(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 105323271 B
(45)授权公告日 2020.04.24

(21)申请号 201410289531.7

(22)申请日 2014.06.24

(65)同一申请的已公布的文献号
申请公布号 CN 105323271 A

(43)申请公布日 2016.02.10

(73)专利权人 中兴通讯股份有限公司
地址 518057 广东省深圳市南山区高新技术产业园科技南路中兴通讯大厦法务部

(72)发明人 莫嫣 高洪 韩银俊

(74)专利代理机构 深圳市世纪恒程知识产权代理事务所 44287
代理人 杨雪梅

(51)Int.Cl.
H04L 29/08(2006.01)

(56)对比文件
CN 103747072 A,2014.04.23,
CN 103763155 A,2014.04.30,
高洪等.云计算分布式缓存技术及其在物联网中的应用.《中兴通讯技术》.2011,第17卷(第4期),

审查员 李腾飞

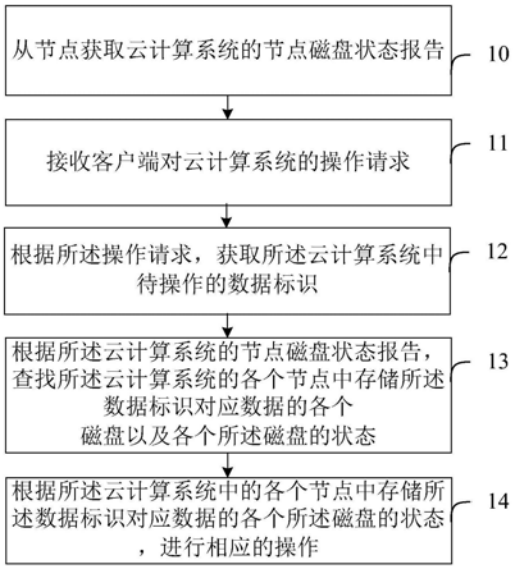
权利要求书3页 说明书9页 附图3页

(54)发明名称

一种云计算系统以及云计算系统的处理方法和装置

(57)摘要

本发明提供一种云计算系统以及云计算系统的处理方法和装置。所述云计算系统的处理方法,包括:接收客户端对云计算系统的操作请求;根据所述操作请求,获取所述云计算系统中待操作的数据标识;根据所述云计算系统的节点磁盘状态报告,查找所述云计算系统的各个节点中存储所述数据标识对应数据的各个磁盘以及各个所述磁盘的状态;所述节点磁盘状态报告包括:所述云计算系统的各个节点中磁盘的状态、所述磁盘中存储的数据所对应的数据标识;根据所述云计算系统中的各个节点中存储所述数据标识对应数据的各个所述磁盘的状态,进行相应的操作。本发明能够提高系统对磁盘故障的容忍性。



CN 105323271 B

1.一种云计算系统的处理方法,其特征在于,包括:

接收客户端对云计算系统的操作请求;

根据所述操作请求,获取所述云计算系统中待操作的数据标识;

根据所述云计算系统的节点磁盘状态报告,查找所述云计算系统的各个节点中存储所述数据标识对应数据的各个磁盘以及各个所述磁盘的状态;所述节点磁盘状态报告包括:所述云计算系统的各个节点中磁盘的状态、所述磁盘中存储的数据所对应的数据标识;

根据所述云计算系统中的各个节点中存储所述数据标识对应数据的各个所述磁盘的状态,进行相应的操作;

其中,所述根据所述云计算系统中的各个节点中存储所述数据标识对应数据的各个所述磁盘的状态,进行相应的操作的步骤包括:

所述操作请求为更新请求;当所述云计算系统中存储所述数据且处于正常状态的所述磁盘的数量大于或等于所述云计算系统预定的一次数据更新请求的最少参与节点数量时,则响应所述更新请求;否则,拒绝所述更新请求;或者

所述操作请求为数据访问请求;当所述云计算系统中存储所述数据且处于正常状态的所述磁盘的数量大于或等于所述云计算系统预定的一次数据访问请求获取的数据副本数量时,则响应所述数据访问请求;否则,拒绝所述数据访问请求。

2.根据权利要求1所述的方法,其特征在于,所述当所述云计算系统中存储所述数据且处于正常状态的所述磁盘的数量大于或等于所述云计算系统预定的一次数据更新请求的最少参与节点数量时,则响应所述更新请求的步骤包括:

当所述操作请求为更新请求,并且存储所述数据的主节点的磁盘的状态为正常时,所述云计算系统的主节点向主节点的所述数据所在磁盘进行数据更新;所述云计算系统的从节点从所述主节点获取待同步的数据,所述从节点向所述从节点的所述数据所在磁盘进行数据更新;

当所述操作请求为更新请求,并且存储所述数据的主节点的磁盘的状态为故障时,所述云计算系统的第一从节点向所述第一从节点的所述数据所在磁盘进行数据更新;所述云计算系统的第二从节点从所述第一从节点获取待同步的数据;所述第二从节点向所述第二从节点的所述数据所在磁盘进行数据更新;所述第一从节点和所述第二从节点的存储所述数据的磁盘的状态为正常。

3.根据权利要求2所述的方法,其特征在于,所述当所述云计算系统中存储所述数据且处于正常状态的所述磁盘的数量大于或等于所述云计算系统预定的一次数据访问请求获取的数据副本数量时,则响应所述数据访问请求的步骤包括:

当所述操作请求为数据访问请求,并且存储所述数据的主节点的磁盘的状态为正常时,从所述云计算系统的主节点的所述数据所在磁盘中获取所述数据的第一副本,从所述云计算系统的至少一个从节点的所述数据所在磁盘中获取所述数据的第二副本;从所述第一副本和所述第二副本中,选取最新版本的副本;并将所述最新版本的副本发送给所述客户端;所述第二从节点的存储所述数据的磁盘的状态为正常;

当所述操作请求为数据访问请求,并且存储所述数据的主节点的磁盘的状态为故障时,从所述云计算系统的至少一个从节点的所述数据所在磁盘中获取所述数据的第三副本;从至少一个所述第三副本中,选取最新版本的副本,并将所述最新版本的副本发送给所

