(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 107168645 B
(45)授权公告日 2020.07.28

(21)申请号 201710218939.9

(22)申请日 2017.03.22

(65)同一申请的已公布的文献号
申请公布号 CN 107168645 A

(43)申请公布日 2017.09.15

(73)专利权人 佛山科学技术学院
地址 528000 广东省佛山市江湾一路18号

(72)发明人 谢建勤 马莉

(51)Int.Cl.
G06F 3/06(2006.01)

(56)对比文件
CN 105530294 A,2016.04.27,
CN 101645039 A,2010.02.10,
CN 103150263 A,2013.06.12,
CN 104050015 A,2014.09.17,
CN 102426545 A,2012.04.25,
CN 103124299 A,2013.05.29,
CN 103116552 A,2013.05.22,
US 2007101082 A1,2007.05.03,
CN 106454959 A,2017.02.22,
马宁.“基于虚拟化的分布式数据容灾保护”.《舰船电子工程》.2008,(第11期),140-143.

审查员 周永翔

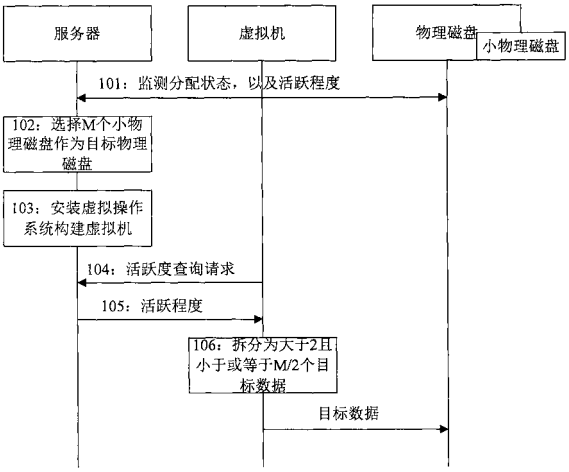
权利要求书3页 说明书9页 附图2页

(54)发明名称

一种分布式系统的存储控制方法及系统

(57)摘要

本发明实施例公开了一种分布式系统的存储控制方法、及系统;应用于包含N个物理磁盘的分布式系统,所述N个物理磁盘中每个物理磁盘的存储空间被分为大小相等的小物理磁盘,每个物理磁盘内的小物理磁盘的序号按照地址从低到高排序;所述N个物理磁盘各自具有大物理磁盘标识,各小物理磁盘具有小物理磁盘标识,所述小物理磁盘标识由所述大物理磁盘标识与所述小物理磁盘的序号组合得到。物理磁盘的标识组成方式,方便后续物理磁盘的查找;虚拟机能够分配到相对较为不活跃的磁盘可以减少拥塞;另外,将需要存储的数据进行了拆分进行数据分配,减少了数据拥塞的可能性,提高数据存储的并行度,可以提高数据存储的安全性。



CN 107168645 B

1.一种分布式系统的存储控制方法,其特征在于,应用于包含N个物理磁盘的分布式系统,所述N个物理磁盘中每个物理磁盘的存储空间被分为大小相等的小物理磁盘,每个物理磁盘内的小物理磁盘的序号按照地址从低到高排序;所述N个物理磁盘各自具有大物理磁盘标识,各小物理磁盘具有小物理磁盘标识,所述小物理磁盘标识由所述大物理磁盘标识与所述小物理磁盘的序号组合得到,所述方法包括:

服务器监测小物理磁盘的分配状态,以及所述N个物理磁盘的活跃程度;

所述服务器确定将要创建的虚拟机的存储空间需求;依据所述小物理磁盘的分配状态确定处于未分配状态的小物理磁盘;从处于未分配状态的小物理磁盘中选择M个小物理磁盘作为目标物理磁盘,所述M个小物理磁盘的存储空间之和满足所述存储空间需求;所述M大于或等于8;所述M个小物理磁盘各自位于不同的物理磁盘;

所述服务器在所述目标物理磁盘中安装虚拟操作系统构建虚拟机;

所述虚拟机启动并运行,在所述虚拟机运行过程中若有数据存储需求,则向所述服务器发送活跃度查询请求,在所述活跃度查询请求中携带所述目标物理磁盘中各小物理磁盘的小物理磁盘标识;

所述服务器在接收到所述活跃度查询请求后,向所述虚拟机返回所述目标物理磁盘中各小物理磁盘所在的物理磁盘的活跃程度;

所述虚拟机将需要存储的数据拆分为大于2且小于或等于M/2个目标数据,按照所述目标物理磁盘中各小物理磁盘所在的物理磁盘的活跃程度从低到高,将各目标数据分别存储到所述目标物理磁盘中的各小物理磁盘。

2.根据权利要求1所述方法,其特征在于,所述大物理磁盘标识为P位的二进制字符串,所述小物理磁盘标识为Q位的二进制字符串;所述小物理磁盘的序号为位于小物理磁盘标识的低位部分,每个小物理磁盘的存储空间为R位;所述方法还包括:

所述虚拟机在确定需要进行访存操作后,确定所述访存操作指定的虚拟地址;所述目标物理磁盘由其包含的各小物理磁盘按照所述各小物理磁盘所在的大物理磁盘标识从低到高依次排序组成,所述虚拟地址以所述目标物理磁盘的起始地址为起始虚拟地址顺序编号获得;在所述虚拟机中存储有地址映射表,所述地址映射表的表项包含:虚拟盘序号、小物理磁盘标识;

所述虚拟机计算所述虚拟地址与所述R的商取整得到所述虚拟地址的虚拟盘序号,计算所述虚拟地址与所述R的商取余得到偏移量;或者,所述虚拟机截取所述虚拟地址的前R位得到所述虚拟盘序号,截取所述虚拟地址的剩余位得到所述偏移量;

