(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 107967164 A
(43)申请公布日 2018.04.27

(21)申请号 201610913494.1

(22)申请日 2016.10.19

(71)申请人 阿里巴巴集团控股有限公司
地址 英属开曼群岛大开曼资本大厦一座四层847号邮箱

(72)发明人 张超

(74)专利代理机构 北京润泽恒知识产权代理有限公司 11319
代理人 赵娟

(51)Int.Cl.
G06F 9/455(2006.01)
G06F 9/50(2006.01)

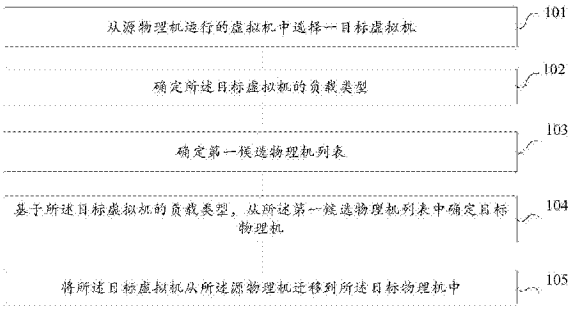
权利要求书4页 说明书19页 附图2页

(54)发明名称

一种虚拟机热迁移的方法及系统

(57)摘要

本申请实施例提供了一种虚拟机热迁移的方法及系统，其中所述方法包括：从源物理机运行的虚拟机中选择一目标虚拟机；确定所述目标虚拟机的负载类型，其中，所述目标虚拟机的负载类型是根据所述目标虚拟机运行在所述源物理机时，所述目标虚拟机或所述源物理机的特定资源利用率来确定的；确定第一候选物理机列表，其中，所述第一候选物理机列表包括存储所述目标虚拟机的多个磁盘数据副本对应的多个物理机；基于所述目标虚拟机的负载类型，从所述第一候选物理机列表中确定目标物理机；将所述目标虚拟机迁移到所述目标物理机中。本申请实施例可以从vm外部依赖的资源关系的相关性，分析得到vm的最优目的nc，并进行迁移操作，从而提高目标vm的性能。



CN 107967164 A

1.一种虚拟机热迁移的方法，其特征在于，所述方法包括：

从源物理机运行的虚拟机中选择一目标虚拟机；

确定所述目标虚拟机的负载类型，其中，所述目标虚拟机的负载类型是根据所述目标虚拟机运行在所述源物理机时，所述目标虚拟机或所述源物理机的特定资源利用率来确定的；

确定第一候选物理机列表，其中，所述第一候选物理机列表包括存储所述目标虚拟机的多个磁盘数据副本对应的多个物理机；

基于所述目标虚拟机的负载类型，从所述第一候选物理机列表中确定目标物理机；

将所述目标虚拟机从所述源物理机迁移到所述目标物理机中。

2.根据权利要求1所述的方法，其特征在于，还包括：

确定所述目标虚拟机的关联虚拟机，所述关联虚拟机为与所述目标虚拟机存在交互关系的虚拟机；

在迁移所述目标虚拟机的同时，将所述关联虚拟机迁移至与所述目标虚拟机相同的目标物理机中。

3.根据权利要求2所述的方法，其特征在于，所述确定所述目标虚拟机的关联虚拟机的步骤包括：

若所述目标虚拟机与某一虚拟机经由源物理机和/或另一物理机进行通信，则判定所述目标虚拟机与所述虚拟机基于同一个物理机或不同物理机存在交互关系；

将与所述目标虚拟机存在交互关系的虚拟机作为关联虚拟机。

4.根据权利要求3所述的方法，其特征在于，所述若所述目标虚拟机与某一虚拟机经由源物理机和/或另一物理机进行通信，则判定所述目标虚拟机与所述虚拟机基于同一个物理机或不同物理机存在交互关系的步骤包括：

采用源物理机中预先布置的探测点对所述目标虚拟机的进出数据包进行探测；

若所述目标虚拟机与某一虚拟机之间具有进出数据包，则判定所述目标虚拟机与某一虚拟机进行通信。

5.根据权利要求1-4任一项所述的方法，其特征在于，所述源物理机包括如下一个或多个：

计算机集群中特定资源利用率高于第一资源负载阈值的物理机或特定资源利用率低于第二资源负载阈值的物理机，其中，所述第一资源负载阈值大于所述第二资源负载阈值；

计算机集群中存在故障的物理机。

6.根据权利要求1所述的方法，其特征在于，所述确定第一候选物理机列表的步骤包括：

分别获取所述目标虚拟机在计算机集群中的每个物理机上运行时访问所述目标虚拟机的不同的磁盘数据副本的访问路径；

将访问路径的长度小于或等于预设路径阈值的物理机组织成第一候选物理机列表。

7.根据权利要求1所述的方法，其特征在于，所述特定资源利用率包括所述目标虚拟机的CPU资源利用率，所述负载类型包括CPU密集型，所述确定所述目标虚拟机的负载类型的步骤包括：

当所述目标虚拟机运行在所述源物理机上时，获取所述目标虚拟机的CPU资源利用率；

获取计算机集群的平均CPU资源利用率；

若预设时间段内所述目标虚拟机的CPU资源利用率均大于所述平均CPU资源利用率，则确定所述目标虚拟机的负载类型为CPU密集型。

8.根据权利要求1所述的方法，其特征在于，所述特定资源利用率包括所述源物理机的I/O队列深度，和/或，所述目标虚拟机的CPU中等待I/O操作的进程数；所述负载类型包括存储密集型；

所述确定所述目标虚拟机的负载类型的步骤包括：

当所述目标虚拟机运行在所述源物理机上时，获取所述源物理机的I/O队列深度，和/或，所述目标虚拟机的CPU中等待I/O操作的进程数；

若所述源物理机的I/O队列深度在预设时间段内均大于预设深度阈值，和/或，所述目标虚拟机的CPU中等待I/O操作的进程数在预设时间段内均大于预设进程数阈值，则确定所述目标虚拟机的负载类型为存储密集型。

9.根据权利要求1所述的方法，其特征在于，所述特定资源利用率包括源物理机中网卡收发数据包的速率；所述负载类型包括网络密集型，所述确定所述目标虚拟机的负载类型的步骤包括：

当所述目标虚拟机运行在所述源物理机上时，获取所述源物理机中网卡收发数据包的速率；

获取计算机集群中收发数据包的平均速率；

若源物理机中网卡收发数据包的速率在预设时间段内均大于所述平均速率，则确定所述目标虚拟机的负载类型为网络密集型。

10.根据权利要求7或8或9所述的方法，其特征在于，所述基于所述目标虚拟机的负载类型，从所述第一候选物理机列表中确定目标物理机的步骤包括：

基于所述目标虚拟机的负载类型，分别判断每个候选物理机中是否存在与所述目标虚拟机存在资源竞争关系的竞争虚拟机；

在所述第一候选物理机列表中删除存在竞争vm的候选物理机，得到第二候选物理机列表；

从所述第二候选物理机列表中确定目标物理机。

11.根据权利要求10所述的方法，其特征在于，所述基于所述目标虚拟机的负载类型，分别判断每个候选物理机中是否存在与所述目标虚拟机存在资源竞争关系的竞争虚拟机的步骤包括：

分别获取所述候选物理机中运行的每个虚拟机的负载类型；

若所述候选物理机中存在与所述目标虚拟机的负载类型相同的虚拟机，则判定所述候选物理机中存在与所述目标虚拟机存在资源竞争关系的竞争虚拟机。

12.根据权利要求10或11所述的方法，其特征在于，所述从所述第二候选物理机列表中确定目标物理机的步骤为：

将第二候选物理机列表中特定资源利用率最小的候选物理机作为目标物理机。

13.一种虚拟机热迁移的系统，其特征在于，所述系统包括：

目标虚拟机选择模块，用于从源物理机运行的虚拟机中选择一目标虚拟机；

负载类型确定模块，用于确定所述目标虚拟机的负载类型，其中，所述目标虚拟机的负

载类型是根据所述目标虚拟机运行在所述源物理机时，所述目标虚拟机或所述源物理机的特定资源利用率来确定的；

候选列表确定模块，用于确定第一候选物理机列表，其中，所述第一候选物理机列表包括存储所述目标虚拟机的多个磁盘数据副本对应的多个物理机；

目标物理机选择模块，用于基于所述目标虚拟机的负载类型，从所述第一候选物理机列表中确定目标物理机；

迁移模块，用于将所述目标虚拟机从所述源物理机迁移到所述目标物理机中。

14.根据权利要求13所述的系统，其特征在于，还包括：

关联虚拟机确定模块，用于确定所述目标虚拟机的关联虚拟机，所述关联虚拟机为与所述目标虚拟机存在交互关系的虚拟机；

关联虚拟机迁移模块，用于在迁移所述目标虚拟机的同时，将所述关联虚拟机迁移至与所述目标虚拟机相同的目标物理机中。

15.根据权利要求14所述的系统，其特征在于，所述关联虚拟机确定模块包括：

交互关系判断子模块，用于若所述目标虚拟机与某一虚拟机经由源物理机和/或另一物理机进行通信，则判定所述目标虚拟机与所述虚拟机基于同一个物理机或不同物理机存在交互关系；

