(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 105069152 B
(45)授权公告日 2019.12.13

(21)申请号 201510524829.6

(22)申请日 2015.08.25

(65)同一申请的已公布的文献号
申请公布号 CN 105069152 A

(43)申请公布日 2015.11.18

(73)专利权人 航天恒星科技有限公司
地址 100086 北京市海淀区知春路82号
专利权人 北京航空航天大学

(72)发明人 肖利民 钟巧灵 焦小超 尹勇 霍志胜 阮利 李书攀 臧媛媛 王培

(74)专利代理机构 北京卓恒知识产权代理事务所(特殊普通合伙) 11394
代理人 唐曙晖

(51)Int.Cl.
G06F 16/182(2019.01)
G06F 16/20(2019.01)

审查员 李梦诗

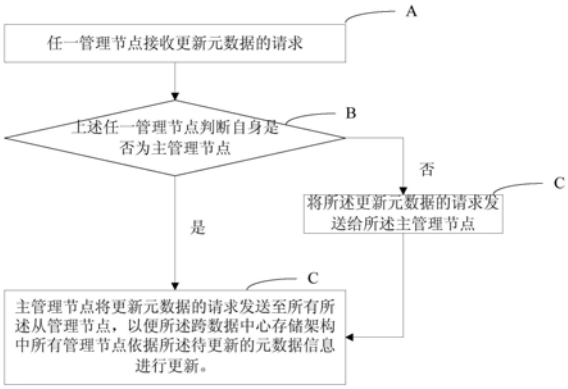
权利要求书3页 说明书13页 附图4页

(54)发明名称

数据处理方法及装置

(57)摘要

本发明实施例提供了一种数据处理方法,数据处理方法包括:A、任一管理节点接收更新元数据的请求;B、任一管理节点判断自身是否为主管理节点,如果是,则进入步骤C,如果否,则将更新元数据的请求发送给主管理节点,进入步骤C;C、主管理节点将更新元数据的请求发送至所有从管理节点,以便跨数据中心存储架构中所有管理节点依据待更新的元数据信息进行更新。采用本发明实施例提供的方法和装置,当需要更新元数据信息时,主管理节点和从管理节点均对其存储的元数据信息作了更新,所以主管理节点和从管理节点存储的元数据信息是一样的,当任一管理节点出现故障时,由于其他管理节点存储的元数据信息相同,所以不会出现现有技术中的空闲时间。



CN 105069152 B

1.一种数据处理方法,其特征在于,应用于跨数据中心存储架构,所述跨数据中心存储架构包括至少两个管理节点,将所述至少两个管理节点中一个管理节点作为主管理节点,其他管理节点作为从管理节点,所述数据处理方法包括:

A、任一所述管理节点接收更新元数据的请求,所述更新元数据的请求包括待更新的元数据信息;

B、所述任一管理节点判断自身是否为主管理节点,如果是,则进入步骤C,如果否,则将所述更新元数据的请求发送给所述主管理节点,进入步骤C;

C、所述主管理节点将所述更新元数据的请求发送至所有所述从管理节点,以便所述跨数据中心存储架构中所有管理节点依据所述待更新的元数据信息进行更新;

从所述跨数据中心存储架构中所有正常运行的管理节点中选取主管理节点的方法,所述跨数据中心存储架构中任一管理节点中存储的各个元数据信息均分配有唯一的事物ID号,且所述事物ID号是随着元数据信息的更新而增大的,对于所述跨数据中心存储架构中任一管理节点,将所述任一管理节点称为第一管理节点,所述方法包括:

D1、设置所述第一管理节点的状态为寻找状态;

D2、将设置第一管理节点为主管理节点的请求发送至所述跨数据中心存储架构中的其他所有管理节点;

D3、向所述跨数据中心存储架构中所有管理节点发送其认可的主管理节点信息;

D4、接收其他管理节点的第二反馈信息,所述第二反馈信息包括管理节点的状态信息、管理节点所认可的主管理节点信息,所述状态信息包括寻找状态信息、主管理节点状态信息、从管理节点状态信息;

D5、当接收到的其他管理节点中第三管理节点的状态信息为寻找状态信息时,比较所述第一管理节点中最大事物ID号与所述第三管理节点中最大事物ID号的大小,当所述第一管理节点的最大事物ID号大于所述第三管理节点的最大事物ID号时,进入步骤D6,否则进入步骤D7;

D6、将所述第一管理节点的状态信息改为主管理节点状态信息,进入步骤D3;

D7、计算认可所述第三管理节点为主管理节点的管理节点的个数,当所述认可所述第三管理节点为主管理节点的管理节点的个数大于等于预设数量时,将所述第一管理节点的状态信息更新为从管理节点状态信息,记录主管理节点信息为所述第三管理节点,结束,当认可所述第三管理节点为主管理节点的管理节点的个数小于所述预设数量时,进入步骤D3;

D8、当接收到的所有所述其他管理节点的所有第二反馈信息为从管理节点状态信息时,更新自身的状态为主管理节点状态,结束;

D9、当接收到的所述其他管理节点中第二管理节点的第二反馈信息为主管理节点状态信息时,更新所述第一管理节点存储的主管理节点信息为所述第二管理节点,结束。

2.根据权利要求1所述数据处理方法,其特征在于,所述步骤C具体包括:

C1、所述主管理节点将所述更新元数据的请求发送至所有所述从管理节点;

C2、所述主管理节点接收所有所述从管理节点的第一反馈信息,所述第一反馈信息包括同意更新信息;

C3、当所述主管理节点接收到的所述第一反馈信息中所述同意更新信息的数量大于预

设值时，向所述跨数据中心存储架构中所有从管理节点发送执行更新元数据信息的命令，以便所述跨数据中心存储架构中所有管理节点依据所述待更新的元数据信息进行更新。

3.根据权利要求2所述数据处理方法，其特征在于，在步骤C3后，还可以包括：

C4、当所述主管理节点接收到的所述第一反馈信息中所述同意更新信息的数量大于等于预设值时，所述主管理节点反馈元数据信息更新成功的消息；

C5、当所述主管理节点接收到的所述第一反馈信息中所述同意更新信息的数量小于预设值时，所述主管理节点反馈元数据信息更新失败的消息。

4.根据权利要求1、2或3所述数据处理方法，其特征在于，还包括获取管理节点中元数据的方法，所述跨数据中心存储架构中所有管理节点之间采用对等互联网技术，所述方法包括：

E1、所述跨数据中心存储架构中任一管理节点接收来自客户端的获取更新元数据的请求，所述获取更新元数据的请求包括元数据事物ID号；

E2、所述任一管理节点将具有所述事物ID号的元数据反馈至所述客户端。

5.一种数据处理装置，其特征在于，应用于跨数据中心存储架构，所述跨数据中心存储架构包括至少两个管理节点，将所述至少两个管理节点中一个管理节点作为主管理节点，其他管理节点作为从管理节点，所述数据处理装置包括：

第一接收模块，用于任一所述管理节点接收更新元数据的请求，所述更新元数据的请求包括待更新的元数据信息；

判断模块，用于所述任一管理节点判断自身是否为主管理节点，如果是，则执行第一发送模块，如果否，则将所述更新元数据的请求发送给所述主管理节点，执行所述第一发送模块；

所述第一发送模块，用于所述主管理节点将所述更新元数据的请求发送至所有所述从管理节点，以便所述跨数据中心存储架构中所有管理节点依据所述待更新的元数据信息进行更新；

所述跨数据中心存储架构中任一管理节点中存储的各个元数据信息均分配有唯一的事物ID号，且事物ID号是随着元数据信息的更新而增大的，所述跨数据中心存储架构中任一管理节点还包括选取主管理节点装置，将所述任一管理节点称为第一管理节点，所述选取主管理节点装置包括：

设置模块，用于设置所述第一管理节点的状态为寻找状态；

第二发送模块，用于将设置第一管理节点为主管理节点的请求发送至所述跨数据中心存储架构中的其他所有管理节点；

第三发送模块，用于向所述跨数据中心存储架构中所有管理节点发送其认可的主管理节点信息；

第二接收模块，用于接收其他管理节点的第二反馈信息，所述第二反馈信息包括管理节点的状态信息、管理节点所认可的主管理节点信息，所述状态信息包括寻找状态信息、主管理节点状态信息、从管理节点状态信息；

