(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 101535964 B
(45)授权公告日 2013.06.12

(21)申请号 200680025690.X
(22)申请日 2006.07.10
(30)优先权数据
11/181,063 2005.07.14 US
(85)PCT申请进入国家阶段日
2008.01.14
(86)PCT申请的申请数据
PCT/US2006/026926 2006.07.10
(87)PCT申请的公布数据
WO2007/011585 EN 2007.01.25
(73)专利权人 微软公司
地址 美国华盛顿州
(72)发明人 C·F·罗伯特 G·瑟里亚纳拉亚南 S·C·哈弗瓦拉 A·H·莫汉德
(74)专利代理机构 上海专利商标事务所有限公司 31100
代理人 顾嘉运
(51)Int.Cl.
G06F 12/00(2006.01)
(56)对比文件
US 2004/0172423 A1,2004.09.02,说明书第7-10、27-29、47-51段,图1-4.
CN 1542637 A,2004.11.03,全文.
审查员 赵晓敏

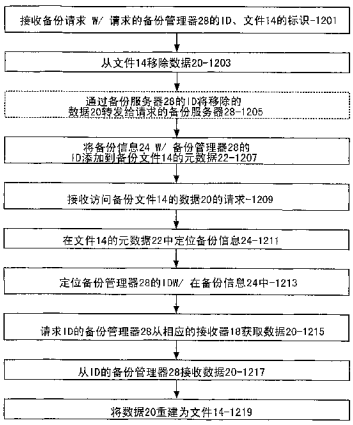
权利要求书3页 说明书18页 附图13页

(54)发明名称

备份文件

(57)摘要

为了重建经备份的文件以供使用,计算设备上的单个备份过滤器定位该经备份的文件的元数据中的备份信息,并在该备份信息中定位该计算设备的特定备份管理器的标识,其中经标识的备份管理器负责经备份的文件。此后,该备份过滤器向经标识的备份管理器发送请求,以从相应的可选位置获取经备份的文件的数据,并且经标识的备份管理器实际上从这种可选位置获取这种请求数据。一旦从经标识的备份管理器接收经标识的文件的请求数据,该备份过滤器将接收到的数据重建为经备份的文件。该单个备份过滤器对于该计算设备的所有备份管理器是通用的。



CN 101535964 B

1. 一种与具有存储卷的计算设备、管理所述存储卷的文件系统以及由所述文件系统存储在所述存储卷上并由该文件系统访问的多个文件有关的方法，每个文件被定义成包括数据和与该数据相关的元数据，所述文件通过下述方式以较小的备份形式被存储在所述存储卷上：所述文件的数据的一部分已从所述文件移除并被存储在多个可选位置中的一个，以便释放所述存储卷上的空间，而所述文件是以较小的备份形式被存储的，所述计算设备还具有用于所述多个可选位置的每一个的相应的备份管理器，每个备份管理器被用于和与之对应的可选位置通信，所述方法通过使用所述计算设备上的单个备份过滤器来重建备份文件以供使用，所述单个备份过滤器执行下述步骤：在所述备份文件的元数据中定位备份信息；

在所述备份信息中定位所述计算设备的特定备份管理器的标识，所述经标识的备份管理器负责所述备份文件；

将一请求传递到所述经标识的备份管理器以获取来自相应的可选位置的所述备份文件的数据，由此所述经标识的备份管理器实际上包含来自所述可选位置的请求的数据；

从所述经标识的备份管理器接收所述备份文件的所述请求的数据；以及

将所接收的数据重建为所述备份文件，

由此单个备份过滤器对所述计算设备的所有备份管理器是通用的。

2. 如权利要求 1 所述的方法，其特征在于，包括将一请求发送给所述经标识的备份管理器以从相应的可选位置获取所述备份文件的数据的一部分。

3. 如权利要求 1 所述的方法，其特征在于，包括将一请求发送给所述经标识的备份管理器以从相应的可选位置获取所述经备份文件的所有数据。

4. 如权利要求 1 所述的方法，其特征在于，所述计算设备是分支服务器，每个可选位置是远离所述分支服务器的集线器服务器，所述集线器服务器为多个所述分支服务器提供服务。

5. 如权利要求 1 所述的方法，其特征在于，所述备份文件的元数据实质上包括在备份之前所述文件的所有元数据，也包括用于从所述可选位置检索所述文件的数据的备份信息。

6. 如权利要求 1 所述的方法，其特征在于，所述文件被定义成包括包含主要 数据和次要数据的数据，所述主要数据在大小上大于所述次要数据，至少将所述文件的主要数据从所述文件移除并将其存储在所述可选位置。

7. 一种与具有存储卷的计算设备、管理所述存储卷的文件系统以及由所述文件系统存储在所述存储卷上并由该文件系统访问的多个文件有关的方法，每个文件被定义成包括数据和与该数据相关的元数据，所述文件通过下述方式以较小的备份形式被存储在所述存储卷上：所述文件的数据的一部分已从所述文件移除并被存储在多个可选位置中的一个，以便释放所述存储卷上的空间，所述计算设备还具有用于所述多个可选位置的每一个的相应的备份管理器，每个备份管理器被用于和与之对应的可选位置通信，所述方法用于重建备份文件以供使用，并且包含：

所述备份文件的元数据中定位备份信息；

在所述备份信息中定位所述计算设备的特定备份管理器的标识，所述经标识的备份管理器负责所述备份文件；

将一请求传递到所述经标识的备份管理器以获取来自相应可选位置的所述备份文件的数据，由此所述经标识的备份管理器实际上包含来自所述可选位置的请求的数据；

从所述经标识的备份管理器接收所述备份文件的所述请求的数据；以及

将所接收到的数据重建为所述备份文件。

8. 如权利要求 7 所述的方法，其特征在于，包括将一请求发送给所述经标识的备份管理器以从相应的可选位置获取经备份的文件的数据的一部分。

9. 如权利要求 7 所述的方法，其特征在于，包括将一请求发送给所述经标识的备份管理器以从相应的可选位置获取经备份的文件的所有数据。

10. 如权利要求 7 所述的方法，其特征在于，所述计算设备是分支服务器，每个可选位置是远离所述分支服务器的集线器服务器，所述集线器服务器为多个所述分支服务器提供服务。

11. 如权利要求 7 所述的方法，其特征在于，所述备份文件的元数据实质上包括在备份之前所述文件的所有元数据，也包括用于从所述可选位置检索所述文件的数据的备份信息。

12. 如权利要求 7 所述的方法，其特征在于，所述文件被定义成包括包含主要数据和次要数据的数据，所述主要数据在大小上大于所述次要数据，至少将所述文件的主要数据从所述文件移除并将其存储在所述可选位置。

13. 一种计算机设备，具有：

存储多个文件的存储卷；

管理所述存储卷并访问所述多个文件的文件系统；以及

其中，每个文件被定义成包括数据和与该数据相关的元数据，所述文件通过下述方式以较小的备份形式被存储在所述存储卷上：所述文件的数据的一部分从所述文件移除并被存储在多个可选位置中的一个，以便释放所述存储卷上的空间，

所述计算设备还具有用于所述多个可选位置的每一个的相应的备份管理器，每个备份管理器被用于和与之对应的可选位置通信，所述计算设备上的每个备份文件在其元数据中具有备份信息，所述备份信息包括所述计算设备的特定备份管理器的标识，所述经标识的备份管理器负责所述备份文件；

由此，每个所述备份文件可通过在其备份信息中定位所述计算设备的特定备份管理器的标识来重建，并将请求传递到所述经标识的备份管理器以获取来自相应的可选位置的所述备份文件的数据，所述经标识的备份管理器实际上从所述可选位置获取所述备份文件的数据，并且将所述获取的数据重建为所述备份文件。

14. 如权利要求 13 所述的设备，其特征在于，所述计算设备是分支服务器，每个可选位置是远离所述分支服务器的集线器服务器，所述集线器服务器为多个所述分支服务器提供服务。

15. 如权利要求 13 所述的设备，其特征在于，所述备份文件的元数据实质上包括在备份之前所述文件的所有元数据，也包括用于从所述可选位置获取所述文件的数据的备份信息。

16. 如权利要求 13 所述的设备，其特征在于，所述文件被定义成包括包含主要数据和次要数据的数据，所述主要数据在大小上大于所述次要数据，至少将所述文件的主要数据

从所述文件移除并将其存储在所述可选位置。

备份文件

技术领域

[0001] 本发明涉及允许来自存储在存储卷上的计算机文件的数据被移动或备份到可选位置以释放存储卷上的空间的结构和方法。更为具体地，本发明涉及这样的结构结构和方法，其中停留在存储卷上的备份文件的剩余部分和在可选位置处的已备份文件可以（如果需要的话）被获取并且被放回被备份的文件中以形成反备份文件。

[0002] 在诸如个人计算机、计算机服务器等计算设备中，正如已知的，数据通常以驻留在计算设备的一个或多个本地存储卷上的本地计算机文件的形式持久地存储在计算设备上。每个这种存储卷可以驻留在计算设备的硬盘等上，这种存储卷可以由在计算设备上运行的文件系统来组织、通过其被访问或者由其控制，这也是已知的。

[0003] 有时，可能对存储卷上的一些、许多、可能甚至是大多数的计算机文件并不感兴趣的情况，并且会被认为是“冷的”。即，举例而言，这种冷文件已经一段时间未被访问和/或可能在未来一段时间也不会被访问，因此将其保留在卷上几乎没有（如果有的话）实际价值。

[0004] 当然，可以简单地将这种冷文件从存储卷上删除，尤其是在这种卷上需要空间时。然而，应该理解，多数用户厌恶简单地删除文件以创建空间。此外，仅因为感觉上废弃而删除冷文件被认为是较差的做法。无论如何，情况可以并可能是虽然认为冷文件是不需要的且预见为不需要的，但是仍然可能在将来某个时间变成需要的。

[0005] 那么，在这种情况下，能够通过将数据从这种冷文件中移动到可选位置，同时仍然允许这种冷文件维持出现在卷上来在卷上创建空间，这会是有用的。即，能够通过将数据从这种冷文件移动到可选位置或将整个文件复制到可选位置，同时将冷文件以较小的、备份形式的留在卷上来备份冷文件会是有用的。因此，如果事实上需要在计算设备上备份文件，那么可以从替代位置重建这种备份的文件，基于此来重建经备份的文件并且接着可以采用这种重建的文件。

[0006] 在另一种情况下，情况可以是组织等的数据文件要从可能是诸如集线器等中央位置复制到该组织的多个分支位置上。例如，结构设计公司会希望其所有的结构设计文件在该公司的若干个分支中的任何一个都可用。

[0007] 在这种情况下，联网的系统可以被构建成将每个文件的副本复制到每个分支，并将所有文件保存在当前所有分支上。在这种系统中，例如，中央集线器服务器会存储每个这样的文件，并且会采用复制服务将集线器处每个文件的副本通过网络发布到每个分支上的分支服务器等上。然而，应该理解，随着组织的文件的数目的增长，随着每个文件的大小的增长，随着分支数目的增长，网络上的话务量也会增长，可能会到超过可用带宽的程度。此外，随着集线器服务器处所有文件的整体大小的增长，实际上可能的情况是每个分支服务器上没有足够的可用空间来存储从集线器服务器复制的所有这种文件。

[0008] 然而，类似于先前的情况，情况可能是对特定分支的分支服务器上的一些、许多或可能甚至大多数计算机文件都不感兴趣，并且它们可以被认为是多余的。这种多余的文件可以是，例如涉及不与特定分支相关的事件，可能不会通过特定分支来访问，因此复制到特

定分支的分支服务器上几乎没有（如果有的话）实际价值。例如，宾夕法尼亚州的威尔克斯一巴里市中的结构设计公司的支局可能几乎不需要（如果有的话）具有与佛罗里达州柏英敦海滩中的公司的支局处理的项目有关的结构设计文件。

[0009] 于是，在这种情况下，类似于先前的情况，能够为该组织仅将与特定分支相关的那些文件全部存储到特定分支的分支服务器上，而仅将来自集线器服务器可用的所有其他非相关文件部分地存储到这种特定分支的这种分支服务器上会是有用的。因此，以与上一情况类似的方式，能够将非相关文件‘备份’到特定分支的分支服务器上，使得每个非相关文件以较小的、‘备份’形式保持在分支服务器上会是有用的。因此，再次，如果在分支服务器处需要经备份的文件，可用从集线器服务器获取这种经备份的文件的数据，并基于此来重建经备份的文件，并且接着可以使用这种经重建的文件。

[0010] 因此，存在对一种方法和机制的需求，该方法和机制用于复制或备份在诸如本地卷或分支服务器的源处的文件使得其上的数据被存储在诸如可选位置或集线器服务器的接收器中，并且源处的文件由此以如有需要可被重建的简化或备份的形式重建。具体地，存在对这种方法和机制的需求，用于按需形成和重建这种经备份的文件。

发明内容

[0011] 本发明至少部分地满足了上述需求，其中连同具有存储卷的计算设备、管理存储卷的文件系统、以及由文件系统存储在存储卷上并由这种文件系统访问的多个文件提供了一种方法。每个文件被定义为包括数据和与该数据相关的元数据，以及从文件中移除并存储在多个可选位置中的一个上的数据的至少一部分，这样使得数据不会实质上占据卷上的任何空间，并且文件是以简化的、备份的形式。此外，对于多个可选位置中的每一个，计算设备具有相应的备份管理器，其中每个备份管理器被用于和与之对应的可选位置通信。

[0012] 为了重建经备份的文件以供使用，计算设备上的单个备份过滤器定位经备份文件的元数据中的备份信息，并在备份信息中定位计算设备的特定备份管理器的标识，其中经标识的备份管理器负责经备份的文件。其后，备份过滤器向经标识的备份管理器传递请求以从相应的可选位置获取经备份文件的数据，并且经标识的备份管理器实际上从这种可选位置获取这种请求数据。一旦从经标识的备份管理器接收经备份文件的请求数据，备份过滤器则将接收到的数据重建为经备份的文件。由此，单个备份过滤器对计算设备的所有备份管理器是通用的。

附图说明

[0013] 当结合附图阅读时，能更好地理解以上概述和以下对本发明的实施例的详细描述。为了示出本发明，在附图中示出了当前优选的实施例。然而，正如应该理解的，本发明并不限于示出的精确的方案和手段。在附图中：

[0014] 图 1 是示出通用计算机系统的框图，其中可以包括本发明的各方面和 / 或其部分；

[0015] 图 2 是依照本发明的实施例示出的在源处备份的文件的框图，这样其数据就存储在接收器中；

[0016] 图 3A 和 3B 是依照本发明的实施例示出的图 2 的文件和数据的框图，其中仅将数

据存储在接收器中（图 3A），以及将整个文件存储在接收器中（图 3B）；

[0017] 图 4 是依照本发明的一个实施例示出的在将文件备份到图 2 的接收器过程中执行的主要步骤的流程图；

[0018] 图 5 是依照本发明的一个实施例示出的从图 2 的接收器重建经备份的文件过程中执行的主要步骤的流程图；

[0019] 图 6 是依照本发明的一个实施例更详细地示出的从图 2 的接收器重建经备份的文件过程中执行的主要步骤的流程图；

[0020] 图 7 是依照本发明的一个实施例示出的从图 2 的接收器部分地重建经备份的文件过程中执行的主要步骤的流程图；

[0021] 图 8 是依照本发明的一个实施例示出的图 2 的文件的数据的各种情况的框图；

[0022] 图 9 是依照本发明的一个实施例示出的基于图 8 中提出的部分的数据段的状况的从图 2 的接收器有效地重建经备份文件的至少一部分的过程中执行的关键步骤的流程图；