关联虚拟机确定子模块，用于将与所述目标虚拟机存在交互关系的虚拟机作为关联虚拟机。

16.根据权利要求15所述的系统，其特征在于，所述交互关系判断子模块包括：

数据包探测单元，用于采用源物理机中预先布置的探测点对所述目标虚拟机的进出数据包进行探测；

通信确定单元，用于在判定所述目标虚拟机与某一虚拟机之间具有进出数据包时，则判定所述目标虚拟机与某一虚拟机进行通信。

17.根据权利要求13-16任一项所述的系统，其特征在于，所述源物理机包括如下一个或多个：

计算机集群中特定资源利用率高于第一资源负载阈值的物理机或特定资源利用率低于第二资源负载阈值的物理机，其中，所述第一资源负载阈值大于所述第二资源负载阈值；

计算机集群中存在故障的物理机。

18.根据权利要求13所述的系统，其特征在于，所述候选列表确定模块包括：

访问路径获取子模块，用于分别获取所述目标虚拟机在计算机集群中的每个物理机上运行时访问所述目标虚拟机的不同的磁盘数据副本的访问路径；

组织子模块，用于将访问路径的长度小于或等于预设路径阈值的物理机组织成第一候选物理机列表。

19.根据权利要求13所述的系统，其特征在于，所述特定资源利用率包括所述目标虚拟机的CPU资源利用率，所述负载类型包括CPU密集型；

所述负载类型确定模块包括：

CPU资源利用率获取子模块，用于当所述目标虚拟机运行在所述源物理机上时，获取所述目标虚拟机的CPU资源利用率；

平均CPU资源利用率获取子模块，用于获取计算机集群的平均CPU资源利用率；

CPU密集型确定子模块，用于若预设时间段内所述目标虚拟机的CPU资源利用率均大于所述平均CPU资源利用率，则确定所述目标虚拟机的负载类型为CPU密集型。

20.根据权利要求13所述的系统，其特征在于，所述特定资源利用率包括所述源物理机的I/O队列深度，和/或，所述目标虚拟机的CPU中等待I/O操作的进程数；所述负载类型包括存储密集型；

所述负载类型确定模块包括：

存储资源获取子模块，用于当所述目标虚拟机运行在所述源物理机上时，获取所述源物理机的I/O队列深度，和/或，所述目标虚拟机的CPU中等待I/O操作的进程数；

存储密集型确定子模块，用于若所述源物理机的I/O队列深度在预设时间段内均大于预设深度阈值，和/或，所述目标虚拟机的CPU中等待I/O操作的进程数在预设时间段内均大于预设进程数阈值，则确定所述目标虚拟机的负载类型为存储密集型。

21.根据权利要求13所述的系统，其特征在于，所述特定资源利用率包括源物理机中网卡收发数据包的速率；所述负载类型包括网络密集型；

所述负载类型确定模块包括：

收发速率获取子模块，用于当所述目标虚拟机运行在所述源物理机上时，获取所述源物理机中网卡收发数据包的速率；

平均速率获取子模块，用于获取计算机集群中收发数据包的平均速率；

网络集型确定子模块，用于若源物理机中网卡收发数据包的速率在预设时间段内均大于所述平均速率，则确定所述目标虚拟机的负载类型为网络密集型。

22.根据权利要求19或20或21所述的系统，其特征在于，所述目标物理机选择模块包括：

竞争虚拟机确定子模块，用于基于所述目标虚拟机的负载类型，分别判断每个候选物理机中是否存在与所述目标虚拟机存在资源竞争关系的竞争虚拟机；

第二候选物理机列表生成子模块，用于在所述第一候选物理机列表中删除存在竞争vm的候选物理机，得到第二候选物理机列表；

目标物理机选取子模块，用于从所述第二候选物理机列表中确定目标物理机。

23.根据权利要求22所述的系统，其特征在于，所述竞争虚拟机确定子模块包括：

负载类型获取单元，用于分别获取所述候选物理机中运行的每个虚拟机的负载类型；

确定单元，用于若所述候选物理机中存在与所述目标虚拟机的负载类型相同的虚拟机，则判定所述候选物理机中存在与所述目标虚拟机存在资源竞争关系的竞争虚拟机。

24.根据权利要求22或23所述的系统，其特征在于，所述目标物理机选取子模块还用于：

将第二候选物理机列表中特定资源利用率最小的候选物理机作为目标物理机。

一种虚拟机热迁移的方法及系统

技术领域

[0001] 本申请涉及数据处理技术领域，特别是涉及一种虚拟机热迁移的方法，以及，一种虚拟机热迁移的系统。

背景技术

[0002] 虚拟机(virtual machine，简称vm)指通过在物理设备中安装虚拟机管理软件(Hypervisor)，由Hypervisor模拟出一个或多个具有完整硬件系统功能的、运行在一个完全隔离环境中的完整系统。虚拟机的热迁移(Live Migration)是虚拟机应用中的一个重要技术，热迁移技术又叫动态迁移、实时迁移，即虚拟机保存/恢复，通常是将整个虚拟机的运行状态完整保存下来，同时可以快速的恢复到原有硬件平台甚至是不同硬件平台上，恢复以后，虚拟机仍旧平滑运行，用户不会察觉到任何差异。

[0003] 现有的热迁移技术通常的使用场景是，当通过监控发现某个物理机nc(即物理计算节点)上的负载过重，vm性能下降，vm的CPU争抢过高时，执行迁移操作，将对应的vm迁移到其余的资源尚有空余的目标nc上。

[0004] 上述方法虽然可以解决vm由于物理资源不足而导致的用户体验下降的问题，但是在迁移的过程中，却没有考虑vm与外部资源的关系，在执行迁移操作后，也许对原有的对该vm造成主要性能下降的问题被解决了，但是在迁移的过程中，可能也会引入一些新的因素，导致该vm在迁移结束后，总体性能表现变差。

[0005] 因此，目前需要本领域技术人员迫切解决的一个技术问题就是：提出一种虚拟机热迁移的机制，用以为虚拟机选择集群范围内最优的物理机进行迁移。

发明内容

[0006] 本申请实施例所要解决的技术问题是提供一种虚拟机热迁移的方法，用以为虚拟机选择集群范围内最优的物理机进行迁移。

[0007] 相应的，本申请实施例还提供了一种虚拟机热迁移的系统，用以保证上述方法的实现及应用。

[0008] 为了解决上述问题，本申请实施例公开了一种虚拟机热迁移的方法，所述方法包括：

[0009] 从源物理机运行的虚拟机中选择一目标虚拟机；

[0010] 确定所述目标虚拟机的负载类型，其中，所述目标虚拟机的负载类型是根据所述目标虚拟机运行在所述源物理机时，所述目标虚拟机或所述源物理机的特定资源利用率来确定的；

[0011] 确定第一候选物理机列表，其中，所述第一候选物理机列表包括存储所述目标虚拟机的多个磁盘数据副本对应的多个物理机；

[0012] 基于所述目标虚拟机的负载类型，从所述第一候选物理机列表中确定目标物理机；

[0013] 将所述目标虚拟机从所述源物理机迁移到所述目标物理机中。

[0014] 优选地，所述方法还包括：

[0015] 确定所述目标虚拟机的关联虚拟机，所述关联虚拟机为与所述目标虚拟机存在交互关系的虚拟机；

[0016] 在迁移所述目标虚拟机的同时，将所述关联虚拟机迁移至与所述目标虚拟机相同的目标物理机中。

[0017] 优选地，所述确定所述目标虚拟机的关联虚拟机的步骤包括：

[0018] 若所述目标虚拟机与某一虚拟机经由源物理机和/或另一物理机进行通信，则判定所述目标虚拟机与所述虚拟机基于同一个物理机或不同物理机存在交互关系；

[0019] 将与所述目标虚拟机存在交互关系的虚拟机作为关联虚拟机。

[0020] 优选地，所述若所述目标虚拟机与某一虚拟机经由源物理机和/或另一物理机进行通信，则判定所述目标虚拟机与所述虚拟机基于同一个物理机或不同物理机存在交互关系的步骤包括：

[0021] 采用源物理机中预先布置的探测点对所述目标虚拟机的进出数据包进行探测；

[0022] 若所述目标虚拟机与某一虚拟机之间具有进出数据包，则判定所述目标虚拟机与某一虚拟机进行通信。

[0023] 优选地，所述源物理机包括如下一个或多个：

[0024] 计算机集群中特定资源利用率高于第一资源负载阈值的物理机或特定资源利用率低于第二资源负载阈值的物理机，其中，所述第一资源负载阈值大于所述第二资源负载阈值；