所述虚拟机查找所述地址映射表获得包含所述虚拟地址的虚拟盘序号的表项,并确定该表项中包含的小物理磁盘标识作为目标小物理磁盘标识;

所述虚拟机截取所述小物理磁盘标识的前P位作为目标大物理磁盘标识,向所述目标大物理磁盘标识对应的物理磁盘发送读请求,在所述读请求中包含所述小物理磁盘标识以及所述偏移量,使所述小物理磁盘标识对应的小物理磁盘返回在所述小物理磁盘的起始位置偏移所述偏移量对应物理地址的数据。

3.根据权利要求1或2所述方法,其特征在于,在所述虚拟机被创建之后,所述方法还包括:

所述服务器接收虚拟机删除请求,所述虚拟机删除请求用于请求删除所述虚拟机;

所述服务器将所述目标物理磁盘中包含的各小物理磁盘的分配状态设置为未分配状态，不删除所述目标物理磁盘中包含的各小物理磁盘已经被写入的数据。

4.根据权利要求3所述方法，其特征在于，在所述服务器将所述目标物理磁盘中包含的各小物理磁盘的分配状态设置为未分配状态之后，所述方法还包括：

所述服务器记录所述目标物理磁盘中包含的各小物理磁盘，在下一次创建新虚拟机时，以随机方式获取所述新虚拟机所需的小物理磁盘，获取到的小物理磁盘中少于或等于两个小物理磁盘属于所述目标物理磁盘中包含的小物理磁盘。

5.一种分布式存储系统，包括：服务器、虚拟机和N个物理磁盘；所述N个物理磁盘中每个物理磁盘的存储空间被分为大小相等的小物理磁盘，每个物理磁盘内的小物理磁盘的序号按照地址从低到高排序；所述N个物理磁盘各自具有大物理磁盘标识，各小物理磁盘具有小物理磁盘标识，所述小物理磁盘标识由所述大物理磁盘标识与所述小物理磁盘的序号组合得到，其特征在于，

所述服务器，用于监测小物理磁盘的分配状态，以及所述N个物理磁盘的活跃程度；确定将要创建的虚拟机的存储空间需求；依据所述小物理磁盘的分配状态确定处于未分配状态的小物理磁盘；从处于未分配状态的小物理磁盘中选择M个小物理磁盘作为目标物理磁盘，所述M个小物理磁盘的存储空间之和满足所述存储空间需求；所述M大于或等于8；所述M个小物理磁盘各自位于不同的物理磁盘；在所述目标物理磁盘中安装虚拟操作系统构建虚拟机；

所述虚拟机，用于启动并运行，在所述虚拟机运行过程中若有数据存储需求，则向所述服务器发送活跃度查询请求，在所述活跃度查询请求中携带所述目标物理磁盘中各小物理磁盘的小物理磁盘标识；

所述服务器，还用于在接收到所述活跃度查询请求后，向所述虚拟机返回所述目标物理磁盘中各小物理磁盘所在的物理磁盘的活跃程度；将需要存储的数据拆分为大于2且小于或等于M/2个目标数据，按照所述目标物理磁盘中各小物理磁盘所在的物理磁盘的活跃程度从低到高，将各目标数据分别存储到所述目标物理磁盘中的各小物理磁盘。

6.根据权利要求5所述系统，其特征在于，所述大物理磁盘标识为P位的二进制字符串，所述小物理磁盘标识为Q位的二进制字符串；所述小物理磁盘的序号为位于小物理磁盘标识的低位部分，每个小物理磁盘的存储空间为R位；

所述虚拟机，还用于在确定需要进行访存操作后，确定所述访存操作指定的虚拟地址；所述目标物理磁盘由其包含的各小物理磁盘按照所述各小物理磁盘所在的大物理磁盘标识从低到高依次排序组成，所述虚拟地址以所述目标物理磁盘的起始地址为起始虚拟地址顺序编号获得；在所述虚拟机中存储有地址映射表，所述地址映射表的表项包含：虚拟盘序号、小物理磁盘标识；计算所述虚拟地址与所述R的商取整得到所述虚拟地址的虚拟盘序号，计算所述虚拟地址与所述R的商取余得到偏移量；或者，截取所述虚拟地址的前R位得到所述虚拟盘序号，截取所述虚拟地址的剩余位得到所述偏移量；查找所述地址映射表获得包含所述虚拟地址的虚拟盘序号的表项，并确定该表项中包含的小物理磁盘标识作为目标小物理磁盘标识；截取所述小物理磁盘标识的前P位作为目标大物理磁盘标识，向所述目标大物理磁盘标识对应的物理磁盘发送读请求，在所述读请求中包含所述小物理磁盘标识以及所述偏移量，使所述小物理磁盘标识对应的小物理磁盘返回在所述小物理磁盘的起始位

置偏移所述偏移量对应物理地址的数据。

7.根据权利要求5或6所述系统,其特征在于,

所述服务器,还用于在所述虚拟机被创建之后,接收虚拟机删除请求,所述虚拟机删除请求用于请求删除所述虚拟机;将所述目标物理磁盘中包含的各小物理磁盘的分配状态设置为未分配状态,不删除所述目标物理磁盘中包含的各小物理磁盘已经被写入的数据。

8.根据权利要求7所述系统,其特征在于,

所述服务器,还用于在将所述目标物理磁盘中包含的各小物理磁盘的分配状态设置为未分配状态之后,记录所述目标物理磁盘中包含的各小物理磁盘,在下一次创建新虚拟机时,以随机方式获取所述新虚拟机所需的小物理磁盘,获取到的小物理磁盘中少于或等于两个小物理磁盘属于所述目标物理磁盘中包含的小物理磁盘。