述客户端;所述第二从节点的存储所述数据的磁盘的状态为正常。

4.根据权利要求1所述的方法,其特征在于,所述接收客户端的操作请求的步骤之前,所述方法还包括:

从节点获取所述云计算系统的节点磁盘状态报告。

5.一种云计算系统的处理装置,其特征在于,包括:

第一接收单元,接收客户端对云计算系统的操作请求;

获取单元,根据所述操作请求,获取所述云计算系统中待操作的数据标识;

查找单元,根据所述云计算系统的节点磁盘状态报告,查找所述云计算系统的各个节点中存储所述数据标识对应数据的各个磁盘以及各个所述磁盘的状态;所述节点磁盘状态报告包括:所述云计算系统的各个节点中磁盘的状态、所述磁盘中存储的数据所对应的数据标识;

操作单元,根据所述云计算系统中的各个节点中存储所述数据标识对应数据的各个所述磁盘的状态,进行相应的操作;

其中,所述操作单元包括:

第一响应子单元,所述操作请求为更新请求;当所述云计算系统中存储所述数据且处于正常状态的所述磁盘的数量大于或等于所述云计算系统预定的一次数据更新请求的最少参与节点数量时,则响应所述更新请求;

第一拒绝子单元,当所述云计算系统中存储所述数据且处于正常状态的所述磁盘的数量小于所述云计算系统预定的一次数据更新请求的最少参与节点数量时,拒绝所述更新请求;

第二响应子单元,所述操作请求为数据访问请求;当所述云计算系统中存储所述数据且处于正常状态的所述磁盘的数量大于或等于所述云计算系统预定的一次数据访问请求获取的数据副本数量时,则响应所述数据访问请求;

第二拒绝子单元,当所述云计算系统中存储所述数据且处于正常状态的所述磁盘的数量小于所述云计算系统预定的一次数据访问请求获取的数据副本数量时,拒绝所述数据访问请求。

6.根据权利要求5所述的装置,其特征在于,还包括:

第二接收单元,从节点接收所述云计算系统的节点磁盘状态报告。

7.一种云计算系统,其特征在于,包括:客户端、处理装置、节点、所述节点对应的磁盘;

所述处理装置,接收所述客户端对云计算系统的操作请求;根据所述操作请求,获取所述云计算系统中待操作的数据标识;根据所述云计算系统的节点磁盘状态报告,查找所述云计算系统的各个所述节点中存储所述数据标识对应数据的磁盘以及各个所述磁盘的状态;所述节点磁盘状态报告包括:所述云计算系统的各个节点中所述磁盘的状态、所述磁盘中存储的数据所对应的数据标识;根据所述云计算系统中的各个节点中存储所述数据标识对应数据的各个所述磁盘的状态,进行相应的操作;

其中,所述根据所述云计算系统中的各个节点中存储所述数据标识对应数据的各个所述磁盘的状态,进行相应的操作的步骤包括:

所述操作请求为更新请求;当所述云计算系统中存储所述数据且处于正常状态的所述磁盘的数量大于或等于所述云计算系统预定的一次数据更新请求的最少参与节点数量时,

则响应所述更新请求;否则,拒绝所述更新请求;或者

所述操作请求为数据访问请求;当所述云计算系统中存储所述数据且处于正常状态的所述磁盘的数量大于或等于所述云计算系统预定的一次数据访问请求获取的数据副本数量时,则响应所述数据访问请求;否则,拒绝所述数据访问请求。

8.根据权利要求7所述的系统,其特征在于,所述节点,给所述处理装置发送节点磁盘状态报告。

一种云计算系统以及云计算系统的处理方法和装置

技术领域

[0001] 本发明涉及云计算技术领域,特别是指一种云计算系统以及云计算系统的处理方法和装置。

背景技术

[0002] 目前,云计算(Cloud Computing)是网格计算(Grid Computing)、分布式计算(Distributed Computing)、并行计算(Parallel Computing)、效用计算(Utility Computing)、网络存储(Network Storage Technologies)虚拟化(Virtualization)、负载均衡(Load Balance)等传统计算机技术和网络技术发展融合的产物。它旨在通过网络把多个成本相对较低的计算实体,整合成一个具有强大计算能力的系统。分布式缓存是云计算范畴中的一个领域,其作用是提供海量数据的分布式存储服务以及高速读写访问的能力。

[0003] 分布式缓存系统是由若干服务器节点和客户端互相连接构成的;服务器节点负责数据的存储,客户端可以对服务器做数据的写入、读取、更新、删除等操作。一般来说,数据不可能只保存在单个服务器节点(以下简称“节点”)上,而是在多台节点上保存同一个数据的副本,互为备份。最常见的存储模式为主备模式,其中一个节点做为主节点(master),其他节点作为备节点(slave),主节点的身份通过选举或其他算法获取。为简化流程,数据更新一般发生在主节点上,备节点从主节点获取数据进行同步,而数据访问可以从主节点中获取数据,也可以从备节点中获取数据,具体看该访问的一致性策略。

[0004] 在分布式缓存系统中,根据一致性及可用性的要求,一般将该数据存储方式按NRW进行分类,其中N表示数据的副本数、R表示一次数据访问请求中获取的数据副本数,W表示一次数据更新请求的最少参与节点数(即多少个节点上的数据更新完成)。

[0005] 当分布式缓存系统实现持久化功能时,分布在该服务器上的数据保存在磁盘上。在实际情况下,如果磁盘发生故障,该服务器就无法提供读写服务了。由于分布式缓存系统数据保存有多个副本的特性,这时,只要其他服务器处于正常状态,系统依然可以通过其他节点的副本正常提供读写服务。

[0006] 如果分布式缓存系统节点挂接了多块磁盘,其中只有一个或者少数几个磁盘由于某种原因损坏,导致该服务器不能正常提供服务,根据前述,由于其他服务器为正常可用,整个集群还是可用的。假定在这段时间内,另一个服务器也发生了类似情况,那个节点也不能正常提供服务,很可能使得副本数无法满足NRW策略,那么分布式缓存集群就彻底无法提供服务了。典型的情况是在比较常用的NRW为3/2/2的条件下,两个节点宕掉,只有一个节点正常,读写操作都无法满足最小在两个副本上操作的要求。

