(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 103136070 A
(43)申请公布日 2013.06.05

(21)申请号 201110391482.4

(22)申请日 2011.11.30

(71)申请人 阿里巴巴集团控股有限公司
地址 英属开曼群岛大开曼资本大厦一座四层847号邮箱

(72)发明人 李圣陶

(74)专利代理机构 北京润泽恒知识产权代理有限公司 11319
代理人 赵娟

(51)Int.Cl.
G06F 11/14(2006.01)

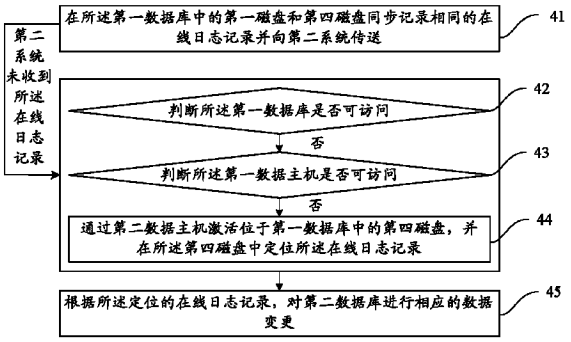
权利要求书2页 说明书11页 附图6页

(54)发明名称

一种数据容灾处理的方法和装置

(57)摘要

本申请提供了一种数据容灾处理的方法和装置，所述数据容灾涉及采用通信链路相连的第一系统与第二系统之间的数据备份处理，所述第一系统包括第一数据主机，第一数据库和第一存储设备，所述第二系统包括第二数据主机，第二数据库和第二存储设备；在所述第一数据库中的第一磁盘和第四磁盘同步记录相同的在线日志记录并向第二系统传送，若第二系统未收到所述在线日志记录，则执行以下步骤：判断所述第一数据库是否可访问；若否，则判断所述第一数据主机是否可访问；若否，则通过第二数据主机激活位于第一数据库中的第四磁盘，并在所述第四磁盘中定位所述在线日志记录。本申请可以在确保主库高可用性的前提下，实现数据的零丢失。



CN 103136070 A

1. 一种数据容灾处理的方法，其特征在于，所述数据容灾涉及采用通信链路相连的第一系统与第二系统之间的数据备份处理，所述第一系统包括第一数据主机，第一数据库和第一存储设备，所述第二系统包括第二数据主机，第二数据库和第二存储设备；所述第一存储设备在第一数据库中分配第一磁盘，在第二数据库中分配第二磁盘；第二存储设备在第二数据库中分配第三磁盘，在第一数据库中分配第四磁盘；

在所述第一数据库中的第一磁盘和第四磁盘同步记录相同的在线日志记录并向第二系统传送，若第二系统未收到所述在线日志记录，则执行以下步骤：

判断所述第一数据库是否可访问；

若否，则判断所述第一数据主机是否可访问；

若否，则通过第二数据主机激活位于第一数据库中的第四磁盘，并在所述第四磁盘中定位所述在线日志记录。

2. 根据权利要求 1 所述的方法，其特征在于，还包括：

当所述第一数据库不可访问，但所述第一数据主机可以访问时，通过所述第一数据主机提取所述在线日志记录的信息，并依据所述在线日志记录的信息在第一数据库中定位所述在线日志记录。

3. 根据权利要求 1 或 2 所述的方法，其特征在于，根据第一数据库的数据变更在所述第一磁盘和第四磁盘同步记录在线日志记录，所述的方法还包括：

根据所述定位的在线日志记录，对第二数据库进行相应的数据变更。

4. 根据权利要求 3 所述的方法，其特征在于，所述第一系统为主系统，所述第一数据库为主库，所述第二系统为备系统，所述第二数据库为备库，所述的方法还包括：

切换所述第二数据库为新的主库。

5. 根据权利要求 4 所述的方法，其特征在于，还包括：

当所述第一数据库和第一数据主机可访问时，切换所述第一数据库为新的备库；

打开所述新的主库接收访问，并重置在线日志记录的传递关系为，在第二数据库的第二磁盘和第三磁盘同步记录相同的在线日志记录并向第一系统传递。

6. 根据权利要求 4 所述的方法，其特征在于，所述第二系统还采用通信链路与第三系统相连，所述第三系统中包括第三数据库、第三数据主机和第三存储设备，第二存储设备在第二数据库中分配第五磁盘，在第三数据库中分配第六磁盘；所述第三存储设备在第三数据库中分配第七磁盘，在第二数据库中分配第八磁盘，所述第七磁盘和第八磁盘被同步写入相同的在线日志记录；

所述的方法，还包括：

切换所述第三数据库为新的备库；

打开所述新的主库接收访问，并重置在线日志记录的传递关系为，在第二数据库的第五磁盘和第八磁盘同步记录相同的在线日志记录并向第三系统传递。

7. 根据权利要求 5 所述的方法，其特征在于，还包括：

若所述第一数据库可以访问，则在作为主库的第一数据库上发起切换请求，依据该请求关闭并重启所述第一数据库；

若作为备库的第二数据库可接收访问，则将所述第二数据库切换为新的主库；

打开所述新的主库接收访问，并重置在线日志记录的传递关系为，在第二数据库的第

二磁盘和第三磁盘同步记录相同的在线日志记录并向第一系统传递。

8. 一种数据容灾处理的装置，其特征在于，包括采用通信链路相连的第一系统与第二系统；

其中，所述第一系统包括第一数据主机，第一数据库和第一存储设备，所述第二系统包括第二数据主机，第二数据库和第二存储设备；所述第一存储设备在第一数据库中分配第一磁盘，在第二数据库中分配第二磁盘；第二存储设备在第二数据库中分配第三磁盘，在第一数据库中分配第四磁盘；

所述装置还包括：

第一日志记录模块，用于在所述第一数据库中的第一磁盘和第四磁盘同步记录相同的在线日志记录；

日志传送模块，用于在所述在线日志记录在第一数据库中记录完后向第二系统传送；

日志恢复模块，用于在第二系统未收到所述在线日志记录时，调用以下子模块：

数据库访问判断子模块，用于判断所述第一数据库是否可访问；若否，则触发数据主机访问判断子模块；

主机访问判断子模块，用于判断所述第一数据主机是否可访问；若否，则触发激活定位子模块；

激活定位子模块，用于通过第二数据主机激活位于第一数据库中的第四磁盘，并在所述第四磁盘中定位所述在线日志记录。

9. 根据权利要求 8 所述的装置，其特征在于，所述日志恢复模块还包括：

日志提取子模块，用于在所述第一数据库不可访问，但所述第一数据主机可以访问时，通过所述第一数据主机提取所述在线日志记录的信息，并依据所述在线日志记录的信息在第一数据库中定位所述在线日志记录。

10. 根据权利要求 9 所述的装置，其特征在于，所述第一系统为主系统，所述第一数据库为主库，所述第二系统为备系统，所述第二数据库为备库，所述的装置还包括：

主库切换模块，用于切换所述第二数据库为新的主库。

一种数据容灾处理的方法和装置

技术领域

[0001] 本申请涉及数据安全的技术领域，特别是涉及一种数据容灾处理的方法和一种数据容灾处理的装置。

背景技术

[0002] 数据容灾，就是指建立一个异地的数据系统，该系统是本地关键应用数据的一个可用复制。在本地数据及整个应用系统出现灾难时，系统至少在异地保存有一份可用的关键业务的数据。该数据可以是与本地生产数据的完全实时复制，也可以比本地数据略微落后，但一定是可用的。其采用的主要技术是数据备份和数据复制技术。数据容灾的处理，实际上是异地数据复制的处理。

[0003] 以 Oracle 数据库 (Oracle Database，又名 Oracle RDBMS，或简称 Oracle) 的数据容灾为例，Oracle 数据库为保证高可用性，高可靠性，在同城的两个机房内分别部署主库（主数据库）和备库（备份数据库），并使用 Oracle DataGuard(Oracle 数据保护）技术进行主库和备库间的数据同步，为了确保主库的高可用性，DataGuard 使用 MaxPerformace 模式，即以异步的方式将主库日志信息写到备库。

