(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 111488198 A
(43)申请公布日 2020.08.04

(21)申请号 202010300695.0

(22)申请日 2020.04.16

(71)申请人 湖南麒麟信安科技有限公司
地址 410000 湖南省长沙市高新区麒云路20号麒麟科技园4楼

(72)发明人 卢刚 孙利杰 杨鹏举 欧阳殷朝 胡智峰 夏华 陈松政 刘文清 杨涛

(74)专利代理机构 湖南兆弘专利事务所(普通合伙) 43008
代理人 谭武艺

(51)Int.Cl.
G06F 9/455(2006.01)

权利要求书2页 说明书8页 附图3页

(54)发明名称

一种超融合环境下的虚拟机调度方法、系统及介质

(57)摘要

本发明公开了一种超融合环境下的虚拟机存储管理方法、系统及介质，本发明方法中准备虚拟机的实施步骤包括：将系统镜像模板文件拷贝到各个主机中；根据各个用户的用户信息和所需使用的系统镜像模板文件信息在分布式存储系统的虚拟存储池中创建对应的目录D；将各个用户运行虚拟机所需的系统镜像差异文件、用户磁盘镜像文件分别写入分布式存储系统的目录D中，使得用户运行虚拟机所需的系统镜像差异文件、用户磁盘镜像文件存储在分布式存储系统的同一台目标主机H中。本发明通过对虚拟机所需的文件进行分布位置优化，将用户的虚拟机调度运行在文件所在主机上，能够提高虚拟机访存性能、减少网络开销。

CN 111488198 A

1.一种超融合环境下的虚拟机调度方法,其特征在于,准备虚拟机的实施步骤包括:

1)将系统镜像模板文件拷贝到各个主机中;

2)根据各个用户的用户信息和所需使用的系统镜像模板文件信息在分布式存储系统的虚拟存储池中创建对应的目录D;

3)将各个用户运行虚拟机所需的系统镜像差异文件、用户磁盘镜像文件分别写入分布式存储系统的目录D中,使得用户运行虚拟机所需的系统镜像差异文件、用户磁盘镜像文件存储在分布式存储系统的同一台目标主机H中。

2.根据权利要求1所述的超融合环境下的虚拟机调度方法,其特征在于,步骤2)中在分布式存储系统的虚拟存储池中创建对应的目录D具体是指在分布式存储系统的虚拟存储池中的指定位置创建用户名、系统镜像模板名称的嵌套两层子目录结构作为目录D,所述用户名、系统镜像模板文件名称的嵌套两层子目录结构具体是指将用户名作为第一层子目录、系统镜像模板名称作为第一层子目录下的第二层子目录,或者将系统镜像模板名称作为第一层子目录、用户名作为第一层子目录下的第二层子目录。

3.根据权利要求1所述的超融合环境下的虚拟机调度方法,其特征在于,步骤3)中写入分布式存储系统的目录D时分布式存储系统处理文件写入请求的步骤包括:获取文件写入请求在虚拟存储池中对应的写入目录名;判断写入目录名是否满足预设的格式要求,所述预设的格式要求是指步骤2)生成目录D的方式,如果满足预设的格式要求,则将写入目录名采用预设的哈希算法计算哈希值h;否则,针对文件写入请求的文件名采用预设的哈希算法计算哈希值h;将哈希值h与预设的哈希值范围进行匹配,获取哈希值范围中匹配的哈希值区间对应的一台或一组主机,根据一台或一组主机确定目标主机H,并将文件写入请求的文件写入目标主机H。

4.根据权利要求1所述的超融合环境下的虚拟机调度方法,其特征在于,所述准备虚拟机之后还包括下述启动虚拟机的步骤:

S1)确定目标用户运行虚拟机的目录D;

S2)确定目标用户运行虚拟机的目标主机H;

S3)基于目录D中系统镜像差异文件中指定的系统镜像模板文件和位置信息在目标主机H找到存储在本地的系统镜像模板文件,将目录D中对应的系统镜像差异文件、用户磁盘镜像文件作为本地的系统镜像模板文件的运行参数,在目标主机H上启动目标用户的虚拟机。

5.根据权利要求4所述的超融合环境下的虚拟机调度方法,其特征在于,步骤S2)的详细步骤包括:

S2.1)根据目标用户运行虚拟机的目录D中的系统镜像差异文件的文件名采用预设的哈希算法计算哈希值h;将哈希值h与预设的哈希值范围进行匹配,如果匹配成功则跳转执行步骤S2.3),否则跳转执行步骤S2.2);

S2.2)根据目录D采用预设的哈希算法计算哈希值h;将哈希值h与预设的哈希值范围进行匹配;