比较模块，用于当接收到的其他管理节点中第三管理节点的状态信息为寻找状态信息时，比较所述第一管理节点中最大事物ID号与所述第三管理节点中最大事物ID号的大小，当所述第一管理节点的最大事物ID号大于所述第三管理节点的最大事物ID号时，执行修改

模块，否则，执行计算模块；

所述修改模块，用于将所述第一管理节点的状态信息改为主管理节点状态信息，返回所述第三发送模块；

所述计算模块，用于计算认可所述第三管理节点为主管理节点的管理节点的个数，当所述认可所述第三管理节点为主管理节点的管理节点的个数大于等于预设数量时，将所述第一管理节点的状态信息更新为从管理节点状态信息，记录主管理节点信息为所述第三管理节点，结束，当认可所述第三管理节点为主管理节点的管理节点的个数小于所述预设数量时，返回所述第三发送模块；

第一更新模块，用于当接收到的所有所述其他管理节点的所有第二反馈信息为从管理节点状态信息时，更新自身的状态为主管理节点状态，结束；

第二更新模块，用于当接收到的所述其他管理节点中第二管理节点的第二反馈信息为主管理节点状态信息时，更新所述第一管理节点存储的主管理节点信息为所述第二管理节点，结束。

6.根据权利要求5所述数据处理装置，其特征在于，所述第一发送模块包括：

发送单元，用于所述主管理节点将所述更新元数据的请求发送至所有所述从管理节点；

接收单元，用于所述主管理节点接收所有所述从管理节点的第一反馈信息，所述第一反馈信息包括同意更新信息；

执行单元，用于当所述主管理节点接收到的所述第一反馈信息中所述同意更新信息的数量大于预设值时，向所述跨数据中心存储架构中所有从管理节点发送执行更新元数据信息的命令，以便所述跨数据中心存储架构中所有管理节点依据所述待更新的元数据信息进行更新。

7.根据权利要求6所述数据处理装置，其特征在于，还包括：

第一反馈单元，用于当所述主管理节点接收到的所述第一反馈信息中所述同意更新信息的数量大于等于预设值时，所述主管理节点反馈元数据信息更新成功的消息；

第二反馈单元，用于当所述主管理节点接收到的所述第一反馈信息中所述同意更新信息的数量小于预设值时，所述主管理节点反馈元数据信息更新失败的消息。

8.根据权利要求5、6或7所述数据处理装置，其特征在于，还包括获取元数据装置，所述跨数据中心存储架构中所有管理节点之间采用对等互联网技术，所述获取元数据装置包括：

第三接收模块，用于所述跨数据中心存储架构中任一管理节点接收来自客户端的获取更新元数据的请求，所述获取更新元数据的请求包括元数据事物ID号；

反馈模块，用于所述任一管理节点将具有所述事物ID号的元数据反馈至所述客户端。

数据处理方法及装置

技术领域

[0001] 本发明涉及计算机技术领域，尤其涉及一种数据处理方法及装置。

背景技术

[0002] 随着计算机技术的发展，各领域的数据量呈海量的趋势增长，对数据存储的需求越来越高，对数据的备份要求也随之提高。例如，卫星遥感等应用的数据正由PB级向EB级发展，如何满足EB级数据存储需求已成为大数据存储平台亟待解决的问题。

[0003] 当前多数据中心成为解决大数据挑战的主要方法，满足了数据容量扩充的需求，例如卫星应用、Google、Facebook等公司在全球建立了多个数据中心。针对多数据中心对数据的管理不便，提出了跨数据中心管理架构，例如Google的Spanner管理系统。但跨数据中心管理架构的管理节点的有时会发生故障，例如断电。

[0004] 对此有如下两种应对策略：一、NFS(Network File System，网络文件系统)远程挂载目录保存多个副本，将本地元数据信息同步到远端的文件服务器上，当发生故障时，可以将远端的文件服务器上的信息同步到正常运行的管理节点上；二、主管理节点定时将自己存储的信息发送至备份管理节点，当主管理节点出现故障时，由备份管理节点替代主管理节点。

[0005] 在实现本发明创造的过程中发明人发现在“将远端的文件服务器上的信息同步到正常运行的管理节点”和“由备份管理节点替代主管理节点”过程中均存在一个空闲时间，在该空闲时间内没有管理节点存储该段时间更新的元数据信息。

发明内容

[0006] 有鉴于此，本发明提供了一种数据处理方法及装置，用以解决现有技术中在该空闲时间内没有管理节点在存储该段时间更新或修改的元数据信息的问题，其技术方案如下：

[0007] 一种数据处理方法，应用于跨数据中心存储架构，所述跨数据中心存储架构包括至少两个管理节点，将所述至少两个管理节点中一个管理节点作为主管理节点，其他管理节点作为从管理节点，所述数据处理方法包括：

[0008] A、任一所述管理节点接收更新元数据的请求，所述更新元数据的请求包括待更新的元数据信息；

[0009] B、所述任一管理节点判断自身是否为主管理节点，如果是，则进入步骤C，如果否，则将所述更新元数据的请求发送给所述主管理节点，进入步骤C；

[0010] C、所述主管理节点将所述更新元数据的请求发送至所有所述从管理节点，以便所述跨数据中心存储架构中所有管理节点依据所述待更新的元数据信息进行更新。

[0011] 其中，所述步骤C具体包括：

[0012] C1、所述主管理节点将所述更新元数据的请求发送至所有所述从管理节点；

[0013] C2、所述主管理节点接收所有所述从管理节点的第一反馈信息，所述第一反馈信

息包括同意更新信息；

[0014] C3、当所述主管理节点接收到的所述第一反馈信息中所述同意更信息的数量大于预设值时，向所述跨数据中心存储架构中所有从管理节点发送执行更新元数据信息的命令，以便所述跨数据中心存储架构中所有管理节点依据所述待更新的元数据信息进行更新。

[0015] 其中，在步骤C3后，还可以包括：

[0016] C4、当所述主管理节点接收到的所述第一反馈信息中所述同意更信息的数量大于等于预设值时，所述主管理节点反馈元数据信息更新成功的消息；

[0017] C5、当所述主管理节点接收到的所述第一反馈信息中所述同意更信息的数量小于预设值时，所述主管理节点反馈元数据信息更新失败的消息。

[0018] 其中，还包括：从所述跨数据中心存储架构中所有正常运行的管理节点中选取主管理节点的方法，所述跨数据中心存储架构中任一管理节点中存储的各个元数据信息均分配有唯一的事物ID号，且事物ID号是随着元数据信息的更新而增大的，对于所述跨数据中心存储架构中任一管理节点，将所述任一管理节点称为第一管理节点，所述方法包括：

[0019] D1、设置所述第一管理节点的状态为寻找状态；

[0020] D2、将设置第一管理节点为主管理节点的请求发送至所述跨数据中心存储架构中的其他所有管理节点；

[0021] D3、向所述跨数据中心存储架构中所有管理节点发送其认可的主管理节点信息；

[0022] D4、接收其他管理节点的第二反馈信息，所述第二反馈信息包括管理节点的状态信息、管理节点所认可的主管理节点信息，所述状态信息包括寻找状态信息、主管理节点状态信息、从管理节点状态信息；

[0023] D5、当接收到的其他管理节点中第三管理节点的状态信息为寻找状态信息时，比较所述第一管理节点中最大事物ID号与所述第三管理节点中最大事物ID号的大小，当所述第一管理节点的最大事物ID号大于所述第三管理节点的最大事物ID号时，进入步骤D6，否则进入步骤D7；

[0024] D6、将所述第一管理节点的状态信息改为主管理节点状态信息，进入步骤D3；

[0025] D7、计算认可所述第三管理节点为主管理节点的管理节点的个数，当所述认可所述第三管理节点为主管理节点的管理节点的个数大于等于预设数量时，将所述第一管理节点的状态信息更新为从管理节点状态信息，记录主管理节点信息为所述第三管理节点，结束，当认可所述第三管理节点为主管理节点的管理节点的个数小于所述预设数量时，进入步骤D3；

[0026] D8、当接收到的所有所述其他管理节点的所有第二反馈信息为从管理节点状态信息时，更新自身的状态为主管理节点状态，结束；

[0027] D9、当接收到的所述其他管理节点中第二管理节点的第二反馈信息为主管理节点状态信息时，更新所述第一管理节点存储的主管理节点信息为所述第二管理节点，结束。

