



(10) 授权公告号 CN 110352410 B

(21) 申请号 201780059741.9

(22) 申请日 2017.09.28

(65) 同一申请的已公布的文献号
申请公布号 CN 110352410 A

(43) 申请公布日 2019.10.18

(30) 优先权数据

15/279694 2016.09.29 US

15/279721 2016.09.29 US

(85) PCT国际申请进入国家阶段日
2019.03.27

(86) PCT国际申请的申请数据
PCT/US2017/053990 2017.09.28

(87) PCT国际申请的公布数据
W02018/064319 EN 2018.04.05

(73) 专利权人 华睿泰科技有限责任公司
地址 美国加利福尼亚州

(72) 发明人 B·帕特尔 F·詹姆斯
M·科塔里 A·班纳吉

(74) 专利代理机构 北京纪凯知识产权代理有限公司 11245

专利代理师 赵蓉民

(51) Int.Cl.
G06F 12/0862 (2016.01)
G06F 16/10 (2019.01)

(56) 对比文件

US 8732406 B1, 2014.05.20

US 8732406 B1, 2014.05.20

US 2008140997 A1, 2008.06.12

US 2008140997 A1, 2008.06.12

US 6973542 B1, 2005.12.06

CN 103916465 A, 2014.07.09

US 2013024428 A1, 2013.01.24

US 2008270706 A1, 2008.10.30

US 2005114289 A1, 2005.05.26

US 2015199269 A1, 2015.07.16

US 8180961 B1, 2012.05.15 (续)

审查员 黄建锋

权利要求书7页 说明书16页 附图16页

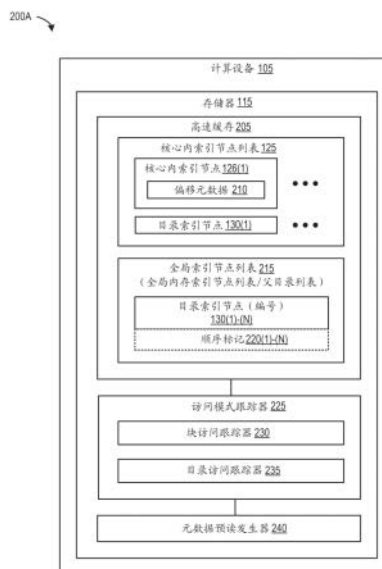
(54) 发明名称

跟踪索引节点的访问模式以及预提取索引节点

(57) 摘要

本文公开的是跟踪索引节点的访问模式并
发出预读指令以将索引节点预提取到存储器中的方法、系统和过程。确定元数据存储区域中元数据单元的位置。确定对应于当前元数据读取操作的元数据存储区域中的另一位置。使用元数据单元的位置和另一位置来确定是否可执行元数据预读操作。响应于确定可执行元数据预读操作而发出元数据预读操作。此外，访问索引节点并且确定索引节点的目录。另外确定目录的条目是否存在于全局索引节点列表中。如果条目存在于全局索引节点列表中，则确定目录的文件结构为顺序的还是非顺序的。如果条目未存在于全局索引节点列表中，则在全局索引节点列表中添加目

录的新条目。



[接上页]

(56) 对比文件

US 2003135514 A1, 2003.07.17

CN 102122284 A, 2011.07.13

US 6266743 B1, 2001.07.24

US 2012259823 A1, 2012.10.11

1. 一种方法,包括:

响应于应用程序发出读取元数据的第一请求,从元数据存储区域读取第一元数据单元;

确定所述元数据存储区域中所述第一元数据单元的末端位置,其中

所述元数据存储区域位于非易失性存储单元中;

响应于所述应用程序发出读取元数据的第二请求,确定所述元数据存储区域中的第二元数据单元的起始位置;

通过比较所述末端位置与所述起始位置来确定是否可执行元数据预读操作;以及

响应于确定所述末端位置和所述起始位置在所述非易失性存储单元中彼此邻近,在所述应用发出读取第三元数据单元的请求之前,通过生成元数据预读操作来执行从所述元数据存储区域读取第三元数据单元。

2. 根据权利要求1所述的方法,其中

所述元数据存储区域包括多个索引节点,

所述第一元数据单元包括所述多个索引节点中的一个或多个,

所述第一元数据单元为结构化文件的部分,并且

所述结构化文件存储在所述非易失性存储单元中。

3. 根据权利要求2所述的方法,还包括:

创建核心内索引节点,其中

所述核心内索引节点对应于所述结构化文件;以及

在所述核心内索引节点中存储所述第一元数据单元的所述末端位置。

4. 根据权利要求3所述的方法,其中另一位置对应于当前元数据读取操作,所述第一元数据单元的所述末端位置标识所述第一元数据单元的末端偏移,并且确定是否可执行所述元数据预读操作包括

访问所述末端偏移,以及

确定所述另一位置是否邻近所述末端偏移。

5. 根据权利要求1所述的方法,其中

发出所述元数据预读操作包括

通过将末端偏移替换为由所述元数据预读操作读取的另一元数据单元的另一末端偏移来更新所述末端偏移,其中

所述另一元数据单元为另一最后读取的元数据块。

6. 根据权利要求4所述的方法,还包括:

如果所述另一位置未邻近所述末端偏移,则通过将所述末端偏移替换为当前元数据读取操作的末端偏移来更新所述末端偏移。

7. 根据权利要求2所述的方法,还包括:

拦截用于读取一个或多个索引节点的命令,其中

所述一个或多个索引节点为所述多个索引节点的部分,

所述拦截响应于输入/输出操作即I/O操作而执行,并且

所述I/O操作导致所述命令;

在所述元数据预读操作中分析所发出的元数据预读值,其中

所述分析包括将所发出的元数据预读值和所述命令中的块总计进行比较；
如果所述分析指示所述I/O操作应当完成，则等待所述I/O操作完成；以及
如果所述分析指示可发出异步元数据预读指令，则发出所述异步元数据预读指令。

8. 根据权利要求7所述的方法，还包括：

确定所述I/O操作是否完成；

如果所述I/O操作完成则生成队列，其中

所述队列包括所述块总计的一个或多个剩余元数据块，并且

所述一个或多个剩余元数据块未包括在所述异步元数据预读指令中；以及

如果所述I/O操作未完成，则更新所述元数据预读操作中的所述块总计。

9. 一种计算机系统，其包括：

一个或多个处理器；

非暂态计算机可读存储介质；

第一读取模块，其响应于应用程序发出读取元数据的第一请求，从元数据存储区域读取第一元数据单元；

第一确定模块，其确定所述元数据存储区域中所述第一元数据单元的末端位置，其中
所述元数据存储区域位于非易失性存储单元中；

第二确定模块，其响应于所述应用程序发出读取元数据的第二请求，确定所述元数据存储区域中的第二元数据单元的起始位置；

第三确定模块，其通过比较所述末端位置与所述起始位置来确定是否可执行元数据预读操作；以及

第一生成模块，其响应于确定所述末端位置与所述起始位置在所述非易失性存储单元中彼此紧邻，在所述应用发出读取第三元数据单元的请求之前，通过生成元数据预读操作来生成所述元数据预读操作以从所述元数据存储区域读取第三元数据单元。