S2.3)获取哈希值范围中匹配的哈希值区间对应的一台或一组主机,根据一台或一组主机确定目标主机H。

6.根据权利要求3或5所述的超融合环境下的虚拟机调度方法,其特征在于,所述获取

哈希值范围中匹配的哈希值区间对应的一台或一组主机时，每一台或一组主机具有静态的哈希值区间；或者每一台或一组主机具有动态的哈希值区间，且将哈希值h与预设的哈希值范围进行匹配之前包括为每一台或一组主机动态分配哈希值区间的步骤：获取该用户运行虚拟机的硬件需求，获取分布式存储系统中满足硬件需求的主机，并将哈希值范围平均分配给满足硬件需求的主机，从而确定各个满足硬件需求的主机的哈希值区间、且使得不满足硬件需求的主机的哈希值区间为空。

7.根据权利要求3或5所述的超融合环境下的虚拟机调度方法，其特征在于，所述根据一台或一组主机确定目标主机H的详细步骤包括：若哈希值区间对应一台主机，则直接将该主机作为目标主机H；若哈希值区间对应多台主机，则在多台主机中选择一台主机作为目标主机H，且选择的方式为下述方式中的任意一种：方式1、随机选择一台主机作为目标主机H；方式2、选择CPU使用率最低的一台主机作为目标主机H；方式3、选择内存使用率最低的一台主机作为目标主机H；方式4、将CPU使用率、内存使用率加权求和且选择加权求和最低的一台主机作为目标主机H。

8.根据权利要求4所述的超融合环境下的虚拟机调度方法，其特征在于，所述启动虚拟机之后还包括下述步骤：检查本地存储的系统镜像模板文件，以及本地运行的虚拟机，如果本地存储的某个系统镜像模板文件没有被任意本地运行的虚拟机所使用，则将该系统镜像模板文件从本地删除。

9.一种超融合环境下的虚拟机调度装置，包括计算机设备，其特征在于，该计算机设备被编程或配置以执行权利要求1～8中任意一项所述超融合环境下的虚拟机调度方法的步骤，或者该计算机设备的存储器上存储有被编程或配置以执行权利要求1～8中任意一项所述超融合环境下的虚拟机调度方法的计算机程序。

10.一种计算机可读存储介质，其特征在于，该计算机可读存储介质上存储有被编程或配置以执行权利要求1～8中任意一项所述超融合环境下的虚拟机调度方法的计算机程序。

一种超融合环境下的虚拟机调度方法、系统及介质

技术领域

[0001] 本发明涉及计算机网络、云计算、虚拟计算、云桌面的超融合技术，具体涉及一种超融合环境下的虚拟机存储管理方法、系统及介质。

背景技术

[0002] 超融合基础架构是基于标准通用的硬件平台(如基于x86的主机、基于arm的主机)，在同一台主机设备中具备计算、网络、存储和虚拟化等资源和技术，多台主机可以通过网络以软件定义的方式实现各主机的计算、存储、网络资源融合，实现以虚拟化为中心的软件定义数据中心的技术架构。

[0003] 超融合基础架构的核心是存储的融合。一般情况下单台主机的存储设备包含多个物理磁盘，其中一部分物理磁盘构成该主机的本地存储，上面安装有主机操作系统和虚拟化软件等，本地存储能够被该主机上运行的应用程序直接访问，其它物理磁盘作为数据存储，多个主机的数据存储组成超融合系统的虚拟存储池，虚拟存储池可以挂载到超融合集群中的所有主机，由任意主机上应用程序访问。数据存储融合成虚拟存储池由分布式存储软件来实现，如各厂商自研的闭源分布式存储软件或基于开源的glusterfs、ceph等，软件中的分布式存储算法是决定文件实际存储在哪些主机存储设备的关键。

[0004] 分布式存储一般采用一致性哈希算法来决定文件实际存储至哪些主机的数据存储，在具体实现中，对一个具体目录下的文件的存储位置，可以在每台主机的数据存储上都创建该目录，每台主机的该目录分配一个不重叠的哈希值范围，所有主机该目录的哈希值范围组成一个完整[0, FFFFFFFF]的4字节32位值空间，每个哈希值范围可以是32位值空间在各主机间平均分配，也可以是依每台主机上数据存储的大小按比例分配，还可以是随机分配;在该目录下保存文件时，如果文件名的哈希值落在某台主机该目录哈希范围则将该文件写到该台主机的数据存储;对多副本分布式存储，复制可以以主机数据存储为单位，同一复制组的主机一般为2至3台，同组主机的相同目录的哈希值范围相同，同一个文件会写入这组主机中的所有成员主机，达到文件多副本目标。

[0005] 虽然超融合基础架构可以带来横向扩展和低成本的优势，但是与传统集中式存储设备相比，标准通用的硬件平台无法使用一些专用存储设备的优化手段，使得存储性能往往成为整个超融合系统的瓶颈和项目成败的关键。为了弥补通用存储硬件与专用存储设备之间的性能差距，超融合系统中对虚拟机所使用的存储进行种种优化，比如把虚拟机系统卷文件分片且保证有一组完整的分片副本保存在同一台主机提供的数据存储上;当虚拟机由多个虚拟机存储组件组成时，将虚拟存储池中各组件的主副本调度存储在同一台主机上;这些优化都是为了能尽量将虚拟机运行在它所需存储相同或最近的主机上，能提高数据读写性能，减少读写过程中网络延迟和带宽开销。

