(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 104054319 A
(43)申请公布日 2014.09.17

(21)申请号 201380005239.1
(22)申请日 2013.01.09
(30)优先权数据
13/348,243 2012.01.11 US
(85)PCT国际申请进入国家阶段日
2014.07.11
(86)PCT国际申请的申请数据
PCT/US2013/020783 2013.01.09
(87)PCT国际申请的公布数据
WO2013/106400 EN 2013.07.18
(71)申请人 阿尔卡特朗讯公司
地址 法国布洛涅-比扬古
(72)发明人 K·普塔斯瓦米纳加 T·南达戈帕尔 Y·马
(74)专利代理机构 北京市中咨律师事务所 11247
代理人 刘丽萍 杨晓光
(51)Int.Cl.
H04L 29/08(2006.01)
G06F 17/30(2006.01)

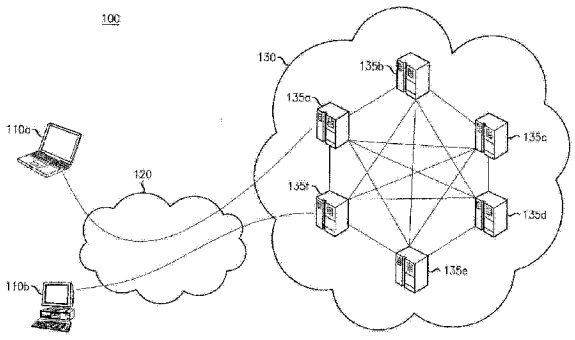
权利要求书2页 说明书9页 附图6页

(54)发明名称

在弹性云文件系统中减少延迟和成本

(57)摘要

各种示例性实施例涉及在包括多个数据中心(135a-f)云系统(130)中存储文件块的方法。该方法可以包括:从客户端110A-b)接收文件块;从文件块生成多个组块,其中每个组块比文件块小并且文件块可以从组块的子集重建;将每个组块分发给多个数据中心(135a-f)中的一个;并将文件块存储在高速缓存中。各种示例性实施例涉及用于存储文件的云系统。该系统可以包括含主数据中心的多个数据中心。主数据中心可以包括:被配置为存储至少一个完整文件块(260)的高速缓存;被配置为对多个文件块(240)的每一个存储组块的组块存储器;文件编码器(230);以及文件解码器(250)。



CN 104054319 A

1. 一种用于在包括多个数据中心 (135a-e) 的云系统 (130) 中存储文件的方法，该方法包括：

在第一数据中心接收来自客户端的文件块 (420)；

从所述文件中块生成多个组块，其中每个组块小于所述文件并且所述文件块能从所述组块的子集重建 (440)；

向所述多个数据中心的至少两个分发所述多个组块；其中所述多个组块的至少第一组块和第二组块被分发给所述多个数据中心的不同数据中心 (450)；以及

在第一数据中心处的高速缓存中存储所述文件块 (430)。

2. 如权利要求 1 所述的方法，还包括：

从客户端接收读取所述文件块的请求 (520)；

确定所述文件块是否存储在高速缓存中 (530)；

如果所述文件块存储在高速缓存中，则向所述客户端发送存储在高速缓存中的所述文件块 (580)。

3. 如权利要求 2 所述的方法，还包括，如果所述文件块没有存储在高速缓存中，则：

从所述多个数据中心请求组块 (540)；

从所述多个数据中心接收所述多个组块的至少一个子集 (550)；

从组块的子集重建所述文件块 (560)；

在第一数据中心处的高速缓存中存储所述文件块 (570)；以及

将所述重建的文件块发送给客户端 (580)。

4. 如权利要求 1-3 任一项所述的方法，其中，从所述文件生成多个组块的步骤包括使用纠删码来生成所述多个组块。

5. 如权利要求 4 所述的方法，其特征在于，所述纠删码是里德所罗门码、MDS 码和 LDPC 码中的至少一个。

6. 如权利要求 1 至 5 任一项所述的方法，还包括：

接收写入所述文件块的请求 (610)；

在高速缓存中 (640) 写入所述文件块；

关闭所述文件 (645)；

从所述文件块生成第二多个组块 (650)；以及

将所述第二多个组块中的每个组块分发给所述多个数据中心的不同数据中心 (655)。

7. 如权利要求 6 所述的方法，其中所述第二多个组块只包括修改的组块。

8. 如权利要求 1-7 任一项所述的方法，还包括：

比较当前高速缓存尺寸的实际存储和传递成本为与先前高速缓存尺寸的假定存储和传递成本 (720)；以及

基于较低存储和传递成本 (735,740,745) 来调整高速缓存尺寸。

9. 如权利要求 1-8 任一项所述的方法，其中所述多个组块根据系统纠删码而生成，其中所述文件块被分成未编码组块的子集和编码组块的的子集。

10. 一种用于存储文件的云系统，该系统包括：

包括主数据中心的多个数据中心 (135a-e)，所述主数据中心包括：

高速缓存 (260)，被配置为存储至少一个完整文件块；

组块存储器 (240)，被配置为对多个文件块的的每一个存储组块；

文件编码器 (230)，被配置为从所述文件块生成多个组块，其中每个组块小于所述文件并且所述文件块能从所述组块的子集重建；以及

文件解码器 (250)，被配置为从所述组块的子集重建完整文件块。

11. 如权利要求 10 所述的云系统，其中，所述高速缓存 (260) 是硬盘。

12. 如权利要求 10 或 11 所述的云系统，其中所述主数据中心还包括客户端接口 (210)，所述客户端接口 (210) 被配置为从客户端接收完整文件块和发送完整文件块到客户端。

13. 如权利要求 10-12 任一项所述的云系统，其中所述主数据中心还包括云接口 (220)，所述云接口 (220) 被配置为将多个组块中的组块分发给所述多个数据中心的每一个，并被配置为从所述多个数据中心接收组块的子集。

14. 如权利要求 10-13 任一项所述的云系统，其中所述文件编码器被配置为使用纠删码来生成所述多个组块。

15. 如权利要求 1 所述的方法或如权利要求 10 所述的云系统，其中子集 (320) 中组块的数目至少比所述文件编码器生成的所述多个组块 (310) 的数目小两个。

在弹性云文件系统中减少延迟和成本

技术领域

[0001] 本文公开的各种示例性实施例一般涉及计算机文件系统。

背景技术

[0002] 云计算可以定义为使用共享资源通过网络传送计算服务。云计算往往需要数据文件的存储，以使它们能够被各种用户访问。文件存储可以被看作是云计算服务。各种最终用户可以访问由云服务提供商存储的文件而不需要不知道文件到底是怎么存储的。

发明内容

[0003] 云计算环境中的文件存储给云服务提供商和用户的带来了各种各样的问题。云服务有时受到延迟问题的影响，因为文件必须位于网络中并通过网络传送给最终用户。用户可能期望以低延迟提供所需的文件的云服务。如果网络的数据中心或部分网络不可用，云服务也可能不可用。用户可能期望当云组件不可用时提供弹性的云服务。对云服务提供商和 / 或用户来说，低延迟和高弹性往往需要额外的成本。

[0004] 鉴于上述情况，人们希望提供用于云计算的文件系统。特别是，人们希望以低成本提供具有低延迟和高回弹性的用于存储文件的方法和系统。

[0005] 鉴于当前对于进行云计算的文件系统的需求，下面提出对各种示例性实施例的简要说明。在以下的简要说明中作了一些简化和省略，这是意在突出和引入各种示例性实施方式的一些方面，而不是限制本发明的范围。后面章节中对优选实施例的详细说明足以使本领域的普通技术人员能够制造和使用本发明的概念。