[0023] 图 10 是依照本发明的一个实施例示出的当将文件重新备份到图 2 的接收机中时执行的主要步骤的流程图；

[0024] 图 11 是依照本发明的一个实施例示出的与多个接收器相关联的图 2 的源的框图，其中源具有单个通用的备份过滤器和与每个接收器对应的备份管理器。

[0025] 图 12 是依照本发明的一个实施例示出的在源处备份和重建文件时由图 11 的备份过滤器执行的主要步骤的流程图。

具体实施方式

[0026] 计算机环境

[0027] 图 1 和以下描述旨在提供对可在其中实现本发明和 / 或其部分的合适的计算环境的简要、一般描述。尽管不是必需的，但本发明将在由诸如客户机工作站或服务器的由计算机执行的诸如程序模块的计算机可执行指令的一般上下文中描述。一般而言，程序模块或软件组件包括例程、程序、对象、组件、数据结构等，它们执行特定任务或实现特定抽象数据类型。此外，应该理解，本发明和 / 或其部分可以使用其他的计算系统配置来实现，包括个手持式设备、多处理器系统、基于微处理器或可编程消费者电子产品、网络 PC、小型机、大型计算机等。本发明也可以在分布式计算环境中实现，其中任务由通过通信网络链接远程处理设备来执行。在分布式计算环境中，程序模块可以位于本地和远程存储器存储设备中。

[0028] 如图 1 中所示，示例性通用计算系统包括常规个人计算机 120 等，包括处理单元 121、系统存储器 122 和将包括系统存储器在内的各种系统组件耦合至处理单元 121 的系统总线 123。系统总线 123 可以是若干种总线结构中的任一种，包括存储器总线或存储器控制器、外围总线和使用各种总线结构中的任一种的局部总线。系统存储器包括只读存储器 (ROM) 124 和随机存取存储器 (RAM) 125。基本输 入 / 输出系统 (BIOS) 126 包含有助于诸如启动时在个人计算机 120 中的元件之间传递信息的基本例程，它可以存储在 ROM 124 中。

[0029] 个人计算机 120 还可以包括从硬盘（未示出）中读取或向其写入的硬盘驱动器 127，从可移动磁盘 129 中读取或向其写入的磁盘驱动器 128，以及从诸如 CDROM 或其它光学介质等可移动光盘 131 中读取或向其写入的光盘驱动器 130。硬盘驱动器 127、磁盘驱动器 128 和光盘驱动器 130 分别由硬盘驱动器接口 132、磁盘驱动器接口 133 和光盘驱动器接口

134 连接至系统总线 123。驱动器及其相关联的计算机可读介质为个人计算机 120 提供了对计算机可读指令、数据结构、程序模块和其它数据的非易失性存储。

[0030] 尽管此处所述的示例性环境采用硬盘、可移动磁盘 129 和可移动光盘 131，但是应该理解，能够存储可由计算机访问的数据的其他类型的计算机可读介质也可用于示例性计算环境中。这种其他类型的介质包括盒式磁带、闪存卡、数字多功能盘、贝努利盒式磁盘、随机存取存储器 (RAM)、只读存储器 (ROM) 等。

[0031] 可以在硬盘、磁盘 129、光盘 131、ROM 124 或 RAM 125 上存储多个程序模块，包括操作系统 135、一个或多个应用程序 136、其它程序模块 137 和程序数据 138。用户可以通过诸如键盘 140 和定点设备 142 的输入设备向个人计算机 120 输入命令和信息。其他输入设备（未示出）可以包括麦克风、操纵杆、游戏垫、圆盘式卫星天线、扫描仪等其它输入设备。这些和其它输入设备通常由耦合至系统总线的串行端口接口 146 连接至处理单元 121，但也可以由其他接口连接，诸如并行端口、游戏端口或通用串行总线 (USB)。监视器 147 或其他类型的显示设备也经由诸如视频适配器 148 的接口连接至系统总线 123。除监视器 147 之外，个人计算机一般包括其它外围输出设备（未示出），诸如扬声器和打印机。图 1 的示例性系统也包括主机适配器 155、小型计算机系统接口 (SCSI) 总线 156 和连接到 SCSI 总线 156 上的外部存储设备 162。

[0032] 个人计算机 120 可以工作在使用到一个或多个诸如远程计算机 149 的远程计算机的逻辑连接的联网环境内。远程计算机 149 可以是另一个人计算机、服务器、路由器、网络 PC、对等设备或其它公共网络节点，并且一般包括与个人计算机 120 相关的许多或所有上述元件，然而图 1 中仅示出了存储器存储设备 150。图 1 所示的逻辑连接包括局域网 (LAN)151 以及广域网 (WAN)152。这种网络环境常见于办公室、企业范围的计算机网络、内联网以及因特网。

[0033] 当用于 LAN 联网环境时，个人计算机 120 通过网络接口或适配器 153 连接到 LAN 151。当用于 WAN 联网环境时，个人计算机 120 一般包括调制解调器 154 或用于在诸如因特网的广域网 152 上建立通信的其它装置。可以是内置或外置的调制解调器 154 经由串行口接口 146 连接到系统总线 123。在联网环境中，相关于个人计算机 120 描述的程序模块或其部分可以存储在远程存储器存储设备中。可以理解，所示网络连接是示例性的，并且可以使用在计算机间建立通信链接的其他装置。

[0034] 备份和重建文件

[0035] 在本发明中，现在转向图 2，诸如个人计算机或计算机服务器等 D 计算设备 10 具有诸如硬盘或永久性 RAM 驱动器等的存储卷 12，卷 12 上存储有多个计算机文件 14，并且卷 12 上的文件 14 是由在计算设备 10 上运行的文件系统 16 组织、通过其访问，或者由其控制的。如可以理解的，计算设备 10、卷 12、文件 14 和文件系统 16 可以是任何类型的计算设备、卷、文件和文件系统，而不会背离本发明的精神和范围。

[0036] 在本发明的一个实施例中，计算设备 10 是个人计算机等，且在其卷 12 上的文件 14 中，至少一些文件 14 被确定为冷的，例如没有被访问已持续一段时间和 / 或可能不会被访问将持续一段时间，因此保留在卷 12 上几乎没有任何（如果有的话）实际价值。注意，虽然可以用任何适当的方式定义文件 14 的冷，但不会背离本发明的精神和范围。

[0037] 无论如何，当被确定为冷的时，文件 14 不会从卷 12 上删除而是通过将它的至少一

部分数据 20 移动到可选位置 18 来减小卷 12 上的大小。这种可选位置 18 可以是计算设备本地的或者远程的。一般地，这种可选位置 18 被适当地耦合到计算设备 10，并且可以是任何适当的存储位置而不会背离本发明的精神和范围。例如，可选位置 18 可以是存储设备 10 上的另一卷 12、另一计算设备 10 上的另一卷 12、服务器处的文件仓库、远程服务器处的长期存储设备等。

[0038] 有了可选位置 18，那么可以通过将上述数据 20 从这种冷文件 14 移动到这种可选位置 18 来释放卷 12 上的空间。重要的是，虽然数据 20 从这种冷文件 14 中移动，但是这种冷文件 14 在卷 12 上保持存在或‘备份’，虽然是以减小或‘经备份’的形式。因此，如果实际上在计算设备 10 上需要经备份的文件 14，那么就通过从可选位置 18 检索其数据 20 并将这种经检索的数据 20 与这种经备份的文件 14 重新关联以形成重建的文件 14 来重建这种经备份的文件 14。如可以理解的，一旦被重建，文件 14 于是可以实际被采用。

[0039] 在本发明的另一实施例中，计算设备 10 是分支服务器等，且其卷 12 上的至少一部分文件 14 已被确定为无关的，因为例如这种文件 14 和与这种分支服务器相关联的支局不相关。当然，关于文件 14 的不相关性可以用任何适当的方式定义，而不会背离本发明的精神和范围。

[0040] 无论如何，如上所述，不相关的文件 14 以简化的格式保留在分支服务器 10 的卷 12 上，而没有至少一些数据 20。此处，这种数据 20 被存储在诸如中央集线器服务器的可选位置 18 处，所述中央集线器服务器由其支局是分支的组织来维护。这种集线器服务器 18 再次被适当地耦合到分支服务器 10 上。

[0041] 有了集线器服务器 18，于是通过来自这种不相关的文件 14 的上述数据 20 存储到计算器服务器 18 处，可以最小化分支服务器 10 的卷 12 上使用的空间，也可以最小化填充该空间所需的带宽。重要的是，并且与之前类似，虽然来自这种不相关的文件 14 的数据 20 不出现在分支服务器 10 上，但是这种不相关的文件 14 在卷 12 上保持存在或‘备份’，虽然是以减小或‘备份’的形式。因此，如果实际上在分支服务器 10 处需要经备份的文件 14，那么就通过从集线器服务器 18 检索其数据 20 并将这种经检索的数据 20 与这种经备份的文件 14 重新关联以形成重建的文件 14 来重建这种经备份的文件 14。而且，一旦被重建，文件 14 于是可以实际被采用。

[0042] 为了推广起见，无论方案中涉及诸如个人计算机的计算设备 10、可选位置 18、分支服务器 10 和集线器服务器 18 或其他方面，‘备份’涉及源 10 处的卷 12 上的文件 14，如图 2 所示，其中已作出确定，文件 14 应该仅以简化的格式出现在卷 12 上，所述简化的格式不包括与这种文件 14 相关联的数据 20 的某些部分。相应地，如图 2 中所示，这种数据 20 被存储在适当地耦合到源 10 处的接收器 18 处。

[0043] 有了这种接收器 18，于是可以最小化源 10 的卷 12 上使用的空间，因为卷 12 上经备份的文件 14 以简化或‘剩余’的形式出现。因此，如果在源 10 处实际上需要经备份的文件 14，那么通过从接收器 18 检索其数据 20 并将这种检索的数据 20 与这种经备份的文件 14 重新关联以形成重建的文件 14 来重建这种经备份的文件 14。而且，一旦被重建，文件 14 于是可以在源 10 处按所需地实际被采用。

[0044] 在本发明的一个实施例中，出现在源 10 的卷上的经备份的文件 14 的剩余被存储在卷上与源文件 14 相同的位置。因此，无论文件系统 16 依照目录格式或其他方式组织卷

12，在特定位置查找被备份的文件 14 的文件系统 16 应该在这样的位置找到文件 14 的剩余，即经备份的文件 14。重要的是，驻留在源 10 的卷 12 上的 经备份的文件 14 包含来自源、未备份的文件 14 的所有元数据 22，并且也包含可用于从接收器 18 为文件 14 检索数据 20 所使用的备份信息 24。如可以理解的，这种备份信息 24 可被添加到文件 14 的元数据 22 中，或被存储在文件 14 的另一位置中。

[0045] 因此，用户等可以通过文件系统 16 浏览源 10 的整个卷 12，即使卷 12 上的一些文件 14 被备份，而卷 12 上的一些文件 14 没有被备份。具体地，在这种浏览期间，可采用与经备份的文件 14 相关联的元数据 22 以标识经备份的文件 14、提供大小信息、日期信息等。当用户试图访问经备份的文件 14 时，基于经备份的文件 14 中出现的备份信息 24 来检索数据 20，将经备份的文件 14 重建成未经备份的文件 14，并且未经备份的文件 14 接着可实际上由用户访问。

[0046] 现在转向图 3A，可以看到可驻留在源 10 处的卷 12 上的未备份文件 14 包括可以被组织到头部等中的元数据 22，其中，这种元数据 22 包括与未备份的文件 14 的数据 20 有关的信息，诸如逻辑文件大小、卷 12 上的大小、创建时间、修改时间、访问时间、访问权限以及各种文件属性。此外，显然地，这种驻留在元 10 的卷 12 上的未备份文件 14 包括文件 14 的实际数据 20。这里注意，这种实际数据 20 可以被组织成主要数据和可选数据，其中主要数据是来自文件 14 可以被应用程序 30 等（图 2）所采用的数据 20，而可选数据是文件 14 的未其他使用所创建的数据。仅作为一个示例，这种可选数据可以包括诸如‘拇指’的图形表示，当显示文件 14 的表示时可采用（可能被应用程序 30 或源 10 处的文件系统 16）该“拇指”。

[0047] 无论如何，情况可能是主要数据是按照大小的大量数据 20。因此，当未备份文件 14 实际上被备份时，实际的情况可能是仅将数据 20 的主要数据从未备份文件 14 的剩余部分移除，以形成经备份的文件 14，并且仅将这些主要数据存储在接收器 18 中，可能与经备份文件 14 的标识一起，如图 3A 所示。当然，文件 14 的其他部分也可以在文件 14 的备份期间被移除，而会不背离本发明的精神和范围。

[0048] 如可以理解的，虽然当未备份文件 14 在分支服务器 10 和集线器服务器 18 的设备环境中实际上被备份时，其中集线器服务器 18 将文件 14 复制到分支服务器 10 中，但实际情况可能是当将数据 20 的主要数据从源 18 处的复制的文件 14 的剩余部分移除以形成已备份文件 14 时，这种文件 14 全部被存储在接收器 18 中，可能与经备份文件 14 的标识一起，如图 3B 所示。

[0049] 为了将未备份的文件 14 转换成经备份的文件 14（即备份文件 14），现在参考图 3 和 4，首先标识文件 14（步骤 401），此后标识要从文件 14 上移除的数据 20（步骤 403）。再者，这种数据 20 通常是数据 20 的主要数据，但可以是来自数据 20 的其他数据。无论如何，然后通过适当的传输机制和渠道将经标识的数据 20 从源 10 移动到接收器 18（步骤 405），然后将这种经移动的数据 20 以适当的格式存储在接收器 18 处（步骤 407）。如可以理解的，可以用任何适当的格式和方式将数据 20 存储在接收器 18 中，而不会背离本发明的精神和范围。例如，如果期望可以压缩和 / 或加密数据 20，可以用某种方式存储这种数据 20，以便收到重建文件 14 的请求时，可以相对简单地检索这种数据 20。例如，可依照为检索所采用的 ID 来存储这种数据 20，如以下将详细描述的。

[0050] 一旦在步骤405处移动了数据20，由这种数据20在源10的卷12上占据的空间无需保持分配给文件14，正如可以理解的。因此，这种分配的空间可被标记为空闲的（步骤409），其结果是现在备份的文件14被标记为稀疏的（sparse）。重要的是，虽然从经备份文件14的使用中释放出这种分配的空间，但现在备份的文件14的逻辑文件大小在这种文件14的元数据22或别处不应该被改变，尽管文件14的卷上的大小可能实际上被改变以反应这种被释放的空间。

[0051] 此外，在本发明的一个实施例中，修改现在备份的文件14的元数据以示出这种文件14现在被备份（步骤411）。举例而言，为此可以在元数据22中设置‘备份’属性。如可以理解的，这种设置备份属性可以主要用作查询应用程序30等的信号，表明文件14实际上被备份了。有了这种信号，查询应用程序30于是可以意识到，不仅文件14实际上被备份了，而且访问这种文件14涉及成本，因为必须首先将这种文件14重建成未备份的形式。如可以理解的，这种成本可以根据从接收器18访问文件14的数据20所需的带宽、卷12上数据20所需的空间和/或在接收器18处访问数据20和基于此来重建文件14的等待时间。

[0052] 同时，在本发明的一个实施例中，修改现在备份的文件14的元数据22以包括上述备份信息24（步骤413）。此外，这种备份信息24包括从接收器18检索文件14的数据20所用的信息等。例如，这种备份信息24可以包括在接收器18处与数据20一起存储并标识该数据的ID，以及可用于存储与经备份的文件14相关的其他信息的存储量，包括如何定位接收器18、如何重建文件14等的信息。在备份信息24上的这种存储可以是可变或固定的量，并且在后一种情况下可能限于16千字节左右。如可以理解的，这种备份信息24主要由用于重建文件14的任何构造所采用，并且一般不由上述查询应用程序30等使用，虽然这种使用仍然可能发生，而不会背离本发明的精神和范围。