发明内容

[0007] 本发明要解决的技术问题是,提供一种云计算系统以及云计算系统的处理方法和装置,能够提高系统对磁盘故障的容忍性。

[0008] 为解决上述技术问题,本发明的实施例提供一种能耗监测系统,包括:

[0009] 一方面，提供一种云计算系统的处理方法，包括：

[0010] 接收客户端对云计算系统的操作请求；

[0011] 根据所述操作请求，获取所述云计算系统中待操作的数据标识；

[0012] 根据所述云计算系统的节点磁盘状态报告，查找所述云计算系统的各个节点中存储所述数据标识对应数据的各个磁盘以及各个所述磁盘的状态；所述节点磁盘状态报告包括：所述云计算系统的各个节点中磁盘的状态、所述磁盘中存储的数据所对应的数据标识；

[0013] 根据所述云计算系统中的各个节点中存储所述数据标识对应数据的各个所述磁盘的状态，进行相应的操作。

[0014] 所述根据各个所述磁盘的状态，进行相应的操作的步骤包括：

[0015] 所述操作请求为更新请求；当所述云计算系统中存储所述数据且处于正常状态的所述磁盘的数量大于或等于所述云计算系统预定的一次数据更新请求的最少参与节点数量时，则响应所述更新请求；否则，拒绝所述更新请求；或者

[0016] 所述操作请求为数据访问请求；当所述云计算系统中存储所述数据且处于正常状态的所述磁盘的数量大于或等于所述云计算系统预定的一次数据访问请求获取的数据副本数量时，则响应所述数据访问请求；否则，拒绝所述数据访问请求。

[0017] 所述当所述云计算系统中存储所述数据且处于正常状态的所述磁盘的数量大于或等于所述云计算系统预定的一次数据更新请求的最少参与节点数量时，则响应所述更新请求的步骤包括：

[0018] 当所述操作请求为更新请求，并且存储所述数据的主节点的磁盘的状态为正常时，所述云计算系统的主节点向主节点的所述数据所在磁盘进行数据更新；所述云计算系统的从节点从所述主节点获取待同步的数据，所述从节点向所述从节点的所述数据所在磁盘进行数据更新；

[0019] 当所述操作请求为更新请求，并且存储所述数据的主节点的磁盘的状态为故障时，所述云计算系统的第一从节点向所述第一从节点的所述数据所在磁盘进行数据更新；所述云计算系统的第二从节点从所述第一从节点获取待同步的数据；所述第二节点向所述第二从节点的所述数据所在磁盘进行数据更新；所述第一从节点和所述第二从节点的存储所述数据的磁盘的状态为正常。

[0020] 所述当所述云计算系统中存储所述数据且处于正常状态的所述磁盘的数量大于或等于所述云计算系统预定的一次数据访问请求获取的数据副本数量时，则响应所述数据访问请求的步骤包括：

[0021] 当所述操作请求为数据访问请求，并且存储所述数据的主节点的磁盘的状态为正常时，从所述云计算系统的主节点的所述数据所在磁盘中获取所述数据的第一副本，从所述云计算系统的至少一个从节点的所述数据所在磁盘中获取所述数据的第二副本；从所述第一副本和所述第二副本中，选取最新版本的副本；并将所述最新版本的副本发送给所述客户端；所述第二从节点的存储所述数据的磁盘的状态为正常；

[0022] 当所述操作请求为数据访问请求，并且存储所述数据的主节点的磁盘的状态为故障时，从所述云计算系统的至少一个从节点的所述数据所在磁盘中获取所述数据的第三副本；从至少一个所述第三副本中，选取最新版本的副本，并将所述最新版本的副本发送给所述客户端；所述第二从节点的存储所述数据的磁盘的状态为正常。

[0023] 所述接收客户端的操作请求的步骤之前，所述方法还包括：

[0024] 从节点获取所述云计算系统的节点磁盘状态报告。

[0025] 另一方面，提供一种云计算系统的处理装置，包括：

[0026] 第一接收单元，接收客户端对云计算系统的操作请求；

[0027] 获取单元，根据所述操作请求，获取所述云计算系统中待操作的数据标识；

[0028] 查找单元，根据所述云计算系统的节点磁盘状态报告，查找所述云计算系统的各个节点中存储所述数据标识对应数据的各个磁盘以及各个所述磁盘的状态；所述节点磁盘状态报告包括：所述云计算系统的各个节点中磁盘的状态、所述磁盘中存储的数据所对应的数据标识；

[0029] 操作单元，根据所述云计算系统中的各个节点中存储所述数据标识对应数据的各个所述磁盘的状态，进行相应的操作。

[0030] 所述操作单元包括：

[0031] 所述操作单元包括：

[0032] 第一响应子单元，所述操作请求为更新请求；当所述云计算系统中存储所述数据且处于正常状态的所述磁盘的数量大于或等于所述云计算系统预定的一次数据更新请求的最少参与节点数量时，则响应所述更新请求；

[0033] 第一拒绝子单元，当所述云计算系统中存储所述数据且处于正常状态的所述磁盘的数量小于所述云计算系统预定的一次数据更新请求的最少参与节点数量时，拒绝所述更新请求；

[0034] 第二响应子单元，所述操作请求为数据访问请求；当所述云计算系统中存储所述数据且处于正常状态的所述磁盘的数量大于或等于所述云计算系统预定的一次数据访问请求获取的数据副本数量时，则响应所述数据访问请求；

[0035] 第二拒绝子单元，当所述云计算系统中存储所述数据且处于正常状态的所述磁盘的数量小于所述云计算系统预定的一次数据访问请求获取的数据副本数量时，拒绝所述数据访问请求。

[0036] 所述的装置，还包括：

[0037] 第二接收单元，从节点接收所述云计算系统的节点磁盘状态报告。

[0038] 另一方面，提供一种云计算系统，包括：客户端、处理装置、节点、所述节点对应的磁盘；

[0039] 所述处理装置，接收所述客户端对云计算系统的操作请求；根据所述操作请求，获取所述云计算系统中待操作的数据标识；根据所述云计算系统的节点磁盘状态报告，查找所述云计算系统的各个所述节点中存储所述数据标识对应数据的各个磁盘以及各个所述磁盘的状态；所述节点磁盘状态报告包括：所述云计算系统的各个节点中所述磁盘的状态、所述磁盘中存储的数据所对应的数据标识；根据所述云计算系统中的各个节点中存储所述数据标识对应数据的各个所述磁盘的状态，进行相应的操作。