10. 根据权利要求9所述的计算机系统，其中

所述元数据存储区域包括多个磁盘上索引节点，

所述第一元数据单元包括所述多个磁盘上索引节点中的一个或多个，

所述第一元数据单元为结构化文件的部分，并且

所述结构化文件存储在所述非易失性存储单元中。

11. 根据权利要求10所述的计算机系统，其中所述计算机系统还包括：

创建模块，其创建核心内索引节点，其中

所述核心内索引节点对应于所述结构化文件；以及

存储模块，其在所述核心内索引节点中存储所述第一元数据单元的所述末端位置。

12. 根据权利要求11所述的计算机系统，其中另一位置对应于当前元数据读取操作，所述末端位置标识所述第一元数据单元的末端偏移，并且所述第三确定模块还包括：

访问模块，其访问所述末端偏移，以及

确定模块，其确定所述另一位置是否邻近所述末端偏移。

13. 根据权利要求12所述的计算机系统，其中所述第一读取模块包括：

第一更新模块，如果所述另一位置邻近所述末端偏移，则所述第一更新模块通过将所述末端偏移替换为由所述元数据预读操作读取的另一元数据单元的另一末端偏移来更新

所述末端偏移,其中

所述另一元数据单元为另一最后读取的元数据块;以及

第二更新模块,如果所述另一位置未邻近所述末端偏移,则所述第二更新模块通过将所述末端偏移替换为当前元数据读取操作的末端偏移来更新所述末端偏移。

14. 根据权利要求10所述的计算机系统,其中所述计算机系统还包括:

拦截模块,其拦截用于读取一个或多个磁盘上索引节点的命令,其中

所述一个或多个磁盘上索引节点为所述多个磁盘上索引节点的部分,

所述拦截模块响应于输入/输出操作即I/O操作,并且

所述I/O操作导致所述命令;

分析模块,其在所述元数据预读操作中分析发出的元数据预读值,其中

所述分析模块包括:

比较模块,其将所发出的元数据预读值和所述命令中的块总计进行比较;

等待模块,如果所述分析模块指示所述I/O操作应当完成,则所述等待模块等待所述I/O操作完成;

发出模块,如果所述分析模块指示可发出异步元数据预读指令,则所述发出模块发出所述异步元数据预读指令;

确定模块,其确定所述I/O操作是否完成;

生成模块,如果所述I/O操作完成则所述生成模块生成队列,其中

所述队列包括所述块总计的一个或多个剩余元数据块,并且

所述一个或多个剩余元数据块未包括在所述异步元数据预读指令中;以及

更新模块,如果所述I/O操作未完成,则所述更新模块更新所述元数据预读操作中的所述块总计。

15. 一种系统,包括:

一个或多个处理器;和

存储器,所述存储器耦接到所述一个或多个处理器,其中所述存储器存储能够由所述一个或多个处理器执行的程序指令,所述程序指令用于:

响应于应用程序发出读取元数据的第一请求,从元数据存储区域读取第一元数据单元;

确定所述元数据存储区域中所述第一元数据单元的末端位置,其中

所述元数据存储区域位于非易失性存储单元中;

响应于所述应用程序发出读取元数据的第二请求,确定所述元数据存储区域中的第二元数据单元的起始位置;

通过比较所述末端位置与所述起始位置来确定是否可执行元数据预读操作;以及

响应于确定所述末端位置与所述起始位置在所述非易失性存储单元中彼此邻近,在所述应用发出读取第三元数据单元的请求之前,通过生成元数据预读操作来执行从所述元数据存储区域读取第三元数据单元。

16. 根据权利要求15所述的系统,其中

所述元数据存储区域包括多个索引节点,

所述第一元数据单元包括所述多个索引节点中的一个或多个,所述第一元数据单元为

结构化文件的部分,并且

所述结构化文件存储在所述非易失性存储单元中。

17. 根据权利要求15所述的系统,其中能够由所述一个或多个处理器执行的所述程序指令还包括能够由所述一个或多个处理器执行以进行以下操作的程序指令:

创建核心内索引节点,其中

所述元数据存储区域包括多个索引节点,

所述索引节点包括所述第一元数据单元,

所述索引节点为结构化文件的部分,

所述结构化文件被存储在所述非易失性存储单元中,

所述核心内索引节点对应于所述结构化文件;以及

在所述核心内索引节点中存储所述第一元数据单元的所述末端位置。

18. 根据权利要求17所述的系统,其中另一位置对应于当前元数据读取操作,所述第一元数据单元的所述末端位置标识所述第一元数据单元的末端偏移,并且能够由所述一个或多个处理器执行以确定是否可执行所述元数据预读操作的所述程序指令还包括能够由所述一个或多个处理器执行以进行以下操作的程序指令:

访问所述末端偏移,以及

确定所述另一位置是否邻近所述末端偏移。

19. 根据权利要求18所述的系统,其中能够由所述一个或多个处理器执行以发出所述元数据预读操作的所述程序指令还包括能够由所述一个或多个处理器执行以进行以下操作的程序指令:

如果所述另一位置邻近所述末端偏移,则通过将所述末端偏移替换为由所述元数据预读操作读取的另一元数据单元的另一末端偏移来更新所述末端偏移,其中

所述另一元数据单元为另一最后读取的元数据块;以及

如果所述另一位置未邻近所述末端偏移,则通过将所述末端偏移替换为当前元数据读取操作的末端偏移来更新所述末端偏移。

20. 根据权利要求16所述的系统,能够由所述一个或多个处理器执行的所述程序指令还包括能够由所述一个或多个处理器执行以进行以下操作的程序指令:

拦截用于读取一个或多个索引节点的命令,其中

所述一个或多个索引节点为所述多个索引节点的部分,

能够由所述一个或多个处理器执行以进行拦截的所述程序指令响应于输入/输出操作即I/O操作而执行,并且

所述I/O操作导致所述命令;

在所述元数据预读操作中分析发出的元数据预读值,其中

能够由所述一个或多个处理器执行以进行分析的所述程序指令包括能够由所述一个或多个处理器执行以将所发出的元数据预读值和所述命令中的块总计进行比较的程序指令;

如果能够由所述一个或多个处理器执行以进行分析的程序指令指示所述I/O操作应当完成,则等待所述I/O操作完成;

如果能够由所述一个或多个处理器执行以进行分析的程序指令指示可发出异步元数

据预读指令,则发出所述异步元数据预读指令;

确定所述I/O操作是否完成;

如果所述I/O操作完成则生成队列,其中

所述队列包括所述块总计的一个或多个剩余元数据块,并且

所述一个或多个剩余元数据块未包括在所述异步元数据预读指令中;以及

如果所述I/O操作未完成,则更新所述元数据预读操作中的所述块总计。

21.一种方法,包括:

确定索引节点的目录;

确定所述目录的文件结构为顺序的还是非顺序的;

响应于确定所述目录的文件结构为顺序的,确定所述目录是否包括顺序标记;

响应于确定所述目录包括所述顺序标记,针对索引节点列表中的多个索引节点中的一个或多个索引节点发出元数据预读操作,其中所述元数据预读操作包括元数据预读值;

拦截用于读取所述索引节点列表中的所述多个索引节点中的一个或多个索引节点中的命令,其中

所述命令与输入/输出操作即I/O操作相关联,并且

所述命令包括要与所述命令一起被读取的块的总数(“块总计”);

将所述元数据预读值与所述块总计进行比较;以及

响应于确定所述元数据预读值小于或等于所述块总计,发出异步元数据预读指令,其中

所述异步元数据预读指令指示要预读的索引节点的数量。

22.根据权利要求21所述的方法,还包括:

访问所述索引节点,其中

在确定所述索引节点的所述目录之前执行所述访问。

23.根据权利要求22所述的方法,其中提取所述索引节点列表。

24.根据权利要求23所述的方法,其中

所述输入/输出操作即I/O操作导致所述命令。

25.根据权利要求24所述的方法,还包括:

如果所述比较指示所述I/O操作应当完成,则等待所述I/O操作完成;以及

确定所述I/O操作是否完成。

26.根据权利要求25所述的方法,还包括:

响应于确定所述I/O操作完成,生成队列,其中

所述队列包括所述块总计的一个或多个剩余元数据块,并且

所述一个或多个剩余元数据块未包括在所述异步元数据预读指令中。

27.根据权利要求21所述的方法,其中

所述索引节点为磁盘上索引节点。

28.一种计算机系统,其包括

一个或多个处理器;

非暂态计算机可读存储介质;

第一确定模块,其确定索引节点的目录;

第二确定模块,如果条目存在于全局索引节点列表中,则所述第二确定模块确定所述目录的文件结构为顺序的还是非顺序的;

第三确定模块,其响应于确定所述目录的文件结构是顺序的,确定所述目录是否顺序标记;

发出模块,其响应于确定所述目录包括所述顺序标记,针对索引节点列表中的多个索引节点中的一个或多个索引节点发出元数据预读操作,其中所述元数据预读操作包括元数据预读值;

拦截模块,其拦截用于读取所述索引节点列表中的所述多个索引节点中的一个或多个索引节点的操作,其中

所述命令与输入/输出操作即I/O操作相关联,并且

所述命令包括要与所述命令一起被读取的块的总数(“块总计”);

比较模块,其将所述元数据预读值与所述块总计进行比较;以及

发出模块,其响应于确定所述元数据预读值小于或等于所述块总计,发出异步元数据预读指令,其中

所述异步元数据预读指令指示要预读的索引节点的数量。

29. 根据权利要求28所述的计算机系统,还包括:

访问模块,其访问所述索引节点;以及

提取模块,其提取所述索引节点列表。

30. 根据权利要求29所述的计算机系统,其中:

输入/输出操作即I/O操作导致所述命令。

31. 根据权利要求30所述的计算机系统,还包括:

等待模块,如果所述比较模块指示所述I/O操作应当完成,则所述等待模块等待所述I/O操作完成;以及

第四确定模块,其确定所述I/O操作是否完成。

32. 根据权利要求31所述的计算机系统,还包括:

生成模块,其响应于确定所述I/O操作完成而生成队列,其中

所述队列包括所述块总计的一个或多个剩余元数据块,并且

所述一个或多个剩余元数据块未包括在所述异步元数据预读指令中。

33. 一种系统,包括:

一个或多个处理器;和

存储器,所述存储器耦接到所述一个或多个处理器,其中所述存储器存储能够由所述一个或多个处理器执行的程序指令,所述程序指令用于执行包括以下操作的方法:

确定索引节点的目录,

确定所述目录的文件结构为顺序的还是非顺序的,

响应于确定所述目录的所述文件结构为顺序的,确定目录是否包括顺序标记;

响应于确定所述目录包括所述顺序标记,针对索引节点列表中的多个索引节点中的一个或多个索引节点发出元数据预读操作,其中所述元数据预读操作包括元数据预读值;

拦截用于读取所述索引节点列表中的所述多个索引节点中的一个或多个索引节点的操作,其中

所述命令与输入/输出操作即I/O操作相关联,并且
所述命令包括要与所述命令一起被读取的块的总数(“块总计”);
将所述元数据预读值与所述块总计进行比较;并且
响应于确定所述元数据预读值小于或等于所述块总计,发出异步元数据预读指令,其中

所述异步元数据预读指令指示要预读的索引节点的数量。

34. 根据权利要求33所述的系统,其中所述方法还包括:

访问所述索引节点,其中

在确定所述索引节点的所述目录之前执行所述访问;以及

提取所述索引节点列表。

35. 根据权利要求34所述的系统,其中:

所述输入/输出操作即I/O操作导致所述命令。

36. 根据权利要求35所述的系统,其中所述方法还包括:

如果所述比较指示所述I/O操作应当完成,则等待所述I/O操作完成;以及

确定所述I/O操作是否完成。

37. 根据权利要求36所述的系统,其中所述方法还包括:

响应于确定所述I/O操作完成,生成队列,其中

所述队列包括所述块总计的一个或多个剩余元数据块,并且

所述一个或多个剩余元数据块未包括在所述异步元数据预读指令中。

跟踪索引节点的访问模式以及预提取索引节点

技术领域

[0001] 本公开涉及数据访问。具体而言，本公开涉及跟踪索引节点的访问模式、以及发出索引节点预读指令以预提取索引节点。

[0002] 相关技术描述

[0003] 文件系统出于计算目的用于控制如何存储和检索数据（例如，用于存储和执行应用程序）。文件系统的数据对象（例如，文件、目录等）具有一个或多个索引节点。索引节点是数据结构，该数据结构用于标识属于文件系统中数据对象的数据。索引节点存储属性（例如，元数据）和数据对象的数据的磁盘块位置。

[0004] 访问文件系统中的文件需要从磁盘（例如，从非易失性存储单元）读取文件的索引节点。诸如备份、定期扫描、管理操作等的数据库操作通常访问磁盘上的多个索引节点。从磁盘读取此类“磁盘上”索引节点可不利地影响应用程序性能。例如，如果底层的磁盘缓慢，则从磁盘读取磁盘上索引节点可导致用所需数据为应用程序提供服务之前不合理和/或显著的输入/输出 (I/O) 等待时间。

[0005] 文件的内容可被载入存储器（例如，随机存取存储器 (RAM)），使得当随后访问文件时，从RAM而非磁盘（例如，硬盘驱动器 (HDD)）读取文件的内容。然而，将文件的内容载入存储器需要索引节点，该索引节点与要从磁盘访问的文件内容对应。

发明内容

[0006] 本文公开的是跟踪索引节点的访问模式以通过将索引节点预提取到存储器中来加快数据访问的方法、系统和过程。一种此类方法涉及确定元数据存储区域中元数据单元的位置，该数据存储区域位于非易失性存储单元中。该方法确定与当前元数据读取操作对应的元数据存储区域中的另一位置，并且使用元数据单元的位置以及另一位置确定是否可执行元数据预读操作。响应于确定可执行元数据预读操作，该方法发出元数据预读操作。

[0007] 在某些实施方案中，元数据存储区域包括磁盘上索引节点。磁盘上索引节点包括元数据单元并且为结构化文件的部分。结构化文件存储在非易失性存储单元中。该方法创建与结构化文件对应的核心内索引节点并将元数据单元的位置存储在核心内索引节点中。在该示例中，元数据单元是最后读取的元数据块，并且元数据单元的位置标识并包括元数据单元的末端偏移。

[0008] 在一些实施方案中，确定是否可执行元数据预读操作包括访问元数据单元的末端偏移并确定另一位置是否邻近末端偏移。在该示例中，发出元数据预读操作包括通过将末端偏移替换为由元数据预读操作读取的另一元数据单元（例如，另一最后读取的元数据块）的另一末端偏移来更新末端偏移，并且如果另一位置未邻近末端偏移，则通过将末端偏移替换为当前元数据读取操作的末端偏移来更新末端偏移。