[0028] 其中，还包括获取管理节点中元数据的方法，所述跨数据中心存储架构中所有管理节点之间采用对等互联网技术，所述方法包括：

[0029] E1、所述跨数据中心存储架构中任一管理节点接收来自客户端的获取元数据请求，所述获取元数据请求包括元数据事物ID号；

[0030] E2、所述任一管理节点将具有所述事物ID号的元数据反馈至所述客户端。

[0031] 一种数据处理装置，应用于跨数据中心存储架构，所述跨数据中心存储架构包括至少两个管理节点，将所述至少两个管理节点中一个管理节点作为主管理节点，其他管理节点作为从管理节点，所述数据处理装置包括：

[0032] 第一接收模块，用于任一所述管理节点接收更新元数据的请求，所述更新元数据的请求包括待更新的元数据信息；

[0033] 判断模块，用于所述任一管理节点判断自身是否为主管理节点，如果是，则执行第一发送模块，如果否，则将所述更新元数据的请求发送给所述主管理节点，执行所述第一发送模块；

[0034] 所述第一发送模块，用于所述主管理节点将所述更新元数据的请求发送至所有所述从管理节点，以便所述跨数据中心存储架构中所有管理节点依据所述待更新的元数据信息进行更新。

[0035] 其中，所述第一发送模块包括：

[0036] 发送单元，用于所述主管理节点将所述更新元数据的请求发送至所有所述从管理节点；

[0037] 接收单元，用于所述主管理节点接收所有所述从管理节点的第一反馈信息，所述第一反馈信息包括同意更新信息；

[0038] 执行单元，用于当所述主管理节点接收到的所述第一反馈信息中所述同意更信息的数量大于预设值时，向所述跨数据中心存储架构中所有从管理节点发送执行更新元数据信息的命令，以便所述跨数据中心存储架构中所有管理节点依据所述待更新的元数据信息进行更新。

[0039] 其中，还包括：

[0040] 第一反馈单元，用于当所述主管理节点接收到的所述第一反馈信息中所述同意更信息的数量大于等于预设值时，所述主管理节点反馈元数据信息更新成功的消息；

[0041] 第二反馈单元，用于当所述主管理节点接收到的所述第一反馈信息中所述同意更信息的数量小于预设值时，所述主管理节点反馈元数据信息更新失败的消息。

[0042] 其中，所述跨数据中心存储架构中任一管理节点中存储的各个元数据信息均分配有唯一的事物ID号，且事物ID号是随着元数据信息的更新而增大的，所述跨数据中心存储架构中任一管理节点还包括选取主管理节点装置，将所述任一管理节点称为第一管理节点，所述选取主管理节点装置包括：

[0043] 设置模块，用于设置所述第一管理节点的状态为寻找状态；

[0044] 第二发送模块，用于将设置第一管理节点为主管理节点的请求发送至所述跨数据中心存储架构中的其他所有管理节点；

[0045] 第三发送模块，用于向所述跨数据中心存储架构中所有管理节点发送其认可的主管理节点信息；

[0046] 第二接收模块，用于接收其他管理节点的第二反馈信息，所述第二反馈信息包括管理节点的状态信息、管理节点所认可的主管理节点信息，所述状态信息包括寻找状态信息、主管理节点状态信息、从管理节点状态信息；

[0047] 比较模块，用于当接收到的其他管理节点中第三管理节点的状态信息为寻找状态

信息时，比较所述第一管理节点中最大事物ID号与所述第三管理节点中最大事物ID号的大小，当所述第一管理节点的最大事物ID号大于所述第三管理节点的最大事物ID号时，执行修改模块，否则，执行计算模块；

[0048] 所述修改模块，用于将所述第一管理节点的状态信息改为主管理节点状态信息，返回所述第三发送模块；

[0049] 所述计算模块，用于计算认可所述第三管理节点为主管理节点的管理节点的个数，当所述认可所述第三管理节点为主管理节点的管理节点的个数大于等于预设数量时，将所述第一管理节点的状态信息更新为从管理节点状态信息，记录主管理节点信息为所述第三管理节点，结束，当认可所述第三管理节点为主管理节点的管理节点的个数小于所述预设数量时，返回所述第三发送模块；

[0050] 第一更新模块，用于当接收到的所有所述其他管理节点的所有第二反馈信息为从管理节点状态信息时，更新自身的状态为主管理节点状态，结束；

[0051] 第二更新模块，用于当接收到的所述其他管理节点中第二管理节点的第二反馈信息为主管理节点状态信息时，更新所述第一管理节点存储的主管理节点信息为所述第二管理节点，结束。

[0052] 其中，还包括获取元数据装置，所述跨数据中心存储架构中所有管理节点之间采用对等互联网技术，所述获取元数据装置包括：

[0053] 第三接收模块，用于所述跨数据中心存储架构中任一管理节点接收来自客户端的获取元数据请求，所述获取元数据请求包括元数据事物ID号；

[0054] 反馈模块，用于所述任一管理节点将具有所述事物ID号的元数据反馈至所述客户端。

[0055] 上述技术方案具有如下有益效果：当需要更新元数据信息时，主管理节点和从管理节点均对其存储的元数据信息作了更新，所以主管理节点和从管理节点存储的元数据信息是一样的，当任一管理节点出现故障时，由于其他管理节点存储的元数据信息相同，所以不会出现现有技术中的空闲时间，从而有效的提高了跨数据中心存储系统中存储服务的可用性。

附图说明

[0056] 图1为本发明实施例提供的一种数据处理方法的一种实现方式的方法流程示意图；

[0057] 图2为本发明实施例提供的一种数据处理方法的中步骤C的一种实现方式的方法流程示意图；

[0058] 图3为本发明实施例提供的一种数据处理方法中从所述跨数据中心存储架构中所有正常运行的管理节点中选取主管理节点的方法的一种实现方式的方法流程示意图；

[0059] 图4为本发明实施例提供的一种数据处理方法中获取管理节点中元数据的方法的一种实现方式的方法流程示意图；

[0060] 图5为本发明实施例提供的一种数据处理装置的一种实现方式的结构示意图；

[0061] 图6为本发明实施例提供的一种数据处理装置中第一发送模块的一种实现方式的结构示意图；

[0062] 图7为本发明实施例提供的一种数据处理装置中选取主管理节点装置的一种实现方式的结构示意图；

[0063] 图8为本发明实施例提供的一种数据处理装置中获取元数据装置的一种实现方式的结构示意图。

具体实施方式

[0064] 为了引用和清楚起见，下文中使用的技术名词的说明、简写或缩写总结如下：

[0065] GFS:Google File System，分布式文件系统；

[0066] HDFS:Hadoop Distributed File System，Hadoop分布式文件系统；

[0067] CRC:Cyclic Redundancy Check，循环冗余校验码；

[0068] NFS:Network File System，网络文件系统；

[0069] TFS:Taobao FileSystem，面向互联网的分布式文件系统；

[0070] Namenode:名字节点。

[0071] 下面将结合本发明实施例中的附图，对本发明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发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，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，都属于本发明保护的范围。

[0072] 随着计算机技术的发展，各领域的数据量呈海量的趋势增长，对数据存储的需求越来越高，对数据的备份要求也随之提高。例如，卫星遥感等应用的数据正由PB级向EB级发展，如何满足EB级数据存储需求已成为大数据存储平台亟待解决的问题。当前多数据中心成为解决大数据挑战的主要方法，满足了数据容量扩充的需求，例如卫星应用、Google、Facebook等公司在全球建立了多个数据中心。针对多数据中心对数据的管理不便，提出了跨数据中心管理架构，例如Google的Spanner管理系统。跨数据中心管理架构的管理节点的单个管理节点故障等瓶颈无法满足EB级数据存储的需求。目前分布式存储系统和多数据中心存储是满足EB级数据存储需求的关键的技术途径。跨数据中心的统一分布式存储系统能够有效解决应用环境中数据分布式存储和管理问题，系统通过广域网互联不同地域的分布式存储系统，提供EB级存储架构服务。