[0006] 各种示例性实施例涉及一种用于在包括多个数据中心的云系统中存储文件的方法，该方法包括：在第一数据中心接收来自客户端的文件块；从文件中块生成多个组块，其中每个组块小于所述文件并且所述文件块能从所述组块的子集重建；向多个数据中心的不同数据中心分发每个组块；以及在第一数据中心处的高速缓存中存储文件块。

[0007] 在各种替代实施方案中，该方法进一步包括从客户端接收读取所述文件块的请求；确定所述文件块是否存储在高速缓存中；如果所述文件块存储在高速缓存中，则向客户端发送存储在高速缓存中的文件块。该方法还可以包括：如果所述文件块没有存储在高速缓存中，则：从所述多个数据中心请求组块；从所述多个数据中心接收所述多个组块的至少一个子集；从组块子集重建所述文件块；在第一数据中心处的高速缓存中存储所述文件块；以及将所述重建的文件块发送给客户端。

[0008] 在各种替代实施例中，从所述文件生成多个组块的步骤包括使用纠删码来生成所述多个组块。

[0009] 在各种替代实施方案中，该方法还包括：接收写入所述文件块的请求；在高速缓存中写入文件块；关闭文件；从所述文件块生成第二多个组块；以及将所述第二多个组块中的每个组块分发给所述多个数据中心的不同数据中心。所述第二多个组块可以只包括修改的组块。

[0010] 在各种替代实施方案中，该方法进一步包括：比较当前高速缓存尺寸的实际存储和传递成本为与先前高速缓存尺寸的假定存储和传递成本；以及基于较低存储和传递成本来调整高速缓存尺寸。

[0011] 在各种替代实施方案中，该方法进一步包括：确定所述高速缓存已满；并从高速缓存中移除文件块。

[0012] 在各种替代实施方案中，所述多个组块根据系统纠删码而生成，其中所述文件块被分成未编码组块的子集和编码组块的子集。

[0013] 在各种替代实施例中，组块的数目比组块子集的数目至大两个。

[0014] 各种示例性实施例涉及编码为有形的、非瞬时的、机器可读存储介质上的指令的上述方法。

[0015] 各种示例性实施例涉及一种用于存储文件的云系统。该系统可以包括含主数据中心的多个数据中心。所述主数据中心包括：被配置为存储至少一个完整文件块的高速缓存；被配置为对多个文件块的每一个存储组块的组块存储器；被配置为从所述文件块生成多个组块的文件编码器，其中每个组块小于所述文件并且所述文件块能从所述组块的子集重建；以及被配置为从所述组块的子集重建完整文件块的件解码器。

[0016] 在各种替代实施例中，高速缓存是硬盘。

[0017] 在各种替代实施例中，主数据中心还包括被配置为从客户端接收完整文件块和发送完整文件块到客户端的客户端接口。

[0018] 在各种替代实施例中，主数据中心还包括被配置成将多个组块中的组块分发给所述多个数据中心的每一个并被配置为从所述多个数据中心接收组块的子集的云接口。

[0019] 在各种替代实施例中，所述文件编码器被配置为使用纠删码来生成所述多个组块。

[0020] 在各种替代实施例中，子集中组块的数目至少比由所述文件编码器生成的多个组块的数目小两个。

[0021] 显而易见的是，在这种方式下，不同的示例性实施例实现了云计算方法及文件系统。特别是，通过分发文件块并在数据中心提供文件高速缓存，高弹性和低延时的目标能够以低成本得到满足。

附图说明

[0022] 为了更好地理解各种示例性实施例中，参考了附图，其中：

[0023] 图 1 示出典型的用于存储文件作为云服务云环境；

[0024] 图 2 示出为云服务提供文件存储的示例性数据中心；

[0025] 图 3 示出用于存储文件块的示例性数据结构；

[0026] 图 4 示出在云服务中存储文件块的示例性的方法流程图；

[0027] 图 5 示出在云服务中用于读取存储的文件块的示例性方法流程图；

[0028] 图 6 示出在云服务中用于写入存储的文件块的示例性方法流程图；以及

[0029] 图 7 示出调整高速缓存的尺寸的示例性方法的流程图。

具体实施方式

[0030] 现在参照附图公开各种示例性实施例的公开的广泛方面，在附图中相同的标号表示同样的部件或步骤。

[0031] 图 1 示出了用于存储文件作为云服务的示例性的云环境 100。文件可以被存储为一个或多个文件块。因此，一个文件块可以是一个文件或整个文件的一部分。示例性的云环境 100 可以包括用户设备 110、网络 120、和云服务 130。云服务 130 可包括多个数据中心 135，用户设备 110 可以与云服务 130 的一个或多个数据中心 135 通过网络 120 进行通信。该云服务 130 可以向用户设备 110 提供数据存储和其他云服务。

[0032] 用户设备 110 可以包括能够通过网络 120 与云服务 130 通信的任何设备。例如，用户设备 110 可以是个人计算机、笔记本电脑、移动电话、智能手机、服务器、个人数字助理、或任何其它电子装置。多个用户设备 110a 和 110b 可访问云服务 130，虽然只有两个用户设备 110 被示出，但应该明白，任何数目的用户设备 110 都可以访问云服务 130。

[0033] 网络 120 可以包括能够处理用户设备 110 和数据中心 135 之间的数字通信的任何通信网络，网络 120 可以是因特网。网络 120 可以提供用户设备 110 和数据中心 135 之间的各种通信路径。

[0034] 云服务 130 可以包括一个或多个向用户设备 110 提供计算服务的计算设备。在各种示例性实施例中，计算设备可以是存储数据文件的数据中心 135。数据中心 135 可以在地理上分布以帮助确保弹性。如果一个数据中心，例如数据中心 135a，由于电力或网络故障不可用，则其他数据中心 135b-f 可以保持可用。数据中心 135 可以相互通信。在各种示例性实施例中，数据中心 135 可以通过专用或租用线路互相通信。可选地，数据中心 135 可以经由网络 120 相互通信。

[0035] 数据中心 135 可以用分布式架构存储数据文件以提供弹性。数据文件可以被划分为可由用户设备 110 请求和访问的一个或多个文件块。例如，文件块可以由一系列的数据文件中的字节定义。文件块可以被划分成多个组块，并存储在多个数据中心 135 的每一个中。可能需要组块中的一个子集来重建文件块。因此，云服务 130 可以能够提供到一个文件块的访问，即使数据中心 135 的一个或多个是不可用的。

[0036] 数据中心 135 的一个可被指定为数据文件的主数据中心。主数据中心可以被选择为在地理上最接近于最初存储的数据文件的用户设备 110 的数据中心 135。例如，数据中心 135a 可能是用户设备 110a 的主数据中心，而数据中心 135f 可能是用户设备 110b 的主数据中心。主数据中心可以包括临时存储文件块的高速缓冲器。高速缓存的文件块可以提供更快的访问，并减少延迟。高速缓存的文件块也可以减少数据中心 135 之间必须发送的数据量。如果用户设备 110 或者数据文件的主数据中心不可用时，用户设备 110 可以从任何其他数据中心检索文件块 135。

[0037] 图 2 示出为云服务提供的文件存储示例性数据中心 135。示例性数据中心 135 可以是云服务 130 的一部分。数据中心 135 可以对一些数据文件充当主数据中心，并对其它的文件充当副数据中心。示例性数据中心 135 可以包括客户端界面 210、云接口 220、文件编码器 230、组块存储器 240、文件解码器 250、文件高速缓存 260 和高速缓存适配器 270。