一种分布式系统的存储控制方法及系统

技术领域

[0001] 本发明涉及信息技术领域,特别涉及一种分布式系统的存储控制方法及系统。

背景技术

[0002] 分布式存储系统,是将数据分散存储在多台独立的设备上。

[0003] 例如:在一个视频监控系统中,选择什么样的存储解决方案直接决定了整个系统的系统架构以及系统的性能和稳定程度。

[0004] 传统的网络存储系统采用集中的存储服务器存放所有数据,存储服务器成为系统性能的瓶颈,也是可靠性和安全性的焦点,不能满足大规模存储应用的需要。分布式网络存储系统采用可扩展的系统结构,利用多台存储服务器分担存储负荷,利用位置服务器定位存储信息,它不但提高了系统的可靠性、可用性和存取效率,还易于扩展。

[0005] 在分布式系统存储系统中,需要将数据存储到多个独立的设备上,在这个过程中,经常出现拥塞导致存储速度慢,因此整个数据存储效率较低。

发明内容

[0006] 本发明实施例提供了一种分布式系统的存储控制方法、及系统,用于减少数据拥塞的几率,并且提高数据存储的安全性。

[0007] 一方面本发明实施例提供了一种分布式系统的存储控制方法,应用于包含N个物理磁盘的分布式系统,所述N个物理磁盘中每个物理磁盘的存储空间被分为大小相等的小物理磁盘,每个物理磁盘内的小物理磁盘的序号按照地址从低到高排序;所述N个物理磁盘各自具有大物理磁盘标识,各小物理磁盘具有小物理磁盘标识,所述小物理磁盘标识由所述大物理磁盘标识与所述小物理磁盘的序号组合得到,所述方法包括:

[0008] 服务器监测小物理磁盘的分配状态,以及所述N个物理磁盘的活跃程度;

[0009] 所述服务器确定将要创建的虚拟机的存储空间需求;依据所述小物理磁盘的分配状态确定处于未分配状态的小物理磁盘;从处于未分配状态的小物理磁盘中选择M个小物理磁盘作为目标物理磁盘,所述M个小物理磁盘的存储空间之和满足所述存储空间需求;所述M大于或等于8;所述M个小物理磁盘各自位于不同的物理磁盘;

[0010] 所述服务器在所述目标物理磁盘中安装虚拟操作系统构建虚拟机;

[0011] 所述虚拟机启动并运行,在所述虚拟机运行过程中若有数据存储需求,则向所述服务器发送活跃度查询请求,在所述活跃度查询请求中携带所述目标物理磁盘中各小物理磁盘的小物理磁盘标识;

[0012] 所述服务器在接收到所述活跃度查询请求后,向所述虚拟机返回所述目标物理磁盘中各小物理磁盘所在的物理磁盘的活跃程度;

[0013] 所述虚拟机将需要存储的数据拆分为大于2且小于或等于M/2个目标数据,按照所述目标物理磁盘中各小物理磁盘所在的物理磁盘的活跃程度从低到高,将各目标数据分别存储到所述目标物理磁盘中的各小物理磁盘。

[0014] 在一个可选的实现方式中，所述大物理磁盘标识为P位的二进制字符串，所述小物理磁盘标识为Q位的二进制字符串；所述小物理磁盘的序号为位于小磁盘标识的低位部分，每个小物理磁盘的存储空间为R位；所述方法还包括：

[0015] 所述虚拟机在确定需要进行访存操作后，确定所述访存操作指定的虚拟地址；所述目标物理磁盘由其包含的各小物理磁盘按照所述各小物理磁盘所在的大物理磁盘标识从低到高依次排序组成，所述虚拟地址以所述目标物理磁盘的起始地址为起始虚拟地址顺序编号获得；在所述虚拟机中存储有地址映射表，所述地址映射表的表项包含：虚拟盘序号、小物理磁盘标识；

[0016] 所述虚拟机计算所述虚拟地址与所述R的商取整得到所述虚拟地址的虚拟盘序号，计算所述虚拟地址与所述R的商取余得到偏移量；

[0017] 所述虚拟机查找所述地址映射表获得包含所述虚拟地址的虚拟盘序号的表项，并确定该表项中包含的小物理磁盘标识作为目标小物理磁盘标识；

[0018] 所述虚拟机截取所述小物理磁盘标识的前P位作为目标大物理磁盘标识，向所述目标大物理磁盘标识对应的物理磁盘发送读请求，在所述读请求中包含所述小物理磁盘标识以及所述偏移量，使所述小物理磁盘标识对应的小物理磁盘返回在所述小物理磁盘的起始位置偏移所述偏移量对应物理地址的数据。

[0019] 在一个可选的实现方式中，所述虚拟机计算所述虚拟地址与所述R的商取整得到所述虚拟地址的虚拟盘序号，计算所述虚拟地址与所述R的商取余得到偏移量包括：

[0020] 所述虚拟机截取所述虚拟地址的前R位得到所述虚拟盘序号，截取所述虚拟地址的剩余位得到所述偏移量。

[0021] 在一个可选的实现方式中，在所述虚拟机被创建之后，所述方法还包括：

[0022] 所述服务器接收虚拟机删除请求，所述虚拟机删除请求用于请求删除所述虚拟机；

[0023] 所述服务器将所述目标物理磁盘中包含的各小物理磁盘的分配状态设置为未分配状态，不删除所述目标物理磁盘中包含的各小物理磁盘已经被写入的数据。

[0024] 在一个可选的实现方式中，在所述服务器将所述目标物理磁盘中包含的各小物理磁盘的分配状态设置为未分配状态之后，所述方法还包括：

[0025] 所述服务器记录所述目标物理磁盘中包含的各小物理磁盘，在下一次创建新虚拟机时，以随机方式获取所述新虚拟机所需的小物理磁盘，并确定获取到的小物理磁盘中少于或等于两个小物理磁盘属于所述目标物理磁盘中包含的小物理磁盘。