[0040] 所述节点，给所述处理装置发送节点磁盘状态报告。

[0041] 本发明的上述技术方案的有益效果如下：

[0042] 本发明针对分布式缓存系统，在有磁盘损坏的情况下，可以充分利用可用的资源，整合出符合一致性和可用性要求的副本资源，尽可能提高系统的可用性，提高系统对故障

的容忍性。

附图说明

[0043] 图1为本发明所述的一种云计算系统的处理方法的流程示意图；

[0044] 图2为本发明所述的一种云计算系统的处理装置的结构示意图；

[0045] 图3为本发明所述的一种云计算系统的结构示意图；

[0046] 图4和图5为本发明所述的一种云计算系统的应用场景的结构示意图。

具体实施方式

[0047] 为使本发明要解决的技术问题、技术方案和优点更加清楚，下面将结合附图及具体实施例进行详细描述。

[0048] 如图1所示，为本发明所述的一种云计算系统的处理方法，包括：

[0049] 步骤11，接收客户端对云计算系统的操作请求；操作请求可以为数据更新请求或者数据访问请求等。

[0050] 步骤12，根据所述操作请求，获取所述云计算系统中待操作的数据标识；例如，操作请求为更新图4中的副本1，副本1为数据标识。

[0051] 步骤13，根据所述云计算系统的节点磁盘状态报告，查找所述云计算系统的各个节点中存储所述数据标识对应数据的各个磁盘以及各个所述磁盘的状态；所述节点磁盘状态报告包括：所述云计算系统的各个节点中磁盘的状态、所述磁盘中存储的数据所对应的数据标识；磁盘的状态为正常或者故障，图4中，节点A的磁盘状态报告为：(节点A：磁盘I，副本1，故障；磁盘II，副本2，正常；磁盘III，副本3，正常)。

[0052] 步骤14，根据所述云计算系统中的各个节点中存储所述数据标识对应数据的各个所述磁盘的状态，进行相应的操作。

[0053] 步骤14之前，所述方法还包括：

[0054] 步骤10，从节点获取所述云计算系统的节点磁盘状态报告。节点检测到存储一数据的磁盘损坏或者发生故障，则发送报告；或者基于请求来发送报告。

[0055] 其中，步骤14步骤包括：

[0056] 所述操作请求为更新请求；当所述云计算系统中存储所述数据且处于正常状态的所述磁盘的数量大于或等于所述云计算系统预定的一次数据更新请求的最少参与节点数量时，则响应所述更新请求；否则，拒绝所述更新请求；或者

[0057] 所述操作请求为数据访问请求；当所述云计算系统中存储所述数据且处于正常状态的所述磁盘的数量大于或等于所述云计算系统预定的一次数据访问请求获取的数据副本数量时，则响应所述数据访问请求；否则，拒绝所述数据访问请求。

[0058] 具体为：

[0059] 当所述操作请求为更新请求，并且存储所述数据的主节点的磁盘的状态为正常时，所述云计算系统的主节点向主节点的所述数据所在磁盘进行数据更新；所述云计算系统的从节点从所述主节点获取待同步的数据，所述从节点向所述从节点的所述数据所在磁盘进行数据更新；

[0060] 当所述操作请求为更新请求，并且存储所述数据的主节点的磁盘的状态为故障

时，所述云计算系统的第一从节点向所述第一从节点的所述数据所在磁盘进行数据更新；所述云计算系统的第二从节点从所述第一从节点获取待同步的数据；所述第二节点向所述第二从节点的所述数据所在磁盘进行数据更新；所述第一从节点和所述第二从节点的存储所述数据的磁盘的状态为正常。

[0061] 当所述操作请求为数据访问请求，并且存储所述数据的主节点的磁盘的状态为正常时，从所述云计算系统的主节点的所述数据所在磁盘中获取所述数据的第一副本，从所述云计算系统的至少一个(也可以为两个或者3个，根据实际情况设定)从节点的所述数据所在磁盘中获取所述数据的第二副本；从所述第一副本和所述第二副本中，选取最新版本的副本；并将所述最新版本的副本发送给所述客户端；所述第二从节点的存储所述数据的磁盘的状态为正常；

[0062] 当所述操作请求为数据访问请求，并且存储所述数据的主节点的磁盘的状态为故障时，从所述云计算系统的至少一个从节点的所述数据所在磁盘中获取所述数据的第三副本；从至少一个所述第三副本中，选取最新版本的副本，并将所述最新版本的副本发送给所述客户端；所述第二从节点的存储所述数据的磁盘的状态为正常。

[0063] 例如，图5为一个由3个节点组成的分布式缓存存储系统，该存储系统每个数据有三个副本，采用322的方式更新及访问数据。云计算系统规定的读请求访问副本数量为2，当有一个磁盘坏掉时，仍然能够响应更新或数据访问操作请求，当有两个磁盘坏掉时，则不能响应操作请求。

[0064] 本发明中，当发生节点磁盘故障，甚至多个节点同时发生故障磁盘，只要集群上剩余的可用磁盘上副本数能满足NRW策略，系统就可以保证一致性和可用性，甚至可能毫不影响所有数据的服务，更不会发生系统彻底无法提供服务的情况，也就尽可能提供了服务。

[0065] 当然，在部分磁盘损坏继续提供服务的情况下，随之带来磁盘恢复后数据的恢复问题，这可以通过分布式缓存数据恢复功能完成，也就是从其他节点上获取副本数据来修复。

[0066] 如图2所示，为本发明所述的一种云计算系统的处理装置，包括：

[0067] 第一接收单元21，接收客户端对云计算系统的操作请求；

[0068] 获取单元22，根据所述操作请求，获取所述云计算系统中待操作的数据标识；

[0069] 查找单元23，根据所述云计算系统的节点磁盘状态报告，查找所述云计算系统的各个节点中存储所述数据标识对应数据的各个磁盘以及各个所述磁盘的状态；所述节点磁盘状态报告包括：所述云计算系统的各个节点中磁盘的状态、所述磁盘中存储的数据所对应的数据标识；