[0038] 客户端接口 210 可以是包括硬件和 / 或编码在机器可读存储介质上的可执行指令的接口，被配置成与用户设备 110 进行通信。客户端接口 210 可以从用户设备 110 接收文件块并启动存储文件块过程。客户端接口 210 可以从用户设备 110 接收文件块请求，并启

动读取文件块的过程。客户端接口 210 可以将完整的文件块或完整的数据文件发送到用户设备 110。

[0039] 云接口 220 是包括硬件和 / 或编码在机器可读存储介质上的可执行指令的接口，被配置成与数据中心 135 通信。云接口 220 可以分发文件块的编码组块到一个或多个其它数据中心 135。在各种示例性实施例中，云接口 220 分发一个组块至多个数据中心 135 的每一个。云接口 220 可以从一个或多个其他数据中心 135 接收文件块的编码组块。云接口 220 可以访问块存储 240 来读取或存储组块。

[0040] 文件编码器 230 可以包括硬件和 / 或编码在机器可读存储介质上的可执行指令，被配置成将文件块编码为多个组块。如下面将进一步关于图 3 至图 7 要详细讨论的，多个组块可以提供用于存储所述文件块的弹性分布格式。在各个示例性实施例中，文件编码器 230 可实施纠删码，用于生成所述多个组块。适于产生多个组块的示例性纠删码可以包括里德所罗门码、最大距离可分离 (MDS) 码和低密度奇偶校验 (LDPC) 码。在各种示例性实施例中，文件编码器 230 可使用系统纠删码，其中原始文件块可以沿用来进行文件恢复的单独的编码组块集被划分成多个未编码组块。在各种替代实施例中，其他编码方案可以被用来生成组块。

[0041] 组块存储器 240 可以包括能存储由例如文件编码器 230 的文件编码器生成的组块的任何机器可读介质。因此，组块存储器 240 可以包括机器可读存储介质，例如随机存取存储器 (RAM)、磁盘存储介质、光存储介质、闪存设备、和 / 或类似的存储介质。在各种示例性实施例中，组块存储器 240 可以提供永久存储，它维护电力或设备故障情况下的存储块。组块存储器 240 可使用日志系统，以保持写操作过程中发生故障时的完整。组块存储器 240 可以存储由文件编码器生成的组块 230 和 / 或通过云接口 220 接收到的组块。

[0042] 文件解码器 250 可以包括硬件和 / 或编码在机器可读存储介质上的可执行指令，被配置成对用于存储文件的多个块的子集进行解码。文件解码器 250 可以从组块存储器 240 和 / 或云接口 220 接收组块的子集。文件解码器 250 可以从块的子集再生文件块。在各种示例性实施例中，文件解码器 250 可以实现文件编码器 230 的逆操作。文件解码器 250 可以使用和文件编码器 230 相同的纠删码。将进一步参考图 3-7 进行详细说明的，再生文件块所需的组块的子集可以比由文件编码器 230 生成的多个组块小。

[0043] 文件高速缓存 260 可以包括能够存储完整文件块的任何机器可读介质。因此，文件高速缓存 260 可以包括机器可读存储介质，例如随机存取存储器 (RAM)、磁盘存储介质、光存储介质、闪存设备、和 / 或类似的存储介质。在各种示例性实施例中，文件高速缓存 260 可提供持久存储用于维护电力或设备故障情况下的文件块。例如，文件高速缓存 260 可以是硬盘。使用硬盘作为文件高速缓存 260，可以使成本最小化同时提供可接受的延迟。文件高速缓存 260 可以使用日志系统以在写操作过程中发生故障时保持完整。文件高速缓存 260 可以存储通过客户端接口 210 接收到的文件块和 / 或文件解码器 250 再生的文件块。

[0044] 文件高速缓存 260 可能具有由物理容量和 / 或缓存适配器 270 确定的有限尺寸。文件高速缓存 260 可以包括高速缓存管理器，用于确定哪些文件块存储在文件高速缓存 260 中。高速缓存管理器可以使用最近最少使用的 (LRU) 高速缓存替换方案。因此，文件高速缓存 260 可以包括那些最近已经由客户端设备访问的文件块。

[0045] 高速缓存适配器 270 可以包括硬件和 / 或编码在机器可读存储介质上的可执行指

令，被配置为调整文件高速缓存 260 的尺寸。缓存适配器 270 可以测量包括文件访问请求数量及频率的云服务 130 的使用情况。高速缓存适配器 270 可以尝试通过调整文件高速缓存 260 的大小，减少云服务 130 的成本。缓存适配器 270 可以考虑存储成本、加工成本和云服务 130 的传输成本。更大的高速缓存可以提高存储成本，同时减少传输和处理成本。较小的高速缓存可以提高传输和处理成本，同时降低存储成本。调整文件高速缓存 260 的尺寸的示例性方法将关于图 7 进行说明。

[0046] 图 3 示出用于存储文件块的示例性数据结构 300。数据结构 300 可包括存储在例如文件高速缓存 260 高速缓存中的文件块 310 以及存储在数据中心 135a-f 的组块存储器 240 中的组块 320a-f。数据结构 300 可以说明如何纠删码可以被用来在云环境提供文件块的弹性分布式存储。所示的示例性数据结构 300 可被用来存储和恢复文件块，即使数据中心的两个或多个不可用。应当认识到，数据结构 300 可以是纠删码的简化。已知的纠删码可用于提供更高效的存储、更佳的弹性、和 / 或更短的延迟。

[0047] 文件块 310 可以被划分成多个段 :A、B、C、D 和 E。组块 320a-f 各自可以包括两个段。例如，块 320a-f 可以分别包括段组合 {A，B}、{C，D}、{E，A}、{B，C}、{D，E} 和 {B，D)。组块 320a-f 的每一个可以被存储在单独的数据中心 135aa-f。文件块 310 可以从块 320a-f 的任意四个再生。因此，即使组块中的两个都不可用，文件块 310 也可以再生。在某些情况下，文件块 310 可以从仅仅三个组块再生。组块 320a-f 可以要求储存总共十二段。相比之下，文件块 310 可以存储在三个数据中心 135 上来提供两个数据中心故障的弹性，但是文件块 310 的三个副本需要存储十五个段。

[0048] 图 4 示出了存储文件块的示例性方法 400 的流程图。方法 400 可以通过数据中心 135 的各种组件来执行。方法 400 可开始于步骤 410 并继续执行步骤 420。

[0049] 在步骤 420，数据中心 135 可以从客户端设备 110 接收存储文件块的请求。数据中心 135 可以确定它是否为该客户端设备的主数据中心。如果数据中心 135 不是主数据中心，数据中心 135 可以将请求转发到主数据中心。可选地，数据中心 135 可以作为副数据中心处理该请求。如果主数据中心不可用，数据中心 135 也可以处理该请求。然后该方法可以前进到步骤 430。

[0050] 在步骤 430，数据中心 135 可以在文件高速缓存 260 中存储所接收到的文件块。如果文件高速缓存 260 是满的，数据中心 135 可以用接收到的文件块替换存储在高速缓存中的文件块。步骤 430 可以是可选的。在各种替代实施例中，数据中心 135 可能不会立即在文件高速缓存 260 中存储接收到的文件块。如果数据中心 135 是副数据中心，数据中心 135 可以跳过步骤 430。则该方法可以前进到步骤 440。

[0051] 在步骤 440，数据中心 135 可以从接收的文件块生成组块。数据中心 135 可使用纠删码来生成组块。在各种示例性实施例中，数据中心 135 可以为包括数据中心 135 的每个可用数据中心生成一个组块。在各种替代实施例中，可以生成任意数目的块。在步骤 450，数据中心 135 可以分发组块至其他数据中心用于存储。在各种示例性实施例中，一个组块可以被分发给包括主数据中心 135 的每个数据中心。在各种替代实施例中，多个组块可被分发给一个数据中心，并且复制组块可以被分发到多个数据中心。所述多个组块中的至少第一组块和第二组块可以分发到在多个数据中心的不同的数据中心。一旦组块已经分发，文件块可弹性地存储在云服务 130 中，并且该方法可以前进到步骤 460。在步骤 460，方法

