

基于物理隔离技术的触发式数据库同步方案

尹飞,熊政,祝永晋

(江苏方天电力技术有限公司,江苏南京 211102)

摘要:数据库同步是不同安全区域应用中必不可少的、最重要的环节。数据库同步包括数据的抽取(Extract)和装载(Load)2个过程。电能量采集涉及的电能表数量多、采集周期短、数据容量较大,不适宜采用全量同步的方式,因此需要研究增量的数据抽取和数据装载。

关键词:物理隔离;数据库同步;数据抽取;数据装载

中图分类号:TM933.4;TP392 **文献标志码:**B

文章编号:1009-0665(2010)01-0035-03

江苏电网电能量计量系统是江苏省电力公司为加强内部管理和输电网内部成本核算,实现集约化经营的需要,并为电力市场的进一步发展提供技术与数据储备而建。该系统主要用于完成下网侧计量关口点的电能量采集、处理,系统的采集全部在 II 区调度数据网,而在此基础上做的应用基本都在 III/IV 区甚至要面向 Internet,因此数据库的同步是必须的环节。电能量采集涉及的电能表数量多、采集周期短,数据容量较大,不适宜采用全量同步的方式,因此需要研究增量式的数据抽取和数据装载。

1 数据增量更新方法比较

为了有效捕捉增量数据变化,对已有数据的扫描是必须面对的问题^[1]。目前有 5 种技术可以达到数据增量更新的要求,如图 1 所示。

方法 1 是扫描已经打上时间戳的数据。当一个应用对记录的最近一次修改打上时间戳时,同步程

收稿日期:2009-10-23;修回日期:2009-11-20

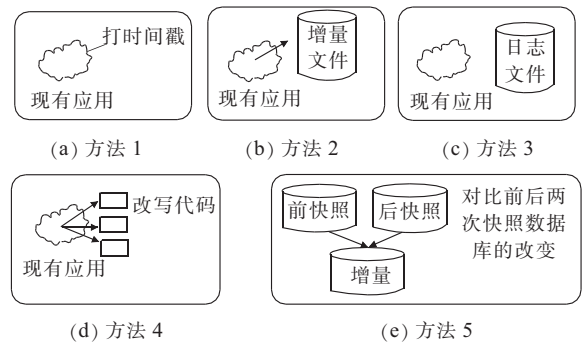


图 1 几种数据增量更新方法

序扫描就能够很有效地进行。

方法 2 是扫描增量文件。增量文件仅记录应用中所发生的改变,有了增量文件,扫描的过程就会高效。增量文件的生成可以由改造应用来完成,但所带来的问题是增加了数据库容量,且不方便人工干预。

方法 3 是扫描日志文件或审计文件。利用日志文件的一个问题就是它的内部格式是针对系统用途构建的,必须使用某种技术手段作为日志文件内容

[3] 蒙健明,曹继丰,王电处,等. 兴仁换流站站用电系统自备定值配合研究[J]. 电力建设,2008,29(8):37-40.

作者简介:

钱波(1983-),男,江苏常州人,助理工程师,研究方向为

500 kV 变电站运行与检修;

何正兴(1967-),男,江苏常州人,技师,从事 500 kV 变电站运行与管理工作;

俞澄一(1958-),男,江苏常州人,工程师,从事 500 kV 变电站运行与管理工作。

Analysis and Improvement for Stand-by Circuit of Self-demand System in Wunan Substation

QIAN Bo, HE Zheng-xing, YU Cheng-yi

(Changzhou Power Supply Company, Changzhou 213003, China)

Abstract: Based on the analysis of the principle and circuit structure of stand-by system, the voltage variation of the low-voltage side at the condition that the high-voltage side loses power supply is discussed in the paper. Focusing on the problems existing in stand-by system, measures for the improvement of the system are proposed. The improved stand-by circuit can be operated successfully when the circuit of high-voltage side is interrupted at any single phase. The reliability of self-demand system is guaranteed.