[0006] 不同规格要求的虚拟机运行还依赖主机上其它资源的可用情况，如cpu，内存，网络，GPU资源等，这些因素也会影响虚拟机调度。当虚拟机所需存储资源所在的主机不能满足这些存储以外的资源时，虚拟机可以被调度到其它主机上运行，此时该虚拟机在存储上

的访问性能会被降低。因此，如何在超融合基础架构系统中从具体技术方案上提高虚拟机访问所需各存储组件的性能，仍然是现有各种超融合实现技术中有待优化提高的问题。

发明内容

[0007] 本发明要解决的技术问题：针对现有技术的上述问题，提供一种超融合环境下的虚拟机存储管理方法、系统及介质，本发明通过对虚拟机所需的文件进行分布位置优化，将用户的虚拟机调度运行在文件所在主机上，能够提高虚拟机访存性能、减少网络开销。

[0008] 为了解决上述技术问题，本发明采用的技术方案为：

一种超融合环境下的虚拟机调度方法，准备虚拟机的实施步骤包括：

1)将系统镜像模板文件拷贝到各个主机中；

2)根据各个用户的用户信息和所需使用的系统镜像模板文件信息在分布式存储系统的虚拟存储池中创建对应的目录D；

3)将各个用户运行虚拟机所需的系统镜像差异文件、用户磁盘镜像文件分别写入分布式存储系统的目录D中，使得用户运行虚拟机所需的系统镜像差异文件、用户磁盘镜像文件存储在分布式存储系统的同一台目标主机H中。

[0009] 可选地，步骤2)中在分布式存储系统的虚拟存储池中创建对应的目录D具体是指在分布式存储系统的虚拟存储池中的指定位置创建用户名、系统镜像模板名称的嵌套两层子目录结构作为目录D，所述用户名、系统镜像模板文件名称的嵌套两层子目录结构具体是指将用户名作为第一层子目录、系统镜像模板名称作为第一层子目录下的第二层子目录，或者将系统镜像模板名称作为第一层子目录、用户名作为第一层子目录下的第二层子目录。

[0010] 可选地，步骤3)中写入分布式存储系统的目录D时分布式存储系统处理文件写入请求的步骤包括：获取文件写入请求在虚拟存储池中对应的写入目录名；判断写入目录名是否满足预设的格式要求，所述预设的格式要求是指步骤2)生成目录D的方式，如果满足预设的格式要求，则将写入目录名采用预设的哈希算法计算哈希值h；否则，针对文件写入请求的文件名采用预设的哈希算法计算哈希值h；将哈希值h与预设的哈希值范围进行匹配，获取哈希值范围中匹配的哈希值区间对应的一台或一组主机，根据一台或一组主机确定目标主机H，并将文件写入请求的文件写入目标主机H。

[0011] 可选地，所述准备虚拟机之后还包括下述启动虚拟机的步骤：

S1)确定目标用户运行虚拟机的目录D；

S2)确定目标用户运行虚拟机的目标主机H；

S3)基于目录D中系统镜像差异文件中指定的系统镜像模板文件和位置信息在目标主机H找到存储在本地的系统镜像模板文件，将目录D中对应的系统镜像差异文件、用户磁盘镜像文件作为本地的系统镜像模板文件的运行参数，在目标主机H上启动目标用户的虚拟机。

[0012] 可选地，步骤S2)的详细步骤包括：

S2.1)根据目标用户运行虚拟机的目录D中的系统镜像差异文件的文件名采用预设的哈希算法计算哈希值h；将哈希值h与预设的哈希值范围进行匹配，如果匹配成功则跳转执行步骤S2.3)，否则跳转执行步骤S2.2)；

S2.2)根据目录D采用预设的哈希算法计算哈希值h;将哈希值h与预设的哈希值范围进行匹配;

S2.3)获取哈希值范围中匹配的哈希值区间对应的一台或一组主机,根据一台或一组主机确定目标主机H。

[0013] 可选地,所述获取哈希值范围中匹配的哈希值区间对应的一台或一组主机时,每一台或一组主机具有静态的哈希值区间;或者每一台或一组主机具有动态的哈希值区间,且将哈希值h与预设的哈希值范围进行匹配之前包括为每一台或一组主机动态分配哈希值区间的步骤:获取该用户运行虚拟机的硬件需求,获取分布式存储系统中满足硬件需求的主机,并将哈希值范围平均分配给满足硬件需求的主机,从而确定各个满足硬件需求的主机的哈希值区间、且使得不满足硬件需求的主机的哈希值区间为空。