[0004] 数据同步的基本流程是当主库产生日志时，通过事先配置的传送方式，以同步或者异步的方式传送到备库，备库通过恢复进程将日志进行应用，实现主库备库间的数据复制。现有的 Oracle DataGurad 技术提供了三种标准数据复制的方式，即 MaxPerformance（最大持久化）模式、MaxAvailability（最大可行性）模式和 MaxProtection（最大保护度）模式。

[0005] 具体而言，采用 MaxPerformance 模式，主库向备库传递日志的方式是异步的，也就是说，当主库的数据产生变化时，主库在保证本地日志文件写完毕后，不会等待远端备库正确、完整地接收了日志文件，就会继续完成后续的数据更新请求，如果此时主数据库发生故障，而备库没有完整的接收日志文件，则会发生数据丢失的情况。

[0006] 采用 MaxProtection 和 MaxAvailability 模式，主库向备库传递日志的方式都是同步的，也就是说，当主库的数据产生变化时，主库在确保本地及远程全部正确接收日志前，是不会进行后续数据处理的，即可实现数据的零丢失。但 MaxProtection 模式下，当备库出现问题时，即备库无法接收日志文件时，主库将自动关闭，即备库的状态将会影响到主库，导致主库的高可用性得不到保障。再者，在 MaxProtection 模式和 MaxAvailability 模式下，日志的传送都是通过网络的，网络具有一定的不稳定性，如延迟现象，在高并发、高压力的应用环境下，主库容易受到备库接收日志的影响，造成主库运行缓慢。

[0007] 综上，采用现有的 Oracle DataGurad 技术，其中的 MaxPerformance 模式将无法保证数据的零丢失，而 MaxAvailability 模式和 MaxProtection 模块无法保证数据库主库在高并发压力下数据库的持续稳定，即无法保证数据库主库的高可用性。

[0008] 因此，目前需要本领域技术人员迫切解决的一个技术问题就是：提出一种全新的数据容灾处理机制，用以在确保主库高可用性的前提下，实现数据的零丢失。

发明内容

[0009] 本申请的目的是提供一种数据容灾处理的方法和装置，用以在确保主库高可用性的前提下，实现数据的零丢失。

[0010] 为了解决上述问题，本申请公开了一种数据容灾处理的方法，所述数据容灾涉及采用通信链路相连的第一系统与第二系统之间的数据备份处理，所述第一系统包括第一数据主机，第一数据库和第一存储设备，所述第二系统包括第二数据主机，第二数据库和第二存储设备；所述第一存储设备在第一数据库中分配第一磁盘，在第二数据库中分配第二磁盘；第二存储设备在第二数据库中分配第三磁盘，在第一数据库中分配第四磁盘；

[0011] 在所述第一数据库中的第一磁盘和第四磁盘同步记录相同的在线日志记录并向第二系统传送，若第二系统未收到所述在线日志记录，则执行以下步骤：

[0012] 判断所述第一数据库是否可访问；

[0013] 若否，则判断所述第一数据主机是否可访问；

[0014] 若否，则通过第二数据主机激活位于第一数据库中的第四磁盘，并在所述第四磁盘中定位所述在线日志记录。

[0015] 优选的，所述的方法，还包括：

[0016] 当所述第一数据库不可访问，但所述第一数据主机可以访问时，通过所述第一数据主机提取所述在线日志记录的信息，并依据所述在线日志记录的信息在第一数据库中定位所述在线日志记录。

[0017] 优选的，根据第一数据库的数据变更在所述第一磁盘和第四磁盘同步记录在线日志记录，所述的方法还包括：

[0018] 根据所述定位的在线日志记录，对第二数据库进行相应的数据变更。

[0019] 优选的，所述第一系统为主系统，所述第一数据库为主库，所述第二系统为备系统，所述第二数据库为备库，所述的方法还包括：

[0020] 切换所述第二数据库为新的主库。

[0021] 优选的，所述的方法，还包括：

[0022] 当所述第一数据库和第一数据主机可访问时，切换所述第一数据库为新的备库；

[0023] 打开所述新的主库接收访问，并重置在线日志记录的传递关系为，在第二数据库的第二磁盘和第三磁盘同步记录相同的在线日志记录并向第一系统传递。

[0024] 优选的，所述第二系统还采用通信链路与第三系统相连，所述第三系统中包括第三数据库、第三数据主机和第三存储设备，第二存储设备在第二数据库中分配第五磁盘，在第三数据库中分配第六磁盘；所述第三存储设备在第三数据库中分配第七磁盘，在第二数据库中分配第八磁盘，所述第七磁盘和第八磁盘被同步写入相同的在线日志记录；

[0025] 所述的方法，还包括：

[0026] 切换所述第三数据库为新的备库；

[0027] 打开所述新的主库接收访问，并重置在线日志记录的传递关系为，在第二数据库的第五磁盘和第八磁盘同步记录相同的在线日志记录并向第三系统传递。

[0028] 优选的，所述的方法，还包括：

[0029] 若所述第一数据库可以访问，则在作为主库的第一数据库上发起切换请求，依据

该请求关闭并重启所述第一数据库；

[0030] 若作为备库的第二数据库可接收访问，则将所述第二数据库切换为新的主库；

[0031] 打开所述新的主库接收访问，并重置在线日志记录的传递关系为，在第二数据库的第二磁盘和第三磁盘同步记录相同的在线日志记录并向第一系统传递。

[0032] 本申请实施例还公开了一种数据容灾处理的装置，包括采用通信链路相连的第一系统与第二系统；

[0033] 其中，所述第一系统包括第一数据主机，第一数据库和第一存储设备，所述第二系统包括第二数据主机，第二数据库和第二存储设备；所述第一存储设备在第一数据库中分配第一磁盘，在第二数据库中分配第二磁盘；第二存储设备在第二数据库中分配第三磁盘，在第一数据库中分配第四磁盘；

[0034] 所述装置还包括：

[0035] 第一日志记录模块，用于在所述第一数据库中的第一磁盘和第四磁盘同步记录相同的在线日志记录；

[0036] 日志传送模块，用于在所述在线日志记录在第一数据库中记录完后向第二系统传送；

[0037] 日志恢复模块，用于在第二系统未收到所述在线日志记录时，调用以下子模块：

[0038] 数据库访问判断子模块，用于判断所述第一数据库是否可访问；若否，则触发数据主机访问判断子模块；

[0039] 主机访问判断子模块，用于判断所述第一数据主机是否可访问；若否，则触发激活定位子模块；

[0040] 激活定位子模块，用于通过第二数据主机激活位于第一数据库中的第四磁盘，并在所述第四磁盘中定位所述在线日志记录。

[0041] 优选的，所述日志恢复模块还包括：

[0042] 日志提取子模块，用于在所述第一数据库不可访问，但所述第一数据主机可以访问时，通过所述第一数据主机提取所述在线日志记录的信息，并依据所述在线日志记录的信息在第一数据库中定位所述在线日志记录。

[0043] 优选的，所述第一系统为主系统，所述第一数据库为主库，所述第二系统为备系统，所述第二数据库为备库，所述的装置还包括：

[0044] 主库切换模块，用于切换所述第二数据库为新的主库。

[0045] 与现有技术相比，本申请包括以下优点：

[0046] 本申请通过对架构的调整，采用通信链路替代原来的以太网络，现有技术中数据库依赖很不稳定的网络层实现，以太网络环境的不稳定将导致本申请实施例在现实中无法有效实施。若通过光纤链路替代原来的以太网络，利用数据库在线日志组在本地同步写的特点和光纤网络高吞吐量低延迟的特点，应用本申请实施例的数据库容灾架构及恢复流程，即可满足在高并发压力下数据零丢失的需求，即兼顾数据的零丢失和高可用性。

