



(12) 发明专利

(10) 授权公告号 CN 113448500 B

(45) 授权公告日 2024.01.05

(21) 申请号 202010810453.6

(51) Int.CI.

(22) 申请日 2020.08.13

G06F 3/06 (2006.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 113448500 A

(56) 对比文件

CN 103384876 A, 2013.11.06

(43) 申请公布日 2021.09.28

CN 105593804 A, 2016.05.18

(30) 优先权数据

US 2008155214 A1, 2008.06.26

2020-055648 2020.03.26 JP

US 2009063556 A1, 2009.03.05

(73) 专利权人 株式会社日立制作所

US 2009300081 A1, 2009.12.03

地址 日本东京都

US 2019171435 A1, 2019.06.06

审查员 潘芳芳

(72) 发明人 野村镇平 早坂光雄 高田昌忠

(74) 专利代理机构 北京尚诚知识产权代理有限公司 11322

专利代理人 龙淳 徐飞跃

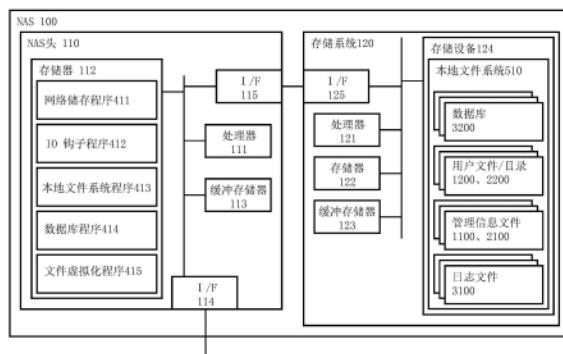
权利要求书2页 说明书13页 附图21页

(54) 发明名称

文件存储系统和文件存储系统的管理方法

(57) 摘要

本发明提供一种文件存储系统和文件存储系统的管理方法。NAS(100)的处理器(111)基于来自应用程序的对文件的操作请求,进行本地文件系统(510)的调用处理,本地文件系统(510)处理对文件的操作请求,I/O钩子程序(412)基于操作请求所涉及的对本地文件系统(510)的输入信息或操作内容,进行文件的状态管理信息的更新处理,文件虚拟化程序(415)基于状态管理信息在NAS(100)与CAS之间进行文件的管理处理。根据本发明,能够不受文件系统的版本升级影响地提供文件虚拟化功能。



1. 一种文件存储系统,其具有提供给应用程序的第一文件系统、能够用所述第一文件系统来存储文件的第一存储系统和处理器,且所述文件存储系统能够使用第二存储系统,该文件存储系统的特征在于:

所述第一文件系统利用具有目录名和文件名的层级结构来管理所述第一存储系统的文件,

所述第一存储系统和所述第二存储系统存储具有相同的识别符的文件,

所述文件存储系统包括:

用户目录,其由所述第一文件系统来管理,在该用户目录中记载了所述文件的目录名和文件名,其中,所述文件是可变更的;

状态信息管理部;

状态管理信息,其由所述状态信息管理部来管理,与所述用户目录相关联,对于存储于所述第二存储系统中的具有相同的识别符的文件,保存有所述第一存储系统的所述文件的状态以及与该相同的识别符不同的识别符;和

管理所述第一存储系统和第二存储系统所存储的文件的文件虚拟化部,

所述处理器基于来自所述应用程序的对所述文件的操作请求,进行所述第一文件系统的调用处理,

所述第一文件系统处理对所述文件的操作请求,以操作存储于所述第一存储系统中的文件,

所述状态信息管理部基于所述操作请求所涉及的对所述第一文件系统的操作内容,进行更新具有与所述第一存储系统的所述文件相同的识别符的所述第二存储系统的所述文件的状态的状态管理信息更新处理,

所述文件虚拟化部基于所述状态管理信息中所包含的所述文件的状态和不同的识别符,进行将对所述第一存储系统的操作反映到所述第二存储系统的具有相同的识别符的所述文件上的文件管理处理。

2. 如权利要求1所述的文件存储系统,其特征在于:

所述文件的管理处理是所述第一存储系统与第二存储系统之间的文件的迁移。

3. 如权利要求2所述的文件存储系统,其特征在于:

在第二存储系统中能够用第二文件系统来存储文件,

第一文件系统具有层级结构,第二文件系统不具有层级结构,

所述文件虚拟化部在第一文件系统与第二文件系统之间进行所述文件的管理处理。

4. 如权利要求1所述的文件存储系统,其特征在于:

所述状态信息管理部进行所述文件的状态管理信息的更新处理,直到要进行所述第一文件系统的调用处理为止。

5. 如权利要求1所述的文件存储系统,其特征在于:

所述状态信息管理部不仅进行所述状态管理信息的更新,而且还创建所述操作请求的日志,

所述文件由不同的识别符来识别,在所述日志中,所述不同的识别符在从所述文件的生成到删除为止的期间没有变更,

所述文件虚拟化部基于所述操作请求的日志来进行所述文件的管理处理。

6. 如权利要求5所述的文件存储系统,其特征在于:

所述层级结构的目录名和文件名是可变更的,

在所述状态管理信息中,所述文件由不同的识别符来识别。

7. 如权利要求1所述的文件存储系统,其特征在于:

所述文件虚拟化部在所述第一存储系统从异常结束后进行复原时,为了更新所述状态管理信息,而使用所述操作请求的日志来进行所述文件的爬虫处理。

8. 如权利要求1所述的文件存储系统,其特征在于,包括:

第一计算机系统,其具有所述第一文件系统、所述第一存储系统、所述处理器、所述状态信息管理部、所述状态管理信息和所述文件虚拟化部;和

第二计算机系统,其具有所述第二存储系统,进行所述第一计算机系统所存储的文件的备份。

9. 如权利要求8所述的文件存储系统,其特征在于:

所述状态信息管理部在所述状态管理信息中管理所述文件的每个部分的状态。

10. 如权利要求8所述的文件存储系统,其特征在于,包括:

所述层级结构的目录名和文件名是可变更的,

所述状态信息管理部在所述第二计算机系统中备份所述文件时,用不同的识别符来管理所述文件。

11. 一种文件存储系统的管理方法,所述文件存储系统具有提供给应用程序的第一文件系统、能够用所述第一文件系统来存储文件的第一存储系统和处理器,且能够使用第二存储系统,所述文件存储系统的管理方法的特征在于,使所述文件存储系统执行如下处理,即:

基于来自所述应用程序的对所述文件的操作请求,进行所述第一文件系统的调用处理,

处理对所述文件的操作请求,来操作存储在所述第一存储系统中的文件,

基于所述操作请求所涉及的对所述第一文件系统的操作内容,进行更新具有与所述第一存储系统的所述文件相同的识别符的所述第二存储系统所保存的所述文件的状态的状态管理信息的更新处理,其中,所述状态管理信息与用户目录相关联,并且保存有所述文件的状态以及与该相同的识别符不同的识别符,

基于所述状态管理信息中的所述文件的状态和不同的识别符,进行将对所述第一存储系统的操作反映到所述第二存储系统的具有相同的识别符的所述文件上的文件管理处理。

文件存储系统和文件存储系统的管理方法

技术领域

[0001] 本发明涉及文件存储系统和文件存储系统的管理方法。

背景技术

[0002] 数字数据特别是文件数据的数据量正在急速增大。NAS (Network Attached Storage: 网络附加储存) 是适合于大量的计算机经由网络来共享文件数据的储存设备。当前,文件数据储存器多使用NAS设备。

[0003] 包含数据文件在内,数字数据例如为了应对各种法令要求,需要为了多种多样的目的而长时间地保存。CAS (Content Addressed Storage: 内容寻址储存) 保证数据的不变性,提供长期的数据存档用的解决方案。一般而言,现用数据是只要正在被使用就保存在NAS设备中,然后,为了存档而转移至CAS设备。

[0004] 已知有如下系统,即:在数据中心配置CAS设备,在各据点(例如,公司的各事业部)配置NAS设备,将CAS设备与NAS设备通过WAN (Wide Area Network: 广域网) 等通信网络连接,在CAS设备上进行数据的集中管理。

[0005] 管理文件数据储存的存储系统对操作文件的客户提供文件系统,并且,将NAS设备中存储的文件在CAS设备中适当地备份。在存储系统所提供的备份功能中,具有检测NAS设备中所生成/更新了的文件并将其非同步地迁移至CAS设备的功能、将未被客户访问的文件从NAS设备删除的存根化功能、和在被客户再次参照时从CAS设备取得文件的恢复功能。以下,在本说明书中,将存储系统提供的迁移功能、存根化功能和恢复功能总称为文件虚拟化功能。

[0006] 作为本技术领域的背景技术,有日本特表2013-524358号(专利文献1)。在该公报中公开有如下方法,即:保持文件系统中的文件操作履历的日志信息,基于该信息来确定文件虚拟化功能的对象数据,判断文件是否需要备份和是否可以存根化。

[0007] 现有技术文献

[0008] 专利文献

[0009] 专利文献1:日本特表2013-524358号公报。

发明内容

[0010] 发明所要解决的技术问题

[0011] 在存储系统中,用于构建文件系统的程序有时由OSS (Open SourceSoftware: 开放源代码软件) 提供。OSS比较频繁地进行版本升级,而且其版本升级的时间不定。因此,为了持续地对存储系统提供文件虚拟化功能,需要在每次OSS的版本升级时进行与该版本升级相伴的文件虚拟化功能的更新。该更新耗时巨大。

[0012] 本发明是鉴于上述的问题而完成的,其目的在于提供能够不受文件系统的版本升级影响地提供文件虚拟化功能的文件存储系统和文件存储系统的管理方法。

[0013] 用于解决问题的技术手段

[0014] 为了解决上述问题,本发明的一个方面的文件存储系统具有通过应用程序来提供的第一文件系统、能够用第一文件系统来存储文件的第一存储系统和处理器,且能够使用第二存储系统,该文件存储系统的特征在于,包括:保存有文件的状态的状态管理信息;管理状态管理信息的状态信息管理部;和管理第一存储系统和第二存储系统所存储的文件的文件虚拟化部,处理器基于来自应用程序的对文件的操作请求,进行第一文件系统的调用处理,第一文件系统处理对文件的操作请求,状态信息管理部基于操作请求所涉及的对第一文件系统的输入信息或操作内容,进行文件的状态管理信息的更新处理,文件虚拟化部基于状态管理信息,在第一存储系统与第二存储系统之间进行文件的管理处理。

[0015] 发明的效果

[0016] 根据本发明,能够实现可不受文件系统的版本升级影响地提供文件虚拟化功能的文件存储系统和文件存储系统的管理方法。

附图说明

- [0017] 图1是表示实施例的文件存储系统的硬件结构的图。
- [0018] 图2是表示实施例的文件存储系统的NAS的概略结构的一个例子的图。
- [0019] 图3是表示实施例的文件存储系统的CAS的概略结构的一个例子的图。
- [0020] 图4是用于说明实施例的文件存储系统的I0钩子(Hook)程序的功能的图。
- [0021] 图5是用于说明实施例的文件存储系统所提供的文件系统的图。
- [0022] 图6是表示实施例的文件存储系统的管理信息文件的一个例子的图。
- [0023] 图7是表示实施例的文件存储系统的管理信息文件的其它例子的图。
- [0024] 图8是表示实施例的文件存储系统的日志文件的一个例子的图。
- [0025] 图9是表示实施例的文件存储系统的数据库的一个例子的图。
- [0026] 图10是用于说明实施例的文件存储系统的文件/目录创建处理的一个例子的流程图。
- [0027] 图11是用于说明实施例的文件存储系统的文件/目录删除处理的一个例子的流程图。
- [0028] 图12是用于说明实施例的文件存储系统的重命名处理的一个例子的流程图。
- [0029] 图13是用于说明实施例的文件存储系统的文件写处理的一个例子的流程图。
- [0030] 图14是用于说明实施例的文件存储系统的文件读处理的一个例子的流程图。
- [0031] 图15是用于说明实施例的文件存储系统的目录读处理的一个例子的流程图。
- [0032] 图16是用于说明实施例的文件存储系统的日志反映处理的一个例子的流程图。
- [0033] 图17是用于说明实施例的文件存储系统的文件迁移处理的一个例子的流程图。
- [0034] 图18是用于说明实施例的文件存储系统的目录迁移处理的一个例子的流程图。
- [0035] 图19是用于说明实施例的文件存储系统的文件存根化处理的一个例子的流程图。
- [0036] 图20是用于说明实施例的文件存储系统的CAS侧文件/目录删除处理的一个例子的流程图。
- [0037] 图21是用于说明实施例的文件存储系统的爬虫处理的一个例子的流程图。
- [0038] 附图标记的说明
- [0039] 1:文件存储系统,10-1:站点,10-2:站点,20:数据中心,30:网络,100:NAS,200:

CAS,411、421:网络储存程序,412:钩子程序,413、422:本地文件系统程序,414:数据库程序,415、423:文件虚拟化程序,510:本地文件系统,520:本地文件系统,530:网络文件系统,600:客户,1100、2100:管理信息文件,1200:用户文件,2200:用户目录,2300: CAS用目录。

具体实施方式

[0040] 以下,参照附图说明本发明的实施方式。以下的记载和图面是用于说明本发明的例示,为了说明的明确化而适当地省略和简化。本发明还能够以其它各种方式实施。只要没有特别限定,各构成要素既可以是单数也可以是复数。

[0041] 另外,在说明实施例的图中,在具有相同的功能之处标注相同的附图标记,省略其重复的说明。

[0042] 图中所示的各构成要素的位置、大小、形状、范围等,为了发明的易于理解,有时不表示实际的位置、大小、形状,范围等。因此,本发明并不一定限定于图中所示的位置、大小、形状、范围等。

[0043] 在以下的说明中,有时以“表格”、“列表”、“队列”等表达方式说明各种信息,各种信息也可以以这些以外的数据结构表达。为了表示并不依赖于数据结构,有时将“XX表格”、“XX列表”等称为“XX信息”。在对识别信息进行说明时,使用“识别信息”、“识别符”、“名称”、“ID”、“序号”等表达方式,它们也能够相互替换。

[0044] 此外,在以下的说明中,各表格的结构为一个例子,1个表格可以分割为2个以上的表格,2个以上的表格的全部或一部分也可以为1个表格。

[0045] 在具有同一或者同样的功能的构成要素存在多个的情况下,有时在同一附图标记标注不同的下标进行说明。不过,在不需要对该多个构成要素进行区别的情况下,有时会省略下标进行说明。

