



(12) 发明专利申请

(10) 申请公布号 CN 115563078 A

(43) 申请公布日 2023. 01. 03

(21) 申请号 202210784190.5

(22) 申请日 2022.06.28

(30) 优先权数据

63/217,714 2021.07.01 US

17/450,409 2021.10.08 US

(71) 申请人 三星电子株式会社

地址 韩国京畿道

(72) 发明人 C-C.C-J.A.吴 F.维克拉姆辛格

(74) 专利代理机构 北京市柳沈律师事务所

11105

专利代理师 刘虹

(51) Int.Cl.

G06F 16/21 (2019.01)

G06F 16/23 (2019.01)

G06F 3/06 (2006.01)

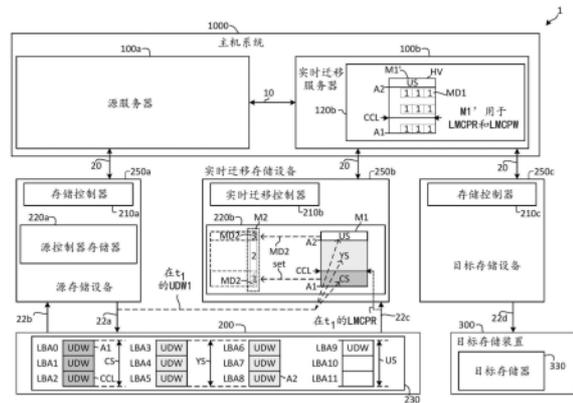
权利要求书4页 说明书25页 附图16页

(54) 发明名称

用于实时迁移的脏页位图中减少脏位跟踪的机制

(57) 摘要

本发明涉及用于实时迁移的脏页位图中减少脏位跟踪的机制。一种用于跟踪实时迁移的数据复制进度的方法,包括由存储控制器将第一数据结构传送到实时迁移服务器,该第一数据结构包括指示要从源存储装置复制到目标存储装置的源数据的位置的第一状态标识符;以及由存储控制器基于实时迁移服务器的第一当前复制位置,在第二数据结构中选择性地生成第二状态标识符或从第二数据结构中选择性地清除第二状态标识符,该第二状态标识符指示第一用户数据写入源存储装置的位置。



1. 一种用于跟踪实时迁移的数据复制进度的方法,所述方法包括:

由存储控制器将第一数据结构传送到实时迁移服务器,所述第一数据结构包括第一状态标识符,所述第一状态标识符指示要从源存储装置复制到目标存储装置的源数据的位置;以及

由存储控制器基于所述实时迁移服务器的第一当前复制位置,在第二数据结构中选择性地生成第二状态标识符或从第二数据结构中选择性地清除第二状态标识符,所述第二状态标识符指示第一用户数据写入到源存储装置的位置。

2. 根据权利要求1所述的方法,还包括:

由存储控制器监视第一当前复制位置,所述第一当前复制位置与基于第一数据结构的源数据的复制相对应;以及

由存储控制器监视第一用户数据写入,

其中,在第二数据结构中选择性地生成第二状态标识符包括:

由存储控制器基于所述第一当前复制位置,跟踪源存储装置的计划但尚未被复制的区段和源存储装置的计划且已复制的区段;以及

基于所述计划但尚未被复制的区段,确定是否在第二数据结构中生成第二状态标识符。

3. 根据权利要求2所述的方法,其中,跟踪计划但尚未被复制的区段和计划且已复制的区段包括:跟踪与连续索引的开始地址和连续索引的结束地址相对应的当前复制索引,所述连续索引与计划要复制的区段相关联,所述计划要复制的区段包括连续的存储位置。

4. 根据权利要求2所述的方法,其中,跟踪计划但尚未被复制的区段和计划且已复制的区段包括:跟踪计划但尚未被复制的位图和计划且已复制的位图,所述计划但尚未被复制的位图和计划且已复制的位图与计划要复制的区段相关联,所述计划要复制的区段包括一个或多个分布式存储位置。

5. 根据权利要求2所述的方法,其中,跟踪计划但尚未被复制的区段和计划且已复制的区段包括:

跟踪与计划要复制的区段相关联的计划但尚未被复制的位图,所述计划要复制的区段包括一个或多个分布式存储位置;以及

使用所述计划但尚未被复制的位图,通过反转用于跟踪所述计划但尚未被复制的区段的位状态,来跟踪计划且已复制的区段。

6. 根据权利要求1所述的方法,其中,从第二数据结构中选择性地清除第二状态标识符包括:

基于所述第一用户数据写入在第二数据结构中生成第二状态标识符;以及

由存储控制器基于所述第二状态标识符与实时迁移服务器的第一当前复制位置相关联,从第二数据结构中清除第二状态标识符。

7. 根据权利要求1所述的方法,还包括:

由存储控制器监视第一当前复制位置,所述第一当前复制位置与基于第一数据结构的源数据的复制相对应;

由存储控制器监视第一用户数据写入;

将第二数据结构传送到实时迁移服务器;

由存储控制器监视实时迁移服务器的第二当前复制位置,所述第二当前复制位置与基于第二数据结构的源数据的复制相对应;

由存储控制器监视对源存储装置的第二用户数据写入,所述第二用户数据写入在第二数据结构的传送期间或之后发生;以及

由存储控制器基于所述实时迁移服务器的第二当前复制位置,在第三数据结构中选择性地生成第三状态标识符或从第三数据结构中选择性地清除第三状态标识符,所述第三状态标识符指示第二用户数据写入的位置。

8. 一种用于跟踪实时迁移的数据复制进度的存储设备,所述存储设备被配置为:

将第一数据结构从存储控制器传送到实时迁移服务器,所述第一数据结构包括第一状态标识符,所述第一状态标识符指示要从源存储装置复制到目标存储装置的源数据的位置;并且

由存储控制器基于所述实时迁移服务器的第一当前复制位置,在第二数据结构中选择性地生成第二状态标识符或从第二数据结构中选择性地清除第二状态标识符,所述第二状态标识符指示对源存储装置的第一用户数据写入的位置。

9. 根据权利要求8所述的存储设备,其中,所述存储设备被配置为:

监视第一当前复制位置,所述第一当前复制位置与基于第一数据结构的源数据的复制相对应;以及

监视第一用户数据写入,

其中,在第二数据结构中选择性地生成第二状态标识符包括:

基于所述第一当前复制位置,跟踪源存储装置的计划但尚未被复制的区段和源存储装置的计划且已复制的区段;以及

基于所述计划但尚未被复制的区段,确定是否在第二数据结构中生成第二状态标识符。

10. 根据权利要求9所述的存储设备,其中,跟踪计划但尚未被复制的区段和计划且已复制的区段包括:跟踪与连续索引的开始地址和连续索引的结束地址相对应的当前复制索引,所述连续索引与计划要复制的区段相关联,所述计划要复制的区段包括连续的存储位置。

11. 根据权利要求9所述的存储设备,其中,跟踪计划但尚未被复制的区段和计划且已复制的区段包括:跟踪计划但尚未被复制的位图和计划且已复制的位图,所述计划但尚未被复制的位图和计划且已复制的位图与计划要复制的区段相关联,所述计划要复制的区段包括一个或多个分布式存储位置。

12. 根据权利要求9所述的存储设备,其中,跟踪计划但尚未被复制的区段和计划且已复制的区段包括:

跟踪与计划要复制的区段相关联的计划但尚未被复制的位图,所述计划要复制的区段包括一个或多个分布式存储位置;以及

使用所述计划但尚未被复制的位图,通过反转用于跟踪所述计划但尚未被复制的区段的位状态,来跟踪计划且已复制的区段。

13. 根据权利要求8所述的存储设备,其中,所述存储设备被配置为:

监视第一当前复制位置,所述第一当前复制位置与基于第一数据结构的源数据的复制

相对应;以及

监视第一用户数据写入,

其中,从第二数据结构中选择性地清除第二状态标识符包括:

基于所述第一用户数据写入在第二数据结构中生成第二状态标识符;以及

基于所述第二状态标识符与实时迁移服务器的第一当前复制位置相关联,从第二数据结构中清除第二状态标识符。

14. 根据权利要求8所述的存储设备,其中,所述存储设备被配置为:

将第二数据结构传送到实时迁移服务器;

监视实时迁移服务器的第二当前复制位置,所述第二当前复制位置与基于第二数据结构的源数据的复制相对应;

监视对源存储装置的第二用户数据写入,所述第二用户数据写入在第二数据结构的传送期间或之后发生;以及

基于所述实时迁移服务器的第二当前复制位置,在第三数据结构中选择性地生成第三状态标识符或从第三数据结构中选择性地清除第三状态标识符,所述第三状态标识符指示第二用户数据写入的位置。

15. 一种用于跟踪实时迁移的数据复制进度的系统,所述系统包括:

实时迁移服务器;以及

存储设备,

其中,所述存储设备被配置为:

将第一数据结构从存储控制器传送到实时迁移服务器,所述第一数据结构包括第一状态标识符,所述第一状态标识符指示要从源存储装置复制到目标存储装置的源数据的位置;并且

由存储控制器基于所述实时迁移服务器的第一当前复制位置,在第二数据结构中选择性地生成第二状态标识符或从第二数据结构中选择性地清除第二状态标识符,所述第二状态标识符指示对源存储装置的第一用户数据写入的位置。

16. 根据权利要求15所述的系统,其中,所述存储设备被配置为:

监视第一当前复制位置,所述第一当前复制位置与基于第一数据结构的源数据的复制相对应;并且

监视第一用户数据写入,

其中,在第二数据结构中选择性地生成第二状态标识符包括:

基于所述第一当前复制位置,跟踪源存储装置的计划但尚未被复制的区段和源存储装置的计划且已复制的区段;以及

基于所述计划但尚未被复制的区段,确定是否在第二数据结构中生成第二状态标识符。

17. 根据权利要求16所述的系统,其中,跟踪计划但尚未被复制的区段和计划且已复制的区段包括:跟踪与连续索引的开始地址和连续索引的结束地址相对应的当前复制索引,所述连续索引与计划要复制的区段相关联,所述计划要复制的区段包括连续的存储位置。

18. 根据权利要求16所述的系统,其中,跟踪计划但尚未被复制的区段和计划且已复制的区段包括:跟踪计划但尚未被复制的位图和计划且已复制的位图,所述计划但尚未被复

制的位图和计划且已复制的位图与计划要复制的区段相关联,所述计划要复制的区段包括一个或多个分布式存储位置。

19. 根据权利要求16所述的系统,其中,跟踪计划但尚未被复制的区段和计划且已复制的区段包括:

跟踪与计划要复制的区段相关联的计划但尚未被复制的位图,所述计划要复制的区段包括一个或多个分布式存储位置;以及

使用所述计划但尚未被复制的位图,通过反转用于跟踪所述计划但尚未被复制的区段的位状态,来跟踪计划且已复制的区段。

20. 根据权利要求15所述的系统,其中,所述存储设备被配置为:

监视第一当前复制位置,所述第一当前复制位置与基于第一数据结构的源数据的复制相对应;并且

监视第一用户数据写入,

其中,从第二数据结构中选择性地清除第二状态标识符包括:

基于所述第一用户数据写入在第二数据结构中生成第二状态标识符;以及

由存储控制器基于所述第二状态标识符与实时迁移服务器的第一当前复制位置相关联,从第二数据结构中清除第二状态标识符。

用于实时迁移的脏页位图中减少脏位跟踪的机制

[0001] 相关申请的交叉引用

[0002] 本申请要求于2021年7月1日提交的美国临时申请序列号63/217,714的优先权和权益,该临时申请题为“ACTIVE ENTRIES REDUCTION ALGORITHM IN THE DIRTY BITMAP FOR LIVE MIGRATION”,其全部内容通过引用结合于此。

技术领域

[0003] 根据本公开的实施例的一个或多个方面涉及计算机存储,并且更具体地,涉及用于跟踪实时迁移操作的数据复制进度(progress)的方法、设备和系统。

背景技术

[0004] 本背景技术部分仅用于提供上下文,并且本部分中任何实施例或概念的公开不构成对所述实施例或概念为现有技术的承认。

[0005] 在计算机存储的领域中,实时迁移(live migration,LM)操作(或实时迁移过程)可以指在源存储装置潜在从源服务器(例如,普通源服务器或普通源主机,其不是实时迁移服务器或实时迁移主机)接收用户数据写入时,将数据从源存储装置复制到目标存储装置。数据结构(例如位图)可以用于标识源存储装置内要复制到目标存储装置的初始数据位置。实时迁移服务器可以使用(例如,参考)数据结构来执行实时迁移过程,该过程可以包括从源存储装置复制相关数据(例如,初始状态下的映射的数据)并将相关数据写入目标存储装置。因为复制和写入过程可能花费大量时间,所以在将初始数据结构传送到实时迁移服务器之后,数据结构(例如,位图)也可以用于跟踪在源存储装置的相关位置处发生的后续用户数据写入的位置。跟踪后续用户数据写入的位置可以允许迁移服务器基于用户数据写入在源存储装置处的位置来更新目标存储装置处的相关数据。然而,用于跟踪后续数据写入的这种数据结构(例如,“脏位图(dirty bitmap)”)可以包括元数据(例如,一个或多个“脏”位或状态标识符),该元数据标识在数据写入被记录在脏位图中时尚未被实时迁移服务器复制的相关源数据的位置。因此,当实时迁移服务器使用初始数据结构(例如,初始脏位图)执行数据从源存储装置到目标存储装置的初始复制时,以及当实时迁移服务器使用跟踪的数据结构(例如,第二脏位图或稍后的脏位图)执行以后的复制时,可能发生冗余复制,这减慢了实时迁移过程。

[0006] 因此,需要高效管理位图使用以更新与实时迁移过程相关联的数据的方法、设备和系统。

发明内容

[0007] 本公开的实施例的各方面涉及计算机存储系统,并且提供了对用于标识源存储装置中具有要复制到目标存储装置的数据更新的位置的元数据的管理和跟踪的改进。

[0008] 根据本公开的一些实施例,提供了一种用于跟踪实时迁移的数据复制进度的方法,该方法包括由存储控制器将第一数据结构传送到实时迁移服务器,该第一数据结构包

括第一状态标识符,该第一状态标识符指示要从源存储装置复制到目标存储装置的源数据的位置;以及由存储控制器基于实时迁移服务器的第一当前复制位置,在第二数据结构中选择性地生成第二状态标识符或从第二数据结构中选择性地清除第二状态标识符,该第二状态标识符指示第一用户数据写入到源存储装置的位置。

[0009] 该方法还可以包括由存储控制器监视第一当前复制位置,该第一当前复制位置与基于第一数据结构的源数据的复制相对应;以及由存储控制器监视第一用户数据写入,其中,在第二数据结构中选择性地生成第二状态标识符包括由存储控制器基于第一当前复制位置,跟踪源存储装置的计划但尚未被复制的区段和源存储装置的计划且已复制的区段;以及基于计划但尚未被复制的区段,确定是否在第二数据结构中生成第二状态标识符。

[0010] 跟踪计划但尚未被复制的区段和计划且已复制的区段可以包括跟踪与连续索引的开始地址和连续索引的结束地址相对应的当前复制索引,该连续索引与计划要复制的区段相关联,该计划要复制的区段包括连续的存储位置。

[0011] 跟踪计划但尚未被复制的区段和计划且已复制的区段可以包括跟踪计划但尚未被复制的位图和计划且已复制的位图,该计划但尚未被复制的位图和计划且已复制的位图与计划要复制的区段相关联,该计划要复制的区段包括一个或多个分布式存储位置。

[0012] 跟踪计划但尚未被复制的区段和计划且已复制的区段可以包括跟踪与计划要复制的区段相关联的计划但尚未被复制的位图,该计划要复制的区段包括一个或多个分布式存储位置;以及使用计划但尚未被复制的位图,通过反转用于跟踪计划但尚未被复制的区段的位状态,来跟踪计划且已复制的区段。

[0013] 从第二数据结构中选择性地清除第二状态标识符可以包括基于第一用户数据写入在第二数据结构中生成第二状态标识符;以及由存储控制器基于第二状态标识符与实时迁移服务器的第一当前复制位置相关联,从第二数据结构中清除第二状态标识符。

[0014] 该方法还可以包括由存储控制器监视第一当前复制位置,该第一当前复制位置与基于第一数据结构的源数据的复制相对应;由存储控制器监视第一用户数据写入;将第二数据结构传送到实时迁移服务器;由存储控制器监视实时迁移服务器的第二当前复制位置,该第二当前复制位置与基于第二数据结构的源数据的复制相对应;由存储控制器监视对源存储装置的第二用户数据写入,该第二用户数据写入在第二数据结构的传送期间或之后发生;以及由存储控制器基于实时迁移服务器的第二当前复制位置,在第三数据结构中选择性地生成第三状态标识符或从第三数据结构中选择性地清除第三状态标识符,该第三状态标识符指示第二用户数据写入的位置。

[0015] 根据本公开的其他实施例,提供了一种用于跟踪实时迁移的数据复制进度的存储设备,该存储设备被配置为将第一数据结构从存储控制器传送到实时迁移服务器,该第一数据结构包括第一状态标识符,该第一状态标识符指示要从源存储装置复制到目标存储装置的源数据的位置;以及由存储控制器基于实时迁移服务器的第一当前复制位置,在第二数据结构中选择性地生成第二状态标识符或从第二数据结构中选择性地清除第二状态标识符,该第二状态标识符指示对源存储装置的第一用户数据写入的位置。

[0016] 该存储设备可以被配置为监视第一当前复制位置,该第一当前复制位置与基于第一数据结构的源数据的复制相对应;以及监视第一用户数据写入,其中,在第二数据结构中选择性地生成第二状态标识符包括:基于第一当前复制位置,跟踪源存储装置的计划但尚