附图说明

[0047] 图 1 是本申请的一种数据容灾处理所涉及的硬件架构的结构框图；

[0048] 图 2 是本申请一种数据容灾处理硬件架构的结构示意图；

[0049] 图 3 是本申请的一种数据容灾处理硬件架构中存储设备与数据库的结构示意图；

[0050] 图 4 是本申请基于上述数据容灾处理的硬件架构提出的一种数据容灾处理的方法实施例 1 的步骤流程图；

[0051] 图 5 是本申请的一种数据容灾处理的方法实施例 2 的步骤流程图；

[0052] 图 6 是本申请基于上述数据容灾处理的硬件架构提出的一种数据容灾处理的方法实施例 3 的步骤流程图；

[0053] 图 7 是本申请的一种数据容灾处理的装置实施例的结构框图。

具体实施方式

[0054] 为使本申请的上述目的、特征和优点能够更加明显易懂，下面结合附图和具体实施方式对本申请作进一步详细的说明。

[0055] 参照图 1，其示出了本申请的一种数据容灾处理所涉及的硬件架构的结构框图，具体可以包括第一系统 11 和第二系统 12，其中，所述第一系统 11 可以包括第一数据主机 111，第一数据库 112 和第一存储设备 113，所述第二系统 12 可以包括第二数据主机 121，第二数据库 122 和第二存储设备 123；所述第一存储设备 113 在第一数据库 112 中分配有第一磁盘 1121，在第二数据库 122 中分配有第二磁盘 1221；第二存储设备 123 在第二数据库 122 中分配有第三磁盘 1222，在第一数据库中分配有第四磁盘 1122。所述第一系统 11 和第二系统 12 之间采用通信延时非常短（如不超过 1 毫秒）的通信链路相连。

[0056] 参照图 2 所示的数据容灾处理硬件架构的结构示意图，在具体实现中，所述第一系统 11 和第二系统 12 可以为同城的两个机房 A 和 B 的系统，所述第一数据库和第二数据库可以为 Oracle 数据库，所述第一数据主机和第一数据库可以设置在第一服务器 110 内，所述第二数据主机和第二数据库可以设置在第二服务器 120 内，这两个机房系统可以通过光纤交换机 13 互联，构成一个大的光纤网络，两个机房系统中的服务器和存储设备通过这个光纤网络实现相互连接，即所述第一服务器 110 通过所述光纤网络与第一存储设备 113 连接，所述第二服务器 120 通过所述光纤网络与第二存储设备 123 连接。

[0057] 在具体实现中，Oracle 数据库可以通过 Online Redo Log file Group（在线日志组）记录数据变化，其中，每个在线日志组包括多个日志成员 (Member)，多个日志成员间的数据（内容）保持一致，而且同一个日志组中的日志成员的写入是同步的。可以将不同的日志成员放在不同的磁盘上，以实现容灾。所述多个日志组 Group 循环使用，如数据库有三组日志组，分别是 A，B，C，则写入顺序为 A- ＞ B- ＞ C- ＞ A- ＞ B- ＞ ..... 。

[0058] 参照图 3 所示的数据容灾处理硬件架构中存储设备与数据库的结构示意图，所述第一存储设备 113 在第一数据库 112 中分配有放在第一磁盘 1121 中的日志成员 redo1，在第二数据库 122 中分配有放在第二磁盘 1221 中的日志成员 redo2；第二存储设备 123 在第二数据库 122 中分配有放在第三磁盘 1222 中的日志成员 redo3，在第一数据库中分配有放在第四磁盘 1122 中的日志成员 redo4。在这种情形中，第一数据库 112 中有两个日志成员 redo1 和 redo4，redo1 是由第一存储设备分配的，redo4 是由第二存储设备分配的，redo1 和 redo4 被第一数据主机同步写入相同的在线日志记录；第二数据库 122 中有两个日志成员 redo2 和 redo3，redo2 是由第一存储设备分配的，redo3 是由第二存储设备分配的，在第二系统被切换为主系统时，redo2 和 redo3 被第二数据主机同步写入相同的在线日志记录。

[0059] 需要说明的是，第一系统通过网络向第二系统传递的日志文件，第二系统并不是保存在第二磁盘和第三磁盘上，而是保存在第二数据库上的其他磁盘上，以供第二系统恢复。第二系统的第二磁盘和第三磁盘，当其为备库角色时是没有使用到的，只有在其是主库角色时，用于写入在线日志文件才起到作用。

[0060] 参照图 4 所示的，本申请基于上述数据容灾处理的硬件架构提出的一种数据容灾处理的方法实施例 1，其步骤包括 ：

[0061] 步骤 41、在所述第一数据库中的第一磁盘和第四磁盘同步记录相同的在线日志记录并向第二系统传送，若第二系统未收到所述在线日志记录，则执行步骤 42 ～ 44 ；

[0062] 在线日志文件可以用于保护数据丢失，数据库在任何数据变更时都会先将变更日志作为在线日志记录写入在线日志文件。参考图 3，采用本步骤，当第一数据库发生数据变更时，会在日志成员 redo1 和 redo4 中同步记录相同内容的在线日志记录，在某个在线日志记录记录完成后，便传送至第二系统。

[0063] 步骤 42、判断所述第一数据库是否可访问 ；若否，则执行步骤 43 ；

[0064] 步骤 43、判断所述第一数据主机是否可访问 ；若否，则执行步骤 44 ；

[0065] 步骤 44、通过第二数据主机激活位于第一数据库中的第四磁盘，并在所述第四磁盘中定位所述在线日志记录。

[0066] 根据上述步骤 41 可以得知，第一系统向第二系统传递在线日志记录的方式是异步的，也就是说，当第一系统的数据产生变化时，在保证第一系统本地的在线日志记录记录完毕后，不会等待第二系统是否正确完整地接收了在线日志记录，如果此时第一数据库发生故障，如电源故障，而第二系统没有完整地接收在线日志记录，则会发生数据丢失的情况。

[0067] 针对这种情况，本申请实施例提出了在第一数据主机发生故障，并且第一数据库也发生故障时的处理机制，简而言之，即在所述第一数据库不可访问，且所述第一数据主机也不可访问的情况下，在第二数据主机上将其对应的第二存储设备分配给第一数据库中的第四磁盘（日志成员 redo4）激活，并在所述第四磁盘中查找到第一数据主机写入的在线日志记录（未传递的在线日志记录），以确保第二数据库能获得对应的在线日志记录，进而实现数据的零丢失。

[0068] 具体而言，由于第二系统的第二存储设备在第一数据库中分配有第四磁盘（日志成员 redo4），即所述第四磁盘虽然是与第二数据主机连接的第二存储设备中的磁盘，但其位于第一数据库中，接受第一数据主机的在线日志记录，即所述第四磁盘中记录了第一数据主机在传递所述在线日志记录之前写入的日志信息。因此，在当在线日志记录未能传递到第二系统，但第一数据库和第一数据主机均不可访问的情况下，可以通过在第二数据主机上执行相应的操作系统命令激活所述第四磁盘。以 IBM AIX 操作系统为例，激活所述磁盘的命令为 varyonvg。

[0069] 激活所述磁盘后，则可以定位当前第二系统缺失的在线日志记录，如通过 Oracle 数据库提供的视图，确定缺失的在线日志记录，具体可以采用如下代码 ：

[0070] SELECT THREAD#, LOW_SEQUENCE#, HIGH_SEQUENCE#FROM V$ARCHIVE_GAP ；

[0071] THREAD#LOW_SEQUENCE#HIGH_SEQUENCE# ；

[0072] 1 90 92

[0073] 根据上述 sequence# 号码，可以通过 ftp 或 scp 的方式，从第一数据库的第四磁盘中定位到相应的在线日志记录。

[0074] 在具体实现中，可能缺失的日志文件主要有历史日志文件和在线日志文件两种。例如，第一数据库当前已有 1 ～ 100 号日志文件，第二数据库当前仅有 1 ～ 97 号日志文件，则 98、99 号日志文件为历史日志文件，100 为当前的在线日志记录。

[0075] 在实际应用中，第一系统向第二系统传递日志信息主要是通过网络异步方式传送，第二系统一般只会缺少第一系统正在写的在线日志文件的一部分，即该在线日志文件中的一部分在先日志记录。首先，历史日志文件第二系统已经获取，比如第一系统目前的在线日志文件为 100，则前 99 个日志文件可以称为历史日志文件，这部分第二系统都已经通过网络方式获取了。历史日志文件缺失及补救不是本申请考虑的重点，在实际中有多种解决方法，本领域技术人员采用现有技术中的任一种方法均可。本申请关注的是当前在线日志记录的缺失及补救的情况。

[0076] 针对第 100 号在线日志文件，第一系统是在不断地边写入边向第二系统传递，但这个动作时异步的，第一系统不能保证写入第 100 号在线日志文件的内容都已经传递到第二系统，本申请就是为了保护这部分还没有传递到第二系统的在线日志记录，即在当第一系统故障时，对于未传递到第二系统的第 100 号日志文件中的相应部分的在线日志记录，通过在第二数据主机上激活第二存储设备分配给第一数据库的第四磁盘，定位对应的在线日志记录。

[0077] 作为本申请实施例具体应用的一种示例，所述第二数据库在第四磁盘中定位所述在线日志记录的操作可以通过如下代码实现：

[0078] recover standby database until cancel；

[0079] Specify log：{<RET> = suggested|filename|AUTO|CANCEL}；

[0080] /u01/oracle/oradata/bmw/redo01_???；

[0081] Log applied.