[0073] 在当前的分布式环境下，系统故障，尤其是针对于低廉的服务器来讲，是比较正常的情况，这其中包括软件、硬件等故障，因此需要系统故障恢复和保障方法来保证系统的运行正常。典型的分布式文件系统GFS、HDFS采用了具体的报障措施，比如对于数据块进行做CRC校验和计算来保证数据块数据的完整和正确，这是针对于数据进行的故障保护机制；除了数据的故障保护机制，HDFS采用了NameNode元数据备份、NameNode主备切换的思想来保障NameNode这个管理节点宕机后系统的正常服务。正常运行的NameNode通过NFS远程挂载目录保存多个副本把本地元数据信息同步到远端的文件服务器上保证元数据信息的多副本正常运行，但是主NameNode宕机后切换至备份的NameNode启动之间有停止服务时间间隔。HDFS新的高可用方案借助第三方系统ZooKeeper来保证HDFS中NameNode的高可用，两个NameNode，一个为激活状态(active)，一个用于备份(standby)，在正常运行情况下standby的NameNode不提供服务，这样系统standby的NameNode的资源没有得到充分的利用。TFS是淘宝针对于电商需求的网络小图片海量存储需求的分布式文件系统，TFS采用基于Linux的

heartbeat机制来保证名字节点的高可用方案,同样实现的是主NameNode发送heartbeat信息给从NameNode,一旦超时没收到主NameNode的heartbeat信息,则从NameNode接管整个元数据管理的服务。Hadoop生态圈中的分布式数据库系统HBase(分布式的、面向列的开源数据库)采用同样的也是基于ZooKeeper来实现HMaster的高可用方案,在HBase中可以拥有多个HMaster,但是只能有一个HMaster作为服务节点,多个HMaster同时竞争成为主HMaster。

[0074] 采用CRC校验码,只能保证数据块的有效性,并不能保证管理节点的有效性。使用NFS进行元数据的挂载保存多个副本,然后同步元数据到本地重新启动管理节点,这种冷备切换的思路中间有一段服务停止时间间隔。新型的基于heartbeat机制以及基于ZooKeeper机制实现的master节点的热备切换思路,也具有服务停止时间间隔。

[0075] 服务停止时间间隔即在“将远端的文件服务器上的信息同步到正常运行的管理节点”和“由备份管理节点替代主管理节点”过程中均存在一个空闲时间,在该空闲时间内没有管理节点在存储该段时间更新或修改的元数据信息。

[0076] 请参阅图1,为本发明实施例提供的一种数据处理方法的一种实现方式的方法流程示意图,该数据处理方法应用于跨数据中心存储架构,所述跨数据中心存储架构包括至少两个管理节点,将所述至少两个管理节点中一个管理节点作为主管理节点,其他管理节点作为从管理节点,所述数据处理方法包括:

[0077] 步骤A:任一管理节点接收更新元数据的请求。

[0078] 更新元数据的请求包括待更新的元数据信息。上述更新元数据信息可以为修改元数据信息、增加元数据信息、删除元数据信息等等。元数据信息又称中介数据、中继数据,为描述数据的数据(data about data),主要是描述数据属性(property)的信息,用来支持如指示存储位置、历史数据、资源查找、文件记录等功能。

[0079] 跨数据中心存储架构中的任一管理节点均可以接收更新元数据的请求,不论该管理节点为主管理节点还是从管理节点。

[0080] 为了实现主管理节点和从管理节点均能够接收更新元数据的请求,则可以设置主管理节点和从管理节点之间为peer to peer的结构等等,本发明实施例并不对主管理节点和从管理节点之间的结构进行限定,主管理节点和从管理节点之间的结构关系可以是能实现“主管理节点和从管理节点均能够接收更新元数据的请求”的任一结构。

[0081] 触发跨数据中心存储架构生成“更新元数据信息请求”的可以是检测到管理节点之间具有不同的元数据信息的情况,也可以是生成了一个元数据信息、删除元数据信息或修改元数据信息,例如新建WORD后,有WORD的建立时间,该建立时间为新增加的元数据信息。

[0082] 步骤B:上述任一管理节点判断自身是否为主管理节点,如果是,则进入步骤C,如果否,则将所述更新元数据的请求发送给所述主管理节点,进入步骤C。

[0083] 上述任一管理节点是指接收到“待更新的元数据信息”的管理节点。

[0084] 本发明实施例中的管理节点可以管理着多个分布式文件系统集群例如HDFS集群,也包括分布式数据库系统集群,例如HBase集群,这样管理节点可以支持分布式文件系统和分布式数据库系统nosql的操作请求。由于分布式文件系统集群和分布式数据库系统集群,都支持集群级别的扩容机制,所以本发明实施例中的跨数据中心存储架构可以实现空间动态的平滑增长和减小,可以使用多个集群数据服务器来进行数据后端的存储。

[0085] 步骤C:主管理节点将更新元数据的请求发送至所有所述从管理节点,以便所述跨数据中心存储架构中所有管理节点依据所述待更新的元数据信息进行更新。

[0086] 主管理节点可以与从管理节点同时更新元数据信息,也可以比从管理节点先一步更新元数据信息,对此本发明实施例不做具体限制。

[0087] 通过设置多个管理节点来保证多个元数据的备份,提高跨数据中存储架构的运行的可靠性。

[0088] 本发明实施例提供了一种数据处理方法,当需要更新元数据信息时,主管理节点和从管理节点均对其存储的元数据信息作了更新,所以主管理节点和从管理节点存储的元数据信息是一样的,当任一管理节点出现故障时,由于其他管理节点存储的元数据信息相同,所以不会出现现有技术中的空闲时间。

[0089] 请参阅图2,为本发明实施例提供的一种数据处理方法的中步骤C的一种实现方式的方法流程示意图,该方法包括:

[0090] C1、所述主管理节点将所述更新元数据的请求发送至所有所述从管理节点。

[0091] C2、所述主管理节点接收所有所述从管理节点的第一反馈信息。

[0092] 所述第一反馈信息包括同意更新信息。

[0093] C3、当主管理节点接收到的所述第一反馈信息中所述同意更信息的数量大于预设值时,向所述跨数据中心存储架构中所有从管理节点发送执行更新元数据信息的命令,以便所述跨数据中心存储架构中所有管理节点依据所述待更新的元数据信息进行更新。

[0094] 上述步骤C的具体方法可以为paxos算法,paxos算法是一个保证多个副本一致性的算法,通过上述方法来保证多个管理节点上的元数据信息同步。

[0095] 一般情况下,从管理节点接收到更新元数据请求时,都会同意更新,此处判断“同意更信息的数量大于预设值”是考虑了时间延迟,因为有的从管理节点将第一反馈信息发送至主管理节点需要的时间较长,为了更加快速的节省时间,只要主管理节点收到的同意更新请求信息超过预设值,则向所述跨数据中心存储架构中所有从管理节点发送执行更新元数据信息的命令。

[0096] 此处判断“同意更新信息的数量大于预设值”还考虑到从管理节点出现故障,无法及时反馈同意信息的情况。

[0097] 上述预设值可以为跨数据中心存储架构中正常运行的管理节点的总数量的一半以上,例如假设跨数据中心存储架构中正常运行的管理节点的总数量为9,预设值可以为5、6、7、8中的任一值。

[0098] 第一反馈信息还可以包括不同意更新信息,例如管理节点检测到自身已经更新了此次需要更新的元数据信息,那么可以拒绝此次更新元数据信息的请求。

[0099] 上述任一数据处理方法中在步骤C3后,还可以包括:C4、当主管理节点接收到的所述第一反馈信息中所述同意更信息的数量大于等于预设值时,所述主管理节点反馈元数据信息更新成功的消息;C5、当主管理节点接收到的所述第一反馈信息中所述同意更信息的数量小于预设值时,所述主管理节点反馈元数据信息更新失败的消息。

[0100] 请参阅图3,为本发明实施例提供的一种数据处理方法中从所述跨数据中心存储架构中所有正常运行的管理节点中选取主管理节点的方法的一种实现方式的方法流程示意图,跨数据中心存储架构中任一管理节点中存储的各个元数据信息均分配有唯一的事物

ID号，且事物ID号是随着元数据信息的更新而增大的，对于所述跨数据中心存储架构中任一管理节点，将所述任一管理节点成为第一管理节点，所述方法包括：

[0101] D1、设置所述第一管理节点的状态为寻找状态。

[0102] 管理节点进行元数据更新操作的时候，采用事物号ID特性来进行每次元数据更新操作管理。例如当前第一管理节点的事物ID号为5，当在第一管理节点新增加一元数据信息时，该新的元数据信息的事物ID为6，所以随着第一管理节点中元数据的不断更新，其存储的元数据的事物ID号越来越大。