[0070] 操作单元24，根据所述云计算系统中的各个节点中存储所述数据标识对应数据的各个所述磁盘的状态，进行相应的操作。

[0071] 所述操作单元24包括：

[0072] 第一响应子单元，所述操作请求为更新请求；当所述云计算系统中存储所述数据且处于正常状态的所述磁盘的数量大于或等于所述云计算系统预定的一次数据更新请求的最少参与节点数量时，则响应所述更新请求；

[0073] 第一拒绝子单元，当所述云计算系统中存储所述数据且处于正常状态的所述磁盘的数量小于所述云计算系统预定的一次数据更新请求的最少参与节点数量时，拒绝所述更

新请求；

[0074] 第二响应子单元，所述操作请求为数据访问请求；当所述云计算系统中存储所述数据且处于正常状态的所述磁盘的数量大于或等于所述云计算系统预定的一次数据访问请求获取的数据副本数量时，则响应所述数据访问请求；

[0075] 第二拒绝子单元，当所述云计算系统中存储所述数据且处于正常状态的所述磁盘的数量小于所述云计算系统预定的一次数据访问请求获取的数据副本数量时，拒绝所述数据访问请求。

[0076] 所述的装置，还包括：

[0077] 第二接收单元25，从节点接收所述云计算系统的节点磁盘状态报告。

[0078] 如图3所示，为本发明所述的一种云计算系统，包括：客户端31、处理装置32、节点33、所述节点33对应的磁盘34；

[0079] 所述处理装置32，接收所述客户端31对云计算系统的操作请求；根据所述操作请求，获取所述云计算系统中待操作的数据标识；根据所述云计算系统的节点磁盘状态报告，查找所述云计算系统的各个所述节点33中存储所述数据标识对应数据的磁盘以及各个所述磁盘34的状态；所述节点磁盘状态报告包括：所述云计算系统的各个节点33中所述磁盘的状态、所述磁盘中存储的数据所对应的数据标识；根据所述云计算系统中的各个节点中存储所述数据标识对应数据的各个所述磁盘34的状态，进行相应的操作。

[0080] 所述节点33，给所述处理装置32发送节点磁盘状态报告。

[0081] 以下描述本发明的两个应用场景。

[0082] 第一应用场景为描述一种云计算分布式缓存系统中、多磁盘路径下当磁盘损坏情况时可用性的实现方法。

[0083] 前置步骤：客户端与分布式缓存系统中多个服务器节点建立连接，服务器节点间之间互相建立连接并且运行正常，每个服务器都有若干块磁盘用于数据的持久化，不同的数据分片持久化在不同的磁盘上。数据副本数为N，读请求访问副本数为R，写请求最少更新副本成功数为W，系统的单次最大故障容忍度为O(表示容许O个节点上的请求发生故障，如单点故障则O＝1，O<W)，一致性要求W+R>N。

[0084] 步骤A：系统正常情况下，每个节点上的所有磁盘正常工作，数据在系统中有N个副本。当客户端发起数据更新请求时，由Master向数据所在磁盘进行数据更新处理，slave从master同步数据，并向slave上数据所在磁盘进行数据更新，当数据更新在W个节点上成功完成后，返回客户端数据更新成功消息；

[0085] 当客户端发起数据访问请求时，由Master/Slave处理请求，从R个节点数据所在磁盘获取访问的数据副本后，从该R个数据副本中选取最新的副本返回给客户端。

[0086] 步骤B：节点A启动时，发现某个磁盘故障无法访问，但其他磁盘仍然正常；或者，节点A运行过程中，发现某个磁盘多次访问失败，判定为该磁盘故障。节点A不切换成节点故障，而是继续提供读写服务，同时记录下故障磁盘和该磁盘上对应的数据副本的标识。

[0087] 步骤C：当客户端发起数据更新请求时，并且该数据恰好分布在步骤B所述节点A的故障磁盘上，则当向该节点更新数据时，节点A直接返回失败；当数据更新在W个节点上(这W个节点中不包含节点A)成功完成后，返回给客户端数据更新成功消息；

[0088] 当客户端发起数据访问请求时，节点A直接返回失败，由Master/Slave处理请求，

从R个节点(这R个节点中不包含节点A)数据所在磁盘获取访问的数据副本后,从该R个数据副本中选取最新的副本返回给客户端。

[0089] 步骤D:当客户端发起数据更新和访问请求,并且该数据不分布在步骤B所述节点A的故障磁盘上,则处理方式同步骤A。

[0090] 步骤E:当节点B在运行过程中,对某个磁盘多次访问失败判定该磁盘为故障。节点B不切换成节点故障,而是继续提供读写服务,同时记录下故障磁盘和该磁盘上对应的数据副本的标识。

[0091] 假定节点B的故障磁盘和节点A的故障磁盘上保存的副本无重合。继续下一步骤。

[0092] 步骤F:当客户端发起数据更新和访问请求,并且该数据恰好分布在步骤E所述节点B的故障磁盘上,基于上述假定,则不在步骤B所述节点A的故障磁盘上,则当向该节点更新数据时,节点B直接返回失败;当数据更新在W个节点上(这W个节点中不包含节点B)成功完成后,返回给客户端数据更新成功消息;

[0093] 当客户端发起数据访问请求时,节点B直接返回失败,由Master/Slave处理请求,从R个节点(这R个节点中不包含节点B)数据所在磁盘获取访问的数据副本后,从该R个数据副本中选取最新的副本,返回给客户端。

[0094] 步骤G:当客户端发起数据更新请求时,并且该数据恰好分布在步骤B所述节点A的故障磁盘上,基于上述假定,则不在步骤E所述节点B的故障磁盘上,则当向该节点更新和访问数据时,处理过程同步骤C,结果是可以正常更新和访问到。