未被复制的区段和源存储装置的计划且已复制的区段;以及基于计划但尚未被复制的区段,确定是否在第二数据结构中生成第二状态标识符。

[0017] 跟踪计划但尚未被复制的区段和计划且已复制的区段可以包括跟踪与连续索引的开始地址和连续索引的结束地址相对应的当前复制索引,该连续索引与计划要复制的区段相关联,该计划要复制的区段包括连续的存储位置。

[0018] 跟踪计划但尚未被复制的区段和计划且已复制的区段可以包括跟踪计划但尚未被复制的位图和计划且已复制的位图,该计划但尚未被复制的位图和计划且已复制的位图与计划要复制的区段相关联,该计划要复制的区段包括一个或多个分布式存储位置。

[0019] 跟踪计划但尚未被复制的区段和计划且已复制的区段可以包括跟踪与计划要复制的区段相关联的计划但尚未被复制的位图,该计划要复制的区段包括一个或多个分布式存储位置;以及使用计划但尚未被复制的位图,通过反转用于跟踪计划但尚未被复制的区段的位状态,来跟踪计划且已复制的区段。

[0020] 该存储设备可以被配置为监视第一当前复制位置,该第一当前复制位置与基于第一数据结构的源数据的复制相对应;以及监视第一用户数据写入,其中,从第二数据结构中选择性地清除第二状态标识符包括:基于第一用户数据写入在第二数据结构中生成第二状态标识符;以及基于第二状态标识符与实时迁移服务器的第一当前复制位置相关联,从第二数据结构中清除第二状态标识符。

[0021] 该存储设备可以被配置为将第二数据结构传送到实时迁移服务器;监视实时迁移服务器的第二当前复制位置,该第二当前复制位置与基于第二数据结构的源数据的复制相对应;监视对源存储装置的第二用户数据写入,该第二用户数据写入在第二数据结构的传送期间或之后发生;以及基于实时迁移服务器的第二当前复制位置,在第三数据结构中选择性地生成第三状态标识符或从第三数据结构中选择性地清除第三状态标识符,该第三状态标识符指示第二用户数据写入的位置。

[0022] 根据本公开的其他实施例,提供了一种用于跟踪实时迁移的数据复制进度的系统,该系统包括实时迁移服务器和存储设备,其中,该存储设备被配置为将第一数据结构从存储控制器传送到实时迁移服务器,该第一数据结构包括第一状态标识符,该第一状态标识符指示要从源存储装置复制到目标存储装置的源数据的位置;并且由存储控制器基于实时迁移服务器的第一当前复制位置,在第二数据结构中选择性地生成第二状态标识符或从第二数据结构中选择性地清除第二状态标识符,该第二状态标识符指示对源存储装置的第一用户数据写入的位置。

[0023] 该存储设备可以被配置为:监视第一当前复制位置,该第一当前复制位置与基于第一数据结构的源数据的复制相对应;并且监视第一用户数据写入,其中,在第二数据结构中选择性地生成第二状态标识符包括基于第一当前复制位置,跟踪源存储装置的计划但尚未被复制的区段和源存储装置的计划且已复制的区段;以及基于计划但尚未被复制的区段,确定是否在第二数据结构中生成第二状态标识符。

[0024] 跟踪计划但尚未被复制的区段和计划且已复制的区段可以包括跟踪与连续索引的开始地址和连续索引的结束地址相对应的当前复制索引,该连续索引与计划要复制的区段相关联,该计划要复制的区段包括连续的存储位置。

[0025] 跟踪计划但尚未被复制的区段和计划且已复制的区段可以包括跟踪计划但尚未

被复制的位图和计划且已复制的位图,该计划但尚未被复制的位图和计划且已复制的位图与计划要复制的区段相关联,该计划要复制的区段包括一个或多个分布式存储位置。

[0026] 跟踪计划但尚未被复制的区段和计划且已复制的区段可以包括跟踪与计划要复制的区段相关联的计划但尚未被复制的位图,该计划要复制的区段包括一个或多个分布式存储位置;以及使用计划但尚未被复制的位图,通过反转用于跟踪计划但尚未被复制的区段的位状态,来跟踪计划且已复制的区段。

[0027] 该存储设备可以被配置为监视第一当前复制位置,该第一当前复制位置与基于第一数据结构的源数据的复制相对应;并且监视第一用户数据写入,其中,从第二数据结构中选择性地清除第二状态标识符包括:基于第一用户数据写入在第二数据结构中生成第二状态标识符;以及由存储控制器基于第二状态标识符与实时迁移服务器的第一当前复制位置相关联,从第二数据结构中清除第二状态标识符。

附图说明

[0028] 参考以下附图描述了本公开的非限制性和非穷尽性实施例,其中,除非另有说明,否则各视图中相同的附图标记表示相同的部分。

[0029] 图1是根据本公开的一些实施例的,描绘用于跟踪实时迁移的数据复制进度的系统的系统图;

[0030] 图2是根据本公开的一些实施例的,描绘被动(passive)第一数据结构的传送和利用主动(active)第二数据结构跟踪用户数据写入的简化概念图;

[0031] 图3是根据本公开的一些实施例的,描绘第二数据结构的被动版本的传送和利用主动第三数据结构跟踪用户数据写入的简化概念图;

[0032] 图4A、图4B、图4C、图4D、图4E、图4F、图4G和图4H(统称为图4)是根据本公开的一些实施例的,描绘用于描绘跟踪实时迁移的数据复制进度的方法的位图视图的细节的简化概念图;

[0033] 图5A、图5B、图5C和图5D(统称为图5)是根据本公开的一些实施例的,描绘使用与连续的存储位置相关联的连续索引和第二数据结构中对第二状态标识符的选择性生成来跟踪实时迁移的数据复制进度的方法的位图视图;

[0034] 图6A、图6B、图6C、图6D和图6E(统称为图6)是根据本公开的一些实施例的,描绘使用与分布式存储位置相关联的计划但尚未被复制的位图和计划且已复制的位图来跟踪实时迁移的数据复制进度的方法的位图视图;

[0035] 图7A、图7B、图7C和图7D(统称为图7)是根据本公开的一些实施例的,描绘使用与图6的分布式存储位置相关联的计划但尚未被复制的位图和计划且已复制的位图,以及在第二数据结构中对第二状态标识符的选择性生成来跟踪实时迁移的数据复制进度的方法的位图视图;

[0036] 图8A、图8B和图8C(统称为图8)是根据本公开的一些实施例的,描绘使用与图7的第二数据结构相对应的计划但尚未被复制的位图和计划且已复制的位图以及在第三数据结构中对第三状态标识符的选择性生成,来在数据复制的后续迭代中跟踪实时迁移的数据复制进度的方法的位图视图;

[0037] 图9A、图9B和图9C(统称为图9)是根据本公开的一些实施例的,描绘使用图8的计

划但尚未被复制的位图以及通过反转用于跟踪计划但尚未被复制的区段的位状态跟踪计划且已复制的区段,来跟踪实时迁移的数据复制进度的方法的位图视图;

[0038] 图10是根据本公开的一些实施例的,包括描绘使用第二数据结构和在第二数据结构中对第二状态标识符的选择性清除来跟踪实时迁移过程的数据复制进度的系统和方法的位图视图的系统图;

[0039] 图11是根据本公开的一些实施例的,使用上文参考图1-图10描述的方法的各方面,描绘跟踪实时迁移过程的数据复制进度的方法的示例操作的简化概要流程图;

[0040] 图12是根据本公开的一些实施例的,描绘跟踪实时迁移过程的数据复制进度的选择性生成方法的示例操作的流程图;以及

[0041] 图13是根据本公开的一些实施例的,描绘跟踪实时迁移过程的数据复制进度的选择性清除方法的示例操作的流程图。

[0042] 贯穿附图的若干个视图,对应的附图标记指示对应的组件。本领域技术人员将会理解,附图中的元素是为了简单和清楚而示出的,并不一定是按比例绘制的。例如,附图中的一些元件、层和区域的尺寸相对于其他元件、层和区域可能被放大,以帮助提高各种实施例的清晰和理解。此外,可能没有示出与实施例的描述不相关的常见且公知的元素和部分,以便于更清楚地理解这些不同的实施例,并使描述清楚。

具体实施方式

[0043] 通过参考实施例的具体实施方式和附图,可以更容易理解本发明构思的特征及其实现方法。在下文中,将参考附图更详细地描述实施例。然而,所描述的实施例可以以各种不同的形式来体现,并且不应该被解释为仅限于本文所示的实施例。相反,这些实施例是作为示例提供的,使得本公开将是彻底和完整的,并且将向本领域技术人员充分传达本发明构思的方面和特征。因此,对于本领域普通技术人员来说,可能没有被描述对于本发明构思的方面和特征的完整理解来说不必要的过程、元素和技术。

[0044] 除非另有说明,否则贯穿附图和书面说明,相同的附图标记、字符或其组合表示相同的元素,因此,不再重复对其的描述。此外,为了使描述清楚,可能没有示出与实施例的描述无关的部分。在附图中,为了清楚起见,可能放大了元件、层和区域的相对大小。

[0045] 在具体实施方式中,出于解释的目的,阐述了多个具体细节,以提供对各种实施例的全面理解。然而,清楚的是,各种实施例可以在没有这些具体细节的情况下或者利用一个或多个等效布置来实践。

[0046] 应当理解,尽管术语“第零”、“第一”、“第二”、“第三”等在本文中可以用来描述各种元件、组件、区域、层和/或区段(section),但是这些元件、组件、区域、层和/或区段不应该受到这些术语的限制。这些术语用于区分一个元件、组件、区域、层或区段与另一个元件、组件、区域、层或区段。因此,在不脱离本公开的精神和范围的情况下,下面描述的第一元件、组件、区域、层或区段可以被称为第二元件、组件、区域、层或区段。

[0047] 应当理解,当元件、层、区域或组件被称为“位于”、“连接到”或“耦合到”另一个元件、层、区域或组件时,其可以直接位于、连接到或耦合到另一个元件、层、区域或组件或者可以存在一个或多个中间元件、层、区域或组件。然而,“直接连接/直接耦合”是指一个组件直接连接或耦合另一个组件,而无需中间组件。同时,描述组件之间关系的其他表述,诸如

“在之间”、“紧接在之间”或“相邻”和“直接相邻”可以类似地解释。此外，还应当理解，当元件或层被称为在两个元件或层“之间”时，其可以是两个元件或层之间的唯一元件或层、或者也可以存在一个或多个中间元件或层。

[0048] 本文使用的术语仅出于描述特定实施例的目的，且无意限制本公开。如本文所使用的，单数形式“一”和“一个”也旨在包括复数形式，除非上下文另有明确指示。还应当理解，当在本说明书中使用术语“包括”、“包含”、“具有”、“有”指定所述特征、整数、步骤、操作、元件和/或组件的存在，但是不排除一个或多个其他特征、整数、步骤、操作、元件、组件和/或其组的存在或添加。如本文所使用的，术语“和/或”包括关联所列项目中的一个或多个的任何和所有组合。

[0049] 如本文所使用的，术语“基本上”、“大约”、“近似”和类似术语用作近似术语，而非程度术语，并且旨在说明本领域普通技术人员可认识到的测量值或计算值的固有偏差。如本文所使用的“大约”或“近似”包括所述值并表示在特定值的可接受偏差范围内，如本领域普通技术人员考虑到所讨论的测量和与特定量的测量相关联的误差（即，对测量系统的限制）所确定的。例如，“大约”可以表示在一个或多个标准偏差内、或者在规定值的 $\pm 30\%$ 、 $\pm 20\%$ 、 $\pm 10\%$ 、 $\pm 5\%$ 内。此外，在描述本公开的实施例时使用“可以”是指“本公开的一个或多个实施例”。

[0050] 当一个或多个实施例可能以不同方式来实现时，特定过程次序可能以不同于所描述的次序的方式来执行。例如，两个连续描述的过程可以基本上同时执行、或者以与所描述的次序相反的次序来执行。

[0051] 所描述的组件中的任一个或组件的任何组合（例如，在本文包括的任何系统图中）可以用于执行本文包括的任何流程图的一个或多个操作。此外，(i) 这些操作是示例操作，并且可以涉及未明确涵盖的各种附加操作，并且(ii) 这些操作的时间次序可以变化。

[0052] 可以利用任何合适的硬件、固件（例如，专用集成电路）、软件或软件、固件和硬件的组合来实现根据在此描述的本公开的实施例的电子或电气设备和/或任何其他相关设备或组件。例如，这些设备的各种组件可以形成在一个集成电路（IC）芯片上或分离的IC芯片上。此外，这些设备的各种组件可以在柔性印刷电路膜、带载封装（TCP）、印刷电路板（PCB）上实现或者形成在一个基底上。

[0053] 此外，这些设备的各种组件可以是在一个或多个计算设备中的一个或多个处理器上运行，执行计算机程序指令，并与其他系统组件交互以执行本文描述的各种功能的进程或线程。计算机程序指令存储在可以在使用标准存储器设备的计算设备中实现的存储器（诸如，例如随机访问存储器（RAM））中。计算机程序指令也可以存储在其他非暂时性计算机可读介质（诸如，例如CD-ROM、闪存驱动等）中。此外，本领域技术人员应当认识到，在不脱离本公开的实施例的精神和范围的情况下，各种计算设备的功能可以被组合或集成到单个计算设备中、或者特定计算设备的功能可以分布在一个或多个其他计算设备上。

[0054] 除非另有定义，否则本文使用的所有术语（包括技术术语和科学术语）的含义与本发明构思所属领域的普通技术人员通常理解的含义相同。还应当理解，术语（诸如在常用词典中定义的那些术语）应当被解释为具有与它们在相关领域和/或本说明书的上下文中的含义一致的含义，并且不应当以理想化或过于正式的意义来解释，除非在本文中明确如此定义。

[0055] 如上所述,在计算机存储的领域中,实时迁移(LM)操作(或实时迁移过程)可以指在源存储装置潜在地从源服务器(例如,普通源服务器或主机)接收用户数据写入时,将数据从源存储装置复制到目标存储装置。实时迁移中可能存在三个操作。第一,可以从源到目的地对所有在用条目(in-use entry)进行初始复制。第二,在在用条目的初始复制期间写入的“脏”条目可以被跟踪并迭代地复制到目的地,以在初始复制之后提供更新。第三,在可以将脏条目的最后列表复制到目的地时,可以在短时间内停止复制,并且可以将提供给用户的数据从源驱动改变为目的地驱动。本公开提供了对与第二操作相关联的脏条目的跟踪的改进。

[0056] 一些跟踪方法可以在跟踪的记录(例如,数据结构)中,将在在用条目的初始复制期间发生的来自用户的所有写入的数据作为脏条目来跟踪,并且跟踪的记录可以由实时迁移服务器在初始复制完成后执行下一次复制迭代时使用。例如,当写入的地址在实时迁移服务器计划要复制但尚未复制的源存储位置(例如,一些脏条目可能位于源存储装置中计划但尚未被复制的区段)中时,这些方法可能不会停止跟踪。在脏位图中跟踪对计划但尚未被复制的区段的用户数据写入可以创建当前复制操作和下一次复制操作的冗余复制和写入操作,从而减慢了实时迁移过程。

[0057] 根据本公开的实施例的各方面的用于跟踪实时迁移的数据复制进度的机制可以:(1)减少在脏位图中跟踪数据复制进度的每次迭代中的脏条目的数量;(2)(例如,通过减少消耗功率的复制写入操作)减少实时迁移时间和功耗;(3)通过基于对实时迁移服务器的当前复制位置的监视来选择性地生成或选择性地清除脏位图条目,从而减少基于脏位图复制源数据的与每次迭代相关联的时延以及每次迭代累积的时延;以及(4)由于减少了复制读取和复制写入的量,因此减少了用户写入(例如,普通用户写入)和用户读取(例如,普通用户读取)的性能干扰。

[0058] 本文所使用的“状态标识符”(或“状态标识符位”)是一种元数据,其用于指示源存储装置的给定存储位置(例如,逻辑块地址(LBA)、字节地址等)是否包括要从源存储装置复制到目标存储装置的数据。例如,设置为“1”的状态标识符可以指示对应的存储位置包括要从源存储装置复制到目标存储装置的数据,而设置为“0”的状态标识符位可以指示源存储装置的给定存储位置不包括要从源存储装置复制到目标存储装置的数据。状态标识符可以被排列为位图,其中位图中的每个状态标识符都对应于不同的存储位置(例如,LBA等)。如本文所使用的,状态标识符可以指示对应的存储位置是否包括有效数据(例如,映射的数据)或者对应的存储位置是否包括“脏”数据(例如,改变的数据)。然而,本公开不限于以上描述。例如,在一些实施例中,1和0的角色可以颠倒、或者除了1和0之外的不同类型的状态标识符(例如,描述性文本)可以用于指示给定存储位置是否包括要从源存储装置复制到目标存储装置的数据。此外,除了位图之外,状态标识符可以以其他数据结构来排列。

[0059] 如以下进一步详细讨论的,参考图1至图13,用于跟踪数据从源存储装置(包括一个或多个存储设备)向目标存储装置的实时迁移的数据复制进度的机制可以允许在实时迁移过程期间使用较少的计算资源(例如,存储器空间、数据总线带宽、功率和处理时间)来跟踪和传送与位图(或其他数据结构)相关的数据。该机制可以涉及在实时迁移服务器将数据(例如,与实时迁移的命名空间相关联的数据)从源存储装置复制到目标存储装置时,由存储控制器(例如,实时迁移存储控制器或实时迁移控制器)跟踪来自源存储控制器(例如,普

通源存储控制器)的用户应用数据写入,并且跟踪实时迁移服务器(例如,与管理实时迁移过程相关联的主机)的实时迁移复制读取(“LMCPR”)操作(例如,过程)状态(例如,当前复制位置)。实时迁移存储控制器可以(1)使用当前复制位置来跟踪相关源存储数据的计划且已复制的区段和计划但尚未被复制的区段,并且(2)创建用于向实时迁移服务器提供关于相关源存储数据在整个实时迁移过程的不同时间的状态的信息的位图。