[0025] 计算机集群中存在故障的物理机。

[0026] 优选地，所述确定第一候选物理机列表的步骤包括：

[0027] 分别获取所述目标虚拟机在计算机集群中的每个物理机上运行时访问所述目标虚拟机的不同的磁盘数据副本的访问路径；

[0028] 将访问路径的长度小于或等于预设路径阈值的物理机组织成第一候选物理机列表。

[0029] 优选地，所述特定资源利用率包括所述目标虚拟机的CPU资源利用率，所述负载类型包括CPU密集型，所述确定所述目标虚拟机的负载类型的步骤包括：

[0030] 当所述目标虚拟机运行在所述源物理机上时，获取所述目标虚拟机的CPU资源利用率；

[0031] 获取计算机集群的平均CPU资源利用率；

[0032] 若预设时间段内所述目标虚拟机的CPU资源利用率均大于所述平均CPU资源利用率，则确定所述目标虚拟机的负载类型为CPU密集型。

[0033] 优选地，所述特定资源利用率包括所述源物理机的I/O队列深度，和/或，所述目标虚拟机的CPU中等待I/O操作的进程数；所述负载类型包括存储密集型；

[0034] 所述确定所述目标虚拟机的负载类型的步骤包括：

[0035] 当所述目标虚拟机运行在所述源物理机上时，获取所述源物理机的I/O队列深度，和/或，所述目标虚拟机的CPU中等待I/O操作的进程数；

[0036] 若所述源物理机的I/O队列深度在预设时间段内均大于预设深度阈值，和/或，所

述目标虚拟机的CPU中等待I/O操作的进程数在预设时间段内均大于预设进程数阈值，则确定所述目标虚拟机的负载类型为存储密集型。

[0037] 优选地，所述特定资源利用率包括源物理机中网卡收发数据包的速率；所述负载类型包括网络密集型，所述确定所述目标虚拟机的负载类型的步骤包括：

[0038] 当所述目标虚拟机运行在所述源物理机上时，获取所述源物理机中网卡收发数据包的速率；

[0039] 获取计算机集群中收发数据包的平均速率；

[0040] 若源物理机中网卡收发数据包的速率在预设时间段内均大于所述平均速率，则确定所述目标虚拟机的负载类型为网络密集型。

[0041] 优选地，所述基于所述目标虚拟机的负载类型，从所述第一候选物理机列表中确定目标物理机的步骤包括：

[0042] 基于所述目标虚拟机的负载类型，分别判断每个候选物理机中是否存在与所述目标虚拟机存在资源竞争关系的竞争虚拟机；

[0043] 在所述第一候选物理机列表中删除存在竞争vm的候选物理机，得到第二候选物理机列表；

[0044] 从所述第二候选物理机列表中确定目标物理机。

[0045] 优选地，所述基于所述目标虚拟机的负载类型，分别判断每个候选物理机中是否存在与所述目标虚拟机存在资源竞争关系的竞争虚拟机的步骤包括：

[0046] 分别获取所述候选物理机中运行的每个虚拟机的负载类型；

[0047] 若所述候选物理机中存在与所述目标虚拟机的负载类型相同的虚拟机，则判定所述候选物理机中存在与所述目标虚拟机存在资源竞争关系的竞争虚拟机。

[0048] 优选地，所述从所述第二候选物理机列表中确定目标物理机的步骤为：

[0049] 将第二候选物理机列表中特定资源利用率最小的候选物理机作为目标物理机。

[0050] 本申请实施例还公开了一种虚拟机热迁移的系统，所述系统包括：

[0051] 目标虚拟机选择模块，用于从源物理机运行的虚拟机中选择一目标虚拟机；

[0052] 负载类型确定模块，用于确定所述目标虚拟机的负载类型，其中，所述目标虚拟机的负载类型是根据所述目标虚拟机运行在所述源物理机时，所述目标虚拟机或所述源物理机的特定资源利用率来确定的；

[0053] 候选列表确定模块，用于确定第一候选物理机列表，其中，所述第一候选物理机列表包括存储所述目标虚拟机的多个磁盘数据副本对应的多个物理机；

[0054] 目标物理机选择模块，用于基于所述目标虚拟机的负载类型，从所述第一候选物理机列表中确定目标物理机；

[0055] 迁移模块，用于将所述目标虚拟机从所述源物理机迁移到所述目标物理机中。

[0056] 优选地，所述系统还包括：

[0057] 关联虚拟机确定模块，用于确定所述目标虚拟机的关联虚拟机，所述关联虚拟机为与所述目标虚拟机存在交互关系的虚拟机；

[0058] 关联虚拟机迁移模块，用于在迁移所述目标虚拟机的同时，将所述关联虚拟机迁移至与所述目标虚拟机相同的目标物理机中。

[0059] 优选地，所述关联虚拟机确定模块包括：

[0060] 交互关系判断子模块，用于若所述目标虚拟机与某一虚拟机经由源物理机和/或另一物理机进行通信，则判定所述目标虚拟机与所述虚拟机基于同一个物理机或不同物理机存在交互关系；

[0061] 关联虚拟机确定子模块，用于将与所述目标虚拟机存在交互关系的虚拟机作为关联虚拟机。

[0062] 优选地，所述交互关系判断子模块包括：

[0063] 数据包探测单元，用于采用源物理机中预先布置的探测点对所述目标虚拟机的进出数据包进行探测；

[0064] 通信确定单元，用于在判定所述目标虚拟机与某一虚拟机之间具有进出数据包时，则判定所述目标虚拟机与某一虚拟机进行通信。

[0065] 优选地，所述源物理机包括如下一个或多个：

[0066] 计算机集群中特定资源利用率高于第一资源负载阈值的物理机或特定资源利用率低于第二资源负载阈值的物理机，其中，所述第一资源负载阈值大于所述第二资源负载阈值；

[0067] 计算机集群中存在故障的物理机。

[0068] 优选地，所述候选列表确定模块包括：

[0069] 访问路径获取子模块，用于分别获取所述目标虚拟机在计算机集群中的每个物理机上运行时访问所述目标虚拟机的不同的磁盘数据副本的访问路径；

[0070] 组织子模块，用于将访问路径的长度小于或等于预设路径阈值的物理机组织成第一候选物理机列表。

[0071] 优选地，所述特定资源利用率包括所述目标虚拟机的CPU资源利用率，所述负载类型包括CPU密集型；

[0072] 所述负载类型确定模块包括：

[0073] CPU资源利用率获取子模块，用于当所述目标虚拟机运行在所述源物理机上时，获取所述目标虚拟机的CPU资源利用率；

[0074] 平均CPU资源利用率获取子模块，用于获取计算机集群的平均CPU资源利用率；

[0075] CPU密集型确定子模块，用于若预设时间段内所述目标虚拟机的CPU资源利用率均大于所述平均CPU资源利用率，则确定所述目标虚拟机的负载类型为CPU密集型。

[0076] 优选地，所述特定资源利用率包括所述源物理机的I/O队列深度，和/或，所述目标虚拟机的CPU中等待I/O操作的进程数；所述负载类型包括存储密集型；

[0077] 所述负载类型确定模块包括：

[0078] 存储资源获取子模块，用于当所述目标虚拟机运行在所述源物理机上时，获取所述源物理机的I/O队列深度，和/或，所述目标虚拟机的CPU中等待I/O操作的进程数；

[0079] 存储密集型确定子模块，用于若所述源物理机的I/O队列深度在预设时间段内均大于预设深度阈值，和/或，所述目标虚拟机的CPU中等待I/O操作的进程数在预设时间段内均大于预设进程数阈值，则确定所述目标虚拟机的负载类型为存储密集型。

[0080] 优选地，所述特定资源利用率包括源物理机中网卡收发数据包的速率；所述负载类型包括网络密集型；

[0081] 所述负载类型确定模块包括：

[0082] 收发速率获取子模块，用于当所述目标虚拟机运行在所述源物理机上时，获取所述源物理机中网卡收发数据包的速率；

[0083] 平均速率获取子模块，用于获取计算机集群中收发数据包的平均速率；

[0084] 网络集型确定子模块，用于若源物理机中网卡收发数据包的速率在预设时间段内均大于所述平均速率，则确定所述目标虚拟机的负载类型为网络密集型。

[0085] 优选地，所述目标物理机选择模块包括：

[0086] 竞争虚拟机确定子模块，用于基于所述目标虚拟机的负载类型，分别判断每个候选物理机中是否存在与所述目标虚拟机存在资源竞争关系的竞争虚拟机；

[0087] 第二候选物理机列表生成子模块，用于在所述第一候选物理机列表中删除存在竞争vm的候选物理机，得到第二候选物理机列表；

[0088] 目标物理机选取子模块，用于从所述第二候选物理机列表中确定目标物理机。

[0089] 优选地，所述竞争虚拟机确定子模块包括：

[0090] 负载类型获取单元，用于分别获取所述候选物理机中运行的每个虚拟机的负载类型；

[0091] 确定单元，用于若所述候选物理机中存在与所述目标虚拟机的负载类型相同的虚拟机，则判定所述候选物理机中存在与所述目标虚拟机存在资源竞争关系的竞争虚拟机。

[0092] 优选地，所述目标物理机选取子模块还用于：

[0093] 将第二候选物理机列表中特定资源利用率最小的候选物理机作为目标物理机。

[0094] 与背景技术相比，本申请实施例包括以下优点：

[0095] 在热迁移过程中利用nc和nc、nc和存储、vm和vm、vm与负载之间的相关性，通过挖掘vm与其周边资源的相关性(网络、存储、cpu等)，来分析该vm在哪个nc上运行是最优的，并进行迁移操作，从而提高vm的性能。

附图说明

[0096] 图1是本申请的一种虚拟机热迁移的方法实施例一的步骤流程图；

[0097] 图2是本申请的一种虚拟机热迁移的方法实施例二的步骤流程图；

[0098] 图3是本申请的一种虚拟机热迁移的系统实施例的结构框图。

具体实施方式

[0099] 为使本申请的上述目的、特征和优点能够更加明显易懂，下面结合附图和具体实施方式对本申请作进一步详细的说明。

[0100] 参照图1，示出了本申请的一种虚拟机热迁移的方法实施例一的步骤流程图，所述方法可以包括如下步骤：

[0101] 步骤101，从源物理机运行的虚拟机中选择一目标虚拟机；

[0102] 步骤102，确定所述目标虚拟机的负载类型；

[0103] 其中，所述目标虚拟机的负载类型是根据所述目标虚拟机运行在所述源物理机时，所述目标虚拟机或所述源物理机的特定资源利用率来确定的。

[0104] 步骤103，确定第一候选物理机列表；

[0105] 其中，所述第一候选物理机列表包括存储所述目标虚拟机的多个磁盘数据副本对

应的多个物理机。

[0106] 步骤104，基于所述目标虚拟机的负载类型，从所述第一候选物理机列表中确定目标物理机；

[0107] 步骤105，将所述目标虚拟机从所述源物理机迁移到所述目标物理机中。

[0108] 在本申请实施例中，在进行目标物理机的选择时，考虑目标虚拟机的负载类型，使得选定的目标物理机为集群内与目标虚拟机最优的物理机，从而使得迁移后的目标虚拟机相对于迁移前而言，性能更好。

[0109] 参照图2，示出了本申请的一种虚拟机热迁移的方法实施例二的步骤流程图，所述方法可以包括如下步骤：

[0110] 步骤201，从源物理机运行的虚拟机中选择一目标虚拟机；

[0111] 本申请实施例可以应用于分布式文件系统(Distributed File System)中，分布式文件系统是指文件系统管理的物理存储资源不一定直接连接在本地节点上，而是通过计算机网络与节点相连。

