



(12) 发明专利

(10) 授权公告号 CN 111723073 B

(45) 授权公告日 2023. 09. 01

(21) 申请号 201910207044.4

G06F 16/22 (2019.01)

(22) 申请日 2019.03.19

G06F 16/23 (2019.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 111723073 A

(56) 对比文件

CN 106503139 A, 2017.03.15

WO 2018021593 A1, 2018.02.01

(43) 申请公布日 2020.09.29

CN 108132830 A, 2018.06.08

(73) 专利权人 北京沃东天骏信息技术有限公司

CN 106572153 A, 2017.04.19

地址 100176 北京市大兴区北京经济技术开发区科创十一街18号院2号楼4层A402室

李法平等.Redis可扩展与高效复制方案的设计与实现.《西南师范大学学报(自然科学版)》.2018,第43卷(第03期),第[83-90]页.

(72) 发明人 王大开

审查员 叶娅娟

(74) 专利代理机构 中国贸促会专利商标事务所

有限公司 11038

专利代理师 方亮

(51) Int. Cl.

G06F 16/21 (2019.01)

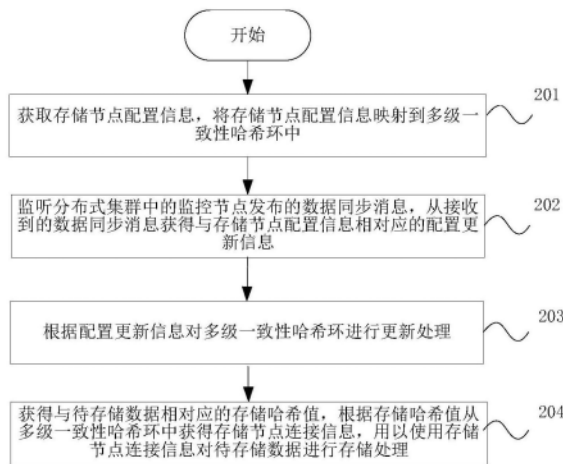
权利要求书3页 说明书10页 附图7页

(54) 发明名称

数据存储处理方法、装置、处理系统以及存储介质

(57) 摘要

本公开提供了一种数据存储处理方法、装置、处理系统以及存储介质,涉及计算机技术领域,其中方法包括:将存储节点配置信息映射到多级一致性哈希环中,监听分布式集群中的监控节点发布的数据同步消息,从接收到的数据同步消息获得与存储节点配置信息相对应的配置更新信息;根据配置更新信息对多级一致性哈希环进行更新处理;根据与待存储数据相对应的存储哈希值从多级一致性哈希环中获得存储节点连接信息。本公开的方法、装置、处理系统以及存储介质,可以结合增量和一致性哈希扩容的两种优点,通过修改与数据库分片信息相对应的优先级,将存储数据引导到性能好的服务器上;在达到写入上限的情况下可以避免继续写入;能够实现数据迁移而且对用户无感知,能够保证配置更新分发的及时性。



1. 一种数据存储处理方法,包括:

获取层次化的存储节点配置信息,将所述存储节点配置信息映射到多级一致性哈希环中;

其中,所述多级一致性哈希环包括三级一致性哈希环,一个一致性哈希环中的一个节点指向一个下一级一致性哈希环;所述层次化的存储节点配置信息包括:数据库分组信息、与所述数据库分组对应的起始哈希值和终止哈希值;所述数据库分组信息包括:数据分片信息、与所述数据分片信息相对应的优先级;所述数据分片信息包括:分片号、与所述分片号相对应的数据库连接信息;

监听分布式集群中的监控节点发布的数据同步消息,从接收到的所述数据同步消息获得与所述存储节点配置信息相对应的配置更新信息;

根据所述配置更新信息对所述多级一致性哈希环进行更新处理;

获得与待存储数据相对应的存储哈希值,根据所述存储哈希值从所述多级一致性哈希环中获得存储节点连接信息,用以使用所述存储节点连接信息对待存储数据进行存储处理。

2. 如权利要求1所述的方法,所述获取层次化的存储节点配置信息包括:

从所述分布式集群中的配置存储节点获得所述层次化的存储节点配置信息并在本地存储。

3. 如权利要求2所述的方法,还包括:

使用树形数据结构存储所述存储节点配置信息;其中,所述树形数据结构包括:TreeMap。

4. 如权利要求2所述的方法,所述根据所述存储哈希值从所述多级一致性哈希环中获得存储节点连接信息包括:

使用所述存储哈希值与所述多级一致性哈希环中的存储信息节点的哈希值分区进行匹配处理,用于获得所述数据库连接信息;

其中,同一级一致性哈希环中的存储信息节点的哈希值分区不重合;所述存储信息节点包括:实际节点、为实际节点在同一级一致性哈希环中分配的虚拟节点。

5. 如权利要求4所述的方法,所述将所述存储节点配置信息映射到多级一致性哈希环中包括:

基于所述起始哈希值和所述终止哈希值将多个数据库分组信息作为第一存储信息节点映射到第一级一致性哈希环中;

基于所述优先级将多个数据分片信息作为第二存储信息节点映射到第二级一致性哈希环中;其中,所述第二存储信息节点的哈希值区间为优先级区间;

基于所述分片号将多个数据库连接信息作为第三存储信息节点映射到第三级一致性哈希环中;其中,所述第三存储信息节点的哈希值区间为分片号区间。

6. 如权利要求5所述的方法,所述使用所述存储哈希值与所述多级一致性哈希环中的存储信息节点的哈希值分区进行匹配处理包括:

将所述存储哈希值与所述第一存储信息节点的哈希值分区进行匹配处理,根据匹配结果从所述第一存储信息节点中选取第一处理节点;

获得匹配优先级值,将与所述第一处理节点对应的第二级一致性哈希环中的第二存储

信息节点的哈希值分区与所述匹配优先级值进行匹配处理,根据匹配结果从所述第二存储信息节点中选取第二处理节点;

将与所述第二处理节点对应的第三级一致性哈希环中的第三存储信息节点的哈希值分区与所述存储哈希值进行匹配处理,根据匹配结果从所述第三存储信息节点中选取第三处理节点;

获取与第三处理节点相对应的数据库连接信息。

7.如权利要求6所述的方法,所述存储节点配置信息还包括:数据分库的状态信息;所述方法还包括:

将所述匹配优先级的初始化值设置为0;

基于所述状态信息判断与选取的第二处理节点相对应的数据分片信息是否能够选取;

如果否,则将所述匹配优先级的值增加预设的数值,并将与所述第一处理节点对应的第二级一致性哈希环中的第二存储信息节点的哈希值分区与所述匹配优先级值进行匹配,直至基于所述状态信息确定此第二处理节点相对应的数据分片信息能够选取。

8.如权利要求7所述的方法,所述存储节点配置信息还包括:数据分库的最大分片数信息;所述方法还包括:

基于所述数据分库的最大分片数信息获得与所述第二处理节点相对应的数据分库的最大分片数;

对所述存储哈希值与最大分片数进行取余运算,获得所述存储哈希值与最大分片数相除后的余数;

将与所述第二处理节点对应的第三级一致性哈希环中的第三存储信息节点的哈希值分区与所述余数进行匹配,根据匹配结果从所述第三存储信息节点中选取第三处理节点。

9.如权利要求8所述的方法,还包括:

当所述监控节点监听到所述存储节点配置信息发生变更时,所述监控节点根据此变更生成所述数据同步消息并在所述分布式集群中的消息队列中发布;

预先在所述消息队列中订阅所述数据同步消息,从所述消息队列中接收所述数据同步消息。

10.如权利要求1至9任一所述的方法,其中,

所述分布式集群包括:Zookeeper集群。

11.一种数据存储处理装置,包括:

配置信息获取模块,用于获取层次化的存储节点配置信息,将所述存储节点配置信息映射到多级一致性哈希环中;其中,所述多级一致性哈希环包括三级一致性哈希环,一个一致性哈希环中的一个节点指向一个下一级一致性哈希环;所述层次化的存储节点配置信息包括:数据库分组信息、与所述数据库分组对应的起始哈希值和终止哈希值;所述数据库分组信息包括:数据分片信息、与所述数据分片信息相对应的优先级;所述数据分片信息包括:分片号、与所述分片号相对应的数据库连接信息;

配置更新获得模块,用于监听分布式集群中的监控节点发布的数据同步消息,从接收到的所述数据同步消息获得与所述存储节点配置信息相对应的配置更新信息;

配置更新同步模块,用于根据所述配置更新信息对所述多级一致性哈希环进行更新处理;

存储目标获得模块,用于获得与待存储数据相对应的存储哈希值,根据所述存储哈希值从所述多级一致性哈希环中获得存储节点连接信息,用以使用所述存储节点连接信息对待存储数据进行存储处理。

12. 一种数据存储处理装置,包括:

存储器;以及耦接至所述存储器的处理器,所述处理器被配置为基于存储在所述存储器中的指令,执行如权利要求1至10中任一项所述的方法。

13. 一种数据处理系统,包括:

多个应用服务器;所述应用服务器包括:如权利要求11或12所述的数据存储处理装置。

14. 一种计算机可读存储介质,所述计算机可读存储介质存储有计算机指令,所述指令被处理器执行如权利要求1至10中任一项所述的方法。

数据存储处理方法、装置、处理系统以及存储介质

技术领域

[0001] 本公开涉及计算机技术领域,尤其涉及一种数据存储处理方法、装置、处理系统以及存储介质。

背景技术

[0002] 水平分库是解决数据库高并发写的一种方案,软件行业中目前流行的分库中间件产品有MyCat等。分库中间件即根据业务字段采用一致性哈希做分片,当数据库遇到瓶颈需要扩容分片时,仍然需要对历史分片进行迁移来保证正确的哈希路由,即使数据迁移是分片级的也会增加不小资源消耗和系统的复杂性。使用MyCat实现水平分库方案如图1所示,单独部署MyCat中间件服务,MyCAT是mysql中间件,配置MyCat分片路由规则,应用直接通过Jdbc连接MyCat,MyCat实现了Mysql协议拦截Sql并解析,根据配置的分库规则,路由到指定实例的分片上;分片扩容前需要re-hash对分片进行迁移,可以在分片迁移后不进行行级数据迁移。修改MyCat分片路由规则,平滑重启MyCat使新路由规则生效,对新增业务数据的行级迁移。

发明内容

[0003] 发明人发现目前的水平分库方案具有下述问题:需要单独部署中间件做代理,中间件可能会成为新的性能瓶颈;扩容时需按分片迁移数据,带来不少的工作量和出错几率;分片数据迁移完成后需要解决迁移过程中产生的新业务数据的行级数据迁移问题。

[0004] 有鉴于此,本公开提供一种数据存储处理方法、数据存储处理装置、数据处理系统以及计算机可读存储介质。

[0005] 根据本公开的一个方面,提供一种数据存储处理方法,包括:获取存储节点配置信息,将所述存储节点配置信息映射到多级一致性哈希环中;监听分布式集群中的监控节点发布的数据同步消息,从接收到的所述数据同步消息获得与所述存储节点配置信息相对应的配置更新信息;根据所述配置更新信息对所述多级一致性哈希环进行更新处理;获得与待存储数据相对应的存储哈希值,根据所述存储哈希值从所述多级一致性哈希环中获得存储节点连接信息,用以使用所述存储节点连接信息对待存储数据进行存储处理。

[0006] 可选地,所述获取存储节点配置信息包括:从所述分布式集群中的配置存储节点获得层次化的所述存储节点配置信息并在本地存储;其中,所述存储节点配置信息包括:数据库分组信息、与所述数据库分组对应的起始哈希值和终止哈希值;所述数据库分组信息包括:数据分片信息、与所述数据库分片信息相对应的优先级;所述数据分片信息包括:分片号、与所述分片号相对应的数据库连接信息。

[0007] 可选地,使用树形数据结构存储所述存储节点配置信息;其中,所述树状存储结构包括:TreeMap。

[0008] 可选地,所述根据所述存储哈希值从所述多级一致性哈希环中获得存储节点连接信息包括:使用所述存储哈希值与所述多级一致性哈希环中的存储信息节点的哈希值分区

进行匹配处理,用于获得所述数据库连接信息;其中,同一级一致性哈希环中的存储信息节点的哈希值分区不重合;所述存储信息节点包括:实际节点、为实际节点在同一级一致性哈希环中分配的虚拟节点。

[0009] 可选地,所述将所述存储节点配置信息映射到多级一致性哈希环中包括:基于所述起始哈希值和所述终止哈希值将多个数据库分组信息作为第一存储信息节点映射到第一级一致性哈希环中;基于所述优先级将多个数据分片信息作为第二存储信息节点映射到第二级一致性哈希环中;其中,所述第二存储信息节点的哈希值区间为优先级区间;基于所述分片号将多个数据库连接信息作为第三存储信息节点映射到第三级一致性哈希环中;其中,所述第三存储信息节点的哈希值区间为分片号区间。

[0010] 可选地,所述使用所述存储哈希值与所述多级一致性哈希环中的存储信息节点的哈希值分区进行匹配处理包括:将所述存储哈希值与所述第一存储信息节点的哈希值分区进行匹配处理,根据匹配结果从所述第一存储信息节点中选取第一处理节点;获得匹配优先级值,将与所述第一处理节点对应的第二级一致性哈希环中的第二存储信息节点的哈希值分区与所述匹配优先级值进行匹配处理,根据匹配结果从所述第二存储信息节点中选取第二处理节点;将与所述第二处理节点对应的第三级一致性哈希环中的第三存储信息节点的哈希值分区与所述存储哈希值进行匹配处理,根据匹配结果从所述第三存储信息节点中选取第三处理节点;获取与第三处理节点相对应的数据库连接信息。

[0011] 可选地,所述存储节点配置信息还包括:数据分库的状态信息;所述方法还包括:将所述匹配优先级的初始化值设置为0;基于所述状态信息判断与选取的第二处理节点相对应的数据分片信息是否能够选取;如果否,则将所述匹配优先级的值增加预设的数值,并将与所述第一处理节点对应的第二级一致性哈希环中的第二存储信息节点的哈希值分区与所述匹配优先级值进行匹配,直至基于所述状态信息确定此第二处理节点相对应的数据分片信息能够选取。

[0012] 可选地,所述存储节点配置信息还包括:数据分库的最大分片数信息;所述方法还包括:基于所述数据分库的最大分片数信息获得与所述第二处理节点相对应的数据分库的最大分片数;对所述存储哈希值与最大分片数进行取余运算,获得所述存储哈希值与最大分片数相除后的余数;将与所述第二处理节点对应的第三级一致性哈希环中的第三存储信息节点的哈希值分区与所述余数进行匹配,根据匹配结果从所述第三存储信息节点中选取第三处理节点。

[0013] 可选地,当所述监控节点监听到所述存储节点配置信息发生变更时,所述监控节点根据此变更生成所述数据同步消息并在所述分布式集群中的消息队列中发布;预先在所述消息队列中订阅所述数据同步消息,从所述消息队列中接收所述数据同步消息。

[0014] 可选地,所述分布式集群包括:Zookeeper集群。

[0015] 根据本公开的另一方面,提供一种数据存储处理装置,包括:配置信息获取模块,用于获取存储节点配置信息,将所述存储节点配置信息映射到多级一致性哈希环中;配置更新获得模块,用于监听分布式集群中的监控节点发布的数据同步消息,从接收到的所述数据同步消息获得与所述存储节点配置信息相对应的配置更新信息;配置更新同步模块,用于根据所述配置更新信息对所述多级一致性哈希环进行更新处理;存储目标获得模块,用于获得与待存储数据相对应的存储哈希值,根据所述存储哈希值从所述多级一致性哈希

