

厂站监控系统中高可用数据库中间件的设计与实现

李纪云, 蒋衍君, 金文凯

(国电南京自动化股份有限公司 研发中心, 江苏 南京 210003)

摘要: 电力系统对数据存储可靠性有极高的要求, 设计并实现了一个高可用数据库中间件 MultiDbGate。它通过对连接和访问的动态维护和调度, 屏蔽了底层连接、多数据库、双网等问题的复杂性, 可以提供廉价而又简单的数据库服务, 具备高可用性和可扩展性。介绍了中间件 MultiDbGate 的结构, 分为单数据库接口、连接池、作业调度和日志服务 4 个模块。单数据库接口模块是对活动数据对象(ADO)的封装; 连接池模块是中间件的核心, 可以让不同的应用共享连接, 并屏蔽了底层的分布式架构; 作业调度模块是管理模块, 响应应用的请求; 日志服务模块通过日志重写保证数据一致性。中间件 MultiDbGate 已经在某电网的多个 220 kV 变电站监控系统工程应用, 效果良好。

关键词: 厂站监控系统; 中间件; 数据库; 高可用

中图分类号: TM 76; TP 31

文献标识码: A

文章编号: 1006-6047(2007)06-0118-04

0 引言

在电力自动化领域, 监控系统时刻都会生成大量历史数据, 历史数据的存储和查询是所有数据采集和监控(SCADA)^[1]系统都面临的问题。商用数据库具有开放性和可靠性的优点, 所以绝大多数 SCADA 系统都把历史数据存入商用库, 以方便历史数据的插入和检索。

现有的数据库客户端数据库编程技术, 如开放数据库互连(ODBC)^[2]、数据库访问对象(DAO)等, 在一定程度上可以简化数据操作, 可是它们没有对高可用性提供支持, 无法满足可靠性要求极高的自动化领域的需求。而且, 对于分布式的系统开发而言, 这些技术的使用也过于复杂。因而, 需要开发一种高可用的数据库访问中间件^[3]解决这些问题。

1 高可用数据库中间件的提出

现有的数据库高可用技术可以分为 2 类, 一类通过硬件实现系统级的高可用, 如独立磁盘冗余阵列(RAID)^[4]技术、集群技术^[5]等。该类技术的特点是通过共享存储保证数据的一致性, 因此不会出现数据不一致的现象, 效率高。但是, 这需要的硬件投入大, 网络附加存储(NAS)、存储域网络(SAN)^[6]等存储设备价格高。另外一类是通过软件方式实现, 一些数据库提供数据的镜像备份功能, 如 SQL Server 就提供快照复制、事务复制和合并复制功能^[7]。该方式存在的问题很多, 其一是配置复杂, 不同数据库提供的功能和配置方式都不同, 而且并不是所有

数据库都支持; 其二是该方式的数据的一致性差, 出现问题后不容易处理, 数据库维护困难, 提供服务的可靠性相对较差。此外, 这 2 种方式都没有很好地解决访问透明的问题, 用户在编程的时候仍然要面对故障处理等一系列复杂问题。

为解决上述问题, 可以在数据库和应用之间增加一个中间层, 来屏蔽高可用带来的复杂性, 即提出高可用数据库中间件多数据库访问网关(MultiDbGate)。它把底层提供 $1+n$ 冗余的多个数据库封装起来, 向上提供统一的接口。MultiDbGate 改变了传统的使用数据库的方式, 屏蔽了连接、多数据库、双网等问题带来的复杂性。用户的所有数据操作都可以通过这个中间件完成, 简单可靠。

2 全局框架模型

作为高可用的数据库中间件, 应该具有 3 个特点。

a. 高可用性^[8]: 应该在系统的每个层次都能够提供 $1+n$ 容错功能, 出现故障时, 系统会自动适应新的环境。故障排除后只要开机就会自动进行数据库的同步, 然后加入服务功能池。