400 可结束。

[0052] 图 5 示出用于读取存储在云服务中的文件块的示例性方法 500 的流程图。方法 500 可以通过数据中心 135 的各种组件来执行。方法 500 可开始于步骤 510 并继续执行步骤 520。

[0053] 在步骤 520，数据中心 135 可以接收读取文件块的请求。数据中心 135 可以确定它是否为客户端设备的主数据中心。如果数据中心 135 不是主数据中心，数据中心 135 可以将请求转发到主数据中心。可选地，数据中心 135 可以作为副数据中心处理该请求。如果主数据中心不可用，数据中心 135 也可以处理该请求。然后该方法可以前进到步骤 530。

[0054] 在步骤 530，数据中心 135 可以确定所请求的文件块是否被存储在该文件高速缓存 260 中。数据中心 135 可以确定对应于文件块的请求字节范围是否被存储在该文件高速缓存 260。如果数据中心 135 不是主数据中心，所请求的文件块可能无法存储在文件高速缓存 260。即使数据中心 135 是主数据中心，因为它最近没有被访问并已被替换，所请求的文件块也可能无法存储在文件高速缓存 260 中。如果所请求的文件块被存储在文件高速缓冲存储器 260 中，该方法可以直接进行到步骤 580，如果所请求的文件块不存储在文件高速缓冲存储器 260 中，该方法可以前进到步骤 540。

[0055] 在步骤 540，数据中心 135 可以从其他数据中心请求组块。在步骤 550，数据中心 135 可以从其他数据中心的一个或多个接收所请求的组块。数据中心 135 可以不从其他数据中心的一个或多个接收请求的组块。例如，其他数据中心可能不可用或可能没有检索到所请求的组块。在任何情况下，当数据中心 135 接收到组块的子集，方法 500 可以前进到步骤 560。

[0056] 在步骤 560，数据中心 135 可以从接收到组块子集再生请求的文件块。数据中心 135 可以根据用于生成组块的纠删码再生文件块。在使用系统纠删码的各种示例性实施例中，文件块可从未编码的组块再生而不对编码组块进行解码。组块接收和基于组块的文件块再生可能会消耗数据中心 135 的处理能力。再生请求的文件块以满足客户端设备 110 的请求所花费的时间也可能会增加延迟。一旦完整的文件块已经被重建，方法 500 可以前进到步骤 570。

[0057] 在步骤 570，数据中心 135 可以在文件高速缓存 260 存储完整的文件块。如果文件高速缓存 260 已满，数据中心 135 可以用再生的文件在文件高速缓存 260 中更换一个或多个现有文件块。使得文件块存储在文件高速缓存 260 中可以让数据中心 135 更迅速地完成涉及该文件块的后续请求。如果数据中心 135 是数据文件的副数据中心，数据中心 135 可以转发完整的文件块至主数据中心用于在主数据中心的文件高速缓存中存储。然后方法 500 可以进入步骤 580。

[0058] 在步骤 580，数据中心 135 可以将文件块发送给发出请求的客户端设备 110。客户端设备 110 可以接收所请求的文件块。采用存储在主数据中心文件的高速缓存中的缓存副本，文件块可以保持弹性存储在云服务 130 中。接着，方法 500 可以前进到步骤 590，其中方法 500 结束。

[0059] 图 6 示出用于写入存储在云服务 130 中的文件块的示例性方法 600 的流程图。方法 600 可以由数据中心 135 的各种组件来执行。方法 600 可开始于步骤 605 并继续执行步骤 610。

[0060] 在步骤610，数据中心135可以从客户端设备110接收写入存储在云服务130中的文件块的请求。这个写请求可以包括文件块的一部分修改而留下其他部分文件不变。数据中心135可以确定它是否为客户端设备110或该文件的主数据中心。如果数据中心135不是主数据中心，数据中心135可以将请求转发到主数据中心。可选地，数据中心135可以作为副数据中心处理该请求。如果主数据中心不可用，那么数据中心135也可以处理该请求。数据中心135可使用日志，以防止在写入过程中的文件损坏。写请求可能包括从潜在的写入错误中恢复。然后方法600可以进入步骤615。

[0061] 在步骤615，数据中心135可以确定文件块是否被存储在文件高速缓存260中。如果文件块最近被读取访问，则它可以被存储在文件高速缓存260中。这样也许是可能的：写请求的文件块将被存储在该文件高速缓存260中，因为客户端设备110总是在修改文件块之前读它然后发送写请求。然而，如果许多文件正被访问，在写请求到达之前，可从该文件高速缓冲存储器260中移除文件块。如果该文件块当前被存储在文件高速缓冲存储器260中，方法600可以前进到步骤640。如果该文件块当前未存储在文件高速缓冲存储器260，方法600可以前进到步骤620。

[0062] 在步骤620中，数据中心135可以从其他数据中心请求组块。如果写请求只影响组块的子集，数据中心135可以仅请求受影响的组块。接着，方法600可以前进到步骤625，其中数据中心135可以接收所请求的组块。例如，如果其他数据中心是由于停电无法使用，数据中心135可以不从其他数据中心接收组块。一旦已经接收到组块的子集，方法600可以前进到步骤630。

[0063] 在步骤630，数据中心135可以从接收到的组块子集再生所请求的文件。数据中心135可以根据用于生成组块的纠删码来再生文件块。组块接收和基于组块的文件块再生可能会消耗数据中心135的处理能力。再生请求的文件块以满足客户端设备110的请求所花费的时间也可能会增加延迟。一旦完整的文件块已经被重建，方法600可以前进到步骤635。

[0064] 在步骤635，数据中心135可以在文件高速缓存260中存储完整的文件块。如果文件高速缓存已满，数据中心135可以用再生的文件在文件高速缓存260中更换一个或多个现有文件块。使得文件块存储在文件高速缓存260中可以让数据中心135更迅速地完成涉及该文件块的后续请求。如果数据中心135是数据文件的副数据中心，数据中心135可以转发完整的文件块至主数据中心用于在主数据中心的文件高速缓存中存储。然后方法600可以进入步骤640。

[0065] 在步骤640，数据中心135可以通过按写请求所要求的写入文件块来更新存储的文件块。写入文件块时，数据中心135可能会打开该文件。写请求可能会修改或替换文件块的任何部分或整个文件。数据中心135可以修改存储在文件高速缓冲存储器260中的文件块的副本。一旦数据中心135已经处理写请求和更新文件块，方法600可以前进到步骤645，在数据中心135可以关闭该文件。关闭文件可以阻止该文件的进一步修改。关闭文件时，数据中心135还可确定如文件尺寸、修改日期、作者等的属性。然后方法600可以进入步骤650。

[0066] 在步骤650，数据中心135可以根据纠删码从更新后的文件块生成组块。在各种示例性实施例中，可以仅仅基于文件块的修改的部分来生成生成组块。一些组块可通过修改

保持不变。在各种替代实施方式中，步骤650可能被延迟，直到该文件块要在文件高速缓存260中被替换。接着，方法600可以前进到步骤655。

[0067] 在步骤655，数据中心135可以分发修改的组块至其他数据中心。数据中心135可以只分发已修改的组块。如果存储在另一个数据中心的组块没有被修改过，数据中心135可以节省时间和通信成本。一旦修改的组块已被分发，修改后的文件块可被弹性地存储在云服务130中，并且方法600可以前进到步骤660，其中方法600结束。