[0009] 在其他实施方案中，该方法响应于输入/输出 (I/O) 操作而拦截用于读取一个或多个磁盘上索引节点的命令。在该示例中，该方法通过将发出的预读值以及命令中的块总计进行比较，分析元数据预读操作中发出的元数据预读值。该方法等待I/O操作完成并且/或

者发出异步元数据预读指令。

[0010] 在某些实施方案中,确定I/O操作是否完成包括在I/O操作完成的情况下生成队列。队列包括未包括在异步元数据预读指令中的块总计的剩余元数据块。如果I/O操作未完成,则该方法更新元数据预读操作中的块总计。

[0011] 一些实施方案涉及访问索引节点、确定索引节点(例如,磁盘上索引节点)的目录、以及确定目录的条目是否存在于全局索引节点列表中。如果条目存在于全局索引节点列表中,则该方法确定目录的文件结构是为顺序的还是非顺序的,并且如果条目未存在于全局索引节点列表中,则该方法会在全局索引节点列表中为目录添加新条目。

[0012] 在某些实施方案中,确定目录是否存在于全局索引节点列表中包括在全局索引节点列表中搜索条目。如果目录的文件结构为顺序,则该方法确定目录是否包括顺序标记。如果目录包括顺序标记,则该方法为目录中的索引节点发出元数据预读操作。如果父目录的文件结构为非顺序的,则该方法提取目录的索引节点列表并为索引节点列表上的索引节点发出元数据预读操作。在一些实施方案中,该方法响应于输入/输出(I/O)操作而拦截用于读取一个或多个磁盘上索引节点的命令。在该示例中,该方法通过将发出的元数据预读值以及命令中的块总计进行比较,分析元数据预读操作中发出的元数据预读值。该方法等待I/O操作完成并且/或者发出异步元数据预读指令。

[0013] 在其他实施方案中,确定I/O操作是否完成包括在I/O操作完成的情况下生成队列。队列包括未包括在异步元数据预读指令中的块总计的剩余元数据块。如果I/O操作未完成,则该方法更新元数据预读操作中的块总计。

[0014] 以上内容是概述,因此必然包含对细节的简化、概括和省略;因此本领域的技术人员将会理解,该概述仅是说明性的,并不意图进行任何限制。如由权利要求单独定义的那样,本公开的其他方面、发明特征和优点将在以下阐述的非限制性详细描述中变得显而易见。

附图说明

[0015] 通过参考附图,可以更好地理解本公开,并且其多个目的、特征和优点对于本领域的技术人员而言是显而易见的。

[0016] 图1A是根据本公开的一个实施方案的跟踪索引节点和预提取索引节点的访问模式的系统的框图。

[0017] 图1B是根据本公开的一个实施方案的结构化文件的框图。

[0018] 图1C是根据本公开的一个实施方案的结构化文件和核心内索引节点的框图。

[0019] 图2A是根据本公开的一个实施方案的跟踪索引节点的访问模式并发出元数据预读指令的系统的框图。

[0020] 图2B是根据本公开的一个实施方案的索引节点的偏移元数据的框图。

[0021] 图2C是根据本公开的一个实施方案的示出全局索引节点列表/父目录列表的内容的表格。

[0022] 图2D是根据本公开的一个实施方案的目录访问跟踪器的框图。

[0023] 图2E是根据本公开的一个实施方案的元数据预读发生器的框图。

[0024] 图3A是根据本公开的一个实施方案的带顺序索引节点的目录的框图。

- [0025] 图3B是根据本公开的一个实施方案的带非顺序索引节点的目录的框图。
- [0026] 图3C是根据本公开的一个实施方案的实现全局索引节点列表的高速缓存的框图。
- [0027] 图4A是根据本公开的一个实施方案的示出了用于执行索引节点预提取的过程的流程图。
- [0028] 图4B是根据本公开的一个实施方案的示出了用于存储与磁盘上索引节点相关的偏移元数据的过程的流程图。
- [0029] 图5A是根据本公开的一个实施方案的示出了用于确定目录的文件结构的过程的流程图。
- [0030] 图5B是根据本公开的一个实施方案的示出了用于为磁盘上索引节点发出元数据预读指令的过程的流程图。
- [0031] 图6A是根据本公开的一个实施方案的示出了用于处理与索引节点预提取相关的输入/输出 (I/O) 操作的过程的流程图。
- [0032] 图6B是根据本公开的一个实施方案的示出了用于处理与索引节点预提取相关的 I/O 操作的过程的流程图。
- [0033] 图7A是根据本公开的一个实施方案的示出了用于处理磁盘上索引节点的访问的过程的流程图。
- [0034] 图7B是根据本公开的一个实施方案的示出了用于处理对访问磁盘上索引节点的请求的过程的流程图。
- [0035] 图8是根据本公开的一个实施方案的计算系统的框图,示出了如何可在软件中实现访问模式跟踪器和元数据预读发生器。
- [0036] 图9是根据本公开的一个实施方案的联网系统的框图,示出了各种计算设备如何可经由网络进行通信。
- [0037] 虽然本公开容许各种修改和替代形式,但是在附图和详细描述中作为示例提供了本公开的特定实施方案。应当理解,附图和详细描述并非旨在将本公开限制于所公开的特定形式。相反,其目的在于覆盖落入由所附权利要求书限定的本公开的实质和范围内的所有修改形式、等同物和替代物。

具体实施方式

[0038] 介绍

[0039] 文件系统(例如,Unix文件系统)用于组织数据并控制数据的存储和检索方式。文件系统负责组织数据对象诸如文件和目录,并负责保持跟踪存储设备(例如,硬盘驱动器(HDD)、固态驱动器(SSD)等)的哪些区域属于哪些数据对象。通常,文件系统每个数据对象(例如,文件、目录等)具有对应的索引节点。

[0040] 索引节点是用于定位文件系统的数据的数据结构。索引节点存储数据对象的数据的属性(例如,元数据)和磁盘块位置,并且可由整数(例如,称为索引节点编号)标识。目录可包括分配至索引节点的名称的列表。目录包含自身的条目、目录的父项的条目、以及目录的子项中的每个项的条目。

[0041] 访问文件系统中的文件通常涉及要从磁盘(例如,从非易失性存储单元)读取的文件的索引节点,例如,由此确定对文件内容的更改和/或修改,并且在某些情况下,还由此验

证所有者和权限信息(例如,组ID、用户ID等)。因此,访问文件的内容之前,必须先从磁盘读取文件的索引节点(元数据)。

[0042] 数据操作(例如,输入/输出(I/O)操作)通常需要访问磁盘上的多个索引节点,并且从磁盘读取此类“磁盘上”索引节点可不利地影响应用程序性能(例如,以给定I/O操作能够完成之前的I/O等待时间的形式)。如前文所述,文件的内容可被“预读”并载入存储器(例如,随机存取存储器(RAM)),使得当随后访问文件时,从RAM而非磁盘(例如,HDD)读取文件的内容。因此,以此方式预提取数据(例如,为了加快数据访问)需要跟踪索引节点的访问模式以“预读”这些索引节点之后,与那些索引节点相关的数据(例如,文件、目录等)可优先载入存储器之前。

[0043] 遗憾的是,和文件数据不同,索引节点元数据访问模式的有效跟踪由于至少两个原因而存在挑战性。首先,多个I/O操作(例如,来自以集群执行的多个应用程序)可同时访问同一索引节点。跟踪多个I/O操作的索引节点访问模式可为存储器和计算资源密集的,并且可导致显著的开销。其次,索引节点访问模式的跟踪还需要此类索引节点的有效“预读”(例如,通过发出预读指令),也不会不利地影响系统性能。