[0053] 如现在应该理解的，经备份的备份文件14可以不受干扰地驻留在源10的卷12上较短的时间、较长的时间、很长的时间或者永远，至少直至卷12不再运作。可是在任何时间点，现在转向图5，应该理解可用从应用程序30、在应用程序30处的用户等接收到访问备份文件14的存储数据20的请求（步骤501）。当然，在这种请求之前，也可能接收到一个或多个要访问这种备份文件14的元数据22的请求。例如，上述应用程序30可以访问这种备份文件14的元数据22中的设置备份属性，以确定文件14实际上被备份了，或者控制卷12的文件系统16可以在编译目录列表等的过程中访问元数据22。在后一种情况下，显著的是，文件系统16、应用程序30或其他实体也可以访问数据22的可选数据以获取当显示文件14的表示时可采用的上述图形。

[0054] 无论如何，相应如在步骤501处访问备份文件14的存储数据20的请求，实际上用以下方式重建备份文件14。首先，文件14位于卷12上（步骤503），随后标识出文件14的元数据22中的备份信息24（步骤505）。如可以理解的，存储在接收器18处的文件14的数据20基于这种标识的备份信息24来定位（步骤507），于是这种定位数据20可以通过适当的传输机制和渠道从接收器18移动到源10处（步骤511）。

[0055] 当然，为了在源10处存储文件14的这种数据20，必须分配由源10的卷12上的这种数据20所占据的空间（步骤509），其结果是不再备份的文件14不再被标记为稀疏的。如现在可以理解的，当实际上将数据20移动到文件14以重建相同的文件后，修改不再备份的文件14的元数据22以示出这种文件14没有被备份（步骤513），诸如通过重置元数据22

中的备份属性。同时，修改不再备份文件 14 的元数据 22 以删除备份信息 24（步骤 515）。相应地，文件 14 现在是未备份形式。

[0056] 如现在应该理解的，有两种主要的情况导致创建备份文件 14。在第一种情况中，直接将文件 14 创建为卷 12 上的备份文件 14。这种备份创建在诸如上述具有分支服务器 10 和集线器服务器 18 的复制结构中是典型的。在这种复制结构中，使用来自中央集线器服务器 18 的所有文件 14 的未备份副本来填充若干分支服务器 10 会需要非常大量的带宽。因此，而是复制引擎会决定只在分支服务器 10 上创建文件 14 的备份副本。于是可以按需和当需要才从集线器服务器 18 检索任何特殊分支服务器 10 处的任何特定的备份文件 14 相关联的数据 20 以将这种特定的备份文件 14 重建成这种特定的分支服务器 10 处的未备份文件 14。

[0057] 在第二种情况中，文件 14 被创建为卷 12 上的未备份文件 14 并在某个较晚的 时间被转换成这种卷 12 上的备份文件 14。这种未备份的创建在诸如上述具有个人计算机等和可选位置 18 的空间节省结构中是典型的。在这种空间节省结构中，通过将与例如很少使用的文件 14 相关联的数据 20 移动到可选位置 18（可以是另一媒体或到另一系统）来收回个人计算机等的卷 12 上的空间。如第一种情况中，于是可以按需和当需要时从可选位置检索与个人计算机处的任何特定备份文件 14 相关联的数据 20 以将这种特定的备份文件 14 重建为未备份文件 14。

[0058] 在本发明的一个实施例中，在源 10 处代表用户或应用程序 30 以及有关备份文件 14 执行的图 5 的操作实际上是在备份过滤器 26 的协助下执行的，如在图 2 中所示。可以透明地或通知用户或应用程序 30 地执行这些操作，虽然透明对用户或应用程序 30 可能是优选的。特别地，当用户或应用程序 30 试图访问备份文件 14 的移除的数据 20 时，当接收到这种访问的请求时，文件系统 16 会发现备份文件 14 并不包含移除的数据 20，并因此返回备份过滤器 26 将会截取的错误。一旦有这种截取，备份过滤器 26 就会利用文件系统 16 来从备份文件 14 的元数据 22 获取备份信息 24，并且基于获取的备份信息 24，这种备份过滤器 26 触发这种备份文件 14 的重建，如以下会更详细描述的，以便访问请求最终得到实现。

[0059] 如可以理解的，备份过滤器 26 可以是低级的构造而没有太多的功能且不能访问诸如接收器 18 的网络资源。在这种情况下，如图 2 所示，备份过滤器 26 会与诸如备份管理器 28 的高级构造连接，其中这种备份管理器 28 包括其他备份功能和对诸如接收器 18 的网络资源的直接访问。此外，在这种情况下，应该理解的是当备份过滤器 26 触发这种文件 14 的这种重建时，备份管理器 28 执行对源 10 的大量的重建功能，如以下将结合图 5 更详细地描述的。这种备份服务器 28 也会执行对源的大量的备份功能，如结合图 4 所示。

[0060] 应用程序 30 从备份文件 14 请求数据 20

[0061] 虽然以上结合图 4 详细地描述了重建备份文件 14，但从希望读取这种备份文件 14 的数据 20 的某些部分的应用程序 30 等的角度看，再次访问这个过程是有益的。如可以理解的，则希望从这种文件 14 读取这种数据 20 的这种应用程序 30 通常通过向与文件 14 相关的文件系统 16 发出打开命令然后向与这种打开的文件 14 有关的文件系统 16 发出读取命令来实现这种功能。在本发明的一个实施例中，这种应用程序 30 连续发出这种打开和读取命令，虽然有时因正在备份的在讨论的文件 14 会略有不同地解释这些命令。这种不同对于应用程序 30 而言是透明的或者不 关心的，尽管应用程序 30 会在重建备份文件 14 的过

程中经历一些等待时间，这种等待时间可能很大程度上取决于将数据 20 从接收器 18 移动到源 10。

[0062] 无论如何，现在转向图 6，应用程序 30 通过发出有关卷 12 上的特定的备份文件 14 的打开命令来开始该过程，这种打开命令最终由文件系统 16 接收到（步骤 601）。显著地，发出这种打开命令的应用程序 30 不期望知道特定的文件 14 实际上被备份，虽然应用程序 30 可能实际上通过命令文件系统 16 报告是否设置了文件 14 的备份属性可作出这样的判定。

[0063] 如以上上所提及的，文件系统 16 在接收到打开命令时注意到备份文件 14 不包含移除的数据 20，由此返回错误（步骤 603），备份过滤器 26 截取这种返回的错误，并基于此认识到在讨论的文件 14 实际上是备份的格式的（步骤 605）。因此，备份过滤器 26 自身命令文件系统 16 从这种备份文件 14 中检索备份信息 24，并且实际上接收这种备份信息 24（步骤 607）。基于这种备份信息 24，备份过滤器 26 接着创建对应于备份文件 14 的缓冲器和句柄，并将该句柄传递给应用程序 30，作为对来自该应用程序的打开命令的（正常）响应（步骤 609）。

[0064] 注意，由应用程序 30 接收到的备份文件 14 的句柄表示打开的备份文件 14 的实例，并且当发出与这种文件 14 有关的诸如读取命令或关闭命令的进一步命令时，由应用程序 30 用作对这种文件 14 的参考。然而，重要的是由备份过滤器 26 创建并且用作来自应用程序 30 的命令的句柄参数的备份文件 14 的这种句柄会使得这种命令被直接传递到备份过滤器 26，而不是到文件系统 16。

[0065] 也要注意，备份过滤器 26 使用备份文件 14 的缓冲器来存储有关备份文件 14 的信息，诸如其备份信息 24。因此，在备份过滤器 26 操作备份文件 14 的过程中，这种备份信息 24 的改变可以被记录在缓冲器中，并无需写回到文件 14 中。当然，如果需要的话，当备份过滤器 26 完成备份文件 14 的操作后，可以将缓冲器中的这种记录的改变写回到文件的备份信息 24 中。

[0066] 无论如何，有了来自备份过滤器 26 的备份文件 14 的句柄，应用程序 30 可通过发出与卷 12 上的特定的备份文件 14 有关的读取命令而继续辖区，其中这种读取命令具有包括句柄、文件的数据 20 中的偏移量、与从该偏移量开始的这种数据 20 有关的读取长度等参数。因此，基于使得这种句柄作为其参数的这种读取命令最终由备份过滤器 26 而不是文件系统 16 接收到（步骤 611）。然而，很显然的是，数据 20 实际上不存在于备份文件 14 中，因此备份过滤器将读取命令存储在队列等中，用于等待进一步的处理（步骤 613）。

[0067] 如应该理解的，这种进一步的处理主要包括从接收器 18 获取备份文件 14 的数据 20。具体地，备份过滤器 26 基于接收器 18 处的它的数据 20 开始重建备份文件 14（步骤 615），其中这种重建会以以下方式发生。

[0068] 记住，备份过滤器 25 会与备份管理器 28 连接以执行高级备份功能和对诸如接收器 18 的网络资源的直接访问，备份过滤器 26 向备份管理器 28 发出请求，以从接收器 18 获取备份文件 14 的数据的一些部分（步骤 615a），其中这种请求包括来自备份文件 14 的备份信息 24 的 ID、到数据 20 中的偏移量以及从这种偏移量开始获取的数据 20 的长度。此后，备份管理器 28 实际上从接收器 18 获取这种数据 20 的请求的部分（步骤 615b）。假设备份管理器 28 包括这样获取的所有必须的功能，并且使用已知或对有关公众显而易见并且无

需在此作出任何详细描述的方式进行。相应地，可以采用这种获取的任何特定的方法，而不会背离本发明的精神和范围。

[0069] 注意，情况可能是如在步骤 615b 处备份管理器 28 从接收器 18 获取数据 20 的请求的部分时，可能有一些延迟。例如，情况可能是通过网络来获取这种数据 29，其中可能有网络延迟，或许类似于若干微秒、若干秒或甚至更多。因此，备份管理器 28、备份过滤器 26 和 / 或另一实体可能希望维护表示从接收器 18 请求的并且还未从中获取的数据 20 的每个部分（可能与相应的请求时间一起）的待决数据列表。如可以理解的，当被如此请求时，数据的每个请求部分的标识被添加到待决数据列表中，并且当被接收到且与备份文件 14 一起存储时从这种待决数据列表中移除。有了这种待决数据列表，则情况可能是监控所有对数据 20 的待决请求的时效，而且在某个时间段内未满足的请求可以超时。

[0070] 一旦备份管理器 28 实际上从接收器 18 获取了这种数据 20 的这种请求部分，那么请求管理器 28 将这种请求部分传递给备份过滤器 26（步骤 615c），并且备份过滤器 26 然后通过对文件系统 16 的适当的命令将这种请求的部分写入在讨论的文件 14 的适当位置（步骤 615d）。将这种请求的部分写入在讨论的文件 14 的适当位置是已知的或者对于相关的公众是显而易见的，因此无需在此作任何详细描述。因此，可以采用这种写入的任何特定方法而不会背离本发明的精神和范围。

[0071] 当然，可能必须多次重复步骤 615a-615d，直至文件 14 的所有数据 20 被写入，因此按照需要重复这些步骤。一旦文件 14 的所有数据 20 实际上被写入，备份过滤器 26 根据需要向向文件系统 14 发布命令以整理 (fix) 这种现在重建的文件 14 的元数据 22，包括移除备份信息 24 和重置备份属性（步骤 617）。此外，备份过滤 器 26 从队列检索文件 14 的所有读取命令，并将这种读取命令传递给文件系统 16 以供进一步处理（步骤 619），由此来自应用程序 30 的读取命令实际上响应于在讨论的数据 20。

[0072] 注意，打开文件 14 的句柄保持与备份过滤器 26 相关联，并继续将来自与这种现在重建的文件 14 相关的应用程序 30 的命令指引到备份过滤器 26。因此，情况可能是备份过滤器 26 将这些命令转发给文件系统 16，情况可能是备份过滤器 26 重新将句柄与文件系统 16 相关联，或者可能是备份过滤器命令另一构造以将手柄与文件系统 16 相关联等。

[0073] 部分重建备份文件 14

[0074] 如可以理解的，在请求的应用程序需要文件 14 少于全部备份数据 20 的情况下，应该不需要完全地重建备份文件 14。因此，如果应用程序 30 可以确定仅需要数据 30 的 2 千兆字节的特定偏移量处的 1、2、12 或 100 千字节，应该不需要从接收器 18 获取数据 20 的 2 千兆字节，而仅需要在特定偏移量处的数据 20 的千字节。此外，在这种情况下，通过仅仅部分重建备份文件 14，避免了将大量不需要的数据 20 从接收器 18 发送到源 10，并且完成这些所需的带宽相应地降低。

[0075] 通过仅仅部分重建备份文件 14，应用程序 30 于是能够仅读取（如果需要）文件的一些字节，而不用触发文件 14 的完全重建。因此，举例而言，如果仅需要视频文件 14 的第一帧，那么会从接收器 18 处的文件 14 的数据 20 获取该第一帧，而不是备份文件 14 的数据 20 的全部量，可能类似于 10 或甚至 100 千兆字节。于是部分重建可仅从接收器 18 获取为备份文件 14 所需的同样多的数据 20 以满足来自应用程序的特定读取请求，就此为止。

[0076] 注意，当部分重建备份文件 14 时，情况可能是或可能不是实际上从接收器 18 移除

从接收器 18 获取的数据 20 的部分。另一方面，获取的部分被存储在源 10 处，因此不再需要在接收器 18 处维持。然而，另一方面，实际上可能需要更多努力以实际上从接收器 18 删除获取的部分。此外，在至少一些情况下，这种获取部分应该保留在接收器 18 处以供其他源 10 的访问，诸如如果接收器是集线器服务器 18，而源是分支服务器 10。

[0077] 如可以理解的，如果备份文件 14 实际上被部分重建，那么必须维护记录以注意备份文件 14 的什么部分实际上被重建，以便备份过滤器 26 可以确定这些部分是否存在于备份文件 14 中。因此，在本发明的一个实施例中，这种记录被维护在备 份文件 14 的元数据 22 中的备份信息 24 中。具体地，对于重建并由此出现在备份文件 14 中的备份文件 14 的数据 20 的每个连续区域 (section)，备份信息 24 包括区域参考，区域参考包括描述区域的开始的偏移量以及该区域的连续量的长度。

[0078] 通常，由备份过滤器 26 维护包括部分重建数据 20 的上述记录的备份文件 14 的备份信息 24。如以上所提及的，虽然备份过滤器 26 可以通过文件系统 16 直接维护和更新备份文件 14 的元数据 22 中的备份信息 24，但这种维护和更新可能干扰文件系统 16 执行的其他操作。因此，备份过滤器 26 在操作备份文件 14 的过程中而是一开始就从元数据 22 获取备份信息 24，并如图 6 的步骤 609，将备份信息 24 存储在与备份文件 14 有关的创建的缓冲器中，接着在操作备份文件 14 的过程中维护和更新备份信息 24，并且在结束时将备份信息 24 从缓冲器写入到备份文件 14 的元数据 22 中。

[0079] 注意，在备份文件 14 的部分重建可以被中断的场合中，例如源 12 处的功率损失或网络连接。同样地，情况也可能是丢失了备份文件 14 的缓冲器，尤其是在功率损失且缓冲器维护在易失性 RAM 等的情况下。在这种情况下，维护和更新的备份信息 24 丢失并且没有从缓冲器写到文件 14 的元数据 22 中，结果不仅丢失了备份信息 24，而且丢失了与之相关联的所有部分重建的数据 20，即使其物理地存在于源 10 处，尤其是因为没有这种备份信息 24 数据 20 不能被定位。因此，在本发明的一个实施例中，被维护和更新的备份信息 24 被周期性地从缓冲器写入到文件 14 的元数据 22 中，诸如大约每分钟一次。因此，由于缓冲器的丢失，最多会丢失 1 分钟左右的这种备份信息 24 以及与之相关联的部分重建的数据 20。