[0112] 在具体实现中，分布式文件系统通常可以包括主控服务器(或称元数据服务器、名字服务器等，通常会配置备用主控服务器以便在故障时接管服务，也可以两个都为主的模式)，多个数据服务器(或称存储服务器，存储节点等)，以及多个客户端，客户端可以是各种应用服务器，也可以是终端用户。

[0113] 物理机nc是相对于虚拟机vm而言的对实体计算机的称呼，物理机提供给虚拟机以硬件环境，有时也称为“寄主”或“宿主”。

[0114] 在一种实施方式中，本申请实施例的源物理机可以包括如下一个或多个：计算机集群中特定资源利用率高于第一资源负载阈值的物理机，或者，特定资源利用率低于第二资源负载阈值的物理机；计算机集群中存在故障的物理机。

[0115] 具体而言，可以采用如下几种方式中的至少一种从计算机集群中确定源nc：

[0116] (1)选择计算机集群中特定资源利用率高于第一资源负载阈值的物理机作为源物理机。

[0117] 具体的，每个物理机的特定资源利用率可以包括CPU资源利用率、存储资源利用率、以及网络资源利用率等的一种。

[0118] 关于特定资源利用率的计算方式将在下述步骤202中进行说明。

[0119] 在获得计算机集群中每个nc的特定资源利用率以后，可以将特定资源利用率高于第一资源负载阈值的物理机作为源物理机，即将计算机集群中负载较重的物理机nc作为源nc进行vm的迁出，可以减轻该nc的负载。

[0120] 例如，检测到计算机集群中某个nc的CPU资源利用率为80％，则可以将该nc作为源nc。

[0121] (2)选择计算机集群中特定资源利用率低于第二资源负载阈值的物理机作为源物理机。

[0122] 获得计算机集群各个nc的特定资源利用率以后，还可以将特定资源利用率低于第二资源负载阈值的物理机作为源物理机，其中，第二资源负载阈值小于第一资源负载阈值。该特定资源利用率低于第二资源负载阈值的物理机即为计算机集群中负载较轻的使用率不高的物理机nc，将此部分nc作为源nc进行vm的迁出，可以释放该nc所占的系统资源。

[0123] 例如，检测到计算机集群中某个nc的CPU资源利用率为10％，则可以将该nc作为源nc。

[0124] 需要说明的是，在从计算机集群中选定源物理机时，不限于用特定资源利用率作为判断指标，也可以采用其他指标进行判断，例如，可以采用该CPU资源利用率、存储资源利用率、以及网络资源利用率进行平均或加权平均后得到的平均值进行判断，即将该平均值高于第一资源负载阈值的物理机作为源物理机。

[0125] (3)将计算机集群中发生故障的物理机作为源物理机，将此部分nc作为源nc可以及时降低故障带来的损失。

[0126] 当然，上述三种源物理机确定的方法仅仅是本申请实施例的示例，本领域技术人员采用其他方式确定源物理机均是可以的，本申请实施例对此不作限制。

[0127] 一台物理机中可以运行多台虚拟机，在一种实施方式中，当确定源物理机后，本申请实施例可以遍历该源物理机中运行的多台虚拟机，判断每个虚拟机是否可以迁移到更优的物理机上。

[0128] 在具体实现中，可以首先从源物理机中任意选取一个虚拟机作为目标虚拟机，以进行后续关于该目标虚拟机的处理。当当前目标虚拟机处理完毕以后，可以继续选取下一虚拟机作为目标虚拟机完成处理过程，直到物理机上所有虚拟机遍历完毕。

[0129] 在另一种实施方式中，还可以获取源物理机中每个虚拟机的特定资源利用率(关于特定资源利用率的获取方式将在步骤202中进行说明)，将特定资源利用率小于某一设定的负载阈值的虚拟机作为目标虚拟机，即将虚拟机压力不太大，对源物理机不那么重要的虚拟机作为目标虚拟机进行迁移，以降低迁移过程中对用户的影响。

[0130] 当然，上述虚拟机的特定资源利用率也可以用CPU资源利用率、存储资源利用率、以及网络资源利用率进行平均或加权平均后得到的平均值代替，本申请实施例对此不作限定。

[0131] 需要说明的是，本申请实施例并不限于上述两种确定目标虚拟机的方式，本领域技术人员采用其他方式均是可以的。

[0132] 步骤202，确定所述目标虚拟机的负载类型；

[0133] 在实际中，可以根据目标vm运行在源nc时，目标vm或源nc的资源负载情况来确定目标vm的负载类型。其中，该资源负载情况可以体现为特定资源利用率。

[0134] 作为本申请实施例的一种示例，该特定资源利用率可以包括CPU资源利用率、存储资源利用率以及网络资源利用率的至少一种。对应的负载类型可以包括如下类型的一种：CPU密集型、存储密集型、网络密集型。

[0135] 具体的，若目标vm在源nc运行时，所负荷的CPU资源、存储资源或网络资源中，CPU资源利用率所占比重最大，则对应的负载类型为CPU密集型；若存储资源利用率所占比重最大，则对应的负载类型为存储密集型；若网络资源利用率所占比重最大，则对应的负载类型为网络密集型。

[0136] 在本发明实施例的一种优选实施例中，若所述特定资源利用率包括所述目标虚拟机的CPU资源利用率，所述目标vm的负载类型为CPU密集型；

[0137] 则步骤202可以包括如下子步骤：

[0138] 子步骤S11，当所述目标虚拟机运行在所述源物理机上时，获取所述目标虚拟机的

CPU资源利用率；

[0139] CPU资源利用率指的是目标vm运行在源nc时的CPU资源占用率，可以在源nc上输入mpstat-P vmware来获得目标vm的CPU资源利用率。

[0140] 子步骤S12，获取计算机集群的平均CPU资源利用率；

[0141] 在具体实现中，可以获得计算机集群中每个虚拟机的CPU资源利用率，然后将所有的CPU资源利用率相加后除以计算机集群中的虚拟机数量，得到平均CPU资源利用率。

[0142] 子步骤S13，若预设时间段内所述目标虚拟机的CPU资源利用率均大于所述平均CPU资源利用率，则确定所述目标虚拟机的负载类型为CPU密集型。

[0143] 在一种实施方式中，如果目标vm的CPU资源利用率大于当前计算机集群中的平均CPU资源利用率，则可以判定该目标vm的负载类型为CPU密集型。

[0144] 在具体实现中，为了更准确地确定目标vm的负载类型，该目标vm的CPU资源利用率大于当前计算机集群中的平均CPU资源利用率可以进一步为目标vm的CPU资源利用率大于或等于当前计算机集群中的平均CPU资源利用率的预设倍数，例如，一个时间段中目标vm的CPU资源利用率均为当前集群的平均CPU资源利用率的1.5倍，则可以判定该目标vm的负载类型为CPU密集型。

[0145] 在另一种实施方式中，为了更准确地确定负载类型，在对CPU资源进行判断时，还可以结合存储资源利用率以及网络资源利用率作为辅助参考，而存储资源利用率可以以源nc中I/O队列深度作为参考指标，网络资源利用率可以以网卡收发包带宽作为参考指标。

[0146] 例如，若当前vm集群中有50个vm，可以获得每个vm的CPU资源资源利用率，并计算当前vm集群中平均CPU资源利用率，若该目标vm的CPU资源利用率大于平均CPU资源利用率的1.5倍，目标vm所运行的源物理机中I/O队列深度长时间小于1，网卡收发包带宽远远小于vm带宽的10％，即该目标vm的CPU资源占用较多，存储资源以及网络资源占用较少，则可以判定该目标vm的负载类型为CPU密集型。

[0147] 在本发明实施例的另一种优选实施例中，若所述特定资源利用率包括所述源物理机的I/O队列深度，和/或，所述目标虚拟机的CPU中等待I/O操作的进程数；所述目标vm的负载类型为存储密集型；

[0148] 则步骤202可以包括如下子步骤：

[0149] 子步骤S21，当所述目标虚拟机运行在所述源物理机上时，获取所述源物理机的I/O队列深度，和/或，所述目标虚拟机的CPU中等待I/O操作的进程数；

[0150] 具体的，存储资源利用率可以体现为目标vm运行在源nc时，源物理机的I/O队列深度，和/或，所述目标虚拟机的CPU中等待I/O操作的进程数。

[0151] 源nc上的I/O队列深度是指目标vm运行在源nc时，源nc中I/O队列中等候的I/O请求的数量，可以通过iostat命令获得I/O队列深度。

[0152] 目标vm的CPU中等待I/O操作的进程数是指目标vm运行在源nc时，目标vm对应的CPU的iowait值，iowait的含义为有进程在等I/O操作结束(备份进程)，并且在等待I/O操作结束的过程中，无其他进程占用CPU，CPU处于空闲状态。

[0153] 子步骤S22，若所述源物理机的I/O队列深度在预设时间段内均大于预设深度阈值，和/或，所述目标虚拟机的CPU中等待I/O操作的进程数在预设时间段内均大于预设进程数阈值，则确定所述目标虚拟机的负载类型为存储密集型。

[0154] 在一种实施方式中，如果源nc中I/O队列深度在预设时间段内均大于预设深度阈值，例如，如果源nc中I/O队列深度长时间段内均大于5，则可以判定目标vm的负载类型为存储密集型。

[0155] 在另一种实施方式中，如果目标vm运行在源nc上时，目标vm的CPU iowait的值在预设时间段内均大于预设进程数阈值，例如，目标vm的CPU iowait的值长时间段内均大于1，则确定目标vm的负载类型为存储密集型。