[0046] 此外,在以下的说明中,有时说明执行程序而进行的处理,而程序是通过由处理器(例如CPU、GPU)执行而适当地使用存储资源(例如存储器)和/或接口设备(例如通信端口)等进行确定的处理,因此处理的主体也可以为处理器。同样,执行程序而进行的处理的主体也可以为具有处理器的控制器、装置、系统、计算机、节点。执行程序而进行的处理的主体只要是运算部即可,也可以包含进行特定的处理的专用电路(例如FPGA、ASIC)。

[0047] 此外,在以下的说明中,“处理器(部)”为1个以上处理器。至少1个处理器典型的是CPU(Central Processing Unit:中央处理器)那样的微处理器,不过也可以是GPU(Graphics Processing Unit:图形处理器)那样的其它处理器。至少1个处理器既可以为单核也可以为多核。

[0048] 此外,至少1个处理器也可以是进行处理的一部分或全部的硬件电路(例如FPGA(Field-Programmable Gate Array:现场可编程门阵列)或ASIC(Application Specific Integrated Circuit:专用集成电路))之类的广义的处理器。

[0049] 在以下的说明中,“接口(部)”为1个以上接口即可。该1个以上接口既可以为1个以上同种类的通信接口设备(例如1个以上NIC(Network Interface Card:网络接口卡)),也可以为2个以上不同种类的通信接口设备(例如NIC和HBA(Host Bus Adapter:主机总线适配器))。

[0050] 此外,在以下的说明中,“存储器部”为1个以上存储器,典型的是主存储设备即可。

存储器部的至少1个存储器既可以是易失性存储器也可以是非易失性存储器。

[0051] 程序也可以从程序源安装至计算机那样的装置。程序源例如也可以是程序分发服务器或计算机可读取的存储介质。在程序源为程序分发服务器的情况下,程序分发服务器也可以包含处理器和存储分发对象程序的存储资源,程序分发服务器的处理器将分发对象程序给其它计算机分发。此外,在以下的说明中,既可以2个以上程序作为1个程序实现,也可以1个程序作为2个以上程序实现。

[0052] 在本发明中,存储设备包括1台HDD(Hard Disk Drive:硬盘驱动器)及SSD(Solid State Drive:固态硬盘)等1台储存驱动器、包含多台储存驱动器的RAID装置和多个RAID装置。此外,在驱动器为HDD的情况下,例如既可以包含SAS(Serial Attached SCSI:串行连接SCSI)HDD,也可以包含NL-SAS(近线SAS)HDD。

[0053] [实施例1]

[0054] 以下,参照附图来说明实施例。

[0055] 图1是表示实施例的文件存储系统的硬件结构的图。

[0056] 实施例的文件存储系统1具有站点10-1、10-2和数据中心20,站点10-1、10-2和数据中心20间通过作为WAN(Wide Area Network:广域网)的网络30连接。另外,在图1中图示2个站点10-1、10-2,不过本实施例中对站点的个数没有特别限制。

[0057] 站点10-1具有NAS100、客户端600和管理终端700,该NAS100、客户端600和管理终端700通过LAN(Local Area Network:局域网)相互连接。

[0058] NAS100的具体的结构后述。客户端600是能够进行各种信息处理的计算机等信息处理装置,其进行各种文件操作,例如在NAS100存储文件,进行文件的读/写处理等。管理终端700进行NAS100的管理,在NAS100出现异常时等对NAS100进行各种操作指示等。

[0059] 站点10-2也具有NAS100和客户端600。另外,图1中图示的站点10-1、10-2的硬件结构仅为示意,只要是至少发布具有1台NAS100和客户端600的结构,对于其台数和具有其它硬件结构方面就没有限制。

[0060] 数据中心20具有CAS200。CAS200作为存储于站点10-1、10-2的NAS100中的文件的备份目的地发挥作用。

[0061] 图2是表示实施例的文件存储系统1的NAS100的概略结构的一个例子的图。

[0062] NAS100具有作为控制器的NAS机头110和存储系统120。

[0063] NAS机头110具有进行NAS机头110和NAS100整体的动作控制的处理器111、暂时存储处理器111的动作控制中使用的程序和数据的存储器112、暂时存储由客户端600读取的数据和由存储系统120写入的数据的缓冲存储器113、进行与站点10-1、10-2内的其它客户端600等之间的通信的接口(I/F)114和进行与存储系统120之间的通信的接口(I/F)115。

[0064] 存储系统120也具有进行存储系统120的动作控制的处理器121、暂时存储处理器121的动作控制中使用的程序和数据的存储器122、暂时存储由NAS机头110读取的数据和由存储设备124写入的数据的缓冲存储器123、存储各种文件的存储设备124和进行与NAS机头110之间的通信的接口(I/F)125。

[0065] 在存储器112存储有网络储存程序411、I/O钩子(Hook)程序412、本地文件系统程序413、数据库程序414和文件虚拟化程序415。

[0066] 网络储存程序411接受来自客户端600等的各种请求,处理该请求中包含的协议。

[0067] I0钩子程序412是进行作为后述的本实施例的特征的I0钩子处理的程序,监视网络储存程序411所发出的系统调用,一旦系统被调用,就替换调用协议处理的程序库。进一步,I0钩子程序412记录日志文件3100。I0钩子程序412的动作的详细情况后述。

[0068] 本地文件系统程序413对客户端600等提供文件系统。数据库程序414管理数据库3200。

[0069] 文件虚拟化程序415监视日志文件3100,进行存储设备124内的文件的迁移、存根化或恢复。

[0070] 在存储设备124中存储有数据库3200、用户文件1200、目录2200、管理信息文件1100、2100和日志文件3100,这些文件由本地文件系统程序413构建的本地文件系统510管理。

[0071] 图3是表示实施例的文件存储系统1的CAS200的概略结构的一个例子的图。

[0072] CAS200具有作为控制器的CAS机头210和存储系统220。

[0073] CAS机头210具有进行CAS机头210和CAS200整体的动作控制的处理器211、暂时存储处理器211的动作控制中使用的程序和数据的存储器212、暂时存储由NAS100写入的数据和由存储系统220写入的数据的缓冲存储器213、进行与站点10-1、10-2之间的通信的接口(I/F)214和进行与存储系统220之间的通信的接口(I/F)215。

[0074] 存储系统220也具有进行存储系统220的动作控制的处理器221、暂时存储处理器221的动作控制中使用的程序和数据的存储器222、暂时存储由CAS机头210写入的数据和由存储设备224写入的数据的缓冲存储器223、存储各种文件的存储设备224和进行与CAS机头210之间的通信的接口(I/F)225。

[0075] 在存储器212存储有网络储存程序421、本地文件系统程序422和文件虚拟化程序423。

[0076] 网络储存程序421接受来自NAS100的各种请求,处理该请求中包含的协议。

[0077] 本地文件系统程序422对NAS100提供文件系统。另外,使用的文件系统程序并不限于本地文件系统程序422,也可以使用分散文件系统。

[0078] 文件虚拟化程序423与NAS机头110的文件虚拟化程序415协同工作,进行NAS100的存储设备124内的文件的迁移、存根化或恢复。

[0079] 在存储设备224存储有用户文件1200和目录2200,这些由文件本地文件系统程序422所构建的本地文件系统520来管理。

[0080] 图4是用于说明实施例的文件存储系统1的I0钩子程序412的功能的图。

[0081] 客户端600具有应用程序601和网络储存客户端602。应用程序601例如包括Excel(注册商标)和Word(注册商标)这样输入输出文件的所谓的软件。网络储存客户端602是用于根据来自应用程序601的要求,与NAS100的网络文件系统程序411进行通信的软件,按NAS100的协议对NAS100发出文件操作的要求。网络储存程序411根据该要求对本地文件系统程序413所提供的本地文件系统510进行文件操作。

[0082] I0钩子程序412监视网络储存程序411发出的该系统调用,在网络储存程序411发出系统调用时中断对本地文件系统510的文件操作的API,进行文件虚拟化管理信息的更新处理,此外,输出日志。另外,中断的对象并不限于系统调用,例如也可以作为本地文件系统510提供的固有的API。

[0083] 图5是用于说明实施例的文件存储系统1所提供的文件系统的图。

[0084] 如已经说明的那样,在NAS100(的存储系统120)构建本地文件系统510,该本地文件系统510作为一个例子具有根目录2200-0、目录2200-1。各目录2200-0、2200-1具有管理信息文件2100-1、2100-2。在目录2200-1中,作为一个例子存储有文件1200-1、1200-2。此外,在目录2200-1中,存储有这些文件1200-1、1200-2的管理信息文件1100-1、1100-2。

[0085] 如果客户端600挂接NAS100,则实现具有根目录2200-0、目录2200-1和文件1200-1、1200-2的网络文件系统530,客户端600能够经该网络文件系统530进行各种文件操作。不过,关于本地文件系统510的管理信息文件,由于I0钩子程序412过滤信息,所以在网络文件系统530上不出现,也不能操作。

[0086] 在CAS200也构建有本地文件系统520。本地文件系统520不具有层级结构,而在根目录下配置有所有目录2300-0、2300-1和文件1200-1、1200-2。在CAS200,各个目录2300-0、2300-1、文件1200-1、1200-2通过UUID(Universally Unique Identifier:通用唯一标识符)唯一地确定。

[0087] 图6是表示实施例1的文件存储系统1的管理信息文件2100的一个例子的图。

[0088] 管理信息文件2100具有用户目录管理信息2110。用户目录管理信息2110按每个UUID具有条目。各条目是被赋予给用户目录2200的UUID2111、用户目录2200的目录状态2112、用户目录2200的主体处理程序2113、迁移有无2114。

[0089] 目录状态2112是表示从上次的备份起该用户目录2200是否被更新的值,脏(Dirty)是表示文件已被更新的值。主体处理程序2113是将用户目录2200唯一地确定的值,是能够用于在系统调用中作为操作对象指定用户目录2200的值。在主体处理程序2113中使用在用户目录2200的生成到删除的期间中不变更的值。迁移有无2114是表示该用户目录2200是否备份过纵使一次的值。

[0090] 用户目录2200具有文件/目录名2201和索引号(Inode#)2202。图6所示的例子是图5的目录(dir1)2200-1,在该目录2200-1存储有2个文件(File1、File2)。索引号2202是唯一地被赋予给各个文件(File1、File2)的索引号。

[0091] CAS用目录2300具有文件/目录名2301和索引号(Inode#)2302。文件/目录名2301与用户目录2200的文件/目录名2201相同,不过索引号2302在从NAS100至CAS200的迁移时被改写为UUID。这是因为,索引号仅在NAS100内被唯一地确定,在迁移时需要分配在CAS200内被唯一地确定的UUID。

[0092] 图7是表示实施例的文件存储系统1的管理信息文件1100的其它例子的图。

[0093] 理信息文件1100具有用户文件管理信息1110和部分管理信息1120。

[0094] 用户文件管理信息1110按每个UUID具有条目。各条目是被赋予给用户文件1200的UUID1111、用户文件1200的文件状态1112、用户文件1200的主体处理程序1113、迁移有无2114。

[0095] 部分管理信息1120按每用户文件1200创建。部分管理信息1120具有偏移量1121、长度1122和部分状态1123。偏移量1121表示用户文件1200被部分地更新处理时的其更新处理的开始位置,长度1122表示从偏移量1121的位置起更新处理了多少数据长度,部分状态1123表示进行了什么样的更新处理。此处,脏(Dirty)1201表示在上次的备份处理后被更新过,存根(Stub)2203表示在备份处理后被从本地(即NAS100)擦除过,已缓存(Cached)2202

表示在本地也具有数据,也有备份。

[0096] 图8是表示实施例的文件存储系统1的日志文件3100的一个例子的图。

[0097] 日志文件3100具有API名称3101、变量3102、返回值3103、类型3104、索引号3105、管理信息文件处理程序3106、上级索引号3107、执行状态3108和时间戳3109。日志文件3100的各行在每当有从客户端600至NAS100的系统调用时创建。

[0098] API名称3101表示系统调用的种类,大致存储有:写(write)、读(read)、打开(open)、关闭(close)的值。变量3102是系统调用的变量,大致具有文件描述符、文件操作开始位置和数据大小。返回值3103是作为系统调用的结果从本地文件系统510返回的值,N.A.表示由于在系统调用执行中尚未有返回值,0表示已被正常执行。在此以外存储由本地文件系统510确定的值。类型3104是表示成为系统调用的对象的是文件还是目录的值。索引号是成为系统调用的对象的文件等的索引号。管理信息文件处理程序3106是将成为系统调用的对象的文件等唯一地确定的值,是能够在系统调用中的文件和目录操作中的操作对象指定中使用的值。管理信息文件处理程序3106从文件和目录的生成到删除都不变更。上级索引号3107是成为系统调用的对象的文件等的上一级(上级)的目录的索引号。这是因为,在通过系统调用移动或删除了文件、目录的情况下,需要作为备份处理的对象确定上级目录。执行状态3108存储表示系统调用的执行状态的值。时间戳3109是系统调用被调用的时刻。

[0099] 图9是表示实施例的文件存储系统1的数据库3200的一个例子的图。

[0100] 数据库3200具有索引号3201、类型3202、管理信息文件处理程序3203、脏(Dirty)部有无3204、非存根部有无3205和删除标志3206。数据库3200的各行按每本地文件系统510内的目录和文件创建。

[0101] 索引号3201是目录或文件的索引号。类型3202是表示按索引号3201确定的是文件还是目录的值。管理信息文件处理程序3203是将成为对象的文件等唯一地确定的值。脏部有无3204存储表示是否在目录中存储的文件或文件自身的纵使一部分具有脏部的值。非存根部有无3205存储表示是否具有在上次的备份处理后数据纵使一部分被改写的部分的值。删除标志3206存储表示是否目录中存储的文件或文件自身被删除的值。

[0102] 接着,参照图10~图21的流程图,说明本实施例的文件存储系统1的动作。

[0103] 图10是用于说明实施例的文件存储系统1的文件/目录创建处理的一个例子的流程图。

[0104] 当文件/目录创建处理开始时(步骤S100),首先,I0钩子程序412在日志文件3100中追加记录创建处理的开始(步骤S101)。

[0105] 然后,I0钩子程序412基于来自客户端600的系统调用,实施用户文件1200/目录2200的创建处理(步骤S102)。接着,I0钩子程序412创建管理信息文件1100、2100(步骤S103)。接着,I0钩子程序412将创建对象的文件/目录的上级目录的管理信息文件2100的目录状态2112更新为脏(Dirty)(步骤S104)。