[0080] 在本发明的一个实施例中，基于来自从备份文件 14 请求数据 20 的应用程序 30 的适当的命令，备份文件 14 仅被部分地重建。因此，应用程序 30 必须首先检查文件系统 16，以确定是否为特定的文件 14 设置了备份属性，以判定这种文件 14 实际上是否被备份，如果是这样，那么应用程序 30 会使用适当的命令请求对这种备份文件 14 的数据 20 的一部分或多部分的部分重建。

[0081] 考虑到部分重建，于是在本发明的一个实施例中，响应于诸如在图 6 的步骤 611 处给出的备份文件 14 有关的读取命令，备份过滤器 26 通过首先查看存储在相应的缓冲器中的备份文件 14 的备份信息 24 中的任何部分参考，接着从该部分参考确定所请求的数据 20 或其部分是否已经存在于源 10 的备份文件 14 中来对此作出响应。如果所有请求的数据 20 实际上存在，那么从源 10 处的备份文件 14 读取这种数据 20，而无需从接收器 18 获取这种数据 20。如果仅存在请求数据 20 的一部 分，那么从源 10 处的备份文件 14 读取数据 20 的存在的部分，而无需从接收器 18 获取这种数据 20，而从接收器 18 处获取数据 20 的剩余部分，并接着如上被读取。如果不存在任何请求的数据 20，就从接收器 18 获取所有这种数据 20，并接着如上被读取。总而言之，现在转向图 7，响应来自应用程序 30 打开可能需要从中

进行部分重建的备份文件 14 的命令（步骤 701），备份过滤器 26 再次创建缓冲器和对应于备份文件 14 的句柄，并将句柄传递给应用程序 30 作为对步骤 609 处来自于此的打开命令的（正常）响应（步骤 703）。此外，备份过滤器 26 一开始从备份文件 14 的元数据 22 获取备份信息 24，并将备份信息 24 存储在创建的缓冲器中（步骤 705）。

[0082] 如上所述，有了来自备份过滤器 26 的过滤文件 14 的句柄，应用程序 30 对卷 12 上特定备份文件 14 的数据 20 的一部分发出读取命令，其中这种读取命令包括句柄以并定义该部分的偏移量和长度，如在步骤 611 处（步骤 707）。此处，响应于读取命令，备份过滤器 26 基于备份文件 14 的缓冲器中的备份信息 24 判定定义的该部分是否至少已部分存在于驻留在源 10 处的备份文件 14 中（步骤 709）。作出这种判定是已知的或者对于相关的公众是显而易见的，因此无需在此作任何详细的描述。因此，可以采用作出这种判定的任何方法而不会背离本发明的精神和范围。

[0083] 再者，如果数据 20 的所有部分实际上是存在，就无需从接收器 18 获取这种数据的该部分（步骤 711a）。如果仅存在数据 20 的部分的一部分，那么从接收器 18 获取数据 20 的剩余部分（步骤 711b）。如果不存在数据 20 的任何部分，那么从接收器 18 获取数据 20 的所有这种部分（步骤 711c）。重要的是，在步骤 711b 和 711c 处从接收器 18 获取数据 20 的任何部分时，备份过滤器 26 更新存储在缓冲器中的备份文件 14 的备份信息 24，以适当地反映数据 20 的该部分现在与备份文件 14 一起驻留并被重建到备份文件 14（步骤 713）。无论如何，由于数据 20 的请求部分现在驻留在源 10 处，这种请求部分现在可能实际上由应用程序 30 读取，如在步骤 619 处（步骤 715）。

[0084] 如上所述，备份过滤器 26 周期性地将备份信息 24 以其最近的形式从缓冲器写到文件 14 的元数据 22 中，这样在缓冲器丢失的事件中不会完全丢失这种备份信息 24 和与之相关联的部分重建的数据 20（步骤 717）。此外，一旦备份文件 14 被关闭，诸如在应用程序 30 的命令时（步骤 719），备份过滤器 26 通过将备份信息 24 以其最近的形式从缓冲器写到文件 14 的元数据 22 中来关闭缓冲器（步骤 721），当然假设备份文件 14 还没有被完全重建。因此，以其最近形式的这种备份 信息 24 可以在以后某个时间再次被获取，如在步骤 705 处。

[0085] 快速读取部分重建备份文件 14

[0086] 发送有关备份文件 14 的读取命令的应用程序 30 不考虑备份文件 14 的状态，尤其不考虑备份文件 14 是否已经被部分重建和 / 或已经处于被部分重建的过程中。即，现在转向图 8，情况可能是与备份文件 14 有关的特定读取命令指定数据 20 的特定部分，这种特定部分对应于包含已经存在于备份文件 14 中的数据 20 的第一分段、包含即将要从接收器 18 复制到备份文件 14 的数据的第二分段，以及包含不存在于备份文件 14 中而是仅存储在接收器 18 处的数据 20 的第三分段。

[0087] 如可以理解的，如果要处理这种读取命令使得数据 20 的所有特定部分从接收器 18 复制到备份文件 14，有关对应于这种特定部分的至少第一和第二分段的这种处理是双重的且浪费的。具体地，且可以理解的，复制对应于第一分段的数据 20 是不必要的，因为这种数据 20 已经存在于备份文件 14 中，复制对应于第二分段的数据 20 是不必要的，因为这种数据 20 已即将从接收器 18 复制到备份文件 14。实际上，仅需要将对应于第三分段的数据 20 从接收器 14 复制到备份文件 14 中，因为这种数据 20 不存在于备份文件 14 中，且还

没有被请求复制到备份文件 14 中。

[0088] 在本发明的一个实施例中，备份过滤器 26 于是通过以下步骤处理对数据 20 的特定部分的读取命令：首先标识对应于已经存在的备份文件 14 中的分段（即第一分段）、已经待决的备份文件 14 中的相应的分段（即第二分段），以及不存在或不是待决的备份文件 14 的相应的分段（即第三分段）的数据 20 的特定部分，然后实际上仅获取第三分段。实际上，备份过滤器 26 然后从读取命令剥除实际上无需从接收器 18 读取的所有分段。通过执行这种剥除动作，能够更快地处理读取命令，因为这种‘快速读取’仅从接收器获取实际上需要的数据 20，而不是已经存在于备份文件 14 中或待决的数据 20。因此，由备份过滤器 26 执行的快速读出造成了对来自于应用程序 30 的读取命令的更快的响应，并降低了这种读取命令所需的带宽量。

[0089] 在本发明的一个实施例中，备份过滤器 26 标识关于为备份文件 14 维护的备份信息已经存在的备份文件 14 中每个第一分段。如上所述，可以参考位于这种备份文件 14 的元数据 22 中的备份信息 24，或者参考位于备份文件 14 的缓冲区中的备份信息 24。同样地，在本发明的一个实施例中，备份过滤器 26 标识已待决的备份文件 14 中的每个第二分段，该备份文件 14 关于以上结合图 6 的步骤 615b 揭示 的待决数据列表中的信息。关于这种备份信息 24 和待决数据列表来标识第一和第二分段应该是已知的或者对于相关的公众是显而易见的，因此无需在此作任何详细的描述。因此，可以用任何适当的方式执行这种标识而不会背离本发明的精神和范围。

[0090] 如现在可以理解的，一旦标识了第一和第二分段，剩余部分就是第三分段，于是通过排除过程，标识出不存在或不是待决的备份文件的第三分段。当然，一旦被标识，可能实际上从接收器 18 处请求第三分段，并在接收到时复制给备份文件 14。注意一旦请求了，每个这种第三分段实际上成为第二、待决分段。也要注意，一旦接收并复制到备份文件 14，每个第二分段实际上变成第一、存在的分段。注意，最后，一旦分段变成了第一分段，那么这种第一分段可用于响应来自于此的读取命令实际上由应用程序 30 读取。

[0091] 至少在某些情况下，应该理解虽然应用程序 30 没有对备份文件 14 中的特定数据 20 发出读取命令，备份过滤器 26 仍然应该预期到这种读取命令将数据 20 从接收器 18 移动到备份文件 14，尤其是如果没有使用其他的方式占据备份过滤器 14。仅作为一个示例，当流动诸如备份音频文件 14、备份视频文件 14 或备份多媒体文件 14 的内容时，合理的是，期望在特定时间段 T0 数据 20 的读取命令会跟随着下一时段 T1 处的数据 20 的读取命令。在这种情况下，备份过滤器 26（如果没有被其他占据）可以使得其自身具有能够从接收器 18 处获取时间段 T1 的数据 20 的机会，甚至无需来自应用程序 30 的特定读取命令。当然，这样做，备份过滤器 26 可执行这种动作，作为用以上所述方法中的快速读取。

[0092] 总而言之，现在转向图 9，备份过滤器 26 用以下方式响应来自应用程序 30 的读取命令而执行快速读取。首先，确实从应用程序 30 接收读取命令，其中这种读取命令指定从备份文件 14 读取出一部分或一定范围的数据 20（步骤 901）。通常，而且这种范围被表示为与数据 20 有关的偏移量和长度。

[0093] 此后，备份过滤器 26 首先标识与已经存在的备份文件 14 中数据 20 的第一分段相应的数据 20 的这种范围（步骤 903）。而且，这种标识是关于为这种备份文件 14 所维护的备份信息 24 而执行的，是在为备份文件 14 的元数据 22 中或者在对应于备份文件 14 的缓

冲器中。如果标识出的第一分段包括数据 20 的所有请求的范围，那么备份文件 14 已经被重建到满足读取命令所需的程度，并且这种读取命令可由此完成，而无需任何进一步地从接收器 18 复制数据 20 以及无需等待任何待决数据 20（步骤 905）。

[0094] 然而，如果标识出的第一分段没有包括数据 20 的所有请求范围，那么必须至少基于待决数据 20，实际上将备份文件 14 重建到满足读取命令所需的程度，因此，备份过滤器 26 计算包括不是第一分段的数据 20 的请求范围的每个分段的范围的第一集合（步骤 907）。此后，类似于先前的，备份过滤器 26 在第一集合中标识出待决的备份文件中 14 数据 20 的相应第二分段（步骤 909）。再者，关于该待决的数据列表来执行这种标识。如果标识出的第二分段包括所有的第一集合，那么备份文件 14 待决地要被重建成满足读取命令所需的程度，并且当将所有的待决数据 20 实际上复制到备份文件 14 后可以完成这种读取命令（步骤 911）。

[0095] 如现在应该理解的，虽然如果标识的第二分段不包含第一集合的全部，但是备份文件 14 必须基于要从接收器 18 复制的数据 20 确实要被重建到满足读取命令所需的程度，因此，备份过滤器 26 计算包括不是第一分段和不是第二分段，即第三分段的数据 20 的请求范围的每个分段的范围的第二集合（步骤 913）。此后，备份过滤器 26 请求将第二集合 / 第三分段从接收器 18 复制到备份文件 14（步骤 915）。

[0096] 经过一定的时间，并且请求的每个第三分段成为第二、待决的分段，实际上复制的每个第二分段成为第一、存在的分段，直至完成所有的复制（步骤 917）。此后，读取命令所请求的范围实际上可用于由应用程序 30 读取，以响应读取命令（步骤 919）。

[0097] 再备份 / 备份文件 14

[0098] 如可以理解的，被重建或部分重建的备份文件 14 在某些时候可以被再备份，例如如果在源 10 处需要更多的空间。此外，如也可以理解的，从来没有被同样备份过的文件 14 在某些点因类似的原因被备份。

[0099] 具体地，一旦部分或完全地重建了备份文件 14，或者一旦文件 14 被安装在源 10 处，那么这种文件 14 继续以这种形式驻留在源 10 处，除非这种文件 14 被再备份或备份（此后成为‘再备份’）。这种再备份是由某些事件触发的，可能是例如确定在源 10 处需要空间、确定文件 14 没有被访问某个时间段等。此外，通过再备份文件 14，尤其在分支服务器 10 和集线器服务器 18 的环境中，在源 10 或接收器 18 处文件 14 的任何改变可以分别被复制到接收器 18 或源 10 处，以保持文件 14 最新。

[0100] 在本发明的一个实施例中，依照再备份算法来执行源 10 处的再备份，再备份算法考虑到诸如最近访问时间、卷上的剩余自由空间、访问数据的频率、在别处发 生的对文件 14 的修改和 / 或类似的因素。考虑到这些因素的再备份算法具有降低备份文件 14 在被再备份后不久被再次重建的可能性的目的。因此，可以最小化（否则为降低）这种备份和再备份所必须的带宽。通常，虽然不是必须的，再备份算法由源 10 处的应用程序 30 或源 10 处的备份管理器 28 执行，虽然其他实体可以执行这种再备份算法，而不会背离本发明的精神和范围。

[0101] 可至少部分基于可配置策略来定义用于再备份源 10 处的文件 14 的再备份算法。具体地，至少可以部分基于可配置策略来触发这种再备份算法，这种再备份算法可至少部分基于可配置策略来决定再备份特定的文件 14。在每个实例中，这种策略可以由源 10 处的

用户、由源 10 的管理员等配置。

[0102] 结合再备份算法采用的可能的触发参数包括但不限于：

[0103] - 当预定时间段过去时激活的周期性触发器；

[0104] - 当源 10 处的卷 12 的自由空间下降至某个量以下，或者当在源 10 处卷 12 处的已用空间超过某个量时激活的空间触发器；

[0105] - 当预定时间段过去，但仅在源 10 处的自由空间下降至某个量以下或源处的已用空间超过某个量时才激活的周期性 - 空间触发器；

[0106] - 当备份过滤器 26 注意到文件系统 16 返回的完整卷错误以表示卷 12 没有剩余的自由空间时激活的完整卷触发器；

[0107] - 当预定数量的字节被下载到源 10 时激活的字节下载触发器；以及

[0108] - 由用户、管理员等激活的手动触发器。

[0109] 当然，可以基于这种单个触发器或触发器的组合发生再备份算法的触发。

[0110] 用于判定是否再备份特定文件 14 的可能的选择因素包括但不限于：

[0111] - 上次文件 14 被访问的时间，由此具有较早的访问时间的文件 14 会被优先再备份；

[0112] - 与文件 14 相关联的下载时间，由此具有较早的下载时间的文件 14 会被优先再备份；

[0113] - 文件 14 的文件大小，由此较大的文件 14 会被优先再备份；

[0114] - 文件 14 的文件类型，由此具有特定扩展的文件 14 会被优先再备份；

[0115] - 文件 14 的文件属性，由此例如，作为系统文件的文件 14 不会被再备份，而作为隐藏的文件 14 被再备份；

[0116] - 是否存在多个类似的文件 14，由此认为是相似的文件 14 会被优先再备份；

[0117] - 是否在接收器 18 处修改过文件 14，由此这种文件 14 被再备份以移除其中 可能被认为过期的数据 20；

[0118] - 文件 14 是否已经在源 10 处被修改过或被创建，由此这种文件 14 可能或者不会被选择作再备份以保持这种修改 / 创建，或者可以被选择再备份以将这种修改 / 创建复制到接收器 18；

[0119] - 在特定时间段访问文件 14 的频率；

[0120] - 在特定时间段访问文件 14 的次数；

[0121] - 诸如在相同的文件夹中、在相同的内容集合中的其他相关文件 14 的访问频率 / 访问次数 / 最近访问时间等；

[0122] - 诸如通过列表、方法、XML 文件等方式的特定文件或文件类型的外部输入被优先再备份。

[0123] 当然，再备份算法可采用这种选择因素中的单个选择因素或其组合。

[0124] 注意，被选择用于再备份的文件 14 实际上可以被再备份，或者而是仅被指定为优先再备份的候选者。尤其在后一种情况下，只有在达到特定的停止触发器时才执行再备份。如可以理解的，这种停止触发器可对应于开始再备份会话的触发参数，可对应于选择候选文件 14 以供再备份的标准，或者可以是另一触发器而不会背离本发明的精神和范围。