[0068] 图7示出调整高速缓存尺寸的示例性方法700的流程图。方法700可由数据中心135或如客户端设备110的其它计算机使用，以减少在云服务130中存储多个文件的成本。方法700可以尝试通过调整高速缓存尺寸以有效地存储文件和处理请求来最小化云服务130的成本。方法700可以测量实际成本和假设成本然后向着更经济的方向调整高速缓存尺寸。方法700可以重复地执行，以不断地调整高速缓存尺寸。方法700可以开始于步骤705，并继续执行步骤710。

[0069] 在步骤710，数据中心135或客户端设备110可以在时间间隔内测量云服务130的实际成本。云服务130的成本可以通过各种函数进行测量。成本函数可取决于各种参数，如，数据存储量、高速缓存大小、请求数量、内部数据中心传送量、以及数据中心的处理量。例如，如果方法700是由云服务的客户进行的，成本可以由云服务提供商收取的费用进行测量。作为另一个例子，如果方法700是由云服务提供商执行，提供商可以评估所云服务130使用的每个系统资源的值。在各种示例性实施例中，高速缓存的大小可以用数据存储量加权平均。成本可以按每个客户、每个数据中心和/或服务范围的基础上确定。也可以使用任何时间间隔。等于一个测量计费周期的时间间隔可以是合适的。例如，如果按照每天的数据存储量对客户收费，一天的时间间隔可能是适当的。然后方法700可以进入步骤715。

[0070] 在步骤715，数据中心135或客户端设备110可以确定云服务130的假定成本。假定成本可以基于不同的高速缓存尺寸。在各种示例性实施例中，假定成本可以基于先前的高速缓存尺寸。用来测量假定成本的函数可以与可以用来测量实际成本的函数相同。因此，步骤715可以类似于步骤710。步骤710和715可能以任何顺序发生。然后该方法可以前进到步骤720。

[0071] 在步骤720，数据中心135或客户端设备110可以确定在步骤710中测得的实际成本是否大于在步骤715中确定的假定成本。如果实际成本大于假定成本，该方法可继续执行步骤730。如果假定成本大于实际成本，则该方法可以前进到步骤725。

[0072] 在步骤725，数据中心135或客户端设备110可以判断当前高速缓存尺寸是否大于用于确定假定成本的老高速缓存尺寸。如果当前高速缓存尺寸较大，方法700可以前进到步骤735。如果老的高速缓存尺寸较大，则该方法可以前进到步骤740。

[0073] 在步骤730，数据中心135或客户端设备110可以判断当前高速缓存尺寸是否大于用于确定假定成本的老高速缓存尺寸。如果当前高速缓存尺寸较大，方法700可以前进到步骤745，如果老高速高速缓存尺寸较大，方法700可以前进到步骤735。换句话说，步骤730可以类似于步骤725，但具有相反的结果。

[0074] 在步骤735，数据中心135可能会增加高速缓存尺寸。数据中心135可能会针对特定客户增加高速缓存尺寸或针对所有客户增加高速缓存尺寸。数据中心135也可能指示

其他数据中心应该增加高速缓存的尺寸。增加的尺寸可以变化。在各种示例性实施例中，高速缓存的尺寸可以按每一千兆字节提高。接着，方法 700 可以前进到步骤 750，其中方法 700 结束。

[0075] 步骤 740 和 745 可以是相同的。在步骤 740 和 / 或 745 中，数据中心 135 可降低高速缓存的尺寸。数据中心 135 可以针对特定客户减小高速缓存尺寸或针对所有客户减小高速缓存尺寸。数据中心 135 也可能指示其他数据中心应该减少高速缓存的尺寸。减少的尺寸可能会有所不同。在各种示例性实施例中，高速缓存的尺寸可以按每一千兆减少。接着，方法 700 可以前进到步骤 750，其中方法 700 结束。

[0076] 根据上述内容，各种示例性实施例提供了一种用于云计算的方法及文件系统。特别是，通过分发文件组块，并在数据中心提供文件高速缓存，高弹性和低延时的目标能够以低成本得到满足。

[0077] 根据上面的描述，显而易见的，本发明的各种示例性实施例可以在硬件和 / 或固件中实现。此外，各种示范性实施例可以被实现为存储在机器可读存储介质上的指令，由至少一个处理器读取和执行所述指令以实施在本文中详细描述的操作。机器可读存储介质可以包括存储机器可读的形式的信息的任何机制，这里的机器诸如个人电脑或笔记本电脑、服务器、或其它计算设备。因此，机器可读存储介质可以包括只读存储器 (ROM)、随机存取存储器 (RAM)、磁盘存储介质、光存储介质、闪存设备、和类似的存储介质。

[0078] 本领域技术人员应当理解，这里的任何框图所代表的只是体现本发明的原理的概念性电路。类似地，可以理解，任何流程图表、流程图、状态转移图、伪代码等表示各种处理，这些处理本质上可以在计算机可读介质中表示并因此由计算机或处理器执行，无论这样的计算机或处理器是否被明确示出。

[0079] 虽然在各种示例性实施例中参考某些示例性方面进行了详细描述，但应当理解，本发明能够具有其他实施例，并且其细节能够在各种明显的方面进行修改。由于本领域技术人员了解了本发明，可以在本发明的精神和范围内受到进行变化和修改。因此，之前的公开、说明和附图仅用于说明的目的，并且不以任何方式限制本发明，本发明仅由权利要求限定。

