



(12) 发明专利申请

(10) 申请公布号 CN 115268785 A

(43) 申请公布日 2022. 11. 01

(21) 申请号 202210891870.7

(22) 申请日 2022.07.27

(71) 申请人 浙江大华技术股份有限公司

地址 310053 浙江省杭州市滨江区滨安路
1187号

(72) 发明人 许克尧 罗心 江文龙 周明伟

(74) 专利代理机构 北京同达信恒知识产权代理
有限公司 11291

专利代理师 赵凯莉

(51) Int.Cl.

G06F 3/06 (2006.01)

G06F 11/20 (2006.01)

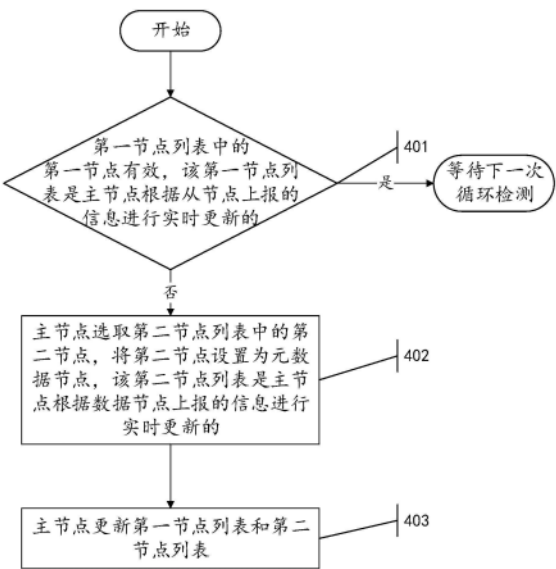
权利要求书2页 说明书9页 附图3页

(54) 发明名称

一种应用于分布式存储系统的管理方法、装置及存储介质

(57) 摘要

本申请公开了一种应用于分布式存储系统的管理方法、装置及存储介质,涉及分布式数据存储技术领域。该方法包括:主节点根据从节点按照设定时间上报的信息更新第一节点列表,所述第一节点列表中包括至少一个元数据节点的地址、上报时间以及元数据服务状态;所述主节点根据所述数据节点按照设定时间上报的信息更新第二节点列表,所述第二节点列表中包括至少一个数据节点的地址以及上报时间;若所述主节点确定所述第一节点列表中的第一节点失效,则选取所述第二节点列表中的第二节点,将所述第二节点设置为元数据节点,并更新所述第一节点列表和所述第二节点列表。



1. 一种应用于分布式存储系统的管理方法,所述分布式存储系统包括元数据节点和数据节点,所述元数据节点包括主节点和从节点,其特征在于,所述方法包括:

所述主节点根据所述从节点按照设定时间上报的信息更新第一节点列表,所述第一节点列表中包括至少一个元数据节点的地址、上报时间以及元数据服务状态;

所述主节点根据所述数据节点按照设定时间上报的信息更新第二节点列表,所述第二节点列表中包括至少一个数据节点的地址以及上报时间;

若所述主节点确定所述第一节点列表中的第一节点失效,则选取所述第二节点列表中的第二节点,将所述第二节点设置为元数据节点,并更新所述第一节点列表和所述第二节点列表。

2. 如权利要求1所述的方法,其特征在于,所述主节点确定所述第一节点列表中的第一节点失效,包括:

若所述第一节点满足以下条件中的至少一项,则确定所述第一节点失效:

所述第一节点的上报时间距当前时间的时间间隔大于第一时长阈值;

所述第一节点的元数据服务状态为异常。

3. 如权利要求1所述的方法,其特征在于,所述选取所述第二节点列表中的第二节点,包括:

若所述第二节点的上报时间距当前时间的时间间隔小于或等于第二时长阈值,则确定所述第二节点有效,并选取所述第二节点。

4. 如权利要求1所述的方法,其特征在于,所述更新所述第一节点列表和所述第二节点列表,包括:

所述主节点将所述第一节点标记为失效,或者将所述第一节点的地址、上报时间以及元数据服务状态从所述第一节点列表中删除,并将所述第二节点的地址以及当前时间添加到所述第一节点列表中;

所述主节点将所述第二节点列表中的所述第二节点的地址以及上报时间删除。

5. 如权利要求1-4任一项所述的方法,其特征在于,所述方法还包括:

若所述主节点确定所述第二节点列表中存在上报时间距当前时间的时间间隔大于所述第二时长阈值的数据节点,则确定所述数据节点失效,并将所述数据节点的地址以及上报时间从所述第二节点列表中删除,或者将所述数据节点标记为失效。

6. 如权利要求1-4任一项所述的方法,其特征在于,所述选取所述第二节点列表中的第二节点,包括:

若所述主节点确定所述分布式存储系统中的元数据节点的数量不满足可用性要求,则选取所述第二节点列表中的第二节点。

7. 一种应用于分布式存储系统的管理装置,其特征在于,包括:

第一节点列表更新模块,用于根据从节点上报的信息更新第一节点列表,所述第一节点列表中包括至少一个元数据节点的地址、上报时间以及元数据服务状态;

第二节点列表更新模块,用于根据数据节点上报的信息更新第二节点列表,所述第二节点列表中包括至少一个数据节点的地址以及上报时间;

节点管理模块,用于若确定所述第一节点列表中的第一节点失效,则选取所述第二节点列表中的第二节点,将所述第二节点设置为元数据节点,并指示所述第一节点列表更新

模块更新所述第一节点列表,指示所述第二节点列表更新模块更新所述第二节点列表。

8.如权利要求7所述的装置,其特征在于,所述节点管理模块,具体用于:

若所述第一节点满足以下条件中的至少一项,则确定所述第一节点失效:

所述第一节点的上报时间距当前时间的时间间隔大于第一时长阈值;

所述第一节点的元数据服务状态为异常。

9.如权利要求7所述的装置,其特征在于,所述节点管理模块,具体用于:

若所述第二节点的上报时间距当前时间的时间间隔小于或等于第二时长阈值,则确定所述第二节点有效,并选取所述第二节点。

10.如权利要求7所述的装置,其特征在于,所述节点管理模块,具体用于:

指示所述第一节点列表更新模块将所述第一节点标记为失效,或者将所述第一节点的地址、上报时间以及元数据服务状态从所述第一节点列表中删除,并指示所述第二节点列表更新模块将所述第二节点的地址以及当前时间添加到所述第一节点列表中;以及指示所述第二节点列表更新模块将所述第二节点列表中的所述第二节点的地址以及上报时间删除。

11.如权利要求7-10任一项所述的装置,其特征在于,所述装置还包括数据节点有效检测块,用于:

若确定所述第二节点列表中存在上报时间距当前时间的时间间隔大于所述第二时长阈值的数据节点,则确定所述数据节点失效,并指示所述第二节点列表更新模块将所述数据节点的地址以及上报时间从所述第二节点列表中删除,或者将所述数据节点标记为失效。

12.如权利要求7-10任一项所述的装置,其特征在于,所述节点管理模块,具体用于:

若确定所述分布式存储系统中的元数据节点的数量不满足可用性要求,则选取所述第二节点列表中的第二节点。

13.一种通信装置,其特征在于,包括:处理器、存储器;

所述存储器,存储计算机指令;

所述处理器,用于读取所述计算机指令,执行如权利要求1-6任一项所述的方法。

14.一种可读计算机存储介质,其特征在于,所述计算机可读存储介质存储有计算机可执行指令,所述计算机可执行指令用于使计算机执行如权利要求1-6任一项所述的方法。

15.一种计算机程序产品,其特征在于,所述计算机程序产品在被计算机调用时,使得所述计算机执行如权利要求1-6任一项所述的方法。

一种应用于分布式存储系统的管理方法、装置及存储介质

技术领域

[0001] 本申请涉及分布式存储技术领域,尤其涉及一种应用于分布式存储系统的管理方法、装置及存储介质。

背景技术

[0002] 随着大数据时代的发展,对数据的安全性、可靠性、数据服务的可用性要求越来越高,在分布式存储的场景下,元数据服务的可用性是指当多个元数据节点中的有部分元数据节点出现故障时,依旧可以提供元数据服务的能力。

[0003] 目前,为了避免一个或多个元数据节点出现故障影响整个业务的正常运作,一般会在元数据集群内指定一定数量的节点,各节点的元数据服务以主备的形式工作;当元数据的主节点出现故障时,便进行主备切换,便可继续提供元数据服务。但是随着出现故障节点的数量增多,其最大容忍的故障节点数量会降低,可用性也随之降低。例如,如果有 n 个元数据节点,最多允许 $n-1$ 次元数据节点故障,当元数据节点故障有 m 个($m < n-1$)后,最大可容忍故障次数会下降为 $n-1-m$,从而降低元数据集群的可用性。

[0004] 因此,如何提高元数据集群的可用性,是目前亟需解决的问题。

发明内容

[0005] 本申请提供一种应用于分布式存储系统的管理方法,用以提高元数据集群的可用性。

[0006] 第一方面,提供一种应用于分布式存储系统的管理方法,所述分布式存储系统包括元数据节点和数据节点,所述元数据节点包括主节点和从节点,所述方法包括:

[0007] 所述主节点根据所述从节点按照设定时间上报的信息更新第一节点列表,所述第一节点列表中包括至少一个元数据节点的地址、上报时间以及元数据服务状态;所述主节点根据所述数据节点按照设定时间上报的信息更新第二节点列表,所述第二节点列表中包括至少一个数据节点的地址以及上报时间;若所述主节点确定所述第一节点列表中的第一节点失效,则选取所述第二节点列表中的第二节点,将所述第二节点设置为元数据节点,并更新所述第一节点列表和所述第二节点列表。

[0008] 在一种可能实现的方式中,所述主节点确定所述第一节点列表中的第一节点失效,包括:

[0009] 若所述第一节点满足以下条件中的至少一项,则确定所述第一节点失效:所述第一节点的上报时间距当前时间的时间间隔大于第一时长阈值;所述第一节点的元数据服务状态为异常。

[0010] 在一种可能实现的方式中,所述选取所述第二节点列表中的第二节点,包括:

[0011] 若所述第二节点的上报时间距当前时间的时间间隔小于或等于第二时长阈值,则确定所述第二节点有效,并选取所述第二节点。

[0012] 在一种可能实现的方式中,所述更新所述第一节点列表和所述第二节点列表,包

括：

[0013] 所述主节点将所述第一节点标记为失效,或者将所述第一节点的地址、上报时间以及元数据服务状态从所述第一节点列表中删除,并将所述第二节点的地址以及当前时间添加到所述第一节点列表中;所述主节点将所述第二节点列表中的所述第二节点的地址以及上报时间删除。