[0125] 也要注意，在只有达到某个停止触发器时才执行再备份的情况下，基于作为要被

再备份以消除文件14的再备份的候选者的文件14的选择标准是可行的。例如，如果基于在源10处具有至少两周的上次访问时间来选择候选文件14，则可行的是首先再备份具有大于两个月的上次访问时间的所有这种候选者，然后如果必须就再备份具有大于一个月的上次访问时间的所有候选者，然后如果必须就再备份三个星期的等，直至激活停止触发器。同样地，如果基于至少具有10兆字节的文件大小来选择候选文件14，则可行的是首先再备份文件大小大于1千兆字节的所有这种候选则，接着如果必须就再备份文件大小大于100兆字节的所有这种候选者，然后如果必须就再备份50兆字节的等，直至激活停止触发器。如可以理解的，在任何一种环境中，需要某些处理来创建候选文件14的列表，以对列表分类来生成一轮或更多轮的再备份等。

[0126] 也要注意，可以采用再备份以维护源10处和接收器18处文件14的一致性，尤其是在不同版本的文件14位于两个位置的情况下。因此，如果在源10处重建了文件14，并且修改了其数据20而没有修改接收器18处对应的数据20，则会执行再备份以将接收器18处的数据20替换成源10处的数据。同样地，如果源10处的 文件14的数据20没有被修改，而接收器18处的相应的数据20被修改，那么可以执行再备份以仅删除源10处的数据20，并期望对这种文件的稍后的重建会将接收器18处的数据20复制到源10处。当然，如果源10处的文件14的数据20被修改，并且也修改了接收器18处相应的数据20，就会存在冲突，那么因此会协商适当的冲突规则以确定是否是这样，并且如果是这样的话，如何再备份文件14。

[0127] 总而言之，现在转向图10，可以依照再备份算法执行对卷12上文件14的再备份，其中这种再备份算法由某些实体用以下方式执行。首先，再备份算法由某些事件触发（步骤1001），其中这种触发器会由实体内部地生成或者由这种实体外部地接收。在这种触发时，再备份算法于是基于某些选择标准从完全和部分重建的文件14和从未被备份的文件14中选择文件14（步骤1003）。

[0128] 此时，再备份算法可以仅从再备份选定的文件14开始（步骤1005），或而是将选定的文件14认为是可能再备份的候选文件14（步骤1007）。在后一种情况下，于是在一轮或多轮选择要再备份的候选文件14直至激活停止触发器。具体地，对于每一轮，选择一组候选文件14（步骤1009），这种选定的组被再备份（步骤1011），并且作出是否已激活停止触发器的判定。如果已经激活，则过程结束（步骤1013）。否则，过程通过返回到步骤1009处选择另一组来继续（步骤1015）。

[0129] 通用备份

[0130] 如本申请中迄今为止所描述的，在特定源10处备份的所有文件14被特别地备份，使得其数据20驻留在单个接收器18处。然而，如可以理解的，情况可能是在特定源10处每个备份文件14的数据20可以驻留在多个这种接收器18中的任何一个，如图11中所示。

[0131] 具体地，如可以理解的，本发明的备份过滤器26无需被限制到与单个备份管理器28一起工作以将来自文件14的数据20备份到单个接收器18。而是，备份过滤器实际上与多个备份管理器28一起工作，其中每个备份管理器28执行与多个接收器18中的特定的一个相关的备份功能。

[0132] 举例而言，在分支服务器10和集线器服务器18的环境中，特定的分支服务器10可以与多个集线器服务器18交互。因此，可能是与特定的分支服务器10交互的一个集线

器服务器 18 具有来自第一源的数据 20，而与这个特定的分支服务器 10 交互的另一集线器服务器 18 具有来自第二源的数据 20。同样地，例如，在计算设备 10 和可选位置 18 的环境中，特定的计算设备 10 可以将数据 20 存储在多个可选位置。因此，可能是特定的计算设备 10 的一个可选位置 18 被指定为存储来自第 一特定类型的文件 14 的数据 20，而该特定的计算设备 10 的另一可选位置被指定为存储来自第二特定类型的文件 14 的数据 20。当然，所采用的接收器 18 的数目以及在接收器 18 之间划分数据 20 的标准可以是任何适当的数目和标准，而不会背离本发明的精神和范围。

[0133] 无论如何，如果对特定的源 18 要采用多个接收器 18，那么需要一种机制来标识哪个接收器 18 具有来自源 10 的特定备份文件 14 的数据 20，并标识必须采用哪个相应的备份管理器 28 来访问接收器 18。因此，在本发明的一个实施例中，这种标识被维护在与这种特定文件 14 相关联的备份信息 24 中，如图 11 中所示。

[0134] 结果是，当遇到任何特定的备份文件 14 并从中读取备份信息 24 时，备份过滤器 26 可以从这种备份信息 24 获取要结合该特定的备份文件 14 使用的备份服务器 28 的标识，并基于此与标识的备份服务器 28 通信以适当地访问接收器 18。于是实际上，备份过滤器 26 对于所有的备份管理器 28 是通用的，并且经标识的备份管理器 28 控制或‘拥有’这种特定的备份文件 14，因为备份过滤器 26 永远不应该与和这种特定备份文件 14 相关的任何其他备份管理器 28 通信。

[0135] 假设，每个备份管理器 28 包括与对应于该备份管理器 28 的接收器 18 通信所需的所有功能和信息，以便备份过滤器 26 无需关心这些问题。于是对于备份过滤器 26，与和特定备份文件 14 相关的经标识的备份服务器 28 通信是访问相应的接收器 18 所需的全部，这种备份过滤器 26 实际上无需关心这种经标识的备份管理器 28 如何与相应的接收器 18 通信，经标识的备份管理器 28 如何定位相应的接收器 18 等。

[0136] 类似于图 4 和 5 中所示的过程，现在参考图 12，使用以下方式执行对文件 14 的通用备份和重建。假设这种备份是由具有特定标识 (ID) 的特定备份管理器 28 开始的，并且实际上由备份过滤器 26 执行，这种备份管理器 28 因此将适当的备份请求发送给具有 ID 的备份过滤器 26 和要备份的文件 14 的标识 (步骤 1201)，此后备份过滤器 26 按需将数据 20 从 14 处移除 (步骤 1203)。备份过滤器 26 然后通过其 ID 将这种按需移除的数据 20 转发给请求的备份管理器 28(步骤 1205)，备份管理器 28 于是基于可以这样做并包含在这种备份管理器 28 中的任何协议将这种已移除的数据转发给相应的接收器 18。

[0137] 注意在根据接收器 18 和文件 14 的类型这点上，备份管理器 28 实际上选择不将移除的数据发送给接收器 18，诸如如果接收器 18 已只读方式将在讨论的文件 14 复制到源 10。当然，在这种情况下，可能是来自备份管理器 28 的备份请求通知备 份过滤器实际上不执行步骤 1205，因为没有必要。

[0138] 无论如何，如上所述，备份过滤器 26 通过设置‘备份’属性并添加备份信息 24 来修改现在备份的文件的元数据 22(步骤 1207)。重要的是，这种备份信息应该包括备份服务器 28 的 ID，以供以后重建备份文件 14 时使用。因此，在稍后某个时间，当备份服务器 28 或应用程序 30 请求访问备份文件 14 的数据 20 时，该请求最终到达备份过滤器 26，如上所述 (步骤 1209)，并且备份过滤器 26 定位文件 14 的元数据 22 中的备份信息 24(步骤 1211)。

[0139] 再者，基于经标识的备份信息 24 来定位存储在接收器 18 处的文件 14 的数据 20，

虽然在该实例中备份过滤器 26 首先在备份信息 24 中定位负责备份文件 14 的备份服务器 28 的 ID(步骤 1213),且使用这种 ID 将请求发送给相应的备份管理器 28,以实际上从相应的接收器 18 获取数据 20(步骤 1215)。假设,备份管理器 28 实际上从接收器 18 获取数据 20,并将其提供给备份过滤器 26(步骤 1217),备份过滤器于是将数据 20 重建为在讨论的文件 14(步骤 1219)。

[0140] 结论

[0141] 实现结合本发明所执行的过程所需的编程是相对简单的,并且对于程序设计公众应该是显而易见的。因此,这种编程没有附在这里。那么,可用采用任何特定的编程来实现本发明而不会背离其精神和范围。

[0142] 在以上描述中,可以看到本发明包括新颖而有用的方法和机制,使用这种方法和机制可以复制或备份位于诸如本地卷 12、计算设备 10 或分支服务器 10 的源 10 处的文件 14,以便其数据 20 被存储在诸如可选位置 18 或集线器服务器 18 的接收器 18 处,因此,如果需要的话,源处的文件 14 是可以被重建的减小或备份的形式。可按需形成和重建经备份的文件 14。

[0143] 应该理解,可以对上述实施例作出改变而不会背离其发明性的理念。于是一般地,应该理解,本发明因此不限于所揭示的特定实施例,它旨在覆盖由所附权利要求所定义的本发明的精神和范围内的修改。

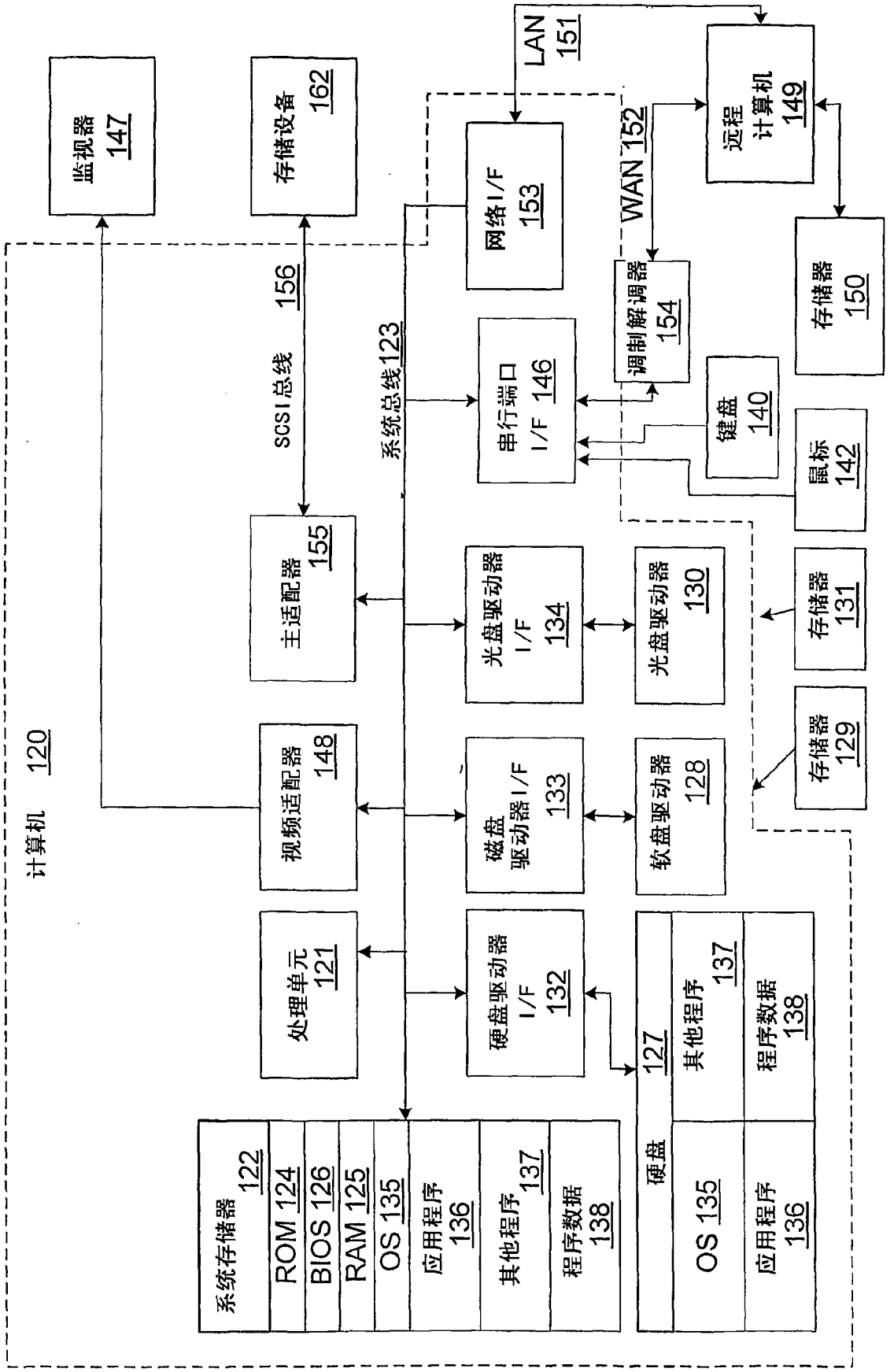

计算机 120
监视器 147
存储设备 162
LAN 151
远程计算机 149
网络 I/F 153
WAN 152
存储器 150
SCSI 总线 156
系统总线 123
调制解调器 154
串行端口 I/F 146
键盘 140
鼠标 142
主适配器 155
光盘驱动器 I/F 134
光盘驱动器 130
存储器 131
视频适配器 148
磁盘驱动器 I/F 133
软盘驱动器 128
存储器 129
处理单元 121
硬盘驱动器 I/F 132
硬盘 127
其他程序 137
程序数据 138
系统存储器 122
ROM 124
BIOS 126
RAM 125
OS 135
应用程序 136
其他程序 137
程序数据 138
硬盘
OS 135
应用程序 136

图 1

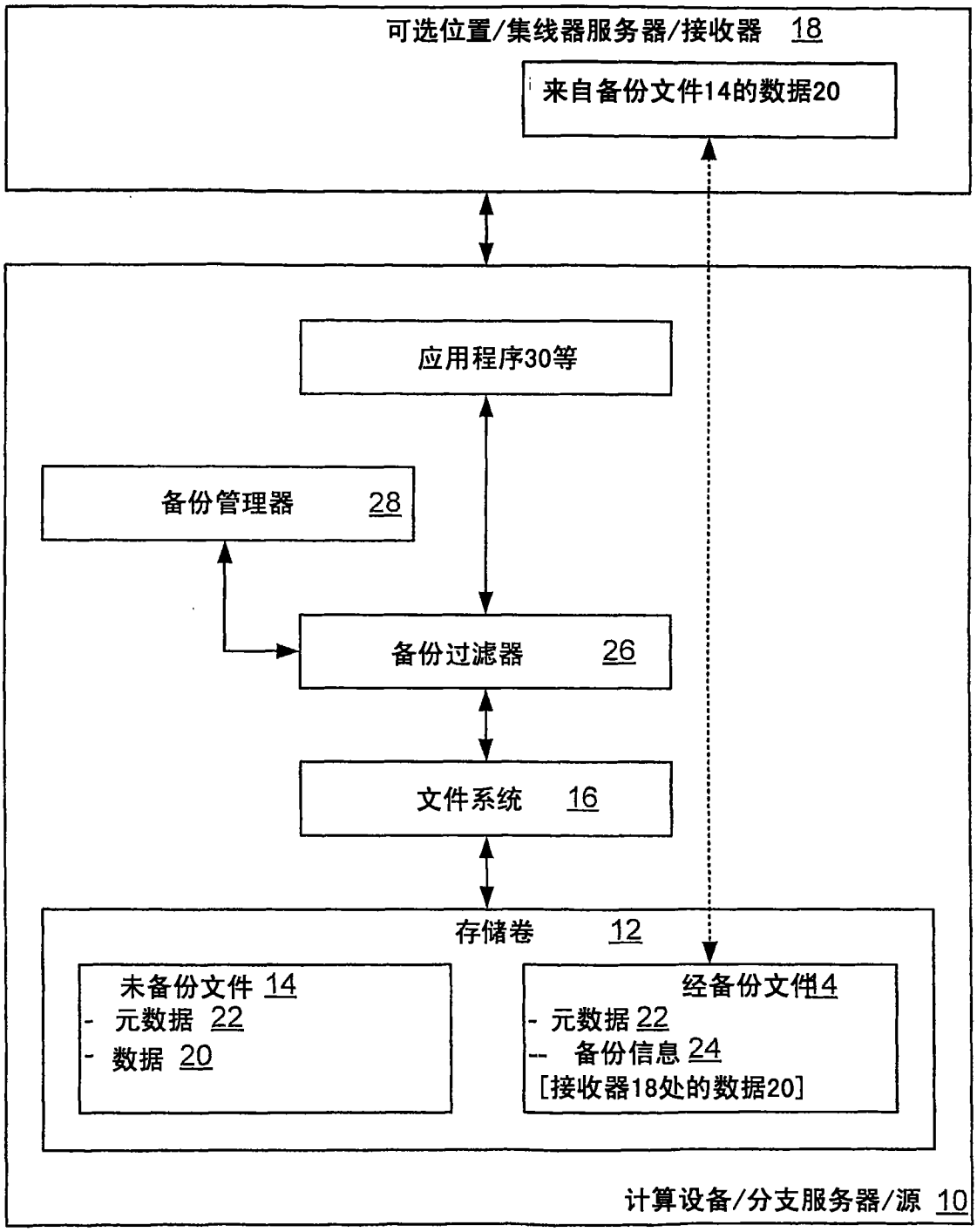

可选位置/集线器服务器/接收器 18
来自备份文件14的数据20
应用程序30等
备份管理器 28
备份过滤器 26
文件系统 16
存储卷 12
未备份文件 14
- 元数据 22
- 数据 20
经备份文件14
- 元数据 22
-- 备份信息 24
[接收器18处的数据20]
计算设备/分支服务器/源 10