环中获得存储节点连接信息,用以使用所述存储节点连接信息对待存储数据进行存储处理。

[0016] 根据本公开的又一方面,提供一种数据存储处理装置,包括:存储器;以及耦接至所述存储器的处理器,所述处理器被配置为基于存储在所述存储器中的指令,执行如上所述的方法。

[0017] 根据本公开的再一方面,提供一种数据处理系统,包括:多个应用服务器;所述应用服务器包括:如上所述的数据存储处理装置。

[0018] 根据本公开的再一方面,提供一种计算机可读存储介质,所述计算机可读存储介质存储有计算机指令,所述指令被处理器执行如上所述的方法。

[0019] 本公开的数据存储处理方法、装置、处理系统以及存储介质,将存储节点配置信息映射到多级一致性哈希环中,监听分布式集群中的监控节点发布的数据同步消息对多级一致性哈希环进行更新处理;根据与待存储数据相对应的存储哈希值从多级一致性哈希环中获得存储节点连接信息;可以结合增量和一致性哈希扩容的两种优点,能够实现数据迁移而且对用户无感知,能够保证配置更新分发的及时性,不存在中间件的性能问题,能够提高工作效率、提高用户的使用感受度。

附图说明

[0020] 为了更清楚地说明本公开实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作一简单地介绍,显而易见地,下面描述中的附图仅仅是本公开的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图。

[0021] 图1为现有技术中使用MyCat实现水平分库方案的示意图;

[0022] 图2为根据本公开的数据存储处理方法的一个实施例的流程示意图;

[0023] 图3为在Zookeeper集群中保存存储节点配置信息的目录结构示意图;

[0024] 图4为层次化的存储节点配置信息的存储示意图;

[0025] 图5为根据本公开的数据存储处理方法的一个实施例中的将存储节点配置信息映射到多级一致性哈希环中的流程示意图;

[0026] 图6为多级一致性哈希环中的示意图;

[0027] 图7为根据本公开的数据存储处理方法的一个实施例中的在多级一致性哈希环中进行匹配处理的流程示意图;

[0028] 图8为多个APP使用本公开的数据存储处理方法的示意图;

[0029] 图9为根据本公开的数据存储处理装置的一个实施例的模块示意图;

[0030] 图10为根据本公开的数据存储处理装置的一个实施例中的配置信息获取模块的模块示意图;

[0031] 图11为根据本公开的数据存储处理装置的一个实施例中的存储目标获得模块的模块示意图;

[0032] 图12为根据本公开的数据存储处理装置的另一个实施例的模块示意图。

具体实施方式

[0033] 下面参照附图对本公开进行更全面的描述,其中说明本公开的示例性实施例。下面将结合本公开实施例中的附图,对本公开实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本公开一部分实施例,而不是全部的实施例。基于本公开中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本公开保护的范围。下面结合各个图和实施例对本公开的技术方案进行多方面的描述。

[0034] 下文中的“第一”、“第二”等仅用于描述上相区别,并没有其他特殊的含义。

[0035] 图2为根据本公开的数据存储处理方法的一个实施例的流程示意图,如图2所示:

[0036] 步骤201获取存储节点配置信息,将存储节点配置信息映射到多级一致性哈希环中。

[0037] 多级一致性哈希环可以为两级、三级一致性哈希环等。一个一致性哈希环中的一个节点指向一个下一级一致性哈希环,非最低一级的一致性哈希环中的节点存储的信息包括下一级一致性哈希环的指针、引用等。形成多级环结构。

[0038] 步骤202,监听分布式集群中的监控节点发布的数据同步消息,从接收到的数据同步消息获得与存储节点配置信息相对应的配置更新信息。

[0039] 步骤203,根据配置更新信息对多级一致性哈希环进行更新处理。

[0040] 步骤204,获得与待存储数据相对应的存储哈希值,根据存储哈希值从多级一致性哈希环中获得存储节点连接信息,用以使用存储节点连接信息对待存储数据进行存储处理。

[0041] 在一个实施例中,从分布式集群中的配置存储节点获得层次化的存储节点配置信息并在本地存储。存储节点配置信息包括:数据库分组信息、与数据库分组对应的起始哈希值和终止哈希值;数据库分组信息包括:数据分片信息、与数据库分片信息相对应的优先级;数据分片信息包括:分片号、与分片号相对应的数据库连接信息。数据库连接信息可以为多种,数据库连接信息包括:url、用户名、密码等,可以使用JDBC方式进行连接等。

[0042] 在分布式存储系统中,数据需要分散存储在多台设备上,数据分片(Sharding)用来确定数据在多台存储设备上分布的技术。数据分片包括水平分片(也叫水平分库)等,水平分片指的是将整体存储在单个数据库中的数据,通过某种策略分摊到多个表结构与其相同的数据库中,这样每个数据库中的数据量就会相对减少很多,并且可以部署在不同物理服务器上,能够实现数据库的无限横向拓展。

[0043] 分布式集群可以有多种,例如Zookeeper集群等。ZooKeeper是一个为分布式应用提供一致性服务的软件,提供的功能包括:配置维护、域名服务、分布式同步、组服务等,是Hadoop和Hbase的重要组件。如图3所示,Zookeeper中的目录结构为多层,例如/root/group1/shard1等,/root/group1节点保存的数据是:hash field的起始结束值,当前group分片数,预设最大分片数,是否有效,优先级;/root/group1/shard1节点保存的数据是数据库连接信息。

[0044] 可以使用树形数据结构存储存储节点配置信息,树状存储结构可以有多种,例如为TreeMap等。如图4所示,层次化的存储节点配置信息的存储结构为:

[0045] TreeMap<HashValue, TreeMap<Priority, TreeMap<Shard, DataSource>>>>。

[0046] 第一层存储节点配置信息为存储节点配置信息,存储节点配置信息的数量可以为

多个。存储节点配置信息包括：数据库分组信息、与数据库分组对应的起始哈希值和终止哈希值。数据库分组信息为TreeMap第一层的value值，即为

[0047] TreeMap<Priority, TreeMap<Shard, DataSource>, 与数据库分组对应的起始哈希值和终止哈希值为TreeMap第一层的key值，例如为40000，则启示哈希值为0，终止哈希值为40000。

[0048] 每个数据库分组信息包括：数据分片信息、与数据库分片信息相对应的优先级。数据分片信息为存储结构第二层的value值，即为两个TreeMap<Shard, DataSourceInfo>, 与数据库分片信息相对应的优先级为TreeMap第二层的key值，例如为1、2等。每个数据分片信息包括：分片号、与分片号相对应的数据库连接信息。与分片号相对应的数据库连接信息为TreeMap第三层的value值，即为DataSourceInfo1、DataSourceInfo1，分片号为存储结构第三层的key值，例如为32、64等。

[0049] 应用在启动时获取所有存储节点配置并缓存到本地，根据存储节点配置初始化存储节点配置信息，存储节点配置信息保存在Java的数据结构TreeMap中，TreeMap基于红黑树实现算法的时间复杂度是 $O(\log 2N)$ ，能够提高运行效率减小运算负载度。

[0050] 在一个实施例中，获得与待存储数据对应的存储哈希值，存储哈希值的获得可以采用多种方式。例如，待存储数据为一个用户订单，获得用户订单中的一个或多个业务信息，将业务信息进行预设的哈希运算获得存储哈希值。使用存储哈希值与多级一致性哈希环中的存储信息节点的哈希值分区进行匹配处理，用于获得数据库连接信息，使用数据库连接信息建立数据库连接，用于存储数据。同一级一致性哈希环中的存储信息节点的哈希值分区不重合，存储信息节点包括：实际节点、为实际节点在同一级一致性哈希环中分配的虚拟节点。