[0156] 在另一种实施方式中，为了更准确地确定负载类型，在对存储资源进行判断时，还可以结合CPU资源利用率以及网络资源利用率作为辅助参考，而网络资源可以以网卡收发包带宽作为参考指标。

[0157] 例如，假设当前计算机集群中有50个vm，若该目标vm的CPU资源利用率长时间小于当前计算机集群中的平均CPU资源利用率，而目标vm所运行的源物理机中I/O队列深度长时间大于5，且目标vm的CPU iowait的值长时间段内均大于1，网卡收发包带宽远远小于vm带宽的10％，即该目标vm的存储资源占用较多，CPU资源以及网络资源占用较少，则可以判定该目标vm的负载类型为存储密集型。

[0158] 在本发明实施例的另一种优选实施例中，若所述特定资源利用率包括源物理机中网卡收发数据包的速率；所述目标vm的负载类型为网络密集型；

[0159] 则步骤202可以包括如下子步骤：

[0160] 子步骤S31，当所述目标虚拟机运行在所述源物理机上时，获取所述源物理机中网卡收发数据包的速率；

[0161] 具体的，源物理机中网卡收发数据包的速率可以为源物理机中网卡产生CPU中断的速率，当网卡接收到数据帧或发送完数据帧时，就会产生一个中断，因此，网卡产生CPU中断的速率也可以为网卡收发数据包的速率。

[0162] 子步骤S32，获取计算机集群中收发数据包的平均速率；

[0163] 在具体实现中，可以获得计算机集群中每个物理机的网卡收发数据包的速率，然后将所有的网卡收发数据包的速率相加后除以计算机集群中的物理机数量，得到计算机集群中收发数据包的平均速率。

[0164] 子步骤S33，若源物理机中网卡收发数据包的速率在预设时间段内均大于所述平均速率，则确定所述目标虚拟机的负载类型为网络密集型。

[0165] 在具体实现中，若源物理机中网卡收发数据包的速率在预设时间段内均大于平均速率，例如，如果源物理机中网卡收发数据包的速率是当前计算机集群中收发数据包的平均速率的5倍，则可以判定目标vm的负载类型为网络密集型。

[0166] 进一步的，在实际应用中，为了更准确地确定负载类型，在对网络资源进行判断时，还可以结合CPU资源利用率以及存储资源利用率作为辅助参考。

[0167] 例如，假设当前vm集群中有50个vm，若该目标vm的CPU资源利用率长时间小于当前计算机集群中的平均CPU资源利用率，目标vm所运行的源物理机中的I/O队列深度长时间小于1，网卡产生CPU中断的速率是当前计算机集群中平均速率的5倍，即该目标vm的网络资源占用较多，CPU资源以及存储资源占用较少，则可以判定该目标vm的负载类型为网络密集型。

[0168] 需要说明的是，在判定网络密集型时，除了可以以上述的源物理机中网卡收发数

据包的速率作为判断条件，还可以以其他条件进行判断，例如，可以以源物理机中网卡收发数据包的带宽作为判断条件，如果源物理机中网卡收发数据包的带宽大于目标vm的带宽的一定比例，例如，大于目标vm的带宽的10％，则说明网络带宽满负荷，此时，可以判定该目标vm的负载类型为网络密集型。

[0169] 在实际中，可能存在目标vm均属于上述三种负载类型或均不属于上述三种负载类型的情况。此时，可以获取源nc上CPU资源利用率、存储资源利用率以及网络资源利用率，并将CPU资源利用率、存储资源利用率以及网络资源利用率进行归一化处理，随后比较归一化后的CPU资源利用率、存储资源利用率以及网络资源利用率的大小，将最大者作为目标vm的负载类型。

[0170] 需要说明的是，本申请实施例并不限于上述负载类型，每个负载类型的判断也不限于上述方式，本领域技术人员采用其他方式进行负载类型的设计和判断均是可以的。

[0171] 步骤203，确定第一候选物理机列表；

[0172] 本申请实施例在选取目标虚拟机以后，可以进一步确定该目标虚拟机对应的第一候选物理机列表。

[0173] 在本申请实施例的一种优选实施例中，该第一候选物理机列表可以包括存储目标虚拟机的多个磁盘数据副本对应的多个物理机。步骤203可以包括如下子步骤：

[0174] 子步骤S41，分别获取所述目标虚拟机在计算机集群中的每个物理机上运行时访问所述目标虚拟机的不同的磁盘数据副本的访问路径；

[0175] 子步骤S42，将访问路径的长度小于或等于预设路径阈值的物理机组织成第一候选物理机列表。

[0176] 在分布式文件系统中，目标虚拟机的磁盘数据可以分多个副本分别存储在不同的存储节点中。

[0177] 具体的，在实际应用中，虚拟机的存储服务可以分为四种：

[0178] 第一种是直接存储(Direct-Attached Storage，DAS)，即将虚拟机所需的数据存储在本地磁盘和SSD(Solid State Drives，固态硬盘)上，以块设备形式直接赋予虚拟机使用，或者以映射文件方式提供。

[0179] 第二种是网络存储系统(Network-Attached Storage，NAS)，例如NFS(Network File System，网络文件系统)，将映射文件置于其上，并提供存储服务。

[0180] 第三种是存储区域网络(Storage Area Network，SAN)，外部通过连接到SAN控制器使用存储服务。

[0181] 第四种是分布式存储，为了保证数据的安全性，虚拟机的磁盘数据通常可以写入多个数据服务器上，在实际中，可以采用如下三种方式将虚拟机的磁盘数据写入数据服务器：第1种方式是客户端分别向多个数据服务器写同一份数据；第2种方式是客户端向主数据服务器写数据，主数据服务器向其他数据服务器转发数据；第3种方式是采用流水复制的方式，客户端向某个数据服务器写数据，该数据服务器向副本链中下一个数据服务器转发数据，依次类推。当有节点宕机或节点间负载不均匀的情况下，主控服务器会制定一些副本复制或迁移计划，而数据服务器实际执行这些计划，将副本转发或迁移至其他的数据服务器。数据服务器也可提供管理工具，在需要的情况下由管理员手动的执行一些复制或迁移计划。

[0182] 本申请实施例优选可以采用上述第四种存储方式对虚拟机的磁盘数据进行存储。例如，若配置的副本数为3，则可以将目标VM中的磁盘数据分别存储于nc1，nc2和nc3三个物理机中。

[0183] 在具体实现中，在vm磁盘数据的存储位置固定的情况下，集群的不同节点访问对应的存储，系统的开销是有差异的，即虚拟机在不同的nc上运行时，其到不同的存储节点(存储VM磁盘数据的nc)的距离是不同的。

[0184] 假设目标vm在集群中每个nc都运行一遍，则在每个nc中，目标vm对不同存储节点中的磁盘数据的访问路径是不一样的。此时，针对每个存储节点中的目标vm的磁盘数据，均可以获得集群中每个nc到该存储节点的访问路径，随后，将访问路径的长度小于或等于预设路径阈值的物理机组织成第一候选物理机列表，该第一候选物理机列表为目标vm在迁移时可选的最优备选nc列表。

[0185] 例如，在下表1中，1和3分别表示访问路径的长度，其中，若访问路径的长度为1，则可以说明目标vm运行的nc为存储目标vm磁盘数据副本的nc。若预设路径阈值为1，则可以得到下述结论：VM1的第一候选物理机列表为nc1-nc2-nc3；VM2的第一候选物理机列表为nc2-nc3-nc4；VM3的第一候选物理机列表为nc1-nc3-nc4。

[0186]

vm所在nc	nc1	nc2	nc3	nc4
vm1	1	1	1	3
vm2	3	1	1	1
vm3	1	3	1	1

[0187] 表1

[0188] 需要说明的是，本发明实施例并不限于上述确定第一候选物理机列表的方式，本领域技术人员还可以采用其他方式确定第一候选物理机列表，例如，

[0189] 人为选定候选nc，并将该人为确定的候选nc记录在配置文件中，当读取配置文件时，可以获得该多个选定的候选nc，并将该多个候选nc组织成第一候选物理机列表。

[0190] 在实际应用中，针对人为设定候选nc的情形，可以应用于如下两种场景：

[0191] 第一种场景是，若当前的nc集群中出现部分nc负载过重的情形时，可以将负载过重的nc上的vm迁移至负载相对较轻的nc上，此时，候选nc的设定可以为负载较轻的nc，例如，当前nc集群中有5个nc负载过重，5个nc负载相对较轻，则可以将5个负载过重的nc上的部分vm迁移至5个负载较轻的nc上，从而达到10个nc的负载均衡。

[0192] 第二种场景是，若当前的nc集群中出现部分nc使用率不高的情形时，可以将使用率不高的nc上的vm集中在某几个nc上，此时，候选nc的设定可以为使用率较高的nc，例如，当前nc集群中有10nc，其中5个nc使用率并不高，5个nc使用率达到平均值，则可以将5个使用率不高的nc上的vm迁移至5个使用率达到平均值的nc上，从而释放5个使用率不高的nc的资源。