[0106] 然后,I0钩子程序412在日志文件3100追加记录创建处理的完成(步骤S105),将创建处理的完成响应给网络储存程序411(步骤S106)。

[0107] 图11是用于说明实施例的文件存储系统1的文件/目录删除处理的一个例子的流程图。

[0108] 当文件/目录删除处理开始时(步骤S200),首先,I0钩子程序412在日志文件3100

追加记录删除处理的开始(步骤S201)。

[0109] 然后, I0钩子程序412判断在删除对象的文件/目录是否有迁移(步骤S202)。迁移的有无能够通过管理信息文件1100、2100的迁移的有无1114、2114来确认。然后,如果判断为肯定(在步骤S202为是),则程序转移至步骤S203,如果判断为否定(在步骤S202为否),则程序转移至步骤S206。

[0110] 在步骤S203,I0钩子程序412将管理信息文件1100、2100和用户文件1200移动至垃圾箱目录,接着,I0钩子程序412将用户文件1200的内部清空(步骤S204)。然后,I0钩子程序412将对应的管理信息文件1100、2100的文件状态1112/目录状态2112更新为被删除(Deleted),擦除部分管理信息1120(步骤S205)。

[0111] 另一方面,在步骤S206,I0钩子程序412将管理信息文件1100、2100擦除,然后,执行用户文件1200/用户目录2200的删除处理(步骤S207)。

[0112] 接着,I0钩子程序412将创建对象的文件/目录的上级目录的管理信息文件2100的目录状态2112更新为脏(Dirty)(步骤S208)。然后,I0钩子程序412在日志文件3100追加记录删除处理的完成(步骤S209),将删除处理的完成响应给网络储存程序411(步骤S210)。

[0113] 图12是用于说明实施例的文件存储系统1的重命名处理的一个例子的流程图。

[0114] 当重命名处理开始时(步骤S300),首先,I0钩子程序412在日志文件3100追加记录重命名处理的开始(步骤S301)。

[0115] 然后,I0钩子程序412实施通常的重命名处理(步骤S302)。接着,I0钩子程序412将与重命名对象的移动目的地目录对应的管理信息文件2100的目录状态2112更新为脏(Dirty)(步骤S303),进一步,I0钩子程序412将与重命名对象的移动出发地目录对应的管理信息文件2100的目录状态2112更新为脏(Dirty)(步骤S304)。

[0116] 然后,I0钩子程序412在日志文件3100追加记录重命名处理的完成(步骤S305),将重命名处理的完成响应给网络储存程序411(步骤S306)。

[0117] 图13是用于说明实施例的文件存储系统1的文件写处理的一个例子的流程图。

[0118] 当文件写处理开始时(步骤S400),首先,I0钩子程序412在日志文件3100追加记录写处理的开始(步骤S401)。

[0119] 接着,I0钩子程序412在用户文件1200实施通常的写处理(步骤S402)。接着,I0钩子程序412将对应的管理信息文件1100的文件状态1112更新为脏(Dirty)(步骤S403)。

[0120] 然后,I0钩子程序412在日志文件3100追加记录写处理的完成(步骤S404),将写处理的完成响应给网络储存程序411(步骤S405)。

[0121] 图14是用于说明实施例的文件存储系统1的文件读处理的一个例子的流程图。

[0122] 当文件读处理开始时(步骤S500),首先,I0钩子程序412取得对应的管理信息文件1100(步骤S501)。

[0123] 接着,I0钩子程序412判断读取对象处是否包含被存根化的部分(步骤S502)。然后,如果判断为肯定(在步骤S502为是),程序转移至步骤S503,如果判断为否定(在步骤S502为否),则程序转移至步骤S506。

[0124] 在步骤S503,I0钩子程序412向CAS200请求读取对象处内的被存根化的部分的数据。CAS200的文件虚拟化程序423基于来自I0钩子程序412的请求将该数据转发至NAS100(步骤S504)。

[0125] 接着, I0钩子程序412将管理信息文件1100内的召回部、即从CAS200转发的数据的部分状态1123更新为已缓存 (Cached) (步骤S505)。

[0126] 然后, I0钩子程序412对用户文件1200实施通常的读处理 (步骤S506), 将读处理的完成响应给网络储存程序411(步骤S507)。

[0127] 图15是用于说明实施例的文件存储系统1的目录读处理的一个例子的流程图。

[0128] 当目录读处理开始时 (步骤S600), 首先, I0钩子程序412取得对应的管理信息文件2100 (步骤S601)。

[0129] 接着, I0钩子程序412判断在读对象目录是否为已被存根化的状态 (步骤S602)。然后, 如果判断为肯定 (在步骤S602为是), 则程序转移至步骤S603, 如果判断为否定 (在步骤S602为否), 则程序转移至步骤S607。

[0130] 在步骤S603, I0钩子程序412将读对象的CAS用目录2300的取得请求转发至 CAS200。CAS200的文件虚拟化程序423基于来自I0钩子程序412的要求将该数据转发至 NAS100 (步骤S604)。

[0131] 接着, I0钩子程序412用由CAS200取得的数据更新用户目录2200 (步骤S605), 将管理信息文件2100的目录状态2112更新为已缓存 (Cached) (步骤S606)。

[0132] 然后, I0钩子程序412对用户目录2200实施通常的读处理 (步骤S607), 以从客户端 600一方看不见管理信息文件2100的方式从读结果擦除管理信息文件2100的信息 (步骤 S608), 将读处理的完成响应给网络储存程序411(步骤S609)。

[0133] 图16是用于说明实施例的文件存储系统1的日志反映处理的一个例子的流程图。

[0134] 当日志反映处理开始时 (步骤S1301), 文件虚拟化程序415参照日志文件3100的执行状态3108, 从日志文件3100取得所完成的操作的列表 (步骤S1302)。

[0135] 接着, 文件虚拟化程序415判断在步骤S1302取得的列表是否为空 (步骤S1303)。其结果, 如果判断为肯定 (在步骤S1303为是) 则程序转移至步骤S1314, 如果判断为否定 (在步骤S1303为否) 则程序转移至步骤S1304。

[0136] 在步骤S1304, 文件虚拟化程序415从在步骤S1302取得的列表中取得1个条目。然后, 文件虚拟化程序415判断在步骤S1304取得的条目是否为写处理 (步骤S1305)。然后, 如果判断为肯定 (在步骤S1305为是) 则程序转移至步骤S1306, 如果判断为否定 (在步骤S1305 为否) 则程序转移至步骤S1307。

[0137] 在步骤S1306, 文件虚拟化程序415令数据库3200的操作对象的条目的脏部有无3204和非存根部有无3205分别为有。

[0138] 在步骤S1307, 文件虚拟化程序415判断在步骤S1304取得的条目是否为创建处理。然后, 如果判断为肯定 (在步骤S1307为是) 则程序转移至步骤S1308, 如果判断为否定 (在步骤S1307为否) 则程序转移至步骤S1310。

[0139] 在步骤S1308, 文件虚拟化程序415创建数据库3200的操作对象的条目, 令所创建的条目的脏部有无3204和非存根部有无3205分别为有, 使删除标志3206的值为失败 (False)。进一步, 文件虚拟化程序415令数据库3200的操作对象的上级目录的条目的脏部有无3204和非存根部有无3205分别为有 (步骤S1309)。

[0140] 在步骤S1310, 文件虚拟化程序415判断在步骤S1304取得的条目为删除处理。然后, 如果判断为肯定 (在步骤S1310为是) 则程序转移至步骤S1311, 如果判断为否定 (在步骤

S1310为否)则程序转移至步骤S1312。

[0141] 在步骤S1311,文件虚拟化程序415令数据库3200的操作对象的条目的脏部有无3204和非存根部有无3205分别为无,进一步,令删除标志3206为成功(True)。

[0142] 在步骤S1312,文件虚拟化程序415判断在步骤S1304取得的条目是否为重命名处理。然后,如果判断为肯定(在步骤S1312为是)则程序转移至步骤S1309,如果判断为否定(在步骤S1312为否)则程序转移至步骤S1313。

[0143] 在步骤S1313,文件虚拟化程序415从在步骤S1302取得的列表中将条目删除。

[0144] 另一方面,在步骤S1314,文件虚拟化程序415将完成处理的日志删除。

[0145] 图17是用于说明实施例的文件存储系统的文件迁移处理的一个例子的流程图。

[0146] 当文件迁移处理开始时(步骤S700),文件虚拟化程序415从数据库3200,作为列表取得脏部有无3204为有而类型3202为文件的条目(步骤S701)。

[0147] 接着,文件虚拟化程序415判断在步骤S701取得的文件列表是否为空(步骤S702)。其结果,如果判断为肯定(在步骤S702为是)则程序转移至步骤S712,如果判断为否定(在步骤S702为否)则程序转移至步骤S703。

[0148] 在步骤S703,文件虚拟化程序415从在步骤S701取得的列表中取得1个条目。接着,文件虚拟化程序415取得在步骤S703取得的条目所示的管理信息文件1100(步骤S704)。接着,文件虚拟化程序415从在步骤S704取得的管理信息文件1100的部分管理信息1120,作为转发部分列表取得脏的条目(步骤S705),从用户文件1200取得所取得的转发部分列表的相应位置(步骤S706)。

[0149] 接着,文件虚拟化程序415伴随向管理信息文件1100中的UUID1111的更新请求,将从在步骤S705取得的转发部分列表和在步骤S706从用户文件1200取得的数据转发至CAS200(步骤S707)。

[0150] CAS200的文件虚拟化程序423从在UUID确定的CAS200内的用户文件1200更新在步骤S707接受的转发部分列表所示的位置(步骤S708),将更新的完成返送回NAS100(步骤S709)。

[0151] 然后,文件虚拟化程序415令管理信息文件1100的文件状态1112和转发部分列表的相应位置的部分状态1123为已缓存(Cached)(步骤S710),从在步骤S701取得的文件列表删除条目(步骤S711)。

[0152] 另一方面,在步骤S712,文件虚拟化程序415令从数据库3200完成了操作的条目的脏部有无3204为无。

[0153] 图18是用于说明实施例的文件存储系统1的目录迁移处理的一个例子的流程图。

[0154] 当目录迁移处理开始时(步骤S800),文件虚拟化程序415从数据库3200,作为列表取得脏部有无3204为有而类型3202为目录的条目(步骤S801)。

[0155] 接着,文件虚拟化程序415判断在步骤S801取得的文件列表是否为空(步骤S802)。其结果,如果判断为肯定(在步骤S802为是)则程序转移至步骤S812,如果判断为否定(在步骤S802为否)则程序转移至步骤S803。

[0156] 在步骤S803,文件虚拟化程序415从在步骤S801取得的列表中取得1个条目。接着,文件虚拟化程序415取得在步骤S803取得的条目所示的管理信息文件2100(步骤S804)。接着,文件虚拟化程序415取得在步骤S804取得的管理信息文件2100所示的用户目录2200(步

骤S805),基于所取得的用户目录2200生成CAS用目录2300的信息(步骤S806)。

[0157] 接着,文件虚拟化程序415随着对管理信息文件2100中的UUID2111的更新请求,将在步骤S806生成的CAS用目录2300的信息转发至CAS200(步骤S807)。

[0158] CAS200的文件虚拟化程序423将按UUID确定的CAS200内的CAS用目录2300更新(步骤S808),将更新的完成返送回NAS100(步骤S809)。

[0159] 然后,文件虚拟化程序415令管理信息文件2100的目录状态2112为已缓存(Cached)(步骤S810),从在步骤S801取得的文件列表将条目删除(步骤S811)。

[0160] 另一方面,在步骤S812,文件虚拟化程序415令从数据库3200完成了操作的条目的脏部有无3204为无。

[0161] 图19是用于说明实施例的文件存储系统1的文件存根化处理的一个例子的流程图。

[0162] 当文件存根化处理开始时(步骤S900),文件虚拟化程序415从数据库3200,作为列表取得脏部有无3204为无而类型3202为文件的条目(步骤S901)。

[0163] 接着,文件虚拟化程序415判断在步骤S901取得的文件列表是否为空(步骤S902)。其结果,如果判断为肯定(在步骤S902为是)则程序转移至步骤S908,如果判断为否定(在步骤S902为否)则程序转移至步骤S903。

[0164] 在步骤S703,文件虚拟化程序415从在步骤S901取得的列表取得1个条目。接着,文件虚拟化程序415取得在步骤S703取得的条目所示的管理信息文件1100(步骤S904)。接着,文件虚拟化程序415取得在步骤S904取得的管理信息文件1100所示的用户文件1200(步骤S905)。

[0165] 然后,文件虚拟化程序415令管理信息文件1100的文件状态1112和转发部分列表的相应位置的部分状态1123为存根(Stub)(步骤S906),从在步骤S901取得的文件列表将条目删除(步骤S907)。

[0166] 另一方面,在步骤S908,文件虚拟化程序415令从数据库3200完成了操作的条目的非存根部有无3205为无。

[0167] 图20是用于说明实施例的文件存储系统1的CAS侧文件/目录删除处理的一个例子的流程图。

[0168] 当CAS侧文件/目录删除处理开始时(步骤S1000),文件虚拟化程序415从数据库3200,作为列表取得删除标志3206为成功(True)的条目(步骤S1001)。

[0169] 接着,文件虚拟化程序415判断在步骤S1001取得的文件列表是否为空(步骤S1002)。其结果,如果判断为肯定(在步骤S1002为是)则程序转移至步骤S1010,如果判断为否定(在步骤S1002为否)则程序转移至步骤S1003。

[0170] 在步骤S1003,文件虚拟化程序415从在步骤S1001取得的列表取得1个条目。接着,文件虚拟化程序415取得在步骤S1003取得的条目所示的管理信息文件1100、2100(步骤S1004)。

[0171] 接着,文件虚拟化程序415将管理信息文件1100、2100所示的UUID1111、2111的删除要求转发至CAS200(步骤S1005)。

[0172] CAS200的文件虚拟化程序423将按UUID确定的CAS200内的用户文件1200/用户目录2200删除(步骤S1006),并将删除的完成返送回NAS100(步骤S1007)。