[0026] 二方面本发明实施例还提供了一种分布式存储系统，包括：服务器、虚拟机和N个物理磁盘；所述N个物理磁盘中每个物理磁盘的存储空间被分为大小相等的小物理磁盘，每个物理磁盘内的小物理磁盘的序号按照地址从低到高排序；所述N个物理磁盘各自具有大物理磁盘标识，各小物理磁盘具有小物理磁盘标识，所述小物理磁盘标识由所述大物理磁盘标识与所述小物理磁盘的序号组合得到；

[0027] 所述服务器，用于监测小物理磁盘的分配状态，以及所述N个物理磁盘的活跃程度；确定将要创建的虚拟机的存储空间需求；依据所述小物理磁盘的分配状态确定处于未分配状态的小物理磁盘；从处于未分配状态的小物理磁盘中选择M个小物理磁盘作为目标物理磁盘，所述M个小物理磁盘的存储空间之和满足所述存储空间需求；所述M大于或等于

8;所述M个小物理磁盘各自位于不同的物理磁盘;在所述目标物理磁盘中安装虚拟操作系统构建虚拟机;

[0028] 所述虚拟机,用于启动并运行,在所述虚拟机运行过程中若有数据存储需求,则向所述服务器发送活跃度查询请求,在所述活跃度查询请求中携带所述目标物理磁盘中各小物理磁盘的小物理磁盘标识;

[0029] 所述服务器,还用于在接收到所述活跃度查询请求后,向所述虚拟机返回所述目标物理磁盘中各小物理磁盘所在的物理磁盘的活跃程度;将需要存储的数据拆分为大于2且小于或等于M/2个目标数据,按照所述目标物理磁盘中各小物理磁盘所在的物理磁盘的活跃程度从低到高,将各目标数据分别存储到所述目标物理磁盘中的各小物理磁盘。

[0030] 在一个可选的实现方式中,所述大物理磁盘标识为P位的二进制字符串,所述小物理磁盘标识为Q位的二进制字符串;所述小物理磁盘的序号为位于小磁盘标识的低位部分,每个小物理磁盘的存储空间为R位;

[0031] 所述虚拟机,还用于在确定需要进行访存操作后,确定所述访存操作指定的虚拟地址;所述目标物理磁盘由其包含的各小物理磁盘按照所述各小物理磁盘所在的大物理磁盘标识从低到高依次排序组成,所述虚拟地址以所述目标物理磁盘的起始地址为起始虚拟地址顺序编号获得;在所述虚拟机中存储有地址映射表,所述地址映射表的表项包含:虚拟盘序号、小物理磁盘标识;计算所述虚拟地址与所述R的商取整得到所述虚拟地址的虚拟盘序号,计算所述虚拟地址与所述R的商取余得到偏移量;查找所述地址映射表获得包含所述虚拟地址的虚拟盘序号的表项,并确定该表项中包含的小物理磁盘标识作为目标小物理磁盘标识;截取所述小物理磁盘标识的前P位作为目标大物理磁盘标识,向所述目标大物理磁盘标识对应的物理磁盘发送读请求,在所述读请求中包含所述小物理磁盘标识以及所述偏移量,使所述小物理磁盘标识对应的小物理磁盘返回在所述小物理磁盘的起始位置偏移所述偏移量对应物理地址的数据。

[0032] 在一个可选的实现方式中,所述虚拟机,用于计算所述虚拟地址与所述R的商取整得到所述虚拟地址的虚拟盘序号,计算所述虚拟地址与所述R的商取余得到偏移量包括:

[0033] 截取所述虚拟地址的前R位得到所述虚拟盘序号,截取所述虚拟地址的剩余位得到所述偏移量。

[0034] 在一个可选的实现方式中,所述服务器,还用于在所述虚拟机被创建之后,接收虚拟机删除请求,所述虚拟机删除请求用于请求删除所述虚拟机;将所述目标物理磁盘中包含的各小物理磁盘的分配状态设置为未分配状态,不删除所述目标物理磁盘中包含的各小物理磁盘已经被写入的数据。

[0035] 在一个可选的实现方式中,所述服务器,还用于在将所述目标物理磁盘中包含的各小物理磁盘的分配状态设置为未分配状态之后,记录所述目标物理磁盘中包含的各小物理磁盘,在下一次创建新虚拟机时,以随机方式获取所述新虚拟机所需的小物理磁盘,并确定获取到的小物理磁盘中少于或等于两个小物理磁盘属于所述目标物理磁盘中包含的小物理磁盘。

[0036] 从以上技术方案可以看出,本发明实施例具有以下优点:特别设定了物理磁盘的标识组成方式,方便后续物理磁盘的查找;另外,虚拟机的物理磁盘分配过程中充分考虑了各物理磁盘的活跃度,使得虚拟机能够分配到较为合适的物理磁盘,相对较为不活跃的磁

盘可以减少拥塞;另外,将需要存储的数据进行了拆分,按照物理磁盘的活跃度再次进行数据分配,一方面进一步减少了数据拥塞的可能性,提高数据存储的并行度,另外还能够降低数据被整体存储到同一小物理磁盘导致可能被窃取的可能性,因此可以提高数据存储的安全性。

附图说明

[0037] 为了更清楚地说明本发明实施例中的技术方案,下面将对实施例描述中所需要使用的附图作简要介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域的普通技术人员来讲,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。

[0038] 图1为本发明实施例方法流程示意图;

[0039] 图2为本发明实施例小物理磁盘标识组成结构示意图;

[0040] 图3为本发明实施例系统结构示意图。

具体实施方式

[0041] 为了使本发明的目的、技术方案和优点更加清楚,下面将结合附图对本发明作进一步地详细描述,显然,所描述的实施例仅仅是本发明一部份实施例,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它实施例,都属于本发明保护的范围。

