(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 102982085 A
(43)申请公布日 2013.03.20

(21)申请号 201210429724.9

(22)申请日 2012.10.31

(71)申请人 北京奇虎科技有限公司
地址 100088 北京市西城区新街口外大街28号D座112室(德胜园区)
申请人 奇智软件(北京)有限公司

(72)发明人 桂勇哲 陈超 代兵 朱超 王超

(74)专利代理机构 北京市德权律师事务所 11302
代理人 刘丽君

(51)Int.Cl.
G06F 17/30(2006.01)
G06F 11/14(2006.01)
H04L 29/08(2006.01)

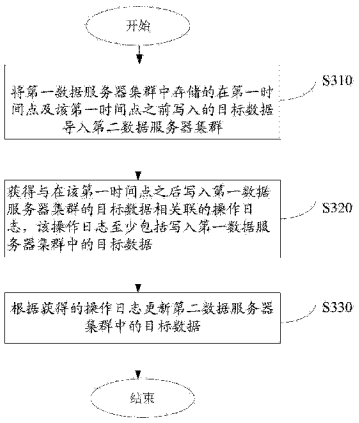
权利要求书 2 页 说明书 10 页 附图 3 页

(54)发明名称

数据迁移系统和方法

(57)摘要

本发明公开了一种数据迁移系统和方法，用于在数据服务器集群之间进行数据迁移，数据为与要迁移的业务相关的目标数据，该方法包括将第一数据服务器集群中存储的在第一时间点及该第一时间点之前写入的目标数据导入第二数据服务器集群；获得与在第一时间点之后写入第一数据服务器集群的目标数据相关联的操作日志；根据获得的操作日志更新第二数据服务器集群中的目标数据；检测第一数据服务器集群和第二数据服务器集群的目标数据是否已同步；以及在第一数据服务器集群和第二数据服务器集群的目标数据已同步之后，将连接数据服务器的入口地址由第一数据服务器集群的入口地址变更为第二数据服务器集群的入口地址。



CN 102982085 A

1. 一种数据迁移系统，用于在数据服务器集群之间进行数据迁移，所述系统至少包括所述第一数据服务器集群和所述第二数据服务器集群，以及迁移设备，其中，

所述数据为与要迁移的业务相关的目标数据；

所述迁移设备包括：

初始数据导入模块，被配置为将第一数据服务器集群中存储的在第一时间点及该第一时间点之前写入的目标数据导入第二数据服务器集群；

同步模块，被配置为获得与在所述第一时间点之后写入所述第一数据服务器集群的目标数据相关联的操作日志，所述操作日志至少包括写入所述第一数据服务器集群中的目标数据的内容；以及

更新模块，被配置为根据所述同步模块获得的操作日志更新所述第二数据服务器集群中的目标数据。

2. 根据权利要求 1 的数据迁移系统，所述初始数据导入模块包括：

第一初始数据导入子模块，被配置为将所述第一数据服务器集群中存储的所述第一时间点及该第一时间点之前写入的目标数据导入一存储介质；

第二初始数据导入子模块，被配置为将所述导入至存储介质中的目标数据导入所述第二数据服务器集群。

3. 根据权利要求 2 的数据迁移系统，其中，所述第一初始数据导入子模块被配置为通过 mongodump 将所述第一数据服务器集群中存储的所述第一时间点及该第一时间点之前写入的目标数据备份至一存储介质；以及

第二初始数据导入子模块被配置为通过 mongorestore 将备份至所述存储介质中的目标数据导入所述第二数据服务器集群。

4. 根据权利要求 1 至 3 中任一项的数据迁移系统，所述第一数据服务器集群中包括主数据服务器和若干从数据服务器，所述迁移设备还包括：

停用处理模块，被配置为在第一时间点之后停止所述第一数据服务器集群中的第一从数据服务器的写入操作；以及

所述初始数据导入模块被配置为将所述第一数据服务器集群的所述第一从数据服务器中存储的第一时间点及该第一时间点之前写入的目标数据导入所述第二数据服务器集群。

5. 根据权利要求 1 至 4 中任一项的数据迁移系统，其中所述迁移设备还包括：

同步检测模块，被配置为检测所述第一数据服务器集群和所述第二数据服务器集群的目标数据是否已同步；

地址更新模块，被配置为在所述同步检测模块检测到所述第一数据服务器集群和所述第二数据服务器集群的目标数据已同步之后，将连接数据服务器的入口地址由所述第一数据服务器集群的入口地址变更为所述第二数据服务器集群的入口地址。

6. 根据权利要求 1 至 5 中任一项的数据迁移系统，所述操作日志还包括如下信息中的一种或多种：

写入数据的时间戳；以及

当写入的数据是对原有数据的更新时，数据更新前的旧值。

7. 根据权利要求 1 至 6 中任一项的数据迁移系统，所述第一数据服务器集群为第一

MongoDB 集群，所述第二数据服务器集群为第二 MongoDB 集群，所述操作日志为 MongoDB 中的 oplog。

8. 一种数据迁移方法，用于在数据服务器集群之间进行数据迁移，所述数据为与要迁移的业务相关的目标数据，该方法包括：

将所述第一数据服务器集群中存储的在第一时间点及该第一时间点之前写入的目标数据导入所述第二数据服务器集群；

获得与在第一时间点之后写入所述第一数据服务器集群的目标数据相关联的操作日志，所述操作日志至少包括写入所述第一数据服务器集群中的目标数据的内容；

根据获得的操作日志更新所述第二数据服务器集群中的目标数据；

检测所述第一数据服务器集群和所述第二数据服务器集群的目标数据是否已同步；以及

在所述第一数据服务器集群和所述第二数据服务器集群的目标数据已同步之后，将连接数据服务器的入口地址由所述第一数据服务器集群的入口地址变更为所述第二数据服务器集群的入口地址。

9. 根据权利要求 8 的方法，所述将第一数据服务器集群中存储的在第一时间点及该第一时间点之前写入的目标数据导入所述第二数据服务器集群的步骤包括：

将所述第一数据服务器集群中存储的第一时间点及该第一时间点之前写入的目标数据导入一存储介质；以及

将所述导入至存储介质中的目标数据导入第二数据服务器集群。

10. 根据权利要求 9 的方法，其中，

通过 mongodump 将所述第一数据服务器集群中存储的第一时间点及该第一时间点之前写入的目标数据备份至一存储介质；以及

通过 mongorestore 将备份至所述存储介质中的目标数据导入所述第二数据服务器集群。

11. 根据权利要求 8 至 10 中任一项的方法，所述第一数据服务器集群中包括主数据服务器和若干从数据服务器，所述方法还包括：在所述第一时间点之后停止所述第一数据服务器集群中的第一从数据服务器的写入操作；以及