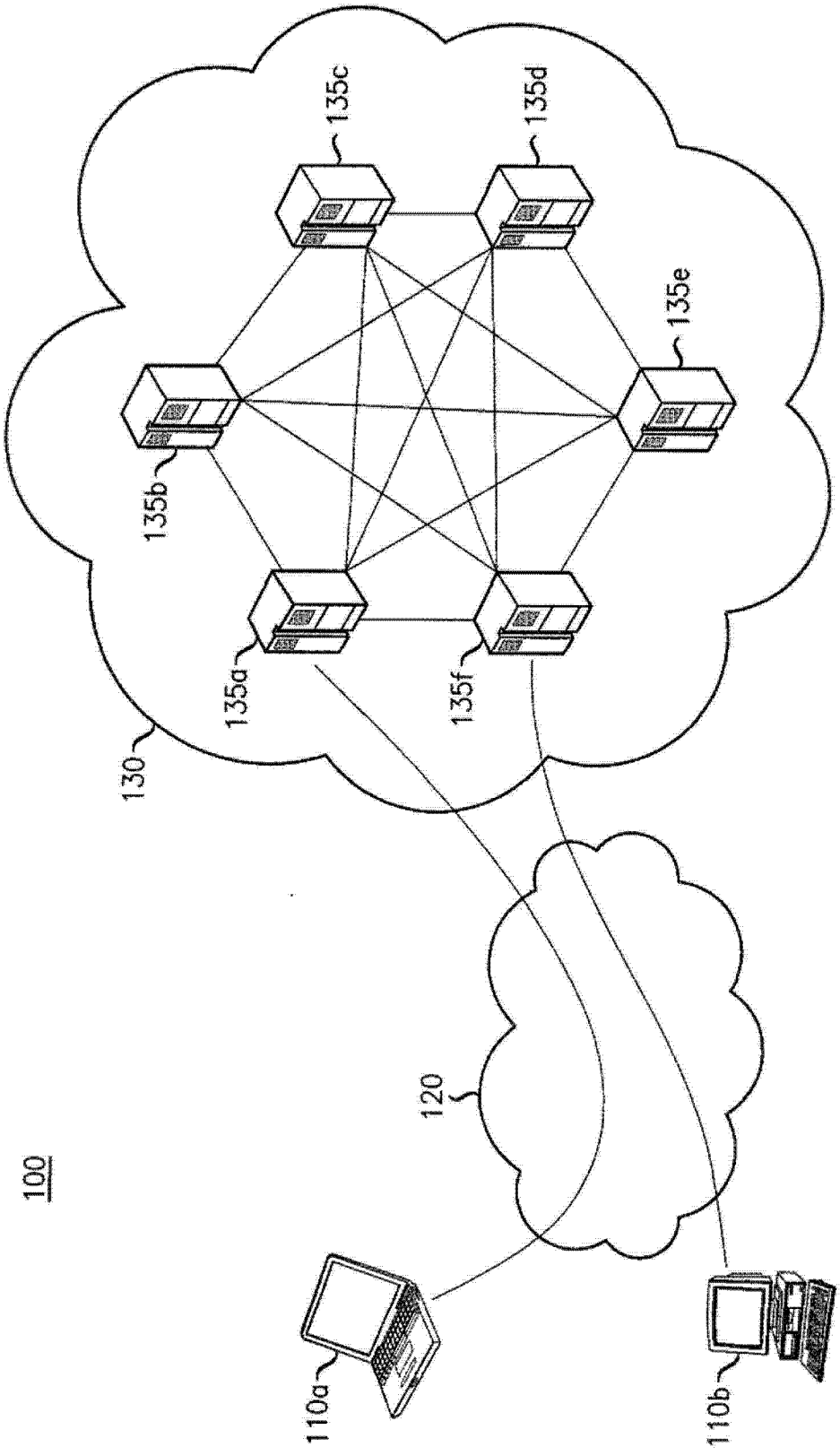


图 1

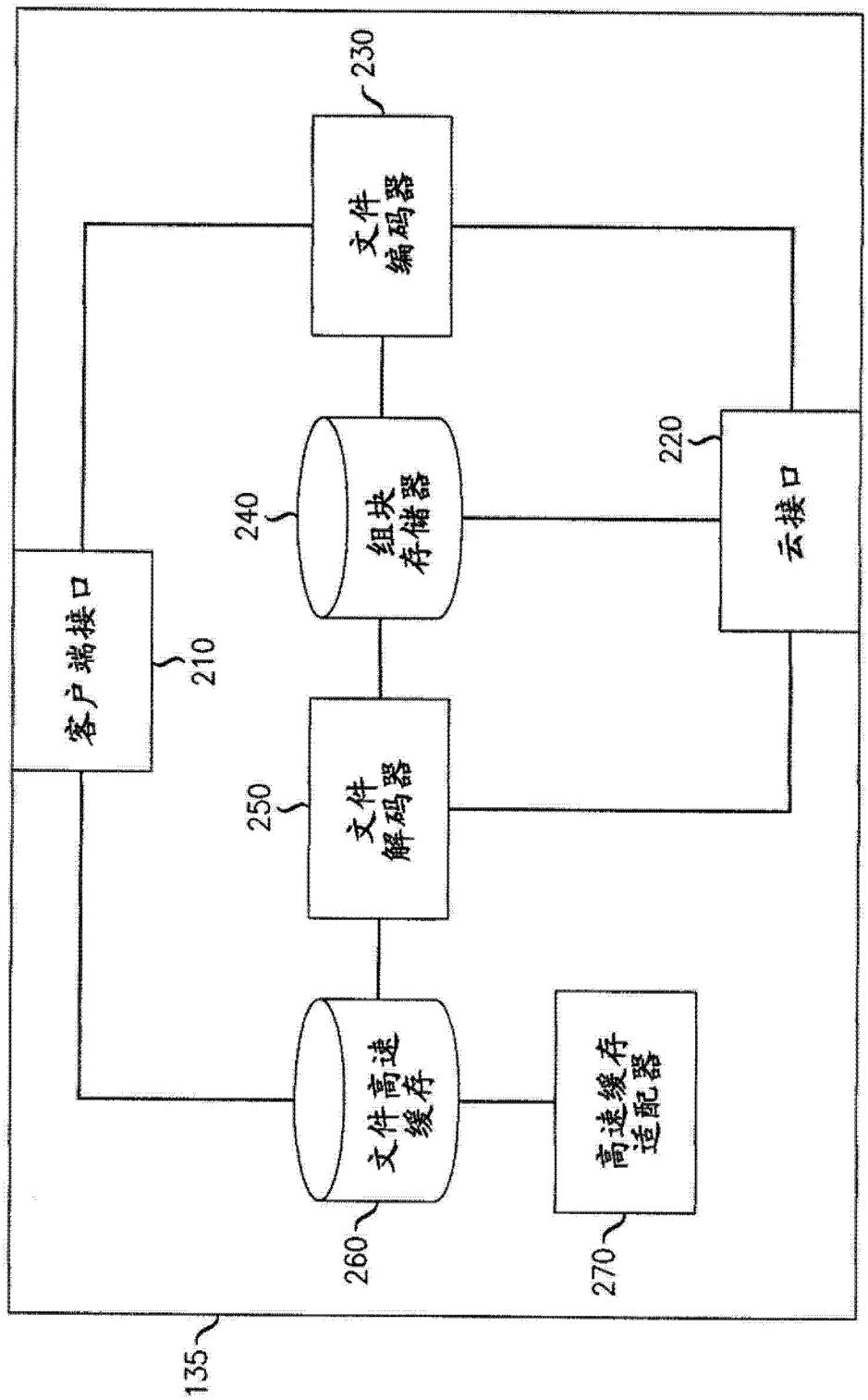

230
文件编码器
240
组块存储器
220
云接口
客户端接口
210
250
文件解码器
260
文件高速缓存
高速缓存适配器
270
135

图 2

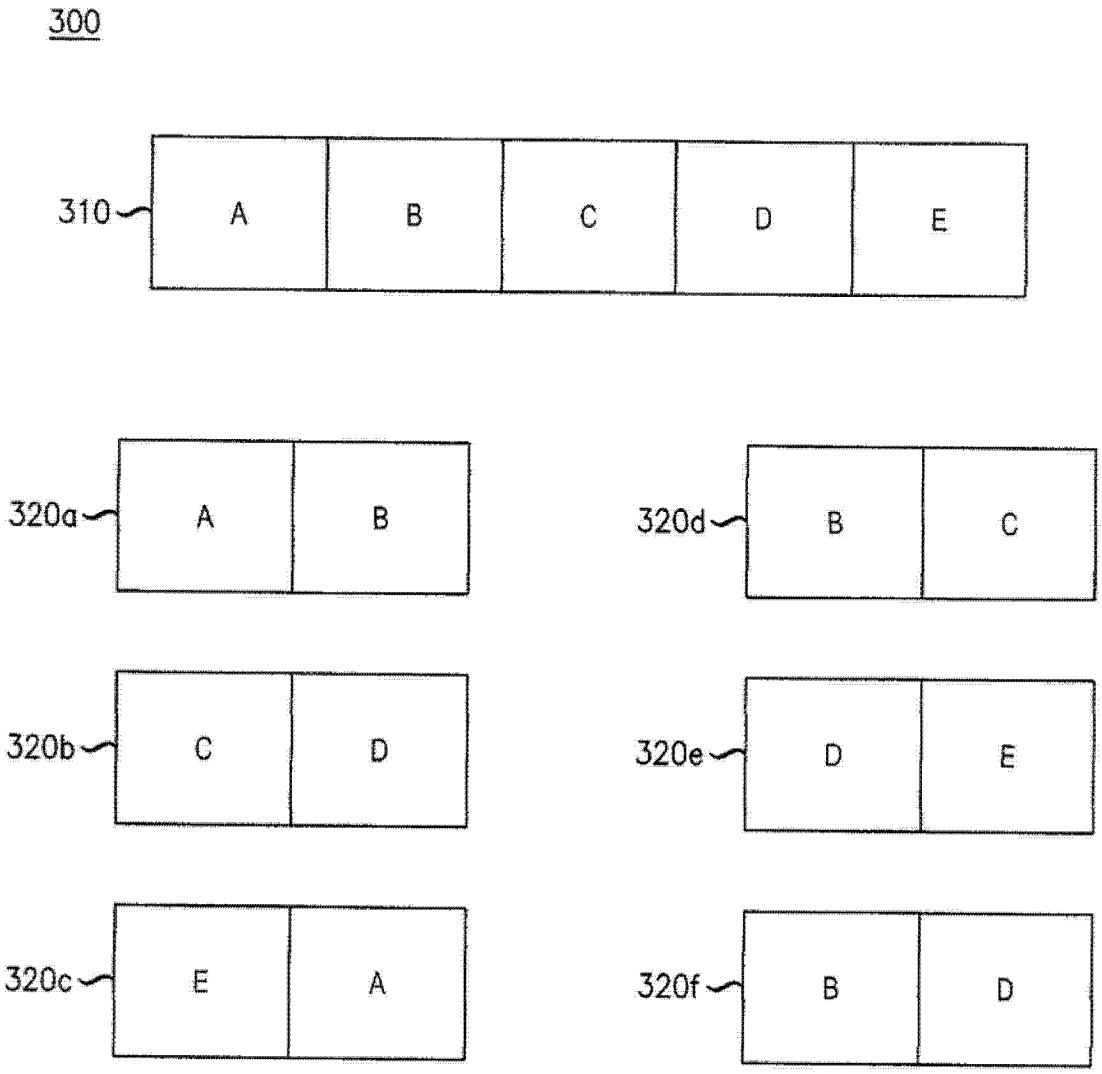


图 3