[0193] 步骤204，基于所述目标虚拟机的负载类型，从所述第一候选物理机列表中确定目标物理机；

[0194] 确定第一候选物理机列表以后，进而可以结合目标vm的负载类型，从第一候选物理机列表中选择一候选物理机作为目标物理机。

[0195] 在本申请实施例的一种优选实施例中，步骤204可以包括如下子步骤：

[0196] 子步骤S51，基于所述目标虚拟机的负载类型，分别判断每个候选物理机中是否存在与所述目标虚拟机存在资源竞争关系的竞争虚拟机；

[0197] 具体的，若目标虚拟机与某一虚拟机共同运行在同一物理机时，共享该物理机的某一资源，则该虚拟机与该目标虚拟机存在资源竞争关系，该虚拟机为竞争虚拟机。

[0198] 在本申请实施例的一种优选实施例中，子步骤S51进一步可以包括如下子步骤：

[0199] 子步骤S511，分别获取所述候选物理机中运行的每个虚拟机的负载类型；

[0200] 候选物理机中运行的每个虚拟机的负载类型也可以包括CPU密集型、存储密集型、网络密集型等类型的至少一种。

[0201] 候选物理机中运行的每个虚拟机的负载类型的确定方式与步骤202中目标vm的负载类型的确定方式相同，具体可以参照步骤202中的描述，此处不再赘述了。

[0202] 子步骤S512，若所述候选物理机中存在与所述目标虚拟机的负载类型相同的虚拟机，则判定所述候选物理机中存在与所述目标虚拟机存在资源竞争关系的竞争虚拟机。在获得候选nc中运行的每个vm的负载类型以后，若该候选nc上存在与目标vm的负载类型相同的vm，则该vm作为目标vm的竞争vm，两者对其负载类型对应的物理机的资源存在竞争关系。

[0203] 子步骤S52，在所述第一候选物理机列表中删除存在竞争vm的候选物理机，得到第二候选物理机列表；

[0204] 在具体实现中，从下表2中可以看出，同一负载类型的两个vm若处于同一nc中，会使得nc性能变差，不同负载类型的两个vm若处于同一nc中，会使得nc性能变好。换言之，对于同一种负载类型的业务，可以把他们调度到不同的nc上，来减少对同一种资源的竞争。对不同负载类型的业务，可以将他们调度到同一个nc上，来提升对nc上空余资源的利用率。

[0205]

负载类型	CPU密集型	存储密集型	网络密集型
CPU密集型	差	好	好
存储密集型	好	差	好
网络密集型	好	好	差

[0206] 表2

[0207] 因此，在本申请实施例中，针对存在竞争vm的候选nc，可以首先将该候选nc从第一候选物理机列表中删除，得到第二候选物理机列表。

[0208] 在实际中，为了更好的提升对nc上空余资源的利用率，可以进一步根据竞争vm在其运行的候选nc上的负载情况来决定是否需要从第一候选物理机列表中删除存在竞争vm的候选物理机，若竞争vm在其运行的候选nc上的负载较重，则从第一候选物理机列表中删除存在竞争vm的候选物理机，否则，若竞争vm在其运行的候选nc上的负载较轻，则可以不删除存在竞争vm的候选物理机，直接将第一候选物理机列表作为第二候选物理机列表。

[0209] 例如，若候选nc上竞争vm的负载类型对应的资源利用率为该nc上对应的资源利用率的80％，若对于同样负载类型的目标vm来说，如果该目标vm迁移入候选nc，则会使得该候选nc超负荷，因此，可以将该存在竞争vm的候选nc从第一候选物理机列表中删除。

[0210] 又如，若候选nc上竞争vm的负载类型对应的资源利用率为该nc上对应的资源利用率的30％，若对于同样负载类型的目标vm来说，如果该目标vm迁移入候选nc，候选nc也不会

超负荷，因此，可以不将该候选nc从第一候选物理机列表中删除。

[0211] 子步骤S53，从所述第二候选物理机列表中确定目标物理机。

[0212] 确定第二候选物理机列表以后，可以进一步从该第二候选物理机列表中确定进行热迁移的目标物理机。

[0213] 在本申请实施例的一种优选实施例中，子步骤S53进一步可以为：

[0214] 将第二候选物理机列表中特定资源利用率最小的候选物理机作为目标物理机。

[0215] 针对源nc负荷较重的场景，获得第二候选物理机列表中每个候选nc的特定资源利用率以后，可以将特定资源利用率最小的候选nc作为目标nc，即将负载最轻的nc作为目标nc，使得将目标vm迁移到目标nc后，可以减轻源nc的负载。

[0216] 针对源nc负荷使用率不高的场景，获得第二候选物理机列表中每个候选nc的特定资源利用率以后，将特定资源利用率最小的候选nc作为目标nc，将目标vm迁移到目标nc后，可以释放源nc的资源。

[0217] 当然，上述物理机的特定资源利用率也可以用CPU资源利用率、存储资源利用率、以及网络资源利用率进行平均或加权平均后得到的平均值代替，本申请实施例对此不作限定。

[0218] 步骤205，确定所述目标虚拟机的关联虚拟机；

[0219] 在本申请实施例的一种优选实施例中，步骤205可以包括如下子步骤：

[0220] 子步骤S61，若所述目标虚拟机与某一虚拟机经由源物理机和/或另一物理机进行通信，则判定所述目标虚拟机与所述虚拟机基于同一个物理机或不同物理机存在交互关系；

[0221] 子步骤S62，将与所述目标虚拟机存在交互关系的虚拟机作为关联虚拟机。

[0222] 应用于本申请实施例，除了可以确定目标vm的竞争vm以外，还可以确定目标vm的关联vm，关联vm是指与目标vm存在交互关系的vm。

[0223] 需要说明的是，该关联vm与所述目标vm可以处于同一源物理机中和/或处于不同源物理机中，本申请实施例对此不作限定。具体的，若目标虚拟机与某一虚拟机经由源物理机进行通信，则判定所述目标虚拟机与所述虚拟机基于同一个源物理机存在交互关系。若目标虚拟机与某一虚拟机经由源物理机和另一物理机进行通信，则判定所述目标虚拟机与所述虚拟机基于不同的物理机存在交互关系。

[0224] 在本发明实施例的一种优选实施例中，子步骤S61进一步可以包括如下子步骤：

[0225] 子步骤S611，采用源物理机中预先布置的探测点对所述目标虚拟机的进出数据包进行探测；

[0226] 子步骤S612，若所述目标虚拟机与某一虚拟机之间具有进出数据包，则判定所述目标虚拟机与某一虚拟机进行通信。

[0227] 在具体实现中，可以通过VIF(Virtual Interface，虚拟网络接口)在源nc上布置探测点以对目标vm的进出数据包进行探测，根据该进出数据包的流向可以获得与目标虚拟机存在交互关系的一个或多个虚拟机，该具有交互关系的一个或多个虚拟机即为目标vm的关联vm。

[0228] 步骤206，将所述目标虚拟机以及对应的关联虚拟机迁移到所述目标物理机中。

[0229] 在本申请实施例中，在得到目标vm以及其对应的关联vm以后，可以将该目标vm以

及关联vm迁移到同一目标物理机中，这是由于两个vm在同一nc上进行通信相比于在不同nc上进行通信，存在如下优势：

[0230] 1、相对于vm之间在同一nc上通信的情况，vm之间的网络通信在跨nc的情况下，通过的网络链路更长，对应所需要的时间也更长。

[0231] 2、在虚拟化框架上，同一nc上的两个vm可以共享内存，则在共享内存的基础上，可以在位于同一nc上的vm之间建立通信通道，该通道包含两个虚拟共享内存环形队列用于接收和发送数据，一个事件通道用于发送事件信号，在虚拟机监视器中为每对vm维护两个共享内存环形队列，通信双方将本地虚拟共享内存环形队列映射到这两个队列，进而实现高带宽的数据通路。

[0232] 3、进一步的，基于共享内存的vm间的通信可以减少网络交换过程中的报文拷贝次数，从而减少系统开销。

[0233] 4、在虚拟化框架上，同一nc上的两个vm还可以进行物理网卡L2层直接交换，两者的交互可以无需host协议栈，缩短报文处理流程的路径，从而减少系统开销。

[0234] 总之，在虚拟化的框架下，两个vm在同一个nc上进行数据交互时，不同的虚拟化平台在这一点上会做一些针对性的优化措施，如：共享内存，零拷贝、物理网卡L2层直接交换等等，这些优化措施可以减少网络交换过程中的报文拷贝次数，从而减少系统开销。而在跨nc通信时，这些优化是没有，因此，在系统开销上比处于同一nc上大。

[0235] 为了使本领域技术人员更好地理解本申请实施例，以下通过一个整体应用场景对本申请实施例加以说明。

[0236] 通过对分布式文件系统做最简化的分析，可以将集群之间的关系抽象为以下几个组合：

[0237] 1、vm和存储；2、vm和网络；3、vm和vm

[0238] 分别说明如下：

[0239] 1、vm和存储

[0240] 现有的云计算存储方案，可以分为存储计算分离，和存储计算混合的场景，不管是哪种场景，可以有一个预定的假设就是，vm在不同的nc上运行时到不同的存储的距离是不同的。例如在某计算机集群下，vm的一个磁盘数据可能会被分配到三个nc(假设为nc1～nc3)上保存三个副本，当vm在nc1～nc3上运行时，对存储的访问路径是最短的，而当vm在其他nc(例如，nc4及后面编号的nc)上运行时，对存储的访问路径是最长的。

[0241] 因此，通过分析vm的磁盘数据在分布式文件系统中的nc的存储位置，可以得到该vm在不同nc运行时，对磁盘数据的存储访问代价，从而确定该vm在迁移时可选的最优备选nc列表。

[0242] 例如，vm在不同nc运行时，对磁盘数据的存储访问路径如下表3所示：

[0243]

vm所在nc	nc1	nc2	nc3	nc4
vm1	1	1	1	3
vm2	3	1	1	1
vm3	1	3	1	1

[0244] 表3

[0245] 从表2可知，vm1的最优备选nc列表为nc1、nc2、nc3；vm2的最优备选nc列表为nc2、nc3、nc4；vm3的最优备选nc列表为nc1、nc3、nc4。

[0246] 2、vm和网络

[0247] VM和网络的关系，主要分析的是，vm和集群内部其它vm之间的产生基于网络的交互时，网络关系对vm的性能的影响。

[0248] 在虚拟化的框架下，两个vm在同一个nc上进行数据交互时，不同的虚拟化平台都会对这一点做针对性的优化，如：共享内存，零拷贝、物理网卡L2层直接交换等等，而两个vm在不同nc上进行数据交互时并不具备上述优化的优势。

