



## (12)发明专利

(10)授权公告号 CN 103970833 B

(45)授权公告日 2017.08.15

(21)申请号 201410131529.7

(22)申请日 2014.04.02

(65)同一申请的已公布的文献号

申请公布号 CN 103970833 A

(43)申请公布日 2014.08.06

(73)专利权人 浙江大学

地址 310027 浙江省杭州市西湖区浙大路  
38号(72)发明人 李莹 姚祥龙 尹建伟 邓水光  
吴健 吴朝晖

(74)专利代理机构 杭州天勤知识产权代理有限公司 33224

代理人 胡红娟

(51)Int.Cl.

G06F 17/30(2006.01)

权利要求书1页 说明书4页 附图1页

## (54)发明名称

一种基于日志的异构数据库同步系统中双向同步数据循环的解决方法

## (57)摘要

本发明公开了一种基于日志的异构数据库同步系统中双向同步数据循环的解决方法，包括以下步骤：在进行双向同步的数据库中分别建立特殊表；当增量数据由源数据库传输到目的数据库时，首先在目的数据库中插入一条特殊表数据，然后在目的数据库中执行完毕所有增量数据，最后在目的数据库手动提交增量数据事务；进行增量数据同步时具体操作如下：从日志中以事务为单位获取增量数据，检查事务中是否包含一条数据库B的特殊表B的插入数据，如果此条插入数据存在且数据的来源为数据库A，则不再将增量数据同步至数据库A；否则，将增量数据同步至数据库A。本发明在不影响数据库性能的同时，利用事务的原子性，简单方便地解决了数据循环的问题。

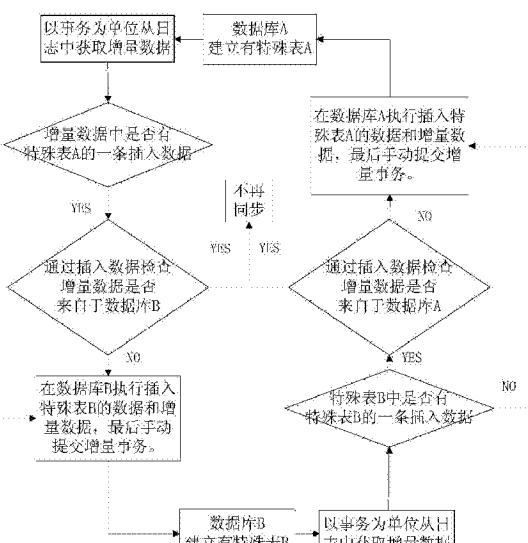
CN 103970833 B

(56)对比文件

WO 0075813 A1, 2000.12.14,  
CN 102436486 A, 2012.05.02,  
CN 103092871 A, 2013.05.08,  
Fateme Hassani et al.. "From unidirectional model transformation to incremental model synchronization".《Software Engineering (MySEC), 2011 5th Malaysian Conference in IEEE》.2011, 全文.

王玉标 等.“异构环境下数据库增量同步更新机制”.《计算机工程与设计》.2011, 第32卷(第3期), 全文.

审查员 郭明亮



1. 一种基于日志的异构数据库同步系统中双向同步数据循环的解决方法，所述异构数据库同步系统包括至少两个能够进行双向同步的数据库，两个数据库进行增量数据传递时，其中一个为提供增量数据的源数据库，另一个为接收增量数据的目的数据库，其特征在于，包括以下步骤：

(1) 在进行双向同步的数据库中分别建立特殊表，各个特殊表中用于记录所在数据库中增量数据的来源信息；

将所有进行双向同步的数据库的事务提交模式设定为手动事务提交模式；

(2) 当增量数据由源数据库传输到目的数据库时，首先在目的数据库中插入一条特殊表数据，该条特殊表数据记载此次增量数据的来源信息，然后在目的数据库中执行完毕所有增量数据，最后在目的数据库手动提交增量数据事务；

(3) 若两个能够进行双向同步的数据库分别为数据库A和数据库B，数据库A中建立有特殊表A，数据库B中建立有特殊表B，判断是否需要将数据库B中的增量数据同步至数据库A，具体操作如下：

从数据库B的日志中以事务为单位获取增量数据，检查事务中是否包含一条数据库B的特殊表B的插入数据，如果此条插入数据存在且此插入数据显示增量数据的来源为数据库A，则不再将增量数据同步至数据库A；否则，将增量数据同步至数据库A；

所述增量数据的来源信息包括三个字段数据，分别为主机IP、端口以及数据库名称；

表述增量数据来源信息的三个字段的数据类型均采用varchar；

判断数据的来源是否为数据库A，将数据来源信息的主机IP、端口以及数据库名称与数据库A的主机IP、端口以及数据库名称做比较，若三个字段均完全一致，则数据的来源为数据库A。