[0014] 可选地,所述根据一台或一组主机确定目标主机H的详细步骤包括:若哈希值区间对应一台主机,则直接将该主机作为目标主机H;若哈希值区间对应多台主机,则在多台主机中选择一台主机作为目标主机H,且选择的方式为下述方式中的任意一种:方式1、随机选择一台主机作为目标主机H;方式2、选择CPU使用率最低的一台主机作为目标主机H;方式3、选择内存使用率最低的一台主机作为目标主机H;方式4、将CPU使用率、内存使用率加权求和且选择加权求和最低的一台主机作为目标主机H。

[0015] 可选地,所述启动虚拟机之后还包括下述步骤:检查本地存储的系统镜像模板文件,以及本地运行的虚拟机,如果本地存储的某个系统镜像模板文件没有被任意本地运行的虚拟机所使用,则将该系统镜像模板文件从本地删除。

[0016] 此外,本发明还提供一种超融合环境下的虚拟机调度装置,包括计算机设备,其特征在于,该计算机设备被编程或配置以执行所述超融合环境下的虚拟机调度方法的步骤,或者该计算机设备的存储器上存储有被编程或配置以执行所述超融合环境下的虚拟机调度方法的计算机程序。

[0017] 此外,本发明还提供一种计算机可读存储介质,该计算机可读存储介质上存储有被编程或配置以执行所述超融合环境下的虚拟机调度方法的计算机程序。

[0018] 和现有技术相比,本发明具有下述优点:本发明将虚拟机运行所需文件分为系统镜像模板文件、系统镜像差异文件、用户磁盘镜像文件,通过系统镜像模板文件在主机本地存储,将系统镜像差异文件、用户磁盘镜像文件存储到虚拟存储池中,且采用特定映射方式生成目录D并将系统镜像差异文件、用户磁盘镜像文件存储在分布式存储系统的同一台目标主机H中,在保持超融合基础架构横向扩展和低成本优势前提下,提高虚拟机运行和访问存储的性能,通过在各主机本地存储存放虚拟机系统镜像模板副本,提高虚拟机访问系统盘效率,通过对分布式存储中哈希算法做优化将系统镜像的差异镜像文件和虚拟机用户盘镜像文件保存在同一主机的数据磁盘给虚拟机调度提供调度基础,通过将虚拟机调度运行在系统镜像差异文件、用户盘镜像文件所在的一台或一组主机,使得虚拟机可以直接访问其所在主机的存储设备,可以提高访存性能、减少网络延迟和带宽占用,通过综合评估虚拟机规格中除存储要求之外的其它因素,使得虚拟机所需存储组件不存放到不满足规格要求的主机的本地存储和数据存储,可以减少存储空间开销和虚拟机不能调度到合适主机的概率。

附图说明

[0019] 图1为本发明实施例中准备虚拟机的流程图。

[0020] 图2为本发明实施例中处理文件写入请求的流程图。

[0021] 图3为本发明实施例中启动虚拟机的流程图。

具体实施方式

[0022] 如图1所示,本实施例超融合环境下的虚拟机调度方法中,准备虚拟机的实施步骤包括:

1)将系统镜像模板文件拷贝到各个主机中;

2)根据各个用户的用户信息和所需使用的系统镜像模板文件信息在分布式存储系统的虚拟存储池中创建对应的目录D;

3)将各个用户运行虚拟机所需的系统镜像差异文件、用户磁盘镜像文件分别写入分布式存储系统的目录D中,使得用户运行虚拟机所需的系统镜像差异文件、用户磁盘镜像文件存储在分布式存储系统的同一台目标主机H中。

[0023] 本实施例中用户运行虚拟机所需的文件(存储组件)包括系统镜像模板文件、系统镜像差异文件、用户磁盘镜像文件。其中,系统镜像模板文件为多个用户共有,系统镜像差异文件的后端文件是系统镜像模板文件,系统镜像差异文件、用户磁盘镜像文件由每个用户虚拟机单独所有。本实施例中,虚拟机运行所需的系统镜像模板文件在每台主机的本地存储中都保存有一个副本,相同系统镜像模板文件在每台主机上所在的目录名和文件名都相同,虚拟机运行所需的系统镜像差异文件、用户磁盘镜像文件则保存在虚拟存储池中,系统镜像差异文件中保存有指向系统镜像模板文件的文件全名(包含主机本地存储目录名);不同主机访问系统镜像差异文件时,系统镜像差异文件的后端系统镜像模板就是该主机上本地存储的系统镜像差异文件,这些系统镜像差异文件是相同的副本。