[0095] 本发明中,当发生节点磁盘故障,甚至多个节点同时发生故障磁盘,只要集群上剩余的可用磁盘上副本数能满足NRW策略,系统就可以保证一致性和可用性,甚至可能毫不影响所有数据的服务,更不会发生系统彻底无法提供服务的情况,也就尽可能提供了服务。

[0096] 当然,在部分磁盘损坏继续提供服务的情况下,随之带来磁盘恢复后数据的恢复问题,这可以通过分布式缓存数据恢复功能完成,也就是从其他节点上获取副本数据来修复。

[0097] 本发明提供了一种在分布式缓存系统在多磁盘损坏情况下提高可用性的实现方法,在一致性不变的情况下,增强了系统的可用性,从而优化了应用体验。

[0098] 以下结合图4和图5,描述第二应用场景。

[0099] 具体为:针对322模式的主备存储系统详细描述单节点出现磁盘损坏和多节点同时出现磁盘损坏下,可用性实现方案。

[0100] 由服务器节点和客户端构成分布式缓存系统,对一个特定的数据,有一个主节点(master)负责处理客户端的更新及访问请求,有若干个备节点用于同步master的数据并接收客户端的数据访问请求(slave不处理数据更新请求)。

[0101] 环境:一个由3个节点组成的分布式缓存存储系统,该存储系统每个数据有三个副本,采用322的方式更新及访问数据。

[0102] 本发明包括如下步骤:

[0103] 步骤1,初始正常阶段,系统接收客户端请求,假定数据位于节点A的磁盘I上副本1(相当于上述的数据标识)、节点B的磁盘I上副本1、以及节点C的磁盘III上副本1上。为描述简化起见,假定节点B上的副本1是master,其他两个节点上的副本是slave。节点A上的副本2是master,其他两个节点上的副本是slave。节点A上的副本3是master,其他两个节点上的

副本是slave。

[0104] 步骤2，当客户端发起数据更新请求时，由B节点Master向磁盘I上副本1进行数据更新，slave从master同步数据，并向slave上数据所在磁盘进行数据更新，当数据更新在W＝2个节点上成功完成后，返回给客户端数据更新成功消息。由于所有磁盘都正常，实际所有副本都更新成功了；当客户端发起数据访问请求时，三个节点都处理请求，从R＝2个节点数据所在磁盘获取访问的数据副本后，返回客户端，实际所有节点副本都读取成功了。

[0105] 步骤3，如图4所示，假定节点A上磁盘I损坏，导致副本1不可用。当客户端发起的更新请求的数据位于节点A副本1上时，由B节点Master向磁盘I上副本1进行数据更新，节点C的slave从master同步数据，并向节点C磁盘III上副本上数据进行数据更新，这时，数据更新在W＝2个节点上成功完成后，返回给客户端数据更新成功消息；

[0106] 当客户端发起数据访问请求的数据位于节点A副本1上时，节点A直接返回失败，从节点B和节点C的副本1上获得数据后，(满足R＝2)返回给客户端。

[0107] 步骤4，在步骤3情况下，当客户端发起的更新和访问请求位于节点A副本2或者副本3上时，由于三个节点的副本均可用，则处理流程同步骤2。

[0108] 步骤5，如图5所示，当节点B上磁盘II损坏，导致节点B的副本3不可用。当客户端发起的更新和访问请求的数据位于节点A副本1上时，节点B和节点C上的副本均可用，满足NRW策略，则处理流程同步骤3。

[0109] 步骤6，在步骤5情况下，当客户端发起的更新和访问请求位于节点A副本2上时，由于三个节点的副本2均可用，则处理流程同步骤2。

[0110] 步骤7，在步骤5情况下，当客户端发起的更新请求的数据位于节点A副本3上时，B节点的副本3损坏，C节点的副本3可用。由A节点Master向磁盘III上副本3进行数据更新，节点C的slave从master同步数据，并向节点C磁盘II上副本3上数据进行数据更新，这时数据更新在W＝2个节点上成功完成后，返回客户端数据更新成功消息；

[0111] 当客户端发起数据访问请求的数据位于节点A副本3上时，节点B直接返回失败，从节点A和节点C的副本3上获得数据后，(满足R＝2)返回客户端。

[0112] 从上面可以看到，即使节点A和节点B都存在磁盘损坏的情况下，只要损坏磁盘的副本不重复，分布式缓存集群还是可以提供全部数据的读写服务。

[0113] 上述应用场景中，如果有两个故障节点，每个节点实际都是部分磁盘损坏，在较乐观的情况下，如果损坏的磁盘上存放的不是同一个数据的副本，则实际整个系统的可用磁盘上，还是保存着所有数据的至少两个副本，完全具备正常提供所有服务的条件。即使在损坏的磁盘上恰好存放着同一个数据的副本，那么其他磁盘上的可用数据，依然可以满足一致性和可用性，可以提供读写服务，仅对同时损坏的这部分数据而言，无法提供读写访问。

[0114] 本发明的有益效果如下：

[0115] 本发明针对分布式缓存系统，在有磁盘损坏的情况下，可以充分利用可用的资源，整合出符合一致性和可用性要求的副本资源，尽可能提高系统的可用性，提高系统对故障的容忍性。也就是说，在云计算领域分布式缓存系统中，提供一种磁盘和数据管理机制，即使在节点部分磁盘发生故障情况下，依然能够尽可能利用可用磁盘上的数据，保持提供服务的能力，使得服务端在较少的磁盘或数据资源的情况下，提供一致性和可用性的存储服务。

[0116] 以上所述是本发明的优选实施方式，应当指出，对于本技术领域的普通技术人员来说，在不脱离本发明所述原理的前提下，还可以作出若干改进和润饰，这些改进和润饰也应视为本发明的保护范围。