所述将所述第一数据服务器集群中存储的在第一时间点及该第一时间点之前写入的目标数据导入所述第二数据服务器集群的步骤包括：

将所述第一数据服务器集群中该第一从数据服务器中存储的第一时间点及该第一时间点之前写入的目标数据导入所述第二数据服务器集群。

12. 根据权利要求 8 至 11 中任一项的方法，所述操作日志还包括如下信息中的一种或多种：

写入数据的时间戳；以及

当写入的数据是对原有数据的更新时，数据更新前的旧值。

13. 根据权利要求 8 至 12 中任一项的方法，所述第一数据服务器集群为第一 MongoDB 集群，所述第二数据服务器集群为第二 MongoDB 集群，所述操作日志为 MongoDB 中的 oplog。

数据迁移系统和方法

技术领域

[0001] 本发明涉及数据存储技术领域，具体涉及一种数据迁移系统和方法。

背景技术

[0002] MongoDB（Data Base，数据库）是一个介于关系数据库和非关系数据库之间的产品，是非关系数据库当中功能最丰富，最像关系数据库的。他支持的数据结构非常松散，因此可以存储比较复杂的数据类型。由于MongoDB本身性能较好，因此在业务发展的早期，开发人员会将多个小的业务部署在一个少量服务器构成的小的Mongodb集群上。当业务逐渐发展，访问量越来越大，比如原有的某个或某几个业务的访问量增长迅速，形成大规模的业务，进而导致少量服务器构成的小集群已经无法满足业务需求了，此时就需要考虑如何增加系统容量，来解决性能问题。

[0003] 目前的第一种解决方案是，基于Mongodb本身支持的动态扩展，可以简单的通过增加数据服务器来提高性能，因此可以通过直接向目前由少量数据服务器构成的Mongodb集群中添加Mongod服务器，来解决性能问题。通过增加新的Mongod服务器，变为更多台数据服务器的集群，mongodb的性能会基本线性提升。但是，这种解决方案也会带来副作用。因为现有的数据服务器集群是为多个业务服务的，而且多个业务中可能既有小规模的业务，也有快速增长起来的大规模的业务，进而在对Mongod服务器的访问过程中，大规模的业务可能会长期占据对Mongod服务器的访问资源，而小规模的业务势必争抢不过大规模的业务，最终导致这些小规模的业务受到快速增长的大规模的业务的影响。

[0004] 为了避免上述第一种解决方案的副作用，逐渐出现了第二种解决方案。即考虑将业务规模扩大了的业务，从原有的mongodb集群迁移到一个全新的mongodb集群上。对于较大规模的业务，使用单独的新集群，而不再与其他业务共用Mongod服务器。这样，原有的多个规模较小的业务还是在原有的数据服务器集群，而快速发展成较大规模的业务单独使用全新的数据服务器集群，于是，较大规模的业务就不会再跟较小规模的业务抢占服务器的访问资源了。

[0005] 在采取上述方案的迁移过程中，首先由该业务停掉所有对mongod服务器的写入操作，然后将现有的mongodb数据库信息备份出来，导入到新的数据服务器集群中。最后，在将业务的数据完全切换到新的数据服务器集群后，再开启业务对mongod服务器的写入操作。由于切换过程中需要停掉业务的所有写入服务，而且当数据量较大的时候，停机备份迁移的过程可能会需要几个小时，因此在停机备份迁移的这段时间内非常影响相关业务的正常运行，影响为用户提供正常的服务。同理，在其他非mongodb的数据服务器集群的应用环境下，也同样存在类似问题。

发明内容

[0006] 鉴于上述问题，提出了本发明以便提供一种克服上述问题或者至少部分地解决上述问题的数据迁移系统和方法。

[0007] 依据本发明的一个方面，提供了一种数据迁移系统，用于在数据服务器集群之间进行数据迁移，该系统至少包括第一数据服务器集群和第二数据服务器集群以及迁移设备，数据为与要迁移的业务相关的目标数据，所述迁移设备包括：初始数据导入模块，被配置为将第一数据服务器集群中存储的在第一时间点及该第一时间点之前写入的目标数据导入第二数据服务器集群；同步模块，被配置为获得与在第一时间点之后写入第一数据服务器集群的目标数据相关联的操作日志，操作日志至少包括写入第一数据服务器集群中的目标数据的内容；更新模块，被配置为根据同步模块获得的操作日志更新第二数据服务器集群中的目标数据。

[0008] 可选的，初始数据导入模块包括：第一初始数据导入子模块，被配置为将第一数据服务器集群中存储的第一时间点及该第一时间点之前写入的目标数据导入一存储介质；第二初始数据导入子模块，被配置为将导入至存储介质中的目标数据导入第二数据服务器集群。

[0009] 可选的，其中第一初始数据导入子模块被配置为通过 mongodump 将第一数据服务器集群中存储的第一时间点及该第一时间点之前写入的目标数据备份至一存储介质；以及第二初始数据导入子模块被配置为通过 mongorestore 将备份至存储介质中的目标数据导入第二数据服务器集群。

[0010] 可选的，第一数据服务器集群中包括主数据服务器和若干从数据服务器，迁移设备还包括：停用处理模块，被配置为在第一时间点之后停止第一数据服务器集群中的第一从数据服务器的写入操作；以及初始数据导入模块，被配置为将第一数据服务器集群的第一从数据服务器中存储的第一时间点及该第一时间点之前写入的目标数据导入第二数据服务器集群。

[0011] 可选的，还包括：同步检测模块，被配置为检测第一数据服务器集群和第二数据服务器集群的目标数据是否已同步；地址更新模块，被配置为在同步检测模块检测到第一数据服务器集群和第二数据服务器集群的目标数据已同步之后，将连接数据服务器的入口地址由第一数据服务器集群的入口地址变更为第二数据服务器集群的入口地址。

[0012] 可选的，操作日志还包括如下信息中的一种或多种：写入数据的时间戳；当写入的数据是对原有数据的更新时，数据更新前的旧值。

[0013] 可选的，第一数据服务器集群为第一 MongoDB 集群，第二数据服务器集群为第二 MongoDB 集群，操作日志为 MongoDB 中的 oplog。