[0024] 本实施例中,步骤2)中在分布式存储系统的虚拟存储池中创建对应的目录D具体是指在分布式存储系统的虚拟存储池中的指定位置创建用户名、系统镜像模板名称的嵌套两层子目录结构作为目录D,用户名、系统镜像模板文件名称的嵌套两层子目录结构具体是指将用户名作为第一层子目录、系统镜像模板名称作为第一层子目录下的第二层子目录,或者将系统镜像模板名称作为第一层子目录、用户名作为第一层子目录下的第二层子目录。例如,假定分布式存储系统的虚拟存储池在各主机上的指定挂载位置(可称为挂载目录或挂载点)为“/home/ kylin-data/”,系统镜像模板采用麒麟操作系统3.3且名称为kylinos33,用户user1基于系统镜像模板kylinos33的虚拟机V生成的系统镜像差异文件“DELTA.IMG”、用户磁盘镜像文件“USER.IMG”保存在目录(目录D)“/home/kylin-data/kylin-desktops/user1/kylinos33”下,即该嵌套两层子目录结构为“user1/kylinos33”,此外嵌套两层子目录结构也可以采用“kylinos33/ user1”。此外,作为一种可选的实施方式,本实施例中系统镜像模板文件“GUEST.IMG”存放在目标主机的本地存储目录/home/desktop-template/kylinos33下。需要说明的是,此处的“DELTA.IMG”、“USER.IMG”、“GUEST.IMG”仅仅是一种命名示例,其实施并不限于此。需要说明的是,目录D是指在虚拟存储池中创建的目录,当虚拟存储池基于文件系统时,比如说虚拟存储池的文件系统可以挂载在主机的/home/kylin-data目录,然后在/home/kylin-data下创建目录kylin-

desktops/user1/kylinos3.3,此时对应中虚拟存储池本身的文件系统中创建的目录名是/kylin-desktop/user1/kylinos3.3。对分布式文件系统内部的一致性哈希定位算法来说,分布式文件系统不能事先知道其挂载目录,前述例子中,主机上的指定挂载位置(可称为挂载目录或挂载点)为“/home/kylin-data”,而目录D是指/kylin-desktops/user1/kylinos3.3,当存储池挂载到/home/kylin-data目录时,此时挂载点目录是指/home/kylin-data,挂载后的D目录全路径变成/home/kylin-data/kylin-desktops/user1/kylinos3.3,但是因为分布式存储子系统并不知道它自己会挂载到哪,所以存储子系统内部程序使用它自身的目录名字信息。

[0025] 如图2所示,步骤3)中写入分布式存储系统的目录D时分布式存储系统处理文件写入请求的步骤包括:获取文件写入请求在虚拟存储池中对应的写入目录名;判断写入目录名是否满足预设的格式要求(满足模式M),预设的格式要求是指步骤2)生成目录D的方式(本实施例中即为模式M),如果满足预设的格式要求,则将写入目录名采用预设的哈希算法A(如DM-hash,也可以是MD5, SHA等常用哈希算法)计算哈希值*h*;否则,针对文件写入请求的文件名采用哈希算法A计算哈希值*h*;将哈希值*h*与预设的哈希值范围进行匹配,获取哈希值范围中匹配的哈希值区间对应的一台或一组主机,根据一台或一组主机确定目标主机H,并将文件写入请求的文件写入目标主机H。由于写入目录名满足模式M时会直接将写入目录名采用预设的哈希算法A计算哈希值*h*,根据该哈希值*h*决定写入的文件存入哪台主机的数据存储,由于同一用户相同系统镜像模板文件的系统镜像差异文件、用户磁盘镜像文件存放的目录是同一目录D,因此则这两个文件会生成相同的哈希值*h*,分布式存储软件将这两个文件存放到相同目标主机H的数据存储上。毫无疑问,将写入目录名采用预设的哈希算法(记为哈希算法A)计算哈希值*h*时,既可以根据需要采用写入目录名的全名或短名采用哈希算法A计算哈希值*h*,只要其采用的全名或短名能够相互区别,就能够实现对不同目标主机H的映射。

[0026] 本实施例中将一致性哈希算法作为分布式存储系统的分布算法,该算法中在每台主机用于存储池的数据存储上都创建存储池上的目录结构,在每个目录结构的扩展属性中写入一个哈希值区间,所有主机的哈希值区间组成一个[0, FFFFFFFF]的4字节32位范围(预设的哈希值范围),每台主机上该目录结构的哈希值区间互不重叠。本实施例中根据目录D是否匹配模式M作为依据使用哈希算法A为文件名或该文件所在目录D的名字计算哈希值h,当哈希值h落在某主机上该目录属性的范围上时,将该文件保存在这台主机的数据存储上;超融合系统管理程序为用户准备虚拟机所需的存储组件时,分布式存储将这些存储组件写入选定主机的数据存储,超融合系统管理程序将虚拟机进程启动在选定的目标主机H上。