[0060] 实时迁移控制器可以创建第一位图,以向实时迁移服务器提供存储位置的元数据(例如,状态标识符)的集合,指示要复制到目标存储装置的相关源存储数据的初始状态。实时迁移存储控制器可以将第一位图传送到实时迁移服务器,以基于该第一位图来执行相关数据的实时迁移。实时迁移存储控制器可以创建第二位图(例如,指示对应的存储位置是否为脏的状态标识符的位图),其最初具有全“0”条目(例如,将所有位清除为“0”),以跟踪从源存储控制器到源存储装置的用户应用(例如,普通)数据写入的位置。实时迁移控制器可以监视(例如,“窥探”)在第一位图被传送到实时迁移服务器时或之后发生的来自源存储控制器的用户应用数据写入(例如,在实时迁移服务器对与第一位图相关联的元数据的复制期间,以及在实时迁移服务器基于第一位图将数据从源存储装置复制到目标存储装置期间)。基于所监视的用户数据写入,实时迁移控制器可以标记脏(例如,设置1)以窥探第二位图中的地址关联位图条目(例如,实时迁移控制器可以使得在第二位图中生成状态标识符位)。在实时迁移服务器根据第一位图完成数据复制之后,第二位图可以被传送到实时迁移服务器,以基于对源存储装置的用户应用数据写入来更新目标存储装置数据。

[0061] 实时迁移存储控制器可以基于实时迁移服务器的当前复制位置来在第二位图中选择性地生成或选择性地清除状态标识符位(例如,用于指示用户应用数据写入的位置的“脏”位),以减少冗余存储数据复制的量。例如,实时迁移控制器可以遵循两种方法(例如,减少脏页位图中的脏位条目的两种单独的方法):(1)禁止/阻止在第二位图中跟踪(例如,过滤跟踪或选择性地跟踪)对源存储装置的计划但尚未被复制的区段的用户应用数据写入的位置或者(2)在第二位图中基于用户数据写入UDW的发生来跟踪和设置状态标识符位(例如,脏位),但是稍后,如果状态标识符位(例如,脏位)对应于实时迁移服务器的当前复制位置,则清除它们。当跟踪但稍后清除脏位时,实时迁移控制器不需要参考计划且已复制的区段或计划但尚未被复制的区段。因此,作为任一方法(选择性地生成或选择性地清除)的结果,第二位图都可以:(1)在要传送到实时迁移服务器的数据结构中包括减少量的脏条目,以及(2)阻止实时迁移服务器从源存储装置中不必要的重复数据复制。这可以减少存储器使用、减少带宽使用、降低功耗、减少处理时间,并且减少用户数据写入UDW和用户数据读取UDR的中断(例如,与它们的通常性能相比)。

[0062] 在跟踪分布式源存储数据的当前复制位置的情况(例如,与要从源存储装置复制的源数据相对应的两个或更多个存储位置彼此不直接相邻的情况)下,实时迁移控制器可以通过使用例如两个分离的位图((1)计划且已复制的位图和(2)计划但尚未被复制的位图)来跟踪实时迁移服务器的当前复制位置。(如上所述,尽管描述涉及位图,但是本公开不限于位图。应当理解,可以使用其他数据结构来跟踪实时迁移的数据复制进度。)可替代地,实时迁移控制器可以仅使用计划但尚未被复制的位图(而不是也在计划且已复制的位图中进行跟踪),并且基于用于跟踪计划但尚未被复制的数据位置以及未映射的(未计划要复制的)数据位置的位状态的反转标识相关的复制的数据(例如,如果“1”位用于跟踪与计划但

尚未被复制的区段相对应的存储位置,则当与该存储位置相对应的源数据已经被复制时,该位可以被实时迁移存储控制器翻转为“0”),来跟踪实时迁移服务器的当前复制状态。

[0063] 图1是根据本公开的一些实施例的,描绘用于跟踪实时迁移的数据复制进度的系统的系统图。

[0064] 参考图1,系统1可以包括源服务器100a(例如,普通源服务器)和实时迁移服务器100b。源服务器100a和实时迁移服务器100b可以是主机系统1000(例如,源主机系统)的一部分。源服务器100a和实时迁移服务器100b可以连接到源存储设备250a和实时迁移存储设备250b。源存储设备250a和实时迁移存储设备250b可以连接到源存储装置200。实时迁移服务器100b还可以连接到目标存储设备250c。目标存储设备250c可以连接到目标存储装置300。源存储装置200和目标存储装置300可以包括存储位置(例如,对于简化的示例,在源存储装置200内描述的包括LBA0至LBA11的LBA),其可以与源存储装置200和目标存储装置300中用于存储数据(例如,用户的数据)的物理位置相关联。

[0065] 实时迁移服务器100b可以经由通信链路20连接到目标存储设备250c(例如,实时迁移目标存储设备)。通信链路20可以包括各种技术(例如,外围组件快速互连(PCIe)、PCIe上非易失性存储器快速(NVMe)、光纤上NVMe(NVMe-oF)、串行ATA(STATA)、小型计算机存储接口(SCSI)等)(例如,可以通过各种技术来实现)。

[0066] 目标存储设备250c可以包括目标存储控制器210c。目标存储装置300可以包括目标存储器330,其可以包括(例如,可以是)非易失性存储器和/或易失性存储器。例如,目标存储器330可以包括非易失性存储器和/或存储器层(tier)(包括易失性和非易失性存储器),并且可以对应于一个或多个目标存储设备250c的非易失性存储器。目标存储设备250c可以经由存储接口22d(例如,服务于实时迁移复制写入(LMCPW)操作(例如,过程)的存储接口)连接到目标存储装置300。存储接口22d可以包括(例如,可以是)闪存接口(例如,NAND闪存接口)或(例如,通过以太网实现的)远程接口。

[0067] 源服务器100a可以服务于来自用户应用的读取和写入请求。例如,用户可以使用应用(例如,文字处理应用)并将用户数据写入UDW(例如,用户应用数据写入)或用户数据读取UDR(例如,用户应用数据读取)(例如,经由源存储设备250a)发送到源存储装置200。

[0068] 源服务器100a可以经由通信链路20连接到源存储设备250a(例如,普通源存储设备)。通信链路20可以包括多种技术(例如,外围组件快速互连(PCIe)、PCIe上非易失性存储器快速(NVMe)、光纤上NVMe(NVMe-oF)、SATA、SCSI等)(例如,可以通过各种技术来实现)。

[0069] 源存储设备250a可以包括源存储控制器210a和源控制器存储器220a。源控制器存储器220a可以包括(例如,可以是)随机访问存储器(RAM)(例如,静态随机访问存储器(SRAM)或动态随机访问存储器(DRAM))。源存储控制器210a可以包括用于处理来自服务器(例如,源服务器100a)的将数据写入源存储装置200的请求的嵌入式逻辑(例如,嵌入在源存储设备250a内)(例如,可以通过嵌入式逻辑来实现)。源存储设备250a可以连接到源存储装置200(或源存储装置200的一部分)。源存储装置200可以包括源存储器230。源存储器230可以包括(例如,可以是)非易失性存储器和/或易失性存储器。例如,源存储器230可以包括非易失性存储器和/或存储器层(包括易失性和非易失性存储器),并且可以连接到一个或多个源存储设备250a。源存储设备250a可以经由存储接口22a(例如,服务于UDW操作的存储接口)和存储接口22b(例如,服务于用户数据读取UDR操作的存储接口)连接到源存储装置

200。存储接口22a和22b可以包括(例如,可以是)闪存接口(例如,NAND闪存接口)或(例如,通过以太网实现的)远程接口。

[0070] 实时迁移服务器100b可以是被分配有与源服务器100a相比相同、较低或较高的访问系统1内数据的权限(privilege)的服务器。可以分配实时迁移服务器100b来管理用于将数据从源存储装置200复制到目标存储装置300的实时迁移操作(例如,实时迁移过程)。实时迁移服务器100b可以包括实时迁移服务器存储器120b。实时迁移服务器存储器120b可以包括(例如,可以是)随机访问存储器(RAM)(例如,静态随机访问存储器(SRAM)或动态随机访问存储器(DRAM))。实时迁移服务器存储器120b可以用于存储用于管理实时迁移操作的元数据。实时迁移服务器100b可以经由通信链路20连接到实时迁移存储设备250b。通信链路20可以包括各种技术(例如,外围组件快速互连(PCIe)、PCIe上非易失性存储器快速(NVMe)、光纤上NVMe(NVMe-oF)、SATA或SCSI等)(例如,可以通过各种技术来实现)。

[0071] 实时迁移存储设备250b可以是被分配有与源存储设备250a或目标存储设备250c相比相同、较低或较高的访问系统1内数据的权限的存储设备。可以分配实时迁移存储设备来辅助实时迁移操作将数据从源存储装置200复制到目标存储装置300。

[0072] 实时迁移存储设备250b可以包括实时迁移控制器210b和实时迁移控制器存储器220b。实时迁移控制器存储器220b可以包括(例如,可以是)随机访问存储器(RAM)(例如,静态随机访问存储器(SRAM)或动态随机访问存储器(DRAM))。实时迁移控制器存储器220b可以用于存储用于管理(例如,监视、跟踪和/或指导)实时迁移操作的元数据。实时迁移控制器210b可以包括用于辅助实时迁移服务器100b将数据从源存储装置200复制到目标存储装置300的嵌入式逻辑(例如,嵌入在实时迁移存储设备250b内)(例如,可以通过嵌入式逻辑来实现)。实时迁移存储设备250b可以经由存储接口22c(例如,服务于LMCPR操作的存储接口)连接到源存储装置200。存储接口22c可以包括(例如,可以是)闪存接口(例如,NAND闪存接口)或(例如,通过以太网实现的)远程接口。

[0073] 实时迁移服务器100b可以经由实时迁移服务器到源链路10连接(例如链接)到源服务器100a。取决于系统1是经由硬件(“HW”)、软件(“SW”)还是硬件和软件的组合来实现的,实时迁移服务器到源链路10可以通过各种技术来实现。例如,应当理解,系统1的服务器和存储设备可以经由HW、SW或HW和SW的组合来实现。例如,在一些实施例中,实时迁移服务器到源链路10可以是物理网络连接(例如,以太网)。在一些实施例中,服务器(例如,源服务器100a或实时迁移服务器100b)中的一个或多个可以是软件实体。例如,服务器中的一个或多个可以由与一个或多个中央处理单元(CPU)相关联的管理程序(hypervisor)管理的虚拟机VM。同样,存储设备中的一个或多个可以通过HW和/或SW技术来实现。例如,源存储设备250a、实时迁移存储设备250b和目标存储设备250c中的一个或多个可以是在一个或多个物理设备上实现的不同虚拟机。例如,源存储装置200和/或目标存储装置300可以是同一物理介质的逻辑分区(例如,源存储装置200可以是第一命名空间,而目标存储装置300可以是第二命名空间)。

[0074] 实时迁移服务器100b和实时迁移控制器210b可以通过传送元数据(例如,状态标识符位)来协调实时迁移过程的管理或跟踪,该元数据(例如,状态标识符位)指示与源存储装置200中的存储位置(例如,LBA)相对应的数据的状态。例如,实时迁移服务器100b可以参考实时迁移服务器存储器120b中的状态标识符位(例如,1)来向实时迁移控制器210b发送

一个或多个复制读取请求。实时迁移控制器210b可以进行到源存储装置200内的要复制到目标存储装置300的相关数据(例如,“有效”数据)的映射。实时迁移控制器210b可以创建与源存储装置相对应的元数据的第一数据结构M1(例如,第一位图)。例如,在实时迁移过程的映射阶段,第一数据结构M1可以被称为映射的页的位图(例如,“BMP”或位图映射的页),每个映射的页(例如LBA)是源存储装置200中的逻辑存储单元(例如页)。可以在实时迁移控制器存储器220b中创建(例如,临时存储)第一数据结构M1。

[0075] 第一数据结构M1可以包括位(例如,第一状态标识符位),第一状态标识符位的每一位指示给定存储位置(例如,LBA)是否包括要复制到目标存储装置300的有效数据。可以包括例如4千字节(KB)数据的每个LBA可以由第一数据结构M1的位图中的一个状态标识符位来表示。例如,“0”位可以指示给定LBA没有要复制到目标存储装置300的有效数据,而“1”位可以指示给定LBA有要复制到目标存储装置300的有效数据。因此,具有大存储容量(例如源存储装置200中的8个太字节(TB)数据)的整个驱动可以通过具有例如2048兆位元数据的位图来表示。例如,实时迁移控制器210b可以基于来自实时迁移服务器100b的请求来映射源存储装置200,并且可以在第一数据结构M1中设置指示要被复制的源数据的位置的第一状态标识符位MD1(例如,1),以标识要从源存储装置200复制到目标存储装置300的数据位置。(指示要被复制的源数据的位置的第一状态标识符位MD1可以是与第一数据结构M1相对应的、指示对应的存储位置包括要从源存储装置复制到目标存储装置的数据的位)

[0076] 当实时迁移控制器210b已经完成了源存储装置200的映射时,实时迁移控制器210b可以将第一数据结构M1(包括指示要被复制的源数据的位置的第一状态标识符位MD1)(例如,经由位图传送请求-响应)传送到实时迁移服务器100b,示为图1的实时迁移服务器存储器120b中的第一数据结构的主机版本M1'(或第一数据结构M1')。实时迁移服务器100b可以使用第一数据结构的主机版本M1'来(例如,经由实时迁移复制(LMCP)过程,包括实时迁移复制读取操作(例如,过程)LM-CP-Read或LMCPRead,以及实时迁移复制写入操作LM-CP-Write或LMCPWrite)将相关数据从源存储装置200复制到目标存储装置300。例如,实时迁移服务器100b可以使用数据结构的主机版本M1'来执行LMCP过程,其中实时迁移服务器100b(例如,在LMCPRead期间)从源存储装置200读取数据,并且实时迁移服务器100b(例如,在LMCPWrite期间)将数据写入目标存储装置300。总之,实时迁移控制器210b可以将第一数据结构M1从实时迁移控制器存储器220b传送(例如,位图传送)到实时迁移服务器存储器120b。第一数据结构M1在实时迁移主机存储器120b中的版本可以被称为第一数据结构的主机版本M1'(例如,位图请求和传送)或者简称为第一数据结构M1'(也被称为“主机快照”)。主机版本HV可以指由实时迁移服务器100b用来将相关数据从源存储装置200复制到目标存储装置300的数据结构的任何主机版本(例如,初始BMP、初始BDP或后续BDP)。

[0077] 在将数据从源存储装置200复制和写入目标存储装置300的过程期间,可以发生对源存储装置200的用户数据写入UDW。为了跟踪用户数据写入UDW的位置,实时迁移控制器210b可以创建元数据的第二数据结构M2(例如,第二位图),其可以被称为“脏”页的位图(“BDP”)。例如,可以创建第二数据结构M2来指示哪些存储位置(例如,LBA)对应于用户数据写入UDW,并且因此在下一次迭代中将数据复制到目标存储装置300。第二数据结构M2可以由实时迁移控制器210b(例如,响应于基于第一数据结构M1的复制的完成)传送(例如,位图传送)到实时迁移服务器100b。实时迁移服务器100b可以使用第二数据结构的主机版本M2'

(或第二数据结构M2') (例如,与由实时迁移控制器210b创建的第二数据结构M2相对应的主机版本HV) 作为在实时迁移服务器100b的稍后迭代中要从源存储装置200复制到目标存储装置300的数据的参考。例如,第二数据结构的主机版本M2' 可以是实时迁移服务器100b可以用来基于源存储装置200处的相关数据的改变来更新目标存储装置300处的相关数据的BDP。

[0078] 总之,可以执行三种不同的操作。第一,实时迁移控制器210b可以将初始数据结构(例如,第一数据结构M1或第二数据结构M2) 传送到实时迁移服务器100b,使得在实时迁移服务器100b中创建第一数据结构的主机版本M1' 或第二数据结构的主机版本M2'。第二,实时迁移服务器100b可以参考第一数据结构的主机版本M1' 或第二数据结构的主机版本M2' 来执行复制过程,同时实时迁移控制器210b在后续数据结构(例如,如果初始数据结构是第一数据结构M1,则是第二数据结构M2、或者如果第二数据结构M2是初始数据结构,则是第三数据结构M3) 中跟踪作为脏位的用户数据写入UDW。第三,当复制过程完成时,实时迁移服务器100b可以向实时迁移控制器210b请求下一个数据结构(例如,下一个位图数据结构) 的传送。当实时迁移服务器100b接收到下一个数据结构的传送时,可以执行这三种不同的操作的另一次迭代。(下面将进一步讨论图1。图2和图3参考第一数据结构M1、第二数据结构M2和第三数据结构M3讨论了关于数据结构的跟踪和传送的细节。)

[0079] 图2是根据本公开的一些实施例的,描绘被动第一数据结构的传送和利用主动第二数据结构跟踪用户数据写入的简化概念图。

[0080] 图3是根据本公开的一些实施例的,描绘第二数据结构的被动版本的传送和利用主动第三数据结构跟踪用户数据写入的简化概念图。

[0081] 参考图2,如上所述,实时迁移控制器210b可以在实时迁移过程快要(例如,在) 开始时创建(例如,提供) 可以是BMP的第一数据结构M1。响应于来自实时迁移服务器100b的请求,实时迁移控制器210b还可以(例如,在从实时迁移控制器210b到实时迁移服务器100b的实时迁移BMP传送期间以及在实时迁移复制过程期间) 创建可以是BDP的第二数据结构M2,用于对目标存储装置300中的数据进行渐进式(incremental) (例如,迭代的) 更新。例如,作为BDP的第二数据结构M2可以跟踪(例如,标识) 来自源存储设备250a的用户数据写入UDW的位置,该用户数据写入UDW可能在作为BMP的第一数据结构M1的传送期间或之后,以及在后续的实时迁移复制操作LMCP(包括实时迁移复制读取操作LMCPR和/或实时迁移复制写操作LMCPW) 期间在源存储装置200中的相关数据位置发生。