[0051] 多级一致性哈希环的级数可以为三级等。例如，图5为根据本公开的数据存储处理方法的一个实施例中的将存储节点配置信息映射到多级一致性哈希环中的流程示意图，如图5所示：

[0052] 步骤501，基于起始哈希值和终止哈希值将多个数据库分组信息作为第一存储信息节点映射到第一级一致性哈希环中。

[0053] 步骤502，基于优先级将多个数据分片信息作为第二存储信息节点映射到第二级一致性哈希环中，第二存储信息节点的哈希值区间为优先级区间。

[0054] 步骤503，基于分片号将多个数据库连接信息作为第三存储信息节点映射到第三级一致性哈希环中，第三存储信息节点的哈希值区间为分片号区间。

[0055] 如图6所示，建立三级一致性哈希环。一致性哈希算法(Consistent Hashing Algorithm)是一种分布式算法，一致性哈希将整个哈希值空间组织成一个虚拟的圆环。对于第一级一致性哈希环，获得全部待存储数据(例如用户订单)的存储哈希值的最大值为K，第一级一致性哈希环061的哈希值空间H1为0-K，将整个H1按照顺时针方向组织，0和K在零点方向重合。

[0056] 基于三个数据库分组信息对应的起始哈希值和终止哈希值，将三个数据库分组信息作为三个第一存储信息节点611、612、613映射到第一级一致性哈希环中061中。

[0057] 对于第二级一致性哈希环的建立，以第一存储信息节点613为例进行说明，建立与第一存储信息节点613相对应的第二级一致性哈希环062。建立与其他第一存储信息节点

611、612相对应的第二级一致性哈希环的方式与建立第二级一致性哈希环062的方法相同。

[0058] 获得与第一存储信息节点613对应的数据分片信息的优先级最大值为L,第二级一致性哈希环063的哈希值空间H2为0-L,将整个H2按照顺时针方向组织,0和L在零点方向重合。

[0059] 基于三个数据分片信息对应的优先级,将三个数据分片信息作为三个第二存储信息节点621、622、623映射到第二级一致性哈希环062中。在第二级一致性哈希环062中,三个第二存储信息节点621、622、623的哈希值区间为优先级区间。

[0060] 例如,获得与第一存储信息节点613对应的三个数据分片信息的优先级分别为1、5、10,优先级最大值为15,第二级一致性哈希环的哈希值空间H2为0-15,将整个H2按照顺时针方向组织,0和15在零点方向重合。基于三个数据分片信息对应的优先级1、5、10,将三个数据分片信息作为三个第二存储信息节点621、622、623映射到第二级一致性哈希环062中。三个第二存储信息节点621、622、623的哈希值区间为优先级区间。

[0061] 对于第三级一致性哈希环的建立,以第二存储信息节点621为例进行说明,建立与第二存储信息节点621相对应的第三级一致性哈希环063。建立与其他第二存储信息节点622、623相对应的第三级一致性哈希环的方式与建立第三级一致性哈希环063的方法相同。

[0062] 获得与第二存储信息节点613对应的数据库连接信息的分片数最大值为N,第三级一致性哈希环的哈希值空间H3为0-N,将整个H3按照顺时针方向组织,0和N在零点方向重合。基于与三个数据库连接信息对应的分片号,将三个数据库连接信息作为三个第三存储信息节点631、632、633映射到第三级一致性哈希环063中。第三存储信息节点631、632、633的哈希值区间为分片号区间。

[0063] 例如,获得与第二存储信息节点613对应的三个数据库连接信息的分片数分别为2、6、10,分片数最大值为16,第三级一致性哈希环的哈希值空间H3为0-16,将整个H3按照顺时针方向组织,0和16在零点方向重合。基于与三个数据库连接信息对应的分片号2、6、10,将三个数据库连接信息作为三个第三存储信息节点631、632、633映射到第三级一致性哈希环063中。第三存储信息节点631、632、633的哈希值区间为分片号区间。

[0064] 如果一致性哈希环的实际节点太少时,容易因为节点分部不均匀而造成数据倾斜问题,一致性哈希算法引入了虚拟节点机制,即对每一个节点计算多个哈希,每个计算结果位置都放置一个此节点,作为虚拟节点。

[0065] 在一个实施例中,根据存储哈希值从多级一致性哈希环中获得存储节点连接信息可以有多种方法。图7为根据本公开的数据存储处理方法的一个实施例中的在多级一致性哈希环中进行匹配处理的流程示意图,如图7所示:

[0066] 步骤701,将存储哈希值与第一存储信息节点的哈希值分区进行匹配处理,根据匹配结果从第一存储信息节点中选取第一处理节点。

[0067] 步骤702,获得匹配优先级值,将与第一处理节点对应的第二级一致性哈希环中的第二存储信息节点的哈希值分区与匹配优先级值进行匹配处理,根据匹配结果从第二存储信息节点中选取第二处理节点。

[0068] 存储节点配置信息还包括:数据分库的状态信息。将匹配优先级的初始化值设置为0,基于状态信息判断与选取的第二处理节点相对应的数据分片信息是否能够选取,如果否,则将匹配优先级的值增加预设的数值,并将与第一处理节点对应的第二级一致性哈希

环中的第二存储信息节点的哈希值分区与匹配优先级值进行匹配,直至基于状态信息确定此第二处理节点相对应的数据分片信息能够选取。

[0069] 步骤703,将与第二处理节点对应的第三级一致性哈希环中的第三存储信息节点的哈希值分区与存储哈希值进行匹配处理,根据匹配结果从第三存储信息节点中选取第三处理节点。

[0070] 存储节点配置信息还包括:数据分库的最大分片数信息。基于数据分库的最大分片数信息获得与第二处理节点相对应的数据分库的最大分片数,对存储哈希值与最大分片数进行取余运算,获得存储哈希值与最大分片数相除后的余数,将与第二处理节点对应的第三级一致性哈希环中的第三存储信息节点的哈希值分区与余数进行匹配,根据匹配结果从第三存储信息节点中选取第三处理节点。

[0071] 步骤704,获取与第三处理节点相对应的数据库连接信息。

[0072] 例如,如图6所示,接收到用户订单,根据预设的哈希函数对用户订单中的业务数据进行哈希运算,获得存储哈希值为hash field,例如hash field为1230000。判断此存储哈希值1230000落在第一级一致性哈希环061中第一存储信息节点611和第一存储信息节点613之间的哈希值分区上,顺着逆时针方向行走遇到第一存储信息节点613,则选取第一处理节点为第一存储信息节点613。

[0073] 将匹配优先级的初始化值设置为0,判断匹配优先级0落在第二级一致性哈希环062中的第二存储信息节点621和第一存储信息节点622之间的哈希值分区上,顺着逆时针方向行走遇到第二存储信息节点621,选取第二处理节点为第二存储信息节点621。

[0074] 存储节点配置信息还包括数据分库的状态信息,基于状态信息判断第二处理节点相对应的第二存储信息节点621是否能够选取。如果第二存储信息节点621的状态信息为繁忙或不可用,则第二存储信息节点622不能够选取,则将匹配优先级的值0增加预设的数值5。

[0075] 将第二级一致性哈希环062中的三个第二存储信息节点621、622、623的哈希值分区与匹配优先级值5进行匹配。判断匹配优先级5落在第二级一致性哈希环062中的第二存储信息节点622和第一存储信息节点623之间的哈希值分区上,顺着逆时针方向行走遇到第二存储信息节点622,选取第二处理节点为第二存储信息节点622,如果第二存储信息节点622的状态信息为正常,则第二存储信息节点622能够选取,如果第二存储信息节点622不能够选取,则将匹配优先级的值5再增加预设的数值5,并重复上述的步骤,直至基于状态信息确定此第二处理节点相对应的数据分片信息能够选取。