[0044] 本文所公开的为:除了其他功能之外,能够基于块访问、顺序访问、和非顺序访问来跟踪索引节点的访问模式,并为索引节点发出预读指令的方法、系统和过程。

[0045] 跟踪索引节点访问模式并发出预读的示例性系统

[0046] 图1A是根据一个实施方案的计算系统100A的框图,该计算系统被配置为跟踪索引节点和预提取索引节点的访问模式。如图1A中所示,计算设备105包括处理器110和存储器115。计算设备105可为任何类型的计算系统,包括服务器、台式电脑、笔记本电脑、平板电脑等,并且经由网络185通信地耦接到存储系统145。网络185可为任何类型的网络和/或互连(例如,局域网(LAN)、广域网(WAN)、存储区域网(SAN)、互联网等)。

[0047] 存储系统145可包括多种不同的存储设备,包括硬盘、光盘、数字通用光盘、SSD存储器诸如闪存存储器等、或者一个或多个逻辑存储设备诸如在一个或多个此类物理存储设备上实现的卷。存储系统145包括此类存储设备(例如,磁盘150)中的一者或多者。在一个实施方案中,磁盘150为非易失性存储单元。在其他实施方案中,磁盘150为HDD或SSD。磁盘150包括文件系统155。文件系统155可为任何类型的文件系统(例如,Unix文件系统、基于范围的文件系统等)。

[0048] 操作系统120、以及应用程序135和140存储在存储器115中,并且由操作系统120执行。操作系统120还包括核心内索引节点列表125(例如,带有多个核心内索引节点126(1)-(N)),该列表包括目录索引节点130(1)-(N)。例如,目录索引节点130(1)-(N)为核心内索引节点列表125(其包括文件、目录等的核心内索引节点)中磁盘上索引节点的父目录表示。文件系统155包括结构化文件160、目录数据结构170、以及数据180。结构化文件160包括磁盘上索引节点165(1)-(N),而目录数据结构170包括索引节点列表175。在一个实施方案中,结构化文件160为iList文件。在该实施方案中,iList文件为保留磁盘上索引节点(例如,磁盘上索引节点165(1)-(N))的列表的文件。

[0049] 文件系统155中的数据对象(例如,文件、目录等)与对应的索引节点(例如,磁盘上索引节点165(1)-(N))相关联。每个磁盘上索引节点具有特定的索引节点编号并且存储在磁盘150上(例如,作为磁盘上索引节点165(1)-(N),并且作为结构化文件160的部分)。核心

内索引节点列表125是一个(或多个)磁盘上索引节点的内存数据结构(或列表)。核心内索引节点列表125包括作为磁盘上索引节点165(1)-(N)的部分存储的元数据、以及其他另外的元数据。

[0050] 文件系统155包括目录数据结构170。目录数据结构170包括索引节点列表175。数据对象(例如,文件、子目录等)的父项为给定数据对象为其部分的目录。例如,如果给定目录(例如,/home/john)包含四个数据对象(例如,分别带有路径名称/home/john/file1、/home/john/file2、/home/john/file3以及/home/john/dir1的file1、file2、file3和dir1),则这些目录条目(例如,file1、file2、file3以及dir1)的父目录为“john”。每个数据对象还具有对应的和/或相关联的磁盘上索引节点(例如,“john”的磁盘上索引节点165(1)、“file1”的磁盘上索引节点165(2)、“file2”的磁盘上索引节点165(3)以及“file3”的磁盘上索引节点165(4)和“dir1”的磁盘上索引节点165(5))。在该情况下,磁盘上索引节点165(2)、165(3)、165(4)和165(5)的父(目录)索引节点编号为磁盘上索引节点165(1)(为了清楚起见显示为核心内索引节点列表125中的目录索引节点130(1))。

[0051] 图1B是根据一些实施方案的结构化文件的框图100B,并且图1C是根据一些实施方案的结构化文件及核心内索引节点的框图100C。文件系统155中的目录包括元组(例如,<文件名称、文件的索引节点编号>)的列表。索引节点列表175是文件系统155中此类元组列表的一个示例。磁盘上索引节点165(1)的数据部分包括<file1、磁盘上索引节点165(2)>、<file2、磁盘上索引节点165(3)>、<file3、磁盘上索引节点165(4)>、以及<dir1、磁盘上索引节点165(5)>。如图1B中所示,这些“磁盘上”索引节点作为结构化文件160(例如,iList文件)的数据保留。因为结构化文件160也为文件,所以结构化文件160还具有其自己的带唯一索引节点编号的索引节点。如图1C中所示,当将结构化文件160(1)引入核心内(例如,引入到存储器115中)时,为结构化文件160(1)创建核心内索引节点(例如,核心内索引节点列表诸如核心内索引节点列表125中的核心内索引节点126(1))。每个结构化文件和磁盘上索引节点具有相关联的核心内索引节点。

[0052] 图2A是根据一个实施方案的跟踪索引节点的访问模式并发出元数据预读指令的计算系统200A的框图。如图2A中所示,计算设备105包括存储器115。存储器115实现高速缓存205、访问模式跟踪器225、以及元数据预读发生器240。高速缓存205实现全局索引节点列表215,该列表包括来自核心内索引节点列表的一个或多个条目。核心内索引节点126(1)是为结构化文件160(1)创建的内存数据结构(或列表)。

[0053] 核心内索引节点126(1)包括偏移元数据210。全局索引节点列表215为全局内存索引节点列表。偏移元数据210包括磁盘上索引节点访问模式的位置信息,而全局索引节点列表215(也称为父目录列表)包括目录索引节点130(1)-(N)(例如,磁盘上索引节点的核心内父目录索引节点编号)以及顺序标记220(1)-(N)。顺序标记220(1)-(N)可作为父目录的核心内索引节点的部分存储(例如,存储器中由核心内父目录索引节点编号标识并且在图2A中显示为目录索引节点130(1)-(N)的父目录)。访问模式跟踪器225包括块访问跟踪器230(例如,以跟踪磁盘上索引节点的块访问)、以及目录访问跟踪器(例如,以跟踪目录中磁盘上索引节点的顺序和非顺序访问)。结合图2B至图2E更详细地描述了图2A的前述元素。

[0054] 跟踪索引节点的块访问模式的示例

[0055] 图2B是根据一个实施方案的偏移元数据的框图200B。偏移元数据210包括最后读

取的元数据块245的末端偏移以及当前元数据读取操作250的起始偏移。因为文件系统155持久地在磁盘150上存储核心内索引节点165(1) - (N)，当在核心内(例如，在高速缓存205内)未找到特定索引节点时，作为访问模式跟踪器225部分的块访问跟踪器230以块(例如，1KB、2KB、4KB、或其他合适的大小)从磁盘150读取结构化文件160中的磁盘上索引节点。这样，块访问跟踪器230可被配置为跟踪元数据块(例如，磁盘上索引节点)的访问模式，并有利于确定应用程序(例如，应用程序135或应用程序140)是否以顺序(或近似顺序)方式访问磁盘上索引节点。