[0173] 然后,文件虚拟化程序415从在步骤S1001取得的列表中将条目删除(步骤S1009)。

[0174] 另一方面,在步骤S1010,文件虚拟化程序415令从数据库3200完成了操作的条目的脏部有无3204为无。

[0175] 图21是用于说明实施例的文件存储系统1的爬虫处理的一个例子的流程图。

[0176] 当爬虫处理开始时(步骤S1100),文件虚拟化程序415对成为文件虚拟化的对象的用户文件1200/用户目录2200的根目录2200执行以下所示的步骤S1200的处理。

[0177] 在步骤S1200,首先,文件虚拟化程序415取得相应的用户文件1200/用户目录2200的管理信息文件1100、2100(步骤S1202)。

[0178] 接着,文件虚拟化程序415判断在步骤S1202取得的管理信息文件1100、2100的文件状态1112/目录状态2112是否为脏(步骤S1203)。然后,如果判断为肯定(在步骤S1203为是),则程序转移至步骤S1204,如果判断为否定(在步骤S1203为否),则程序转移至步骤S1205。

[0179] 在步骤S1204,脏部有无3204为有,非存根部有无3205为有,删除标志3206作为失败(False),在数据库3200登记对象的条目。

[0180] 在步骤S1205,文件虚拟化程序415判断在步骤S1202取得的管理信息文件1100、2100的文件状态1112/目录状态2112是否为已缓存(Cached)。然后,如果判断为肯定(在步骤S1205为是),则程序转移至步骤S1206,如果判断为否定(在步骤S1205为否),则程序转移至步骤S1207。

[0181] 在步骤S1206,脏部有无3204为无,非存根部有无3205为有,删除标志3206作为失败(False),在数据库3200登记对象的条目。

[0182] 在步骤S1207,文件虚拟化程序415判断在步骤S1202取得的管理信息文件1100、2100的文件状态1112/目录状态2112是否为被删除(Deleted)。然后,如果判断为肯定(在步骤S1207为是),则程序转移至步骤S1208,如果判断否定(在步骤S1207为否),则程序转移至步骤S1209。

[0183] 在步骤S1208,脏部有无3204为无,非存根部有无3205为无,删除标志3206作为成功(True),在数据库3200登记对象的条目。

[0184] 在步骤S1209,文件虚拟化程序415判断爬虫处理的对象是否为目录。然后,如果判断为肯定(在步骤S1209为是),则程序转移至步骤S1210,如果判断为否定(在步骤S1209为否),则结束程序。

[0185] 在步骤S1210,对目录内的各文件/目录执行步骤S1200的处理。

[0186] 根据这样构成的本实施例,文件存储系统1的NAS100在来自客户端600的文件的操作请求至文件系统的调用处理为止的期间进行中断,基于对文件系统的输入信息或操作内容,追加作为文件的状态管理信息的管理信息文件1100、2100的更新处理。

[0187] 因此,根据本实施例,能够不受文件系统的版本升级影响地提供文件虚拟化功能。

[0188] 此外,NAS100在日志文件3100中登记从文件的生成到删除为止的期间没有变更的、文件的访问所需的信息。由此,与作为访问必要信息登记从文件生成到删除为止的期间产生变更的路径的方法相比较,由于不需要进行追踪各文件的路径变更的处理,因此能够抑制在进行文件虚拟化处理即迁移处理和/或存根化处理时分析日志文件3100的负载的增大。

[0189] 另外,上述的实施例为了将本发明说明得容易明白而对结构进行了详细的说明,但是并不一定限于包括所说明的所有结构。能够对各实施例的结构的一部分进行其它结构的追加、删除、替换。

[0190] 此外,上述的各结构、功能、处理部、处理单元等例如也可以通过利用集成电路等进行设计、利用硬件实现其一部分或全部。此外,此外,本发明还能够通过实现实施例的功能的软件的程序代码来实现。在这种情况下,在计算机提供记录有程序代码的存储介质,其计算机所具备的处理器读出存储介质中存储的程序代码。在这种情况下,从存储介质读出的程序代码自身实现上述的实施例的功能,其程序代码自身和存储有该程序代码的存储介质构成本发明。作为用于供给这样的程序代码的存储介质,例如使用软盘、CD-ROM、DVD-ROM、硬盘、SSD (Solid State Drive: 固态硬盘)、光盘、光磁盘、CD-R、磁带、非易失性的存储卡、ROM等。

[0191] 此外,实现本实施例中记载的功能的程序代码例如能够以汇编程序、C/C++、perl、Shell、PHP、Java (注册商标) 等大范围的程序或脚本语言来安装。