[0076] 如果选取的第二处理节点为第二存储信息节点621,存储节点配置信息包括:数据分库的最大分片数信息。基于数据分库的最大分片数信息获得与第二处理节点相对应的数据分库的最大分片数cur_shards;对存储哈希值hash field与最大分片数cur_shards进行取余运算,获得存储哈希值hash field与最大分片数cur_shards相除后的余数。例如,hash field%cur_shards取到的值为6,判断余数6落在第三级一致性哈希环063中的第三存储信息节点633和第三存储信息节点632之间的哈希值分区上,顺着逆时针方向行走遇到第三存储信息节点632,则选取第三处理节点为第三存储信息节点632,获取第三存储信息节点632对应的数据库连接信息,建立数据库连接并存储数据。

[0077] 如图8所示,多个APP应用用于对待存储数据进行存储。APP应用可以封装

DataSource接口实现类,在getConnection()实现方法中获得数据库连接信息。多个APP应用监听分布式集群中的监控节点发布的数据同步消息,从接收到的数据同步消息获得与存储节点配置信息相对应的配置更新信息。

[0078] 多个APP可以应用监听Zookeeper中的每一层节点create delete dataChange事件,当有变更通知时维护存储节点配置信息。或者,多个APP应用预先在消息队列中订阅数据同步消息,从消息队列中接收数据同步消息,当监控节点监听到存储节点配置信息发生变更时,监控节点根据此变更生成数据同步消息并在分布式集群中的消息队列中发布。

[0079] 上述实施例中的数据存储方法,在对数据进行存储时,可以结合增量和一致性哈希扩容的两种优点,能够根据机器性能来控制每个数据库分片存储的数量级,可以通过修改与数据库分片信息相对应的优先级,将存储数据引导到性能好的服务器上。例如,当前有10台性能一般的一组服务器Group1和2台性能非常好的一组服务器Group2,可以将服务器Group2上的数据库分片信息相对应的优先级修改为较小值,则数据可以优先存储在服务器Group2上。

[0080] 上述实施例中的数据存储方法,可以自由配置处理的数据量级,根据与数据库分组对应的起始哈希值和终止哈希值,提前规划在达到写入上限的情况下可以避免继续写入;对于写满的数据库分片,当历史数据被归档后,可以修改数据库分片信息相对应的优先级,使其再次被利用;通过配置数据分库的状态信息的有效状态和与分片相对应的优先级的机制,即使在数据库扩容时,也做到数据迁移的逻辑简单而且对用户无感知;使用Zookeeper的分布式集群做集群统一配置管理,可以在存储节点配置信息配置变动时异步通知到每台应用,保证配置更新分发的及时性;使用去中心化的思想每台应用缓存一套路由逻辑,不存在MyCat的单点性能问题。

[0081] 在一个实施例中,如图9所示,本公开提供一种数据存储处理装置90,包括:配置信息获取模块91、配置更新获得模块92、配置更新同步模块93和存储目标获得模块94。

[0082] 配置信息获取模块91获取存储节点配置信息,将存储节点配置信息映射到多级一致性哈希环中。配置更新获得模块92监听分布式集群中的监控节点发布的数据同步消息,从接收到的数据同步消息获得与存储节点配置信息相对应的配置更新信息。

[0083] 配置更新同步模块93根据配置更新信息对多级一致性哈希环进行更新处理。存储目标获得模块94获得与待存储数据相对应的存储哈希值,根据存储哈希值从多级一致性哈希环中获得存储节点连接信息,用以使用存储节点连接信息对待存储数据进行存储处理。

[0084] 当监控节点监听到存储节点配置信息发生变更时,监控节点根据此变更生成数据同步消息并在分布式集群中的消息队列中发布。配置获得单元912预先在消息队列中订阅数据同步消息,从消息队列中接收数据同步消息。分布式集群包括:Zookeeper集群等。

[0085] 在一个实施例中,如图10所示,配置信息获取模块91包括:配置获得单元911、配置保存单元912、第一哈希环构建单元913、第二哈希环构建单元914和第三哈希环构建单元915。

[0086] 配置获得单元911从分布式集群中的配置存储节点获得层次化的存储节点配置信息并在本地存储。存储节点配置信息包括:数据库分组信息、与数据库分组对应的起始哈希值和终止哈希值。数据库分组信息包括:数据分片信息、与数据库分片信息相对应的优先级。数据分片信息包括:分片号、与分片号相对应的数据库连接信息。

[0087] 配置保存单元912使用树形数据结构存储存储节点配置信息,树状存储结构包括:TreeMap等。存储目标获得模块94使用存储哈希值与多级一致性哈希环中的存储信息节点的哈希值分区进行匹配处理,用于获得数据库连接信息。同一级一致性哈希环中的存储信息节点的哈希值分区不重合,存储信息节点包括:实际节点、为实际节点在同一级一致性哈希环中分配的虚拟节点。

[0088] 第一哈希环构建单元913基于起始哈希值和终止哈希值将多个数据库分组信息作为第一存储信息节点映射到第一级一致性哈希环中。第二哈希环构建单元914基于优先级将多个数据分片信息作为第二存储信息节点映射到第二级一致性哈希环中,第二存储信息节点的哈希值区间为优先级区间。第三哈希环构建单元914基于分片号将多个数据库连接信息作为第三存储信息节点映射到第三级一致性哈希环中,第三存储信息节点的哈希值区间为分片号区间。

[0089] 在一个实施例中,如图11所示,存储目标获得模块94包括:第一匹配单元941、第二匹配单元942和第三匹配单元943。第一匹配单元941将存储哈希值与第一存储信息节点的哈希值分区进行匹配处理,根据匹配结果从第一存储信息节点中选取第一处理节点。第二匹配单元942获得匹配优先级值,将与第一处理节点对应的第二级一致性哈希环中的第二存储信息节点的哈希值分区与匹配优先级值进行匹配处理,根据匹配结果从第二存储信息节点中选取第二处理节点。

[0090] 第三匹配单元943将与第二处理节点对应的第三级一致性哈希环中的第三存储信息节点的哈希值分区与存储哈希值进行匹配处理,根据匹配结果从第三存储信息节点中选取第三处理节点。第三匹配单元943获取与第三处理节点相对应的数据库连接信息。

[0091] 在一个实施例中,存储节点配置信息还包括:数据分库的状态信息;第二匹配单元942将匹配优先级的初始化值设置为0,基于状态信息判断与选取的第二处理节点相对应的数据分片信息是否能够选取;如果否,则第二匹配单元942将匹配优先级的值增加预设的数值,并将与第一处理节点对应的第二级一致性哈希环中的第二存储信息节点的哈希值分区与匹配优先级值进行匹配,直至基于状态信息确定此第二处理节点相对应的数据分片信息能够选取。

[0092] 存储节点配置信息还包括:数据分库的最大分片数信息。第三匹配单元943基于数据分库的最大分片数信息获得与第二处理节点相对应的数据分库的最大分片数。第三匹配单元943对存储哈希值与最大分片数进行取余运算,获得存储哈希值与最大分片数相除后的余数。第三匹配单元943将与第二处理节点对应的第三级一致性哈希环中的第三存储信息节点的哈希值分区与余数进行匹配,根据匹配结果从第三存储信息节点中选取第三处理节点。

[0093] 图12为根据本公开的数据存储处理系统的另一个实施例的模块示意图。如图12所示,该装置可包括存储器1201、处理器1202、通信接口1203以及总线1204。存储器1201用于存储指令,处理器1202耦合到存储器1201,处理器1202被配置为基于存储器1201存储的指令执行实现上述的数据存储处理方法。

[0094] 存储器1201可以为高速RAM存储器、非易失性存储器(non-volatile memory)等,存储器1201也可以是存储器阵列。存储器1201还可能被分块,并且块可按一定的规则组合成虚拟卷。处理器1202可以为中央处理器CPU,或专用集成电路ASIC(Application