[0014] 根据本发明的另一个实施例，提供了一种数据迁移方法，用于在数据服务器集群之间进行数据迁移，数据为与要迁移的业务相关的目标数据，包括：将第一数据服务器集群中存储的在第一时间点及该第一时间点之前写入的目标数据导入第二数据服务器集群；获得与在第一时间点之后写入第一数据服务器集群的目标数据相关联的操作日志，操作日志至少包括写入第一数据服务器集群中的目标数据的内容；根据获得的操作日志更新第二数据服务器集群中的目标数据；检测所述第一数据服务器集群和所述第二数据服务器集群的目标数据是否已同步；以及，在所述第一数据服务器集群和所述第二数据服务器集群的目标数据已同步之后，将连接数据服务器的入口地址由所述第一数据服务器集群的入口地址变更为所述第二数据服务器集群的入口地址。

[0015] 根据本发明的数据迁移系统和方法，一方面通过备份的方式将某个时间点之前的

目标数据直接备份至新的数据服务器集群，另一方面利用操作日志将第一时间点之后写入旧数据服务器集群的目标数据同步写入到第二数据服务器集群，使得使新旧数据服务器集群基本实现了目标数据的同步，进而后续被迁移业务可以直接连接新数据服务器集群进行数据的写入和读取即可，在此过程中不需要停止要迁移的业务，由此解决了现有必须停机备份才能实现业务数据迁移的问题，取得了在不影响被迁移业务对外正常服务的情况下能够完成业务数据迁移的有益效果。

[0016] 上述说明仅是本发明技术方案的概述，为了能够更清楚了解本发明的技术手段，而可依照说明书的内容予以实施，并且为了让本发明的上述和其它目的、特征和优点能够更明显易懂，以下特举本发明的具体实施方式。

附图说明

[0017] 通过阅读下文优选实施方式的详细描述，各种其他的优点和益处对于本领域普通技术人员将变得清楚明了。附图仅用于示出优选实施方式的目的，而并不认为是对本发明的限制。而且在整个附图中，用相同的参考符号表示相同的部件。在附图中：

[0018] 图 1 示出了根据本发明一个实施例的用于数据服务器集群之间进行数据迁移的第一系统示意图；

[0019] 图 2 示出了根据本发明一个实施例的用于数据服务器集群之间进行数据迁移的第二系统示意图；以及

[0020] 图 3 示出了根据本发明一个实施例的数据迁移方法的流程图。

具体实施方式

[0021] 下面将参照附图更详细地描述本公开的示例性实施例。虽然附图中显示了本公开的示例性实施例，然而应当理解，可以以各种形式实现本公开而不应被这里阐述的实施例所限制。相反，提供这些实施例是为了能够更透彻地理解本公开，并且能够将本公开的范围完整的传达给本领域的技术人员。

[0022] 请参阅图 1，其为根据本发明一个实施例的用于数据服务器集群之间进行数据迁移的第一系统示意图。该系统包括用于数据服务器集群之间进行数据迁移的迁移设备 100、第一数据服务器集群 200 和第二数据服务器集群 300。迁移设备 100 包括初始数据导入模块 102、同步模块 104 和更新模块 106。第一数据服务器集群 200 包括多个数据服务器，在图中只示意性的给出了其中的第一数据服务器 202 和第二数据服务器 204，实际应用中还可以根据需要包含更多的数据服务器，本发明对此并没有限制。类似的，第二数据服务器集群 300 也包括多个数据服务器，在图中示意性给出了第三数据服务器 302、第四数据服务器 304 以及第五数据服务器 306。

[0023] 通常而言，每个数据服务器集群中的各数据服务器之间互为数据备份，一般多个数据服务器中会有一个是主数据服务器，其余为从数据服务器，应用服务器 400 多数情况下是直接向主数据服务器中写入数据，其余从数据服务器一般不直接接受应用服务器 400 的数据写入，而是从主数据服务器备份数据。下面结合各部件对数据的处理过程，以及各部件之间的关联关系进行详细说明。为后续叙述方便，将需要从第一数据服务器集群 200 迁移至第二数据服务器集群 300 的业务数据，称为与要迁移的业务相关的目标数据。

[0024] 在一个实施例中，在目标数据被迁移之前，应用服务器 400 都是将要迁移业务的数据，即目标数据，随时写入第一数据服务器集群 200。比如第一数据服务器集群 200 中的第一数据服务器 202 和第二数据服务器 204 中都存储有已经写入的目标数据。进而，在准备进行数据迁移时，首先选择一个已经是过去时的时间点，简称为第一时间点，然后由初始数据导入模块 102 将应用服务器 400 在该第一时间点和第一时间之前写入到第一数据服务器集群 200 中的目标数据，导入到第二数据服务器 300 中。后续为了更方便的说明，以第一数据服务器 200 中的第二数据服务器为主数据服务器、第一数据服务器为从数据服务器为例进行说明。

[0025] 具体而言，首先，初始数据导出模块 102 从第一数据服务器集群 200 中的多个数据服务器中选择一个数据服务器进行数据导出，通常，可以选择一个从数据服务器进行数据导出，以便在数据导出的过程中，不影响应用服务器 400 后续向主数据服务器的写入，例如，选择作为从数据服务器的第一数据服务器 202 进行数据导出。此外，由于应用服务器 400 不直接向作为从数据服务器的第一数据服务器 202 写入数据，而是在应用服务器 400 向作为主数据服务器的第二数据服务器 204 写入数据之后，第一数据服务器 202 再从第二数据服务器 204 获得数据，而且由于从第一数据服务器 202 导出的第一时间点和第一时间点之前写入第一数据服务器集群 200 的目标数据的数据量可能比较大，因此，为了更好的避免在第一数据服务器 202 中同时进行数据导出与数据写入可能出现的各种意外情况，迁移设备 100 可以包括停用处理模块，用于在第一时间点之后可以先停止第一数据服务器 202 的写入操作，然后才开始进行数据的导出操作。待初始数据导入模块 102 从第一数据服务器 202 成功导出第一时间点和第一时间之前写入第一数据服务器集群 200 的数据之后，可以再恢复第一数据服务器 202 的写入操作。应当注意的是，上面有关在导出目标数据之前先停止第一数据服务器 202 的写入操作的示例仅仅是可选的，完全可以在导出目标数据的同时不停止从第二数据服务器 204 同步数据到第一数据服务器 202。