[0082] Media recovery complete.

[0083] 在本申请的一种优选实施例中，还可以包括如下步骤：

[0084] 步骤 46、根据所述定位的在线日志记录，对第二数据库进行相应的数据变更。

[0085] 如前所述，定位在线日志记录的目的是为了实现数据库中数据的零丢失。在本实施例中，在线日志记录是根据第一数据库的数据变更在所述第一磁盘和第四磁盘同步记录的，当所述在线日志记录未能传递到第二系统时，只要第二数据主机能够定位到所述在线日志记录，即可根据该在线日志记录对第二数据库中的相应数据进行对应变更。

[0086] 在实际中，可以先激活第二数据库，如采用如下代码激活第二数据库：alter database activate standby database；再根据当前所定位的在线日志记录对第二数据库的相应数据进行对应的变更。

[0087] 为使本领域技术人员更好地理解本发明，以下通过一个具体应用的示例进行说明。

[0088] 假设有两个机房，分别为机房 A 和机房 B，两个机房通过光纤交换机互连，在机房 A 中部署有数据主机 A，数据库 A 和存储设备 A，所述数据主机 A，数据库 A 和存储设备 A 通过光纤网络相连，在机房 B 中部署有主机 B，数据库 B 和存储设备 B，所述数据主机 B，数据库 B

和存储设备 B 通过光纤网络相连。

[0089] 两个存储设备分别向两个数据库分配两个磁盘，其中一个磁盘分配给本地机房数据库，另一个磁盘分配给远程机房数据库。即所述存储设备 A 在数据库 A 中分配第一磁盘，在数据库 B 中分配第二磁盘；所述存储设备 B 在数据库 B 中分配第三磁盘，在 A 数据库中分配第四磁盘；数据主机在创建在线日志组时，确保其中一个日志成员位于本地存储设备分配的磁盘，另一个日志成员位于远程存储设备分配的磁盘。在这种情况下，数据库 A 有两个日志成员 redo1 和 redo4，redo1 位于存储设备 A 分配的第一磁盘，redo4 位于存储设备 B 分配的第四磁盘，redo1 和 redo4 被数据主机 A 同步写入相同的在线日志记录；数据库 B 中有两个日志成员 redo2 和 redo3，redo2 位于存储设备 A 分配的第二磁盘，redo3 位于存储设备 B 分配的第三磁盘。

[0090] 根据 Oracle 数据库的特性，任何数据库中的数据修改都会在修改真实数据前，将数据修改的内容写入在线日志记录，同时一个在线日志组内的各个日志成员间是同步写入，其内容完全一致。

[0091] 当机房 A 发生故障时，机房 B 的数据主机 B 可以激活当初分配给机房 A 的第四磁盘，获取 redo4 中记录的在线日志记录，然后根据该在线日志记录恢复数据库 B，从而实现数据库 B 与数据库 A 中的数据完全一致，实现数据零丢失。

[0092] 参照图 5，其示出了本申请的一种数据容灾处理的方法实施例 2 的步骤流程图，所述数据容灾涉及采用通信链路相连的主系统与备系统之间的数据备份处理，所述主系统包括主数据主机，主库和主存储设备，所述备系统包括备数据主机，备库和备存储设备；所述主存储设备在主库中分配第一磁盘，在备库中分配第二磁盘；所述备存储设备在备库中分配第三磁盘，在主库中分配第四磁盘。

[0093] 本实施例具体可以包括如下步骤：

[0094] 步骤 51、在所述主库中的第一磁盘和第四磁盘同步记录相同的在线日志记录并向备系统传送，若备系统未收到所述在线日志记录，则执行步骤 52 ～ 56；

[0095] 步骤 52、判断所述主库是否可访问；若否，则执行步骤 54；

[0096] 步骤 53、判断所述主数据主机是否可访问；若否，则执行步骤 55；

[0097] 步骤 54、通过备数据主机激活位于主库中的第四磁盘，并在所述第四磁盘中定位所述在线日志记录；

[0098] 步骤 55、根据所述定位的在线日志记录，对备库进行相应的数据变更；

[0099] 步骤 56、切换所述备库为新的主库。

[0100] 本实施例可以在当前主库和主数据主机不可访问的情况下，通过备数据主机定位到未传递到备系统的在线日志记录，并根据该在线日志记录对备库进行数据变更，然后将正常工作的备库切换为新的主库。

[0101] 在具体实现中，本实施例还可以包括如下步骤：

[0102] 步骤 57、当所述主数据库和主数据主机可访问时，切换所述主库为新的备库；

[0103] 步骤 58、打开所述新的主库接收访问，并重置在线日志记录的传递关系为，在新主库的第二磁盘和第三磁盘同步记录相同的在线日志记录并向新备库传递。

[0104] 应用本实施例，可以置换系统的主备关系，即在主系统出现的故障的情况下，将原来的备系统切换为主系统，而在原来的主系统恢复正常（可接受访问）后，将其切换为备系

统，并重置其日志传递关系。

[0105] 在实际中，若主系统在一定时间内未能恢复正常或在其他情况下，备系统也可以与其它相连的系统重置主备关系。例如，所述备系统还采用通信链路与第三系统相连，所述第三系统中包括第三数据库、第三数据主机和第三存储设备，备存储设备在备库中分配有第五磁盘，在第三数据库中分配有第六磁盘；所述第三存储设备在第三数据库中分配有第七磁盘，在备库中分配有第八磁盘，所述第七磁盘和第八磁盘被同步写入相同的在线日志记录；

[0106] 在这种情况下，则可以通过以下步骤重置系统间的主备关系：

[0107] 切换所述备库为新的主库，切换所述第三数据库为新的备库；

[0108] 打开所述新的主库接收访问，并重置在线日志记录的传递关系为，在备库的第五磁盘和第八磁盘同步记录相同的在线日志记录并向第三系统传递。

[0109] 可以理解，本申请实施例不仅适用于同城双机房的部署，还适用于不限位置范围的多机房部署，或者单机房内多服务器之间的部署，但需要保证机房之间通信链路的延时非常短，如在 1 毫秒之内。就目前的技术而言，所述通信链路可以采用光纤链路以保证所述延时，主库与备库可以采用通用的 Oracle 数据库。现有技术中 Oracle 数据库依赖很不稳定的网络层实现，以太网络环境的不稳定将导致本申请实施例在现实中无法有效实施。若通过光纤链路替代原来的以太网络，利用 Oracle 数据库在线日志组在本地同步写的特点和光纤网络高吞吐量低延迟的特点，应用本申请实施例的数据库容灾架构及恢复流程，即可满足在高并发压力下数据零丢失的需求，即兼顾数据的零丢失和高可用性。

[0110] 参照图 6 所示的，本申请基于图 1、图 2 和图 3 所示的数据容灾处理的硬件架构，提出了一种数据容灾处理的方法实施例 3，其步骤包括：

[0111] 步骤 61、在所述第一数据库中的第一磁盘和第四磁盘同步记录相同的在线日志记录并向第二系统传送，若第二系统未收到所述在线日志记录，则执行步骤 62 ～ 67；

[0112] 步骤 62、判断所述第一数据库是否可访问；若是，则执行步骤 63；若否，则执行步骤 64；

[0113] 在实际中，可以通过一个测试账号，循环访问所述第一数据库，查询某个测试数据，当可以取得数据时，即可判定该第一数据库可访问；当无法取得数据时，即可判定该第一数据库发生故障，不可访问。