[0042] 本发明实施例提供了一种分布式系统的存储控制方法,应用于包含N个物理磁盘的分布式系统,上述N个物理磁盘中每个物理磁盘的存储空间被分为大小相等的小物理磁盘,每个物理磁盘内的小物理磁盘的序号按照地址从低到高排序;上述N个物理磁盘各自具有大物理磁盘标识,各小物理磁盘具有小物理磁盘标识,上述小物理磁盘标识由上述大物理磁盘标识与上述小物理磁盘的序号组合得到,如图1所示,包括:

[0043] 小物理磁盘标识的结构请参阅图2所示;

[0044] 101:服务器监测小物理磁盘的分配状态,以及上述N个物理磁盘的活跃程度;

[0045] 活跃程度可以用物理磁盘当前或者综合历史数据统计得到的平均数据吞吐量,或者平均数据吞吐量占对应物理磁盘的数据存储能力的比例。越活跃的物理磁盘其数据存储压力越大,形成拥塞的可能性也将会越大。

[0046] 102:确定将要创建的虚拟机的存储空间需求;依据上述小物理磁盘的分配状态确定处于未分配状态的小物理磁盘;从处于未分配状态的小物理磁盘中选择M个小物理磁盘作为目标物理磁盘,上述M个小物理磁盘的存储空间之和满足上述存储空间需求;上述M大于或等于8;上述M个小物理磁盘各自位于不同的物理磁盘;

[0047] 服务器可以在接收到虚拟机创建请求后执行上述步骤102,其中虚拟机创建请求可以是任意设备发出的,假定我们的系统应用于大型公司员工的虚拟机创建,那么可以是管理者发出的。对于不同的虚拟机可能会有不同的存储空间需求,例如:做业务的员工和做软件开发的员工,对存储空间的需求是不一样的。在本实施例中,可以假定每个小物理磁盘是300M,假定需要3000M的存储空间,那么可以分成10个小物理磁盘。这10个小物理磁盘从那些相对较为空闲的物理磁盘中选取。由于不同的虚拟机被使用的可能性是不同的,因此

通过对小物理磁盘的选择性分配可以达到第一次均衡。

[0048] 103:在上述目标物理磁盘中安装虚拟操作系统构建虚拟机;

[0049] 在虚拟机被创建以后,虚拟机的操作系统被安装,那么将会成为一个真正的虚拟机。虚拟机会获知自己被分配的目标物理磁盘,以及这些目标物理磁盘所处的位置。

[0050] 104:上述虚拟机启动并运行,在上述虚拟机运行过程中若有数据存储需求,则向上述服务器发送活跃度查询请求,在上述活跃度查询请求中携带上述目标物理磁盘中各小物理磁盘的小物理磁盘标识;

[0051] 105:上述服务器在接收到上述活跃度查询请求后,向上述虚拟机返回上述目标物理磁盘中各小物理磁盘所在的物理磁盘的活跃程度;

[0052] 106:上述虚拟机将需要存储的数据拆分为大于2且小于或等于M/2个目标数据,按照上述目标物理磁盘中各小物理磁盘所在的物理磁盘的活跃程度从低到高,将各目标数据分别存储到上述目标物理磁盘中的各小物理磁盘。

[0053] 本步骤中,设定目标数据的数量,一方面可以保持数据被分的目标数据较多,提高安全性和存储并行度;另一方面考虑到需要存储到那些比较空闲的物理磁盘中,减少出现拥塞的可能性。

[0054] 在本发明实施例中,特别设定了物理磁盘的标识组成方式,方便后续物理磁盘的查找;另外,虚拟机的物理磁盘分配过程中充分考虑了各物理磁盘的活跃度,使得虚拟机能够分配到较为合适的物理磁盘,相对较为不活跃的磁盘可以减少拥塞;另外,将需要存储的数据进行了拆分,按照物理磁盘的活跃度再次进行数据分配,一方面进一步减少了数据拥塞的可能性,提高数据存储的并行度,另外还能够降低数据被整体存储到同一小物理磁盘导致可能被窃取的可能性,因此可以提高数据存储的安全性。

[0055] 优选地,如图2所示,上述大物理磁盘标识为P位的二进制字符串,上述小物理磁盘标识为Q位的二进制字符串;上述小物理磁盘的序号为位于小磁盘标识的低位部分,每个小物理磁盘的存储空间为R位;上述方法还包括:

[0056] 上述虚拟机在确定需要进行访存操作后,确定上述访存操作指定的虚拟地址;上述目标物理磁盘由其包含的各小物理磁盘按照上述各小物理磁盘所在的大物理磁盘标识从低到高依次排序组成,上述虚拟地址以上述目标物理磁盘的起始地址为起始虚拟地址顺序编号获得;在上述虚拟机中存储有地址映射表,上述地址映射表的表项包含:虚拟盘序号、小物理磁盘标识;

[0057] 上述虚拟机计算上述虚拟地址与上述R的商取整得到上述虚拟地址的虚拟盘序号,计算上述虚拟地址与上述R的商取余得到偏移量;

[0058] 上述虚拟机查找上述地址映射表获得包含上述虚拟地址的虚拟盘序号的表项,并确定该表项中包含的小物理磁盘标识作为目标小物理磁盘标识;

[0059] 上述虚拟机截取上述小物理磁盘标识的前P位作为目标大物理磁盘标识,向上述目标大物理磁盘标识对应的物理磁盘发送读请求,在上述读请求中包含上述小物理磁盘标识以及上述偏移量,使上述小物理磁盘标识对应的小物理磁盘返回在上述小物理磁盘的起始位置偏移上述偏移量对应物理地址的数据。

[0060] 在本实施例中,设定了一个特别的大物理磁盘标识、小物理磁盘标识,这样可以设定一个地址映射表,方便后续迅速查到到对应的物理磁盘。由于在目标物理磁盘中,虚拟机