图 2

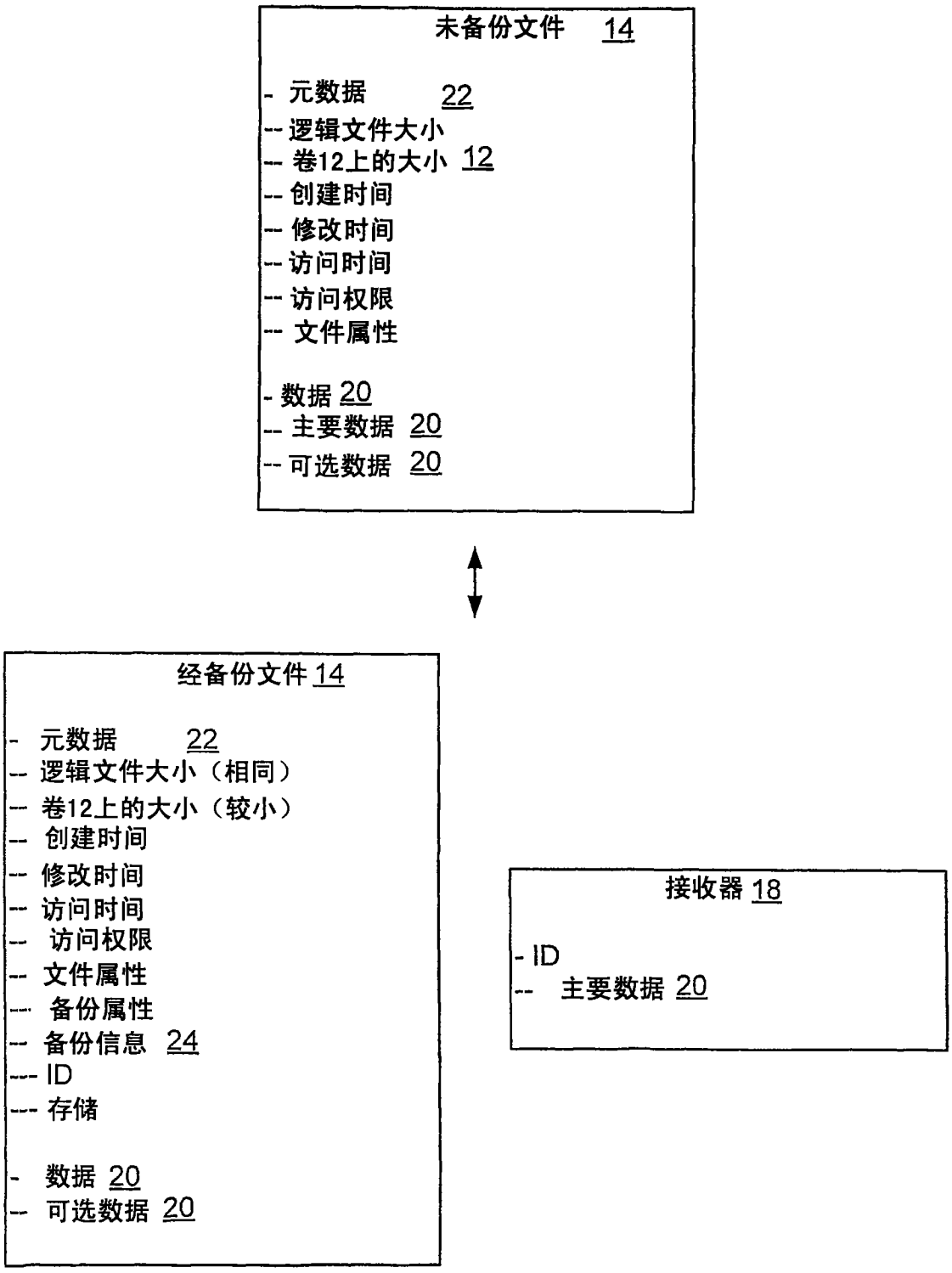

未备份文件 14
- 元数据 22
-- 逻辑文件大小
-- 卷12上的大小 12
-- 创建时间
-- 修改时间
-- 访问时间
-- 访问权限
-- 文件属性
- 数据 20
-- 主要数据 20
-- 可选数据 20
经备份文件 14
- 元数据 22
-- 逻辑文件大小（相同）
-- 卷12上的大小（较小）
-- 创建时间
-- 修改时间
-- 访问时间
-- 访问权限
-- 文件属性
--- 备份属性
-- 备份信息 24
--- ID
--- 存储
- 数据 20
-- 可选数据 20
接收器 18
- ID
-- 主要数据 20

图 3A

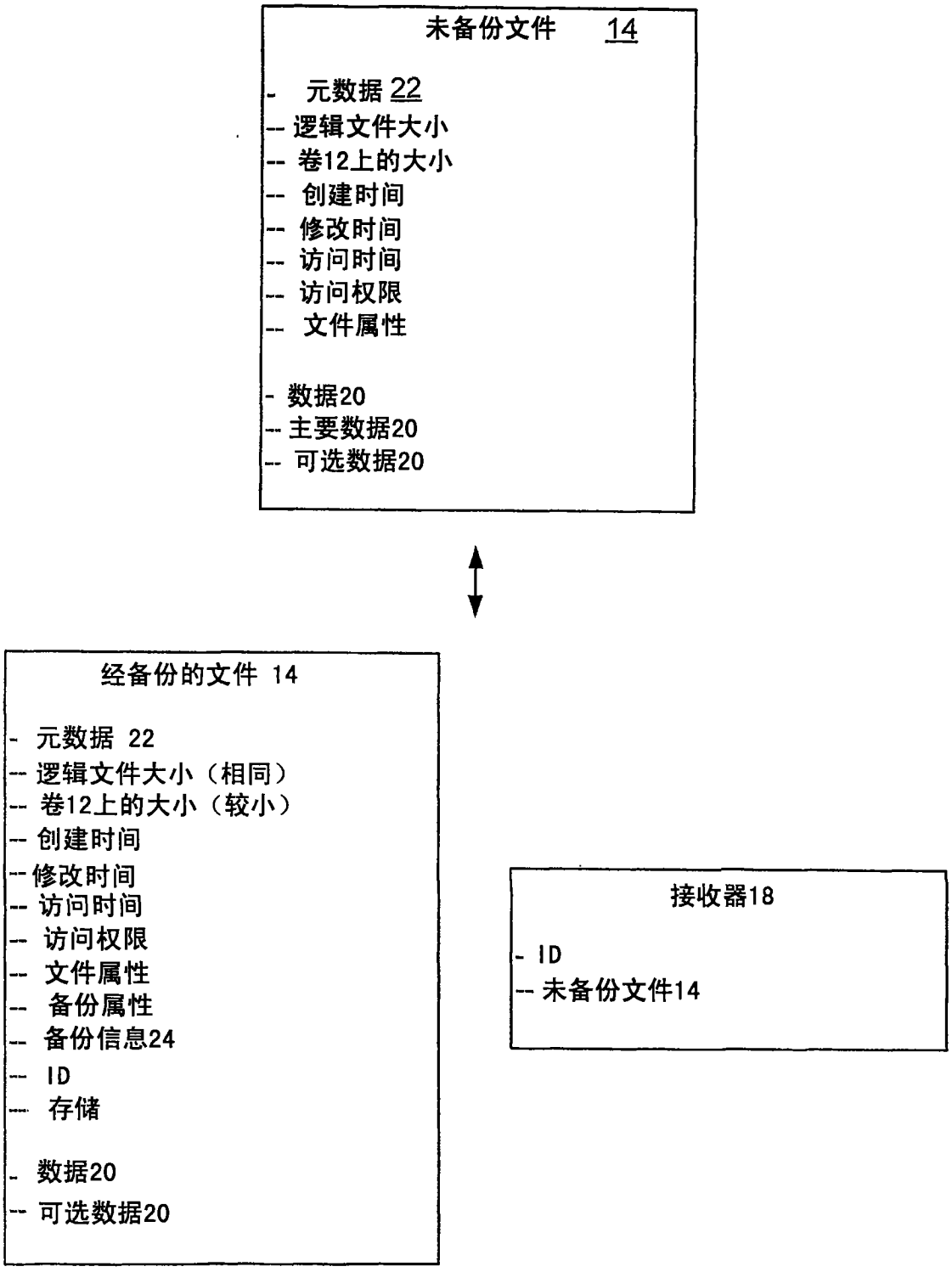

未备份文件 14
- 元数据 22
-- 逻辑文件大小
-- 卷12上的大小
-- 创建时间
-- 修改时间
-- 访问时间
-- 访问权限
-- 文件属性
- 数据20
-- 主要数据20
-- 可选数据20
经备份的文件 14
- 元数据 22
-- 逻辑文件大小（相同）
-- 卷12上的大小（较小）
-- 创建时间
-- 修改时间
-- 访问时间
-- 访问权限
-- 文件属性
-- 备份属性
-- 备份信息24
-- ID
-- 存储
- 数据20
-- 可选数据20
接收器18
- ID
-- 未备份文件14

图 3B

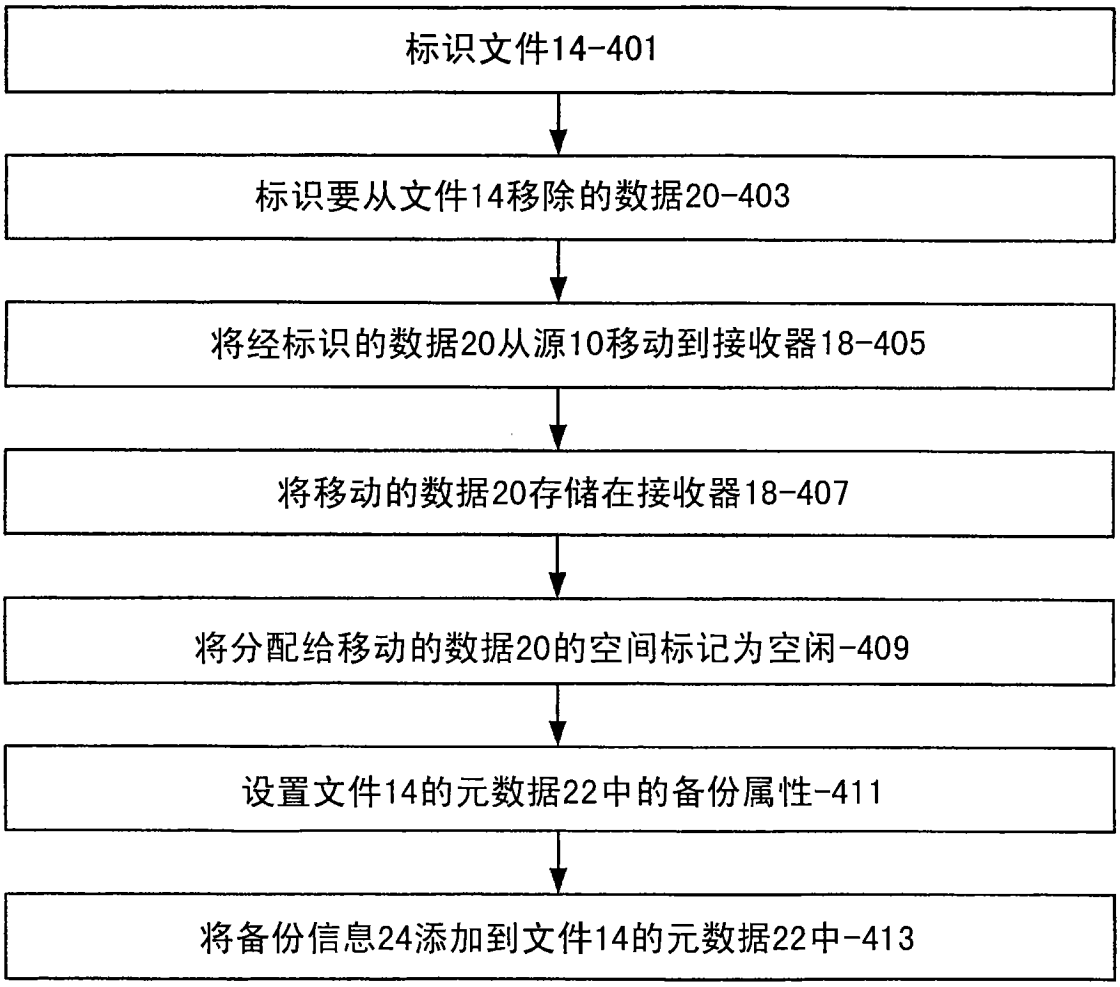

标识文件14-401
标识要从文件14移除的数据20-403
将经标识的数据20从源10移动到接收器18-405
将移动的数据20存储在接收器18-407
将分配给移动的数据20的空间标记为空闲-409
设置文件14的元数据22中的备份属性-411
将备份信息24添加到文件14的元数据22中-413