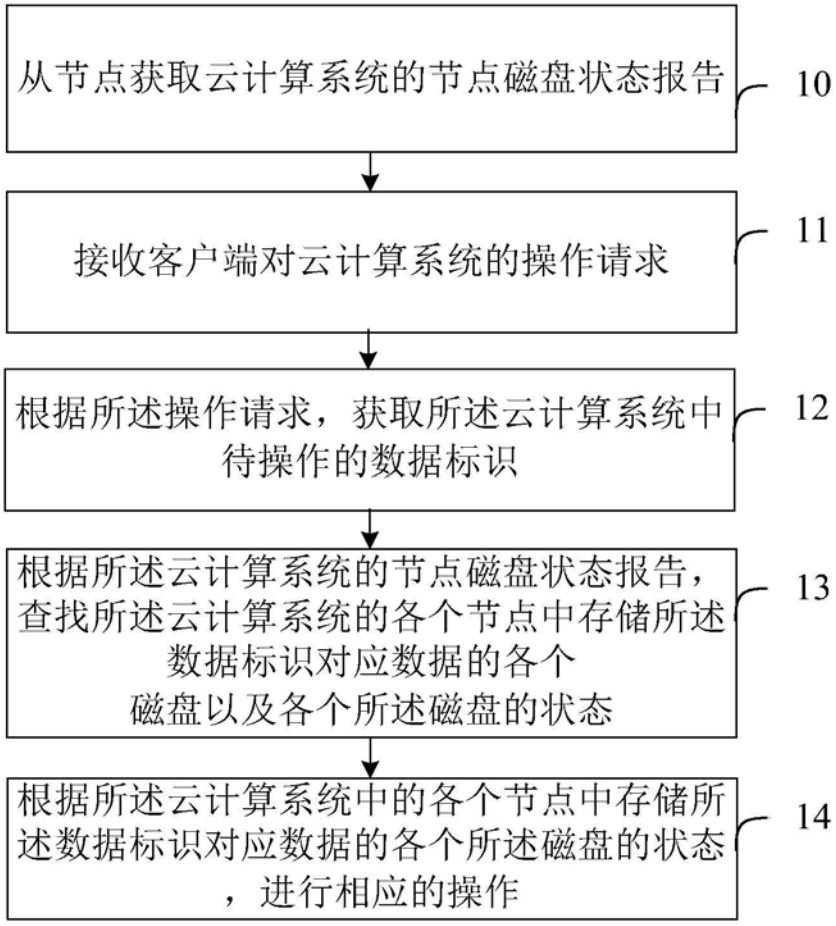

从节点获取云计算系统的节点磁盘状态报告
10
接收客户端对云计算系统的操作请求
11
根据所述操作请求，获取所述云计算系统中待操作的数据标识
12
根据所述云计算系统的节点磁盘状态报告，查找所述云计算系统的各个节点中存储所述数据标识对应数据的各个磁盘以及各个所述磁盘的状态
13
根据所述云计算系统中的各个节点中存储所述数据标识对应数据的各个所述磁盘的状态，进行相应的操作
14