b. 可扩展性^[9]: 包括数据库的可扩展性和规模的可扩展性。能够支持不同的数据库, 甚至是异构的数据库系统; 在系统规模扩展时, 系统性能不能显著的下降。

c. 使用的简单性: 中间件的目的是为了使用的简单性, 所以配置和使用都应该很简单。

针对上述目标, 高可用数据库中间件的全局框架模型如图 1 所示。

MultiDbGate 实现的是数据服务的高可用, 对于应用程序的高可用可通过它们之间的心跳监护实现,

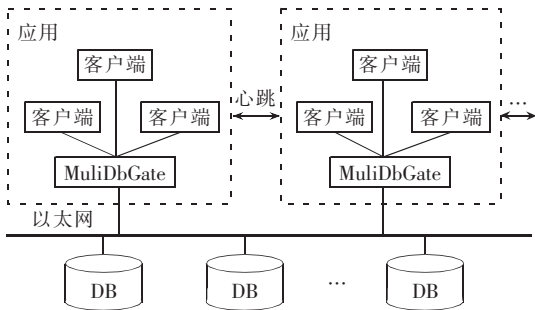


图 1 高可用数据库中间件的全局框架模型
Fig.1 Framework model of MultiDbGate

系统从下到上的每个层次都可以是 $1+n$ 容错的。可以根据工程需要,选配不同类型不同数量的商用数据库。当然,数据库也可以和应用程序运行在同一台机器上,该框架完全可满足大多数情况下对高可用数据库服务的需求,特别是电力自动化领域的需求。

3 设计与实现

3.1 总体结构设计

MultiDbGate 的结构如图 2 所示,一共分为单数据库接口、连接池^[10]、作业调度^[11]和日志服务 4 个模块^[12]。

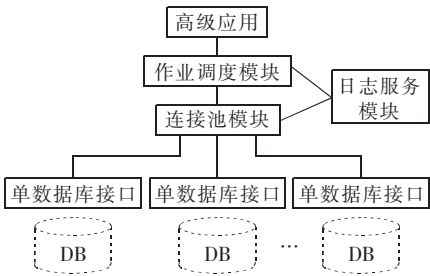


图 2 高可用数据库中间件模块结构图
Fig.2 Structure of MultiDbGate

其中,单数据库访问接口模块是对活动数据对象(ADO)^[13]的封装,它是所有功能实现的基础,对数据库的具体操作都最终落在它的身上。通过 3 个组件对象模型(COM)对象实现,即 Connection、Recordset 和 Command。其中,Connection 是最重要的,它负责和数据库建立连接,设置连接模式、等待时间、游标类型等。当然也可以通过它执行一些不需要返回记录的结构化查询语言(SQL)操作^[14]。只要选用不同的 ADO 数据库驱动,就可以适用于不同的数据库甚至是文件。

连接池模块是中间件的核心,它负责管理和维护数据库连接,屏蔽底层的分布式架构,向上提供一个可靠的统一的数据平台。由于建立连接的过程是一个耗费时间和资源的过程,不可能当用户有操作请求时才建立连接,所以需要该模块实时维护所有的连接,处理新连接的加入和无效连接的剔除。

日志服务模块是为了保证多数据库的数据一致性,它是实现高可用的前提。如果由于故障有些数据库退出服务,那么它就会把进行的操作临时写入

日志,等到故障排除重新提供服务时,就会根据日志记录完成和其他数据库的同步。

作业调度模块是协调日志服务线程和连接池维护线程,同时完成用户接口的封装。通过它的协调调度,不同客户端的操作可以共享相同的连接,提高效率。新数据库加入后,它首先调用日志模块进行日志处理,然后才允许加入连接池。

3.2 单数据库接口的设计与实现

单数据库接口模块是对 ADO 的封装。目前的数据库客户端技术很多,如 ODBC、DAO、OLEDB^[15]等,从综合效率和可扩展性考虑,选择了 ADO。功能上是根据 ADO 所提供的接口,把 Connection、Recordset 和 Command 3 个对象进行封装(见图 3),屏蔽 COM 的复杂性及与 ADO 紧密相关的数据类型。