[0192] 在上述的实施例中,对于控制线和信息线,仅展示在说明上被认为需要的部分,并不一定展示制品上的所有的控制线和信息线。所有的结构也可以相互连接。

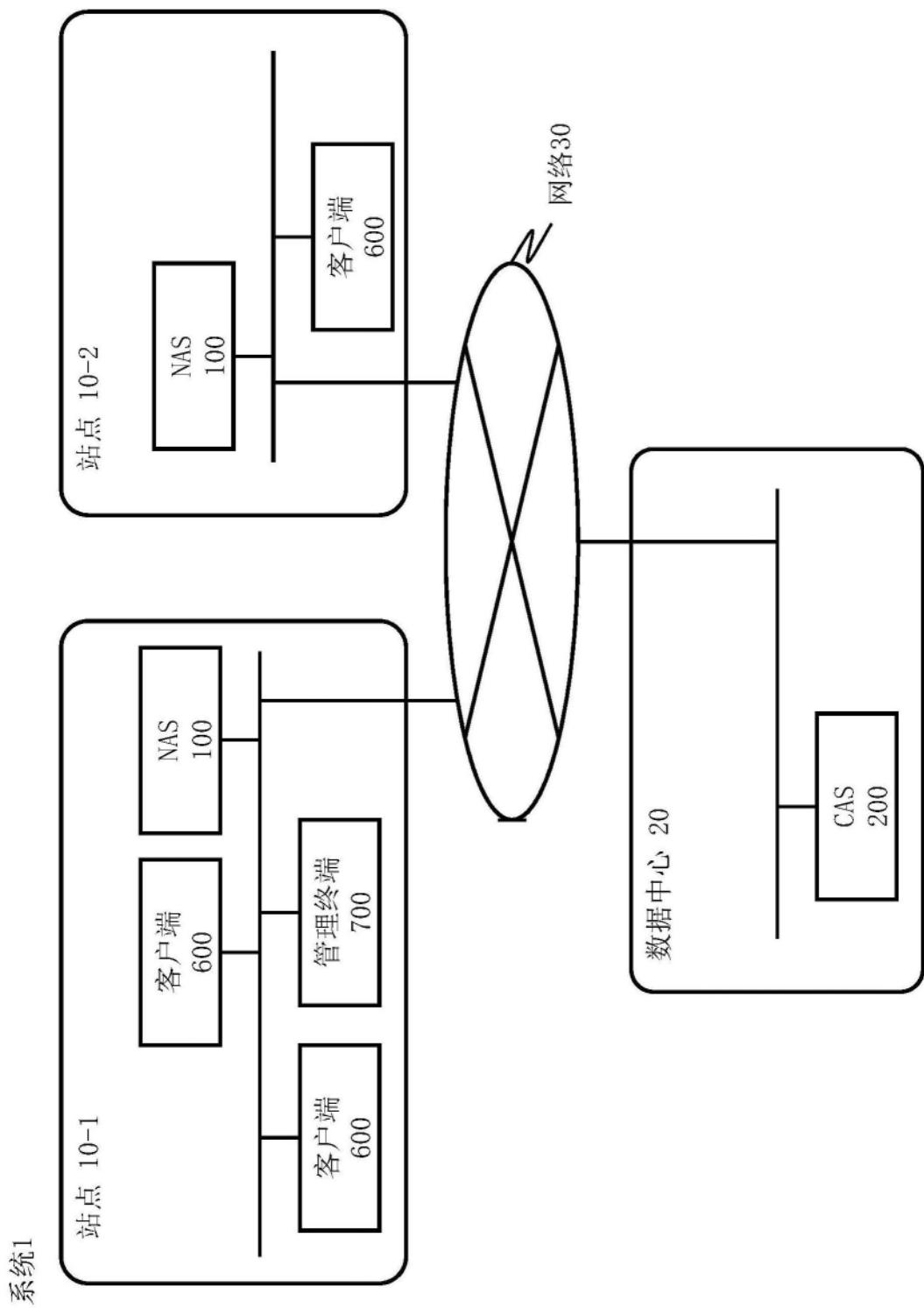


图1

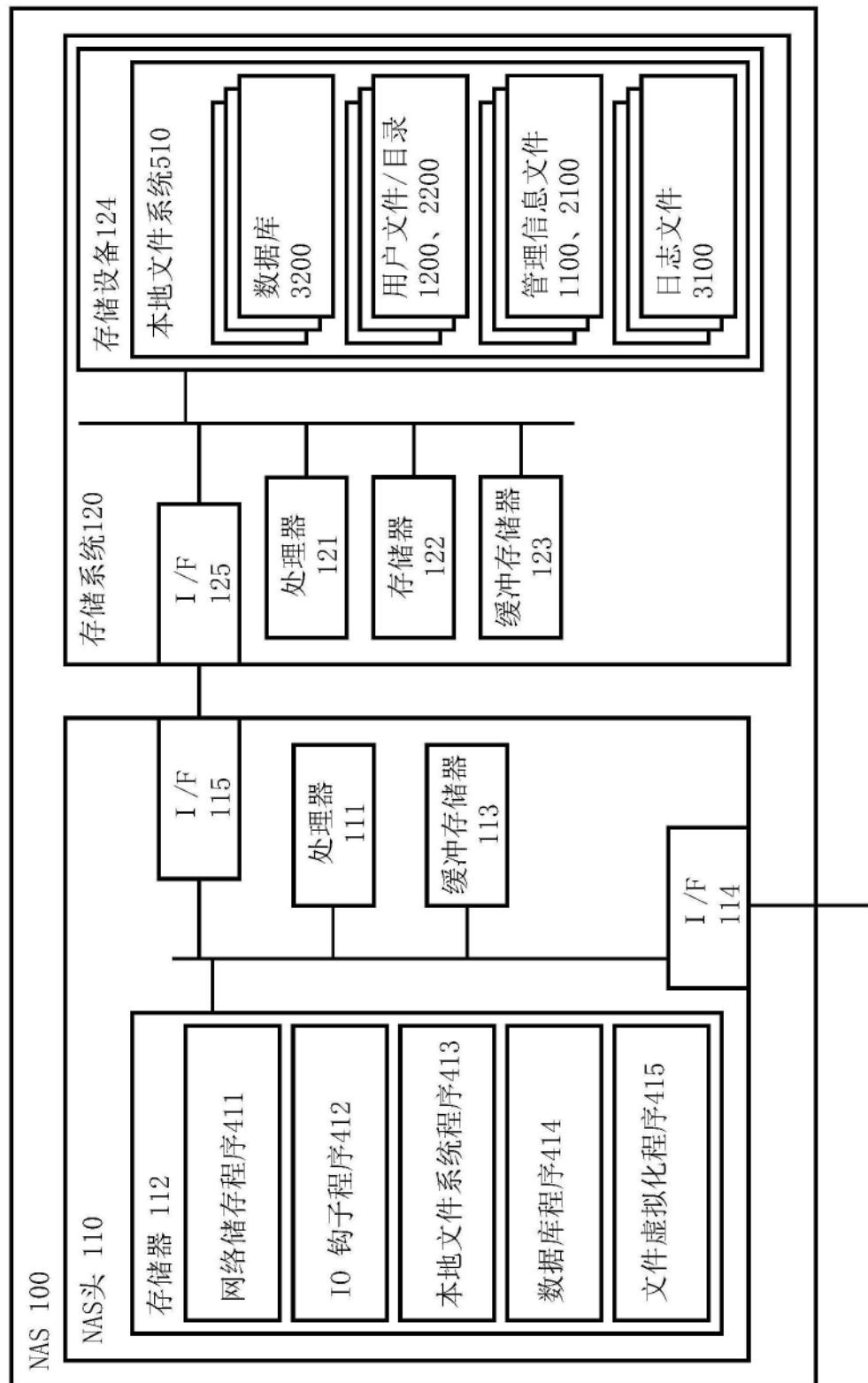


图2

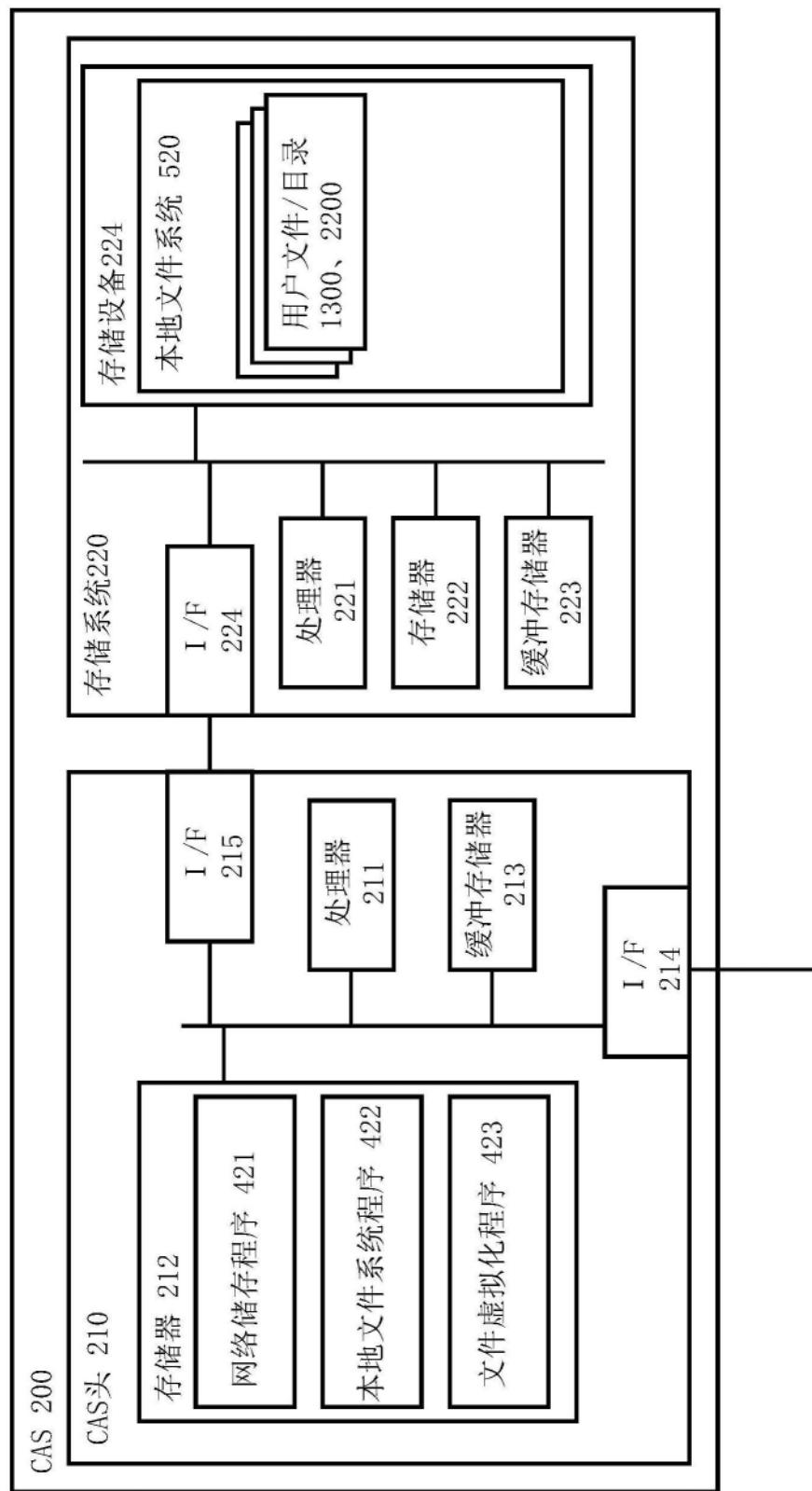


图3

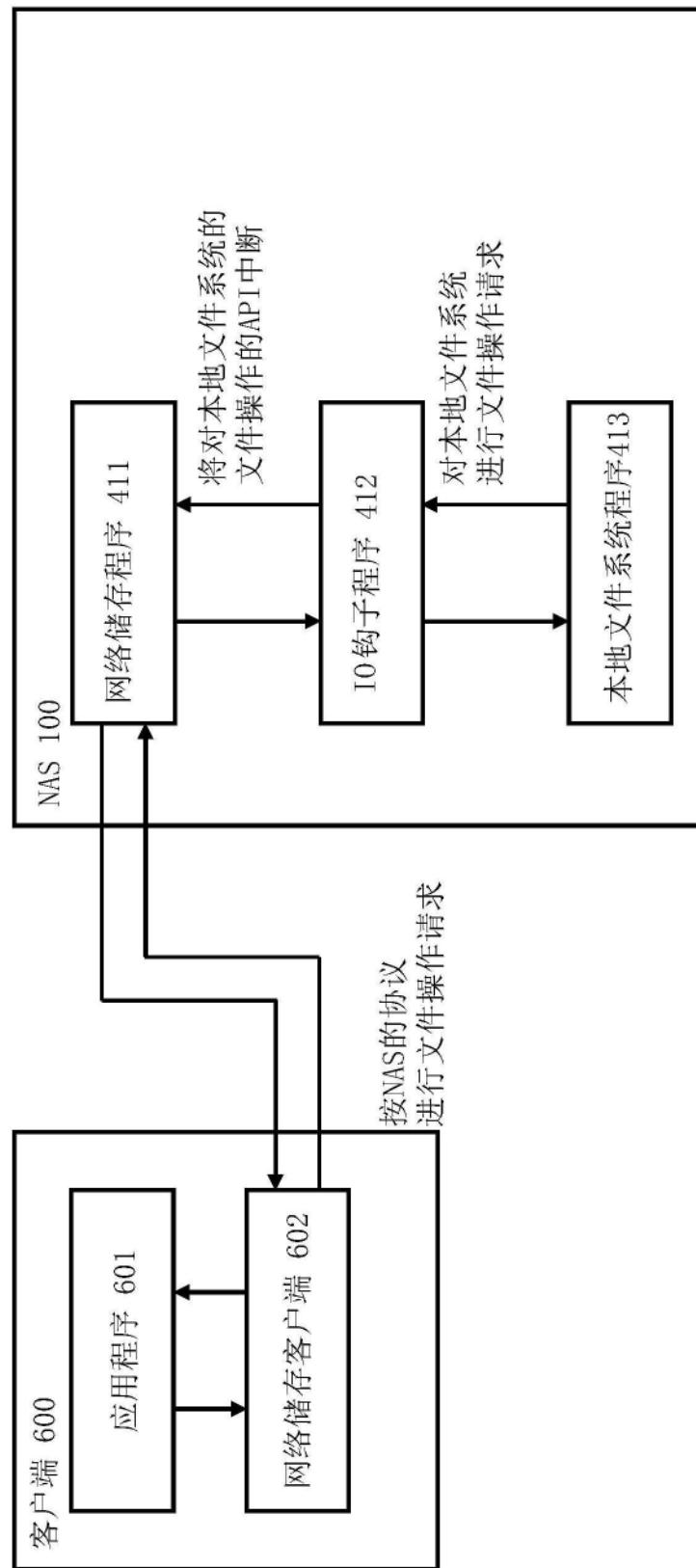


图4

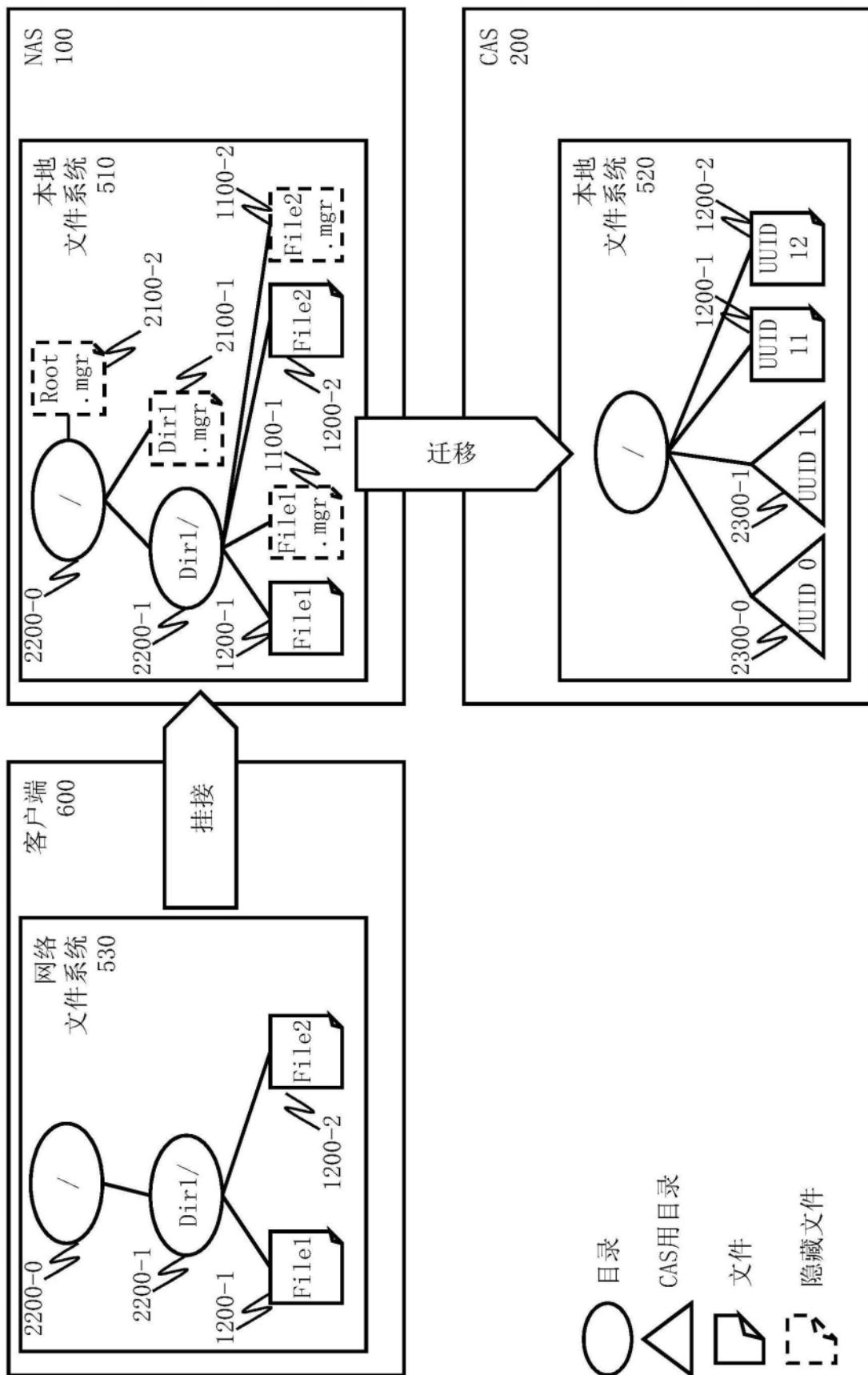


图5

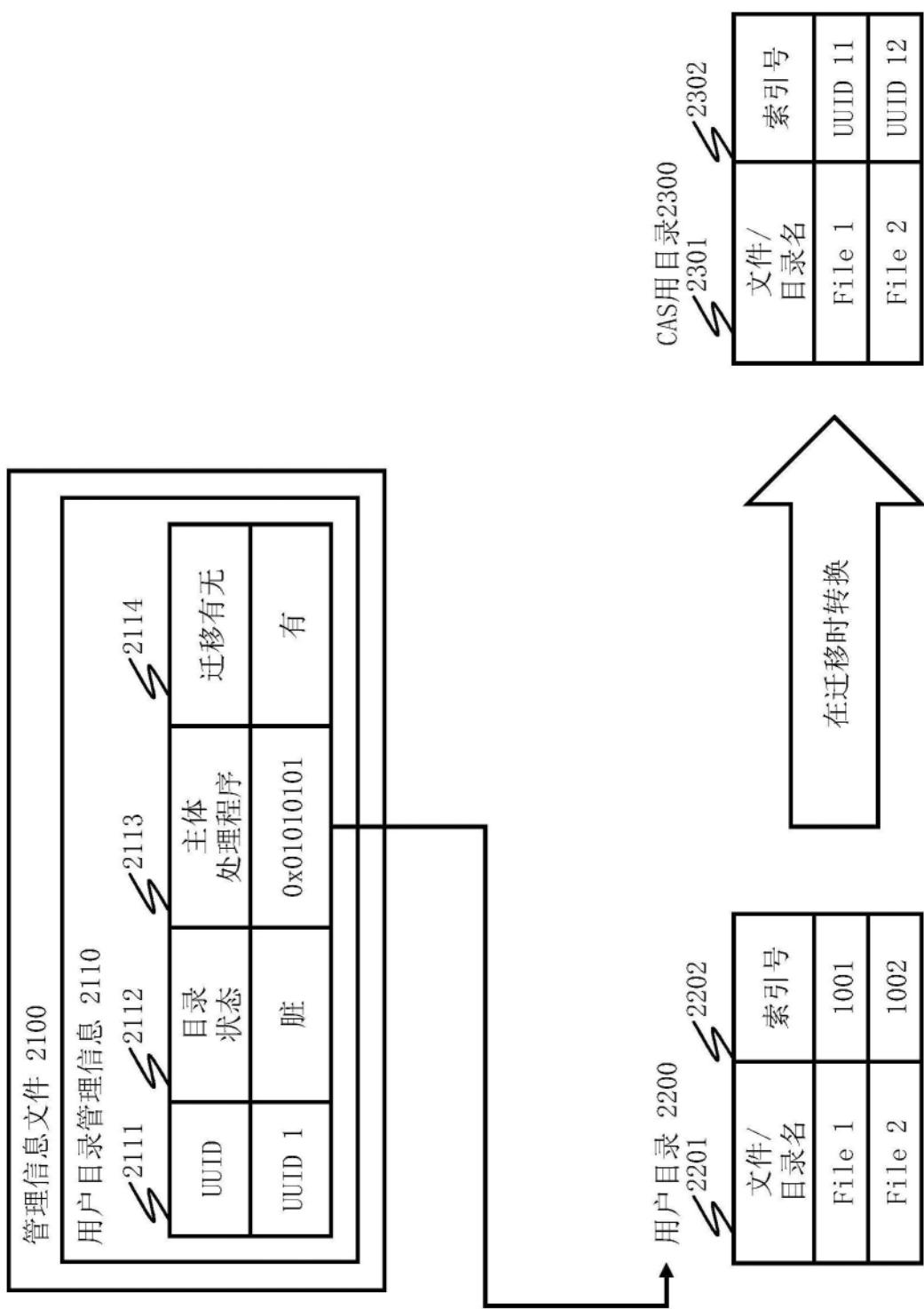