[0103] D2、将设置第一管理节点为主管理节点的请求发送至所述跨数据中心存储架构中的其他所有管理节点。

[0104] 在最初时，第一管理节点推荐自己为主管理节点。

[0105] D3、向所述跨数据中心存储架构中所有管理节点发送其认可的主管理节点信息。

[0106] D4、接收其他管理节点的第二反馈信息，所述第二反馈信息包括管理节点的状态信息、管理节点所认可的主管理节点信息。

[0107] 所述状态信息包括寻找状态信息、主管理节点状态信息、从管理节点状态信息。

[0108] D5、当接收到的其他管理节点中第三管理节点的状态信息为寻找状态信息时，比较所述第一管理节点中最大事物ID号与所述第三管理节点中最大事物ID号的大小，当所述第一管理节点的最大事物ID号大于所述第三管理节点的最大事物ID号时，进入步骤D6，否则，进入步骤D7。

[0109] 第三管理节点为跨数据中心存储架构中除第一管理节点以外的任一管理节点，这里为了与第一管理节点进行区分，所以称为第三管理节点。

[0110] D6、将所述第一管理节点的状态信息改为主管理节点状态信息，进入步骤D3。

[0111] D7、计算认可所述第三管理节点为主管理节点的管理节点的个数，当认可所述第三管理节点为主管理节点的管理节点的个数大于等于预设数量时，将所述第一管理节点的状态信息更新为从管理节点状态信息，记录主管理节点信息为所述第三管理节点，结束，当认可所述第三管理节点为主管理节点的管理节点的个数小于预设数量时，进入步骤D3。

[0112] 预设数量可以为于跨数据中心存储架构中正常运行的管理节点的总数量的一半以上，例如假设跨数据中心存储架构中正常运行的管理节点的总数量为9，预设数量可以为5、6、7、8中任一值。

[0113] D8、当接收到的所有所述其他管理节点的所有第二反馈信息为从管理节点状态信息时，更新自身的状态为主管理节点状态，结束。

[0114] D9、当接收到的所述其他管理节点中第二管理节点的第二反馈信息为主管理节点状态信息时，更新所述第一管理节点存储的主管理节点信息为所述第二管理节点，结束。

[0115] 上述步骤D1至D9的主语均为第一管理节点，即适用于跨数据中心存储架构中的任一管理节点。

[0116] 本发明实施例提供的数据处理方法中采用了主管理节点提出重新选取主管理节点的方式，避免了上述选取主管理节点的方法中的活锁问题。活锁指的是任务或者执行者没有被阻塞，由于某些条件没有满足，导致一直重复尝试，失败，尝试，失败。

[0117] 现有技术中作为备份的管理节点并不能充分发挥它本身拥有的价值，即同一时刻主管理节点和从管理节点不能同时服务于客户端。为此本发明提供以下方法。

[0118] 请参阅图4，为本发明实施例提供的一种数据处理方法中获取管理节点中元数据的方法的一种实现方式的方法流程示意图，所述跨数据中心存储架构中所有管理节点之间采用对等互联网技术，该方法包括：

[0119] E1、所述跨数据中心存储架构中任一管理节点接收来自客户端的获取元数据请求，所述获取元数据请求包括元数据事物ID号。

[0120] E2、所述任一管理节点将具有所述事物ID号的元数据反馈至所述客户端。

[0121] 本发明实施例采用了对等互联网技术，这样不论是主管理节点还是从管理节点都可以正常服务请求，这样在客户端可以选择其中一个管理节点进行访问，充分利用多个管理节点的资源。每一个管理节点都可以服务来自客户端的请求，如果是元数据的读请求，则管理节点直接响应回复请求；如果是元数据更新请求，则从管理节点将该请求发往主管理节点。

[0122] 本发明实施例可以用在软件上，此时操作系统可以为Linux系统，运行在Linux机群中提供文件IO服务的软件之上，如HDFS等分布式文件系统和HBase等NoSQL分布式数据库系统，并且HDFS分布式文件系统配置多个DataNode。

[0123] 请参阅图5，为本发明实施例提供的一种数据处理装置的一种实现方式的结构示意图，该数据处理装置应用于跨数据中心存储架构，所述跨数据中心存储架构包括至少两个管理节点，将所述至少两个管理节点中一个管理节点作为主管理节点，其他管理节点作为从管理节点，所述数据处理装置包括：第一接收模块501、判断模块502、第一发送模块503，其中：

[0124] 第一接收模501，用于任一所述管理节点接收更新元数据的请求。

[0125] 更新元数据的请求包括待更新的元数据信息。上述更新元数据信息可以为修改元数据信息、增加元数据信息、删除元数据信息等等。元数据信息又称中介数据、中继数据，为描述数据的数据(data about data)，主要是描述数据属性(property)的信息，用来支持如指示存储位置、历史数据、资源查找、文件记录等功能。

[0126] 跨数据中心存储架构包括至少两个管理节点，具体可以为大于等于3个管理节点。

[0127] 跨数据中心存储架构中的任一管理节点均可以接收更新元数据的请求，不论该管理节点为主管理节点还是从管理节点。

[0128] 为了实现主管理节点和从管理节点均能够接收更新元数据的请求，则可以设置主管理节点和从管理节点之间为peer-to-peer的结构等等，本发明实施例并不对主管理节点和从管理节点之间的结构进行限定，主管理节点和从管理节点之间的结构关系可以是能实现“主管理节点和从管理节点均能够接收更新元数据的请求”的任一结构。

[0129] 触发跨数据中心存储架构生成“更新元数据信息请求”的可以是检测到管理节点之间具有不同的元数据信息的情况，也可以是生成了一个元数据信息、删除元数据信息或修改元数据信息，例如新建WORD后，有WORD的建立时间，该建立时间为新增加的元数据信息。

[0130] 判断模块502，用于所述任一管理节点判断自身是否为主管理节点，如果是，则执行第一发送模块503，如果否，则将所述更新元数据的请求发送给所述主管理节点，执行所述第一发送模块503。

[0131] 上述任一管理节点是指接收到“待更新的元数据信息”的管理节点。

[0132] 本发明实施例中的管理节点可以管理着多个分布式文件系统集群例如HDFS集群，也包括分布式数据库系统集群，例如HBase集群，这样管理节点可以支持分布式文件系统和分布式数据库系统NoSQL的操作请求。由于分布式文件系统集群和分布式数据库系统集群，都支持集群级别的扩容机制，所以本发明实施例中的跨数据中心存储架构可以实现空间动态的平滑增长和减小，可以使用多个集群数据服务器来进行数据后端的存储。

[0133] 所述第一发送模块503，用于所述主管理节点将所述更新元数据的请求发送至所有所述从管理节点，以便所述跨数据中心存储架构中所有管理节点依据所述待更新的元数据信息进行更新。

[0134] 主管理节点可以与从管理节点同时更新元数据信息，也可以比从管理节点先一步更新元数据信息，对此本发明实施例不做具体限制。

[0135] 通过设置多个管理节点来保证多个元数据的备份，提高跨数据中存储架构的运行的可靠性。

[0136] 本发明实施例提供了一种数据处理装置，当需要更新元数据信息时，主管理节点和从管理节点均对其存储的元数据信息作了更新，所以主管理节点和从管理节点存储的元数据信息是一样的，当任一管理节点出现故障时，由于其他管理节点存储的元数据信息相同，所以不会出现现有技术中的空闲时间。

[0137] 请参阅图6，为本发明实施例提供的一种数据处理装置中第一发送模块的一种实现方式的结构示意图，第一发送模块可以包括：发送单元601、接收单元602、执行单元603，其中：

[0138] 发送单元601，用于所述主管理节点将所述更新元数据的请求发送至所有所述从管理节点。

[0139] 接收单元602，用于所述主管理节点接收所有所述从管理节点的第一反馈信息，所述第一反馈信息包括同意更新信息。

[0140] 执行单元603，用于当主管理节点接收到的所述第一反馈信息中所述同意更信息的数量大于预设值时，向所述跨数据中心存储架构中所有从管理节点发送执行更新元数据信息的命令，以便所述跨数据中心存储架构中所有管理节点依据所述待更新的元数据信息进行更新。

[0141] 上述第一发送模块的具体实现方式可以为paxos算法，paxos算法是一个保证多个副本一致性的算法，通过上述方法来保证多个管理节点上的元数据信息同步。

[0142] 一般情况下，从管理节点接收到更新元数据请求时，都会同意更新，此处判断“同意更信息的数量大于预设值”是考虑了时间延迟，因为有的从管理节点将第一反馈信息发送至主管理节点需要的时间较长，为了更加快速的节省时间，只要主管理节点收到的同意更新请求信息超过预设值，则向所述跨数据中心存储架构中所有从管理节点发送执行更新元数据信息的命令。