[0082] 例如,在实时迁移过程的映射阶段,第一数据结构M1可以是BMP。实时迁移服务器100b可以请求BMP的复制(例如,实时迁移服务器100b可以向实时迁移控制器210b发送“查询BMP”命令(例如,主机位图获取212))。实时迁移控制器210b可以为该BMP准备作为第一数据结构M1的缓冲区(buffer) (例如,图1所示的实时迁移控制器存储器220b中的位置)。实时迁移控制器210b可以在第一数据结构M1中将指示要被复制的源数据的位置的第一状态标识符位MD1设置为映射的位(例如,一个或多个1),该第一数据结构M1是具有有效数据的映射操作的每一页(或LBA) 的BMP缓冲区。当映射操作完成时,作为BMP的第一数据结构M1可以成为被动位图(例如,BMP不再可用于更新位图的位),并且实时迁移控制器210b可以将第一数据结构M1内容、指示要被复制的源数据的位置的第一状态标识符位MD1传送(例如,位图传送) (例如,将BMP发送) 到实时迁移服务器100b中的实时迁移服务器存储器120b。实时迁

移服务器100b可以接收作为BMP的第一数据结构的主机版本M1'并将其存储在实时迁移服务器存储器120b中,参考第一数据结构的主机版本M1'从源存储装置200读取映射的页,并将相关数据复制到目标存储装置300(参见图1)。

[0083] 如上所述,实时迁移控制器210b可以创建可以是第一BDP的第二数据结构M2,以跟踪在第一数据结构M1(例如,BMP)的传送(例如,位图传送)和后续的实时迁移复制操作LMCP期间或之后发生的对源存储装置200的相关位置的更新。例如,实时迁移控制器210b可以准备最初将所有位都清除为“0”的第二数据结构M2(例如,它是主动BDP缓冲区或主动位图缓冲区)。位图采集器(logger)211可以监视(例如,“窥探”)来自源存储设备250a的用户数据写入UDW,并且可以在当前的实时迁移复制LMCP操作期间,在主动第二数据结构M2中设置指示要被复制的源数据的位置的第二状态标识符位MD2(例如,设置为“1”位),以标识用户数据写入UDW的相关位置。(指示要被复制的源数据的位置的第二状态标识符位MD2可以是与第二数据结构M2相对应的、指示对应的存储位置包括要从源存储装置复制到目标存储装置的数据的位)

[0084] 在一些实施例中,当实时迁移服务器100b已经完成了与被动位图(例如,被动第一数据结构M1或后续的被动第二数据结构M2的缓冲区)相对应的实时迁移复制操作时,实时迁移服务器100b可以发送命令(例如,“交换BDP”命令),以交换主动位图缓冲区和被动位图缓冲区。

[0085] 例如,参考图3,在一些实施例中,与被动第一数据结构M1相对应的缓冲区可以成为主动第三数据结构M3缓冲区(例如,第二位图或后续的BDP位图的缓冲区),并且第一主动第二数据结构M2缓冲区可以成为被动第二数据结构M2缓冲区。实时迁移服务器100b可以向实时迁移控制器210b发送命令(例如,“查询BDP”命令),来将被动缓冲区以第二数据结构的主机版本M2'传送(例如,位图传送)(例如,复制与被动缓冲区相对应的被动第二数据结构M2)到实时迁移服务器100b。在一些实施例中,当被动位图已经被传送到实时迁移服务器100b(例如,作为位图传送、或主机版本HV)时,实时迁移控制器210b可以重置被动位图缓冲区(例如,将所有的位都清除为“0”),并且准备将其作为下一个BDP(例如,具有指示要被复制的源数据的位置的第三状态标识符位MD3的第三数据结构M3)的主动位图缓冲区进行跟踪。(指示要被复制的源数据的位置的第三状态标识符位MD3可以是与第三数据结构M3相对应的、指示对应的存储位置包括要从源存储装置复制到目标存储装置的数据的位。)实时迁移服务器100b可以解析第二数据结构的主机版本M2'(例如,BDP),(参考第二数据结构的主机版本M2')从源存储装置200读取脏页,并将相关数据复制到目标存储装置300。

[0086] 再次参考图1,在一些实施例中,如以下参考图5-图13所述的,实时迁移控制器210b可以监视(例如,“窥探”)从源存储设备250a到源存储装置200的用户数据写入UDW。实时迁移控制器210b可以监视与实时迁移控制器100b的LMCPR和LMCPW操作相关联的当前复制位置CCL。

[0087] 在一些实施例中,实时迁移控制器210b可以使用当前复制位置CCL、开始(start/starting)地址A1(例如,计划要复制的区段的开始地址)和结束(end/ending)地址A2(例如,计划要复制的区段的结束地址)来跟踪源存储装置200的计划且已复制的区段CS和计划但尚未被复制的区段YS。例如,如图1所示,LBA0至LBA8可以对应于计划要复制的区段(例如,源存储装置200中实时迁移服务器100b已经被指派要复制到目标存储装置300的区段)。

LBA0至LBA2可以对应于在第一时间 t_1 计划且已复制的区段CS, LBA3至LBA8可以对应于在第一时间 t_1 计划但尚未被复制的区段YS。LBA9至LBA11可以对应于未映射的区段US (例如, 源存储装置200中未计划要被复制的区段)。例如, 当前复制位置CCL可以与对应于连续索引的开始地址A1和连续索引的结束地址A2的当前复制索引相关联, 该连续索引与计划要复制的区段相关联, 该计划要复制的区段包括开始地址A1和结束地址A2之间连续的存储位置 (例如, LBA)。

[0088] 在一些实施例中, 当计划要被复制的存储位置 (例如, LBA) 并非全部彼此相邻 (例如, 并非全部是连续的、或者是分布式的) 时, 实时迁移控制器210b可以使用当前复制位置CCL以及一个或多个位图符号标记 (notation) /表示 (例如, 代替开始地址A1和结束地址A2) 来跟踪源存储装置200的计划且已复制的区段CS和计划但尚未被复制的区段YS。位图可以是包括具有位置信息的1和/或0的集合的数据结构 (参见例如下面的图4和相关描述)。

[0089] 当实时迁移复制正在进行 (例如, 在LMCPR和LMCPW过程期间) 并且在所有数据都被复制之前, 实时迁移控制器210b可以监视在第一数据结构M1到实时迁移服务器100b的位图传送期间和之后发生的用户数据写入UDW。实时迁移控制器210b可以创建第二数据结构M2, 并且在第二数据结构M2中利用指示要被复制的源数据的位置的第二状态标识符位MD2 (例如, “脏”位) 来跟踪 (例如, 记录) 用户数据写入UDW的位置。在第一LMCP过程 (例如, 包括LMCPR和LMCPW操作) 已经完成之后, 第二数据结构M2可以被传送到实时迁移服务器100b, 以向实时迁移服务器100b指示相关源数据已经改变的位置 (例如, 用户数据写入UDW已经被写入的位置), 使得实时迁移服务器100b可以执行另一次LMCP迭代 (例如, 稍后的迭代), 其包括在源存储装置200处 (从由一个或多个指示要被复制的源数据的位置的第二状态标识符位MD2所标识的位置) 复制源数据的另一个LMCPR操作和在目标存储装置300处更新数据的另一个LMCPW操作。

[0090] 例如, 第一用户数据写入UDW1可以在第一时间 t_1 被写入LBA0至LBA9 (例如, 当当前复制地址CCL对应于LBA2时)。在一些实施例中, 实时迁移控制器210b可以在第二数据结构M2中生成指示要被复制的源数据的位置的第二状态标识符位MD2 (该第二状态标识符位MD2与将被包括在第二数据结构M2中的每个LBA地址LBA0至LBA9相对应) 以供实时迁移服务器100b在下一个LMCP过程期间从源存储装置200复制 (与第一数据结构的主机版本M1' 相关联的) 有效数据时跟踪对源存储设备250a的用户数据写入UDW。然而, 在 t_1 , 基于第一数据结构M1', 地址LBA3至LBA8被调度为被复制和写入到目标存储器330, 但是尚未被复制和写入。因此, 如果地址LBA3至LBA8被包括在第二数据结构M2中并且被发送到实时迁移服务器100b, 则与LBA3至LBA8相对应的源数据将从源存储装置200被复制两次 (例如, 从源存储装置200读取两次并向目标存储装置300写入两次) ——一次是基于第一数据结构M1' 的, 而一次是基于第二数据结构M2' 的。因此, 如果从源存储装置200到目标存储装置300再次复制LBA3至LBA8, 则下一次LMCPR和LMCPW迭代可能包括冗余的读取和写入。消耗时间和功率来冗余地复制与LBA3至LBA8相对应的源数据可能是浪费的, 因为源数据先前在前一次迭代中被复制到了目标存储装置300。此外, 由于创建更多的BDP运行 (例如, LMCPR和LMCPW操作的更多次迭代以在目标存储装置300处更新数据), 与冗余复制相关联的缺点可能具有累积影响。

[0091] 在一些实施例中, 如下文参考图5-图13所述的, 为了减少 (例如, 消除) 冗余复制, 实时迁移控制器210b可以在第二数据结构M2 (和/或LMCP过程的后续迭代) 中选择性地生成

或选择性地清除指示要被复制的源数据的位置的第二状态标识符位MD2(例如,脏位)。

[0092] 例如,在涉及选择性生成方法的一些实施例中,实时迁移控制器210b可以使用对计划且已复制的区段CS和/或计划但尚未被复制的区段YS的跟踪,来过滤在第二数据结构M2(和/或LMCP过程的后续迭代)中指示要被复制的源数据的位置的第二状态标识符位MD2(例如,脏位)的生成,使得对计划但尚未被复制的区段的跟踪(例如,对于与计划但尚未被复制的位图YBM重叠的那些当前复制位置CCL)禁止(或阻止)实时迁移控制器210在第二数据结构M2中生成指示要被复制的源数据的位置的第二状态标识符位MD2。例如,实时迁移控制器210b可以在第二数据结构M2中选择性地生成指示要被复制的源数据的位置的第二状态标识符位MD2,使得当参考由实时迁移控制器210b传送的第二数据结构M2位图时,实时迁移服务器100b可以减少(例如,避免)下一次LMCP迭代中的冗余复制。

[0093] 在涉及选择性清除方法的一些实施例中,代替使用对计划且已复制的区段CS和/或计划但尚未被复制的区段YS的跟踪来过滤在第二数据结构M2中对指示要被复制的源数据的位置的第二状态标识符位MD2的生成,实时迁移控制器210可以基于用户数据写入UDW的发生,在第二数据结构M2中跟踪和设置指示要被复制的源数据的位置的第二状态标识符位MD2,但是,稍后,如果指示要被复制的源数据的位置的第二状态标识符位MD2对应于实时迁移服务器的当前复制位置CCL,则从第二数据结构M2中清除指示要被复制的源数据的位置的第二状态标识符位MD2。

[0094] 如本文所使用的,“跟踪”区段(例如,源存储装置200的计划但尚未被复制的区段YS和/或计划且已复制的区段CS)可以指保持对与该区段相关联的存储位置的集合(例如,LBA)内的存储位置(例如,LBA)的知晓或者维护与该区段相关联的存储位置的集合(例如,LBA)内的存储位置(例如,LBA)的记录,和/或保持与该区段相关联的存储位置的集合(例如,LBA)外(不在其内)的存储位置(例如,LBA)的知晓或者维护与该区段相关联的存储位置的集合(例如,LBA)外(不在其内)的存储位置(例如,LBA)的记录。

[0095] 如本文所使用的,“过滤”状态标识符(例如,指示要被复制的源数据的位置的第二状态标识符位MD2)的生成可以指在一些情况下生成状态标识符,而在一些其他情况下不生成状态标识符。

[0096] 例如,在一些实施例中,在第一数据结构M1已经被传送到了实时迁移服务器后,第一数据结构M1的实时迁移(LM)复制状态数据结构(例如,第一数据结构M1、或者基于第一数据结构M1且与第二数据结构M2分离且不同的另一个数据结构或其他数据结构)可以在在第二数据结构M2中设置状态标识符位(例如,指示要被复制的源数据的位置的第二状态标识符位MD2)以标识对源存储装置200的用户数据写入的位置之前,有效地用作过滤器。

[0097] 例如,LM复制状态数据结构可以是实时迁移控制器210b处的数据结构。LM复制状态数据结构可以基于M1,但是LM复制状态数据结构可以是与M1不同的数据结构。在一些实施例中,LM复制状态数据结构可以包括(例如,可以是)三个不同的位图和一个索引(例如,指示当前复制位置CCL在哪里的一个索引)。在涉及连续区段的一些实施例中,LM复制状态数据结构可以包括三个索引(例如,开始索引、当前索引和结束索引)。在涉及非连续(或分布式)区段的一些实施例中,LM复制状态数据结构可以包括三个位图(例如,计划且已复制的位图CBM、计划但尚未被复制的位图YBM和未映射的区段位图UBM)。在一些实施例中,LM复制状态数据结构可以包括(例如,可以是)一个位图和若干个索引。例如,一个索引可以指示

映射的区段开始的位置(所以该索引之前的位置可以是未映射的位置),一个索引可以指示当前复制位置CCL在哪里,以及一个索引可以指示最终映射的条目在哪里(所以该索引之后的位置可以再次是未映射的位置)。应当理解,其他数据结构和索引实施例也是可能的。在一些实施例中,其中LM复制状态数据结构是不同于第一数据结构M1的数据结构,可以基于第一数据结构M1来初始化计划但尚未被复制的区段YS。

[0098] 例如,在一些实施例中,实时迁移控制器210b可以初始化第一数据结构M1的LM复制状态数据结构,以跟踪三个区段以及当前复制位置CCL:(1)计划且已复制的区段CS(或计划且已复制的位图CBM);(2)计划但尚未被复制的区段YS(或计划但尚未被复制的位图YBM);以及(3)未映射的区段US(或未映射的区段位图UBM)。例如,实时迁移控制器210b可以将第一数据结构M1(例如,第一数据结构的LM复制状态数据结构)中与计划且已复制的区段CS相对应的所有状态标识符位都初始化为0;实时迁移控制器210b可以将第一数据结构M1中与计划但尚未被复制的区段YS相对应的所有状态标识符位都初始化为全1/有效,以匹配M1(例如,将在计划但尚未被复制的区段YS中的状态标识符位初始化为1);并且实时迁移控制器210b可以将第一数据结构M1中与未映射的区段US相对应的所有状态标识符位都初始化为1(例如,未映射的区段US中的所有状态标识符位都可以是1,并且可以在整个LMCP迭代中保持静态(例如,保持1)。

[0099] 在一些实施例中,用于跟踪当前复制位置CCL的第一数据结构M1(或基于第一数据结构M1的另一个数据结构)可以包括一个以上的用于跟踪分离的区段CS/CBM、YS/YBM和/或US/UBM的分离的数据结构(例如,参见图4、图6、图7和图8)。

[0100] 在一些实施例中,用于跟踪当前复制位置CCL的第一数据结构M1(或基于第一数据结构M1的另一个数据结构)可以包括一个用于跟踪分离的区段CS/CBM、YS/YBM和/或UM的数据结构(例如,参见图9)。

[0101] 再次参考图1,在一些实施例中,实时迁移控制器210b可以(例如,并行)监视用户数据写入UDW和LMCPR读取LBA。随着当前复制位置CCL改变(例如,前进),实时迁移控制器210b可以清除计划但尚未被复制的区段YS中的状态标识符位(例如,指示要被复制的源数据的位置的计划但尚未被复制的区段状态标识符位YS.MD1)(将这些位设置为0),并且在计划且已复制的区段CS中将对应的状态标识符位(例如,指示已复制的源数据的位置的计划且已复制的区段状态标识符位CS.MD1)设置为1(例如,参见下面的图4B)。(指示要被复制的源数据的位置的计划但尚未被复制的区段状态标识符位YS.MD1可以是指示对应的存储位置包括尚未从源存储装置复制到目标存储装置的数据的位,其与LM复制状态数据结构的计划但尚未被复制的区段YS相对应。指示已复制的源数据的位置的计划且已复制的区段状态标识符位CS.MD1可以是与LM复制状态数据结构的计划且已复制的区段CS相对应的、指示对应的存储位置包括已经从源存储装置复制到目标存储装置的数据的位。)如果监视的用户数据写入UDW对应于与计划但尚未被复制的区段YS中具有被设置为1的状态标识符位的第一数据结构M1的条目相对应的LBA,则可以(例如,通过不将第二状态标识符位设置为有效,即设置为1)不在第二数据结构M2中跟踪与该LBA地址相对应的用户数据写入UDW。另一方面,如果监视的用户数据写入UDW对应于与计划但尚未被复制的区段YS中具有被设置为0的状态标识符位的第一数据结构M1的条目相对应的LBA,则可以(例如,通过将第二状态标识符位设置为有效,即设置为1)在第二数据结构M2中跟踪与该LBA地址相对应的用户数据写

入UDW。因为来自用户数据写入UDW的更新的数据可以由实时迁移服务器基于第一数据结构M1中的映射来复制,所以可以通过过滤对第二数据结构M2的更新(例如,阻止脏位的生成)来阻止实时迁移服务器100b的冗余复制操作。例如,通过不将指示要被复制的源数据的位置的第二状态标识符位MD2设置为有效/1(例如,通过过滤对第二数据结构M2的更新),可以阻止实时迁移服务器100b的冗余复制操作。例如,在未应用选择性生成方法的情况下,指示要被复制的源数据的位置的第二状态标识符位MD2可以被设置为1,这可能导致冗余复制。

[0102] 例如,如图1所示的,第一时间 t_1 的第一用户数据写入UDW1可以写入与第一数据结构M1(或基于第一数据结构M1的另一个数据结构)中的三个区段之一相关的LBA位置(计划且已复制的区段CS、计划但尚未被复制的区段YS和未映射的区段US)。实时迁移控制器210b可以参考第一数据结构M1(或基于第一数据结构M1的另一个数据结构),并且仅允许与计划且已复制的区段CS或未映射的区段US相对应的用户数据写入UDW被设置为第二数据结构M2中指示要被复制的源数据的位置的第二状态标识符位MD2(例如,1),同时不允许与计划但尚未被复制的区段YS的用户数据写入UDW被设置为第二数据结构M2中指示要被复制的源数据的位置的第二状态标识符位MD2(也参见图12的流程图。)