[0027] 例如,在某由3台服务器组成超融合云桌面系统中,每台主机平分该目录的32位哈希值范围,第1、2、3台的哈希值区间依次为[0xAAAAAAAA, 0xFFFFFFFF]、[0x00000000, 0x55555554]、[0x55555555, 0xAAAAAAA9](起点可以在任一台,此例中为第2台为范围0起点),一致性哈希算法使用模式M为/kylin-desktops/.*/.*/,其中“.”代表的是单个字符,“*” 代表的是任意字符串(可以是空串),".*"代表至少一个字符的任意字符串,第一个“.*”用于匹配任意用户名,第二个“.*” 用于匹配任意系统镜像模板文件名称。云桌面管理系统在/home/kylin-data/user1/kylinos33目录创建DELTA.IMG、USER.IMG,由于可以匹配

模式M,则直接使用字符串/home/kylin-data/user1/kylinos33而不是字符串DELTA.IMG或USER.IMG计算哈希值,计算得到哈希值*h*为0x91c3e25f,则落在第3台主机的哈希值区间,分布式存储子系统将这两个IMG文件存储在第3台主机的数据存储,用户请求启动虚拟机V时,云桌面管理系统获取目录/home/kylin-data/user1/kylinos33的哈希值得到第3台主机,然后将虚拟机V调度在第3台主机上运行。

[0028] 例如某由8台服务器(序号为1-8)组成超融合云桌面系统,分布式存储将其分为4组(1,2)(3,4)(5,6)(7,8),同组两台服务器的数据存储为复制组,保存有相同文件副本,kylinos33的镜像模板文件GUEST.IMG内安装的操作系统要求有GPU支持,8台服务器在1、2、3、4、7、8上有GPU资源,5、6上没有GPU资源,在除5、6每台服务器的本地存储目录/home/desktop_template/kylinos33下保存镜像模板文件GUEST.IMG,服务器数据存储组成的虚拟存储池挂载在/home/kylin-data目录,用户user1基于kylinos33镜像模板的虚拟机V生成的系统镜像差异文件DELTA.IMG和用户磁盘镜像文件USER.IMG保存在/home/kylin-data/user1/kylinos33目录下,各组主机在该目录的32位哈希值范围分布为:第1组(1,2)为[0x55555555, 0xAAAAAAA9],第2组(3,4)为 [0xAAAAAAAA, 0xFFFFFFFF],第3组(5,6)为(0,0),第4组为[0x00000000, 0x55555554],一致性哈希算法使用模式M为 /kylin-desktops/.*/.*/,云桌面管理系统在/home/kylin-data/kylin-desktops/user1/kylinos33目录创建DELTA.IMG、USER.IMG,由于可以匹配模式M,则直接使用字符串/kylin-desktops/user1/kylinos33而不是字符串DELTA.IMG或USER.IMG计算哈希值,计算得到哈希值为0x91c3e25f,则落在第1组主机的哈希范围,分布式存储子系统将这两个IMG文件存储在第1组服务器1,2的数据存储,用户请求启动虚拟机V时,云桌面管理系统获取目录/kylin-desktops/user1/kylinos33的哈希值范围分布,计算目录名的哈希值,得到第1组主机,然后将虚拟机V调度在第1台或第2台主机上运行。

[0029] 如图3所示,准备虚拟机之后还包括下述启动虚拟机的步骤:

S1)确定目标用户运行虚拟机的目录D;

S2)确定目标用户运行虚拟机的目标主机H;

S3)基于目录D中系统镜像差异文件中指定的系统镜像模板文件和位置信息在目标主机H找到存储在本地的系统镜像模板文件,将目录D中对应的系统镜像差异文件、用户磁盘镜像文件作为本地的系统镜像模板文件的运行参数,在目标主机H上启动目标用户的虚拟机。

[0030] 本实施例中,步骤S2)的详细步骤包括:

S2.1)根据目标用户运行虚拟机的目录D中的系统镜像差异文件的文件名采用预设的哈希算法计算哈希值*h*;将哈希值*h*与预设的哈希值范围进行匹配,如果匹配成功则跳转执行步骤S2.3),否则跳转执行步骤S2.2);

S2.2)根据目录D采用预设的哈希算法计算哈希值*h*;将哈希值*h*与预设的哈希值范围进行匹配;

S2.3)获取哈希值范围中匹配的哈希值区间对应的一台或一组主机,根据一台或一组主机确定目标主机H。

[0031] 作为一种可选的实施方式,本实施例中,获取哈希值范围中匹配的哈希值区间对应的一台或一组主机时,每一台或一组主机具有静态的哈希值区间。

[0032] 作为另一种可选的实施方式，考虑到某些虚拟机对于硬件资源有特定要求，例如需要GPU或其他硬件加速器等，针对上述问题，可采用根据硬件需求来动态分配哈希值区间的方式：每一台或一组主机具有动态的哈希值区间，且将哈希值h与预设的哈希值范围进行匹配之前包括为每一台或一组主机动态分配哈希值区间的步骤：获取该用户运行虚拟机的硬件需求，获取分布式存储系统中满足硬件需求的主机，并将哈希值范围平均分配给满足硬件需求的主机，从而确定各个满足硬件需求的主机的哈希值区间、且使得不满足硬件需求的主机的哈希值区间为空(即不会匹配任何哈希值h)。