[0026] 然后，初始数据导入模块 102 将第一数据服务器 202 中存储的在第一时间点和第一时间点之前写入的目标数据导入到第二数据服务器集群 300。例如，初始数据导入模块 102 具体包括第一初始数据导入子模块和第二初始数据导入子模块，第一初始数据导入子模块先从第一数据服务器 202 中取出在第一时间点和第一时间点之前写入的目标数据，并将目标数据导入一存储介质，如磁盘中(例如，可以以文件形式)。随后第二初始数据导入子模块再将已导入该存储介质的目标数据导入第二数据服务器集群 300 中的各数据服务器(302-306)。在初始数据导入模块 102 将目标数据导入第二数据服务器集群 300 的过程中，可以先将目标数据导入第二数据服务器集群 300 中的主数据服务器，然后其他从数据服务器再从主数据服务器获取目标数据，这样第二数据服务器集群 300 中的主数据服务器和从数据服务器，即第三数据服务器 302、第四数据服务器 304 和第五数据服务器 306 都成功获得了第一时间点及第一时间点之前写入第一数据服务器集群 100 的目标数据。

[0027] 在初始数据导入模块 102 将第一时间点和第一时间点之前写入第一数据服务器集群 200 的目标数据成功导入第二数据服务器集群 300 的过程中以及成功导入后，应用服务器 400 并没有停止向第一数据服务器集群 200 中的作为主数据服务器的第二数据服务器 204 写入数据，因此在第一时间点之后，仍然有目标数据被写入第一数据服务器集群 200 中。具体而言，第二数据服务器 204 先被应用服务器 400 写入第一时间点之后的目标数据，

随后第一数据服务器202可以再从第二数据服务器204同步获得第一时间点之后写入的目标数据。

[0028] 而且，在应用服务器400向第二数据服务器204每写入一条目标数据时，基于数据服务器集群自身的特性，数据服务器集群都会同时产生一条与该目标数据相关联的操作日志(oplog,operation log)，在该操作日志中记录着每次写入的具体数据的内容。而且，一般操作日志中还记录着每次写入的数据的时间信息，即时间戳，如果写入的数据是对以前数据的更新，则不但会记录数据更新后的新值，而且还会记录数据更新前的旧值。换而言之，根据操作日志，就可以知道应用服务器400在哪个时间、具体写入了哪些内容的数据。

[0029] 进而，迁移设备100中的同步模块104可以获得与在第一时间点之后写入第一数据服务器集群200的目标数据相关联的操作日志。具体而言，由于应用服务器400在向第一数据服务器集群200写入目标数据的过程中，多是逐条数据进行写入的，相应的，也会逐条产生操作日志，而且操作日志中包括写入数据的时间戳。因此可选的，同步模块104可以实时去获取与在第一时间点之后写入第一数据服务器集群200的目标数据相关联的操作日志。当然，也可以不必实时去获取操作日志，而是每间隔一定时间去获取一次操作日志，不过，为了尽快实现两个数据服务器集群的尽快同步，同步模块104获取操作日志的时间间隔尽量短。

[0030] 然后，同步模块104将获得的与在第一时间点之后写入第一数据服务器集群200的目标数据相关联的操作日志提供给更新模块106。更新模块106获得与在第一时间点之后写入第一数据服务器集群200的数据相关联的操作日志之后，就可以根据所获得的操作日志更新第二数据服务器集群300中的目标数据。之前已经提过，操作日志中至少包含应用服务器400在第一时间点之后每次写入第一数据服务器集群200的具体数据的内容，因此更新模块106可以根据每条操作日志更新第二数据服务器集群300中相应的目标数据，使得第二数据服务器集群300也存储了第一时间点之后写入第一数据服务器集群200的目标数据。

[0031] 由于单条操作日志涉及到的数据更新内容很少，因此更新模块106可以很及时的根据一条操作日志更新完毕第二数据服务器集群300中相关联的目标数据，进而，基本可以达到应用服务器400在第一时间点之后每向第一数据服务器集群200写入一条目标数据或几条目标数据，更新模块106就可以相应的将同样的目标数据写入第二数据服务器集群300，所以，经过很短暂的时间之后，第二数据服务器300就可以达到与第一数据服务器200中的目标数据一致、同步的目的。

[0032] 至此，在第一数据服务器集群200中存储的与要迁移业务相关的目标数据已经全部迁移到第二数据服务器集群300。

[0033] 在另一个实施例中，为了目标数据迁移成功后，后续应用服务器400可以直接向第二数据服务器集群300写入与要迁移业务相关的数据，在迁移设备100中还可以包括同步检测模块和地址更新模块。具体而言，在更新模块106根据同步模块104获得的与在第一时间点之后写入第一数据服务器集群200的目标数据相关联的操作日志更新第二数据服务器集群300中的目标数据之后，同步检测模块可以对第一数据服务器集群200和第二数据服务器集群300中的目标数据进行对比，从而检测出两个数据服务器集群中的目标数据是否已经同步成功。检测是否同步成功，除了可以参考两个数据服务器集群中的目标数

据的自身内容，还可以参考与目标数据相关联的操作日志，因为操作日志中一般有写入目标数据的时间戳和目标数据更新前、后的值等辅助信息，因此，同步检测模块也可以参考这些辅助信息，更快速、准确的检测出两个数据服务器集群是否已经成功同步。

[0034] 在同步检测模块判定第一数据服务器集群 200 和第二数据服务器集群 300 已经同步成功之后，可以通知地址更新模块两个数据服务器集群已经同步成功，进而地址更新模块就可以将应用服务器 400 连接数据服务器的入口地址由第一数据服务器集群 200 的入口地址变更为第二数据服务器集群 300 的入口地址。后续应用服务器 400 如果需要再向数据服务器集群写入与要迁移业务相关的数据或者读取该业务相关的数据，就会因为连接数据服务器的入口地址已变更为第二数据服务器集群的入口地址，直接向第二数据服务器集群 300 进行数据的写入和读取。此时，需要被迁移的业务已经成功从第一数据服务器集群 200 迁移至第二数据服务器集群 300。而后，第一数据服务器集群 200 中与被迁移业务相关的数据就可以删除了。