[0103] 在一些实施例中,如下文参考图10和图13所述的,为了减少(例如,消除)冗余复制,实时迁移控制器210b可以跟踪当前复制位置CCL,而无需参考计划且已复制的区段CS或计划但尚未被复制的区段YS。在这样的实施例中,实时迁移控制器210b可以监视和跟踪在第一数据结构M1到实时迁移服务器100b的位图传送期间或之后(例如,在实时迁移服务器复制期间,以及在实时迁移服务器复制完成之前)发生的用户数据写入UDW。同时,实时迁移服务器100b可以根据用户数据写入UDW在第二数据结构M2中生成指示要被复制的源数据的位置的第二状态标识符位MD2,而不过滤或阻止与计划但尚未被复制的区段相对应的、指示要被复制的源数据的位置的第二状态标识符位MD2的生成。在这样的实施例中,实时迁移控制器210b可以监视当前复制位置CCL,并且可以从第二数据结构M2中清除与当前复制位置CCL相对应的指示要被复制的源数据的位置的第二状态标识符位MD2(例如,“脏”位),因为相关联的源数据已经被实时迁移服务器100b复制(如当前复制位置CCL所指示的)。因此,实时迁移控制器210b可以在第二数据结构M2中选择性地清除指示要被复制的源数据的位置的第二状态标识符位MD2,使得当参考由实时迁移控制器210b传送的第二数据结构M2时,实时迁移服务器100b可以减少(例如,避免)在下一LMCPR和LMCPW迭代中的冗余复制。(也参见图13的流程图。)

[0104] 在一些实施例中,实时迁移控制器210b可以生成指示要被复制的源数据的位置的第二状态标识符位MD2,以指示在计划要复制的区段之外(例如,在未映射的区段US中)发生的用户数据写入UDW的位置。

[0105] 图4A、图4B、图4C、图4D、图4E、图4F、图4G和图4H(统称为图4)是根据本公开的一些实施例的,描绘用于描绘跟踪实时迁移的数据复制进度的方法的位图视图的细节的简化概念图。

[0106] 参考图4A,第一数据结构的主机版本M1'可以在实时迁移服务器存储器120b中,用于执行LMCP操作(参见图1)。位图可以表示源存储装置200中与例如LBA0至LBA49相关联的源数据。指示要被复制的源数据的位置的第一状态标识符位MD1(例如,“1”位)可以指示用于从源存储装置200复制到目标存储装置300的源数据(例如有效或脏数据)。当前复制位置

CCL可以在第零时间 t_0 和第三时间 t_3 之间从开始地址A1前进到结束地址A2。结束地址A2上方顶行的位图位置可以对应于未映射的区段US。

[0107] 参考图4B,第一数据结构M1(或一个或多个其他数据结构,如上文参考图1所述的)可以位于实时迁移控制器存储器220b中,以跟踪当前复制位置CCL的进度。例如,如图4B中针对第一时间 t_1 所描绘的,实时迁移控制器210b可以使用第一数据结构M1通过从第一数据结构M1中清除与计划但尚未被复制的区段YS相对应的状态标识符位(例如,指示要被复制的源数据的位置的计划但尚未被复制的区段状态标识符位YS.MD1)(将这些位清除为0)并且向第一数据结构M1添加(例如,分别添加)与计划且已复制的区段CS相对应的状态标识符位(例如,指示已复制的源数据的位置的计划且已复制的区段状态标识符位CS.MD1)来跟踪当前复制位置CCL的进度。例如,随着当前复制位置CCL前进,指示已复制的源数据的位置的计划且已复制的区段状态标识符位CS.MD1(例如,1)的量可以在计划且已复制的区段CS中增加,并且指示要被复制的源数据的位置的计划但尚未被复制的区段状态标识符位YS.MD1(例如,1)的量可以在计划但尚未被复制的区段YS中减少(例如,分别减少)。未映射的区段状态标识符位US.MD1的量可能不在给定迭代中增加或减少。(未映射的区段状态标识符位US.MD1可以是与LM复制状态数据结构的未映射的区段US相对应的、指示对应的存储位置在计划要复制的区段之外的位。)

[0108] 参考图4C,第二数据结构M2可以在实时迁移控制器存储器220b中并被激活,其中每个位初始设置为“0”,用于跟踪在第零时间 t_0 和第三时间 t_3 之间发生的来自源存储设备250a的用户数据写入UDW。在图4C至图4F中,当前复制位置CCL区段(例如,YS、CS和UM)在第二数据结构M2旁边示出,以描绘计划且已复制的区段CS和计划但尚未被复制的区段YS如何分别在大小上改变,以及如何对应于在第二数据结构M2中对指示要被复制的源数据位置的第二个状态标识符位MD2(例如,“1”位)的生成的过滤。

[0109] 在第零时间 t_0 ,开始地址A1和结束地址A2之间的每一位都可以对应于计划但尚未被复制的区段YS。

[0110] 参考图4D,在第一时间 t_1 ,第一用户数据写入UDW1可以在LBA0至LBA9发生。根据一些实施例,可以在计划且已复制的区段CS和未映射的区段US中与LBA相对应的第二数据结构M2位置中生成指示要被复制的源数据的位置的第二状态标识符位MD2(例如,BDP中的脏位),同时可以不在计划但尚未被复制的区段YS中与LBA(例如,LBA3至LBA8)相对应的第二数据结构M2位置中生成指示要被复制的源数据的位置的第二状态标识符位MD2(例如,脏位或1)。

[0111] 参考图4E,基于第二时间 t_2 的第二用户数据写入UDW2(在LBA10至LBA19发生的),可以类似地对脏位的生成进行过滤。例如,可以不在计划但尚未被复制的区段YS中与LBA(例如,LBA16至LBA18)相对应的第二数据结构M2位置中生成指示要被复制的源数据的位置的第二状态标识符位MD2(例如,脏位或1)。参考图4F,在第三时间 t_3 发生的第三数据写入UDW3(在LBA20至LBA29发生的)可以贯穿与LBA20至LBA29相对应的位图位置生成一个或多个指示要被复制的源数据的位置的第二状态标识符位MD2,因为在第三时间 t_3 每个位置都对应于的计划且已复制的区段CS。

[0112] 参考图4G和图4H,可以以位图视图(例如,用于描绘数据结构中状态标识符的选择性生成或选择性清除的图,该数据结构用于指示对源存储装置的用户数据写入的位置)来

表示用户数据写入UDW时间和位置的序列以及最终的过滤后的状态标识符(例如,脏位)。例如,图4G描绘了在图4C至图4F中生成的第二数据结构M2的最终版本(例如,BDP),其中示出了“1”和“0”位。图4H描绘了在图4C至图4F中生成的第二数据结构M2的最终版本,其叠加在第一数据结构M1上,为了简单起见,每个被跟踪的用户数据写入UDW被分为多个区段。在位图视图中,第二数据结构M2被示为叠加在第一数据结构M1上,以描绘当前复制位置CCL的进度如何对应于对第二数据结构M2中指示要被复制的源数据的位置的第二状态标识符位MD2(例如,“1”位)的生成的过滤。

[0113] 例如,第一用户数据写入UDW1被分为三个区段—1、2和3;第二用户数据写入UDW2被分为三个区段—4、5和6;并且第三用户数据写入UDW3被分为两个区段—7和8。区段1、4和7描绘了在它们各自的用户数据写入UDW时间在计划且已复制的区段CS中与LBA相对应的位图位置;区段2和5描绘了在它们各自的用户数据写入UDW时间在计划但尚未被复制的区段YS中与LBA相对应的位图位置;并且区段3、6和8描绘了在用户数据写入UDW发生时在未映射的区段US中与LBA相对应的位图位置。

[0114] 图5A、图5B、图5C和图5D(统称为图5)是根据本公开的一些实施例的,描绘使用与连续的存储位置相关联的连续索引和第二数据结构中对第二状态标识符的选择性生成来跟踪实时迁移的数据复制进度的方法的位图视图。

[0115] 参考图5,可以在一系列位图视图中描绘前述对图4C至图4F的描述。总之,在一些实施例中,实时迁移控制器210b可以监视第一数据结构M1(例如,BMP或先前的BDP)的复制索引(例如,当前复制位置CCL),并且可以禁止或阻止在计划要复制的区段(与计划但尚未被复制的区段YS)中当前复制位置CCL之上与LBA相对应的第二数据结构M2的位置中生成指示要被复制的源数据的位置的第二状态标识符位。因此,第二数据结构M2的最终版本可以被过滤掉可能导致冗余复制的脏位(例如,如图5B至图5D中描绘的,用户数据写入UDW区段2和5所描绘的)。

[0116] 图6至图10所示的本公开的实施例的各方面使用位图视图来讨论。

[0117] 图6A、图6B、图6C、图6D和图6E(统称为图6)是根据本公开的一些实施例的,描绘使用与分布式存储位置相关联的计划但尚未被复制的位图和计划且已复制的位图来跟踪实时迁移的数据复制进度的方法的位图视图。

[0118] 在一些实施例中,第一数据结构M1可以对应于源存储装置200中不位于连续块中的源数据。例如,与源存储装置200中要复制到目标存储装置300的相关数据相关联的存储位置(例如,LBA地址)可以彼此不相邻(例如,可以是分布式的)。

[0119] 参考图6,已将间隙GAP添加到图6A至图6D的位图视图,以描绘与此类分布式(例如,非相邻)存储位置(例如,LBA地址)相对应的非相邻位图位置。在这种情况下,可以通过计划但尚未被复制的位图YBM和计划且已复制的位图CBM来跟踪当前复制位置CCL。可以通过与指示第一分布式地址D1、第二分布式地址D2、第三分布式地址D3和第四分布式地址D4的位置的位图位(例如,状态标识符)相对应的阴影部分来查看当前复制位置CCL的进度(参见图6E中描绘的放大的位图)。

[0120] 参考图6A,与第一分布式地址D1、第二分布式地址D2、第三分布式地址D3和第四分布式地址D4中的每一个相关联的状态标识符(例如,1位)可以在第零时间 t_0 都在计划但尚未被复制的位图YBM中。例如,基于指示分布式存储位置(例如,D1至D4)的第一数据结构M1

的LMCP过程可能在图6A中描绘的时间尚未开始,使得所有的分布式存储位置都由计划但尚未被复制的位图YBM来指示(例如,在其内)。参考图6B,在第一时间 t_1 ,基于指示分布式存储位置(例如,D1至D4)的第一数据结构M1的LMCP过程可能正在进行,其中与第一分布式地址D1相关联的数据已经被复制。因此,与第一分布式地址D1相关联的对应的状态标识符可以从计划但尚未被复制的位图YBM中被移除,并且被添加到计划且已复制的位图CBM中。过程可以继续,直到与第一分布式地址D1、第二分布式地址D2、第三分布式地址D3和第四分布式地址D4相关联的所有状态标识符都从计划但尚未被复制的位图YBM中被移除(例如,被清除)并且被添加到计划且已复制的位图CBM中为止。

[0121] 例如,参考图6C,在第二时间 t_2 ,基于指示分布式存储位置(例如,D1至D4)的第一数据结构M1的LMCP过程可能正在进行,其中与第一分布式地址D1和第二分布式地址D2相关联的数据已经被复制。因此,与第二分布式地址D2相关联的对应的状态标识符可以从计划但尚未被复制的位图YBM中被移除,并且被添加到计划且已复制的位图CBM中。参考图6D,在第三时间 t_3 ,基于指示分布式存储位置(例如,D1至D4)的第一数据结构M1的LMCP过程可以完成,其中与第一分布式地址D1、第二分布式地址D2、第三分布式地址D3和第四分布式地址D4相关联的数据已经被复制。因此,与第三分布式地址D3和第四分布式地址D4相关联的对应的状态标识符可以从计划但尚未被复制的位图YBM中被移除,并且被添加到计划且已复制的位图CBM中。

[0122] 图7A、图7B、图7C和图7D(统称为图7)是根据本公开的一些实施例的,描绘使用与图6的分布式存储位置相关联的计划但尚未被复制的位图和计划且已复制的位图以及在第二数据结构中对第二状态标识符的选择性生成来跟踪实时迁移的数据复制进度的方法的位图视图。

[0123] 参考图7,实时迁移控制器210b可以通过在第二数据结构M2中选择性地生成指示要被复制的源数据的位置的第二状态标识符位MD2来跟踪在第一数据结构M1的传送期间或之后发生的分布式用户数据写入UDW。例如,应用了选择性生成方法的第二数据结构M2可以比没有应用选择性生成方法的其他类似的第二数据结构XM2产生更少的脏位(如图7B、图7C和图7D中的每一个在下面所示的)。

[0124] 参考图7A,在第零时间 t_0 ,基于指示分布式存储位置的第一数据结构M1的LMCP过程可能尚未开始,使得所有的分布式存储位置都在计划但尚未被复制的位图YBM内被指示。

[0125] 参考图7B,在第一时间 t_1 ,基于指示分布式存储位置的第一数据结构M1的LMCP过程可能正在进行,其中与分布式存储位置中的一些相关联的数据已经被复制。第一用户数据写入UDW1可以在第一时间 t_1 发生,使得第一用户数据写入UDW1的第一区段可以在与计划且已复制的位图CBM的分布式存储位置处发生,第一用户数据写入UDW1的第二区段可以在与计划但尚未被复制的位图YBM的分布式存储位置处发生,并且第一用户数据写入UDW1的第三区段可以在与未映射的区段US的分布式存储位置处发生。因此,可以在第二数据结构M2中选择性地生成指示要被复制的源数据的位置的第二状态标识符位MD2,使得可以不在第二数据结构M2中跟踪第一用户数据写入UDW1的第二区段。

[0126] 参考图7C,在第二时间 t_2 ,基于指示分布式存储位置的第一数据结构M1的LMCP过程可能正在进行,其中与分布式存储位置中的更多位置相关联的数据已经被复制。第二用户数据写入UDW2可以在第二时间 t_2 发生,使得第二用户数据写入UDW2的第四区段可以在与

计划且已复制的位图CBM的分布式存储位置处发生,第二用户数据写入UDW2的第五区段可以在与计划但尚未被复制的位图YBM的分布式存储位置处发生,并且第二用户数据写入UDW2的第六区段可以在与未映射的区段US的分布式存储位置处发生。因此,可以在第二数据结构M2中选择性地生成指示要被复制的源数据的位置的第二状态标识符位MD2,使得可以不在第二数据结构M2中跟踪第二用户数据写入UDW2的第五区段。

[0127] 参考图7D,在第三时间 t_3 ,基于指示分布式存储位置的第一数据结构M1的LMCP过程可以完成,其中与所有的分布式存储位置相关联的数据都已经被复制。第三用户数据写入UDW3可以在第三时间 t_3 发生,使得第三用户数据写入UDW3的第七区段可以在与计划且已复制的位图CBM的分布式存储位置处发生,并且第三用户数据写入UDW3的第八区段可以在与未映射的区段US的分布式存储位置处发生。因此,可以在第二数据结构M2中选择性地生成指示要被复制的源数据的位置的第二状态标识符位MD2,使得可以在第二数据结构M2中跟踪第三用户数据写入UDW3的所有区段。

[0128] 在选择性清除方法的情况下(以下参考图10和图13进行讨论),第二数据结构M2和其他类似的第二数据结构XM2最初(例如,在第一时间 t_1)看起来相同,但是第二数据结构M2具有稍后(例如,在第一时间 t_1 之后)将被选择性地清除的、指示要被复制的源数据的位置的第二状态标识符位MD2。

[0129] 图8A、图8B和图8C(统称为图8)是根据本公开的一些实施例的,描绘使用与图7的第二数据结构相对应的计划但尚未被复制的位图和计划且已复制的位图以及在第三数据结构中对第三状态标识符的选择性生成,来在数据复制的后续迭代中跟踪实时迁移的数据复制进度的方法的位图视图。

[0130] 参考图8,相同的选择性生成方法可以应用于后续的LMCP迭代。例如,如图2所示,第二数据结构M2可以在(在第二数据结构M2中)主动地跟踪在第一数据结构M1被传送到实时迁移服务器100b用于第一LMCP过程期间或之后发生的对源存储装置200的用户数据写入UDW。在第一LMCP过程之后,第二数据结构M2可以被传送到实时迁移服务器100b,如图3所示,以用于第二次LMCP迭代。实时迁移控制器210b可以在(在第三数据结构M3中)跟踪在第二数据结构M2被传送到实时迁移服务器100b用于第二LMCP过程期间或之后发生的用户数据写入UDW。因此,如图8B和图8C所示的,应用了选择性生成方法的第三数据结构M3可以比没有应用选择性生成方法的类似的第三数据结构XM3产生更少的脏位。

[0131] 例如,参考图8A,在第零时间 t_0 ,基于指示要被复制的存储位置(例如,分布式存储位置)的第二数据结构M2的LMCP过程可能尚未开始,使得所有的存储位置都在计划但尚未被复制的位图YBM内被指示。此外,因为可能已经在先前的LMCP迭代期间基于选择性生成(或选择性清除)技术生成了第二数据结构M2,所以图8A指示了第二数据结构中在计划要复制的区段P内的存储位置和第二数据结构中在计划要复制的区段O外的存储位置。

[0132] 参考图8B,在第一时间 t_1 ,基于第二数据结构M2的LMCP过程可能正在进行,其中与存储位置中的一些相关联的数据已经被复制。第一用户数据写入UDW1可以在第一时间 t_1 发生,使得第一用户数据写入UDW1的第一区段可以在与计划且已复制的位图CBM相对应的一些存储位置和与第二数据结构中在计划要复制的区段O外的一些存储位置处发生,第一用户数据写入UDW1的第二区段可以仅在第二数据结构中在计划要复制的区段O外的存储位置处发生,并且第一用户数据写入UDW1的第三区段可以仅在与计划但尚未被复制的位图YBM