[0143] 此处判断“同意更新信息的数量大于预设值”还考虑到从管理节点出现故障，无法及时反馈同意信息的情况。

[0144] 上述预设值可以为跨数据中心存储架构中正常运行的管理节点的总数量的一半以上，例如假设跨数据中心存储架构中正常运行的管理节点的总数量为9，预设值可以为5、6、7、8中的任一值。

[0145] 第一反馈信息还可以包括不同意更新信息，例如管理节点检测到自身已经更新了此次需要更新的元数据信息，那么可以拒绝此次更新元数据信息的请求。

[0146] 上述任一数据处理装置还可以包括：第一反馈单元，用于当主管理节点接收到的所述第一反馈信息中所述同意更信息的数量大于等于预设值时，所述主管理节点反馈元数据信息更新成功的消息；第二反馈单元，用于当主管理节点接收到的所述第一反馈信息中所述同意更信息的数量小于预设值时，所述主管理节点反馈元数据信息更新失败的消息。

[0147] 客户端为触发此次“更新元数据请求”的客户端。

[0148] 请参阅图7，为本发明实施例提供的一种数据处理装置中选取主管理节点装置的一种实现方式的结构示意图，所述跨数据中心存储架构中任一管理节点中存储的各个元数据信息均分配有一事物ID号，且事物ID号是随着元数据信息的更新而增大的，所述跨数据中心存储架构中任一管理节点还包括选取主管理节点装置，将所述任一管理节点称为第一管理节点，所述选取主管理节点装置包括：设置模块701、第二发送模块702、第三发送模块703、第二接收模块704、比较模块705、修改模块706、计算模块707、第一更新模块708、第二更新模块709，其中：

[0149] 设置模块701，用于设置所述第一管理节点的状态为寻找状态。

[0150] 管理节点进行元数据更新操作的时候，采用事物号ID特性来进行每次元数据更新操作管理。例如当前第一管理节点的事物ID号为5，当在第一管理节点新增加一元数据信息时，该新的元数据信息的事物ID为6，所以随着第一管理节点中元数据的不断更新，其存储的元数据的事物ID号越来越大。

[0151] 第二发送模块702，用于将设置第一管理节点为主管理节点的请求发送至所述跨数据中心存储架构中的其他所有管理节点。

[0152] 第三发送模块703，用于向所述跨数据中心存储架构中所有管理节点发送其认可的主管理节点信息。

[0153] 在最初时，第一管理节点推荐自己为主管理节点。

[0154] 第二接收模块704，用于接收其他管理节点的第二反馈信息，所述第二反馈信息包括管理节点的状态信息、管理节点所认可的主管理节点信息。

[0155] 所述状态信息包括寻找状态信息、主管理节点状态信息、从管理节点状态信息。

[0156] 比较模块705，用于当接收到的其他管理节点中第三管理节点的状态信息为寻找状态信息时，比较所述第一管理节点中最大事物ID号与所述第三管理节点所存储的最大事物ID号的大小，当所述第一管理节点的最大事物ID号大于所述第三管理节点中最大事物ID号时，执行修改模块706，当所述第一管理节点的最大事物ID号小于所述第三管理节点的最大事物ID号时，执行计算模块707。

[0157] 第三管理节点为跨数据中心存储架构中除第一管理节点以外的任一管理节点，这里为了与第一管理节点进行区分，所以称为第三管理节点。所述修改模块706，用于将所述第一管理节点的状态信息改为主管理节点状态信息，返回所述第三发送模块703。

[0158] 所述计算模块707，用于计算认可所述第三管理节点为主管理节点的管理节点的个数，当认可所述第三管理节点为主管理节点的管理节点的个数大于等于预设数量时，将所述第一管理节点的状态信息更新为从管理节点状态信息，记录主管理节点信息为所述第三管理节点，结束，当认可所述第三管理节点为主管理节点的管理节点的个数小于预设数

量时，返回所述第三发送模块703。

[0159] 预设数量可以为于跨数据中心存储架构中正常运行的管理节点的总数量的一半以上，例如假设跨数据中心存储架构中正常运行的管理节点的总数量为9，预设数量可以为5、6、7、8中任一值。

[0160] 第一更新模块708，用于当接收到的所有所述其他管理节点的所有第二反馈信息为从管理节点状态信息时，更新自身的状态为主管理节点状态，结束。

[0161] 第二更新模块709，用于当接收到的所述其他管理节点中第二管理节点的第二反馈信息为主管理节点状态信息时，更新所述第一管理节点存储的主管理节点信息为所述第二管理节点，结束。

[0162] 本发明实施例提供的数据处理方法中采用了主管理节点提出重新选取主管理节点的方式，避免了上述选取主管理节点的方法中的活锁问题。活锁指的是任务或者执行者没有被阻塞，由于某些条件没有满足，导致一直重复尝试，失败，尝试，失败。

[0163] 上述设置模块701、第二发送模块702、第三发送模块703、第二接收模块704、比较模块705、修改模块706、计算模块707、第一更新模块708、第二更新模块709均属于第一管理节点，即适用于跨数据中心存储架构中的任一管理节点。

[0164] 本发明实施例提供的数据处理方法中采用了主管理节点提出重新选取主管理节点的方式，避免了上述选取主管理节点的方法中的活锁问题。活锁指的是任务或者执行者没有被阻塞，由于某些条件没有满足，导致一直重复尝试，失败，尝试，失败。

[0165] 现有技术中作为备份的管理节点并不能充分发挥它本身拥有的价值，即同一时刻主管理节点和从管理节点不能同时服务于客户端。为此本发明提供以下装置。

[0166] 请参阅图8，为本发明实施例提供的一种数据处理装置中获取元数据装置的一种实现方式的结构示意图，所述跨数据中心存储架构中所有管理节点之间采用对等互联网技术，获取元数据装置包括：

[0167] 第三接收模块801，用于所述跨数据中心存储架构中任一管理节点接收来自客户端的获取元数据请求，所述获取元数据请求包括元数据事物ID号。

[0168] 反馈模块802，用于所述任一管理节点将具有所述事物ID号的元数据反馈至所述客户端。

[0169] 本发明实施例采用了对等互联网技术，这样不论是主管理节点还是从管理节点都可以正常服务请求，这样在客户端可以选择其中一个管理节点进行访问，充分利用多个管理节点的资源。每一个管理节点都可以服务来自客户端的请求，如果是元数据的读请求，则管理节点直接响应回复请求；如果是元数据更新请求，则有管理节点发往主管理节点。

[0170] 本发明实施例可以用在软件上，此时操作系统可以为Linux系统，运行在Linux机群中提供文件IO服务的软件之上，如HDFS等分布式文件系统和HBase等nosql分布式数据库系统，并且HDFS分布式文件系统配置多个datanode。

[0171] 本说明书中各个实施例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部分互相参见即可。

[0172] 对所提供的实施例的上述说明，使本领域专业技术人员能够实现或使用本发明。对这些实施例的多种修改对本领域的专业技术人员来说将是显而易见的，本文中所定义的一般原理可以在不脱离本发明的精神或范围的情况下，在其它实施例中实现。因此，本发明