[0035] 应该注意的是，迁移设备 100 在具体实现过程中，可以独立于第一数据服务器集群 200 和第二数据服务器集群 300 而单独实现，也可以置于某个数据服务器集群，比如第二数据服务器集群 300 的数据管理服务器中予以实现。该数据管理服务器可以是目前很多数据服务器集群中都具有的、担当该集群管理角色的服务器，比如 MongoDB 集群中的 mangos。

[0036] 在一个实施例中，上面第一系统的第一数据服务器集群 200 是第一 MongoDB 集群 500，第二数据服务器集群 300 是第二 MongoDB 集群 600，其中的迁移设备 100 置于第二 MongoDB 集群的 mangos 中予以实现。具体请参阅图 2，其为根据本发明一个实施例的用于数据服务器集群之间进行数据迁移的第二系统示意图。该第二系统可以理解为是上面第一系统应用于 MongoDB 这种类型的数据服务器集群的具体应用实例，因此第二系统中的第一 MongoDB 集群 500 与第一系统中的第一数据服务器集群 200 雷同，类似的，第二系统中的第二 MongoDB 集群 600 与第一系统中的第二数据服务器集群 300 雷同，第二系统中的初始数据导入模块 702 与第一系统中的初始数据导入模块 102 雷同，第二系统中的同步模块 704 与第一系统中的同步模块 104 雷同，以及第二系统中的更新模块 706 与第一系统中的更新模块 106 雷同，因此对第二系统中的各部件的具体实现基本不再赘述，可以参考第一系统中相关联部件的具体实现方式。仅针对涉及到具体 MongoDB 集群特性的个别部件予以说明。

[0037] 比如，初始数据导入模块 702 中的第一初始数据导入子模块可以先通过调用 mongo 自带的一种备份工具 mongodump，将第一 mongod 集群 500 中存储的第一时间点及第一时间点之前写入的目标数据备份到磁盘上，生成一个数据文件。然后初始数据导入模块 702 中的第二初始数据子模块再使用 mongo 自带的一种恢复工具 mongorestore 将该数据文件导入到第二 mongod 集群 600 中。同步模块 704 获得的操作日志具体是 MongoDB 中的 oplog，例如一条具体 oplog 实例内容如下所示：

[0038] {" ts" :{" t" :1339660240000," i" :8},

[0039] " h" :NumberLong(" -7936072258265513667")," op" :" i" ," ns" :" test.method" ,

[0040] " o" :{" _id" :" testid" ," v" :”test”}}

[0041] 其中，" ts" 记录的是该操作的时间戳；" op" 记录的是该操作的类型，比如类

型“i”表明是插入操作；" h" 记录的是本条 oplog 的哈希值，" ns" 记录的是该操作的命名空间；“o”记录的是文件内容，即具体写入的数据的内容。

[0042] 从上述一条 oplog 实例可以看出，在 MongoDB 集群的 oplog 中不但包含写入的目标数据的内容，而且还包括时间戳等其他辅助信息，因此更新模块 706 可以根据同步模块 704 获得与在第一时间点之后写入第一 MongoDB 集群 500 的目标数据相关联的操作日志，更新第二 MongoDB 集群 600 中的目标数据的内容。

[0043] 本领域技术人员可以理解，在其他非 MongoDB 集群的分布式数据存储系统，例如 Cassandra 等其他分布式数据存储系统中，也存在数据服务器集群之间需要进行数据迁移的类似问题，而且也具有与 MongoDB 集群的操作日志类似的操作日志，因此本发明的技术方案不仅适用于 MongoDB 集群之间的数据迁移，也同样适用于其他类型的数据服务器集群之间的数据迁移。

[0044] 请参阅图 3，其为根据本发明一个实施例的数据迁移方法示意图，该数据为与要迁移的业务相关的目标数据，进行数据迁移的两个数据服务器集群，其例如可以为上面图 1 描述的第一数据服务器集群 200 和第二数据服务器集群 300，也可以为上面图 2 描述的第一 MongoDB 集群 500 和第二 MongoDB 集群 600。

[0045] 该数据迁移方法始于步骤 S310，在步骤 S310 中，将第一数据服务器集群中存储的在第一时间点及该第一时间点之前写入的目标数据导入第二数据服务器集群。具体而言，在准备进行数据迁移时，首先选择一个已经是过去时的第一时间点，然后将第一数据服务器集群中存储的第一时间点和第一时间点之前被写入的目标数据导入到第二数据服务器集群。例如，可以先将第一数据服务器集群中的目标数据备份到类似磁盘的存储介质，比如在 MongoDB 集群中可以使用 mongodump 这一备份工具进行数据导入；然后再将存储介质中的目标数据导入第二数据服务器集群，比如在 MongoDB 集群中可以使用 mongorestore 这一恢复工具。可选的，如果第一数据服务器集群中存在主数据服务器和从数据服务器，那么优选在停掉一个从数据服务器的写入操作之后，再从该从数据服务器中导出目标数据。本步骤可以通过前述图 1 中的初始数据导入模块 102 或图 2 中的初始数据导入模块 702 执行，相关的技术实现可以参考前述初始数据导入模块在各实施例中的相关描述，此处不再赘述。

[0046] 在上面步骤 S310 中从第一数据服务器集群备份至第二数据服务器集群的目标数据，是在第一时间点及第一时间点之前写入到第一数据服务器集群的，因此后续需要迁移的就是在第一时间之后写入第一数据器集群的目标数据。进而，在步骤 S320 中，首选获得与在该第一时间点之后写入第一数据服务器集群的目标数据相关联的操作日志，操作日志是记录每一次数据写入这一操作具体内容的信息记录，其中包括每一次写入第一数据服务器集群的目标数据的内容，通常还包括写入的时间戳，如果该操作是对以前写入过的数据内容的更新，那么在操作日志中不但记录更新后的新值，还记录更新前的旧值。由此可见，根据操作日志就可以知道每次写入数据的具体内容。在获得操作日志的过程中，由于数据是逐条被写入的，进而操作日志也是逐条产生的，因此可以实时获取操作日志，即每写入一条或几条操作日志，就去获取一次；也可以定期获取操作日志。步骤 S320 可以通过前述图 1 中的同步模块 104 或图 2 中的同步模块 704 执行，相关的技术实现可以参考前述同步模块在各实施例中的相关描述，此处不再赘述。