的存储位置处发生。因此,可以在第三数据结构M3中选择性地生成指示要被复制的源数据的位置的第三状态标识符位MD3,使得不在第三数据结构M3中跟踪第一用户数据写入UDW1的第三区段(并且可以在第三数据结构M3中跟踪第一用户数据写入UDW1的第一区段和第二区段)。

[0133] 参考图8C,在第二时间 t_2 ,基于第二数据结构M2的LMCP过程可以完成,其中与所有的存储位置相关联的数据都已经被复制。第二用户数据写入UDW2可以在第二时间 t_2 发生,使得第二用户数据写入UDW2的第四区段、第五区段和第六区段都可以在与计划且已复制的位图CBM相对应的存储位置或第二数据结构中在计划要复制的区段0外的存储位置处发生。因此,可以在第三数据结构M3中选择性地生成指示要被复制的源数据的位置的第三状态标识符位MD3,使得可以在第三数据结构M3中跟踪第二用户数据写入UDW2的所有区段。

[0134] 图9A、图9B和图9C(统称为图9)是根据本公开的一些实施例的,描绘使用图8的计划但尚未被复制的位图以及通过反转用于跟踪计划但尚未被复制的区段的位状态跟踪计划且已复制的区段,来跟踪实时迁移的数据复制进度的方法的位图视图。

[0135] 参考图9,可以创建单个位图、计划但尚未被复制的位图/计划且已复制的位图YBM/CBM(“YBM/CMB位图”),并由实时迁移控制器210b用来跟踪两种区段。例如,实时迁移控制器210b可以利用第一位状态(例如“1”)来标识位图中与源存储装置200的计划但尚未被复制的区段YS相对应的位置。参考图9A,在第零时间 t_0 ,与YBM/CMB位图相关联的所有的状态标识符都可以与源存储装置200的计划但尚未被复制的区段相关联。参考图9B和图9C,与计划且已复制的位图相对应的位图位置可以通过在与该位图位置相关联的源数据已经被实时迁移服务器100b复制之后,将第一位状态反转为第二位状态(例如“0”)来标识。因此,可以减少用于跟踪当前复制位置的位图的存储器大小。例如,可以使用单个计划但尚未被复制的位图YBM,而不是使用三个位图(例如,计划且已复制的位图CBM、计划但尚未被复制的位图YBM和未映射的位图)。

[0136] 图10是根据本公开的一些实施例的,包括描绘使用第二数据结构和在第二数据结构中对第二状态标识符的选择性清除来跟踪实时迁移过程的数据复制进度的系统和方法的位图视图的系统图。

[0137] 参考图10,在一些实施例中,实时迁移控制器210b可以通过从第二数据结构M2中选择性地清除(例如,将脏位从“1”重置或改变为“0”)指示要被复制的源数据的位置的第二状态标识符位MD2(而不参考计划但尚未被复制的区段YS或计划且已复制的区段CS)来减少实时迁移服务器100b的冗余复制。例如,实时迁移控制器210b可以(例如,并行地)监视用户数据写入UDW和当前复制位置CCL。例如,实时迁移控制器210b可以监视第一时间 t_1 的第一用户数据写入UDW1,并且在第二数据结构中生成与第一数据写入UDW1的所有位置(例如,区段1、2和3)相对应的、指示要被复制的源数据的位置的第二状态标识符位MD2(例如,设置为“1”)。因此,在第一时间 t_1 ,与第一用户数据写入UDW1的位置相对应的所有位图位置(即使在区段2处)都可以具有指示要被复制的源数据的位置的第二状态标识符位MD2(例如,脏位)。基于指示要被复制的源数据的位置的第二状态标识符位MD2(例如,脏位或“1”)在第二时间 t_2 (比第一时间 t_1 晚的时间)被标识为与当前复制位置CCL相对应,实时迁移控制器210b可以从第二数据结构M2中清除与当前复制位置CCL(例如,第二区段2中的脏位)相关联的、指示要被复制的源数据的位置的第二状态标识符位MD2(例如,可以将“1”位清除为“0”位),

因为已经从源存储装置200中复制了与脏位(1)相关联的数据,因此不需要在稍后的LMCP迭代中再对其进行复制。在一些实施例中,实时迁移控制器210b可以清除(例如,自动地清除)第二数据结构M2中与当前复制位置CCL相对应的条目(例如,而不用确定对应的条目是被设置为“1”位还是“0”位)。

[0138] 在一些实施例中,使用选择性清除方法,系统1可以不结合第一数据结构M1(或替代的数据结构),以跟踪当前复制位置CCL/计划且已复制的区段CS/计划但尚未被复制段YS、或者跟踪当前复制位置CCL/计划且已复制的位图CBM/计划但尚未被复制的位图YBM。例如,可以仅使用第二数据结构M2和第一数据结构M1的一小段时间来(例如,在第一数据结构M1到实时迁移服务器100b的传送期间)跟踪用户数据写入UDW。这可以得到更高效的存储器使用,因为第一数据结构M1可以在其已经被传送到实时迁移服务器100b之后(例如,紧接着之后)在实时迁移控制器存储器220b中被回收。根据这样的选择性清除实施例,在大部分实时迁移操作时间期间,除了在传送第二数据结构M2(或脏位图的后续迭代,例如,第三数据结构M3)时,可以仅执行脏位图跟踪。

[0139] 因此,与其他方法相比,选择性清除方法可以提供若干的改进和优势。例如,在一些实施例中,选择性生成方法可以使用细分的第一数据结构M1位图存储器的一个或三个复制。在一些实施例中,选择性生成方法可以读取当前计划但尚未被复制的区段YS/计划但尚未被复制的位图YBM状态,将状态清除为0,并且将“1”位写入计划且已复制的区段CS/计划且已复制的位图CBM。相反,在一些实施例中,选择性清除方法可以不使用第一数据结构M1。因此,在大部分操作时间内,可能会使用较少的存储器。此外,在一些实施例中,选择性清除可以不对计划且已复制的区段CS/计划且已复制的位图CBM和计划但尚未被复制的YS/计划但尚未被复制的位图YBM数据库执行额外的读取和写入。

[0140] 图11是根据本公开的一些实施例的,使用上文参考图1-图10描述的方法的各方面,描绘跟踪实时迁移过程的数据复制进度的方法的示例操作的简化概要流程图。

[0141] 参考图11,跟踪实时迁移的数据复制进度的方法1100可以包括:从迁移服务器接收将数据从源存储装置传送到目标存储装置的请求(操作1101);将第一数据结构从存储控制器传送到实时迁移服务器,该第一数据结构包括第一状态标识符(操作1102);清除第二数据结构中的所有位置(操作1103);基于实时迁移服务器的当前复制位置,在第二数据结构中选择性地生成第二状态标识符(例如,将位置设置为“1”)(操作1104A);或者在第二数据结构中非选择性地生成第二状态标识符,并且基于实时迁移服务器的当前复制位置,从第二数据结构中选择性地清除第二状态标识符(例如,将位置清除为“0”)(操作1104B);以及将第二数据结构传送到实时迁移服务器(操作1105)。

[0142] 图12是根据本公开的一些实施例的,描绘跟踪实时迁移过程的数据复制进度的选择性生成方法的示例操作的流程图。

[0143] 参考图12,跟踪实时迁移的数据复制进度的方法1200可以包括:由实时迁移(“LM”)服务器向实时迁移控制器发送请求(操作1201);由实时迁移控制器生成第一数据结构M1(例如,映射的页的位图(“BMP”))(操作1202);初始化第二数据结构M2(例如,初始或后续BDP)并且将所有的状态标识符位设置(例如,清除)为0(操作1203);确定用户写入请求(例如,普通用户写入请求)是否已经从源存储设备发出(操作1204)(注意:流程图左侧的操作可以与第二数据结构M2相关联,而流程图右侧的操作可以与第一数据结构M1相关联;流

程图左侧和右侧的操作可以基本上并行发生);由源存储控制器生成对源存储装置的写入命令(操作1205);将用户数据写入UDW写入源存储装置(操作1206);由实时迁移控制器(例如,通过提取与用户数据写入UDW相关联的LBA地址)窥探(例如,监视)到源存储装置的存储位置(例如,LBA)的用户数据写入UDW(操作1207);由实时迁移控制器基于确定用户数据写入UDW是否在计划但尚未被复制的区段YS/YBM中发生来确定是否在第二数据结构M2中生成脏位(例如,通过确定与用户数据写入UDW的LBA地址相对应的计划但尚未被复制的位图YS/YBM的状态标识符位是否被设置为“1”,从而指示该LBA地址在计划但尚未被复制的区段YS/YBM内)(操作1208)(也参见相关操作1218);基于确定用户数据写入UDW已经在计划但尚未被复制的区段YS/YBM中发生,在第二数据结构M2中生成脏位(操作1209A);基于确定用户数据写入UDW没有在计划但尚未被复制的区段YS/YBM中发生,禁止(或者阻止或放弃)在第二数据结构M2中生成脏位(操作1209B);基于实时迁移服务器100b中的第一数据结构的主机版本M1'(例如,BMP),确定数据复制是否已经完成(并且基于确定数据复制尚未完成,循环回到操作1204)(操作1210);由实时迁移控制器(例如,通过将第三数据结构M3的所有状态标识符位都清除为“0”)初始化第三数据结构M3(操作1211);将第二数据结构M2传送到实时迁移服务器(操作1212);将第一数据结构M1(例如,BMP)传送到实时迁移服务器(操作1213);由实时迁移控制器(例如,通过将所有状态标识符位都设置为0)初始化计划且已复制的区段CS/计划且已复制的位图CBM,并且(例如,通过将所有状态标识符位都设置为对应于第一数据结构M1的状态标识符位(例如,1))初始化计划但尚未被复制的区段YS/计划但尚未被复制的位图YBM(操作1214);确定LM复制读取(LM Copy Read)是否发生(并且基于确定LM复制读取没有发生而循环回到操作1215)(操作1215);响应于确定LM复制读取已经发生,(例如,通过提取与LM CPR读取相关联的LBA地址)窥探(例如,监视)LM CPR读取LBA地址(操作1216);由实时迁移服务器(基于确定LM复制读取已经发生)从源存储装置读取(操作1217);(例如,基于窥探的LBA地址)在计划但尚未被复制的区段YS/计划但尚未被复制的位图YBM中清除状态标识符位条目(例如,指示要被复制的源数据的位置的计划但尚未被复制的区段状态标识符位YS.MD1或YS/YBM.MD1)(操作1218)(注意:虚线1218x指示在操作1208设置为1的状态标识符位可以基于对应的数据正被复制而在操作1218被清除);(例如,基于窥探的LBA地址)在计划且已复制的区段CS/计划且已复制的位图CBM中添加(例如,设置)状态标识符位条目(例如,指示已复制的源数据的位置的计划且已复制的区段状态标识符位CS.MD1或CS/CBM.MD1)(操作1219);确定第一数据结构M1的数据复制是否已经完成(操作1220)(注意:虚线1220x指示操作1210和1220两者中都可以包括(例如,在相同的数据库中)检查相同的BMP COPY END(BMP复制结束)结果);以及基于确定由实时迁移服务器进行的第一数据结构M1的数据复制已经完成,停止当前迭代的LM CPR窥探操作(操作1221)。

[0144] 图13是根据本公开的一些实施例的,描绘跟踪实时迁移过程的数据复制进度的选择性清除方法的示例操作的流程图。

[0145] 参考图13,跟踪实时迁移的数据复制进度的方法1300可以包括:由实时迁移(“LM”)服务器向实时迁移控制器发送请求(操作1301);由实时迁移控制器生成第一数据结构M1(例如,映射的页的位图(“BMP”))(操作1302);初始化(例如,生成或创建)第二数据结构M2(例如,初始或后续BDP)并且将所有的状态标识符位都设置(例如,清除)为0(操作1303);确定用户写入请求(例如,普通用户写入请求)是否已经从源存储设备发出(例如,发

生) (并且基于确定没有用户写入请求发出而循环回到操作1304) (操作1304) (注意:流程图左侧的操作可以与第二数据结构M2相关联,而流程图右侧的操作可以与第一数据结构M1相关联;流程图左侧和右侧的操作可以基本上并行发生);由源存储控制器生成对源存储装置的写入命令(操作1305);将用户数据写入UDW写入源存储装置(操作1306);由实时迁移控制器(例如,通过提取与用户数据写入UDW相关联的LBA地址)窥探(例如,监视)用户数据写入UDW到源存储装置的存储位置(例如,LBA)(操作1307);在第二数据结构M2中生成脏位(操作1308)(也参见相关操作1317);确定基于第一数据结构M1(例如,BMP)的数据复制是否已经完成(并且基于确定基于第一数据结构M1的数据复制尚未完成,循环回到操作1304)(操作1309);基于确定基于第一数据结构M1的数据复制已经完成,由实时迁移控制器(例如,通过将第三数据结构M3的所有状态标识符位都设置为“0”)初始化第三数据结构M3(操作1310);将第二数据结构M2传送到实时迁移服务器(操作1311);将第一数据结构M1(例如,BMP)传送到实时迁移服务器(操作1312);确定LM复制读取是否发生(并且基于确定LM复制读取没有发生而循环回到操作1313)(操作1313);基于确定LM复制读取已经发生,(例如,通过提取与LMCPR读取相关联的LBA地址)窥探(例如,监视)LMCPR读取LBA地址(操作1314);由实时迁移服务器从源存储装置进行读取(操作1315);(例如,基于窥探的LBA)在第二数据结构M2中清除状态标识符位(例如,脏位)条目(操作1316)(注意:虚线1316x指示在操作1308在第二数据结构M2中生成的脏位可以在操作1316被选择性地清除);确定第一数据结构M1的数据复制是否已经完成(操作1317)(注意:虚线1317x指示操作1309和1317两者都可以包括(例如,在相同的数据库中)检查相同的BMP COPY END结果);以及基于确定由实时迁移服务器进行的第一数据结构M1的数据复制已经完成,停止当前迭代的LMCPR窥探操作(操作1318)。

[0146] 因此,本公开的实施例提供了针对在实时迁移过程期间跟踪数据复制进度的改进和优点。通过在脏位图中选择性地生成或选择性地清除脏位,可以减少源数据的冗余复制,并且可以更高效地使用计算资源(例如,存储器、带宽、功率以及主机或存储设备处理)。

[0147] 尽管已参考本文所述的实施例具体示出和描述了本公开的实施例,但本领域普通技术人员应当理解,在不脱离所附权利要求书及其等同物阐述的本公开精神和范围的情况下,可以在形式和细节上对本公开进行各种改变。