Specific Integrated Circuit), 或者是被配置成实施本公开的数据存储处理方法的一个或多个集成电路。

[0095] 在一个实施例中, 本公开提供一种数据处理系统, 包括: 多个应用服务器; 应用服务器包括如上任一实施例的数据存储处理装置。

[0096] 根据本公开的再一方面, 提供一种计算机可读存储介质, 计算机可读存储介质存储有计算机指令, 指令被处理器执行如上的方法。

[0097] 上述实施例中的数据存储处理方法、装置、处理系统以及存储介质, 将存储节点配置信息映射到多级一致性哈希环中, 监听分布式集群中的监控节点发布的数据同步消息对多级一致性哈希环进行更新处理; 获得与待存储数据相对应的存储哈希值, 根据存储哈希值从多级一致性哈希环中获得存储节点连接信息; 可以结合增量和一致性哈希扩容的两种优点, 控制每个数据库分片存储的数量级, 可以通过修改与数据库分片信息相对应的优先级, 将存储数据引导到性能好的服务器上; 根据与数据库分组对应的起始哈希值和终止哈希值, 提前规划在达到写入上限的情况下可以避免继续写入; 对于写满的数据库分片, 可以修改数据库分片信息相对应的优先级, 使其再次被利用; 通过配置数据分库的状态信息的有效状态和与分片相对应的优先级的机制, 即使在数据库扩容时, 也做到数据迁移的逻辑简单而且对用户无感知; 使用Zookeeper的分布式集群做集群统一配置管理, 可以在存储节点配置信息配置变动时异步通知到每台应用, 保证配置更新分发的及时性; 不存在中间件的性能问题, 能够提高工作效率、提高用户的使用感受度。

[0098] 可能以许多方式来实现本公开的方法和系统。例如, 可通过软件、硬件、固件或者软件、硬件、固件的任何组合来实现本公开的方法和系统。用于方法的步骤的上述顺序仅是为了进行说明, 本公开的方法的步骤不限于以上具体描述的顺序, 除非以其它方式特别说明。此外, 在一些实施例中, 还可将本公开实施为记录在记录介质中的程序, 这些程序包括用于实现根据本公开的方法的机器可读指令。因而, 本公开还覆盖存储用于执行根据本公开的方法的程序的记录介质。

[0099] 本公开的描述是为了示例和描述起见而给出的, 而并不是无遗漏的或者将本公开限于所公开的形式。很多修改和变化对于本领域的普通技术人员而言是显然的。选择和描述实施例是为了更好说明本公开的原理和实际应用, 并且使本领域的普通技术人员能够理解本公开从而设计适于特定用途的带有各种修改的各种实施例。