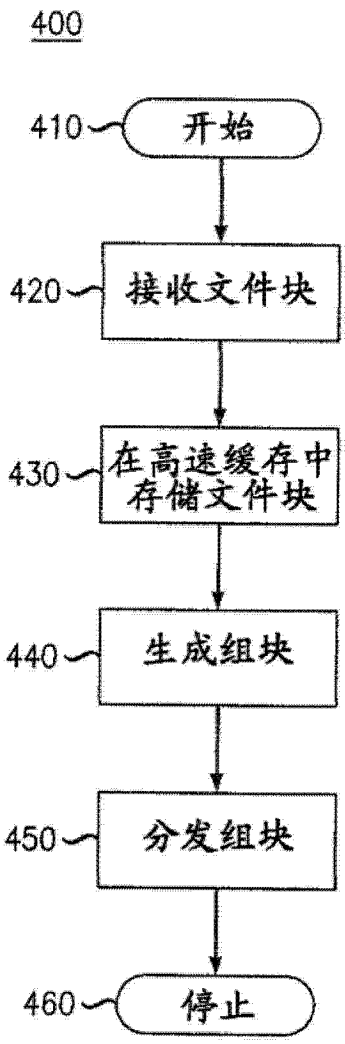

400
410
开始
420
接收文件块
430
在高速缓存中存储文件块
440
生成组块
450
分发组块
460
停止

图 4

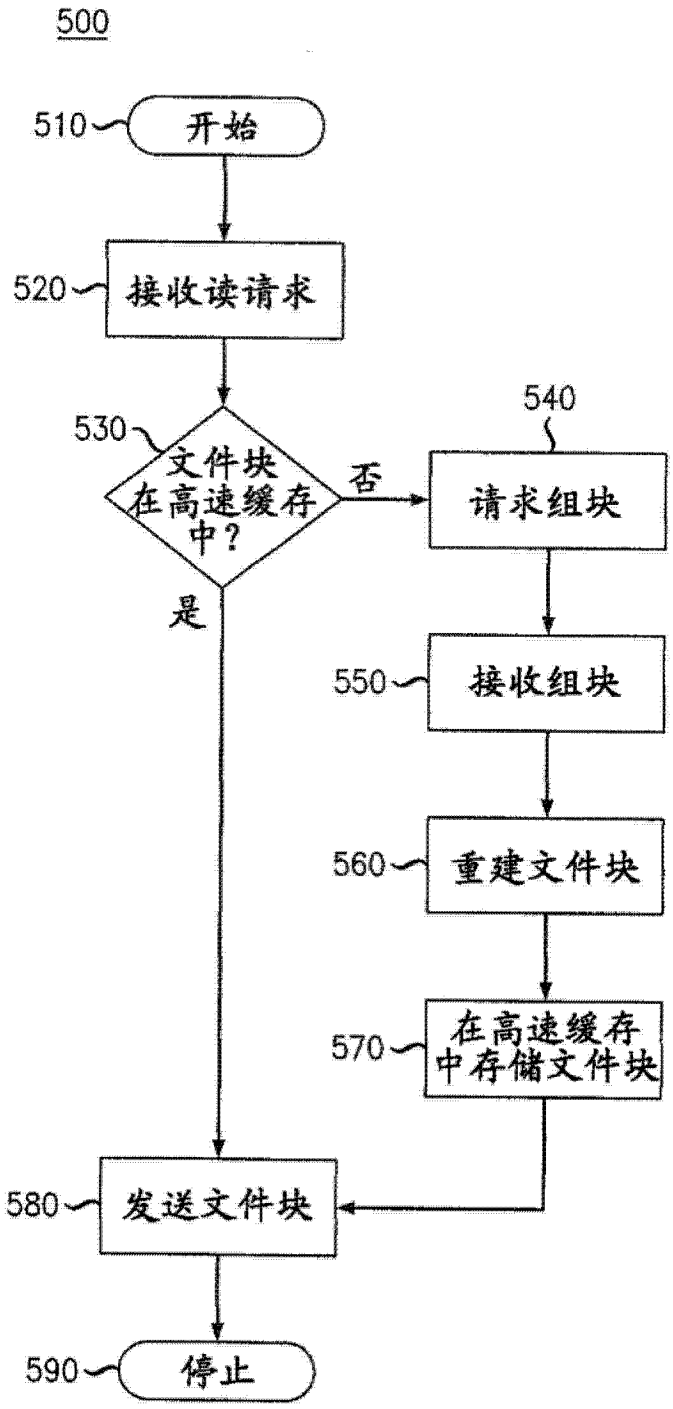

500
510
开始
520
接收读请求
530
文件块在高速缓存中？
否
是
540
请求组块
550
接收组块
560
重建文件块
570
在高速缓存中存储文件块
580
发送文件块
590
停止