[0014] 在一种可能实现的方式中,所述方法还包括：

[0015] 若所述主节点确定所述第二节点列表中存在上报时间距当前时间的时间间隔大于所述第二时长阈值的数据节点,则确定所述数据节点失效,并将所述数据节点的地址以及上报时间从所述第二节点列表中删除,或者将所述数据节点标记为失效。

[0016] 在一种可能实现的方式中,所述选取所述第二节点列表中的第二节点,包括：

[0017] 若所述主节点确定所述元数据分布式系统中的元数据节点的数量不满足可用性要求,则选取所述第二节点列表中的第二节点。

[0018] 第二方面,提供一种应用于分布式存储系统的管理装置,包括：

[0019] 第一节点列表更新模块,用于根据从节点上报的信息更新第一节点列表,所述第一节点列表中包括至少一个元数据节点的地址、上报时间以及元数据服务状态;第二节点列表更新模块,用于根据数据节点上报的信息更新第二节点列表,所述第二节点列表中包括至少一个数据节点的地址以及上报时间;节点管理模块,用于若确定所述第一节点列表中的第一节点失效,则选取所述第二节点列表中的第二节点,将所述第二节点设置为元数据节点,并指示所述第一节点列表更新模块更新所述第一节点列表,指示所述第二节点列表更新模块更新所述第二节点列表。

[0020] 在一种可能实现的方式中,所述节点管理模块,具体用于：

[0021] 若所述第一节点满足以下条件中的至少一项,则确定所述第一节点失效:所述第一节点的上报时间距当前时间的时间间隔大于第一时长阈值;所述第一节点的元数据服务状态为异常。

[0022] 在一种可能实现的方式中,所述节点管理模块,具体用于：

[0023] 若所述第二节点的上报时间距当前时间的时间间隔小于或等于第二时长阈值,则确定所述第二节点有效,并选取所述第二节点。

[0024] 在一种可能实现的方式中,所述节点管理模块,具体用于：

[0025] 指示所述第一节点列表更新模块将所述第一节点标记为失效,或者将所述第一节点的地址、上报时间以及元数据服务状态从所述第一节点列表中删除,并指示所述第二节点列表更新模块将所述第二节点的地址以及当前时间添加到所述第一节点列表中;以及指示所述第二节点列表更新模块将所述第二节点列表中的所述第二节点的地址以及上报时间删除。

[0026] 在一种可能实现的方式中,所述装置还包括数据节点有效检测块,用于：

[0027] 若确定所述第二节点列表中存在上报时间距当前时间的时间间隔大于所述第二时长阈值的数据节点,则确定所述数据节点失效,并指示所述第二节点列表更新模块将所述数据节点的地址以及上报时间从所述第二节点列表中删除,或者将所述数据节点标记为失效。

[0028] 在一种可能实现的方式中,所述节点管理模块,具体用于：

[0029] 若确定所述元数据分布式系统中的元数据节点的数量不满足可用性要求,则选取所述第二节点列表中的第二节点。

[0030] 第三方面,提供一种通信装置,包括:

[0031] 处理器、存储器;所述存储器,存储计算机指令;所述处理器,用于读取所述计算机指令,执行如第一方面任一项所述的方法。

[0032] 第四方面,提供一种可读计算机存储介质,所述计算机可读存储介质存储有计算机可执行指令,所述计算机可执行指令用于使计算机执行如第一方面任一项所述的方法。

[0033] 第五方面,提供一种计算机程序产品,所述计算机程序产品在被计算机调用时,使得所述计算机执行如第一方面任一项所述的方法。

[0034] 本申请实施例中,由于主节点可根据从节点按照设定时间上报的信息更新第一节点列表,主节点还可根据数据节点按照设定时间上报的信息更新第二节点列表;当主节点确定出第一节点列表中的第一节点失效时,可选取第二节点列表中的第二节点,并将第二节点设置为元数据节点,便可提供元数据服务,因此可以保证提供元数据服务的元数据节点的数量保持不变,从而提高了元数据集群的可用性。

附图说明

[0035] 图1为本申请实施例提供的一种分布式存储系统架构图;

[0036] 图2为本申请实施例提供的一种基于分布式存储系统的第一节点列表更新方法的流程图;

[0037] 图3为本申请实施例提供的一种基于分布式存储系统的第二节点列表更新方法的流程图;

[0038] 图4为本申请实施例提供的一种应用于分布式存储系统的管理方法的流程图;

[0039] 图5为本申请实施例提供的一种应用于分布式存储系统的管理装置的结构示意图;

[0040] 图6为本申请实施例提供的一种通信装置的结构示意图。

具体实施方式

[0041] 为了使本申请的目的、技术方案和优点更加清楚,下面将结合附图对本申请作进一步地详细描述,显然,所描述的实施例仅仅是本申请一部分实施例,而不是全部的实施例。基于本申请中的实施例,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它实施例,都属于本申请保护的范围。

[0042] 以下对本申请实施例中的部分用语进行解释说明,以便于本领域技术人员理解。

[0043] (1) 本申请实施例中,名词“网络”和“系统”经常交替使用,但本领域的技术人员可以理解其含义。

[0044] (2) 本申请实施例中术语“多个”是指两个或两个以上,其它量词与之类似。

[0045] (3) “和/或”,描述关联对象的关联关系,表示可以存在三种关系,例如,A和/或B,可以表示:单独存在A,同时存在A和B,单独存在B这三种情况。字符“/”一般表示前后关联对象是一种“或”的关系。