图6

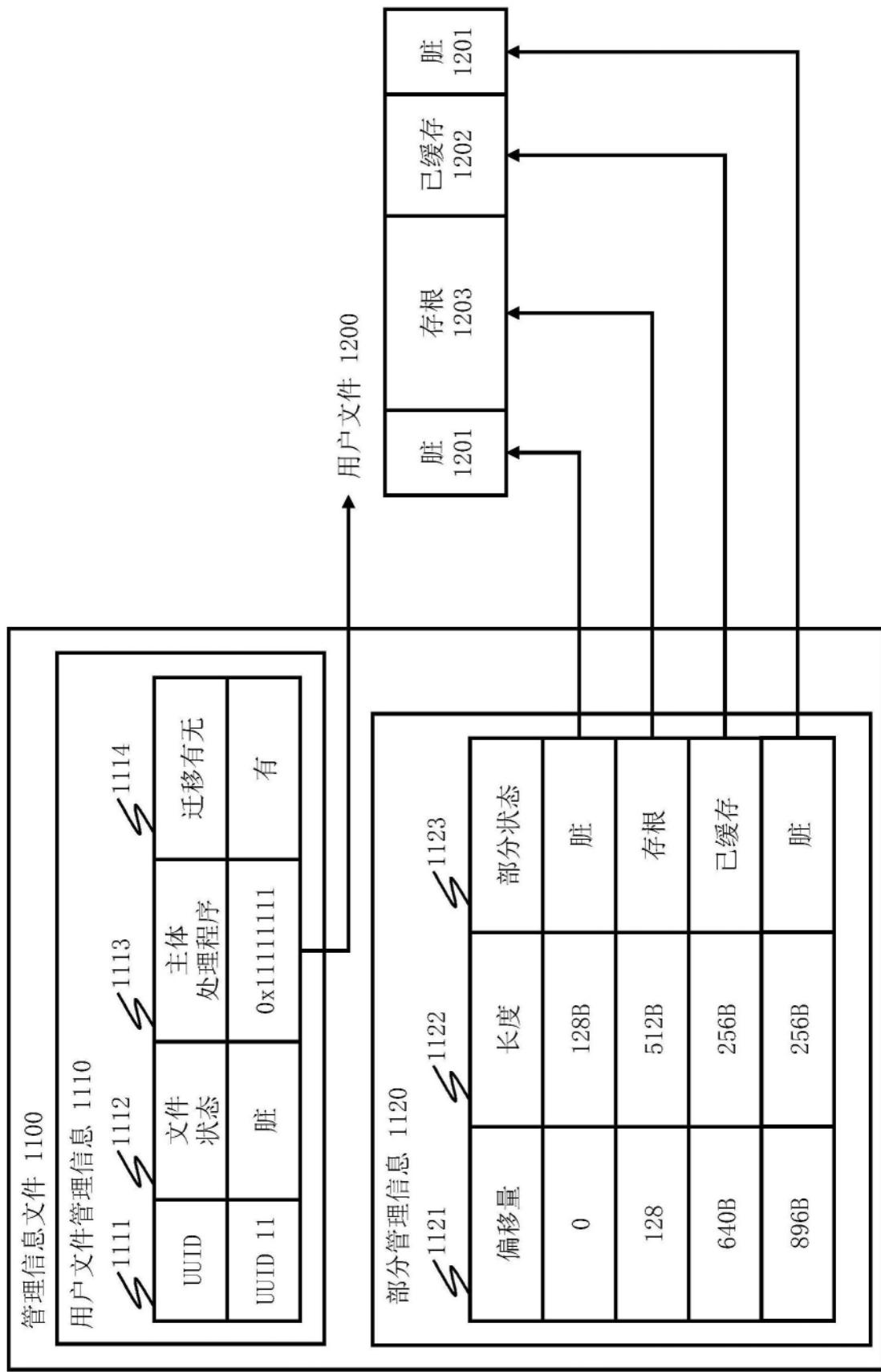


图7

日志文件 3100

API 名称	变量	返回值	类型	索引 号	文件 管理信息 程序	上 级 索引 号	执行状态	时间戳
写	fd=1, buf=0xxxxxxxx, nbytes=20	N.A.	文件	1001	0x11111111	-	开始	
写	fd=1, buf=0xxxxxxxx, nbytes=20	0	文件	1001	0x11111111	-	完成	

图8

索引号	类型	管理信息文件 处理程序	脏部 有无	非存根部 有无	删除标志
101	目录	0x01010101	有	无	失败
1001	文件	0x11111111	有	有	失败
1002	文件	0x12121212	无	有	失败