2. 如权利要求1所述的基于日志的异构数据库同步系统中双向同步数据循环的解决方法，其特征在于，依据预定时间间隔清除特殊表中的记录。

## 一种基于日志的异构数据库同步系统中双向同步数据循环的 解决方法

### 技术领域

[0001] 本发明涉及数据库同步技术领域,具体涉及一种基于日志的异构数据库同步系统中双向同步数据循环的解决方法。

### 背景技术

[0002] 数据库是IT架构中的一种基础构件,是对数据进行存储、管理的一种有效设施,现在几乎所有的应用都离不开数据库的支持。由于现在海量数据和数据访问地区分布不均衡,单个数据库已经不能承载所有的访问,一般需要在多个不同的地方分别部署一个数据库来提供服务,其中每一个数据库对外界来说都是数据完整和服务完整的,但是事实上每两个数据库之间的数据都是不一致的,所以这些不同地方的数据库之间就需要进行增量数据的同步来保证所有数据库的数据一致性和服务完整性。

[0003] 增量数据是指一个数据库在某一段时间内产生或变化的数据,按照操作类型区分又分为插入数据、删除数据、修改数据,分别对应SQL中insert、delete、update操作。

[0004] 数据库增量同步就是把增量数据从主数据库传输到从数据库中,以达到主从数据库的数据一致性,过程依次如下:获取A数据库的增量数据、传输到B数据库、在B数据库中执行SQL语句进行数据载入。

[0005] 不同地方的活动数据库之间需要进行增量双向同步,即A数据库的增量数据需要同步到B数据库,同时B数据库的增量数据也需要同步到A数据库。

[0006] 增量双向同步过程中会面临数据循环的问题,数据循环问题的产生是因为A数据库的增量数据同步到B数据库之后,这部分增量数据又会成为B数据库的增量数据再次同步到A数据库,然后这部分数据加上B数据库的增量数据又再次成为A数据库的增量数据,以此类推,增量数据就在不断增多、循环过程中。

[0007] 解决增量双向同步过程中数据循环问题的关键是在A数据库的增量数据同步到B数据库之后,不允许这部分数据再次同步回A数据库,而解决问题的难点在于识别B数据库的增量数据到底是不是A数据库产生的。

### 发明内容

[0008] 本发明提供了一种基于日志的异构数据库同步系统中双向同步数据循环的解决方法,在不影响数据库性能的同时,利用数据库事务的原子性,简单方便地解决了数据循环的问题。

[0009] 一种基于日志的异构数据库同步系统中双向同步数据循环的解决方法,所述异构数据库同步系统包括至少两个能够进行双向同步的数据库,两个数据库进行增量数据传递时,其中一个为提供增量数据的源数据库,另一个为接收增量数据的目的数据库,包括以下步骤:

[0010] (1)在进行双向同步的数据库中分别建立特殊表,各个特殊表中用于记录所在数

数据库中增量数据的来源信息；

[0011] 将所有进行双向同步的数据库的事务提交模式设定为手动事务提交模式；

[0012] (2)当增量数据由源数据库传输到目的数据库时,首先在目的数据库中插入一条特殊表数据,该条特殊表数据记载此次增量数据的来源信息,然后在目的数据库中执行完毕所有增量数据,最后在目的数据库手动提交增量数据事务；

[0013] (3)若两个能够进行双向同步的数据库分别为数据库A和数据库B,数据库A中建立有特殊表A,数据库B中建立有特殊表B,判断是否需要将数据库B中的增量数据同步至数据库A,具体操作如下：

[0014] 从数据库B的日志中以事务为单位获取增量数据,检查事务中是否包含一条数据库B的特殊表B的插入数据,如果此条插入数据存在且此插入数据显示增量数据的来源为数据库A,则不再将增量数据同步至数据库A;否则,将增量数据同步至数据库A。

[0015] 本发明的目的在于解决增量数据双向同步过程中的数据循环问题,因此,对于异构数据库中仅能实现单向同步的数据库的同步过程可以采用现有技术。

[0016] 本发明所述的异构数据库同步系统中能够进行双向同步的数据库不允许3个及以上相互之间构成双向数据循环回路,例如,数据库A和数据库B进行双向同步,数据库B和数据库C进行双向同步,数据库C和数据库A进行双向同步,则数据库A、数据库B和数据库C构成3个数据库的双向数据循环回路。