图 5

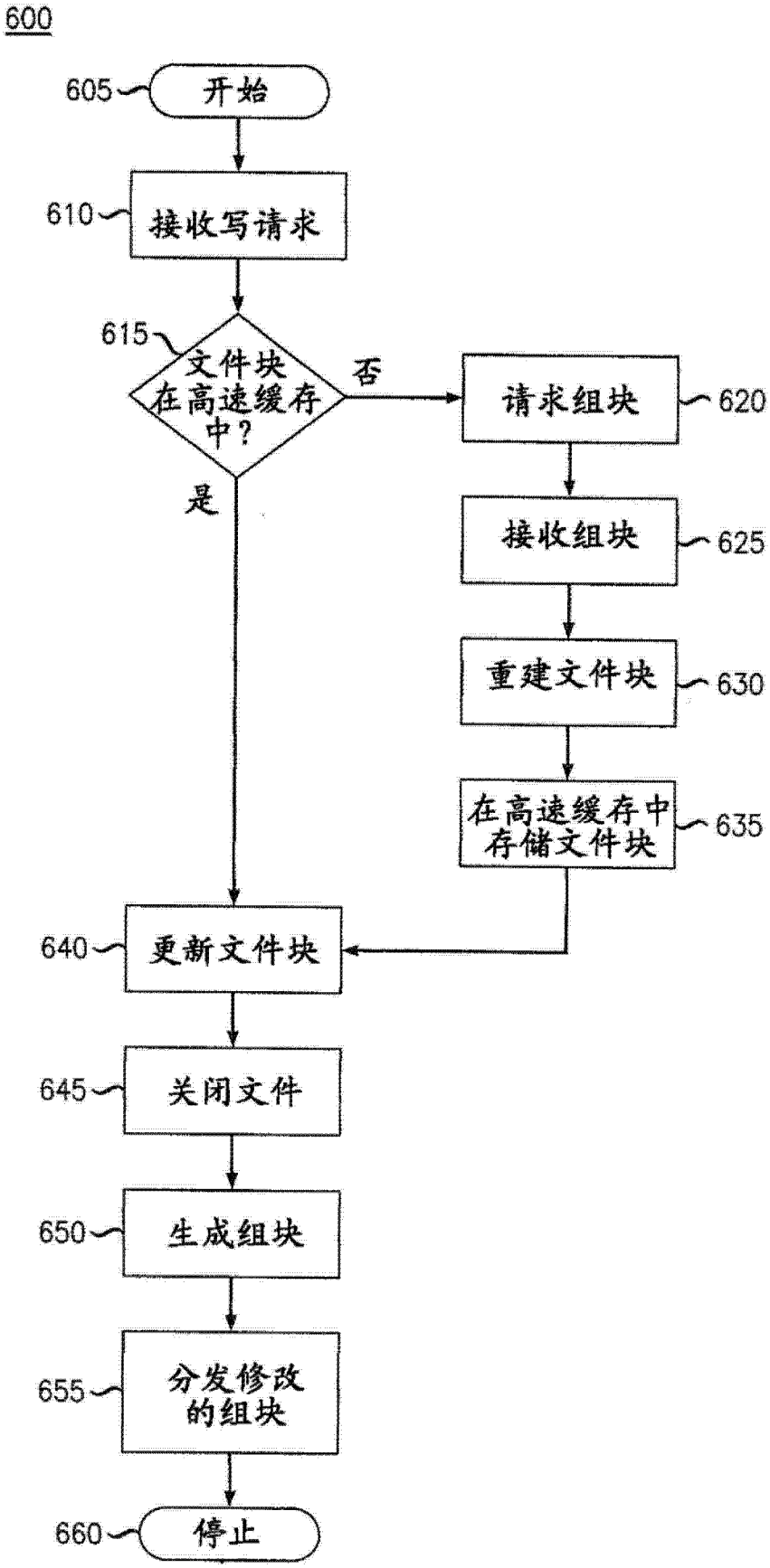

600
605
开始
610
接收写请求
615
文件块在高速缓存中?
否
是
620
请求组块
625
接收组块
630
重建文件块
635
在高速缓存中存储文件块
640
更新文件块
645
关闭文件
650
生成组块
655
分发修改的组块
660
停止

图 6

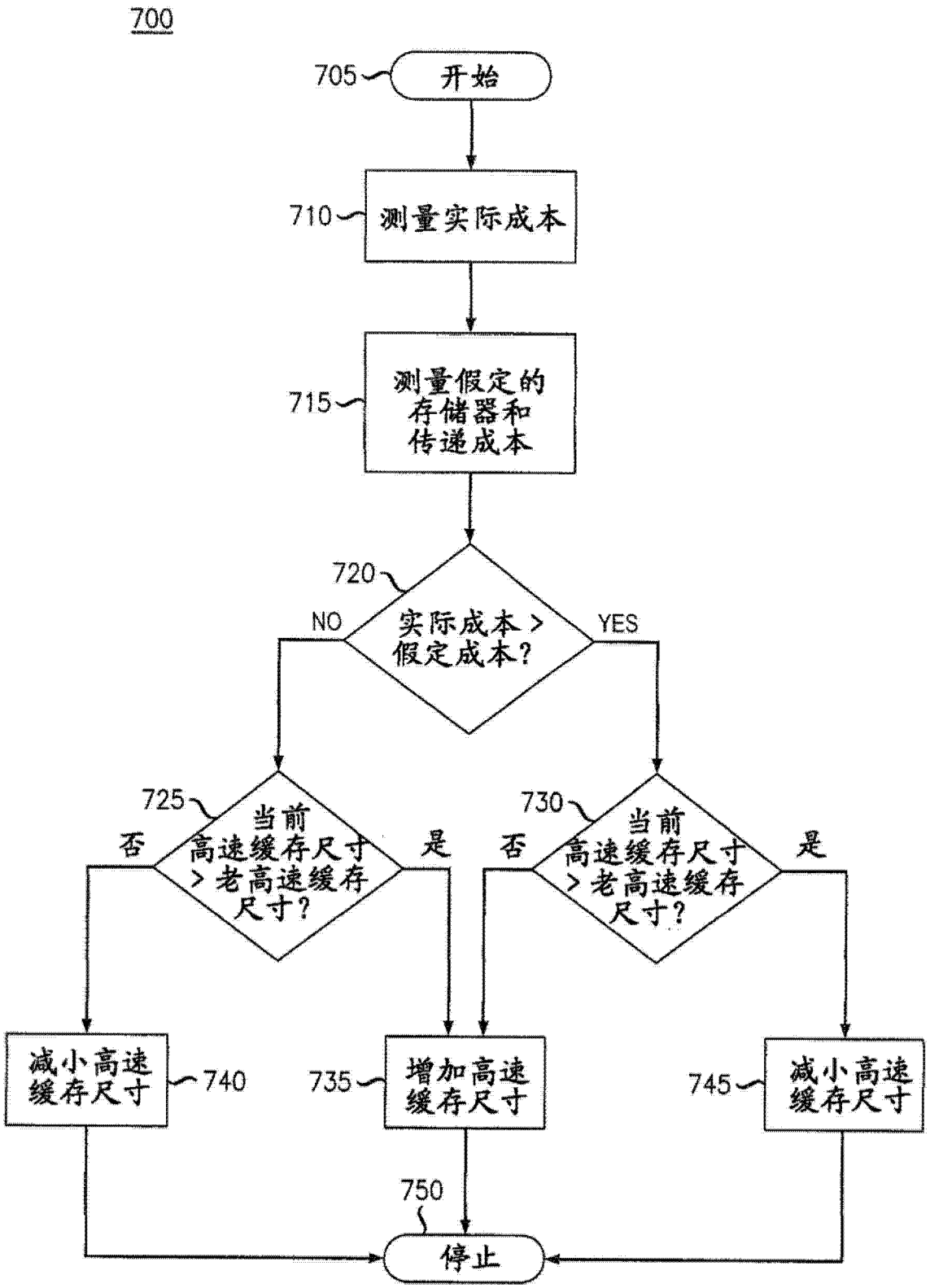

700
705
开始
710
测量实际成本
715
测量假定的存储器和传递成本
720
实际成本 > 假定成本?
NO
YES
725
当前高速缓存尺寸 > 老高速缓存尺寸?
否
是
730
当前高速缓存尺寸 > 老高速缓存尺寸?
否
是
减小高速缓存尺寸
740
735
增加高速缓存尺寸
745
减小高速缓存尺寸
750
停止

图 7