[0047] 然后，在步骤 S330 中，根据步骤 S320 获得的在第一时间点之后写入第一数据服务

器集群的操作日志，更新第二数据服务器集群中的目标数据。具体而言，可以根据操作日志中记录的每次写入数据的具体内容，将相应的数据也写入到第二数据服务器集群，从而使第二数据服务器集群中也成功存储了在第一时间点之后写入第一数据服务器集群的目标数据。由于每条操作日志的内容较少，因此步骤 S330 执行完毕更新操作的速度也就很快，近乎可以达到每写入第一数据服务器集群一条目标数据，随之也被写入第二数据服务器集群中。至此，第一数据服务器集群中的目标数据，已经全部被迁移到了第二数据服务器集群，并且达到了两个数据服务器集群中的目标数据同步的目的。步骤 S330 可以通过前述图 1 中的更新模块 106 或图 2 中的更新模块 706 执行，相关的技术实现可以参考前述更新模块在各实施例中的相关描述，此处不再赘述。

[0048] 此后，还可以检测两个数据服务器集群中存储的目标数据是否一致，以及借助操作日志中的时间戳和更新前后的数值内容等辅助信息，一并来判断两个数据服务器集群是否已经同步成功。该步骤可以通过前文的迁移设备 100 中描述的同步检测模块执行。如果检测后确定已经同步成功，那么就可以将连接数据服务器的入口地址由第一数据服务器集群的入口地址变更为第二数据服务器集群的入口地址。该步骤可以通过前文的迁移设备中的地址更新模块执行。此后，无论应用服务器是需要对目标数据进行写入，还是对目标数据进行读取，都会由第二数据服务器集群执行相应的操作。在观察一段时间两个数据服务器的运行完全稳定之后，就可以将第一数据服务器集群中的目标数据删除了。

[0049] 应当注意的是，上述方法中的各步骤的顺序是可以调整的，例如步骤 S320 可以不必等步骤 S310 完全执行完之后才开始，可能出现这两个步骤同时执行或部分时段同时执行的情况。由于操作日志是实时产生的，因此步骤 S320 和步骤 S330 也可以重复执行，即每当一条或几条操作日志新产生后，就执行一遍步骤 S320 和步骤 S330，以便及时获得最新的操作日志并根据操作日志及时更新第二数据服务器集群中的目标数据，从而可以使两个数据服务器尽快达到目标数据的同步写入。再例如，前文提到的同步检测步骤，也可以重复执行，比如第一次执行完毕之后发现没有成功同步，那么间隔一段时间之后还可以再次检测，直到检测出两个数据服务器集群同步为止。

[0050] 通过以上各实施例的描述可知，采用本发明实施例提供的技术方案，在迁移数据量可能较大的第一时间点和第一时间点之前的目标数据的过程中，不需要停止要迁移业务的数据的正常写入和读取，比如应用服务器仍然可以向旧数据服务器集群（如前文描述的第一数据服务器集群）中的一个数据服务器进行数据写入和读取，同时从旧数据服务器集群中的另一个数据服务器导出第一时间点和第一时间点之前的目标数据到新数据服务器集群（如前文描述的第二数据服务器集群）；在迁移第一时间点之后写入旧数据服务器集群的目标数据过程中，是根据旧数据服务器集群的操作日志将第一时间之后的目标数据同步写入到新数据服务器中，在此过程中仍然不需要停止要迁移的业务，在两个新旧数据服务器集群完全同步之后，要迁移的业务就可以直接向新数据服务器集群正常写入和读取数据了。由此可见，采用本发明的技术方案，在整个业务的迁移过程中，都不需要停止要迁移的业务，进而就不会影响该业务对外的正常服务和运行，从而实现了不需要停机备份即可实现数据迁移的有益效果。

[0051] 在此提供的算法和显示不与任何特定计算机、虚拟系统或者其它设备固有相关。各种通用系统也可以与基于在此的示教一起使用。根据上面的描述，构造这类系统所要求

的结构是显而易见的。此外，本发明也不针对任何特定编程语言。应当明白，可以利用各种编程语言实现在此描述的本发明的内容，并且上面对特定语言所做的描述是为了披露本发明的最佳实施方式。

[0052] 在此处所提供的说明书中，说明了大量具体细节。然而，能够理解，本发明的实施例可以在没有这些具体细节的情况下实践。在一些实例中，并未详细示出公知的方法、结构和技术，以便不模糊对本说明书的理解。

[0053] 类似地，应当理解，为了精简本公开并帮助理解各个发明方面中的一个或多个，在上面对本发明的示例性实施例的描述中，本发明的各个特征有时被一起分组到单个实施例、图、或者对其的描述中。然而，并不应将该公开的方法解释成反映如下意图：即所要求保护的本发明要求比在每个权利要求中所明确记载的特征更多的特征。更确切地说，如下面的权利要求书所反映的那样，发明方面在于少于前面公开的单个实施例的所有特征。因此，遵循具体实施方式的权利要求书由此明确地并入该具体实施方式，其中每个权利要求本身都作为本发明的单独实施例。

[0054] 本领域那些技术人员可以理解，可以对实施例中的设备中的模块进行自适应性地改变并且把它们设置在与该实施例不同的一个或多个设备中。可以把实施例中的模块或单元或组件组合成一个模块或单元或组件，以及此外可以把它们分成多个子模块或子单元或子组件。除了这样的特征和 / 或过程或者单元中的至少一些是相互排斥之外，可以采用任何组合对本说明书（包括伴随的权利要求、摘要和附图）中公开的所有特征以及如此公开的任何方法或者设备的所有过程或单元进行组合。除非另外明确陈述，本说明书（包括伴随的权利要求、摘要和附图）中公开的每个特征可以由提供相同、等同或相似目的的替代特征来代替。

[0055] 此外，本领域的技术人员能够理解，尽管在此的一些实施例包括其它实施例中所包括的某些特征而不是其它特征，但是不同实施例的特征的组合意味着处于本发明的范围之内并且形成不同的实施例。例如，在下面的权利要求书中，所要求保护的实施例的任意之一都可以以任意的组合方式来使用。

[0056] 本发明的各个部件实施例可以以硬件实现，或者以在一个或者多个处理器上运行的软件模块实现，或者以它们的组合实现。本领域的技术人员应当理解，可以在实践中使用微处理器或者数字信号处理器（DSP）来实现根据本发明实施例的用于数据服务器集群之间进行数据迁移的迁移设备中的一些或者全部部件的一些或者全部功能。本发明还可以实现为用于执行这里所描述的方法的一部分或者全部的设备或者设备程序（例如，计算机程序和计算机程序产品）。这样的实现本发明的程序可以存储在计算机可读介质上，或者可以具有一个或者多个信号的形式。这样的信号可以从因特网网站上下载得到，或者在载体信号上提供，或者以任何其他形式提供。