[0056] 在一个实施方案中，块访问跟踪器230确定元数据存储区域中(例如，结构化文件160中)元数据单元的位置(例如，如图1B中所示的磁盘上索引节点165(1) - (4)的1KB块的位置/末端偏移)。块访问跟踪器230确定对应于当前元数据读取操作的元数据存储区域中(例如，结构化文件160中)的另一位置(例如，起始偏移)。元数据预读发生器240然后使用数据块的位置以及对应于当前元数据读取操作的另一位置确定是否需要元数据预读操作。如果需要元数据预读操作，则元数据预读发生器240发出元数据预读操作。

[0057] 块访问跟踪器230在与结构化文件160(1)相关联的核心内索引节点126(1)中保留来自磁盘150的最后读取的元数据块245的末端偏移(例如，结构化文件160(1)中逻辑偏移的末端)。例如，如果应用程序135和/或应用程序140尝试访问致使(并且需要)读取磁盘上索引节点的1KB块的磁盘150中的数据180，则块访问跟踪器230在结构化文件160(1)的核心内索引节点126(1)中将磁盘上索引节点的1KB块的末端偏移(例如，如图1B中所示的磁盘上索引节点165(1) - (4)的1KB块的末端偏移)存储为“存储的值”。

[0058] 在一些实施方案中，如果当前元数据读取操作250的起始偏移(例如，如图1B中所示的磁盘上索引节点165(5))邻近存储的值(例如，磁盘上索引节点165(1) - (4)的末端偏移至最后读取的元数据块245的末端偏移)，则元数据预读发生器240发出元数据预读指令以在结构化文件160中从1KB到2KB将磁盘上索引节点(例如，如图1B中所示的磁盘上索引节点165(5) - (8))提取到高速缓存205中。因为块访问跟踪器230确定磁盘上索引节点访问是以顺序方式发生，元数据预读发生器240发出元数据预读指令以从磁盘150将磁盘上索引节点的下个1KB块提取至核心内索引节点列表125(例如，1KB至2KB)，因此加快(未来)索引节点访问。

[0059] 如果触发了上述元数据预读指令(例如，如果当前元数据读取操作250的起始偏移紧靠最后读取的元数据块245的末端偏移)，则访问模式跟踪器225在核心内索引节点列表125中通过以下方式更新存储的值(例如，磁盘上索引节点165(1) - (4)的末端偏移至最后读取的元数据块245的末端偏移)：将高速缓存205中存储的值替换为由元数据预读操作读取的另一最后读取的元数据块的另一末端偏移(例如，如图1B中所示的磁盘上索引节点165(5) - (8)的末端偏移(例如，在2KB处)，因为元数据预读操作由于(发出的)元数据预读指令而从1KB至2KB预读磁盘上索引节点)。然而，如果当前元数据读取操作250的起始偏移未紧靠最后读取的元数据块245的末端偏移，并且因此没有触发元数据预读，则访问模式跟踪器225利用当前元数据读取操作的末端偏移重置存储的值(例如，磁盘上索引节点165(1) - (4)的末端偏移至最后读取的元数据块245的末端偏移)。

[0060] 应当理解，可跟踪索引节点的块访问模式，并且可发出元数据预读指令和/或元数据预读操作以从磁盘将可适用的索引节点预提取到存储器中，从而加快这些索引节点的后

续访问。接下来描述的是跟踪作为目录的部分的索引节点的索引节点访问模式的方法、系统和过程。

[0061] 使用父目录列表来跟踪索引节点的访问模式的示例

[0062] 应当理解,索引节点分配策略可使磁盘上索引节点在相同目录中保持彼此靠近,在本文中称为靠近位址。例如,经常访问的文件的磁盘上索引节点可一起保留在相同目录中。目录中磁盘上索引节点的该靠近位址可用于跟踪索引节点的目录访问。例如,可创建并保留全局索引节点列表215(或父目录列表)以跟踪磁盘上索引节点的访问。

[0063] 图2C是根据某些实施方案的表200C,示出了此类全局索引节点列表的内容,并且图2D是根据某些实施方案的使用全局索引节点列表的目录访问跟踪器235的框图200D。全局索引节点列表215(其为目录265(1)-(N)的父目录列表)包括目录索引节点字段255以及顺序标记字段260。全局索引节点列表215在存储器中创建和保留,并且包括目录索引节点130(1)-(N)和顺序标记220(1)-(N)。顺序标记可作为父目录的核心内索引节点的部分存储。目录访问跟踪器235包括顺序目录访问跟踪器270和非顺序目录访问跟踪器275。

[0064] 在一个实施方案中,应用程序访问索引节点(例如,磁盘上索引节点165(4))。目录访问跟踪器235确定索引节点的父目录(例如,目录265(1)),以及目录的条目是否存在于全局索引节点列表215中。如果目录的条目存在于全局索引节点列表215中,则目录访问跟踪器235确定目录的文件结构是顺序还是非顺序(例如,通过确定该目录中的磁盘上索引节点以顺序方式还是非顺序方式被列出,以及因此以顺序方式还是非顺序方式被访问)。如果目录的条目未存在于全局索引节点列表215中,则目录访问跟踪器235在全局索引节点列表215中为父目录索引节点添加新条目。应当指出的是,如图1A、图2A、图2C和图3C中所示,目录索引节点(例如,目录索引节点130(1)-(N))仅为表示父目录的磁盘上索引节点的特定于目录的表示(例如,替代各个文件)。例如,目录索引节点130(1)是表示一个或多个磁盘上索引节点的父目录的内存数据结构。

[0065] 为了跟踪磁盘上索引节点的顺序目录访问,顺序目录访问跟踪器270首先查找给定磁盘上索引节点上的父目录(例如,查找父目录的索引节点编号)。例如,如果从磁盘150读取磁盘上索引节点165(4),则顺序目录访问跟踪器270查找磁盘上索引节点165(4)的父目录(例如,目录索引节点130(1))。顺序目录访问跟踪器270然后在全局索引节点列表215中搜索给定磁盘上索引节点的父目录(索引节点)的条目(例如,父目录索引节点编号是否存在于高速缓存205中)。

[0066] 如果父目录(索引节点)的条目存在于全局索引节点列表215中,则顺序目录访问跟踪器270检查全局索引节点列表215以确定父目录是否设置了顺序标记(例如,目录索引节点130(1),其在该情况下为父目录索引节点编号,并且如图2C中所示设置了顺序标记)。如果设置了顺序标记,则元数据预读发生器240发出元数据预读指令(例如,以提取目录265(1)中的所有剩余磁盘上索引节点,因为磁盘上索引节点访问为顺序的)。如果顺序目录访问跟踪器270未在全局索引节点列表215中找到父目录的条目,则顺序目录访问跟踪器270将父目录的索引节点编号的新条目添加至全局索引节点列表215。

[0067] 为了跟踪磁盘上索引节点的非顺序目录访问,非顺序目录访问跟踪器275首先查找给定磁盘上索引节点上的父目录(例如,父目录的索引节点编号)。非顺序目录访问跟踪器275然后在全局索引节点列表215中搜索父目录的现有条目(例如,父目录索引节点编号

是否存在于高速缓存205中)。如果父目录的条目存在于全局索引节点列表215中,则非顺序目录访问跟踪器275从磁盘150将父目录的索引节点列表(例如,适用于所考虑的父目录的索引节点列表175的一部分或部分)提取(或检索)到高速缓存205中,并且元数据预读发生器240为在检索的索引节点列表上列出的磁盘上索引节点(例如,与父目录相关联并且作为其部分的磁盘上索引节点)发出元数据预读指令。如果父目录的条目未存在于全局索引节点列表215中,则非顺序目录访问跟踪器275向全局索引节点列表215添加父目录索引节点编号的新条目。