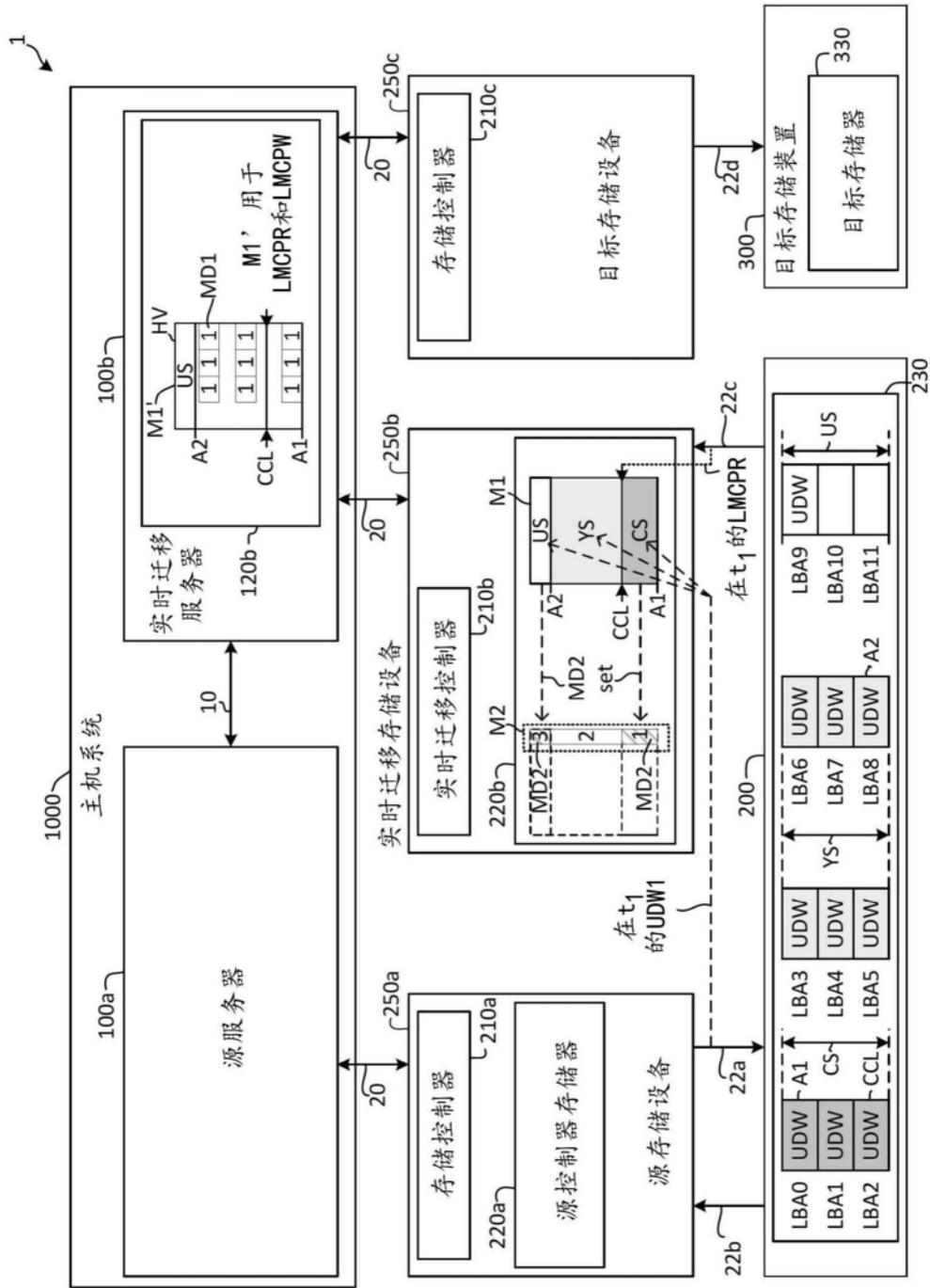


图1

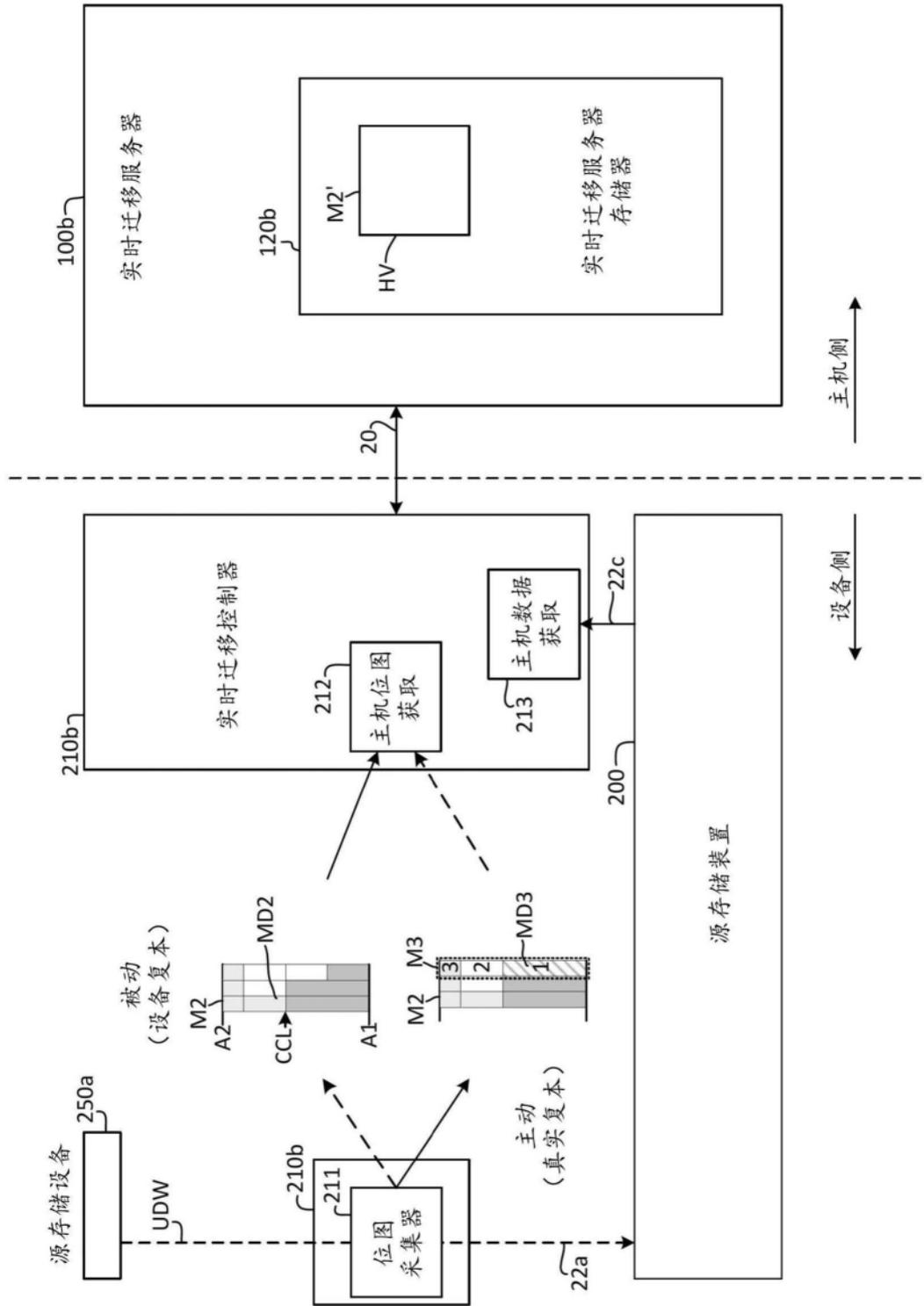


图3

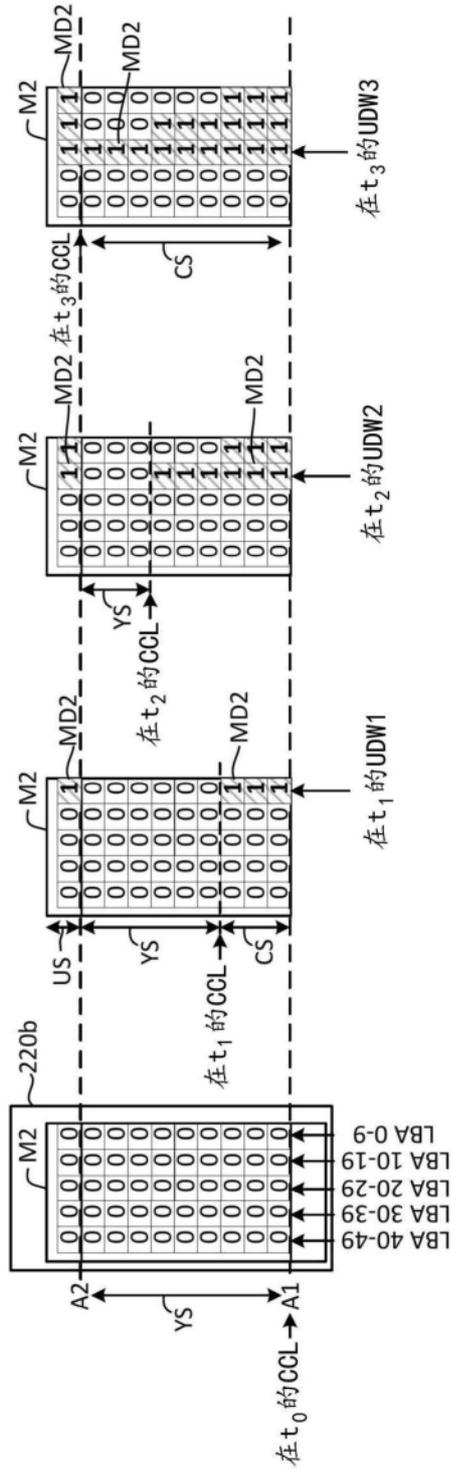


图 4C

图 4D

图 4E

图 4F

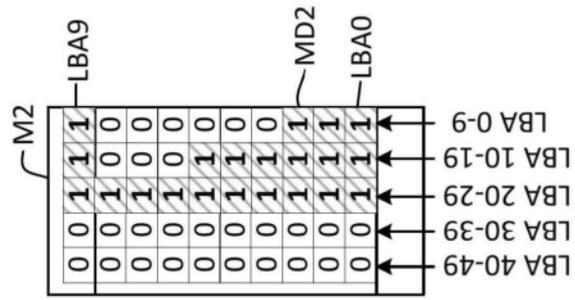


图4G

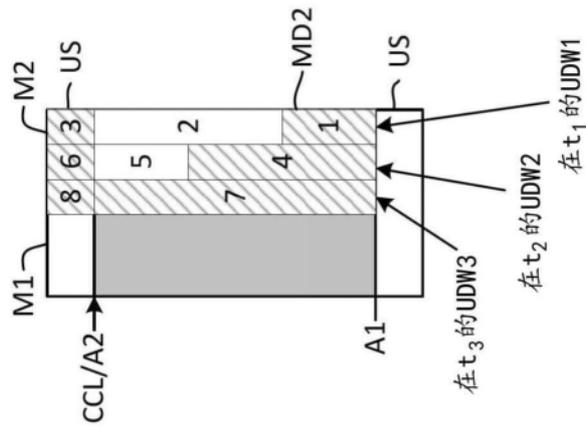


图4H

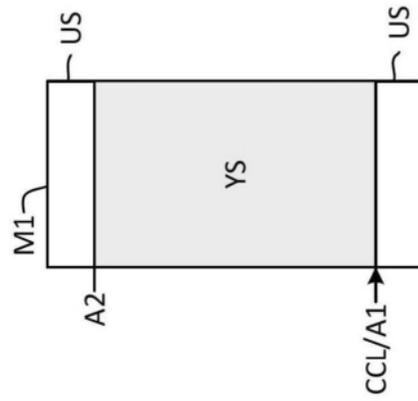


图5A

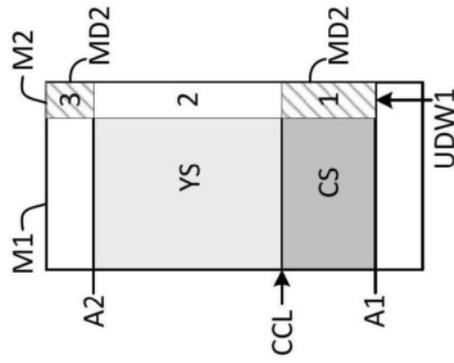


图5B

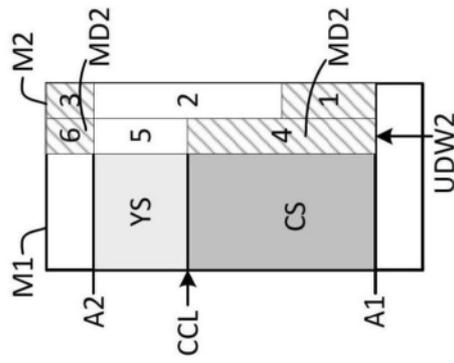


图5C

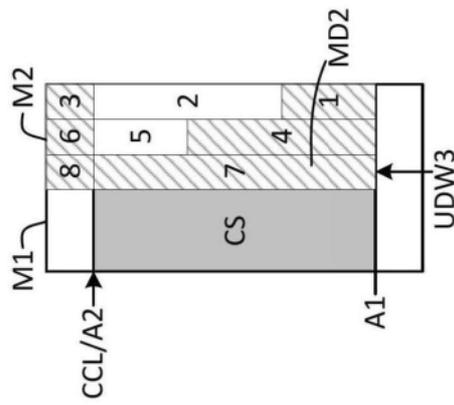


图5D

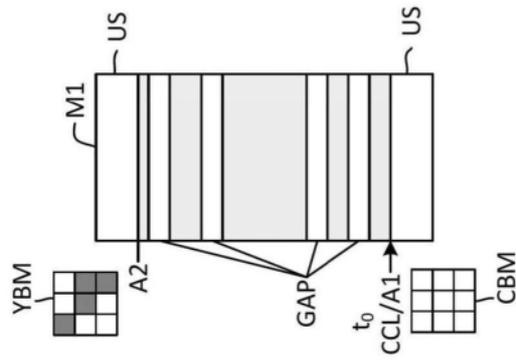


图6A

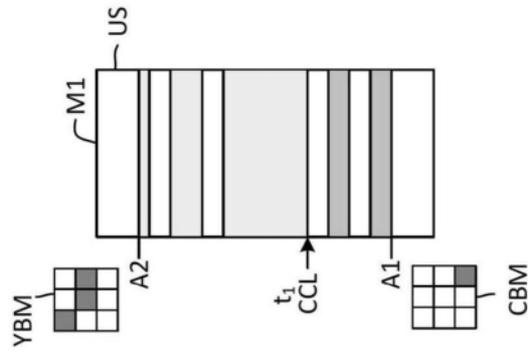


图6B

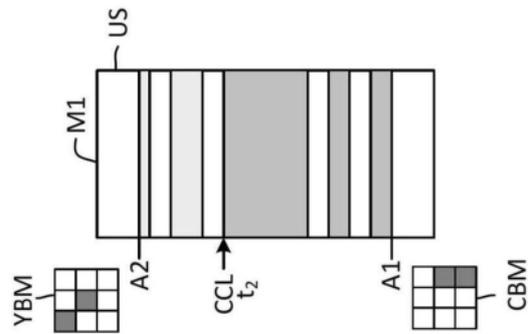


图6C

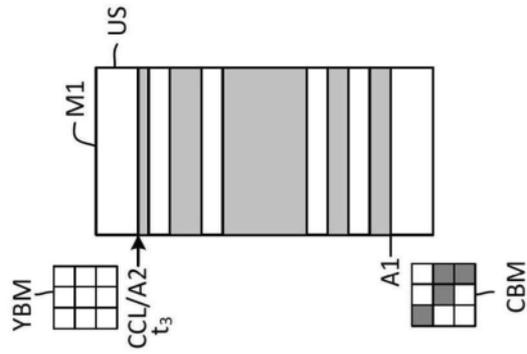


图6D

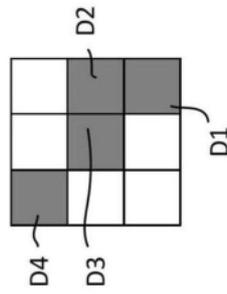