会认为该目标物理磁盘是一个真实的物理磁盘，这样地址应该是目标物理磁盘中是连续的，然而实际上目标物理磁盘中的存储空间位于不同的物理磁盘，因此物理地址实际上是不同的；因此需要对虚拟地址进行转换；虚拟地址的使用是为了在虚拟机中方便应用，例如：软件编程等。虚拟地址是把上述目标物理磁盘看作是一个整体的物理磁盘后获得的地址，因为该虚拟地址并不与实际的物理磁盘地址对应，因此称为虚拟地址。通过本发明实施例方案，可以迅速查到的对应的物理磁盘和对应的物理地址，因此可以快速存储数据，相应地，也可以快速读取数据。

[0061] 进一步地，鉴于本发明实施例所设定的特殊映射表，本发明实施例可以使用如下方式进行计算：上述虚拟机计算上述虚拟地址与上述R的商取整得到上述虚拟地址的虚拟盘序号，计算上述虚拟地址与上述R的商取余得到偏移量包括：

[0062] 上述虚拟机截取上述虚拟地址的前R位得到上述虚拟盘序号，截取上述虚拟地址的剩余位得到上述偏移量。

[0063] 本发明实施例使用截取的方式得到结果，可以减少大量的逻辑运算，因此可以减少运算量，提高数据存储的效率。

[0064] 进一步地，本发明实施例还提供了虚拟机删除的方案，如下：在上述虚拟机被创建之后，上述方法还包括：

[0065] 上述服务器接收虚拟机删除请求，上述虚拟机删除请求用于请求删除上述虚拟机；

[0066] 上述服务器将上述目标物理磁盘中包含的各小物理磁盘的分配状态设置为未分配状态，不删除上述目标物理磁盘中包含的各小物理磁盘已经被写入的数据。

[0067] 在本发明实施例中，由于数据存储的方式是将数据拆分了存储，这样安全性较高，在删除虚拟机的时候，可以仅标记小物理磁盘的分配状态，不进行数据删除操作；一方面可以保证数据安全性，另一方面还可以减少物理磁盘的擦写次数，提高物理磁盘的寿命。

[0068] 本发明实施例还提供了后续再次分配物理磁盘的可选实现方案，如下：在上述服务器将上述目标物理磁盘中包含的各小物理磁盘的分配状态设置为未分配状态之后，上述方法还包括：

[0069] 上述服务器记录上述目标物理磁盘中包含的各小物理磁盘，在下一次创建新虚拟机时，以随机方式获取上述新虚拟机所需的小物理磁盘，并确定获取到的小物理磁盘中少于或等于两个小物理磁盘属于上述目标物理磁盘中包含的小物理磁盘。

[0070] 采用本实施例方案，可以进一步提高数据安全性。这是基于数据被存储到多个物理磁盘，虽然这些小物理磁盘中的数据并不具有连续性，但是如果这些物理磁盘又被分配到同一虚拟机，鉴于本发明实施例所使用的特殊小物理磁盘组成目标物理磁盘的方案，则有可能被恢复出来；为了避免这种情况的发生，提出了本实施例实现方案。

[0071] 本发明实施例还提供了一种分布式存储系统，如图3所示，可以一并参阅图1所示，包括：服务器、虚拟机和N个物理磁盘；上述N个物理磁盘中每个物理磁盘的存储空间被分为大小相等的小物理磁盘，每个物理磁盘内的小物理磁盘的序号按照地址从低到高排序；上述N个物理磁盘各自具有大物理磁盘标识，各小物理磁盘具有小物理磁盘标识，上述小物理磁盘标识由上述大物理磁盘标识与上述小物理磁盘的序号组合得到，其特征在于，

[0072] 上述服务器，用于监测小物理磁盘的分配状态，以及上述N个物理磁盘的活跃程

度;确定将要创建的虚拟机的存储空间需求;依据上述小物理磁盘的分配状态确定处于未分配状态的小物理磁盘;从处于未分配状态的小物理磁盘中选择M个小物理磁盘作为目标物理磁盘,上述M个小物理磁盘的存储空间之和满足上述存储空间需求;上述M大于或等于8;上述M个小物理磁盘各自位于不同的物理磁盘;在上述目标物理磁盘中安装虚拟操作系统构建虚拟机;

[0073] 上述虚拟机,用于启动并运行,在上述虚拟机运行过程中若有数据存储需求,则向上述服务器发送活跃度查询请求,在上述活跃度查询请求中携带上述目标物理磁盘中各小物理磁盘的小物理磁盘标识;

[0074] 上述服务器,还用于在接收到上述活跃度查询请求后,向上述虚拟机返回上述目标物理磁盘中各小物理磁盘所在的物理磁盘的活跃程度;将需要存储的数据拆分为大于2且小于或等于M/2个目标数据,按照上述目标物理磁盘中各小物理磁盘所在的物理磁盘的活跃程度从低到高,将各目标数据分别存储到上述目标物理磁盘中的各小物理磁盘。

[0075] 小物理磁盘标识的结构请参阅图2所示;

[0076] 活跃程度可以用物理磁盘当前或者综合历史数据统计得到的平均数据吞吐量,或者平均数据吞吐量占对应物理磁盘的数据存储能力的比例。越活跃的物理磁盘其数据存储压力越大,形成拥塞的可能性也将会越大。

[0077] 服务器可以在接收到虚拟机创建请求后执行上述“确定将要创建的虚拟机的存储空间需求”及其之后的步骤,其中,虚拟机创建请求可以是任意设备发出的,假定我们的系统应用于大型公司员工的虚拟机创建,那么可以是管理者发出的。对于不同的虚拟机可能会有不同的存储空间需求,例如:做业务的员工和做软件开发的员工,对存储空间的需求是不一样的。在本实施例中,可以假定每个小物理磁盘是300M,假定需要3000M的存储空间,那么可以分成10个小物理磁盘。这10个小物理磁盘从那些相对较为空闲的物理磁盘中选取。由于不同的虚拟机被使用的可能性是不同的,因此通过对小物理磁盘的选择性分配可以达到第一次均衡。