[0068] 为索引节点发出元数据预读指令的示例

[0069] 图2E是根据一个实施方案的元数据预读发生器的框图200E。元数据预读发生器240通过计算设备105实现并且存储发出的元数据预读值280和异步元数据预读指令285,并且包括队列发生器290。尽管目录访问跟踪器235标识一个或多个磁盘上索引节点以预提取,如果没有预提取这些磁盘上索引节点,则无法完成与这些磁盘上索引节点相关联的I/O操作。

[0070] 因此,在一个实施方案中,元数据预读发生器240响应于I/O操作(例如,读取操作或写入操作)拦截用于读取磁盘上索引节点的命令。访问数据的I/O操作可导致(或致使)访问和读取与该数据关联的磁盘上索引节点(例如,元数据)的命令(例如,以确定何时以及如何修改了请求的数据等)。元数据预读发生器240通过将发出的元数据预读值280和命令中的块总计进行比较,分析元数据预读操作中发出的元数据预读值280。

[0071] 发出的元数据预读值280包括应当预读(例如,基于顺序/近似顺序块访问、和/或目录中磁盘上索引节点顺序或非顺序访问来检测)的所有磁盘上索引节点。块总计为要预读的磁盘上索引节点的块的总计数目(例如,在本文出于讨论目的表示为整数“N”)。例如,块访问跟踪器230和顺序目录访问跟踪器270可标识并确定必须预读磁盘上索引节点的1KB块(例如,磁盘上索引节点165(1)-(4))或磁盘上索引节点的2KB块(例如,磁盘上索引节点165(1)-(8)) (例如,基于顺序/近似顺序块访问、和/或目录中磁盘上索引节点顺序或非顺序访问)。然而,如上文所讨论的,非顺序目录访问跟踪器275可标识要预读的数个非顺序磁盘上索引节点(例如,其可为各个不同块的部分)(例如,图3B中在目录265(2)的情况中所示)。因此,应当理解,在某些情况下,发出的元数据预读值280可以或可以不等于N。

[0072] 在一些实施方案中,基于将发出的元数据预读值280和命令中的块总计进行比较,元数据预读发生器240等待I/O操作完成,或发出异步元数据预读指令285。如果I/O操作完成,则队列发生器290生成队列并且包括未包括在异步元数据预读指令285中的块总计的剩余元数据块。然而,如果I/O操作未完成,则队列发生器290更新元数据预读操作中的块总计。

[0073] 例如,前往磁盘150的读取磁盘上索引节点的命令(或调用)在发出I/O操作(例如,由应用程序130发出)之后被拦截。该命令触发上文所述索引节点访问模式检测方法(例如,顺序/近似顺序块访问、和/或目录中磁盘上索引节点顺序或非顺序访问)。如果索引节点访问模式检测方法没有触发磁盘上索引节点的预读,则元数据预读发生器240仅等待I/O操作完成。然而,如果索引节点访问模式检测方法触发了磁盘上索引节点的预读,则元数据预读发生器240确定总计发出的元数据预读(例如,发出的元数据预读值280)是否小于或等于N(例如,要预读的磁盘上索引节点的块的总计数目)。

[0074] 如果总计发出的元数据预读小于或等于N,则元数据预读发生器240利用异步元数据读取的下个块(例如,N后发出的元数据预读值280中的下个块)发出异步元数据预读指令285。如果初始I/O操作完成,则队列发生器290生成单独的线程,该单独的线程利用异步元数据读取的剩余块(例如,N后发出的元数据预读值280中的剩余块)发出异步元数据预读指令285。如果初始I/O操作未完成,元数据预读发生器240针对发出的元数据预读值280递增计数器,确定发出的元数据预读值280是否等于N,并等待初始I/O操作完成。

[0075] 应当理解,元数据预读检测和异步元数据预读指令的发出是在阻止线程的上下文中执行的,而初始I/O操作正在等待在后台完成。这些方法减少对系统性能的索引节点访问模式检测开销,并且还避免单独线程的创建和计划,这可能推迟后续读取的块的可用性。

[0076] 跟踪索引节点的顺序和非顺序目录访问模式的示例

[0077] 图3A是根据一些实施方案的具有顺序索引节点结构300A的目录的框图,图3B是根据一些实施方案的具有非顺序索引节点结构300B的目录的框图,并且图3C是根据一些实施方案的实现全局索引节点列表(例如,父目录列表或全局内存索引节点列表)的高速缓存300C的框图。应当理解,目录的列出首先由应用程序执行(例如,由应用程序135、应用程序140、或一些其他应用程序执行)。在目录的列出期间,返回与每个目录条目相关联的磁盘上索引节点编号。如果给定目录中磁盘上索引节点编号是顺序的(例如,如图3A中所示的目录265(1)),则顺序目录访问跟踪器270在该目录的核心内索引节点中为该特定目录设置顺序标记(例如,目录265(1)的顺序标记如图2C和图3C中所示是在全局索引节点列表215中设置的)。

[0078] 例如,由于目录265(1)的磁盘上索引节点是顺序的(例如,磁盘上索引节点165(4) - (9)被顺序地列出),顺序目录访问跟踪器270在目录索引节点130(1)中(例如,在核心内索引节点中)为目录265(1)设置顺序标记(例如,在图2C和图3C中由全局索引节点列表215的顺序标记字段260中的“1”指示)。相反,因为目录265(2)的磁盘上索引节点为非顺序的(例如,磁盘上索引节点165(4)、165(9)、165(15)、165(11)、165(19)、以及165(6)被非顺序地列出),非顺序目录访问跟踪器275不会在目录索引节点130(2)中(例如,在核心内索引节点中)为目录265(2)设置顺序标记(例如,在图2C和图3C中由全局索引节点列表215的顺序标记字段260中的“0”指示)。

[0079] 如前文所述,执行元数据预读操作的元数据预读指令可在执行目录的列出之后发出。在一个实施方案中,应用程序执行作为磁盘150的部分的目录的列出。例如,如图3B中所示的目录265(2)的列出返回带磁盘上索引节点165(4)的文件315(1)、带磁盘上索引节点165(9)的文件315(2)、带磁盘上索引节点165(15)的文件315(3)、带磁盘上索引节点165(11)的文件315(4)、带磁盘上索引节点165(19)的文件315(5)、以及带磁盘上索引节点165(6)的文件315(6)。在该情况下,目录访问跟踪器235创建并保留内存数据结构(例如,内存索引节点列表),该内存数据结构包括前述索引节点编号(例如,磁盘上索引节点165(4)、165(9)、165(15)、165(11)、165(19)、以及165(6))的列表,并将该内存索引节点列表与目录265(2)的核心内索引节点(例如,目录索引节点130(2))相关联。

[0080] 例如,如果应用程序130访问文件315(1),并且因此需要访问和读取磁盘上索引节点165(4),则非顺序目录访问跟踪器275从磁盘150读取磁盘上索引节点165(4),并且确定磁盘上索引节点165(4)的父目录索引节点编号为目录索引节点130(2)。非顺序目录访问跟

踪器275然后检查目录索引节点130(2)是否存在于高速缓存205中,如图3C中所示。如果目录索引节点130(2)未存在于高速缓存205中,则非顺序目录访问跟踪器275向高速缓存205添加目录索引节点130(2)(例如,由图2C和3C中的粗体指示)。