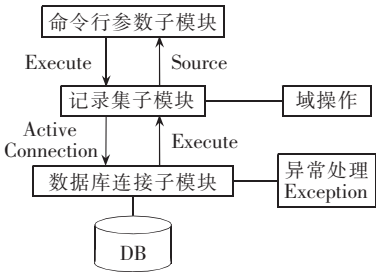


图 3 单数据库访问接口模块软件结构图
Fig.3 Structure of single database access module

a. 数据库连接子模块:主要是建立和数据库的连接。功能包括连接字符串设置、连接的建立和销毁、连接模式的设置、事务的启动和回撤以及不需要返回记录的 SQL 操作等。

b. 记录集子模块:它是数据库操作的主体,所有的操作都可以通过它来完成,使用该模块可以把结果集返回到本地内存中进行处理。例如,记录的插入、删除、查询等。

c. 命令行参数子模块:它是操作数据的 SQL 语句,为了便于各部分的修改,一个语句由多个参数组成。用于在执行时根据环境来动态调整 SQL 语句,例如调整查询条件等。

d. 异常处理子模块:因为数据操作出错的原因很多,使用该模块专门进行异常处理,方便获取出错原因的描述。这样,就可以在访问异常时判断出错原因,然后自动调整把问题都在底层给过滤掉。

3.3 连接池模块的设计与实现

连接池模块是中间件的核心,它负责管理和维护所有的连接,屏蔽底层的分布式架构,向上提供一个可靠的统一的数据平台。它提供的主要功能包括:动态维护所有的连接,包括加入和删除;根据请求,向上返回需要的连接供上层模块使用;让指定的连接执行对应的数据库操作,如存储操作。

连接具有数据库连接字符串、连接状态等属性。连接字符串是 ADO 连接数据库使用的参数集合,包括数据库所在主机的 IP 地址、ADO 驱动、用户名和

密码等信息。连接状态是当前连接所处的状态,根据数据库提供服务的情况,把连接分为 4 个状态:活动态、恢复态、失败态和关机(断网)态。把所有的数据库连接通过一个链表组织起来,然后通过启动独立线程对当前所有连接进行维护,动态修改这个链表,就可以实现对连接池的维护。

活动态是指数据库连接处于正常服务状态,能够对数据操作进行响应。恢复态就是日志处理状态,当一个新节点加入时需要处理日志进行数据同步,处于这个过程连接被称为恢复态,处于恢复态的连接不对外提供服务。失败态就是无法和数据库建立连接状态,但是数据库所在的主机是可达的,原因可能是数据库系统出现了问题。关机(断网)就是无法建立连接,同时数据库所在的主机也是不可达的,也就是无法连通。将不能提供服务时的状态细化为失败和关机(断网) 2 个状态,是因为在主机不可达的情况下建立连接非常消耗资源,因此在尝试建立连接前一定要确认数据库所在主机是可达的。

连接维护线程认为失败态和关机态是不稳定的,所以它会定时探测处于这 2 个状态的节点,发现它们恢复后就加入服务功能池。

它们之间的状态转换关系如图 4 所示。

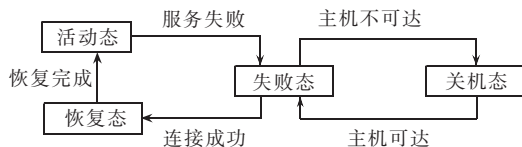


图 4 连接状态转换图

Fig.4 State machine of connection

3.4 日志服务模块设计与实现

对于非活动态的数据库写操作过程其实就是写日志的过程,读操作不需要写日志。日志处理过程就是把日志的操作重新写入数据库。日志处理过程可以被突然中断和恢复,即处理日志过程中突然断电,下次可以从出错的位置继续进行。

日志的格式:出于对日志扩展性和开放性的考虑,选用记录 SQL 的方式记录日志。把对数据库中所有会引起数据库变化的操作用标准的 SQL 语句记录下来,这样就可以用于不同的数据库系统。为了方便日志操作,针对每个数据库都存在对应的日志文件,这样日志中就可以不包含节点信息。考虑到不同机器的时间不一致性,时间信息用流水号替代,同时,这个流水号也是处理日志过程中需要的一个重要标记。这样,日志就是一个流水号和 SQL 语句的二元组,即 LOG: <No., SQL>。

使用日志同步过程:使用日志同步过程就是读入日志文件,然后对不同步数据库顺序进行 SQL 操作的过程。每一步或几步操作以后,记录处理到的流水号,用于出错恢复^[16]。

3.5 作业调度模块设计与实现

作业调度模块是管理模块,它管理和协调连接池

模块和日志模块,同时对用户的请求进行响应,完成用户接口的封装。加载中间件 MultiDbGate 后它立即工作。运行流程为:首先初始化 COM 环境,然后初始化连接池,初始化成功后启动连接池维护线程和日志服务线程,等待用户调用。