[0078] 在虚拟机被创建以后,虚拟机的操作系统被安装,那么将会成为一个真正的虚拟机。虚拟机会获知自己被分配的目标物理磁盘,以及这些目标物理磁盘所处的位置。

[0079] 本实施例中,设定目标数据的数量,一方面可以保持数据被分的目标数据较多,提高安全性和存储并行度;另一方面考虑到需要存储到那些比较空闲的物理磁盘中,减少出现拥塞的可能性。

[0080] 在本发明实施例中,特别设定了物理磁盘的标识组成方式,方便后续物理磁盘的查找;另外,虚拟机的物理磁盘分配过程中充分考虑了各物理磁盘的活跃度,使得虚拟机能够分配到较为合适的物理磁盘,相对较为不活跃的磁盘可以减少拥塞;另外,将需要存储的数据进行了拆分,按照物理磁盘的活跃度再次进行数据分配,一方面进一步减少了数据拥塞的可能性,提高数据存储的并行度,另外还能够降低数据被整体存储到同一小物理磁盘导致可能被窃取的可能性,因此可以提高数据存储的安全性。

[0081] 优选地,如图2所示,上述大物理磁盘标识为P位的二进制字符串,上述小物理磁盘标识为Q位的二进制字符串;上述小物理磁盘的序号为位于小磁盘标识的低位部分,每个小物理磁盘的存储空间为R位;

[0082] 上述虚拟机,还用于在确定需要进行访存操作后,确定上述访存操作指定的虚拟

地址;上述目标物理磁盘由其包含的各小物理磁盘按照上述各小物理磁盘所在的大物理磁盘标识从低到高依次排序组成,上述虚拟地址以上述目标物理磁盘的起始地址为起始虚拟地址顺序编号获得;在上述虚拟机中存储有地址映射表,上述地址映射表的表项包含:虚拟盘序号、小物理磁盘标识;计算上述虚拟地址与上述R的商取整得到上述虚拟地址的虚拟盘序号,计算上述虚拟地址与上述R的商取余得到偏移量;查找上述地址映射表获得包含上述虚拟地址的虚拟盘序号的表项,并确定该表项中包含的小物理磁盘标识作为目标小物理磁盘标识;截取上述小物理磁盘标识的前P位作为目标大物理磁盘标识,向上述目标大物理磁盘标识对应的物理磁盘发送读请求,在上述读请求中包含上述小物理磁盘标识以及上述偏移量,使上述小物理磁盘标识对应的小物理磁盘返回在上述小物理磁盘的起始位置偏移上述偏移量对应物理地址的数据。

[0083] 在本实施例中,设定了一个特别的大物理磁盘标识、小物理磁盘标识,这样可以设定一个地址映射表,方便后续迅速查到到对应的物理磁盘。由于在目标物理磁盘中,虚拟机会认为该目标物理磁盘是一个真实的物理磁盘,这样地址应该是目标物理磁盘中是连续的,然而实际上目标物理磁盘中的存储空间位于不同的物理磁盘,因此物理地址实际上是不同的;因此需要对虚拟地址进行转换;虚拟地址的使用是为了在虚拟机中方便应用,例如:软件编程等。虚拟地址是把上述目标物理磁盘看作是一个整体的物理磁盘后获得的地址,因为该虚拟地址并不与实际的物理磁盘地址对应,因此称为虚拟地址。通过本发明实施例方案,可以迅速查到的对应的物理磁盘和对应的物理地址,因此可以快速存储数据,相应地,也可以快速读取数据。

[0084] 进一步地,鉴于本发明实施例所设定的特殊映射表,本发明实施例可以使用如下方式进行计算:上述虚拟机,用于计算上述虚拟地址与上述R的商取整得到上述虚拟地址的虚拟盘序号,计算上述虚拟地址与上述R的商取余得到偏移量包括:

[0085] 截取上述虚拟地址的前R位得到上述虚拟盘序号,截取上述虚拟地址的剩余位得到上述偏移量。

[0086] 本发明实施例使用截取的方式得到结果,可以减少大量的逻辑运算,因此可以减少运算量,提高数据存储的效率。

[0087] 进一步地,本发明实施例还提供了虚拟机删除的方案,如下:上述服务器,还用于在上述虚拟机被创建之后,接收虚拟机删除请求,上述虚拟机删除请求用于请求删除上述虚拟机;将上述目标物理磁盘中包含的各小物理磁盘的分配状态设置为未分配状态,不删除上述目标物理磁盘中包含的各小物理磁盘已经被写入的数据。

[0088] 在本发明实施例中,由于数据存储的方式是将数据拆分了存储,这样安全性较高,在删除虚拟机的时候,可以仅标记小物理磁盘的分配状态,不进行数据删除操作;一方面可以保证数据安全性,另一方面还可以减少物理磁盘的擦写次数,提高物理磁盘的寿命。

[0089] 本发明实施例还提供了后续再次分配物理磁盘的可选实现方案,如下:上述服务器,还用于在将上述目标物理磁盘中包含的各小物理磁盘的分配状态设置为未分配状态之后,记录上述目标物理磁盘中包含的各小物理磁盘,在下一次创建新虚拟机时,以随机方式获取上述新虚拟机所需的小物理磁盘,并确定获取到的小物理磁盘中少于或等于两个小物理磁盘属于上述目标物理磁盘中包含的小物理磁盘。