图6E

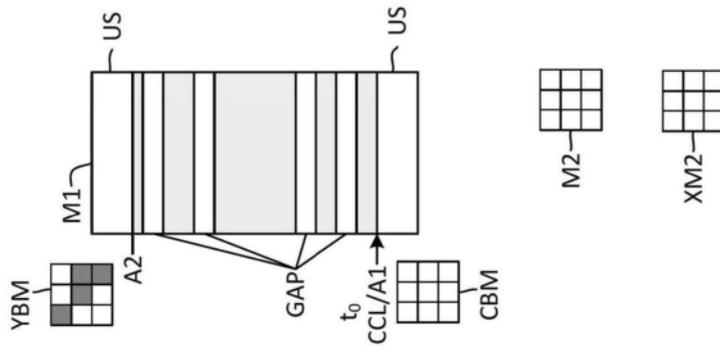


图7A

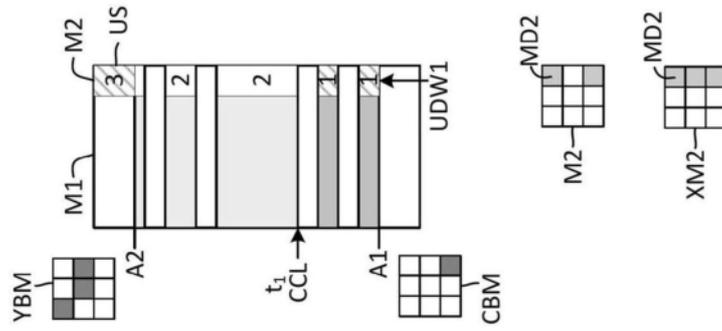


图7B

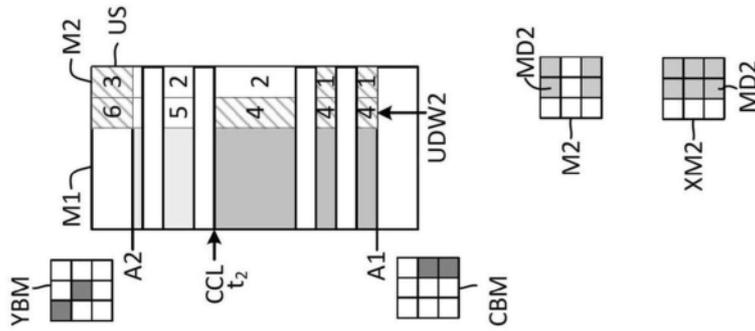


图7C

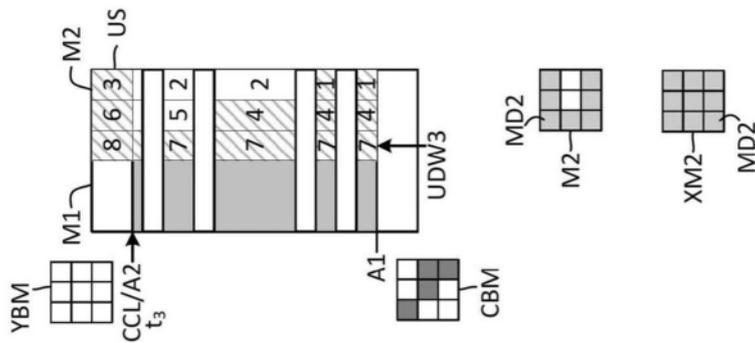


图7D

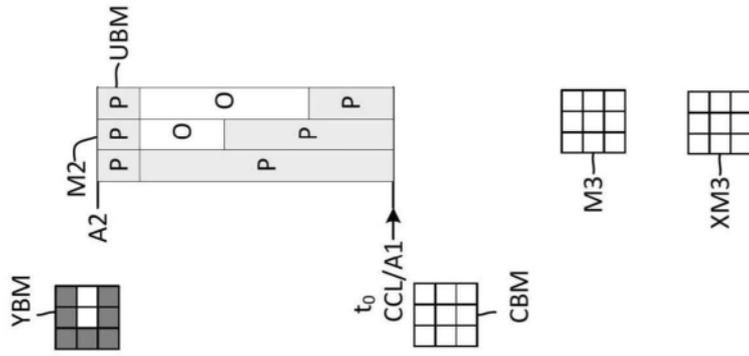


图8A

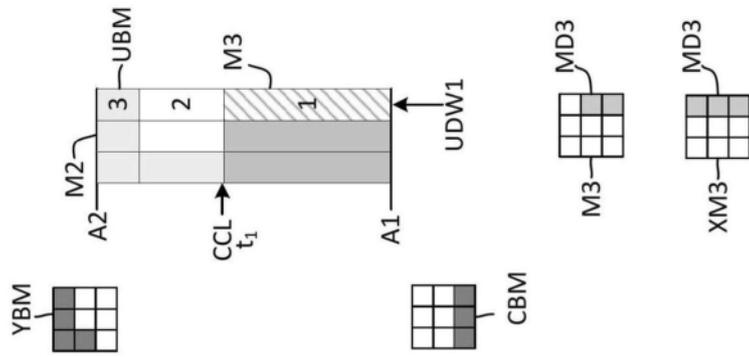


图8B

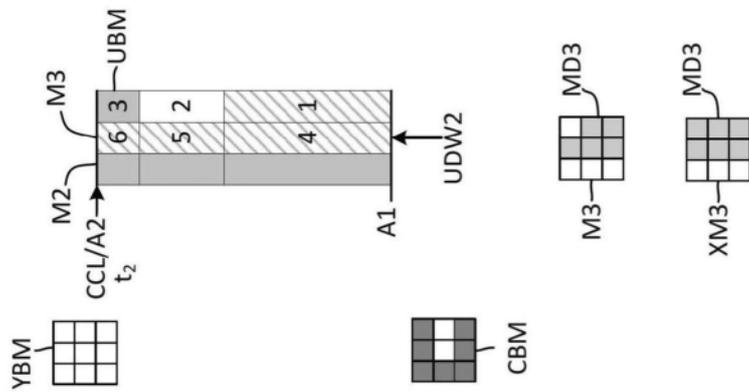


图8C

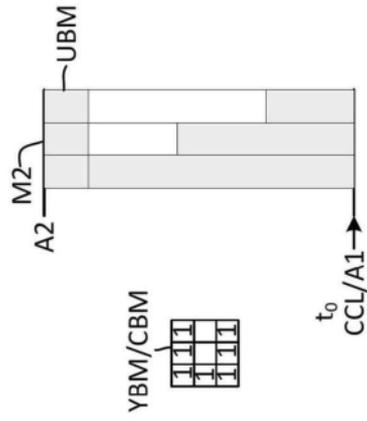


图9A

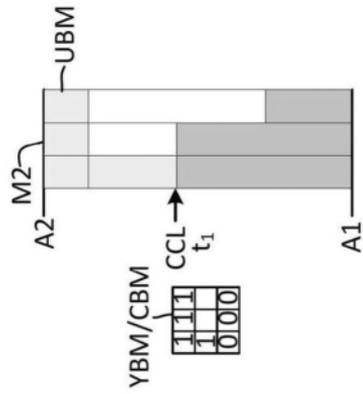


图9B

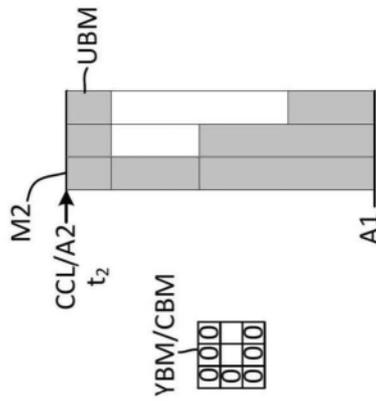


图9C

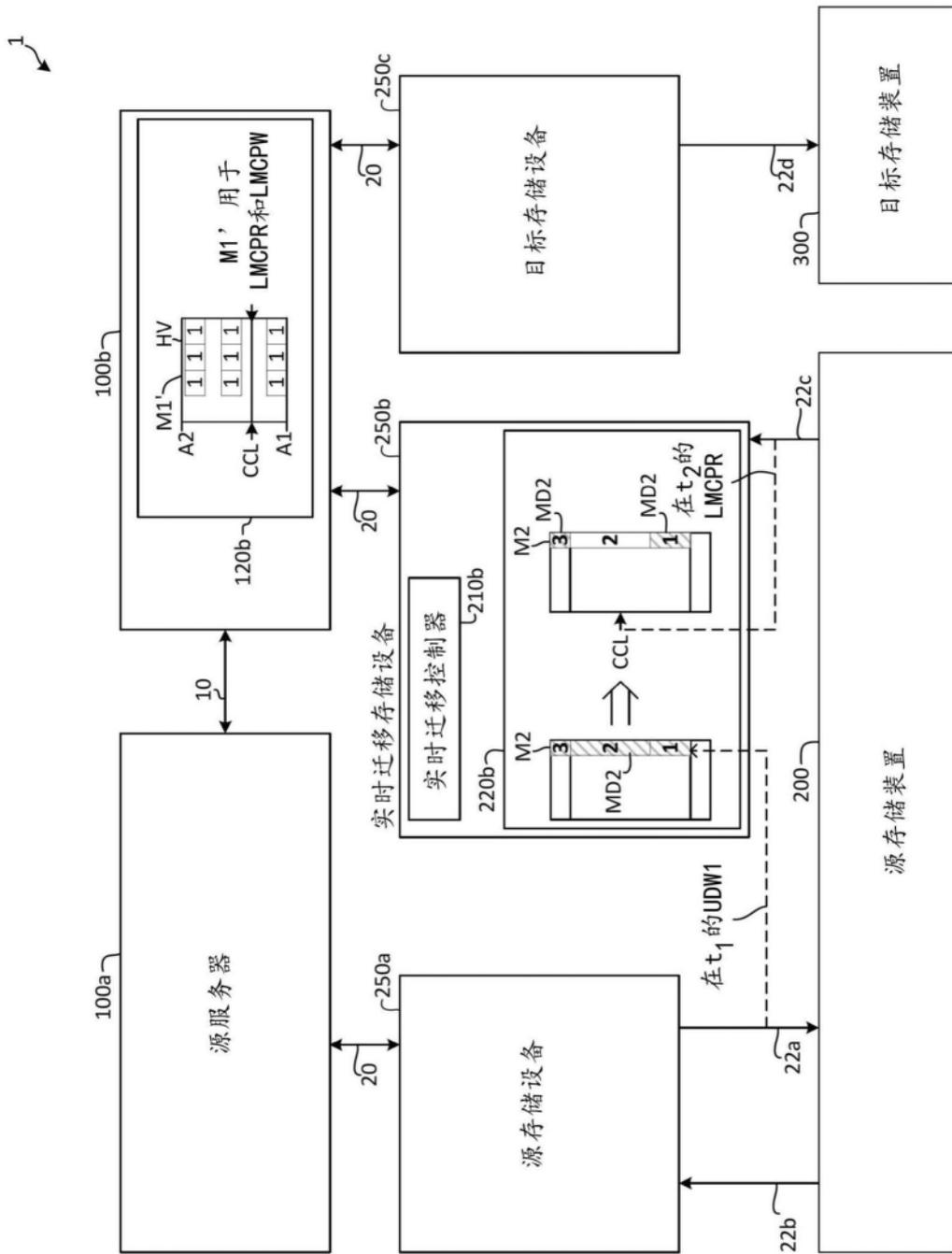


图10

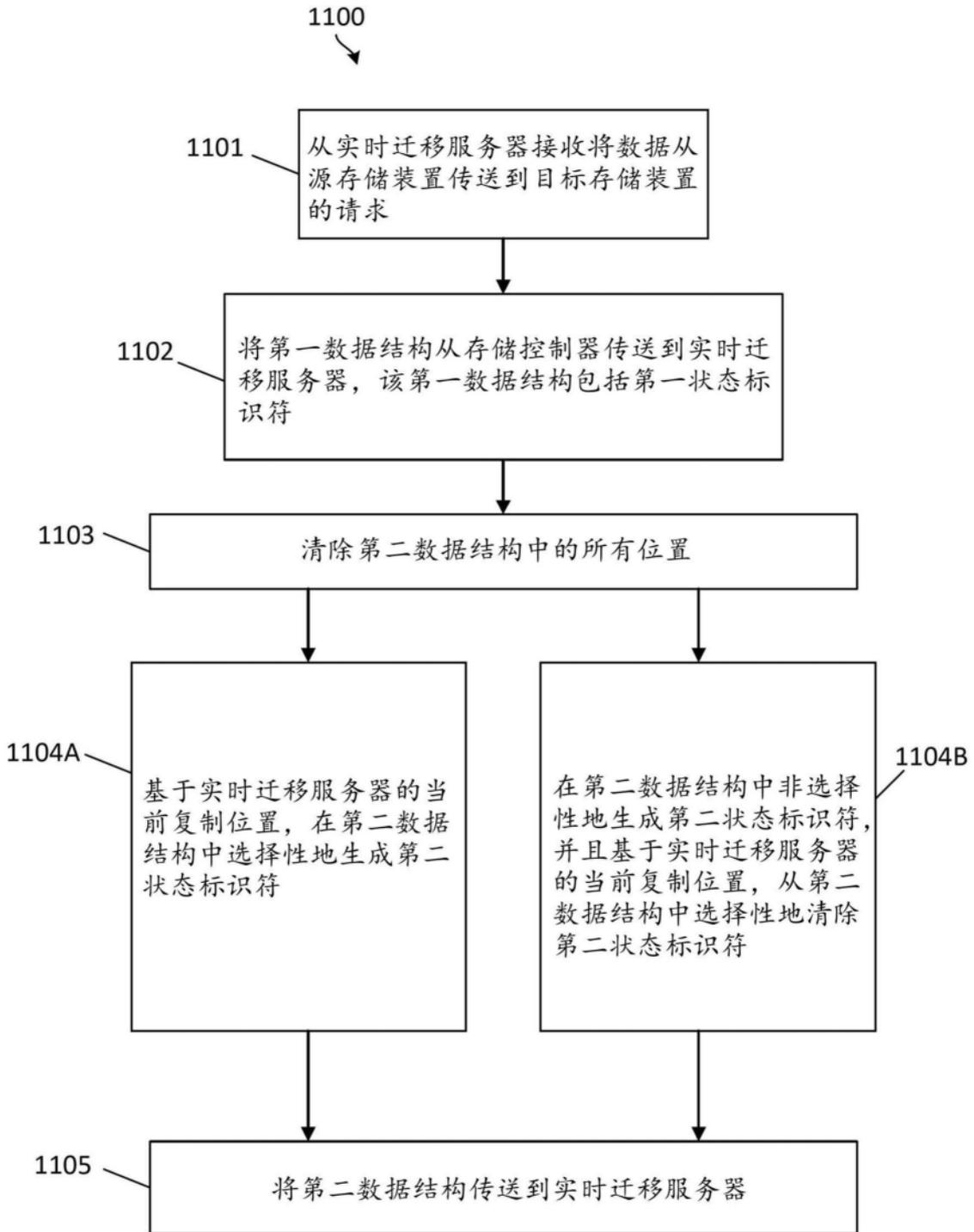


图11

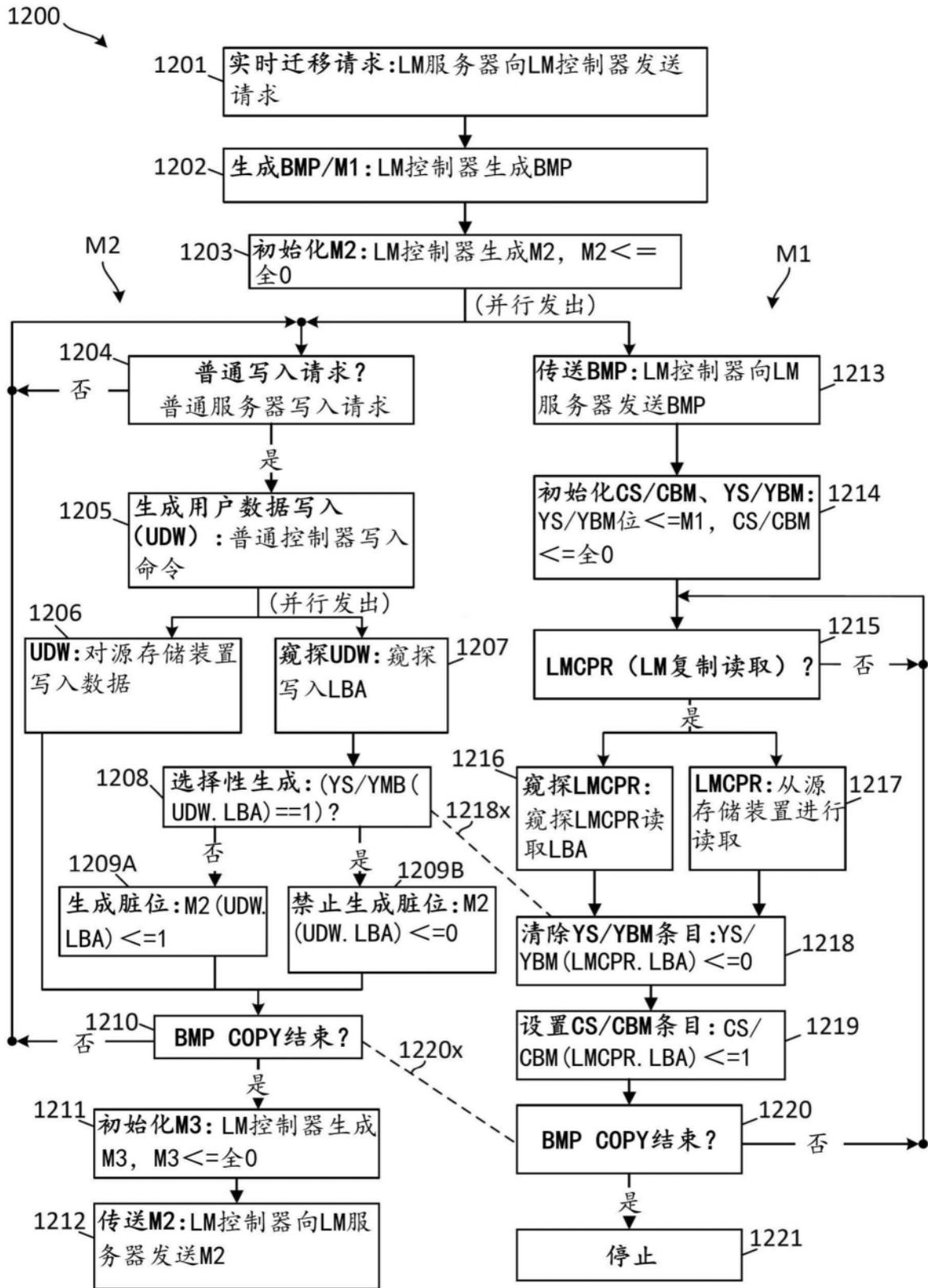


图12

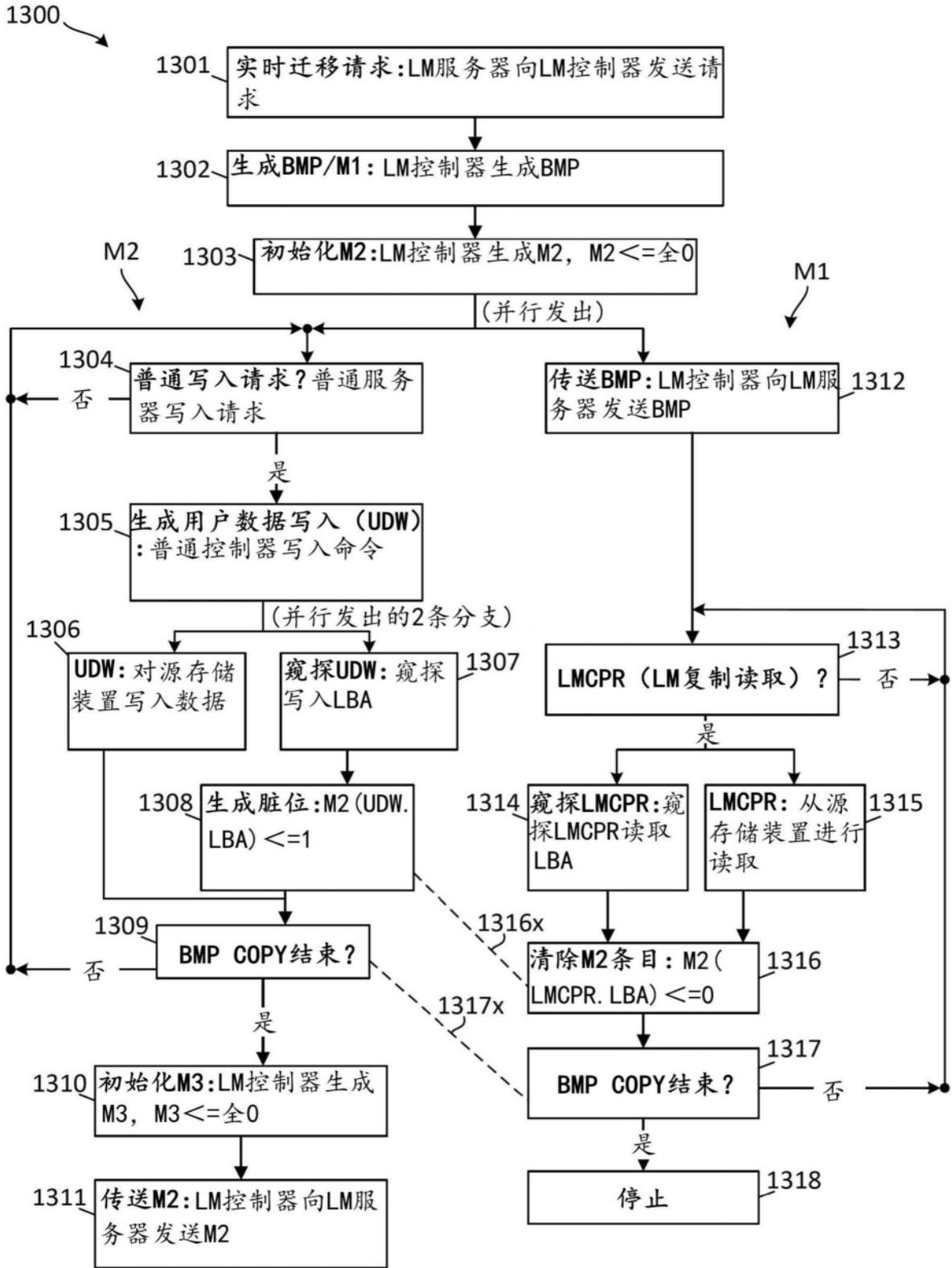


图13