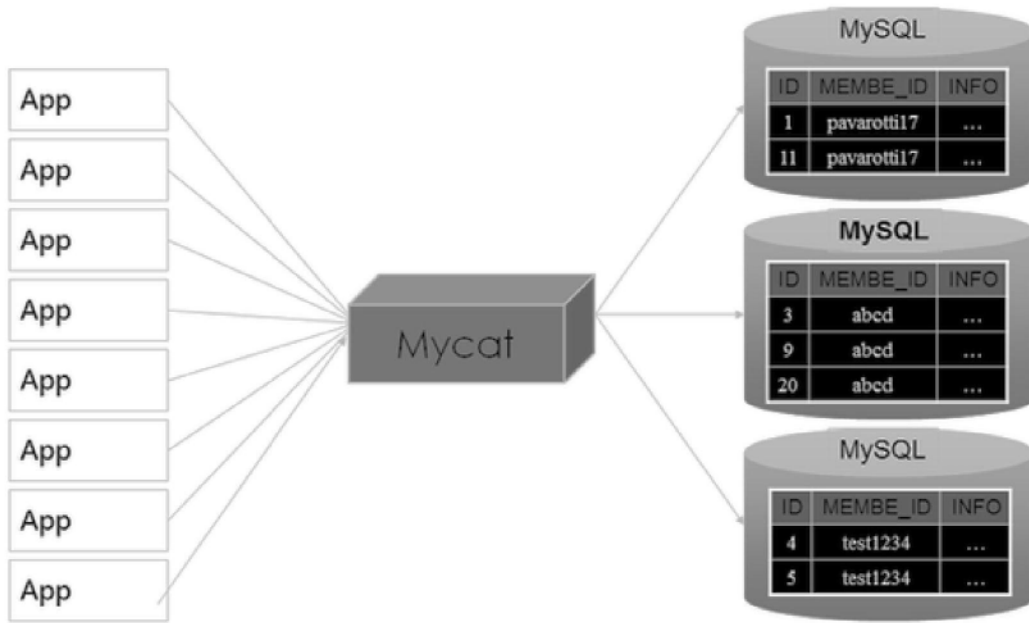


图1

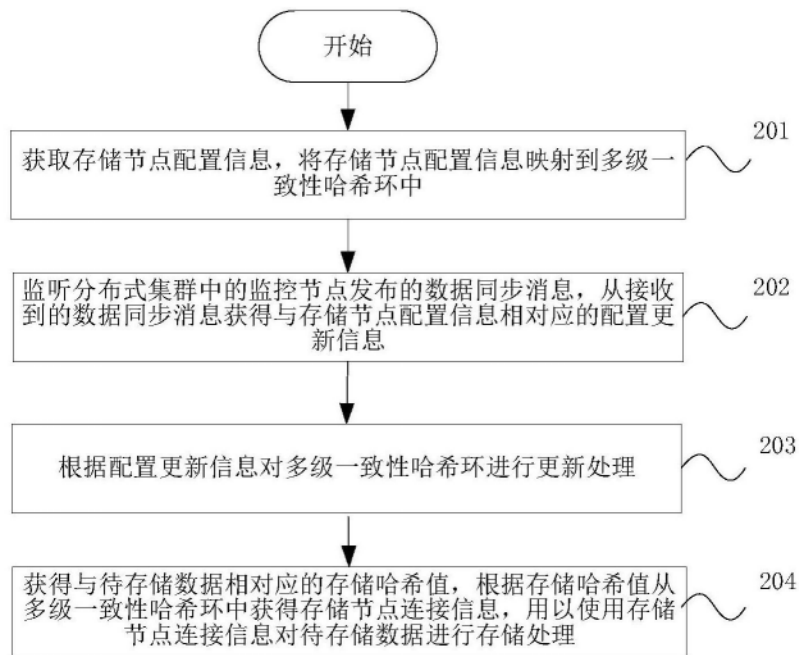


图2

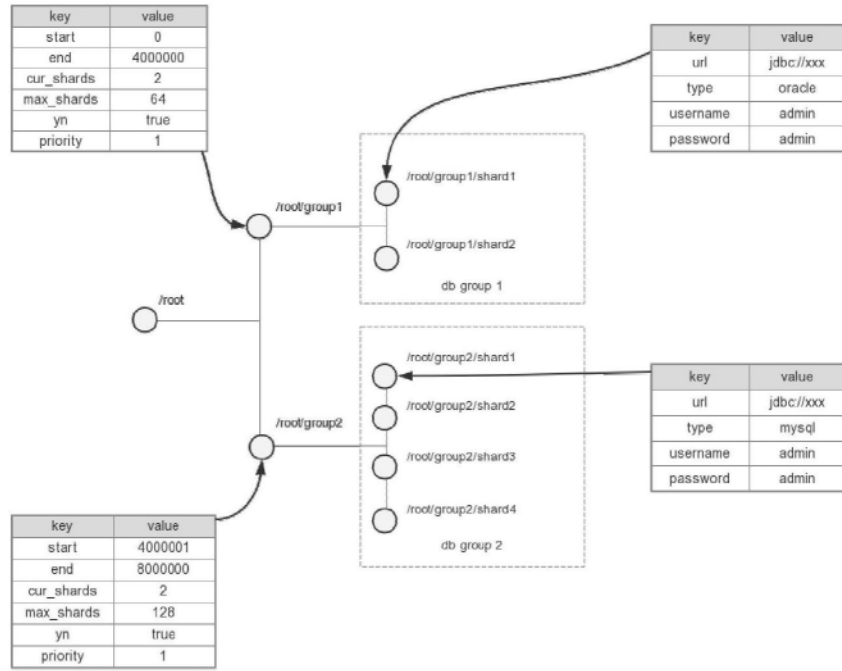


图3

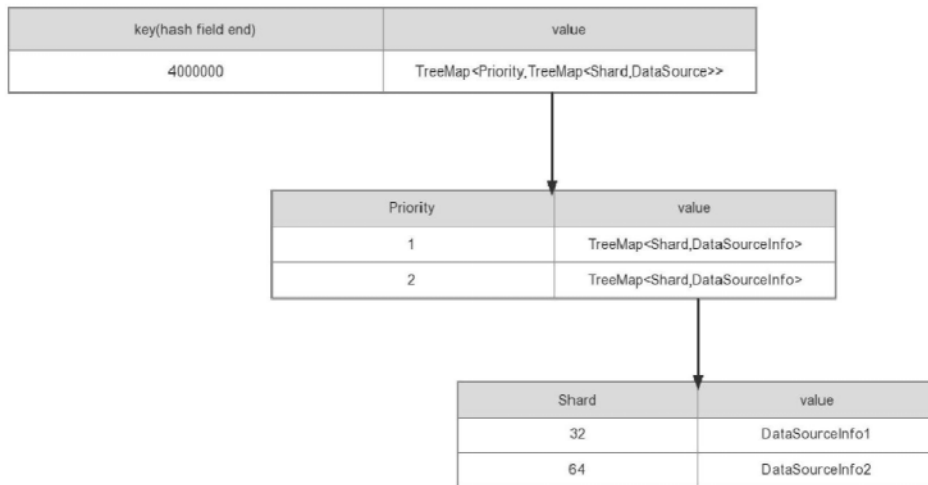


图4

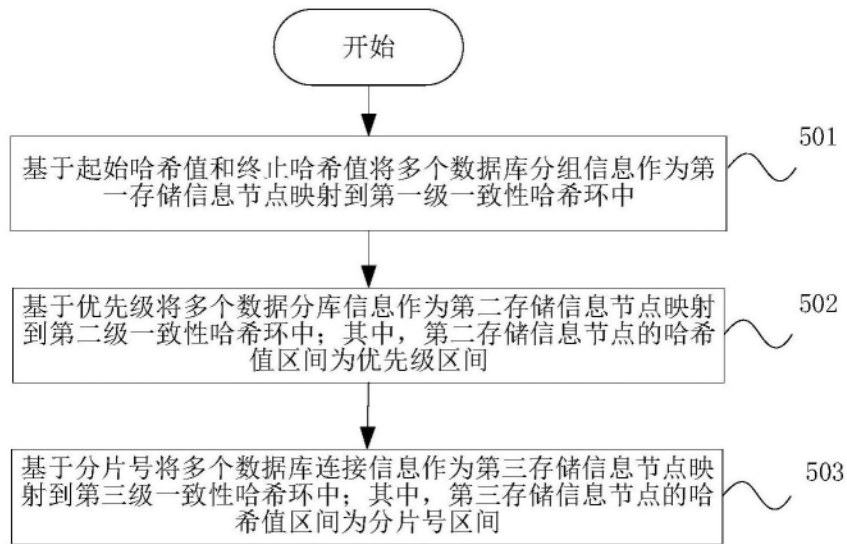


图5

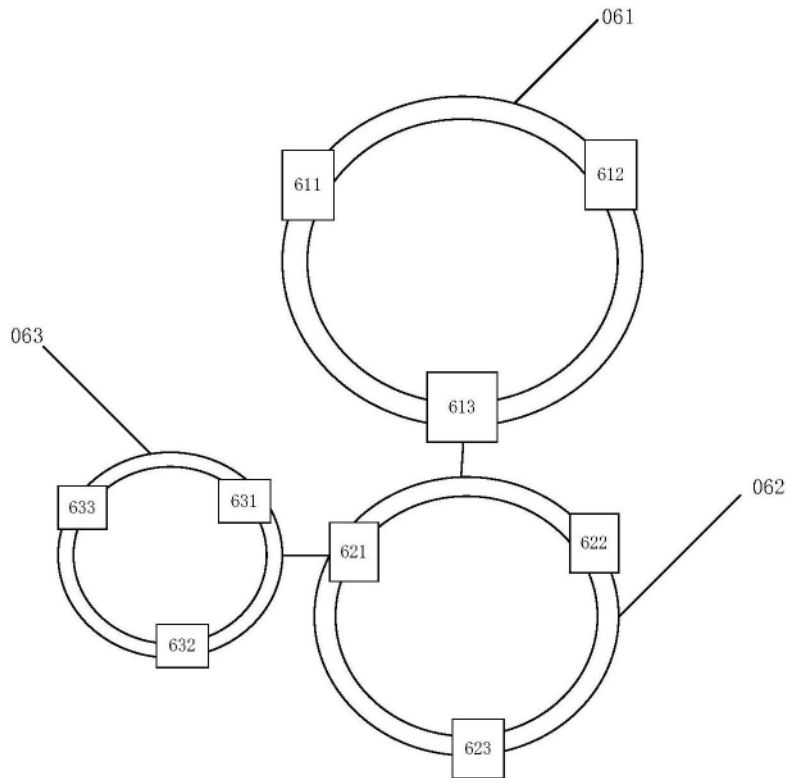


图6