[0057] 应该注意的是上述实施例对本发明进行说明而不是对本发明进行限制，并且本领域技术人员在不脱离所附权利要求的范围的情况下可设计出替换实施例。在权利要求中，不应将位于括号之间的任何参考符号构造成对权利要求的限制。单词“包含”不排除存在未列在权利要求中的元件或步骤。位于元件之前的单词“一”或“一个”不排除存在多个这样的元件。本发明可以借助于包括有若干不同元件的硬件以及借助于适当编程的计算机来实现。在列举了若干设备的单元权利要求中，这些设备中的若干个可以是通过同一个硬件

项来具体体现。单词第一、第二、以及第三等的使用不表示任何顺序。可将这些单词解释为名称。

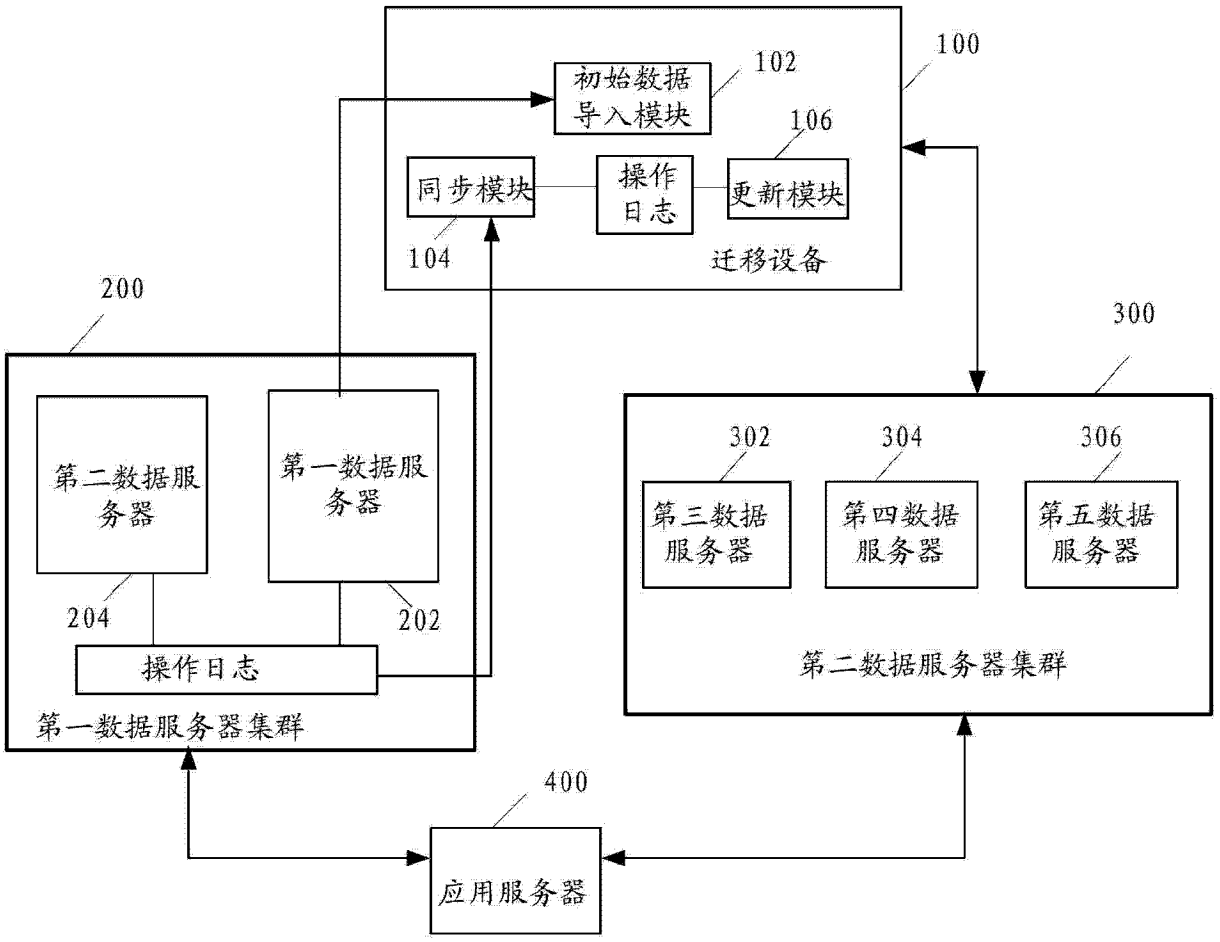

100
初始数据导入模块
102
106
同步模块
操作日志
更新模块
104
迁移设备
200
300
第二数据服务器
第一数据服务器
302
304
306
第三数据服务器
第四数据服务器
第五数据服务器
204
202
操作日志
第一数据服务器集群
第二数据服务器集群
400
应用服务器