[0114] 步骤 63、执行主备库切换 (switchover) 的操作，具体可以包括如下执行子步骤：

[0115] 子步骤 S11、若所述第一数据库可以访问，则在作为主库的第一数据库上发起切换请求，依据该请求关闭并重启所述第一数据库；

[0116] 子步骤 S12、若作为备库的第二数据库可接收访问，则将所述第二数据库切换为新的主库；

[0117] 子步骤 S13、打开所述新的主库接收访问，并重置在线日志记录的传递关系为，在第二数据库的第二磁盘和第三磁盘同步记录相同的在线日志记录并向第一系统传递。

[0118] 以 Oracle 数据库执行 switchover 为例，需要在主库和备库上执行的命令及操作如下：

[0119] 1) 在主库上发起切换：

[0120] ALTER DATABASE COMMIT TO SWITCHOVER TO PHYSICALSTANDBY；

[0121] 2) 关闭并重启原主库：

[0122] SQL ＞ SHUTDOWN IMMEDIATE；

[0123] SQL ＞ STARTUP MOUNT；

[0124] 3) 确认备库可切换：

[0125] SELECT SWITCHOVER_STATUS FROM V$DATABASE；

[0126] 4) 切换备库为新主库：

[0127] ALTER DATABASE COMMIT TO SWITCHOVER TO PRIMARY；

[0128] 5) 打开新主库接收访问：

[0129] ALTER DATABASE OPEN；

[0130] 6) 如果需要，可以重新配置日志传送关系为从新主库到新备库。

[0131] 步骤 64、判断所述第一数据主机是否可访问；若是，则执行步骤 65；若否，则执行步骤 66；

[0132] 一般而言，数据主机可以包括主机名，主机 IP，数据库名称等信息，在实际中，可以通过 ssh 主机名，验证数据主机是否可以登录访问。

[0133] 步骤 65、通过所述第一数据主机提取所述在线日志记录的信息，并依据所述在线日志记录的信息在第一数据库中定位所述在线日志记录，然后转步骤 67；

[0134] 在第一数据主机未发生故障的情况下，可以直接通过第一数据主机从操作系统层面定位当前未传递给第二系统的在线日志记录，如通过 Oracle 数据库提供的视图，确定缺少的在线日志记录，具体可以采用如下代码：

[0135] SELECT THREAD#, LOW_SEQUENCE#, HIGH_SEQUENCE#FROM V$ARCHIVE_GAP；

[0136] THREAD#LOW_SEQUENCE#HIGH_SEQUENCE#；

[0137] 1 90 92

[0138] 根据上述 sequence# 号码，可以通过 ftp 或 scp 的方式，从第一数据库的第一磁盘中定位到相应的在线日志记录。

[0139] 步骤 66、通过第二数据主机激活位于第一数据库中的第四磁盘，并在所述第四磁盘中定位所述在线日志记录，然后转步骤 67；

[0140] 步骤 67、根据所述定位的在线日志记录，对第二数据库进行相应的数据变更。

[0141] 需要说明的是，对于方法实施例，为了简单描述，故将其都表述为一系列的动作组合，但是本领域技术人员应该知悉，本申请并不受所描述的动作顺序的限制，因为依据本申请，某些步骤可以采用其他顺序或者同时进行。其次，本领域技术人员也应该知悉，说明书中所描述的实施例均属于优选实施例，所涉及的动作和模块并不一定是本申请所必须的。

[0142] 再者，上述各个方法实施例均采用递进的方式描述，每个实施例重点说明的都是与其他实施例的不同之处，各个实施例之间相同相似的部分互相参见即可。

[0143] 参照图 7，其示出了本申请的一种数据容灾处理的装置实施例的结构框图，本装置实施例基于如图 1 所示的硬件架构实现，主要可以包括以下模块：

[0144] 第一日志记录模块 71，用于在所述第一数据库中的第一磁盘和第四磁盘同步记录相同的在线日志记录；

[0145] 日志传送模块 72，用于在所述在线日志记录在第一数据库中记录完后向第二系统传送；

[0146] 日志恢复模块 73，用于在第二系统未收到所述在线日志记录时，调用以下子模块：

[0147] 数据库访问判断子模块 731，用于判断所述第一数据库是否可访问；若否，则触发数据主机访问判断子模块 733；

[0148] 主机访问判断子模块 733，用于判断所述第一数据主机是否可访问；若否，则触发激活定位子模块 735；

[0149] 激活定位子模块 735，用于通过第二数据主机激活位于第一数据库中的第四磁盘，并在所述第四磁盘中定位所述在线日志记录。

[0150] 在本申请的一种优选实施例中，所述日志恢复模块还可以包括：

[0151] 日志提取子模块 734，用于在所述第一数据库不可访问，但所述第一数据主机可以访问时，通过所述第一数据主机提取所述在线日志记录的信息，并依据所述在线日志记录的信息在第一数据库中定位所述在线日志记录。

[0152] 在具体实现中，所述在线日志记录根据第一数据库的数据变更在第一磁盘和第四磁盘中同步记录，在这种情况下，所述装置实施例还可以包括以下模块：

[0153] 数据更新模块 74，用于根据所述定位的在线日志记录，对第二数据库进行相应的数据变更。

[0154] 当所述第一系统为主系统，所述第一数据库为主库，所述第二系统为备系统，所述第二数据库为备库时，所述装置实施例还可以包括以下模块：

[0155] 主库切换模块，用于切换所述第二数据库为新的主库。

[0156] 在本申请的一种优选实施例中，所述装置实施例还可以包括以下模块：

[0157] 第一备库切换模块，用于在所述第一数据库和第一数据主机可访问时，切换所述第一数据库为新的备库；

[0158] 第一重置模块，用于打开所述新的主库接收访问，并重置在线日志记录的传递关系为，在第二数据库的第二磁盘和第三磁盘同步记录相同的在线日志记录并向第一系统传递。

[0159] 本申请实施例还可以应用于多机房的数据容灾部署方案中，在这种应用中，所述第二系统还可以采用通信延时非常短（不超过 1 毫秒）的通信链路与第三系统相连，所述第三系统中包括第三数据库、第三数据主机和第三存储设备，第二存储设备在第二数据库中分配第五磁盘，在第三数据库中分配第六磁盘；所述第三存储设备在第三数据库中分配第七磁盘，在第二数据库中分配第八磁盘，所述第七磁盘和第八磁盘被同步写入相同的在线日志记录；

[0160] 在这种情况下，作为本申请的另一种优选实施例，所述装置实施例还可以包括以下模块：

[0161] 第二备库切换模块，用于切换所述第三数据库为新的备库；

[0162] 第二重置模块，用于打开所述新的主库接收访问，并重置在线日志记录的传递关系为，在第二数据库的第五磁盘和第八磁盘同步记录相同的在线日志记录并向第三系统传递。

[0163] 在本申请的一种优选实施例中，所述日志恢复模块还可以包括：

[0164] 主备切换子模块 732，用于在所述第一数据库可以访问时，依次调用以下单元：

[0165] 切换发起单元，用于在作为主库的第一数据库上发起切换请求；

[0166] 数据库重启单元，用于依据所述请求关闭并重启所述第一数据库；

[0167] 主库调整单元，用于在作为备库的第二数据库可接收访问时，将所述第二数据库切换为新的主库；

[0168] 主库打开单元，用于打开所述新的主库接收访问；

[0169] 日志传递关系重置单元，用于重置在线日志记录的传递关系为，在第二数据库的第二磁盘和第三磁盘同步记录相同的在线日志记录并向第一系统传递。

[0170] 作为本申请实施例具体应用的一种示例，所述通信链路可以为光纤通信链路，所述第一数据库、第二数据库和第三数据库均为 Oracle 数据库。

[0171] 对于系统实施例而言，由于其与方法实施例基本相似，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

[0172] 本申请可用于众多通用或专用的计算系统环境或配置中。例如：个人计算机、服务器计算机、手持设备或便携式设备、平板型设备、多处理器系统、基于微处理器的系统、置顶盒、可编程的消费电子设备、网络 PC、小型计算机、大型计算机、包括以上任何系统或设备的分布式计算环境等等。