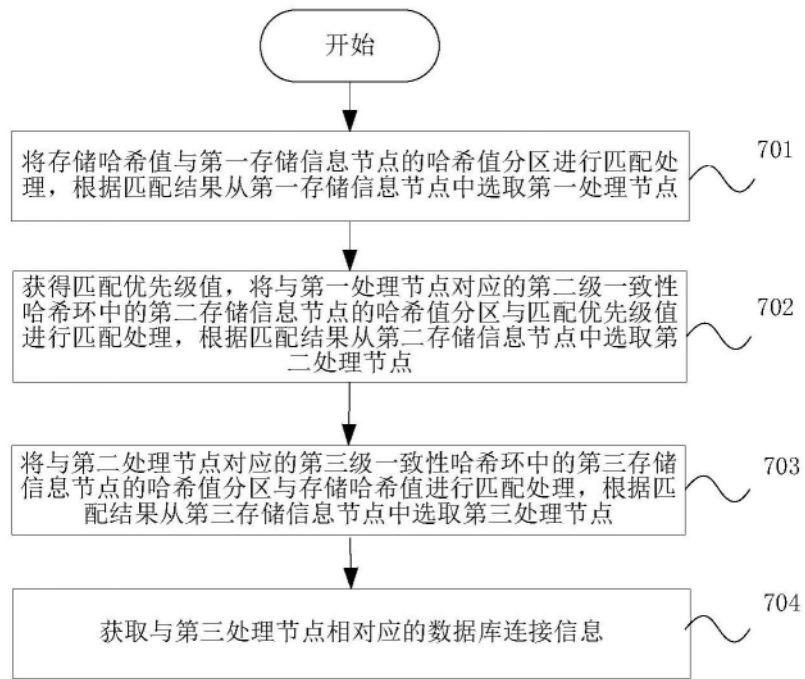


图7

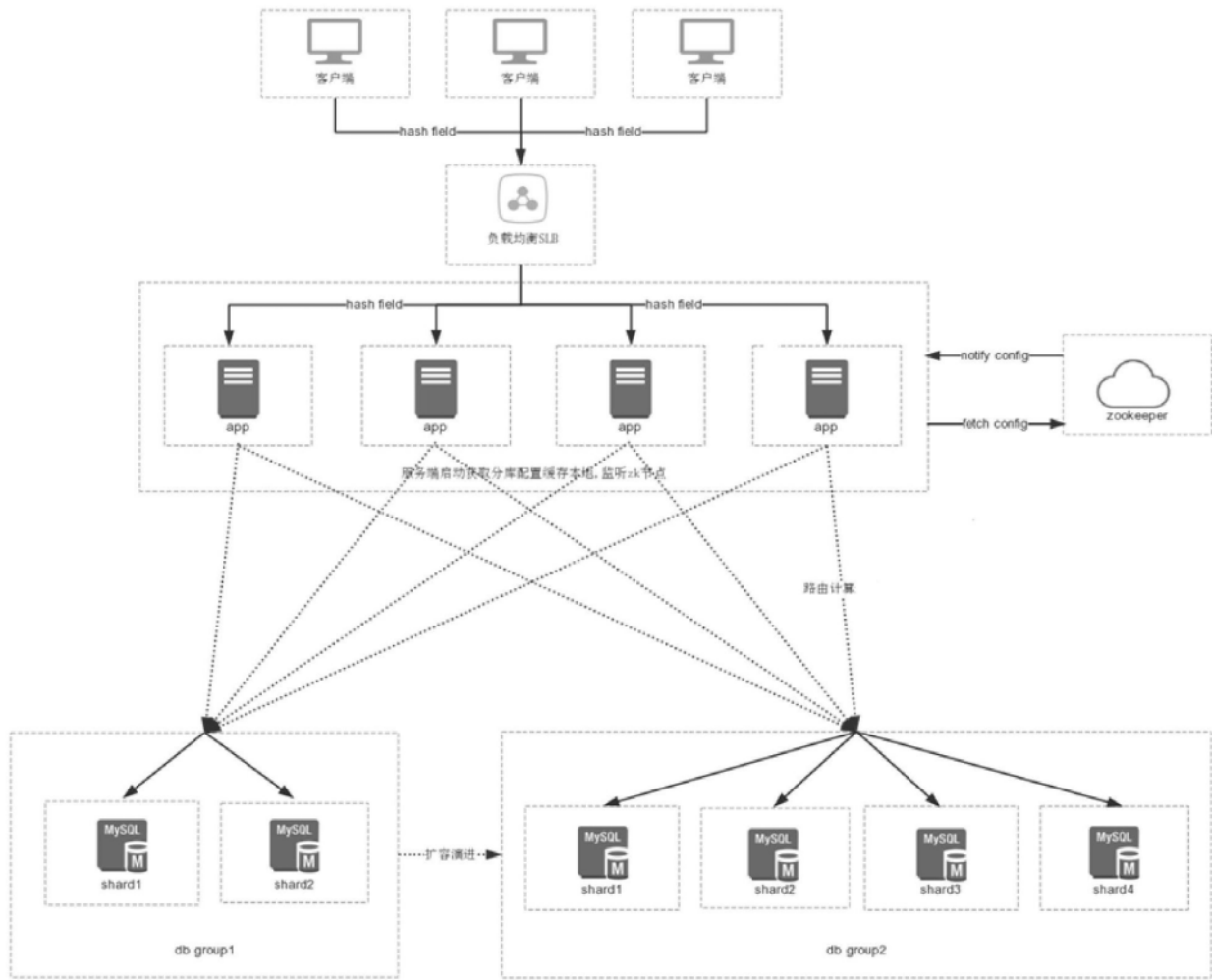


图8

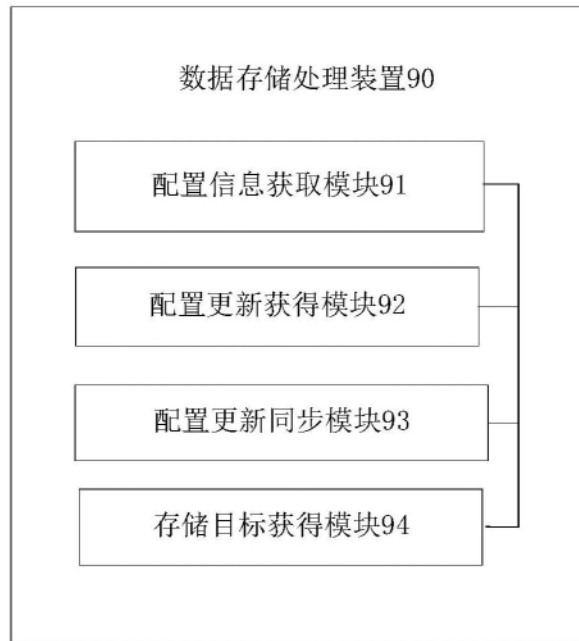


图9

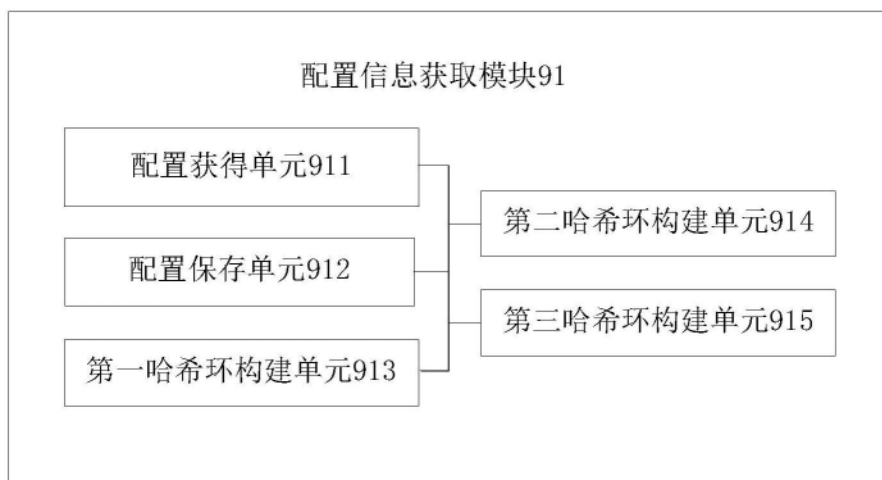


图10

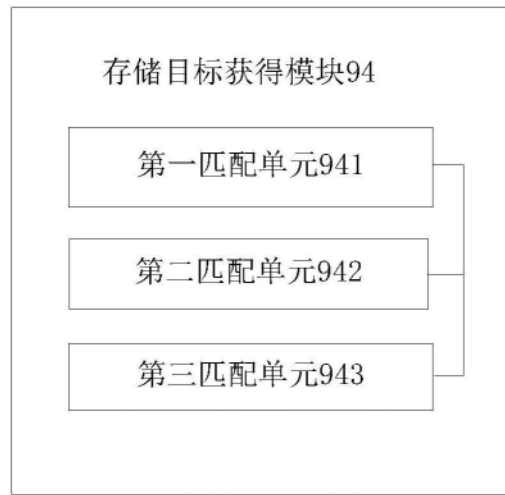


图11

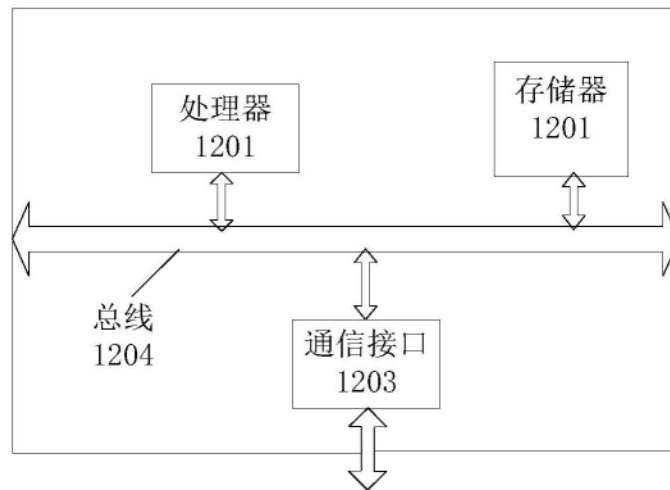


图12