将不会被限制于本文所示的这些实施例，而是要符合与本文所提供的原理和新颖特点相一致的最宽的范围。

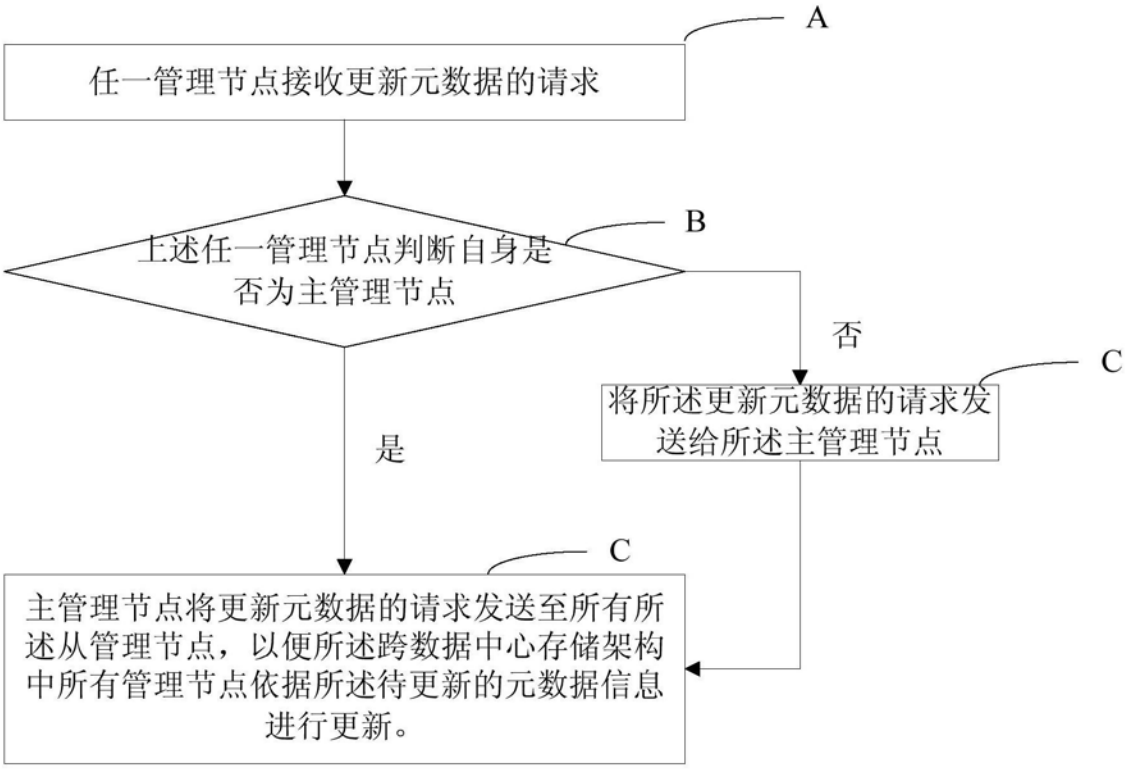

A
任一管理节点接收更新元数据的请求
B
上述任一管理节点判断自身是否为主管理节点
否
C
将所述更新元数据的请求发送给所述主管理节点
是
C
主管理节点将更新元数据的请求发送至所有所述从管理节点，以便所述跨数据中心存储架构中所有管理节点依据所述待更新的元数据信息进行更新。

图1

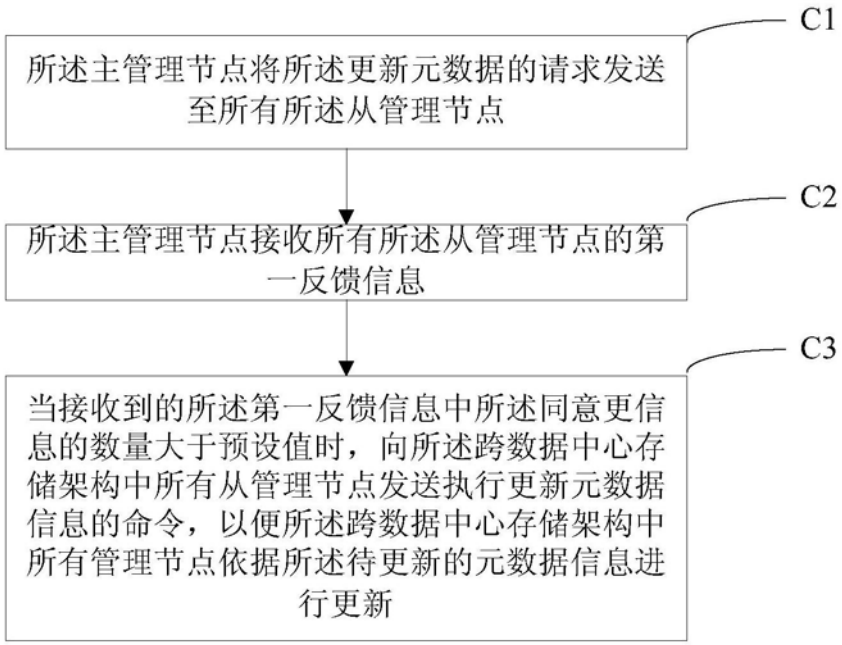

C1
所述主管理节点将所述更新元数据的请求发送至所有所述从管理节点
C2
所述主管理节点接收所有所述从管理节点的第一反馈信息
C3
当接收到的所述第一反馈信息中所述同意更信息的数量大于预设值时，向所述跨数据中心存储架构中所有从管理节点发送执行更新元数据信息的命令，以便所述跨数据中心存储架构中所有管理节点依据所述待更新的元数据信息进行更新

图2

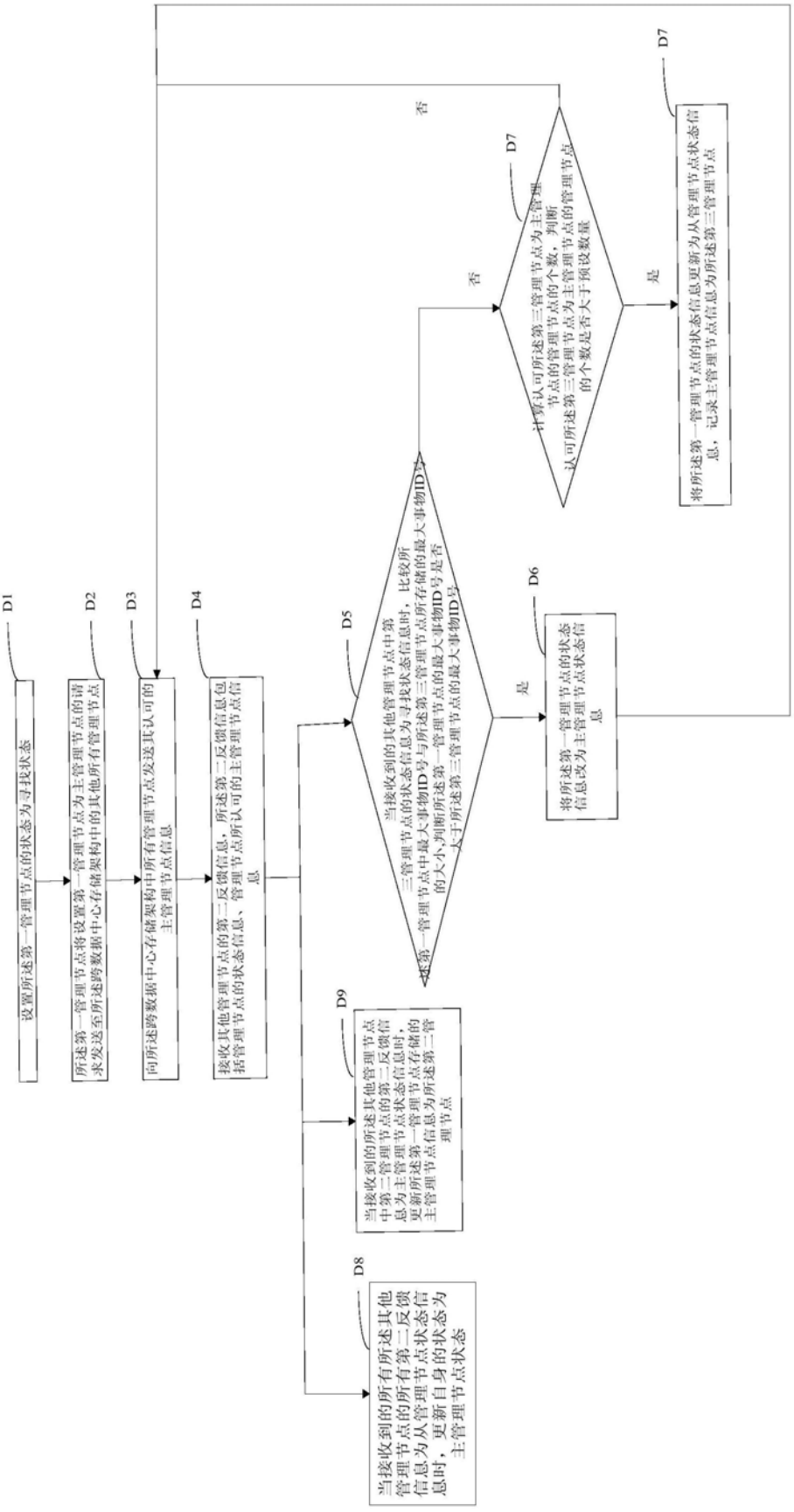

D1
设置所述第一管理节点的状态为寻找状态
D2
所述第一管理节点将设置第一管理节点为主管理节点的请求发送至所述跨数据中心存储架构中的其他所有管理节点
D3
向所述跨数据中心存储架构中所有管理节点发送其认可的主管理节点信息
D4
接收其他管理节点的第二反馈信息，所述第二反馈信息包括管理节点的状态信息、管理节点所认可的主管理节点信息
D5
当接收到的其他管理节点中第三管理节点的状态信息为寻找状态信息时，比较所述第一管理节点中最大事物ID号与所述第三管理节点所存储的最大事物ID号的大小，判断所述第一管理节点的最大事物ID号是否大于所述第三管理节点的最大事物ID号
否
是
D6
将所述第一管理节点的状态信息改为主管理节点状态信息
D7
计算认可所述第三管理节点为主管理节点的管理节点的个数，判断认可所述第三管理节点为主管理节点的管理节点的个数是否大于预设数量
否
是
D7
将所述第一管理节点的状态信息更新为从管理节点状态信息，记录主管理节点信息为所述第三管理节点
D8
当接收到的所有所述其他管理节点的所有第二反馈信息为从管理节点状态信息时，更新自身的状态为主管理节点状态
D9
当接收到的所述其他管理节点中第三管理节点的第二反馈信息为主管理节点状态信息时，更新所述第一管理节点存储的主管理节点信息为所述第三管理节点