图9

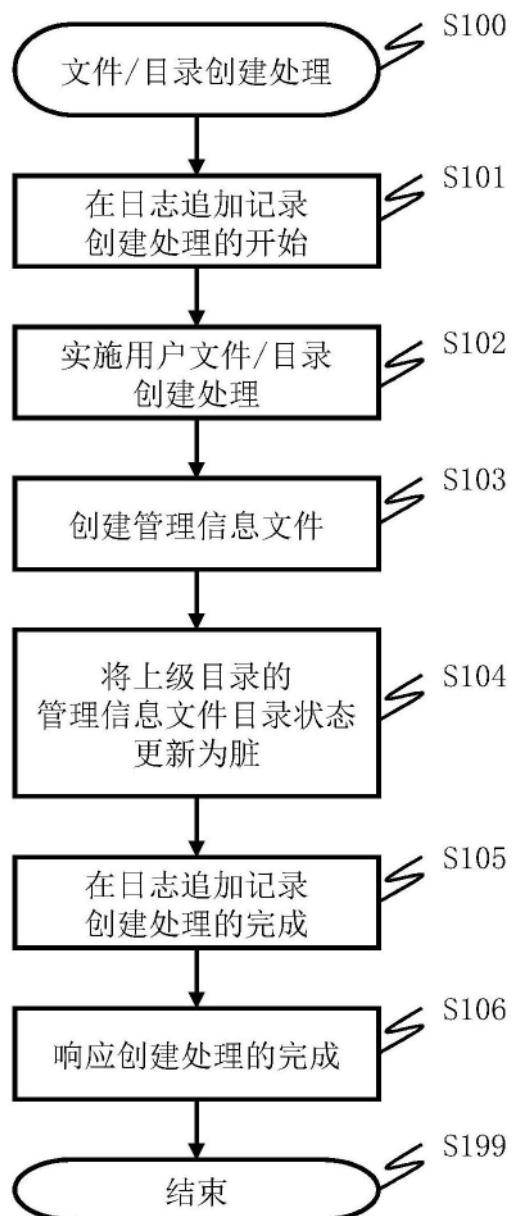


图10

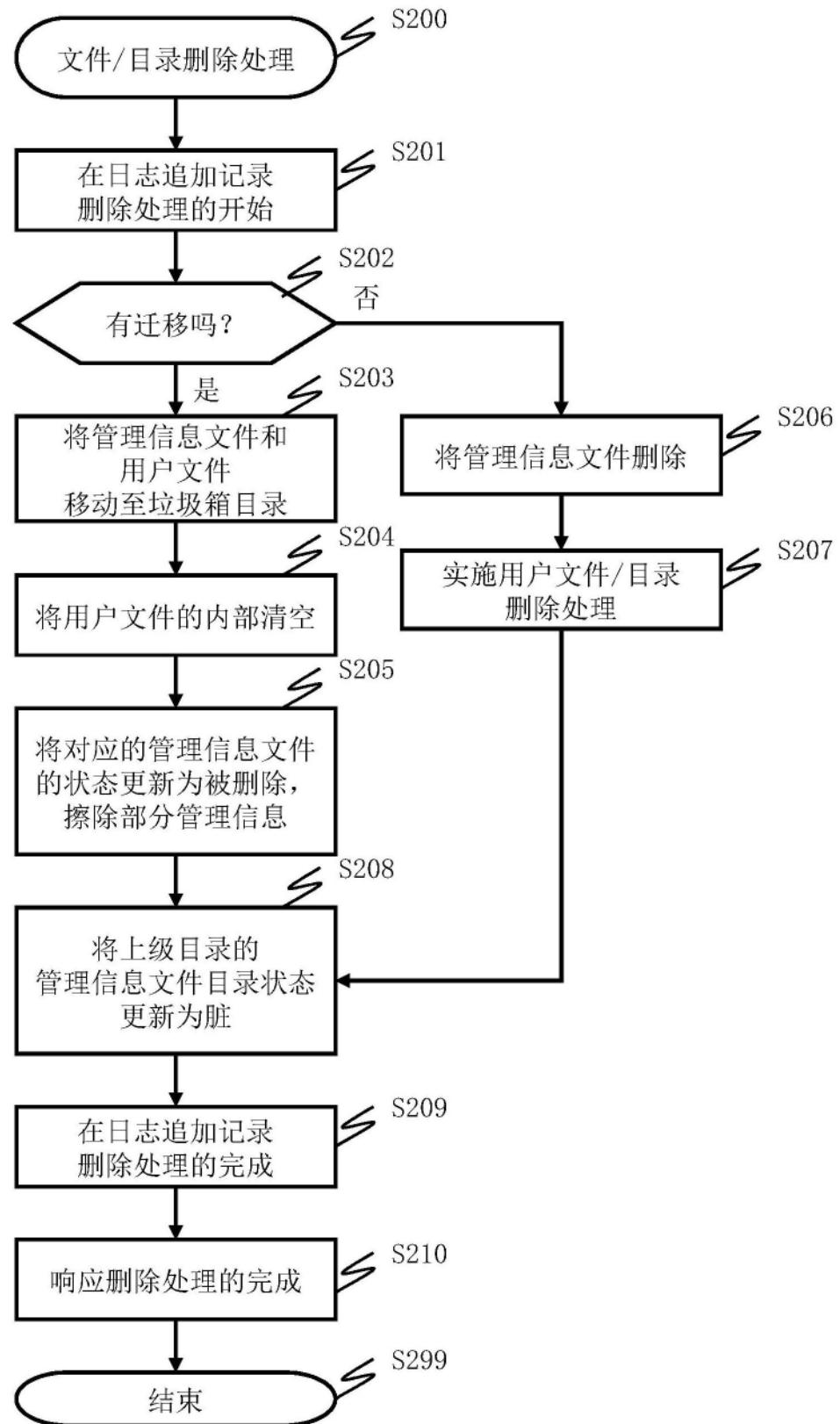


图11

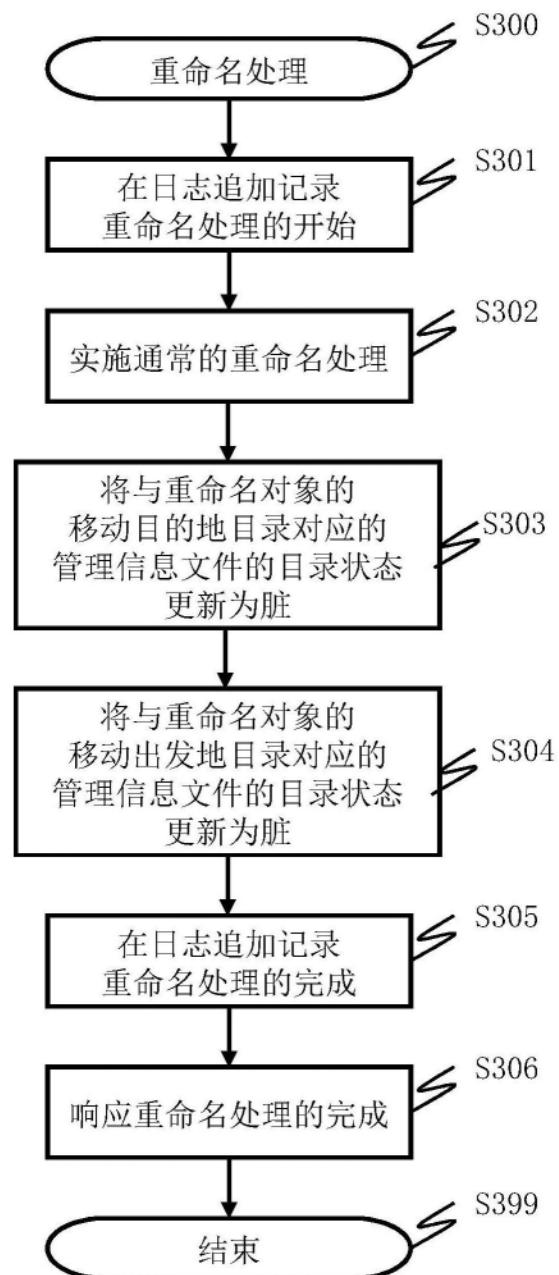


图12

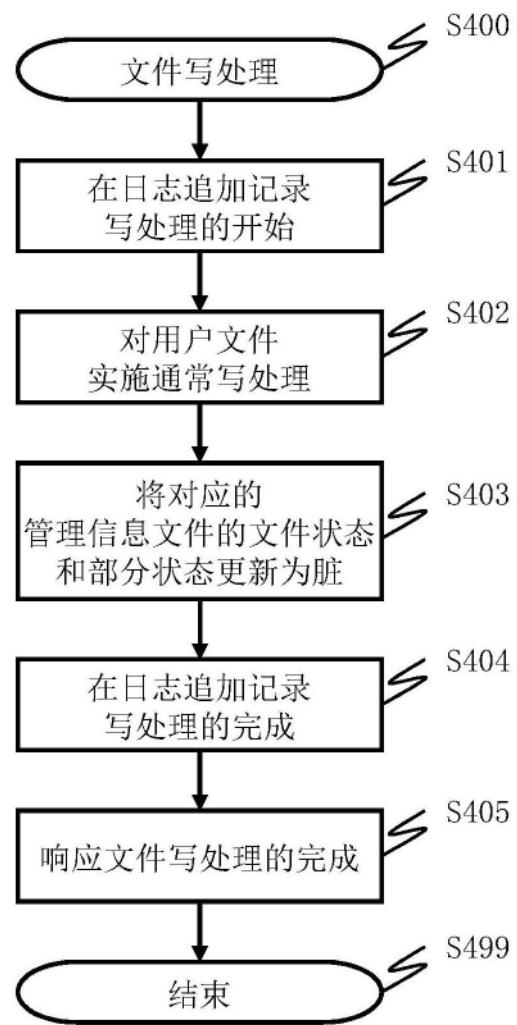


图13

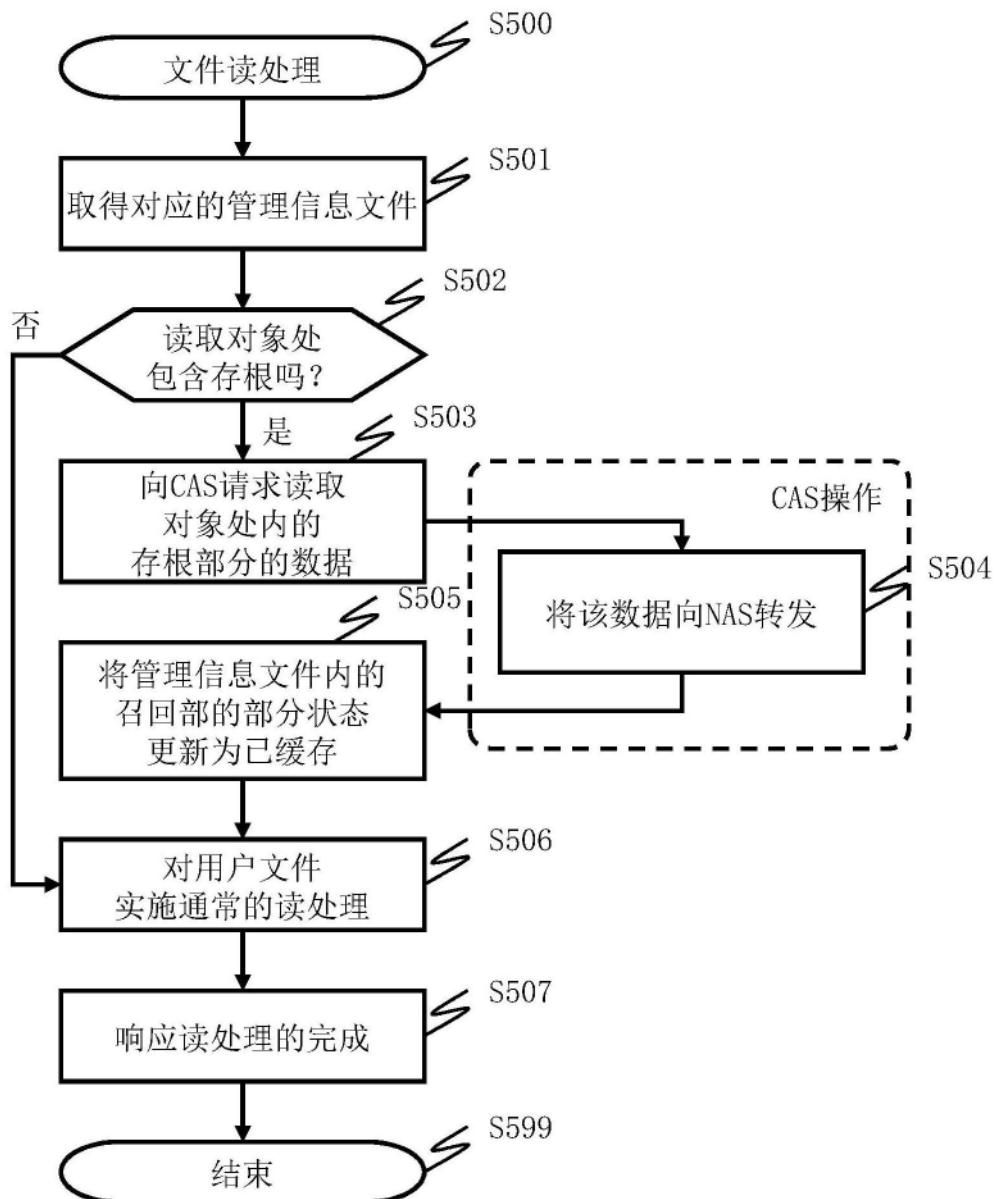


图14

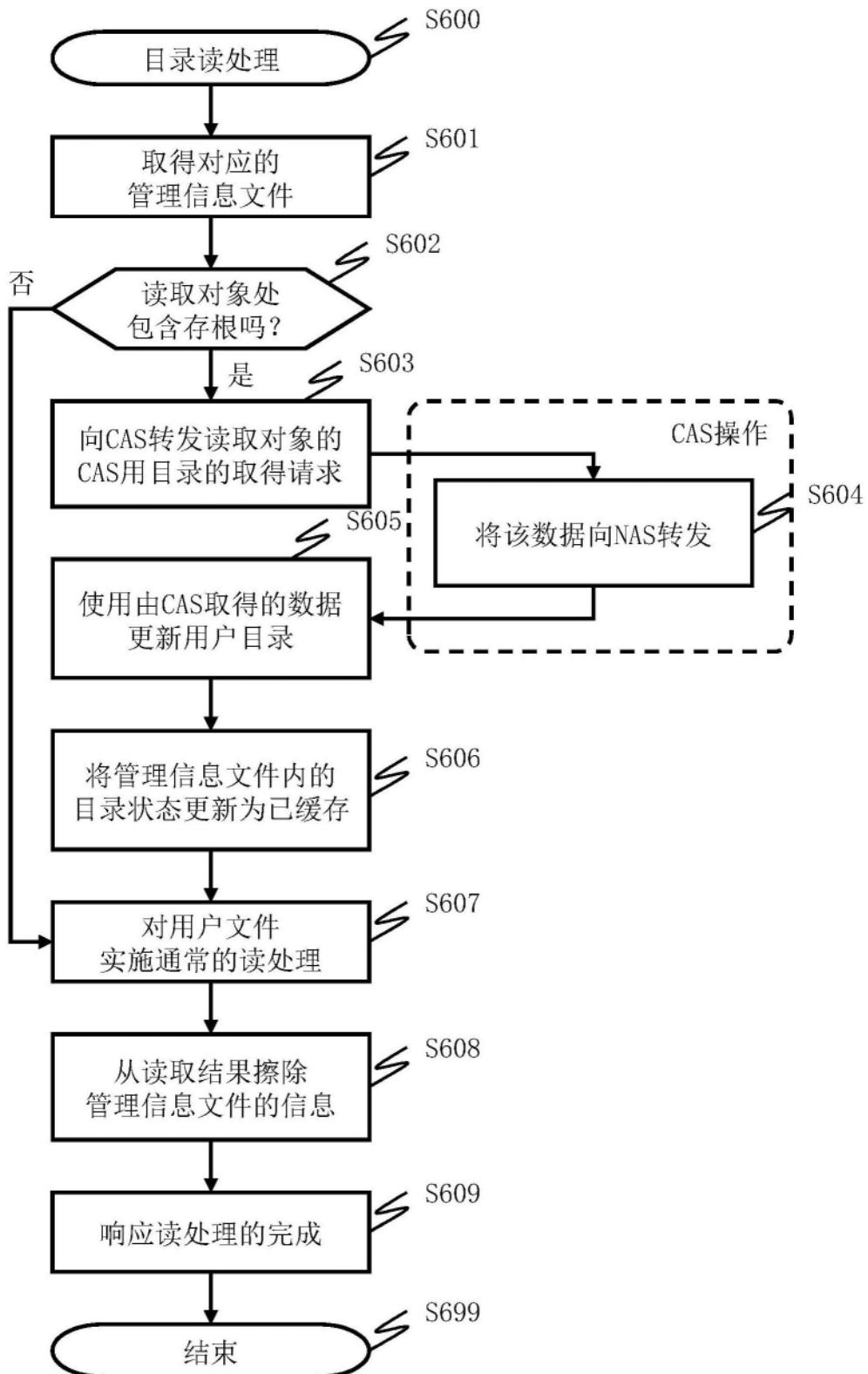