[0046] (4) 元数据集群是指存在元数据服务的节点所成的集群。

[0047] 本申请实施例中,分布式存储系统包括元数据节点和数据节点,元数据节点是指存在元数据服务、可存储元数据的节点,数据节点是指未启动或未设置有元数据服务、可存储业务数据的节点,该元数据节点包括主节点和从节点。为了区分元数据节点和数据节点,可预先设置一个IP地址列表,该IP地址列表用于存储各元数据节点的地址。

[0048] 图1为本申请实施例提供的一种分布式存储系统架构图。如图所示,该分布式存储系统100主要由主节点10、从节点(1,2,……,N)、数据节点(1,2,……,M)组成。其中,主节点10与从节点(1,2,……,N)为元数据节点,组成元数据集群,主节点10可以分别与从节点(1,2,……,N)进行有线通信,主节点10可以与数据节点(1,2,……,M)进行有线通信,从节点(1,2,……,N)可与数据节点(1,2,……,M)进行有线通信。

[0049] 从节点(1,2,……,N)可将各自的信息实时上报给主节点10。

[0050] 数据节点(1,2,……,M)可将各自的信息实时上报给主节点10。

[0051] 主节点10可将从节点(1,2,……,N)上报的信息进行实时记录、更新、检测等,还可将数据节点(1,2,……,M)上报的信息进行实时记录、更新、检测等。

[0052] 图2为本申请实施例提供的一种基于分布式存储系统的第一节点列表更新方法的流程图。该流程具体可由主节点(如图1中的主节点10)所执行。如图所示,该流程包括如下步骤:

[0053] 201:主节点接收从节点按照设定时间上报的信息。

[0054] 该设定时间可以是周期性的,也可以是非周期性的,本申请实施例在此不做限制。可选的,从节点上报的信息可包括元数据节点的地址、上报时间、元数据服务状态,还可包括元数据节点标志、版本号、型号等。该地址可以是业务地址、

[0055] 202:主节点根据从节点按照设定时间上报的信息更新第一节点列表。

[0056] 该第一节点列表可存储为文件,例如存储为元数据节点信息配置文件。

[0057] 可选的,该第一节点列表中包括元数据节点的地址、上报时间以及元数据服务状态,还可包括元数据节点的标志、版本号、型号等,可用于主节点实时记录、更新元数据节点上报的信息。表1示例性示出了第一节点列表的示例表如表1所示:

[0058] 表1:第一节点列表

[0059]	元数据节点	地址	上报时间	元数据服务状态
	从节点 1	192.168.0.1	8: 00	正常
	从节点 2	192.168.0.2	8: 10	正常
	从节点 3	192.168.0.3	8: 00	正常
				

[0060] 举例来说,该更新过程可通过以下方式:从节点(例如,图1中的从节点1)按照设定时间(例如,每隔一小时)将本节点的信息上报给主节点,主节点接收到该从节点发送的信息后,根据该从节点的地址,确定该第一节点列表中是否已经记录过该从节点的信息,若是,则对该从节点的上报时间以及服务状态进行一次更新,若否,则将该从节点的地址、上报时间以及服务状态记录在第一节点列表中。

[0061] 图3为本申请实施例提供的一种基于分布式存储系统的第二节点列表更新方法的流程图。该流程可由主节点(如图1中的主节点10)所执行。如图所示,该流程包括如下步骤:

[0062] 301:主节点接收数据节点按照设定时间上报的信息。

[0063] 该设定时间可以是周期性的,也可以是非周期性的,本申请实施例在此不做限制。可选的,数据节点上报的信息可包括数据节点的地址、上报时间,还可包括元数据节点标志、版本号、型号等,该地址可以是业务地址。

[0064] 302:主节点根据数据节点按照设定时间上报的信息更新第二节点列表。

[0065] 该第二节点列表可存储为文件,例如存储为数据节点信息配置文件。

[0066] 可选的,该第二节点列表中可包括数据节点的地址以及上报时间,还可包括数据节点的标志、版本号、型号等,可用于主节点实时记录、更新数据节点上报的信息。表2示例性示出了第二节点列表的示例表,如表2所示:

[0067] 表2:第二节点列表

[0068]	数据节点	地址	上报时间
	数据节点 1	192.168.1.2	8: 10
	数据节点 2	192.168.1.3	8: 30
	数据节点 3	192.168.1.4	8: 20
			

[0069] 举例来说,该更新过程可通过以下方式:数据节点(如图1种的数据节点1)按照设定时间将信息上报给主节点,主节点接收到该数据节点的信息后,确定第二节点列表中是否已经记录过该数据节点的信息,若是,则对该数据节点的上报时间以及进行一次更新,若否,则将该数据节点的地址、上报时间记录在第二节点列表中。

[0070] 本申请实施中,通过上述方式将元数据节点和数据节点的信息进行分别记录,也便于后续用户对信息的查询和管理。

[0071] 在分布式存储的场景中,基于图2对第一节点列表的实时更新,以及图3对第二节点列表的实时更新,还需对各列表中的节点进行故障排查,以及解决故障问题,从而维持分布式存储系统提供元数据服务的能力。

[0072] 鉴于此,图4为本申请实施例提供的一种应用于分布式存储系统的管理方法的流程图。该流程可由主节点(如图1中的主节点10)所执行。如图所示,该流程包括如下步骤:

[0073] 由于第一节点列表中的从节点数量较多,为了方便下文描述,以一个第一节点进行步骤的阐述。

[0074] 401:主节点判断该第一节点列表中的第一节点是否有效,若否,则转入402,否则,等待下一次循环检测,或者对第二节点列表中的数据节点进行有效检测。

[0075] 该第一节点列表是主节点根据从节点上报的信息进行实时更新的,该第一节点列表的更新过程可参照图2的相关描述,在此不再重复。

[0076] 可选的,主节点确定第一节点是否有效,可通过以下方式:

[0077] 方式1:若主节点确定出第一节点列表中记录的第一节点的上报时间距当前时间

的时间间隔大于第一时长阈值,则确定该第一节点失效。例如,若第一节点(如表1中的从节点1)的前一次上报时间为8:00,当前时间为9:30,设定的第一时长阈值为一个小时,则确定该第一节点失效(出现故障)。

[0078] 方式2:若主节点确定出第一节点列表中记录的第一节点的服务状态为异常,则确定该第一节点失效。

[0079] 方式3:若主节点确定出第一节点列表中记录的第一节点的服务状态为异常,且该第一节点的上报时间距当前时间的的时间间隔大于第一时长阈值,则确定该第一节点失效。

[0080] 在一些实施例中,主节点还可对该第二节点列表中的所有数据节点实时进行有效性检测,可包括如下情况:

[0081] 情况1:若主节点确定第二节点列表中存在上报时间距当前时间的的时间间隔大于第二时长阈值的数据节点,则确定该数据节点失效,并将该数据节点的地址以及上报时间从第二节点列表中删除,进一步的,还可将该数据节点的配置信息进行清理,从而可以实时保证第二节点列表中的数据节点是可用的数据节点。

[0082] 情况2:若主节点确定第二节点列表中存在上报时间距当前时间的的时间间隔大于第二时长阈值的数据节点,则确定该数据节点失效,并将该数据节点标记为不可用,便于在第二节点列表中区分可用的数据节点和不可用的数据节点。

[0083] 通过上述方式,主节点可以实时监测第一节点列表,以及第二节点列表中是否有出现故障节点,便于后续用户及时采取相应的补救措施。

[0084] 402:当判断出该第一节点失效后,主节点选取第二节点列表中的第二节点,将第二节点设置为元数据节点。

[0085] 该第二节点列表是主节点根据数据节点上报的信息进行实时更新的,该第二节点列表的更新过程可参照图3的相关描述,在此不再重复。

[0086] 可选的,在选取该第二节点设置为元数据节点时,主节点可对该第二节点进行有效性判断,避免选取的第二节点无法提供元数据服务,从而无法替代上述第一节点。若主节点确定该第二节点的前一次上报时间距当前时间的的时间间隔小于或等于第二时长阈值,则确定该第二节点有效,并选取该第二节点,将第二节点设置为元数据节点。举例来说,若第二节点(如表2中的数据节点2)的前一次上报时间为8:30,当前时间为9:30,设定的第二时长阈值为一个半小时,则确定该第二节点有效,然后将第二节点设置为元数据节点。

[0087] 可选的,当主节点对该第二节点列表中的所有数据节点实时进行有效性检测时,其选取的过程可包括:

[0088] 针对主节点采用上述情况1的方式对所有数据节点实时进行有效性检测,在选取过程中,主节点可直接从第二列表中任意选择一个第二节点。

[0089] 针对主节点采用上述情况2的方式对所有数据节点实时进行有效性检测,在选取的过程中,主节点可根据标记,从第二节点列表中选取可用的第二节点。

[0090] 通过上述方式,在选取过程中可以快速感知哪些数据节点不可用,从而防止主节点将其故障的数据节点作为元数据节点添加到第一节点列表中。

[0091] 元数据集群中元数据节点的数量决定了元数据集群的可用性,在进行元数集群部署时,会设定最多允许 m ($m < n$) 个元数据节点发生故障,还会在上述IP地址列表中实时记录元数据节点的地址。在该场景中,元数集群中元数据节点的总数量可能会发生变化,例如,

之前出现故障的元数据节点被修复、元数据集群中又新添了元数据节点,这时元数据集群中元数据节点的总数量大于 n 个,IP地址列表中记录的元数据节点地址也会大于 n 个,这样会造成资源浪费。

[0092] 鉴于此,可选的,在选取该第二节点时,还可对元数集群的可用性进行判断,若确定出元数据节点的数量不满足可用性要求,则选取第二节点列表中的第二节点,否则,就不从第二节点列表中选取第二节点。举例来说,假设在元数据集群部署时,设定有元数据节点为5个,允许故障点有2个,当第一节点列表中的第一节点失效时,主节点查询上述IP地址列表,确定出元数据节点的数量是否大于或等于3,若否,则表明元数据节点的数量不满足可用性要求,并选取第二节点列表中的第二节点,若是,则表明元数据节点的数量满足可用性要求,就不从第二节点列表中选取第二节点。

[0093] 上述步骤,将第二节点设置为元数据节点后,同步元数据,启动元数据服务,可以在短时间内替代出现故障的第一节点,无需要重新组建新的元数据集群,数据可以及时同步,降低了整个元数据集群内元数据服务不可用的风险,也保证元数据集群内提供元数据服务的元数据节点的数量不变,提高了元数据集群的可用性。

[0094] 403:当完成第二节点的选取时,主节点更新上述第一节点列表和第二节点列表。

[0095] 可选的,主节点更新该第一节点列表主要可包括以下内容:

[0096] 在一些实施例中主节点将第一节点列表中的该第一节点标记为失效,并将该第二节点的地址以及当前时间添加到该第一节点列表中。在另一些实施例中,主节点将第一节点的地址、上报时间以及元数据服务状态从第一节点列表中删除,进一步的,还可将该第一节点的配置信息从元数据集群中删除,并将该第二节点的地址以及当前时间添加到该第一节点列表中。

