, Fault-tolerant

1 引言

随着网络信息传输量的急剧扩大,数据库已经成为计算机科学的重要研究与应用领域。对数据库而言,除了记录正确的处理结果之外,还面临着提高处理速度、增加数据安全性、保证数据的可用性和数据集的可扩展性的挑战。随着 IT 应用的深入和有线、无线网络的快速增长,电子事务处理业务对以上 4 个方面提出了更高的要求。数据库机群是实现上述需求的很好途径。

本文通过对数据库机群体系结构的研究与分析,阐述了由 PC 机组成的数据库服务器机群获得高可用性的技术方法,提出了构建具有高可用性的 PCs 数据库机群系统的设计模型并实践性地加以论证。

2 数据库机群系统

机群是一组相互独立的、通过高速网络互联的计算机,它们构成了一个组,并以单一系统的模式加以管理。数据库机群系统是由多个有独立操作系统的服务器共享一个数据库而成。当机群系统中个别节点发生故障时,可将其任务分散到其它正常节点上,继续完成其工作。机群系统通过虚拟服务对外提供一个单一的系统映射,屏蔽了内部的节点信息,用户通过机群的虚拟 IP 和机器名访问机群系统。

3 数据库机群服务器的体系结构

当前数据库机群有两类体系结构:一种是基于数据库引擎的机群技术(共享磁盘 Shared disk 或非共享磁盘 Non-Shared di,如图 1;另一种是机于数据库网关(中间件 Middleware)的机群技术(非共享 Shared nothing),如图 2。

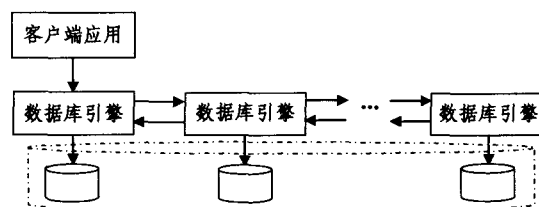


图 1 基于数据库引擎的机群技术(共享磁盘或非共享磁盘)

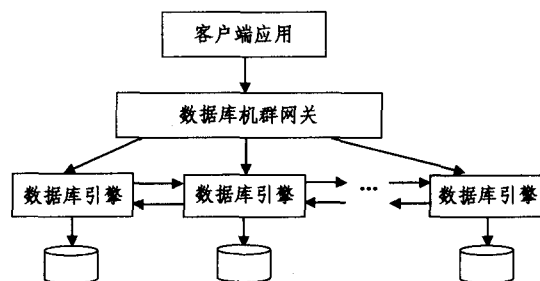


图 2 基于数据库网关(中间件)的机群技术(非共享)

基于数据库引擎的数据库机群技术有 Oracle's Real Application Cluster; Microsoft SQL Cluster Server; IBM's DB2 UDB High Availability Cluster; Sybase ASE High Availability Cluster; MYSQL High Availability Cluster。

基于数据库网关的机群技术有 Parallel Computers Technology Inc. 's ICX-UDS middleware^[1]。

4 数据库机群服务器的高可用性

4.1 高可用性概念

^{*} 山东省教育厅高等学校实验技术改革项目(编号:2005-400)。祝永志 教授,硕士生导师,主要研究方向为网络与分布式系统;孔令鑫 硕士研究生,主要研究方向为分布式计算;侯秀杰 硕士研究生,主要研究方向为分布式计算。

高可用性机群 (High Availability Cluster), 简称 HA Cluster 或者 HA 高可用性机群, 是指以减少服务中断(宕机)时间为目标的服务器机群技术。高可靠机群主要使用在数据库系统 (Oracle, Sybase, SQL Server) 以及关键应用系统 (Notes, Exchange) 中。

数据库机群可用性是指系统无故障使用的时间百分比 (MTBF 平均无故障时间)。一个数据库机群系统, 除了要处理单节点故障情况, 还要处理与节点或网络相关的故障情况, 以保证在故障处理时的数据库可用性。

数据库机群高可用性的宗旨是最大化地提高系统 MTBF, 减少服务中断的时间, 并能够自动检测数据库服务器的故障^[2]。

4.2 PC 机数据库机群服务器的可用性分析

由于 PCs 数据库机群是由廉价的不太可靠的 PC 机构建的, 因此可以从机群硬件、软件及通讯链路的冗余等方面来保证其可用性。

理想的数据库机群系统应该具有提高处理速度、增强安全性、提升数据高可用性和数据集可扩展性能力。

基于数据库引擎的数据库机群系统, 无论是共享或是非共享磁盘, 要么因只有一个物理数据映像, 在提高了系统性能的同时因为缺少数据冗余而降低了可用性, 要么有了数据冗余但因使用了 IPC 协议而使两者都有所降低。