图3

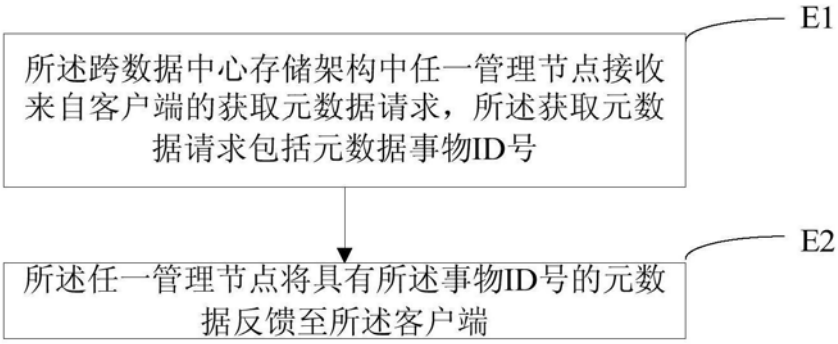

E1
所述跨数据中心存储架构中任一管理节点接收来自客户端的获取元数据请求，所述获取元数据请求包括元数据事物ID号
E2
所述任一管理节点将具有所述事物ID号的元数据反馈至所述客户端

图4

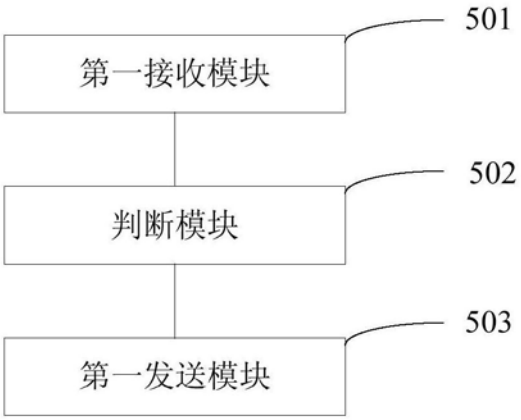

501
第一接收模块
502
判断模块
503
第一发送模块

图5

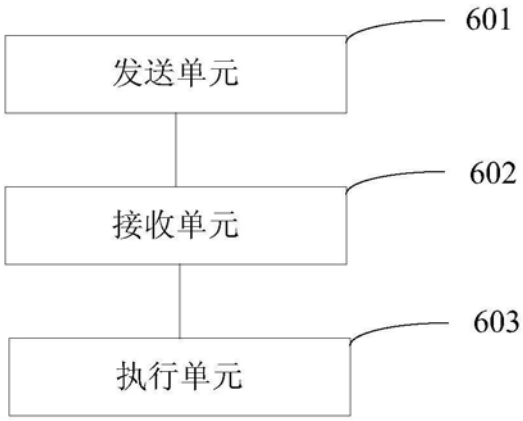

601
发送单元
602
接收单元
603
执行单元

图6

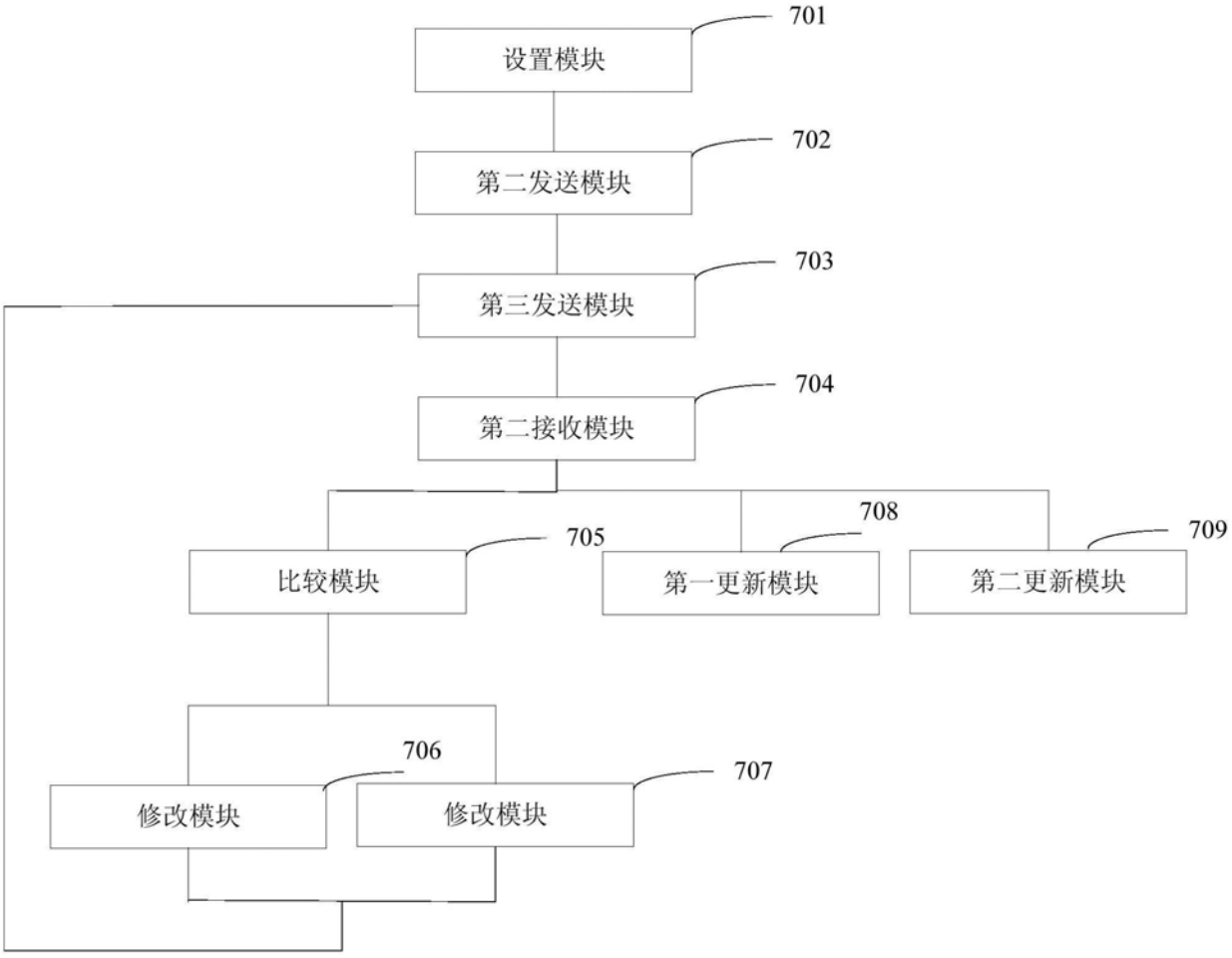

701
设置模块
702
第二发送模块
703
第三发送模块
704
第二接收模块
708
709
705
比较模块
第一更新模块
第二更新模块
706
707
修改模块
修改模块

图7

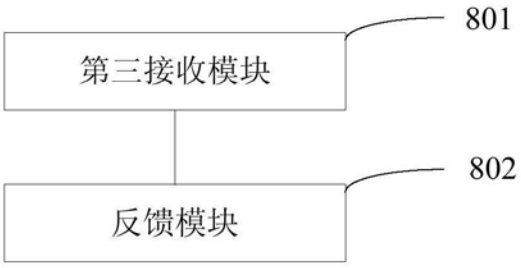

801
第三接收模块
802
反馈模块

图8