图 1

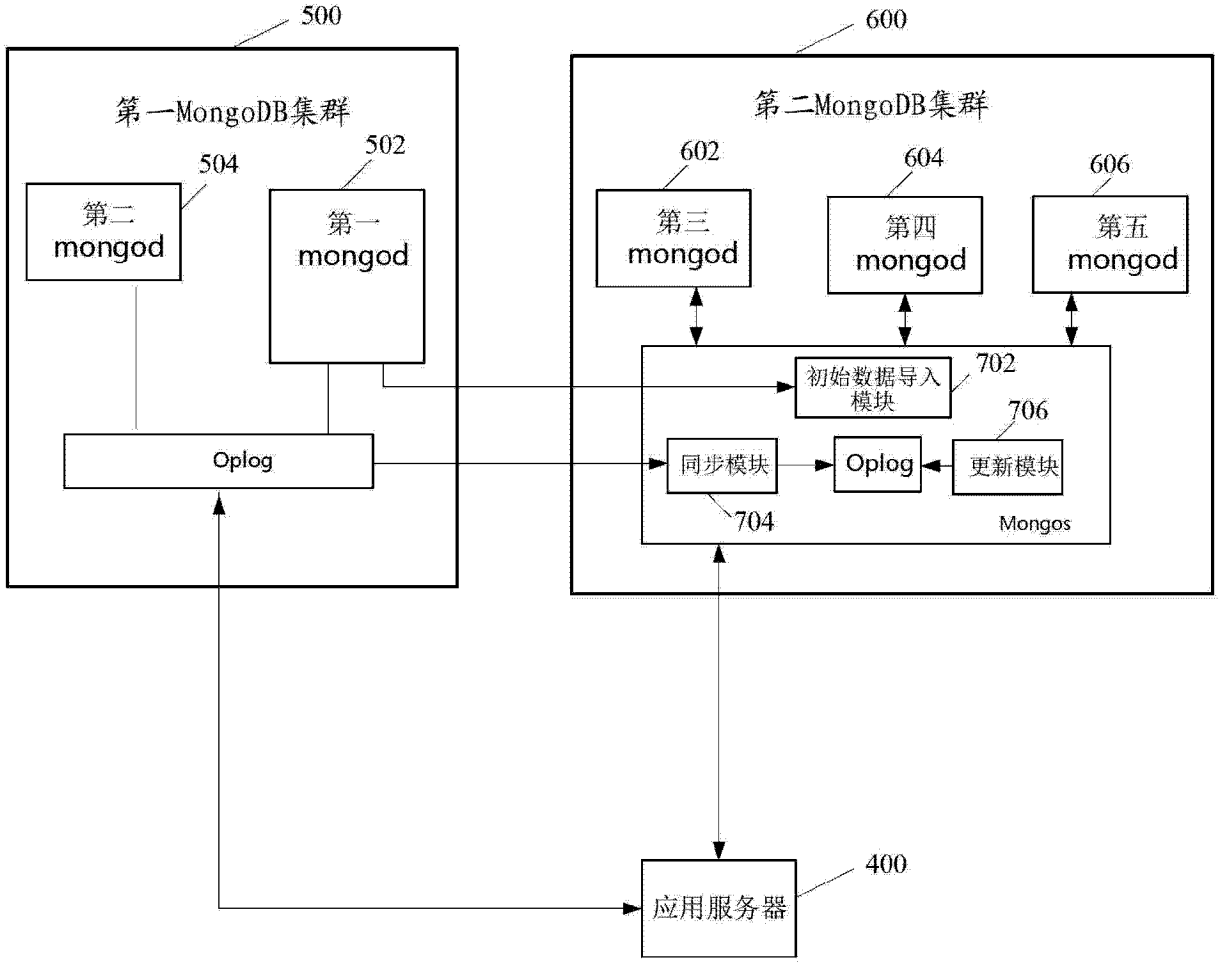

500
第一MongoDB集群
504
第二
mongod
502
第一
mongod
Oplog
600
第二MongoDB集群
602
第三
mongod
604
第四
mongod
606
第五
mongod
初始数据导入
模块
702
706
同步模块
Oplog
更新模块
704
Mongos
400
应用服务器

图 2

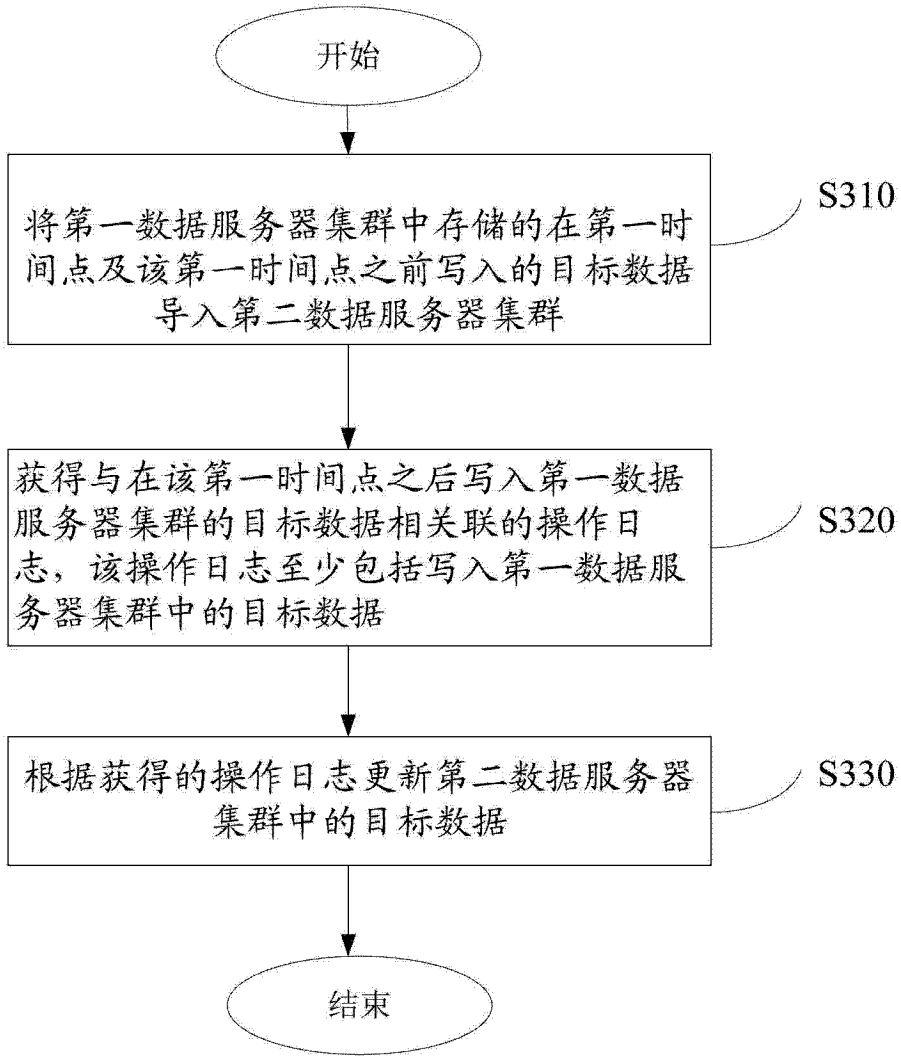

开始
将第一数据服务器集群中存储的在第一时间点及该第一时间点之前写入的目标数据导入第二数据服务器集群
S310
获得与在该第一时间点之后写入第一数据服务器集群的目标数据相关联的操作日志，该操作日志至少包括写入第一数据服务器集群中的目标数据
S320
根据获得的操作日志更新第二数据服务器集群中的目标数据
S330
结束

图 3