[0017] 所述特殊表为利用标准SQL建立的名称固定的表,各个特殊表中用于记录相应数据库中增量数据的来源信息,能够表征增量数据的来源信息的字段可以依据需要进行选择,例如,所述增量数据的来源信息包括三个字段,分别为主机IP、端口以及数据库名称。

[0018] 为了保持异构数据库中各个特殊表的兼容性,表述增量数据来源信息的三个字段的数据类型均采用varchar。

[0019] 每次记录增量数据的来源信息时,在特殊表中插入一条数据,数据包括能够表述该增量数据的来源信息的三个字段数据,分别为主机IP、端口以及数据库名称。

[0020] 将增量数据事务提交模式设定为手动提交模式是为了保持数据库中增量数据和增量数据来源信息的一致性,执行整个增量数据事务后,由用户手动将增量数据提交到数据库中,使数据库发生变化,增量数据和增量数据来源信息一起被写入数据库。

[0021] 首先从数据库B的日志中以事务为单位获取增量数据,然后检查事务中是否包含特殊表的插入数据,如果不存在,那么增量数据肯定不是来源于数据库A,如果存在,那么根据此条插入数据判断增量数据的来源是否为数据库A的方法为将数据来源信息的主机IP、端口以及数据库名称与数据库A的主机IP、端口以及数据库名称做比较,若三个字段均完全一致,则增量数据的来源为数据库A。

[0022] 在特殊表中插入记录之后,所有操作包括特殊表中插入的记录,都会被记录在日志文件中,所以不需要保存特殊表的数据,为了防止数据表中的数据量过大,依据预定时间间隔清除特殊表中的记录。

[0023] 本发明一种基于日志的异构数据库同步系统中双向同步数据循环的解决方法,在不影响数据库性能的同时,利用事务的原子性,简单方便地解决了数据循环的问题。

## 附图说明

[0024] 图1为本发明基于日志的异构数据库同步系统中双向同步数据循环的解决方法的流程图。

## 具体实施方式

[0025] 下面结合附图,对本发明一种基于日志的异构数据库同步系统中双向同步数据循环的解决方法做详细描述。

[0026] 本发明适用的异构数据库同步系统包括至少两个能够进行双向同步的数据库,若某两个数据库仅能进行单向同步,则采用现有技术,若某两个数据库能够进行双向同步,则采用本发明的方法避免产生数据增量双向同步中的数据循环问题。

[0027] 为便于叙述,以下以能够进行双向同步的数据库A和数据库B为例,详述同步方法,如图1所示,具体包括以下步骤:

[0028] (1)在数据库A中建立特殊表A,特殊表A中用于记录数据库A中增量数据的来源信息;在数据库B中建立特殊表B,特殊表B中用于记录数据库B中增量数据的来源信息。

[0029] 特殊表为标准SQL建立的一个名称固定的表,包括能够表述增量数据来源信息的主机IP、端口、数据库名称三个字段,为了数据库之间的兼容性,使用varchar作为特殊表中所有字段的数据类型。

[0030] 建立特殊表的SQL语句为:create table database\_sync (host\_str varchar (50),port\_str varchar (50),database\_str varchar (50))。

[0031] (2)来自数据库A的增量数据传输到数据库B,将数据库B中的事务提交模式设定为手动提交模式,首先在数据库B中执行插入一条特殊表数据(该条数据记载此次增量数据的来源信息),然后执行所有的增量数据,最后手动将事务提交到数据库B。

[0032] 插入数据记录增量数据的来源信息,插入数据包括增量数据来源数据库所对应的主机IP、端口和数据库名称,分别对应特殊表B中的三个字段,每个字段均为字符串类型。

[0033] (3)首先从数据库B中的日志中以事务为单位获取增量数据,然后检查事务中是否包含一条特殊表B的插入数据,若需要将数据库B中的增量数据同步到数据库A,则进行以下判断:

[0034] a)、若事务中不包含一条特殊表B中的插入数据,则将相应数据同步至数据库A;

[0035] 增量数据可能来自于数据库的同步,也可能来自于用户的操作,即用户对数据库中的数据做插入、删除或者修改,若增量数据来自于数据库的同步,则该增量数据有来源记录;若增量数据来自于用户的操作,则该增量数据没有来源记录,所有没有来源信息的增量数据都需要同步至数据库A。

[0036] b)、若事务中包含一条特殊表B中的插入数据记录,但此插入数据显示增量数据的来源并非数据库A,则将相应数据同步至数据库A;

[0037] 若特殊表B中的插入数据记录的增量数据来源信息的主机IP、端口以及数据库名称与数据库A的主机IP、端口以及数据库名称不完全一致或者完全不一致,则该增量数据不是数据库A的增量数据,需要同步至数据库A;