[0033] 本实施例中，一个哈希值区间既可以对应一台主机，也可以根据需要对应一组主机，例如超融合架构所使用的分布式存储以主机数据存储为单位存在多个(2至3个)副本时，则上述方法中将相同数据存储的主机视为一组，用户的系统镜像差异文件和用户磁盘镜像文件以组为单位来计算实际存放的主机，组内每台主机都存放有相同的差异文件和磁盘镜像文件副本。本实施例中，根据一台或一组主机确定目标主机H的详细步骤包括：若哈希值区间对应一台主机，则直接将该主机作为目标主机H；若哈希值区间对应多台主机，则在多台主机中选择一台主机作为目标主机H，且选择的方式为下述方式中的任意一种：方式1、随机选择一台主机作为目标主机H；方式2、选择CPU使用率最低的一台主机作为目标主机H；方式3、选择内存使用率最低的一台主机作为目标主机H；方式4、将CPU使用率、内存使用率加权求和且选择加权求和最低的一台主机作为目标主机H。

[0034] 此外，为了进一步防止系统镜像模板文件在主机中浪费空间，本实施例中，启动虚拟机之后还包括下述步骤：检查本地存储的系统镜像模板文件，以及本地运行的虚拟机，如果本地存储的某个系统镜像模板文件没有被任意本地运行的虚拟机所使用，则将该系统镜像模板文件从本地删除。

[0035] 综上所述，1、本实施例通过识别虚拟机运行所需的三种镜像文件，将系统镜像模板放在主机本地存储，系统镜像差异文件、用户磁盘镜像文件存放在存储集群且保存在特定模式目录名的目录；2、本实施例的一致性哈希算法与特定模式相关联，分布式存储保存文件时按一致性算法在目录上分配哈希区间，使用文件名哈希值计算得到文件应分配的数据存储节点，但是当文件目录被匹配到特定模式时，不使用文件名而使用其路径名来计算哈希值区间，以此保证同一用户的同一系统镜像虚拟机所需的系统镜像差异文件、用户磁盘镜像文件位于相同主机的数据存储；3，本实施例的虚拟机调度时取得目录的哈希值区间，并计算路径哈希值，可以得到主机节点；4，本实施例的将虚拟机规格要求、服务器主机资源与镜像文件存储位置相关联，当虚拟机规格中有特殊要求而某些主机不能满足时，可以不在这些主机存储上保存该虚拟机的存储组件文件；5，多副本环境下，各副本节点的数据存储内容一致，本实施例的方法中可方便地虚拟机可调度至复制节点中任一或优选节点。本实施例超融合环境下的虚拟机调度方法能够在保持超融合基础架构横向扩展和低成本优势前提下，提高虚拟机运行和访问存储的性能，通过在各主机本地存储存放虚拟机系统镜像模板副本，提高虚拟机访问系统盘效率，通过对分布式存储中一致性哈希算法做优化将系统镜像的差异镜像文件和虚拟机用户盘镜像文件保存在同一主机的数据磁盘给虚拟机调度提供调度基础，通过将虚拟机调度运行在系统镜像差异文件、用户盘镜像文件所在的一台或一组主机，使得虚拟机可以直接访问其所在主机的存储设备，可以提高访存性能、减少网络延迟和带宽占用，通过综合评估虚拟机规格中除存储要求之外的其它因素，使

得虚拟机所需存储组件不存放到不满足规格要求的主机的本地存储和数据存储，可以减少存储空间开销和虚拟机不能调度到合适主机的概率。

[0036] 此外，本发明还提供一种超融合环境下的虚拟机调度装置，包括计算机设备，该计算机设备被编程或配置以执行前述超融合环境下的虚拟机调度方法的步骤，或者该计算机设备的存储器上存储有被编程或配置以执行前述超融合环境下的虚拟机调度方法的计算机程序。

[0037] 此外，本发明还提供一种计算机可读存储介质，该计算机可读存储介质上存储有被编程或配置以执行前述超融合环境下的虚拟机调度方法的计算机程序。

[0038] 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。本申请是参照根据本申请实施例的方法、设备(系统)、和计算机程序产品的流程图和/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。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。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。

[0039] 以上所述仅是本发明的优选实施方式，本发明的保护范围并不仅局限于上述实施例，凡属于本发明思路下的技术方案均属于本发明的保护范围。应当指出，对于本技术领域的普通技术人员来说，在不脱离本发明原理前提下的若干改进和润饰，这些改进和润饰也应视为本发明的保护范围。