图 4

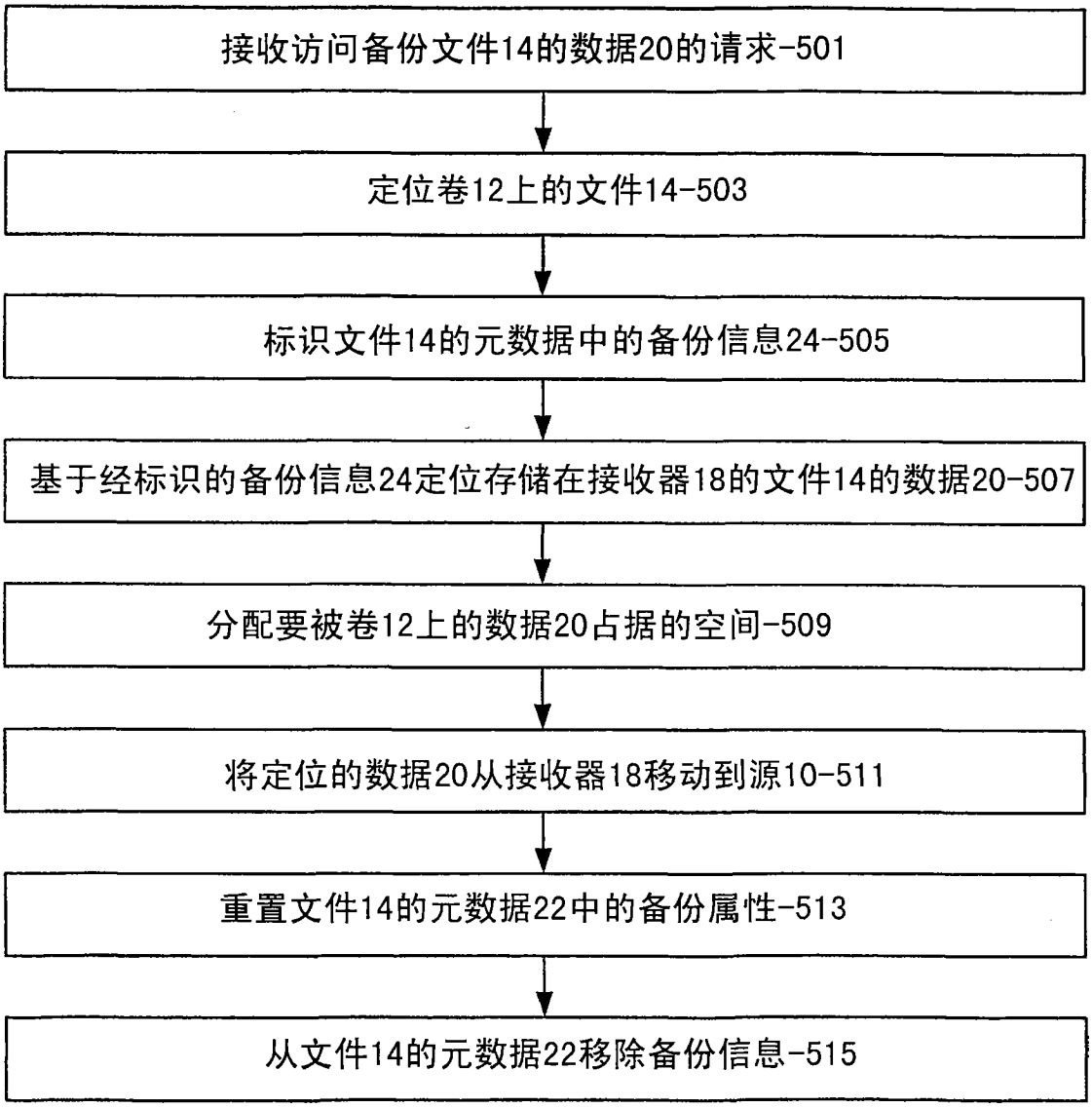

接收访问备份文件14的数据20的请求-501
定位卷12上的文件14-503
标识文件14的元数据中的备份信息24-505
基于经标识的备份信息24定位存储在接收器18的文件14的数据20-507
分配要被卷12上的数据20占据的空间-509
将定位的数据20从接收器18移动到源10-511
重置文件14的元数据22中的备份属性-513
从文件14的元数据22移除备份信息-515

图 5

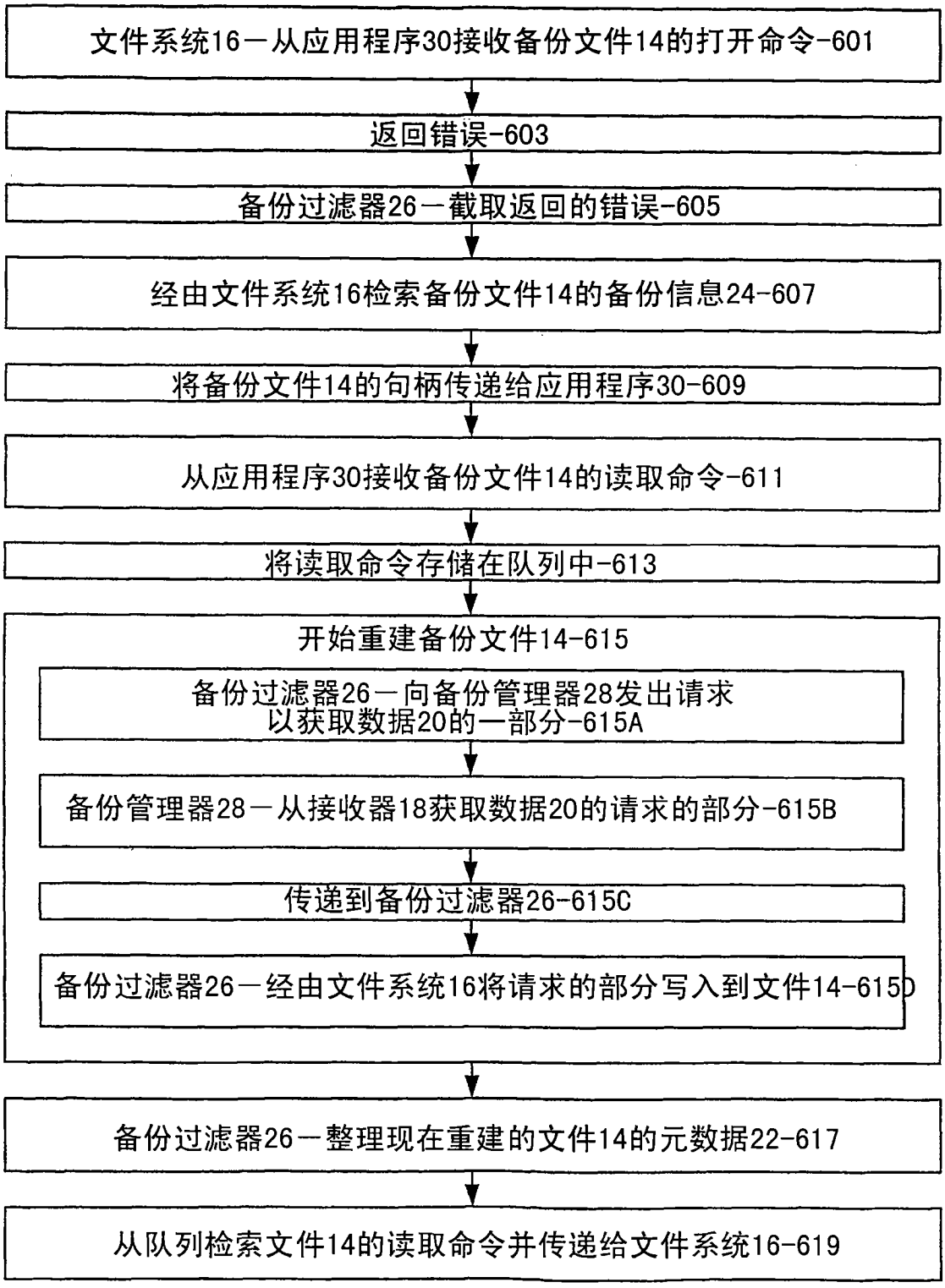

文件系统16－从应用程序30接收备份文件14的打开命令-601
返回错误-603
备份过滤器26－截取返回的错误-605
经由文件系统16检索备份文件14的备份信息24-607
将备份文件14的句柄传递给应用程序30-609
从应用程序30接收备份文件14的读取命令-611
将读取命令存储在队列中-613
开始重建备份文件14-615
备份过滤器26－向备份管理器28发出请求以获取数据20的一部分-615A
备份管理器28－从接收器18获取数据20的请求的部分-615B
传递到备份过滤器26-615C
备份过滤器26－经由文件系统16将请求的部分写入到文件14-615D
备份过滤器26－整理现在重建的文件14的元数据22-617
从队列检索文件14的读取命令并传递给文件系统16-619

图 6

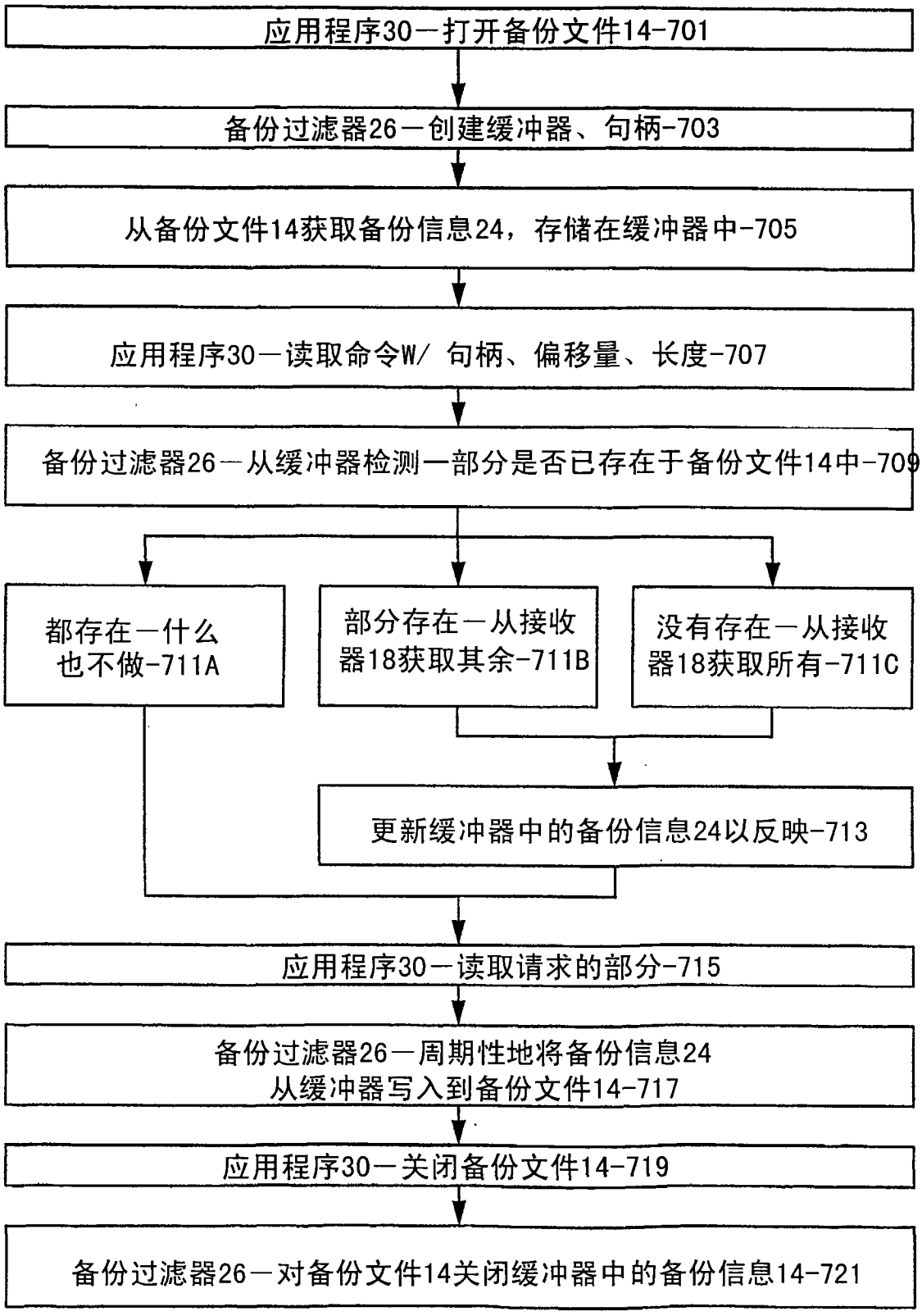

应用程序30－打开备份文件14-701
备份过滤器26－创建缓冲器、句柄-703
从备份文件14获取备份信息24，存储在缓冲器中-705
应用程序30－读取命令W/ 句柄、偏移量、长度-707
备份过滤器26－从缓冲器检测一部分是否已存在于备份文件14中-709
都存在－什么也不做-711A
部分存在－从接收器18获取其余-711B
没有存在－从接收器18获取所有-711C
更新缓冲器中的备份信息24以反映-713
应用程序30－读取请求的部分-715
备份过滤器26－周期性地将备份信息24从缓冲器写入到备份文件14-717
应用程序30－关闭备份文件14-719
备份过滤器26－对备份文件14关闭缓冲器中的备份信息14-721