[0173] 本申请可以在由计算机执行的计算机可执行指令的一般上下文中描述，例如程序模块。一般地，程序模块包括执行特定任务或实现特定抽象数据类型的例程、程序、对象、组件、数据结构等等。也可以在分布式计算环境中实践本申请，在这些分布式计算环境中，由通过通信网络而被连接的远程处理设备来执行任务。在分布式计算环境中，程序模块可以位于包括存储设备在内的本地和远程计算机存储介质中。

[0174] 最后，还需要说明的是，在本文中，诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来，而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

[0175] 以上对本申请所提供的一种数据容灾处理的方法和一种数据容灾处理的装置进行了详细介绍，本文中应用了具体个例对本申请的原理及实施方式进行了阐述，以上实施例的说明只是用于帮助理解本申请的方法及其核心思想；同时，对于本领域的一般技术人员，依据本申请的思想，在具体实施方式及应用范围上均会有改变之处，综上所述，本说明书内容不应理解为对本申请的限制。

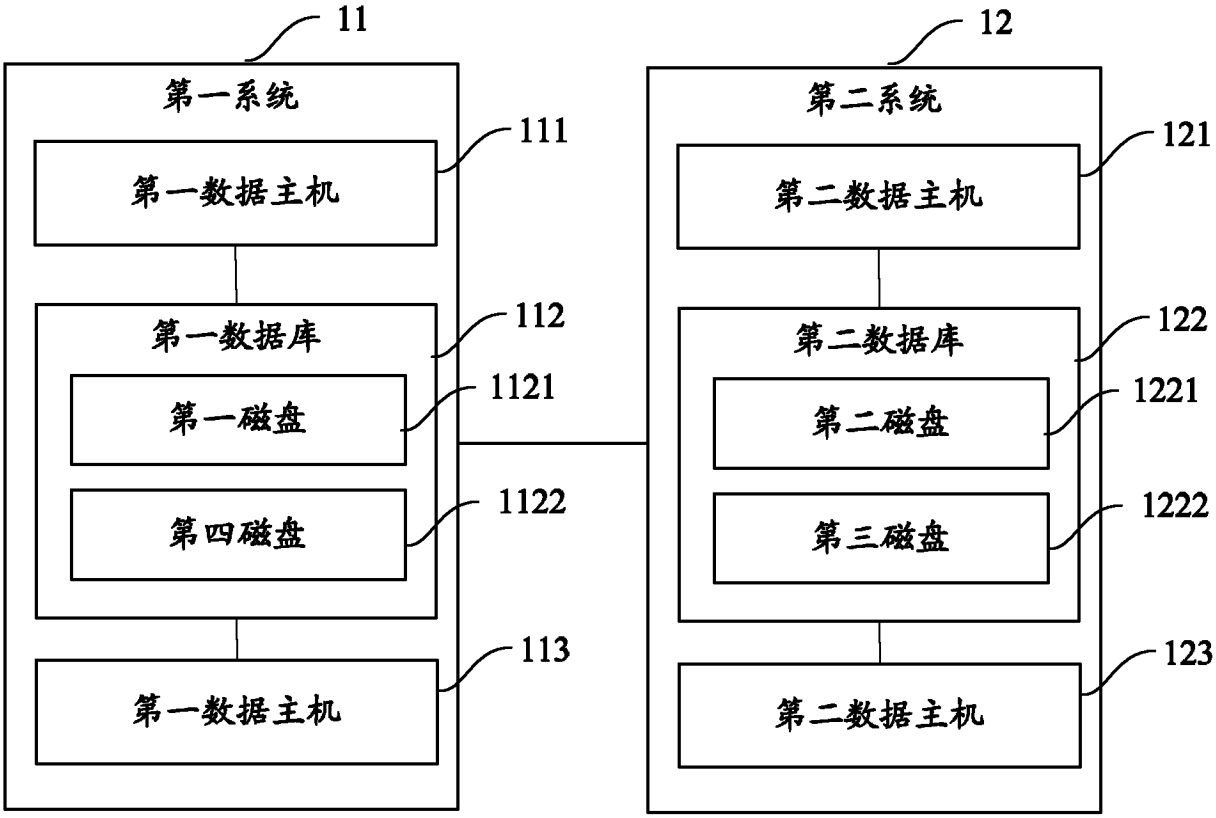

11
第一系统
111
第一数据主机
112
第一数据库
1121
第一磁盘
1122
第四磁盘
113
第一数据主机
12
第二系统
121
第二数据主机
122
第二数据库
1221
第二磁盘
1222
第三磁盘
123
第二数据主机