Key words: stand-by system; stand-by circuit; single phase interruption

输出的接口,因而使用完全依赖于专门的数据库厂商,依赖性太大,不灵活。

方法4就是修改所有应用程序的数据更新代码,统一生成增量数据拷贝。但是,这种方法通常行不通,主要是因为基于数据的应用太多,要实现此方法必须修改所有程序。

方法5是将“前”,“后”两个映像文件进行比较。使用此方法需在第1次抽取时就对数据库进行快照(Snapshot)。进行另外一个抽取时,就进行另外一个快照。再将两个快照逐次比较来确定数据的修改。

根据电能量计量系统数据库的现状,结合了第1和第2这2种方法,制定了增量数据识别方案。

2 方案概述

2.1 总体思路

在全量数据源上打时间戳,在数据插入、更新时写入时间戳,在删除数据时记录下该条记录的删除时间戳、表名、主键,根据系统的配置定期扫描表内时间戳和删除记录时间戳,判断时间顺序优化生成增量数据,生成数据包,在目标数据处匹配/装载。由于基于结果的同步方式,数据装载时的顺序可能不满足源数据库数据表之间的约束条件,在目标数据库上应当关闭数据表之间的约束^[2]。程序框架如图2所示。

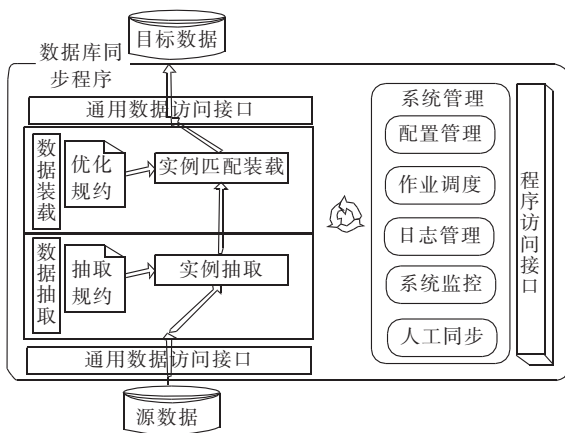


图2 程序框架

2.2 基本原则

不影响源数据安全,尽可能少影响源数据库性能;数据同步过程可以控制,同步时间粒度由用户自行设置,同步进程可以监控;要同步的数据表可配置,同步时间粒度可单独设置;支持物理隔离装置和网络两种方式同步,支持向多个目标数据库同步;支持监视新建的数据表,自动在目标数据库上新建新表,默认同步新表。同步程序分为服务端程序和客户端程序两个部分,服务端程序负责增量数据规则设置、自动创建触发器、自动添加时戳字段、定期

自动抽取数据、优化抽取的数据、生成同步数据包,客户端程序负责定期获取同步数据包、解析同步数据包、匹配/装载同步数据。程序数据流程见图3。

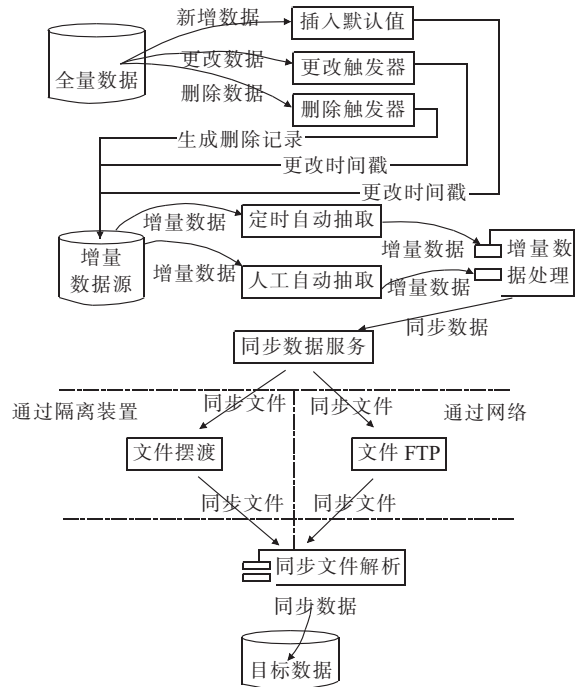


图3 程序数据流程

3 方案介绍

3.1 现有数据库改动

为了能在现有数据库上捕获数据改动,知道数据更新时间,需对现有的数据库进行改动,增加对应的域及触发器。各种改动情况的说明如表1所示。

表1 数据库改动情况

改动名	内 容	作 用
加更新管理表	增加一张表 FA_Table_Manage	记录针对每一张表的个性配置及最后同步时间
加删除记录表	增加一张表 FA_Delete_Record	记录系统中删除记录的时间、主键
加时间戳字段	针对现有的数据表,增加字段 FA_Data_Time	在插入数据时生成默认的时间戳
加更新触发器	针对现有的数据表,增加更新触发器 FA_TG_UP_XXXX	在数据表更新时间时写入时间戳
加删除触发器	针对现有的数据表,增加删除触发器 FA_TG_DL_XXXX	在数据删除时,插入删除记录到删除记录表