对于用户的读操作,它会选用一个活动数据库进行操作,把结果返回。写操作根据连接状态的不同分为 2 种操作,一种是数据库操作,一种是日志操作。只有等所有的连接都操作完成,它才返回写操作成功。如果数据库操作失败,自动转换为写日志操作。不同的连接可以并行处理,提高效率。

4 结果测试

MultiDbGate 已经在南方电网的多个 220 kV 变电站监控系统^[17]投入工程应用,结果表明其高可用性完全可以满足电力领域的苛刻要求。与不提供高可用的数据库直接访问比较,读操作性几乎没有任何影响;写操作随着数据库个数的增多会有所下降,如图 5 所示(图中, n 为数据库个数)。

测试环境:数据库

选用 SQL Server 2000 标准版;5 台 Dell PC 机组成的百兆以太网(双网),CPU 主频 2.8 GHz,内存 512 M。

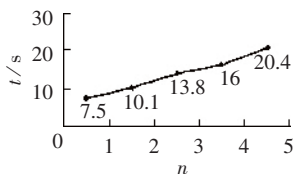


图 5 MultiDbGate 写操作性能曲线

Fig.5 Writing performance curve of MultiDbGate

a. 插入一万条记录测试(每条插入记录长均为 3 500 Byte)。在

只有一个数据库的情况下,和直接使用 ADO 访问数据库性能差别不大,约为 7~8 s。但是,随着数据库个数的增多,花费的时间也逐渐增多,在数据库个数小于 5 的情况下呈类线性关系。所以,在满足高可用需求的前提下,尽可能减少数据库个数。一般,大多数的厂站监控系统工程应用选用 1+1 冗余,也就是双机备份就够了。

b. 检索记录测试。在数据库建立的测试表中存在 10 万条记录,建有索引,检索满足条件的 10 条记录生成报表,结果用时 1~2 s,和数据库个数没有任何关系。

由此可见,使用 MultiDb Gate 访问数据库对性能没有显著影响,特别是对监控系统而言,它写历史库在时间上是分散的,对性能要求不高,但是检索操作是集成的,性能要求高,这恰好符合该中间件的性能特点。

5 结语

由于中间件技术可以简化分布式系统的开发和应用,所以得到了越来越广泛的应用。针对电力自动化领域对高可用存储服务的需求,设计并实现了一个高可用的数据库中间件 MultiDbGate,实现了廉价

而又简单的高可用数据服务。经过测试和现场工程应用表明,该中间件的性能和可靠性都表现良好,可以满足电力自动化监控系统对数据存储的需求。

参考文献:

- [1] DANEELS A, SALTER W. What is SCADA[C]//International Conference on Accelerator and Large Experimental Physics Control Systems. Trieste, Italy: ICALEPCS 99, 1999: 339-343.
- [2] 李佐卿. 开放型数据库互连 ODBC[J]. 计算机应用研究, 1995, 12(2): 1-3.
- LI Zuo-qing. Open database connectivity[J]. Application Research of Computers, 1995, 12(2): 1-3.
- [3] 李纪云, 董小社, 童端. 分布式对象中间件技术研究[J]. 计算机工程与设计, 2004, 25(2): 170-173.
- LI Ji-yun, DONG Xiao-she, TONG Duan. Research on distributed objects middleware[J]. Computer Engineering & Design, 2004, 25(2): 170-173.
- [4] JIN Hai, HWANG Kai. Stripped mirroring RAID architect[J]. Journal of Systems Architecture, 2000, 46(6): 543-550.
- [5] BUYA R. High performance cluster computing: architectures and systems[M]. NJ, USA: Prentice Hall PTR, 1999.
- [6] PRESTON W C. Using SANs and NAS[M]. Cambridge, MA: O'Reilly & Associates, 2002.
- [7] 耿冲. SQL Server 2000 数据库管理[M]. 北京: 机械工业出版社, 2001.
- [8] 朱立谷. 基于备份、复制和镜像的数据库高可用技术的研究[J]. 计算机应用研究, 2002, 19(4): 41-43.
- ZHU Li-gu. Evaluation of backup, replication and mirroring technology for high availability databases[J]. Application Research of Computers, 2002, 19(4): 41-43.
- [9] 陈军, 莫则尧, 李晓梅, 等. 大规模并行应用程序的可扩展性研究[J]. 计算机研究与发展, 2000, 37(11): 1382-1388.
- CHEN Jun, MO Ze-yao, LI Xiao-mei, et al. Research on the scalability of the large scale parallel application programs[J]. Journal of Computer Research and Development, 2000, 37(11): 1382-1388.
- [10] 黄汛, 程治刚. 数据库连接池技术的应用研究[J]. 武汉大学学报: 工学版, 2002, 35(1): 98-101.
- HUANG Xun, CHENG Zhi-gang. Research on application of connection pool technique for database[J]. Engineering Journal