图 1

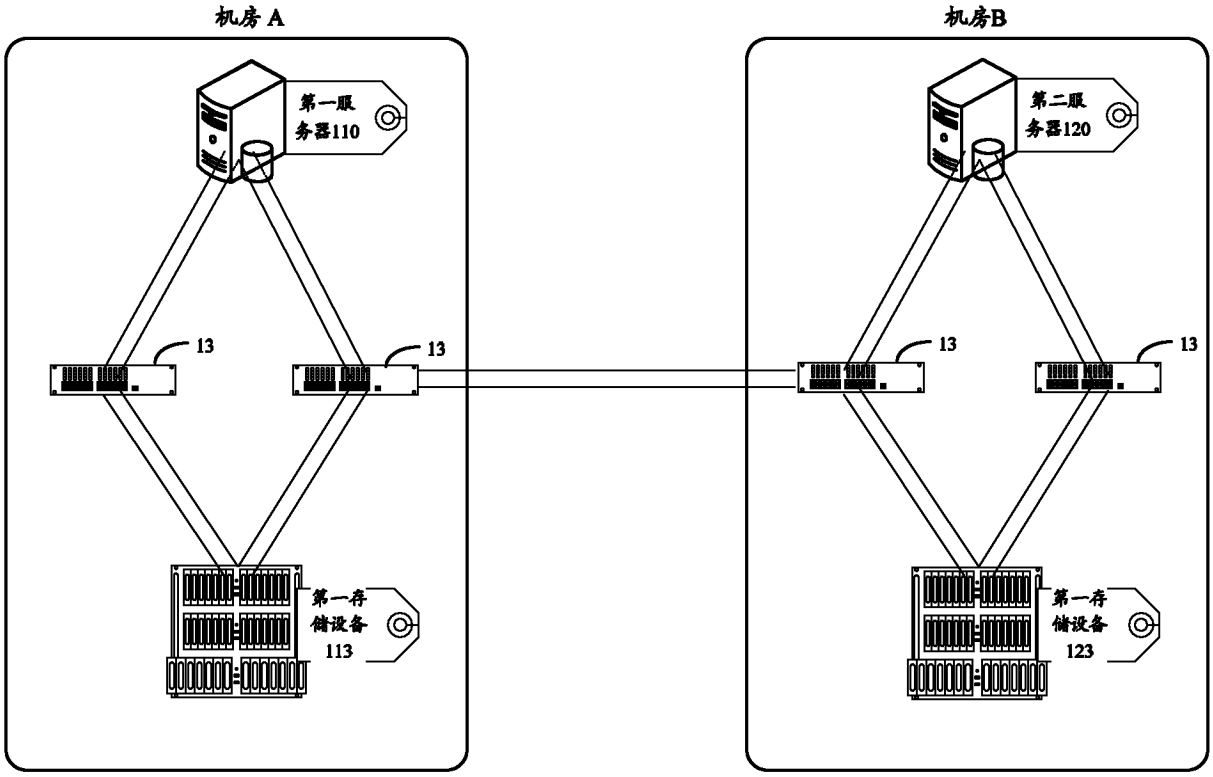


图 2

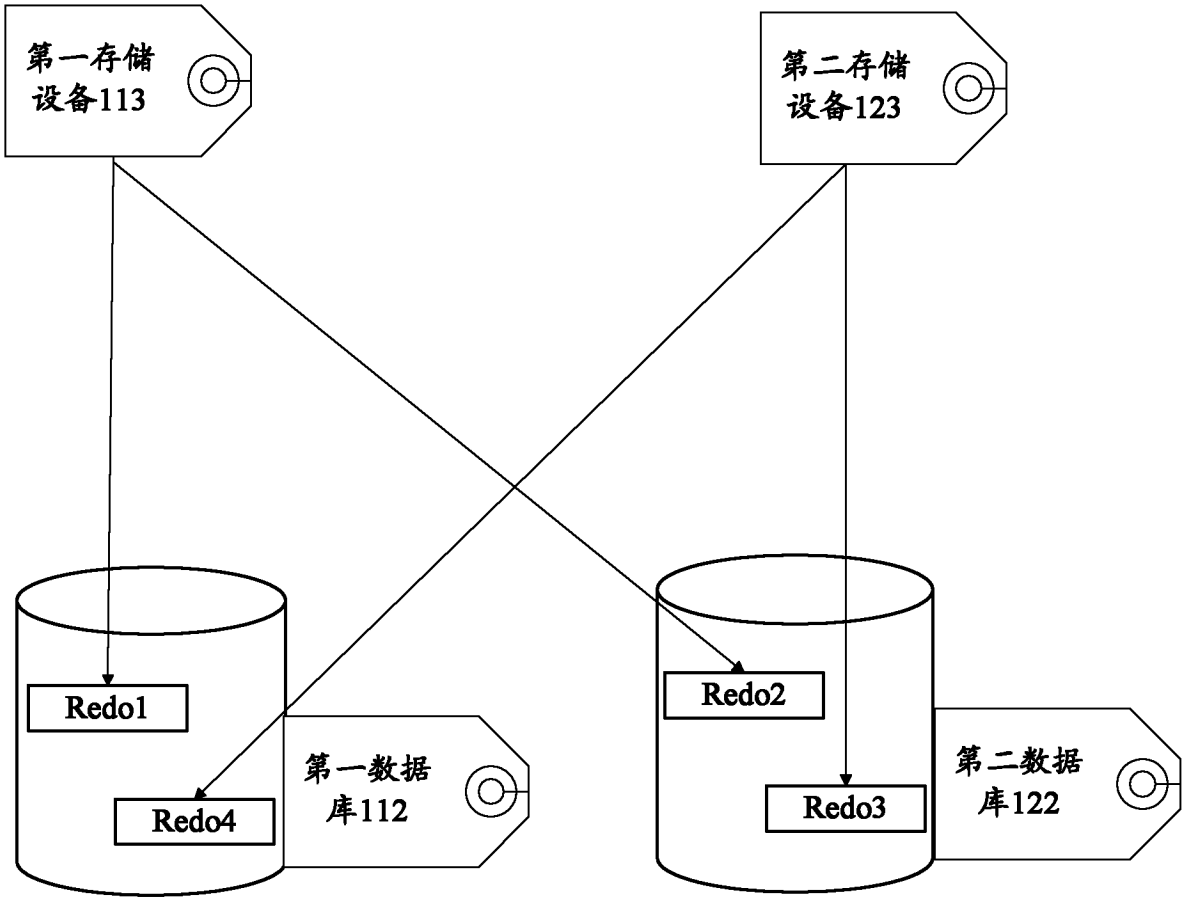

第一存储设备113
第二存储设备123
Redo1
Redo4
第一数据库112
Redo2
Redo3
第二数据库122

图 3

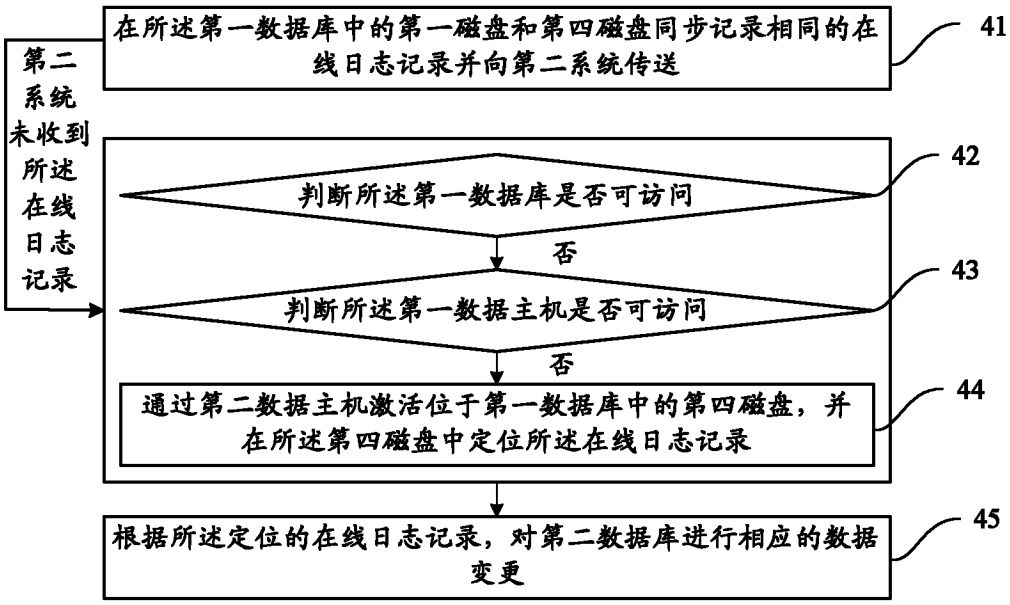

在所述第一数据库中的第一磁盘和第四磁盘同步记录相同的在线日志记录并向第二系统传送
41
第二系统未收到所述在线日志记录
判断所述第一数据库是否可访问
42
否
判断所述第一数据主机是否可访问
43
否
通过第二数据主机激活位于第一数据库中的第四磁盘，并在所述第四磁盘中定位所述在线日志记录
44
根据所述定位的在线日志记录，对第二数据库进行相应的数据变更
45