[0097] 可选的,主节点更新第二节点列表主要可包括:在一些实施例中,主节点将第二节点列表中的第二节点的地址以及上报时间删除,进一步的,对该第二节点的配置信息进行清理。

[0098] 在一些实施例中,上述删除操作,可将第一节点列表中第一节点需要删除的内容(第一节点的地址、上报时间以及元数据服务状态、第一节点的配置、驱动程序文件等),以及第二节点列表中第二节点需要删除的内容,统一移出到删除列表中进行删除,可以对其进行监测,各内容是否清理完成。

[0099] 在一些实施例中,当主节点更新完上述第一节点列表和第二节点列表后,可进入休眠期,等待下一次循环检测,从而实时监控各元数据节点以及数据节点,保证整个元数据集群的可用性。

[0100] 本申请实施例中,由于主节点可根据从节点按照设定时间上报的信息更新第一节点列表,主节点还可根据数据节点按照设定时间上报的信息更新第二节点列表;当主节点确定出第一节点列表中的第一节点失效时,可选取第二节点列表中的第二节点,并将第二节点设置为元数据节点,便可提供元数据服务,因此可以保证提供元数据服务的元数据节点的数量保持不变,从而提高了元数据集群的可用性。

[0101] 基于相同的技术构思,本申请实施例还提供了一种应用于分布式存储系统的管理装置,该装置可实现本申请实施例中上述应用于分布式存储系统的管理方法的流程。

[0102] 图5为本申请实施例提供的一种应用于分布式存储系统的管理装置的结构示意

图。如图所示,该装置包括:第一节点列表更新模块501、第二节点列表更新模块502、节点管理模块503。

[0103] 第一节点列表更新模块501,用于根据从节点上报的信息更新第一节点列表,所述第一节点列表中包括至少一个元数据节点的地址、上报时间以及元数据服务状态。

[0104] 第二节点列表更新模块502,用于根据数据节点上报的信息更新第二节点列表,所述第二节点列表中包括至少一个数据节点的地址以及上报时间。

[0105] 节点管理模块503,用于若所述第一节点列表中的第一节点失效,则选取所述第二节点列表中的第二节点,将所述第二节点设置为元数据节点,并指示所述第一节点列表更新模块更新所述第一节点列表,指示所述第二节点列表更新模块更新所述第二节点列表。

[0106] 可选的,节点管理模块503,具体用于:若所述第一节点满足以下条件中的至少一项,则确定所述第一节点失效:所述第一节点的上报时间距当前时间的时间间隔大于第一时长阈值;所述第一节点的元数据服务状态为异常。

[0107] 可选的,节点管理模块503,具体用于若所述第二节点的上报时间距当前时间的时间间隔小于或等于第二时长阈值,则确定所述第二节点有效,并选取所述第二节点。

[0108] 可选的,节点管理模块503,具体用于:指示所述第一节点列表更新模块501将所述第一节点标记为失效,或者将所述第一节点的地址、上报时间以及元数据服务状态从所述第一节点列表中删除,并指示所述第二节点列表更新模块502将所述第二节点的地址以及当前时间添加到所述第一节点列表中;以及指示所述第二节点列表更新模块502将所述第二节点列表中的所述第二节点的地址以及上报时间删除。

[0109] 可选的,节点管理模块503,具体用于:若确定所述元数据分布式系统中的元数据节点的数量不满足可用性要求,则选取所述第二节点列表中的第二节点。

[0110] 在此需要说明的是,本申请实施例提供的上述装置,能够实现上述分布式存储系统管理方法实施例中的所有方法步骤,且能够达到相同的技术效果,在此不再对本实施例中与方法实施例相同的部分及有益效果进行具体赘述。

[0111] 基于相同的技术构思,本申请实施例还提供了一种通信装置,该装置可实现本申请实施例中上述应用于分布式存储系统的管理方法的流程。

[0112] 图6为本申请实施例提供的通信装置的结构示意图。

[0113] 如图所示,该装置可包括:处理器601、存储器602以及总线接口603。

[0114] 处理器601负责管理总线架构和通常的处理,存储器602可以存储处理器601在执行操作时所使用的数据。

[0115] 总线架构可以包括任意数量的互联的总线和桥,具体由处理器601代表的一个或多个处理器和存储器602代表的存储器的各种电路链接在一起。总线架构还可以将诸如外围设备、稳压器和功率管理电路等之类的各种其他电路链接在一起,这些都是本领域所公知的,因此,本文不再对其进行进一步描述。总线接口提供接口。处理器601负责管理总线架构和通常的处理,存储器602可以存储处理器601在执行操作时所使用的数据。

[0116] 本申请实施例揭示的流程,可以应用于处理器601中,或者由处理器601实现。在实现过程中,信号处理流程的各步骤可以通过处理器601中的硬件的集成逻辑电路或者软件形式的指令完成。处理器601可以是通用处理器、数字信号处理器、专用集成电路、现场可编程门阵列或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件,可以实现

或者执行本申请实施例中的公开的各方法、步骤及逻辑框图。通用处理器可以是微处理器或者任何常规的处理器等。结合本申请实施例所申请的方法的步骤可以直接体现为硬件处理器执行完成,或者用处理器中的硬件及软件模块组合执行完成。软件模块可以位于随机存储器,闪存、只读存储器,可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器602,处理器601读取存储器602中的信息,结合其硬件完成信息处理流程的步骤。