图1

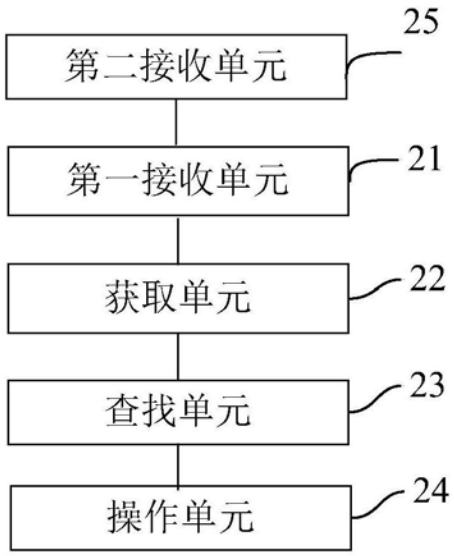

第二接收单元
25
第一接收单元
21
获取单元
22
查找单元
23
操作单元
24

图2

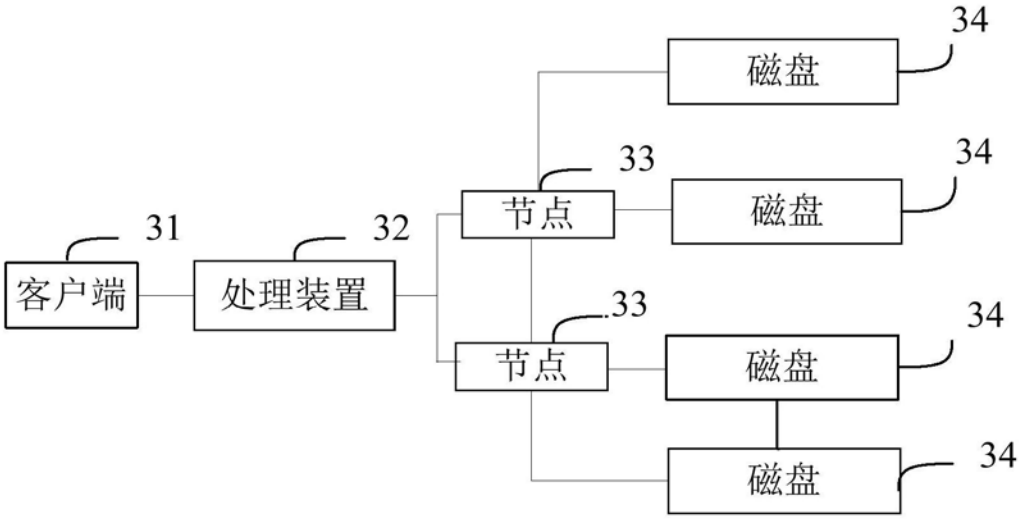

34
磁盘
33
34
节点
磁盘
31
32
客户端
处理装置
33
34
节点
磁盘
34
磁盘

图3

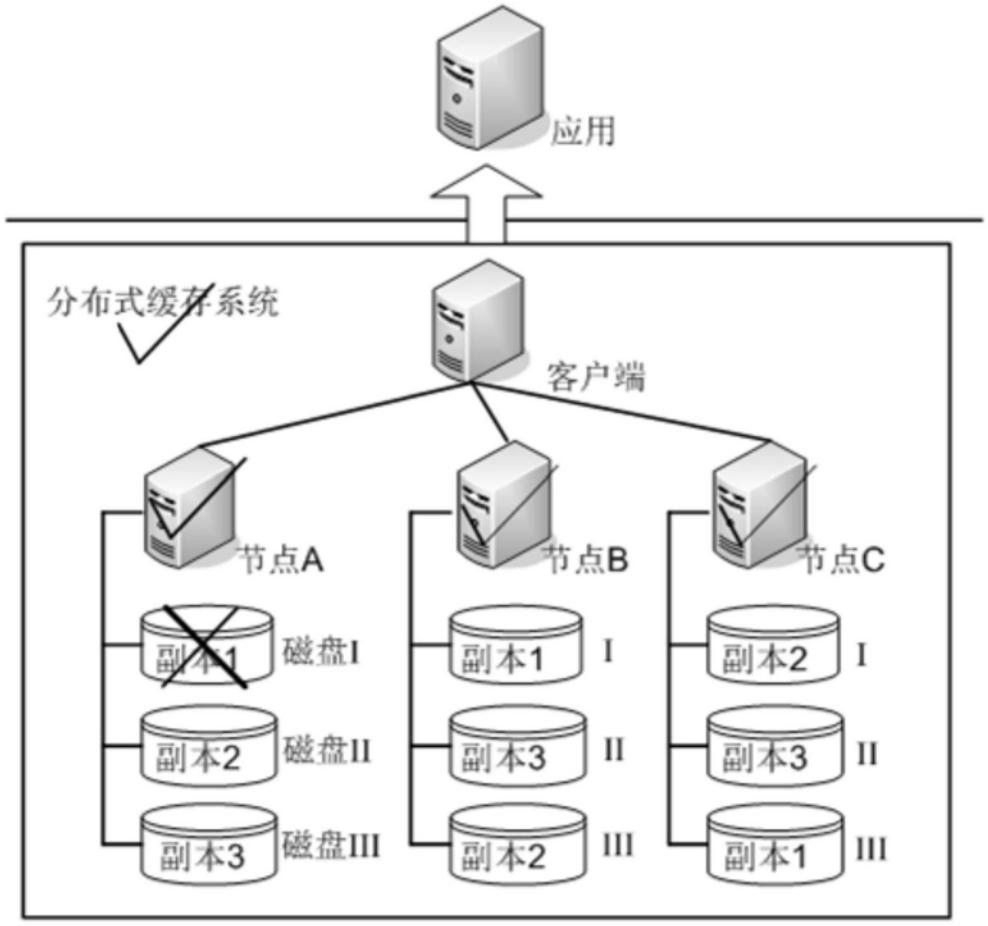

应用
分布式缓存系统
客户端
节点A
节点B
节点C
副本1
磁盘I
副本2
磁盘II
副本3
磁盘III
副本1
I
副本3
II
副本2
III
副本2
I
副本3
II
副本1
III

图4

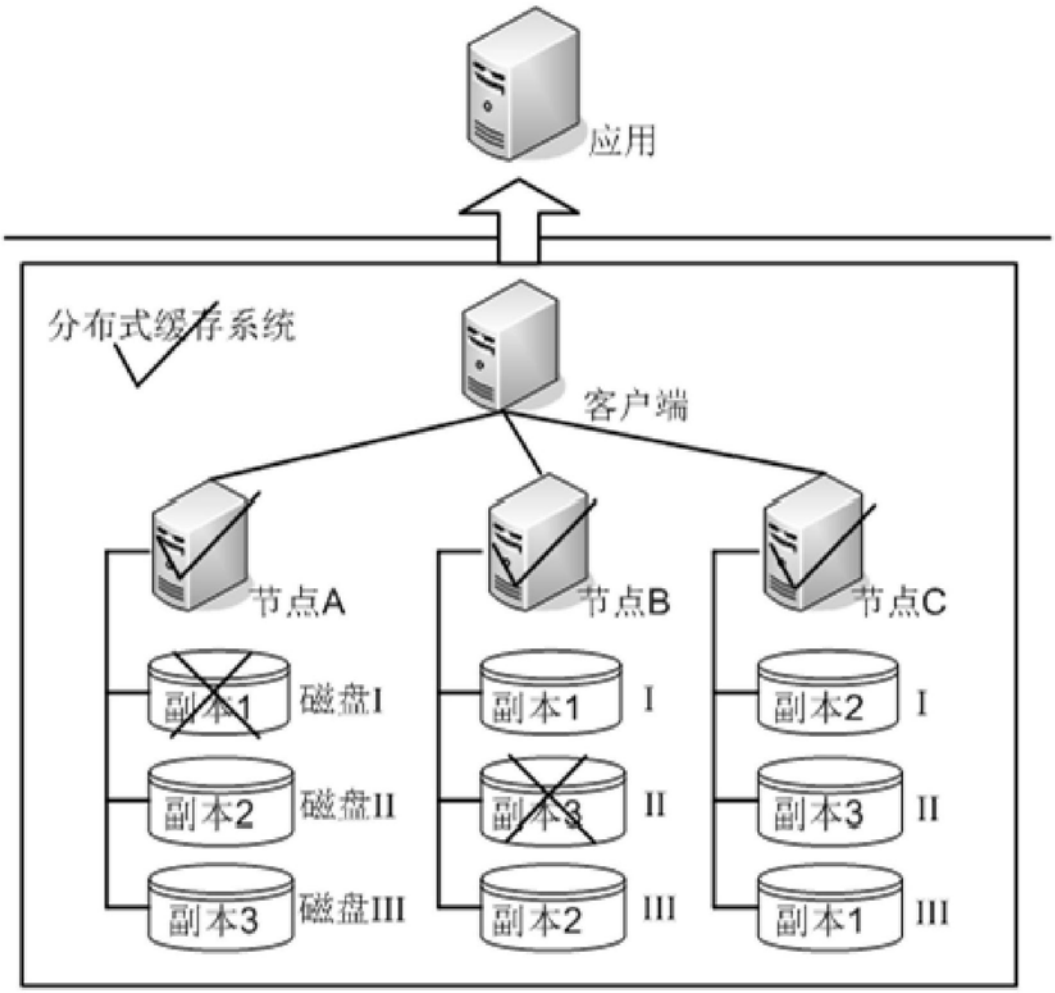

应用
分布式缓存系统
客户端
节点A
节点B
节点C
副本1
磁盘I
副本2
磁盘II
副本3
磁盘III
副本1
I
副本3
II
副本2
III
副本2
I
副本3
II
副本1
III

图5