of Wuhan University, 2002, 35(1): 98-101.

- [11] 童端, 董小社, 李纪云, 等. 基于 OpenPBS 的机群作业管理系统的设计与实现[J]. 计算机工程与应用, 2004, 40(13): 123-125.
- TONG Duan, DONG Xiao-she, LI Ji-yun, et al. The design and implementation of OpenPBS-based cluster job management system[J]. Computer Engineering and Applications, 2004, 40(13): 123-125.
- [12] 李玉荣, 杨树强, 贾焰, 等. 分布式日志服务关键技术研究[J]. 计算机工程与应用, 2006, 42(7): 116-118.
- LI Yu-rong, YANG Shu-qiang, JIA Yan, et al. Research on the key technology of the distributed log service[J]. Computer Engineering and Applications, 2006, 42(7): 116-118.
- [13] 周治平. ADO 数据存取技术[J]. 计算机应用, 1999, 7(1): 23-25.
- ZHOU Zhi-ping. ADO data access technology[J]. Computer Applications, 1999, 7(1): 23-25.
- [14] SHELDON R. LQL: a beginner's guide [M]. 2nd ed. Berkeley: Osborne/Mcgraw, 2005.
- [15] 冯华. Visual C++ 数据库开发技巧与实例[M]. 北京: 机械工业出版社, 2000.
- [16] 宋广华, 杨长生. 基于混合日志的内存数据库恢复子系统[J]. 浙江大学学报: 理学版, 2001, 28(2): 164-168.
- SONG Guang-hua, YANG Chang-sheng. A main-memory database recovery subsystem based on hybrid logging[J]. Journal of Zhejiang University: Sciences Edition, 2001, 28(2): 164-168.
- [17] 李苇. 500 kV 变电站计算机监控系统的设计[J]. 电力系统自动化, 2001, 25(12): 60-62.
- LI Wei. Computerized supervisory control system in 500 kV substation[J]. Automation of Electric Power systems, 2001, 25(12): 60-62.

(责任编辑: 汪仪珍)

作者简介:

李纪云(1979-), 男, 山东邹城人, 工程师, 硕士研究生, 主要研究方向为电力监控系统、数据库技术(E-mail: leejiyun@163.com);

蒋衍君(1975-), 男, 黑龙江尚志人, 工程师, 硕士研究生, 主要研究方向为电网自动化系统;

金文凯(1969-), 男, 江苏苏州人, 副主任, 高级工程师, 主要研究方向为电网自动化系统。

Design and realization of high-availability database middleware for substation supervisory control system

LI Ji-yun, JIANG Yan-jun, JIN Wen-kai

(R & D Center, Guodian Nanjing Automation Co., Ltd., Nanjing 210003, China)

Abstract: Power system demands high reliability of data storage. A database middleware with high-availability, MultiDbGate, is designed and realized, which uses dynamic maintenance and dispatch of connection and access to mask the complexity of bottom layer connection, multi database, dual nets, etc., and provides cheap and simple database services with high-availability and expansibility. The structure of MultiDbGate is divided into four modules: single database access, connection pool, job management and log service. The single database access module is the encapsulation of ADO (Active Data Objects). As the core, the connection pool module makes applications sharing the connections and masks the distributed architecture of bottom layer. The job management module schedules the requests of applications. The log service module ensures the data consistency by log records. The MultiDbGate runs well in supervisory control systems of several 220 kV substations.

Key words: substation supervisory control system; middleware; database; high-availability