[0117] 具体地,处理器601,用于读取存储器602中的计算机指令并执行本申请实施例中的一种应用于分布式存储系统的管理方法。

[0118] 在此需要说明的是,申请本申请实施例提供的上述通信装置,能够实现上述方法实施例所实现的所有方法步骤,且能够达到相同的技术效果,在此不再对本实施例中与方法实施例相同的部分及有益效果进行具体赘述。

[0119] 本申请实施例还提供了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可执行指令,计算机可执行指令用于使计算机执行上述实施例中的一种应用于分布式存储系统的管理方法。

[0120] 本申请实施例还提供了一种计算机程序产品,所述计算机程序产品在被计算机调用时,使得所述计算机执行上述实施例中的一种应用于分布式存储系统的管理方法。

[0121] 本领域内的技术人员应明白,本申请的实施例可提供为方法、系统、或计算机程序产品。因此,本申请可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且,本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0122] 本申请是参照根据本申请的方法、设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器,使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0123] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品,该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0124] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0125] 显然,本领域的技术人员可以对本申请进行各种改动和变型而不脱离本申请的精神和范围。这样,倘若本申请的这些修改和变型属于本申请权利要求及其等同技术的范围之内,则本申请也意图包含这些改动和变型在内。

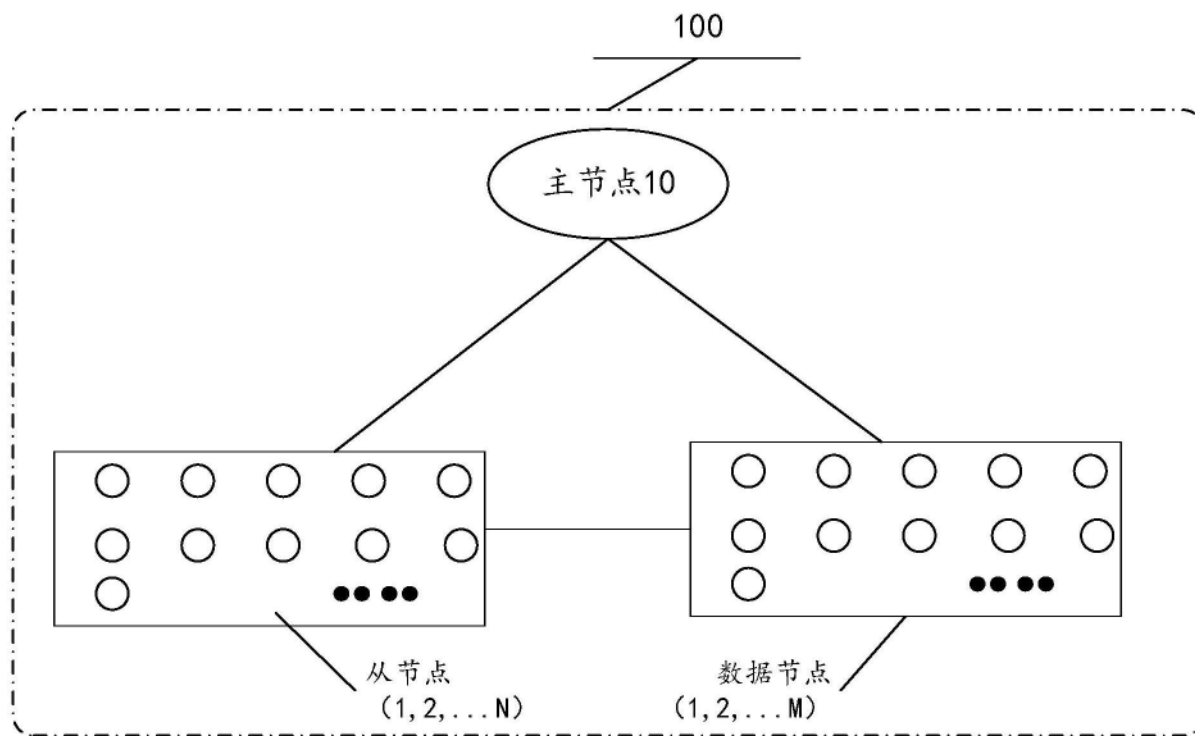


图1

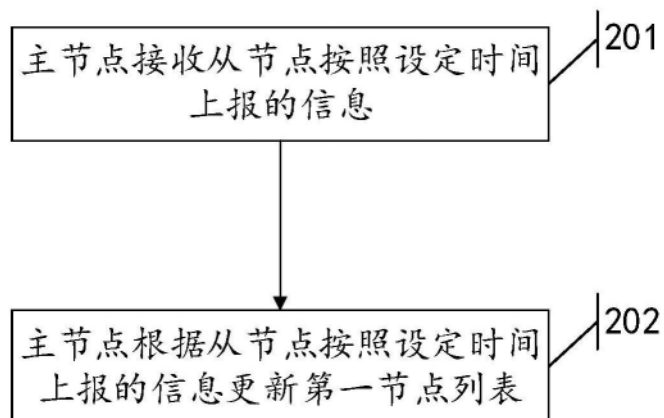


图2

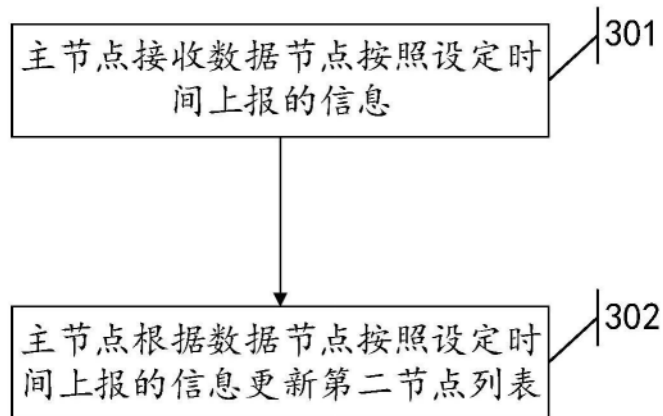


图3

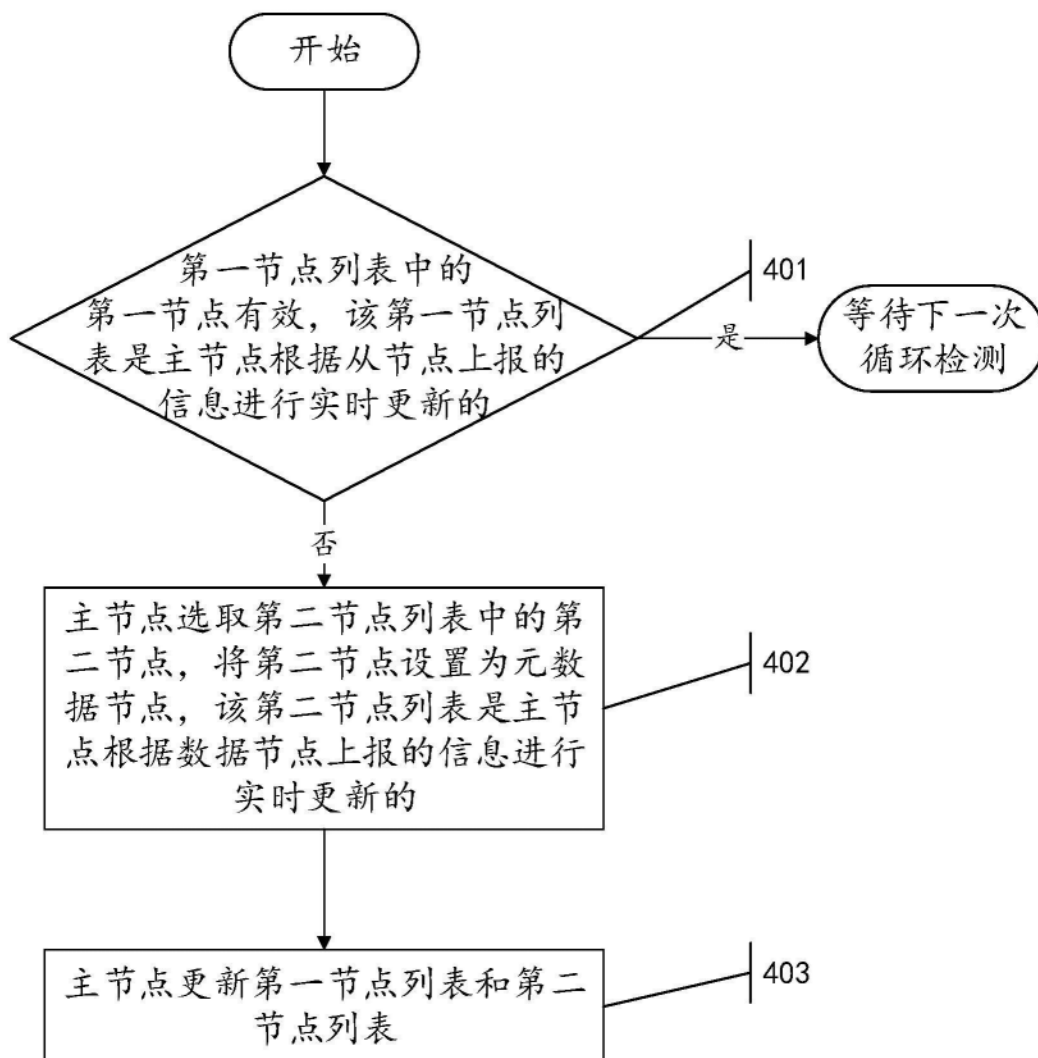


图4

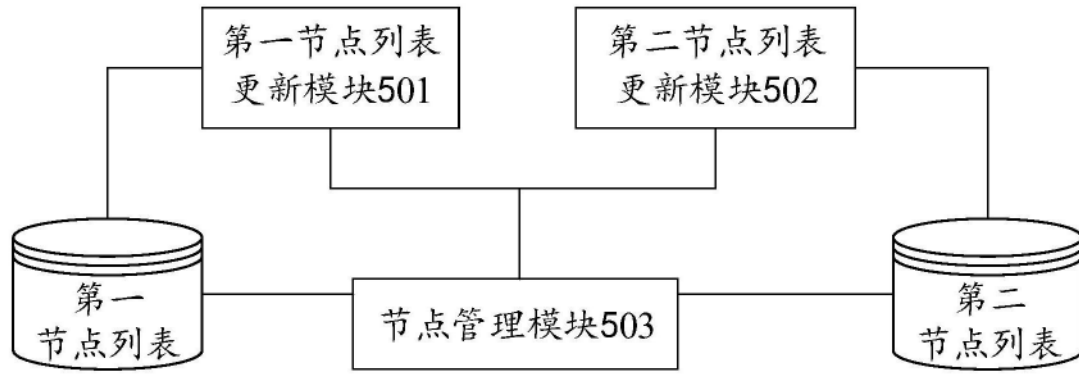


图5



图6