图 4

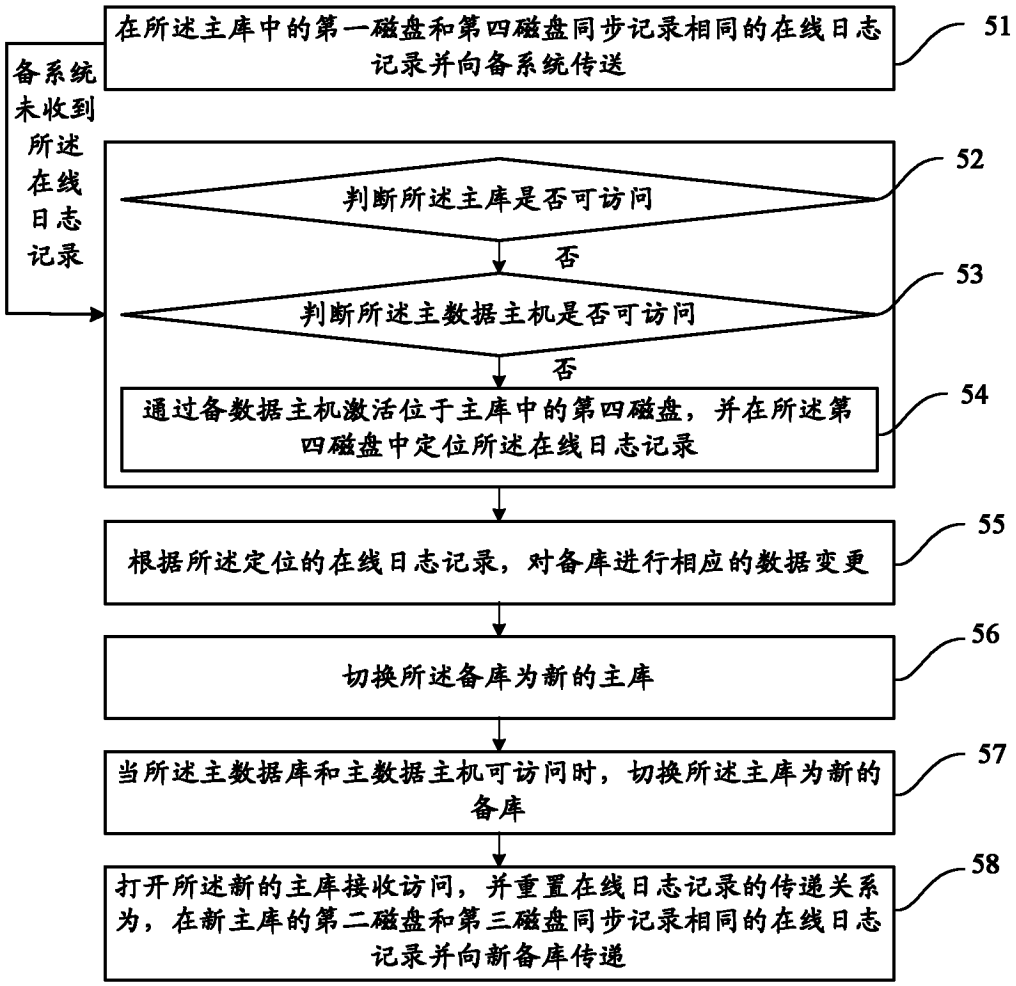

在所述主库中的第一磁盘和第四磁盘同步记录相同的在线日志记录并向备系统传送
51
备系统未收到所述在线日志记录
判断所述主库是否可访问
52
否
判断所述主数据主机是否可访问
53
否
通过备数据主机激活位于主库中的第四磁盘，并在所述第四磁盘中定位所述在线日志记录
54
根据所述定位的在线日志记录，对备库进行相应的数据变更
55
切换所述备库为新的主库
56
当所述主数据库和主数据主机可访问时，切换所述主库为新的备库
57
打开所述新的主库接收访问，并重置在线日志记录的传递关系为，在新主库的第二磁盘和第三磁盘同步记录相同的在线日志记录并向新备库传递
58

图 5

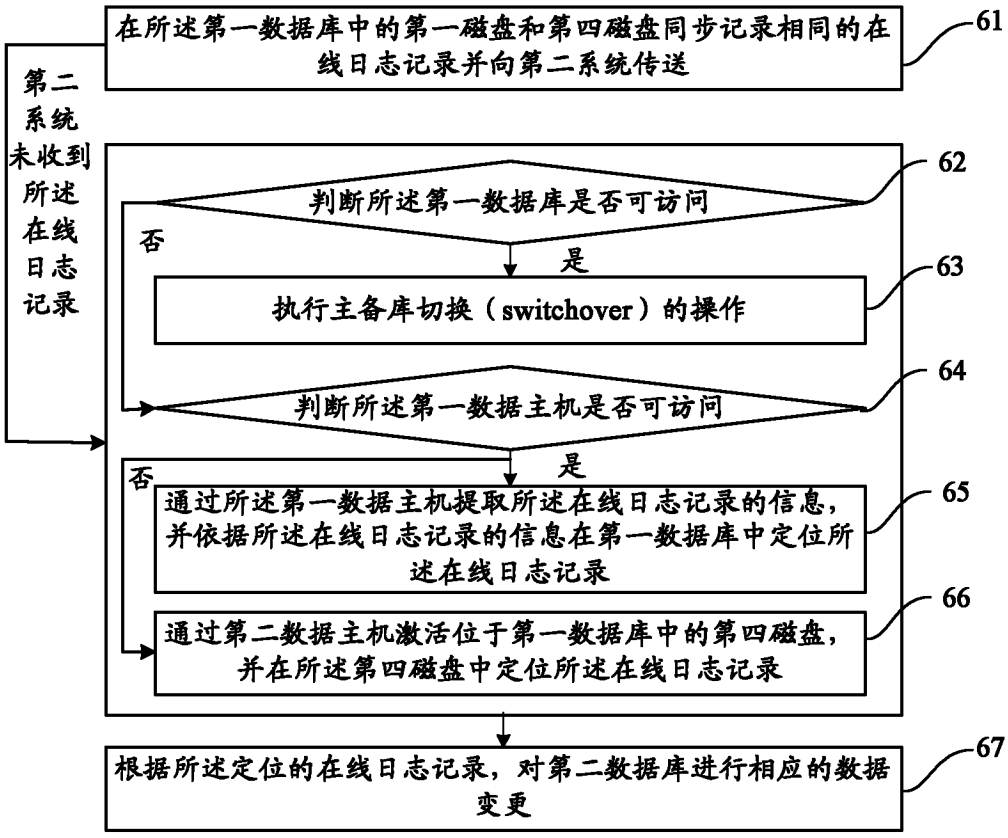

在所述第一数据库中的第一磁盘和第四磁盘同步记录相同的在线日志记录并向第二系统传送
61
第二系统未收到所述在线日志记录
判断所述第一数据库是否可访问
62
否
是
执行主备库切换（switchover）的操作
63
判断所述第一数据主机是否可访问
64
否
是
通过所述第一数据主机提取所述在线日志记录的信息，并依据所述在线日志记录的信息在第一数据库中定位所述在线日志记录
65
通过第二数据主机激活位于第一数据库中的第四磁盘，并在所述第四磁盘中定位所述在线日志记录
66
根据所述定位的在线日志记录，对第二数据库进行相应的数据变更
67

图 6

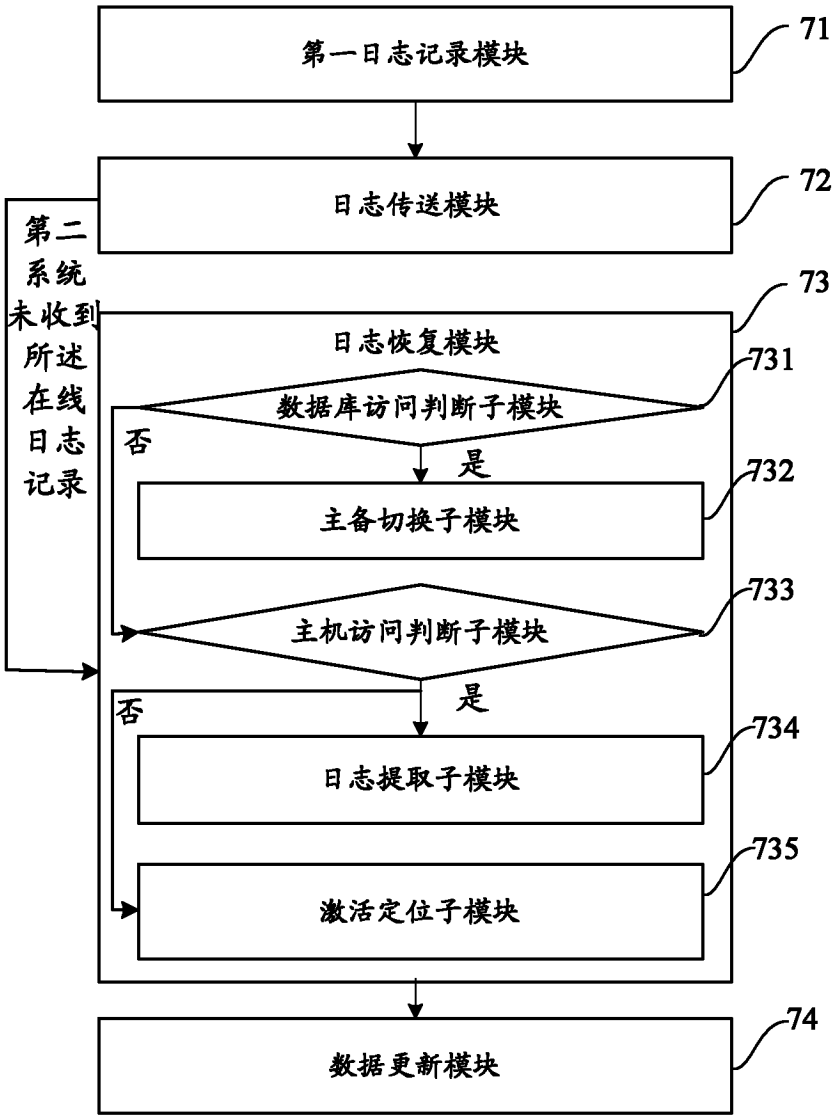

第一日志记录模块
71
日志传送模块
72
第二系统未收到所述在线日志记录
日志恢复模块
73
数据库访问判断子模块
731
否
是
主备切换子模块
732
主机访问判断子模块
733
否
是
日志提取子模块
734
激活定位子模块
735
数据更新模块
74

图 7