[0249] 因此，通过vif布置探测点，可以获取到当前vm网络进出包中和其它vm关系最密切的vm关系列表，在迁移中，可以将当前vm和最密切的vm同时迁移到1个目的nc上。

[0250] 例如：vm2是与vm1关系最密切的vm，当源nc资源不足时，将vm1迁移到nc2之后，需要将vm2也同时迁移到nc2。

[0251] 3、vm和vm

[0252] vm和vm之间的关系，主要考虑的是vm负载之间的相互影响关系。在总体上可以将vm的负载分为CPU密集型、存储密集型和网络密集型。

[0253] 对上述三种类型，可以采用如下指标进行划分：

[0254] CPU密集型：体现着vm CPU整体占用率在整个集群中高于平均值，vm对应的后端设备的I/O队列长时间小于1个，网卡收发包带宽远远小于vm带宽的10％。

[0255] 存储密集型：体现在vm的CPU占用率整体小于整个集群中的平均值，vm对应的CPU io wait值长期大于1，后端I/O设备中I/O队列中的个数长期大于5，网卡收发包远远小于vm带宽的10％。

[0256] 网络密集型：体现在后端vm网卡中断速率高于集群中平均值5倍，网络带宽长期满负荷。

[0257] 需要说明的是，除了上述指标，本领域技术人员还可以采用其他指标，达到同样的目的。

[0258] 相同负载类型与不同负载类型的vm之间的关系可以如下表4所示，

[0259]

负载类型	CPU密集型	存储密集型	网络密集型
CPU密集型	差	好	好
存储密集型	好	差	好
网络密集型	好	好	差

[0260] 表4

[0261] 从表4可以看出，对于同一种类型的业务，应该尽可能的把他们调度到不同的nc上，来减少对同一种资源的竞争。对不同类型的业务，可以将他们调度到同一个nc上，来提升对nc上空余资源的利用率。

[0262] 在实际中，选择目标nc时，需要保证在目标nc上的vm对CPU、存储、网络的密集程度加权平均后，该nc在存储和网络、CPU上的加权平均处于整个集群的平均水平。

[0263] 基于上述三种组合的综合考虑，目标nc选择过程可以包括如下步骤：

[0264] 1.根据vm和存储的关系确定vm对应的基于存储的最优待选nc列表。

[0265] 2.根据vm和网络的关系确定vm是否有密切关联的vm用户组需要同时进行vm的迁移。

[0266] 3.根据步骤1,2中选定的最优待选nc列表以及根据vm和vm的关系,计算nc负载加权值,确定目标nc,以确保vm迁移到目标nc后,目标nc的负载资源不会超限。

[0267] 4.执行迁移,将一个或多个vm迁移到目的nc。

[0268] 本实例从vm外部依赖的资源关系(存储、网络,vm之间)的相关性,分析得到vm的最优目的nc,并进行迁移操作。

[0269] 在本申请实施例中,在热迁移过程中利用nc和nc、nc和存储、vm和vm、vm与负载之间的相关性,通过挖掘vm与其周边资源的相关性(网络、存储、cpu等),来分析该vm在哪个nc上运行是最优的,并进行迁移操作,从而提高目标vm的性能。

[0270] 需要说明的是,对于方法实施例,为了简单描述,故将其都表述为一系列的动作组合,但是本领域技术人员应该知悉,本申请实施例并不受所描述的动作顺序的限制,因为依据本申请实施例,某些步骤可以采用其他顺序或者同时进行。其次,本领域技术人员也应该知悉,说明书中所描述的实施例均属于优选实施例,所涉及的动作并不一定是本申请实施例所必须的。

[0271] 参照图3,示出了本申请的一种虚拟机热迁移的系统实施例的结构框图,所述系统可以包括如下模块:

[0272] 目标虚拟机选择模块301,用于从源物理机运行的虚拟机中选择一目标虚拟机;

[0273] 负载类型确定模块302,用于确定所述目标虚拟机的负载类型,其中,所述目标虚拟机的负载类型是根据所述目标虚拟机运行在所述源物理机时,所述目标虚拟机或所述源物理机的特定资源利用率来确定的;

[0274] 候选列表确定模块303,用于确定第一候选物理机列表,其中,所述第一候选物理机列表包括存储所述目标虚拟机的多个磁盘数据副本对应的多个物理机;

[0275] 目标物理机选择模块304,用于基于所述目标虚拟机的负载类型,从所述第一候选物理机列表中确定目标物理机;

[0276] 迁移模块305,用于将所述目标虚拟机从所述源物理机迁移到所述目标物理机中。

[0277] 在本申请实施例的一种优选实施例中,所述系统还可以包括如下模块:

[0278] 关联虚拟机确定模块,用于确定所述目标虚拟机的关联虚拟机,所述关联虚拟机为与所述目标虚拟机存在交互关系的虚拟机;

[0279] 关联虚拟机迁移模块,用于在迁移所述目标虚拟机的同时,将所述关联虚拟机迁移至与所述目标虚拟机相同的目标物理机中。

[0280] 在本申请实施例的一种优选实施例中,所述关联虚拟机确定模块包括如下子模块:

[0281] 交互关系判断子模块,用于若所述目标虚拟机与某一虚拟机经由源物理机和/或另一物理机进行通信,则判定所述目标虚拟机与所述虚拟机基于同一个物理机或不同物理机存在交互关系;

[0282] 关联虚拟机确定子模块,用于将与所述目标虚拟机存在交互关系的虚拟机作为关联虚拟机。

[0283] 在本申请实施例的一种优选实施例中,所述交互关系判断子模块包括:

[0284] 数据包探测单元，用于采用源物理机中预先布置的探测点对所述目标虚拟机的进出数据包进行探测；

[0285] 通信确定单元，用于在判定所述目标虚拟机与某一虚拟机之间具有进出数据包时，则判定所述目标虚拟机与某一虚拟机进行通信。

[0286] 在本申请实施例的一种优选实施例中，所述源物理机包括如下一个或多个：

[0287] 计算机集群中特定资源利用率高于第一资源负载阈值的物理机或特定资源利用率低于第二资源负载阈值的物理机，其中，所述第一资源负载阈值大于所述第二资源负载阈值；

[0288] 计算机集群中存在故障的物理机。

[0289] 在本申请实施例的一种优选实施例中，所述候选列表确定模块303可以包括如下子模块：

[0290] 访问路径获取子模块，用于分别获取所述目标虚拟机在计算机集群中的每个物理机上运行时访问所述目标虚拟机的不同的磁盘数据副本的访问路径；

[0291] 组织子模块，用于将访问路径的长度小于或等于预设路径阈值的物理机组织成第一候选物理机列表。

[0292] 在本申请实施例的一种优选实施例中，所述特定资源利用率包括所述目标虚拟机的CPU资源利用率，所述负载类型包括CPU密集型；

[0293] 所述负载类型确定模块包括：

[0294] CPU资源利用率获取子模块，用于当所述目标虚拟机运行在所述源物理机上时，获取所述目标虚拟机的CPU资源利用率；

[0295] 平均CPU资源利用率获取子模块，用于获取计算机集群的平均CPU资源利用率；

[0296] CPU密集型确定子模块，用于若预设时间段内所述目标虚拟机的CPU资源利用率均大于所述平均CPU资源利用率，则确定所述目标虚拟机的负载类型为CPU密集型。

[0297] 在本申请实施例的一种优选实施例中，所述特定资源利用率包括所述源物理机的I/O队列深度，和/或，所述目标虚拟机的CPU中等待I/O操作的进程数；所述负载类型包括存储密集型；

[0298] 所述负载类型确定模块包括：

[0299] 存储资源获取子模块，用于当所述目标虚拟机运行在所述源物理机上时，获取所述源物理机的I/O队列深度，和/或，所述目标虚拟机的CPU中等待I/O操作的进程数；

[0300] 存储密集型确定子模块，用于若所述源物理机的I/O队列深度在预设时间段内均大于预设深度阈值，和/或，所述目标虚拟机的CPU中等待I/O操作的进程数在预设时间段内均大于预设进程数阈值，则确定所述目标虚拟机的负载类型为存储密集型。

[0301] 在本申请实施例的一种优选实施例中，所述特定资源利用率包括源物理机中网卡收发数据包的速率；所述负载类型包括网络密集型；

[0302] 所述负载类型确定模块包括：

[0303] 收发速率获取子模块，用于当所述目标虚拟机运行在所述源物理机上时，获取所述源物理机中网卡收发数据包的速率；

[0304] 平均速率获取子模块，用于获取计算机集群中收发数据包的平均速率；

[0305] 网络集型确定子模块，用于若源物理机中网卡收发数据包的速率在预设时间段内

均大于所述平均速率，则确定所述目标虚拟机的负载类型为网络密集型。

[0306] 在本申请实施例的一种优选实施例中，所述目标物理机选择模块304可以包括如下子模块：

[0307] 竞争虚拟机确定子模块，用于基于所述目标虚拟机的负载类型，分别判断每个候选物理机中是否存在与所述目标虚拟机存在资源竞争关系的竞争虚拟机；

[0308] 第二候选物理机列表生成子模块，用于在所述第一候选物理机列表中删除存在竞争vm的候选物理机，得到第二候选物理机列表；

[0309] 目标物理机选取子模块，用于从所述第二候选物理机列表中确定目标物理机。

[0310] 在本申请实施例的一种优选实施例中，所述竞争虚拟机确定子模块可以包括如下单元：

[0311] 负载类型获取单元，用于分别获取所述候选物理机中运行的每个虚拟机的负载类型；

[0312] 确定单元，用于若所述候选物理机中存在与所述目标虚拟机的负载类型相同的虚拟机，则判定所述候选物理机中存在与所述目标虚拟机存在资源竞争关系的竞争虚拟机。