(1) 更新管理表:表中记录数据表的 Schema、名称、是否需同步、同步时间粒度、最后同步时间,配合定期扫描数据库中的新建表,可自动同步建新表到目标数据库。

(2) 删除记录表:因为不是基于日志方式,所以必须记录下被删除的记录,相对于日志方式的优势在于无需记录下该条记录的所有信息,只要记录下

主键信息。我们仔细研究了电能量计量系统中所有数据表的结构,只有极个别表没有主键,因此只需对一两张表进行改动。此表将根据系统的配置要求仅保留一段时间以免数据量大而影响性能。

(3) 时间戳字段:为了能提取出增量,时间戳是必须的,配合删除记录表就可以抽取出完整的增量数据,同时也可以实现人工抽取增量。

3.2 增量数据抽取

针对电能量计量系统的 DB2 数据库的增量数据抽取可以采用 2 种方法:DB2 提供的 Export 方法和语句生成法^[3]。

(1) Export 方法:适用于增量数据非常大,对同步速度要求非常高的情况,将增量的数据用 DB2 的 CLI 指令 Export 导出到文件,再将文件中的数据通过 Import 命令导入到目标数据库中,从而完成数据从数据源到目标库的同步,一旦有数据插入失败,后面的数据将无法继续。

(2) 语句生成法:此方法是将增量数据逐条生成语句,并对其中具有相关性的进行优化合并,生成一个 Command List,在目标数据库上,分组采用事务的方式提交,对于执行失败的语句将记录在日志中,后面的语句继续执行。

对比 2 种方法,其中 Export 方法对数据库的结合较紧,性能较好,但不具有通用性,同时程序可控性较差,因此采用语句生成法,便于对数据同步的全过程控制。

3.3 同步数据包

考虑到同步可能面对隔离装置和网络两种情况,统一采用数据包文件的方式来同步数据。同步数据包中主要包括生成的 Command List 和大值字段(LOB)文件。Command List 包括 2 种格式,一种是标准的 SQL 语句,客户端拿到后直接执行即可,一种是约定格式的装载大值字段的语句,该语句需要客户端程序进行解析后执行。为了提高执行效率,建议两种格式的语句分开以指定的文件名单独存放。

3.4 数据装载

客户端程序拿到数据包之后,需对指定文件名

的 Command List 进行解析,顺序是先执行纯 SQL 的 List,再执行加载大值字段的 List。在执行过程中出现错误的语句记录在日志中,在整个过程结束后统一在监控界面显示,可重试装载,并可根据用户配置自动重试及重试失败后发送相关反馈至服务端。

4 方案特点

(1) 基于结果的增量抽取,无需对原有的应用程序做改动,没有基于过程同步时的顺序性,减少了不同数据表之间的干扰。

(2) 由于各个数据表之间同步的相对独立性,可以对要同步表进行个性化配置,设置是否同步及同步的时间粒度。

(3) 采用同步数据包来传递同步数据,兼容了物理隔离装置及网络方式,适用范围更广。

(4) 采用语句生成法同步数据,具有较强的可控性,同时也具有较强的通用性。

5 结束语

基于结果的数据库增量同步,抓住了数据变动的源头,考虑了电能量计量系统应用现状,提高了抽取的准确性、数据装载的效率,同时提供了健壮的异常处理能力,并且具有很强的灵活性和适用性。当然数据装载的效率还需要进一步提高以便满足数据量日益增加的需要。

参考文献:

- [1] 史晶波. 在 DB2 中提取增量数据的一种方法[J]. 计算机与数字工程, 2004, 32(6): 15-16.
- [2] 萨师煊. 数据库系统概论(第 3 版)[M]. 北京: 高等教育出版社, 2002.
- [3] 李志伟. DB2 基础教程[M]. 北京: 清华大学出版社, 2003.

作者简介:

- 尹飞(1978-),男,安徽巢湖人,工程师,从事计量自动化相关工作;
熊政(1976-),男,江西南昌人,工程师,从事计量自动化相关工作;
祝永晋(1981-),男,江苏南京人,工程师,从事计量自动化相关工作。

The Synchronization Program of Trigger Database Based on Physical Isolation Technology

YIN Fei, XIONG Zheng, ZHU Yong-jin

(Jiangsu Frontier Electric Technologies Co., Ltd., Nanjing 211102, China)

Abstract: The synchronization of database is an indispensable and important part in different implementation fields. The process of database synchronization contains data extract and data load. Because of numerous sheets, short acquisition cycle and large capacity of power data, total synchronization is not suitable. Thus, data extract and load based on increment is studied in this paper.

Key words: physical isolation; database synchronization; data extract; data load