Parallel Computers Technology Inc. 's ICX-UDS middle-ware 是一种基于复制中间件的数据库机群系统。这种中间件位于网络关键路径上, 监听所有进出数据库系统的流量, 能方便地提供防火墙和其它安全服务, 保护物理的数据库服务器。中间件复制技术通过多服务器的并发处理屏蔽了处理延迟, 通过对硬件、软件和通讯链路的高度冗余, 提供了系统高可用性^[1]。

5 基于非共享磁盘高可用性的 PCs 数据库机群系统的构建

5.1 系统硬件环境、软件环境以及网络环境的构建

图 3 中, 1 台客户机及数据库机群的 4 个节点计算机均为方正电脑: Pentium (R) 4 CPU 2.66GHz; 内存: DDR 256M; 硬盘 7200R 40G。客户机操作系统是 Windows X PService Pack2 简体中文版, IE7.0 浏览器; 数据库机群节点操作系统是 Windows Server2003, 数据库机群节点的 DBMS 是 MS SQL Server2005。

5.2 非共享磁盘高可用性的 PCs 数据库机群系统工作原理

机群由一台主服务器和 3 台备份服务器组成。

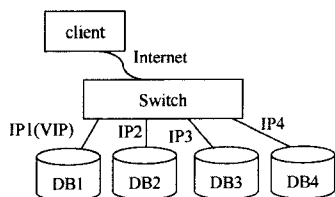


图 3 非共享机群 (Share-nothing Cluster) 结构图

主服务器 IP1 是数据库机群入口, 机群虚拟 VIP: 219.218.26.215, 其余 3 个备份服务器的 IP 地址分别是 219.218.

26.217, 219.218.26.221, 219.218.26.220, 和 VIP 处于同一子网。

早期共享磁盘服务器允许所有服务器访问所有磁盘, 该方式需要昂贵的网络设备并要求使用专用软件和应用 (协调共享磁盘访问的专用软件称为分布式锁定管理器-DLM)。虽然 SCSI 标准已经消除了对昂贵的网络设备 (如线缆、交换机) 的需求, 但鉴于其固有的伸缩性局限, 仍不能广泛使用。共享磁盘解决方案有 Digital VAX 机群和 Oracle 并行服务器。

为解决共享磁盘机群的局限性, 除了采用磁盘镜像技术外, 很多机群采用非共享体系结构, 机群中每一台服务器都拥有自己的磁盘资源。当一台服务器发生故障时, 非共享机群软件能将控制所有权从该服务器转换到另一台服务器上。非共享方式因没有特殊的磁盘访问需求, 可完全支持标准化应用。非共享机群解决方案范例有 Tandem Nonstop, Informix Onlin/XPS 和 Microsoft Cluster Server^[3]。

在图 3 中, 当主服务器发生故障, 在本文开发的全局事务管理器的作用下, 机群系统将从剩余的正常节点中选一个具有最大消息写操作排序号的节点作为新的主服务器。新的主服务器通过 IP 强占方法将机群的入口 IP (219.218.26.215) 绑定在自己的网络接口上。

5.3 系统采用的冗余技术

提高数据库机群可用性通常有 4 种方法: 硬件级冗余、通信链路级冗余、软件级冗余和数据冗余^[4]。

被动更新数据集的数据冗余方式由于容灾能力差、资源占用多, 已渐成过去时。以一台或多台备份数据库服务器管理数据集的主动更新数据集方式可用于故障恢复。

图 3 采用了较先进的同步主动复制数据集技术, 以保证机群高可用性。即所有并发事务处理在所有数据库服务器上同时完成, 较好地解决了队列管理问题, 同时通过负载均衡实现更高可伸缩性。

结束语 本文阐述了构建数据库机群的意义, 论述了数据库机群可用性概念和理论, 分析了共享数据库机群和非共享数据库机群的结构特征和性能比较。

本文实践性地构造了基于非共享的 MSCS 数据库机群系统, 该系统结构简单、性价比高、支持底层数据库节点的异构性^[5], 并且具有一定的可扩展性和高可用性, 对力图使用廉价的普通微机构建具有高可用性数据库机群提供了很好的范例。

参考文献

- [1] 数据库机群技术分析和比较. 并行计算技术公司 900 Eighth Avenue, Suite 300 King of Prussia, PA 19406, USA. August 2005
- [2] 王婉菲, 王欣, 张志浩. 数据库机群系统的研究与实施[J]. 微型电脑应用, 2003, 19(10): 31-33
- [3] 邵志远, 金海, 唐晓辉. 基于主动 TCP 连接复制的高性能高可用 MySQL 数据库机群机群 [J]. 计算机研究与发展, 2005, 42(6): 1006-1012
- [4] http://www.hlib.cn/article/014/14720_2.asp
- [5] 金蓉, 龚卫华. 基于无共享数据库集群系统结构的设计[J]. 浙江理工大学学报, 2007, 24(4): 448-450