[0313] 在本申请实施例的一种优选实施例中，所述目标物理机选取子模块还用于：

[0314] 将第二候选物理机列表中特定资源利用率最小的候选物理机作为目标物理机。

[0315] 对于系统实施例而言，由于其与上述方法实施例基本相似，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

[0316] 本说明书中的各个实施例均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的部分互相参见即可。

[0317] 本领域内的技术人员应明白，本申请实施例的实施例可提供为方法、装置、或计算机程序产品。因此，本申请实施例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，本申请实施例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。

[0318] 本申请实施例是参照根据本申请实施例的方法、终端设备(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操作指令实现流程图和/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操作指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理终端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终端设备的处理器执行的操作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。

[0319] 这些计算机程序操作指令也可存储在能引导计算机或其他可编程数据处理终端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操作指令产生包括操作指令装置的制造品，该操作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。

[0320] 这些计算机程序操作指令也可装载到计算机或其他可编程数据处理终端设备上，使得在计算机或其他可编程终端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终端设备上执行的操作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。

[0321] 尽管已描述了本申请实施例的优选实施例，但本领域内的技术人员一旦得知了基本创造性概念，则可对这些实施例做出另外的变更和修改。所以，所附权利要求意欲解释为包括优选实施例以及落入本申请实施例范围的所有变更和修改。

[0322] 最后，还需要说明的是，在本文中，诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来，而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者终端设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者终端设备所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括所述要素的过程、方法、物品或者终端设备中还存在另外的相同要素。

[0323] 以上对本申请所提供的一种虚拟机热迁移的方法及系统进行了详细介绍，本文中应用了具体个例对本申请的原理及实施方式进行了阐述，以上实施例的说明只是用于帮助理解本申请的方法及其核心思想；同时，对于本领域的一般技术人员，依据本申请的思想，在具体实施方式及应用范围上均会有改变之处，综上所述，本说明书内容不应理解为对本申请的限制。

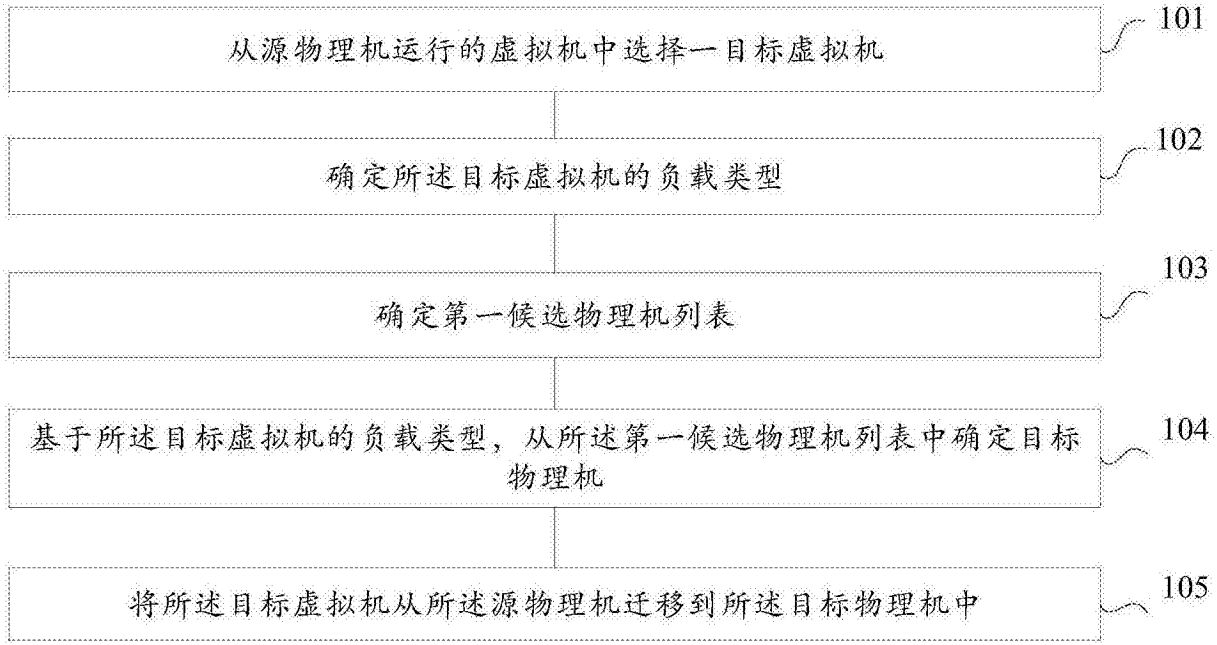

从源物理机运行的虚拟机中选择一目标虚拟机
101
确定所述目标虚拟机的负载类型
102
确定第一候选物理机列表
103
基于所述目标虚拟机的负载类型，从所述第一候选物理机列表中确定目标物理机
104
将所述目标虚拟机从所述源物理机迁移到所述目标物理机中
105

图1

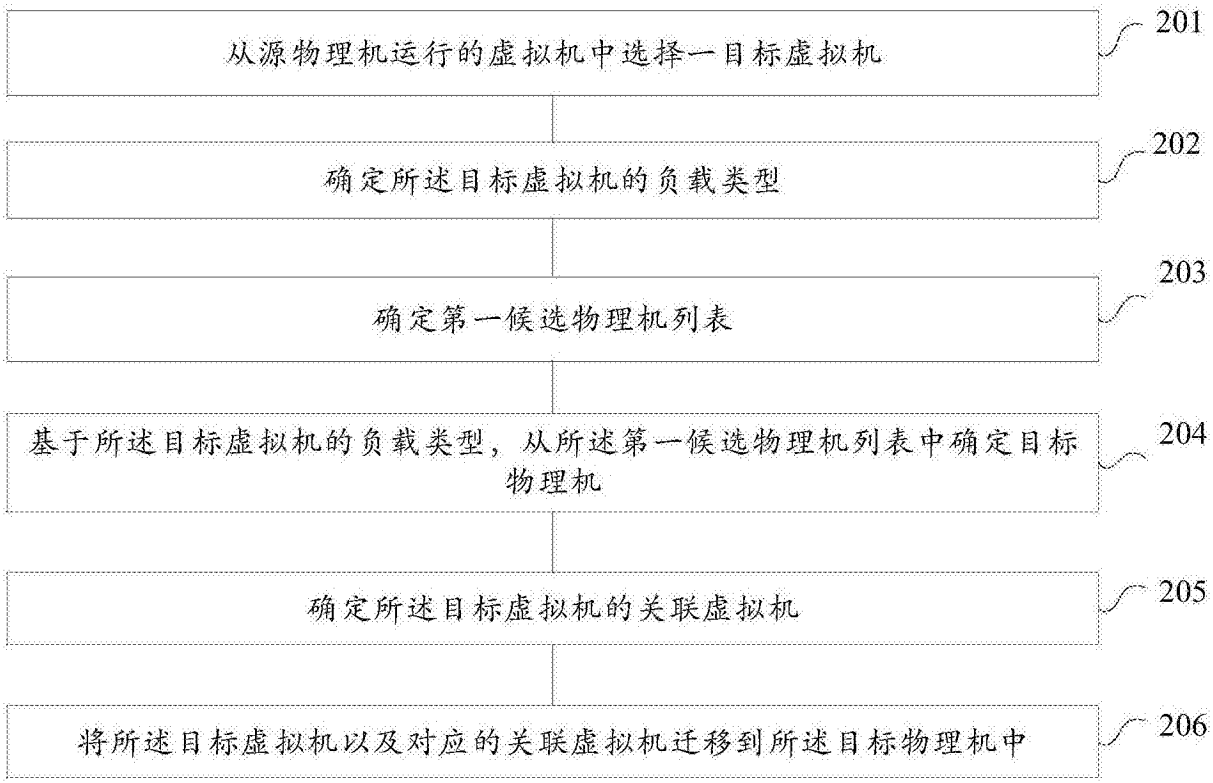

从源物理机运行的虚拟机中选择一目标虚拟机
201
确定所述目标虚拟机的负载类型
202
确定第一候选物理机列表
203
基于所述目标虚拟机的负载类型，从所述第一候选物理机列表中确定目标物理机
204
确定所述目标虚拟机的关联虚拟机
205
将所述目标虚拟机以及对应的关联虚拟机迁移到所述目标物理机中
206

图2

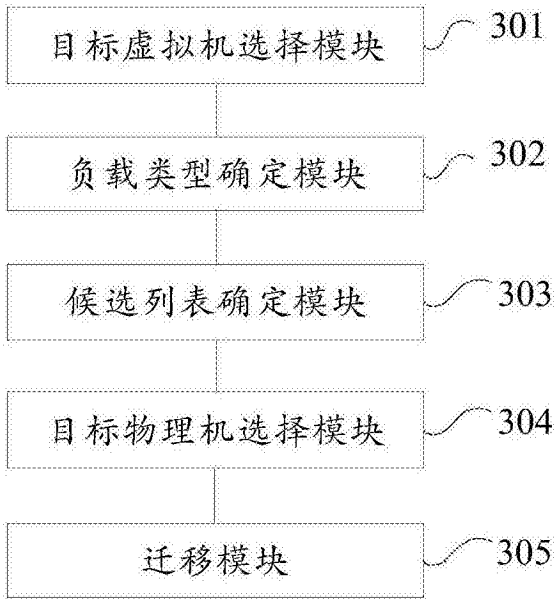

目标虚拟机选择模块
301
负载类型确定模块
302
候选列表确定模块
303
目标物理机选择模块
304
迁移模块
305

图3