[0090] 采用本实施例方案,可以进一步提高数据安全性。这是基于数据被存储到多个物

理磁盘，虽然这些小物理磁盘中的数据并不具有连续性，但是如果这些物理磁盘又被分配到同一虚拟机，鉴于本发明实施例所使用的特殊小物理磁盘组成目标物理磁盘的方案，则有可能被恢复出来；为了避免这种情况的发生，提出了本实施例实现方案。

[0091] 本领域普通技术人员可以理解实现上述各方法实施例中的全部或部分步骤是可以通过程序来指令相关的硬件完成，相应的程序可以存储于一种计算机可读存储介质中，上述提到的存储介质可以是只读存储器，磁盘或光盘等。

[0092] 以上仅为本发明较佳的具体实施方式，但本发明的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本发明实施例揭露的技术范围内，可轻易想到的变化或替换，都应涵盖在本发明的保护范围之内。因此，本发明的保护范围应该以权利要求的保护范围为准。

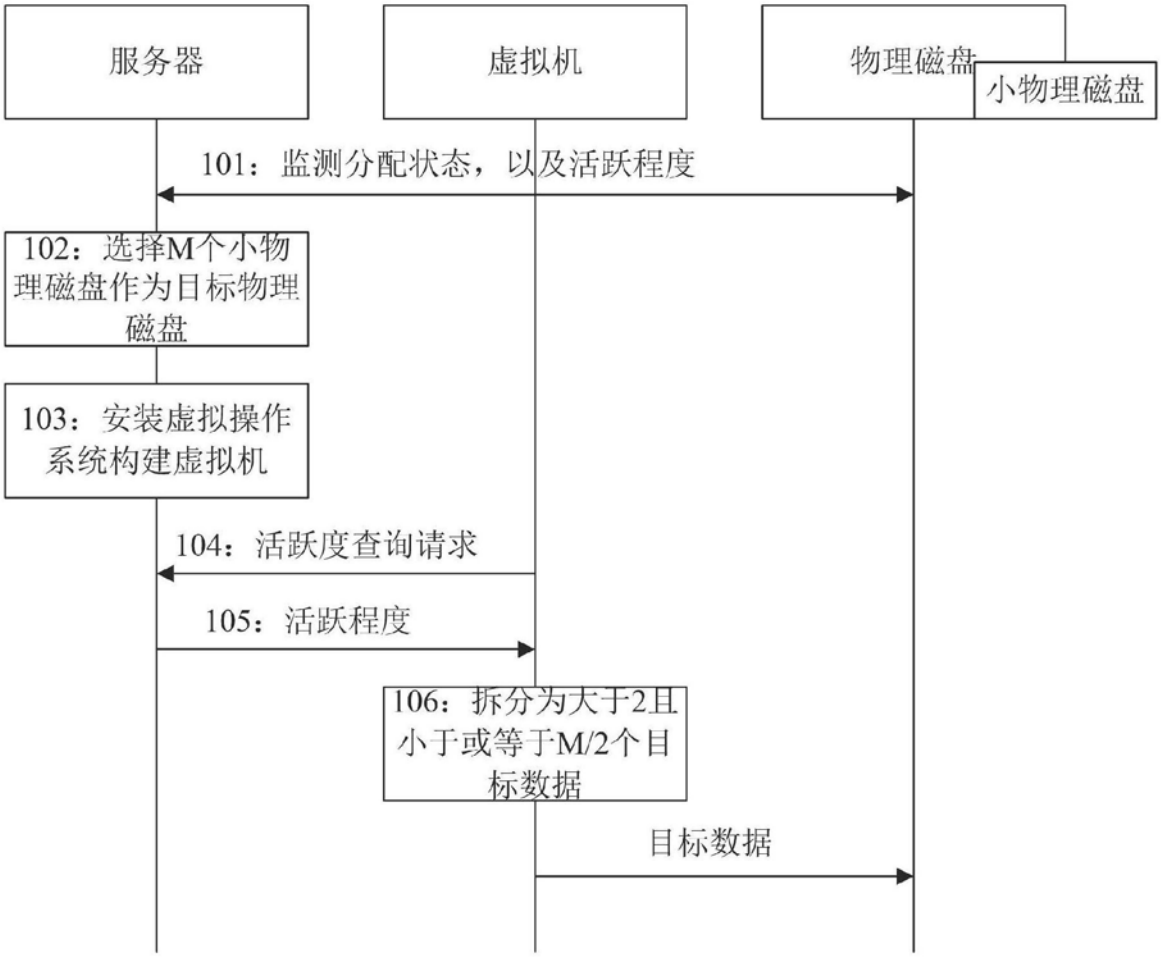

服务器
虚拟机
物理磁盘
小物理磁盘
101：监测分配状态，以及活跃程度
102：选择M个小物理磁盘作为目标物理磁盘
103：安装虚拟操作系统构建虚拟机
104：活跃度查询请求
105：活跃程度
106：拆分为大于2且小于或等于M/2个目标数据
目标数据

图1


大物理磁盘标识
小物理磁盘的序号
Q位
P位

图2

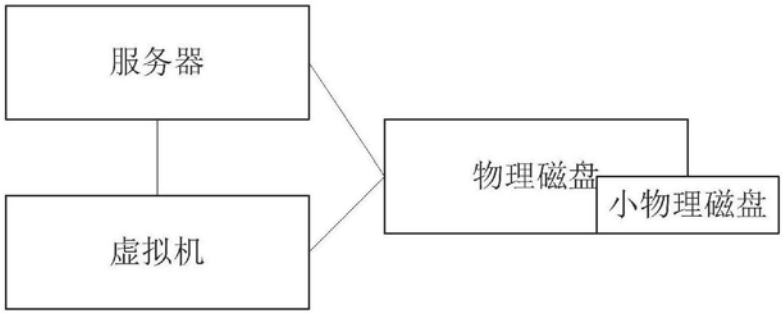


图3