图 7

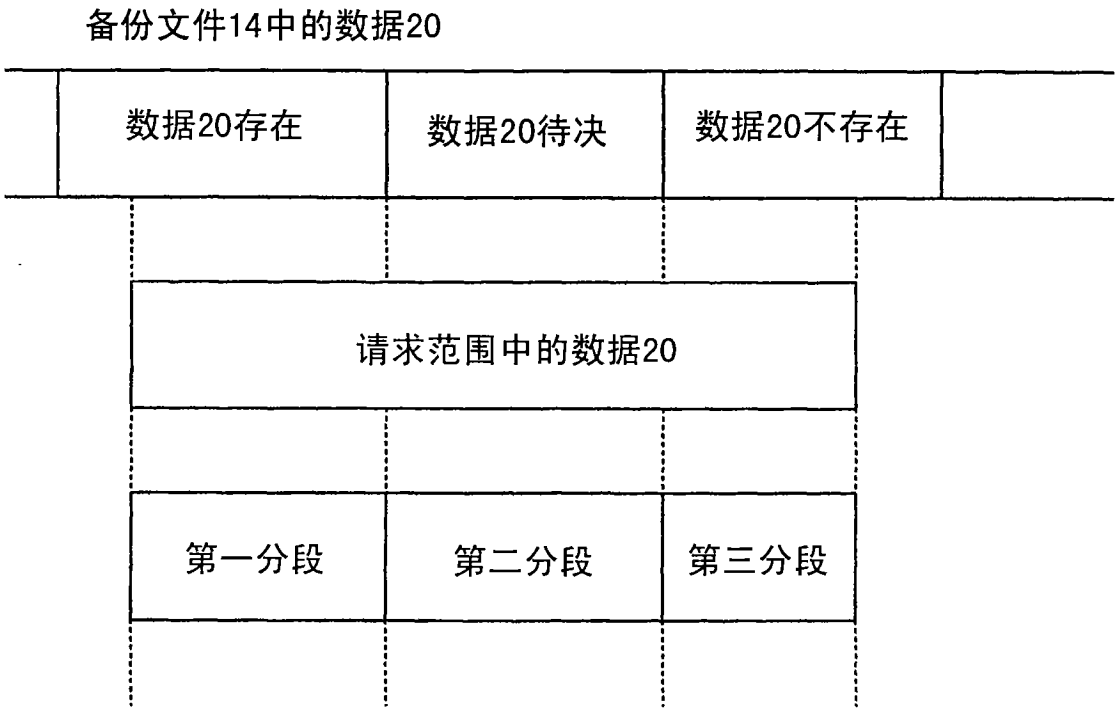

备份文件14中的数据20
数据20存在
数据20待决
数据20不存在
请求范围中的数据20
第一分段
第二分段
第三分段

图 8

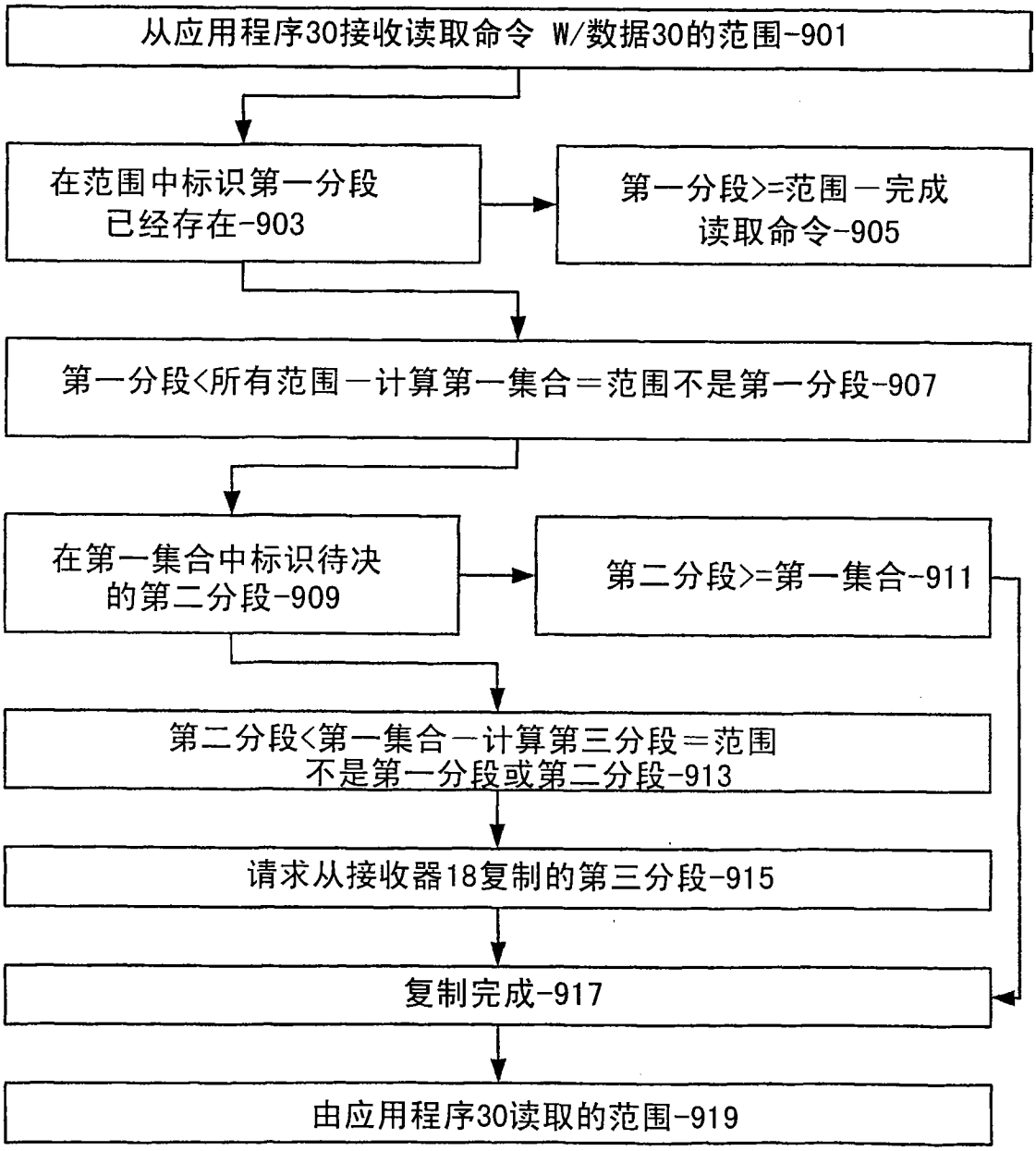

从应用程序30接收读取命令 W/数据30的范围-901
在范围中标识第一分段已经存在-903
第一分段>=范围－完成读取命令-905
第一分段<所有范围－计算第一集合＝范围不是第一分段-907
在第一集合中标识待决的第二分段-909
第二分段>=第一集合-911
第二分段<第一集合－计算第三分段＝范围不是第一分段或第二分段-913
请求从接收器18复制的第三分段-915
复制完成-917
由应用程序30读取的范围-919

图 9

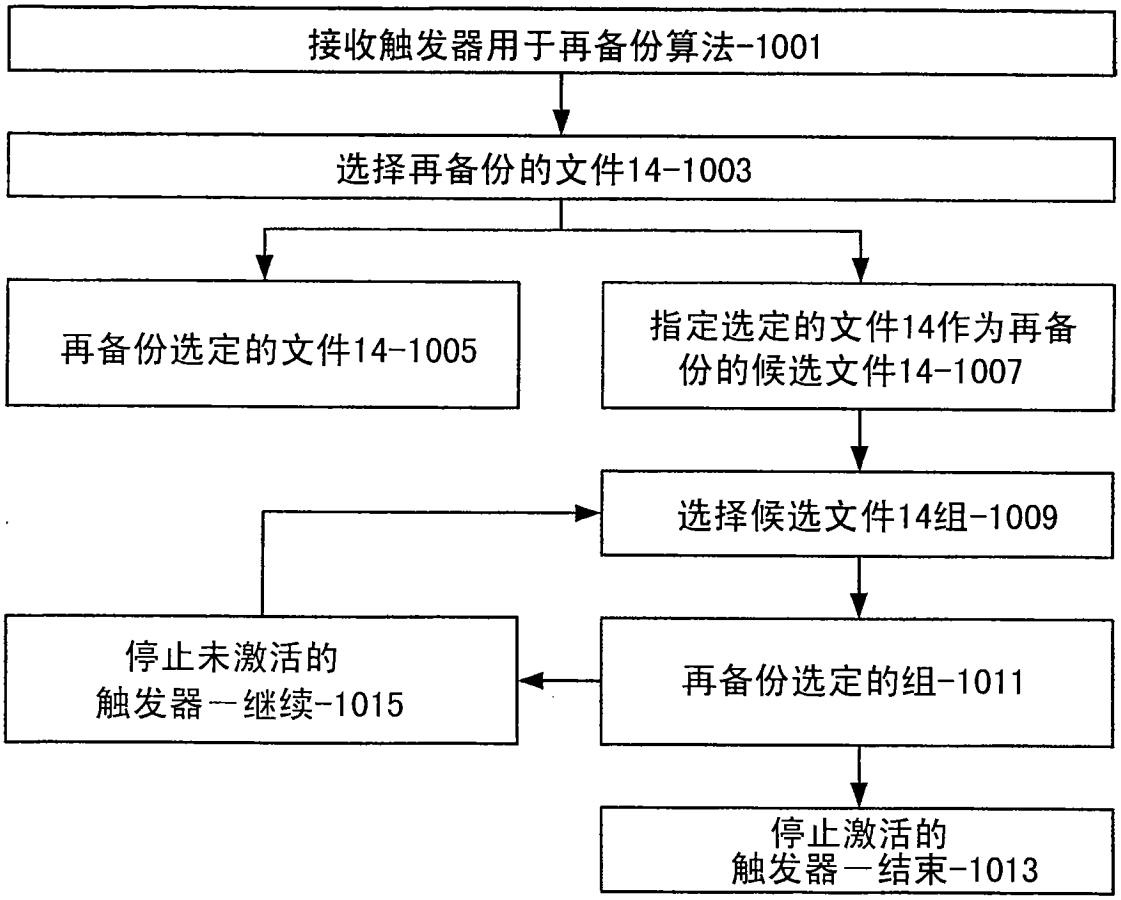

接收触发器用于再备份算法-1001
选择再备份的文件14-1003
再备份选定的文件14-1005
指定选定的文件14作为再备份的候选文件14-1007
选择候选文件14组-1009
停止未激活的触发器—继续-1015
再备份选定的组-1011
停止激活的触发器—结束-1013

图 10

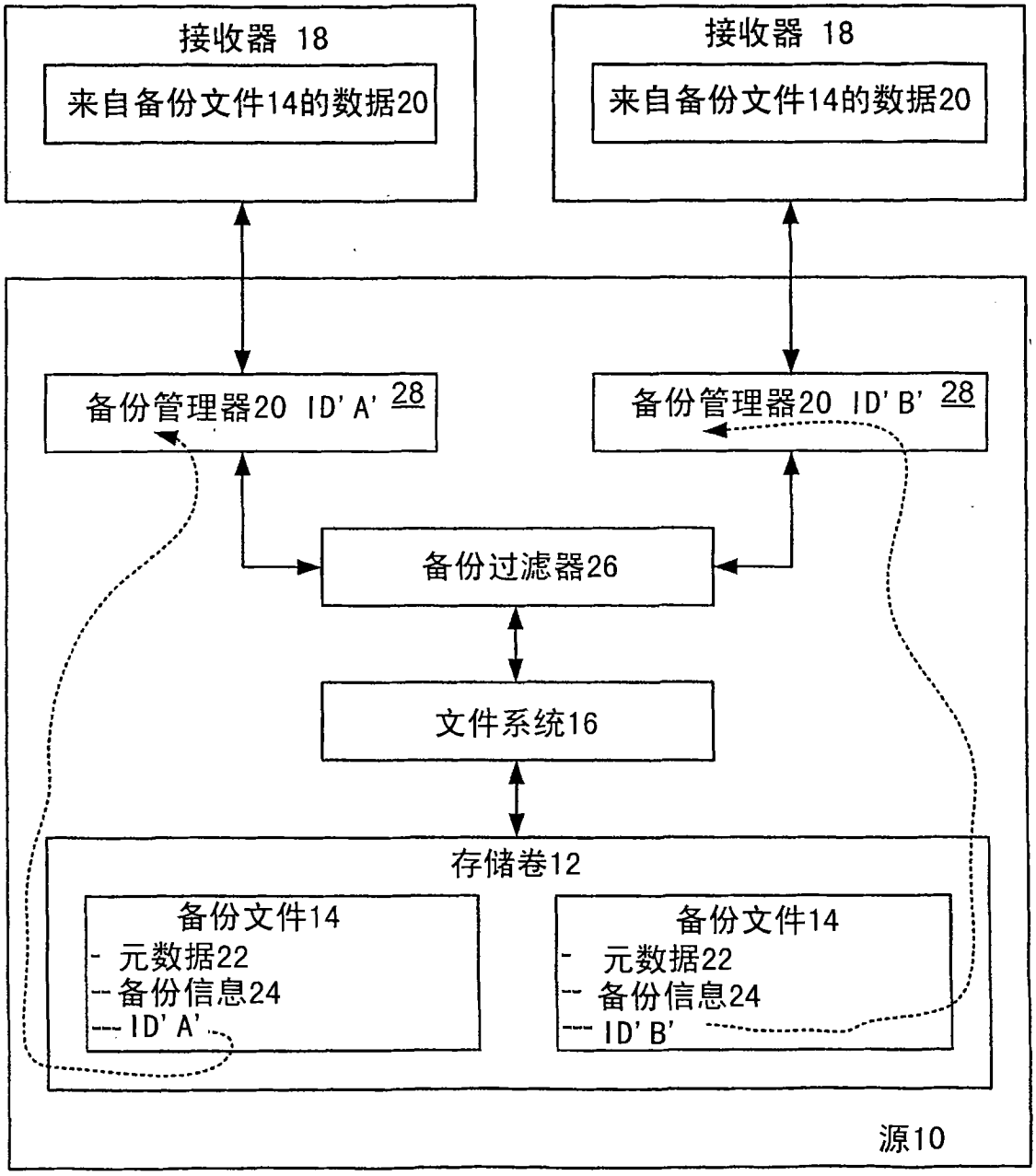

接收器 18
来自备份文件14的数据20
接收器 18
来自备份文件14的数据20
备份管理器20 ID'A' 28
备份管理器20 ID'B' 28
备份过滤器26
文件系统16
存储卷12
备份文件14
- 元数据22
-- 备份信息24
--- ID'A'
备份文件14
- 元数据22
-- 备份信息24
--- ID'B'
源10

图 11

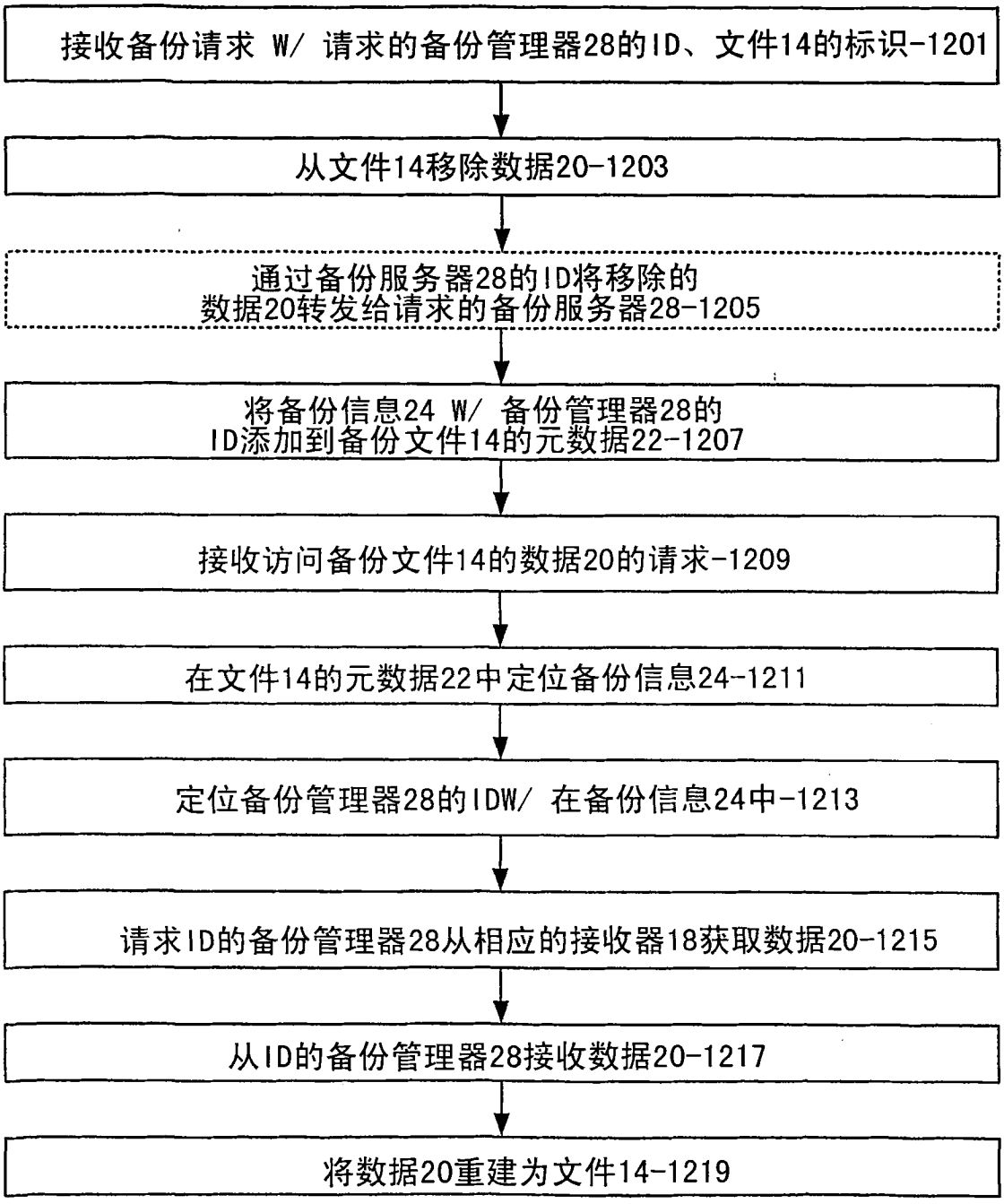

接收备份请求 W/ 请求的备份管理器28的ID、文件14的标识-1201
从文件14移除数据20-1203
通过备份服务器28的ID将移除的
数据20转发给请求的备份服务器28-1205
将备份信息24 W/ 备份管理器28的
ID添加到备份文件14的元数据22-1207
接收访问备份文件14的数据20的请求-1209
在文件14的元数据22中定位备份信息24-1211
定位备份管理器28的IDW/ 在备份信息24中-1213
请求ID的备份管理器28从相应的接收器18获取数据20-1215
从ID的备份管理器28接收数据20-1217
将数据20重建为文件14-1219

图 12