[0038] c)、若事务中包含一条特殊表B中的插入数据记录,且此插入数据显示增量数据的来源为数据库A,则不再将数据同步至数据库A;

[0039] 若特殊表B中的插入数据记录的增量数据来源信息的主机IP、端口以及数据库名

称与数据库A的主机IP、端口以及数据库名称做比较,若三个字段均完全一致,则数据的来源为数据库A。

[0040] (4) 若需要将数据库A中的增量数据同步到数据库B中,则采取和步骤(3)相类似的操作。

[0041] (5) 启动一个守护线程定时清除特殊表的数据,防止特殊表中的数据量过大。

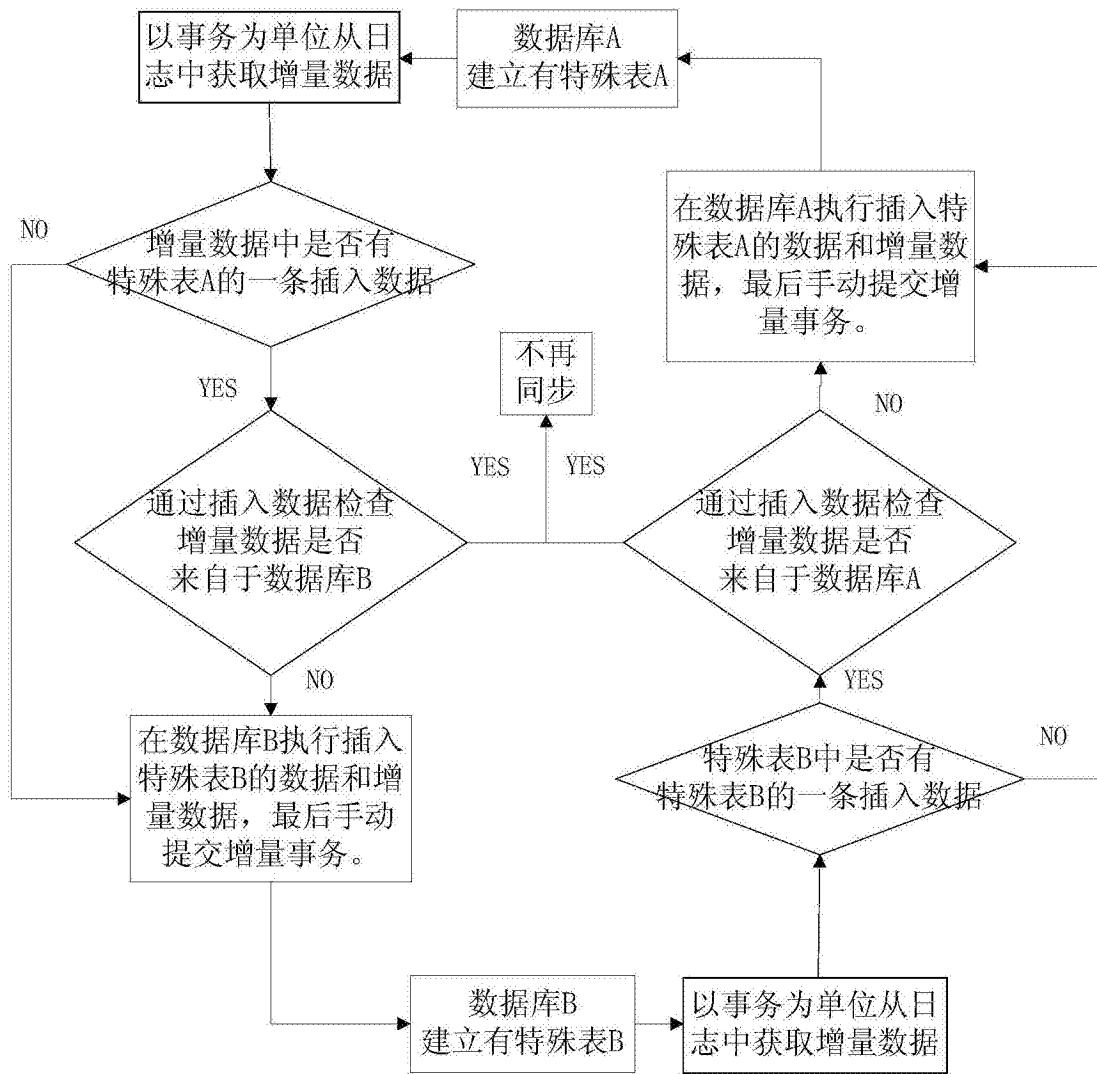


图1