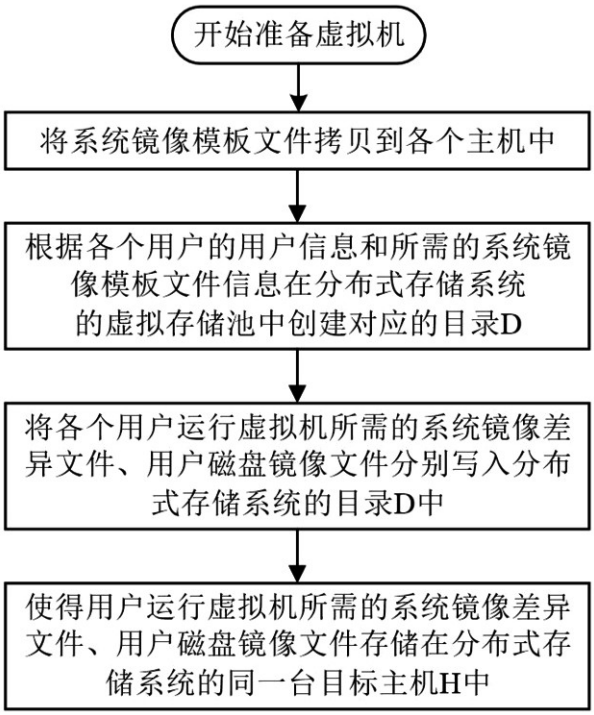
开始准备虚拟机
将系统镜像模板文件拷贝到各个主机中
根据各个用户的用户信息和所需的系统镜像模板文件信息在分布式存储系统的虚拟存储池中创建对应的目录D
将各个用户运行虚拟机所需的系统镜像差异文件、用户磁盘镜像文件分别写入分布式存储系统的目录D中
使得用户运行虚拟机所需的系统镜像差异文件、用户磁盘镜像文件存储在分布式存储系统的同一台目标主机H中

图1

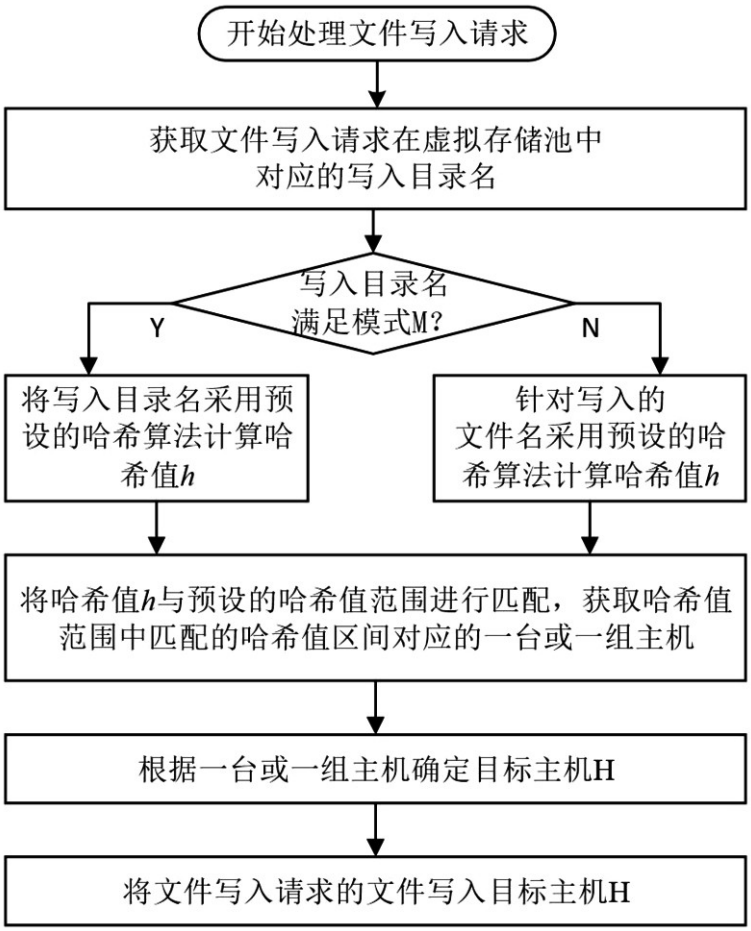
开始处理文件写入请求
获取文件写入请求在虚拟存储池中对应的写入目录名
写入目录名满足模式M?
Y
N
将写入目录名采用预设的哈希算法计算哈希值h
针对写入的文件名采用预设的哈希算法计算哈希值h
将哈希值h与预设的哈希值范围进行匹配，获取哈希值范围中匹配的哈希值区间对应的一台或一组主机
根据一台或一组主机确定目标主机H
将文件写入请求的文件写入目标主机H

图2


开始启动虚拟机
确定目标用户运行虚拟机的目录D
确定目标用户运行虚拟机的目标主机H
基于目录D中系统镜像差异文件中指定的虚拟机系统镜像模板文件和位置信息在目标主机H找到存储在本地的虚拟机系统镜像模板
将目录D中对应的系统镜像差异文件、用户磁盘镜像文件作为本地的虚拟机系统镜像模板的运行参数，在目标主机H上启动目标用户的虚拟机

图3