图15

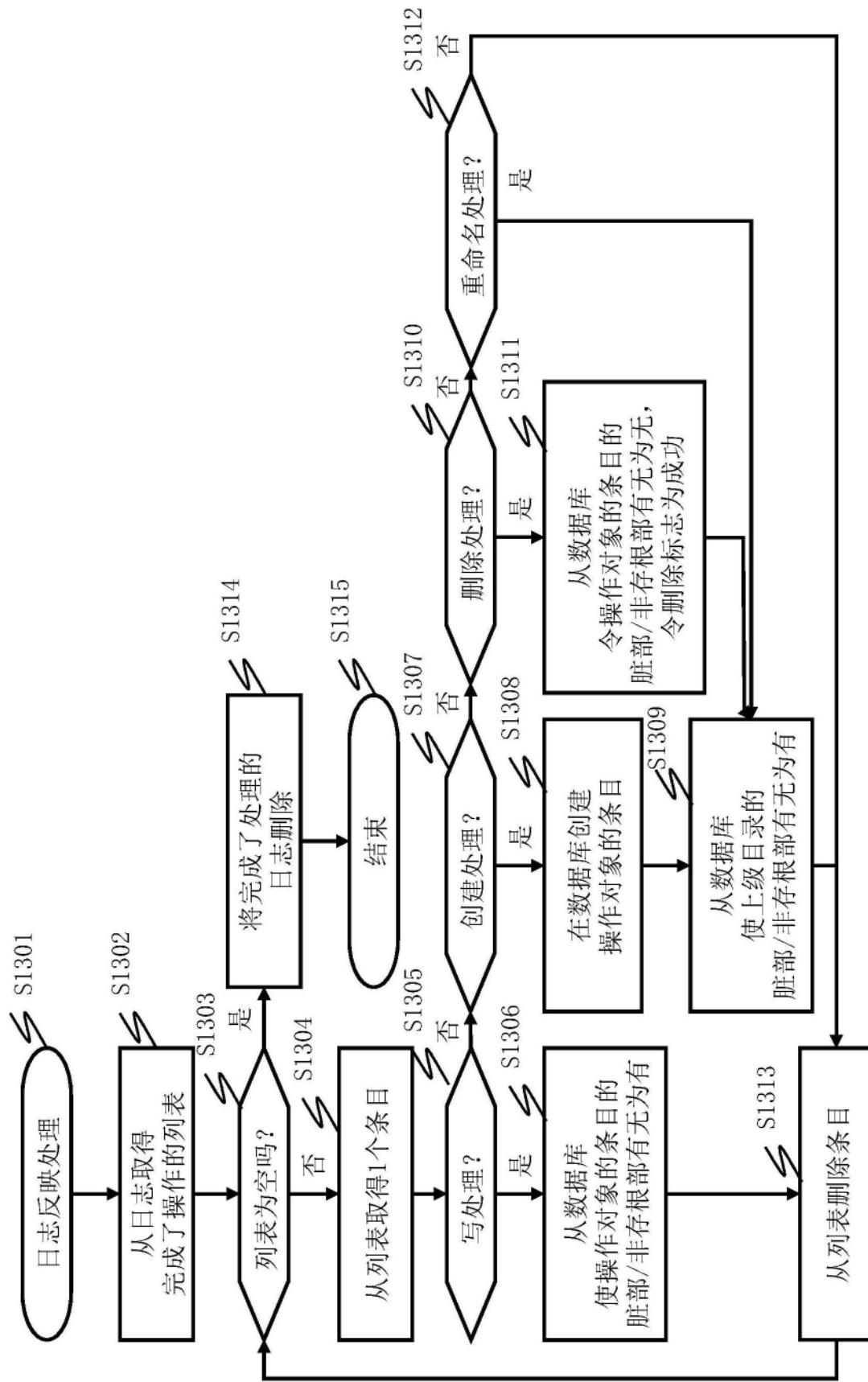


图16

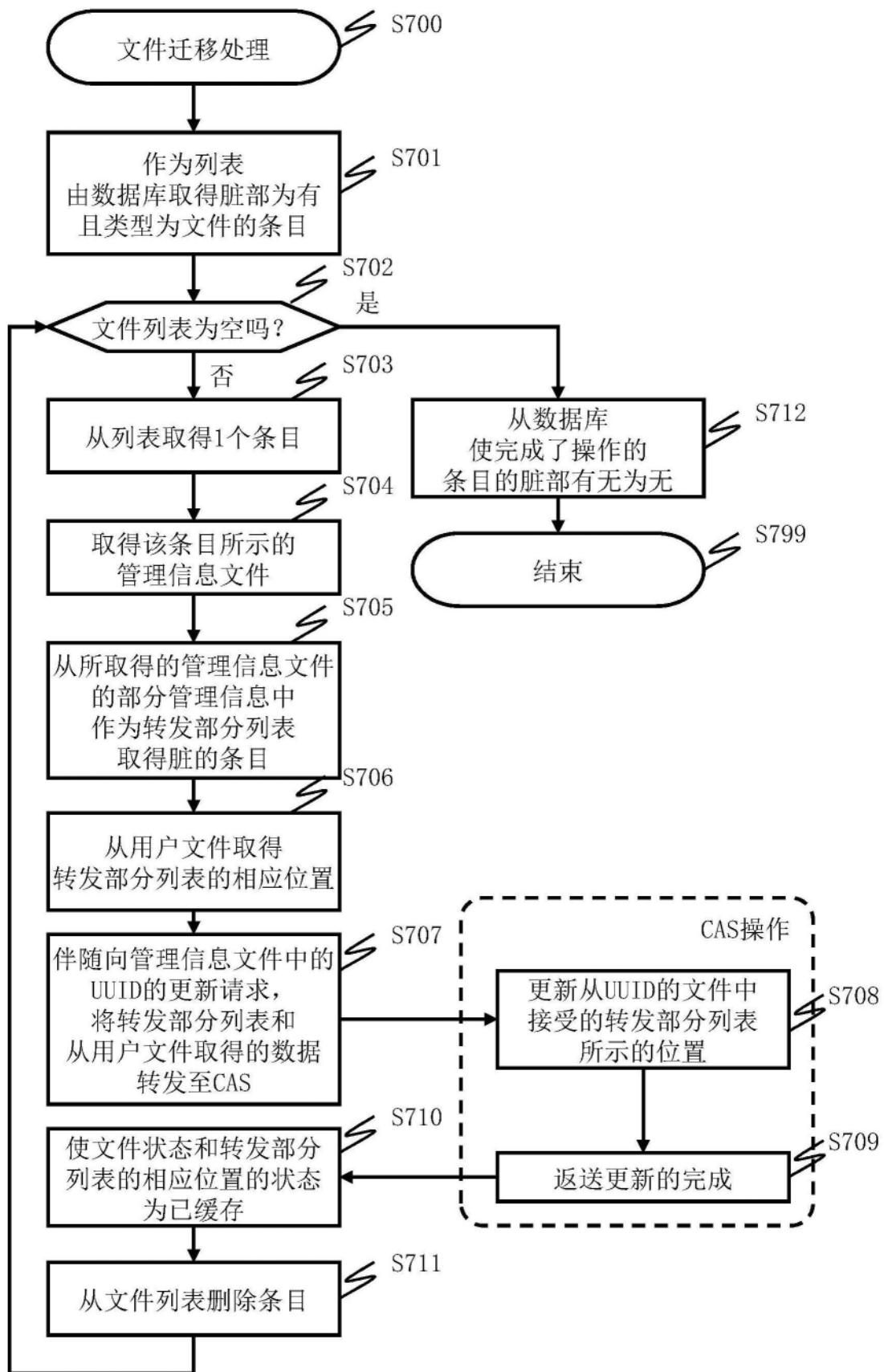


图17

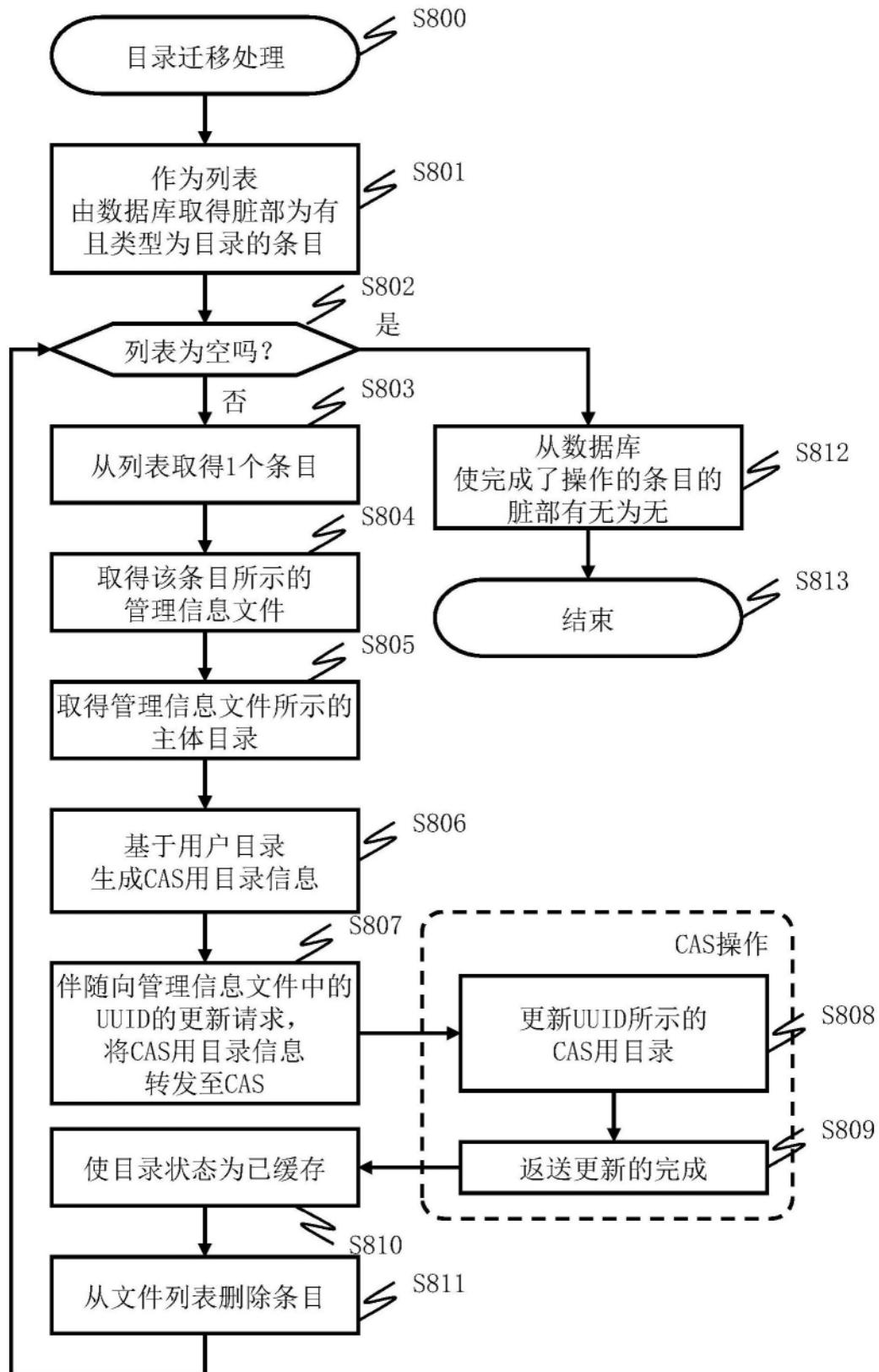


图18

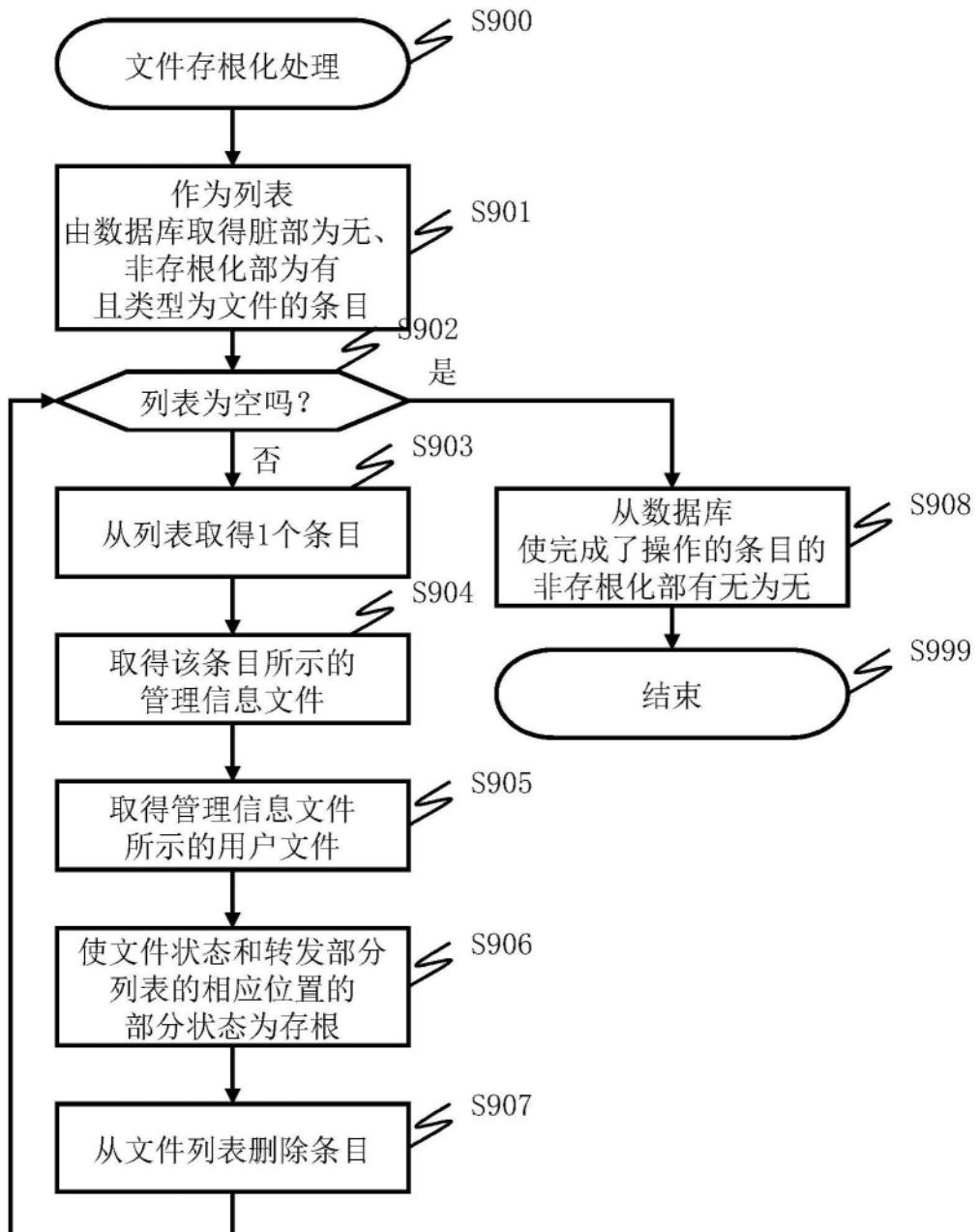


图19

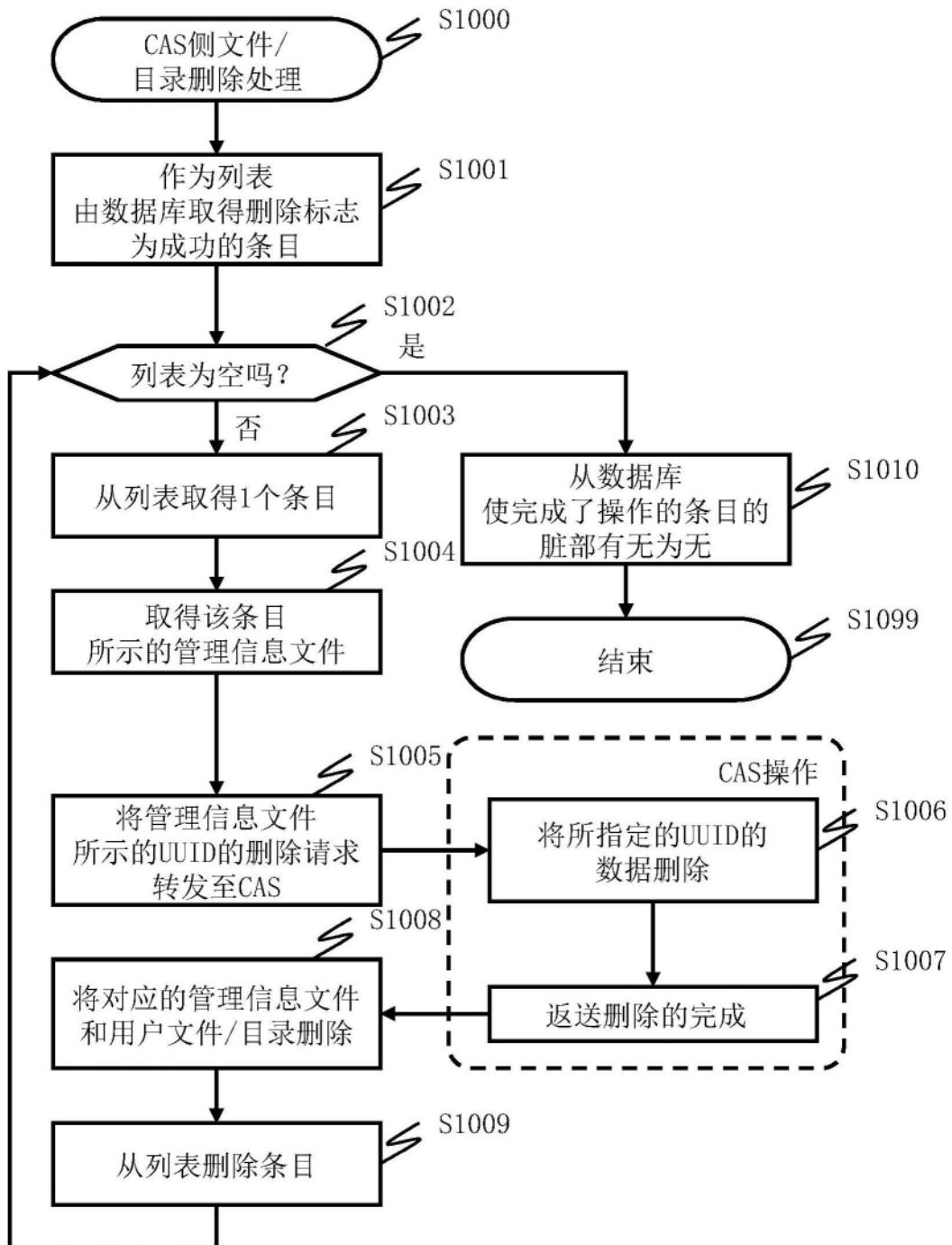


图20

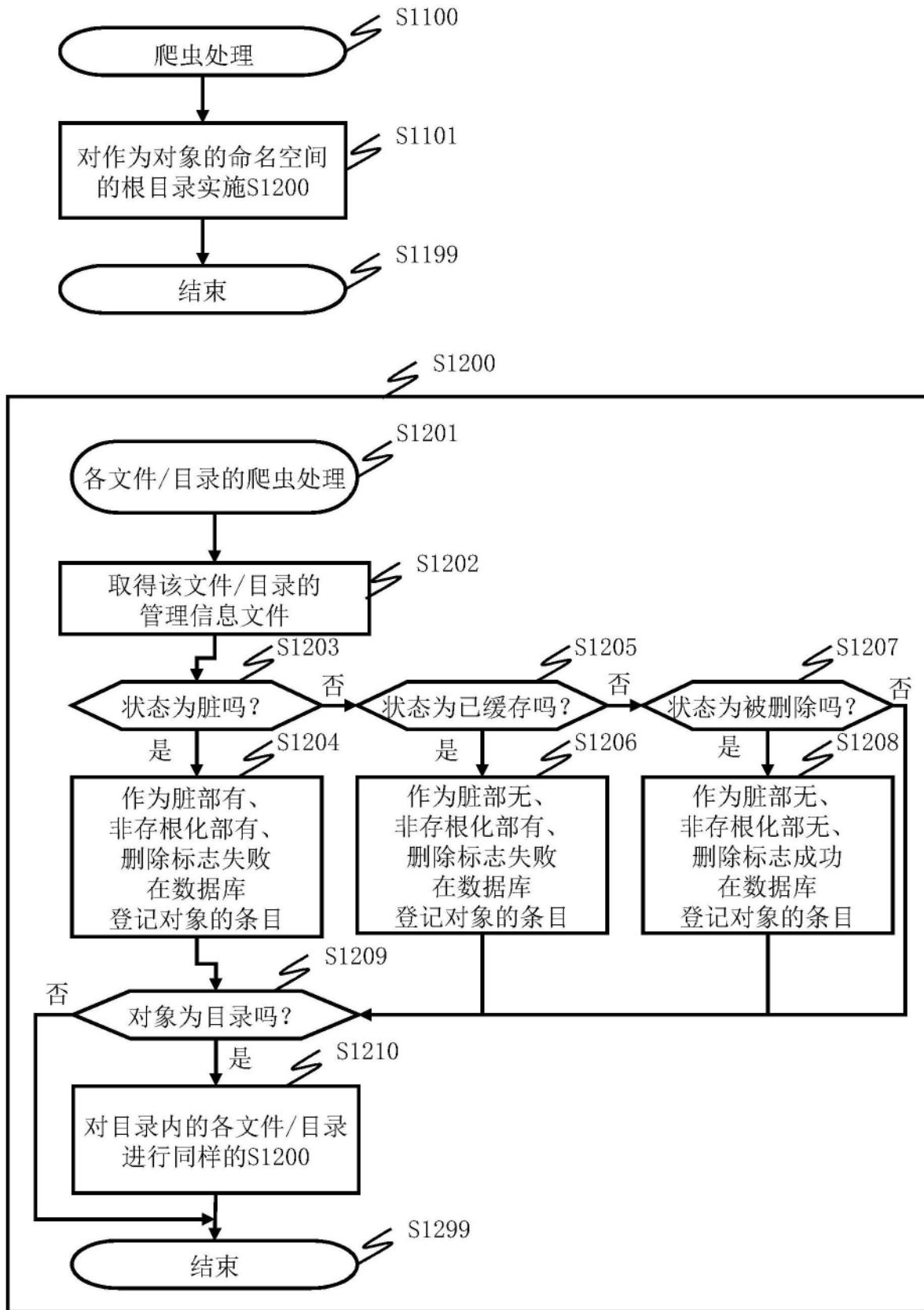


图21