[0081] 接下来,如果应用程序130访问文件315(2),并且因此从磁盘150访问和读取磁盘上索引节点165(9),则非顺序目录访问跟踪器275确定磁盘上索引节点165(9)的父目录索引节点编号也为目录索引节点130(2)。因为目录索引节点130(2)已经被添加至高速缓存205,所以元数据预读发生器240确定可预读目录265(2)下的剩余文件(例如,文件315(3)-(6))。元数据预读发生器240然后访问内存索引节点列表并标识与文件315(3)-(6)相关联的剩余磁盘上索引节点编号(例如,磁盘上索引节点165(15)、165(11)、165(19)、以及165(6)),并生成元数据预读指令,该指令执行元数据预读操作以从磁盘150将磁盘上索引节点165(15)、165(11)、165(19)、以及165(6)提取到存储器115。

[0082] 应当理解,目录访问跟踪器235跟踪磁盘上索引节点的块访问模式、以及作为目录的部分的磁盘上索引节点的顺序访问模式和非顺序访问模式,以标识作为加快索引节点和数据访问的元数据预读操作的候选项的磁盘上索引节点。

[0083] 跟踪索引节点访问模式并发出元数据预读指令的过程

[0084] 图4A是根据一个实施方案的示出用于发出预读指令以从磁盘将磁盘上索引节点预提取至存储器的过程的流程图400A。该过程通过访问元数据存储区域(例如,结构化文件160)开始于405。在410处,该过程确定最后读取的元数据块的位置(例如,最后读取的元数据块245的末端偏移)。在415处,该过程确定是否已收到命令(或调用)(例如,读取磁盘上索引节点)。如果尚未收到命令,则该过程循环回415。然而,如果收到命令,则在420处,该过程确定元数据存储区域中命令的对象的位置(例如,当前元数据读取操作250的起始偏移)。

[0085] 在425处,该过程确定是否可执行元数据预读(或者是否需要元数据预读或元数据预读是否可行)。如果无法执行元数据预读,则在430处,该过程允许正常处理(例如,不执行元数据预读操作并且不从磁盘将磁盘上索引节点预提取至存储器中)。然而,如果可执行元数据预读(并且/或者需要元数据预读和/或元数据预读可行),则在435处,该过程发出元数据预读操作(或例如使用元数据预读发生器240发出元数据预读指令,该指令致使计算设备105执行元数据预读操作以从磁盘将磁盘上索引节点的块提取到存储器中)。在440处,该过程确定是否存在新命令(例如,以访问和/或读取磁盘上索引节点)。如果存在访问和/或读取磁盘上索引节点的新命令,则过程循环回405。否则,过程结束。

[0086] 图4B是根据一个实施方案的示出了用于存储与磁盘上索引节点关联的偏移元数据的过程的流程图400B。该过程通过确定元数据存储区域(例如,结构化文件160)中元数据块的偏移位置(例如,最后读取的元数据块245的末端偏移)开始于445。在450处,该过程在存储器(例如,存储器115)中创建核心内索引节点(例如,核心内索引节点列表125),并且在455处,该过程在核心内索引节点中存储元数据块的偏移位置。

[0087] 在455处,该过程确定当前元数据读取操作的位置(例如,当前元数据读取操作250的起始偏移)是否紧靠(或邻近)元数据块的偏移位置。如果当前元数据读取操作的位置未紧靠元数据块的偏移位置,则在465处,该过程存储由当前元数据读取操作读取的元数据块的偏移位置(例如,当前元数据读取操作的末端偏移)。然而,如果当前元数据读取操作的位置紧靠元数据块的偏移位置,则在470处,该过程发出元数据预读操作(或发出元数据预读

指令),并且在475处,存储由元数据预读操作读取的元数据块的偏移位置。在480处,该过程确定是否存在新的读取调用(例如,应用程序I/O操作引起的读取磁盘上索引节点的命令)。如果存在读取磁盘上索引节点的新读取调用,则该过程循环回460。否则,过程结束。

[0088] 应当理解,图4A的流程图400A以及图4B的流程图400B中示出的过程是跟踪磁盘上索引节点的块访问模式的示例。因为磁盘上索引节点持久性存储在磁盘(例如,磁盘150)上,所以块访问跟踪器230可访问结构化文件160,以确定最后读取的元数据块的末端偏移以及当前元数据读取操作的起始偏移。因为块访问跟踪器230可保存该位置信息,所以块访问跟踪器230可确定磁盘上索引节点由应用程序135还是应用程序140以顺序方式还是近似顺序的方式访问。基于该存储的位置信息,可能访问的磁盘上索引节点可被预读并从磁盘预提取到存储器中,因此加快那些磁盘上索引节点的后续索引节点访问。

[0089] 图5A是根据一个实施方案的示出了用于确定目录的文件结构的过程的流程图500A。该过程通过访问目录中的文件(例如,如图3A中所示的目录265(1)中的文件135(1)或如图3B中所示的目录265(2)中的文件135(1))开始于505。在510处,该过程在全局索引节点列表中搜索目录的条目(例如,索引节点编号)。在515处,该过程确定目录是否存在于全局索引节点列表中(例如,作为目录265(1)的索引节点编号的目录索引节点130(1)是否存在并且列在全局索引节点列表215中)。

[0090] 如果目录未存在于全局索引节点列表中,则在520处,该过程在全局索引节点列表中为磁盘上索引节点添加新条目(例如,添加父目录索引节点编号)(例如,图2C和图3C中相对于目录265(3)所示)。如果目录存在于全局索引节点列表中,则在525处,该过程确定目录的文件结构(例如,目录中的磁盘上索引节点是顺序地列出还是非顺序地列出)。在530处,该过程确定是否存在磁盘上索引节点的另一访问。如果存在磁盘上索引节点的另一访问,则该过程循环回505。否则,过程结束。

[0091] 图5B是根据一个实施方案的示出了用于为磁盘上索引节点发出元数据预读指令的过程的流程图500B。该过程通过确定给定目录中条目的磁盘上索引节点编号为顺序的还是非顺序的(例如,由于执行目录的列出而顺序地列出还是非顺序地列出)来开始于535。例如,在图3A中,条目的磁盘上索引节点编号为顺序地列出,并且在图3B中,条目的磁盘上索引节点编号为非顺序地列出。

[0092] 如果条目的磁盘上索引节点编号为非顺序地列出,则在540处,该过程访问全局索引节点列表(例如,如图3C中所示的父目录列表)。在545处,该过程标识全局索引节点列表上的父目录索引节点(例如,使用磁盘上索引节点的父目录索引节点编号)。在550处,该过程提取与所考虑的目录相关联的索引节点列表(例如,从磁盘150或从存储器提取,因为在列出过程期间索引节点列表可存储和保留在存储器中),并且在555处,该过程针对索引节点列表上磁盘上索引节点发出元数据预读指令。

[0093] 然而,如果条目的磁盘上索引节点编号为顺序地列出,则在560处,该过程访问全局索引节点列表,并且在565处,标识在全局索引节点列表上访问的文件的父目录的内存索引节点(例如,使用磁盘上索引节点的父目录索引节点编号)。在570处,该过程确认目录设置了顺序标记,并且在575处,为所考虑的目录中列出的剩余磁盘上索引节点发出元数据预读指令。在580处,该过程确定是否存在磁盘上索引节点的另一访问。如果存在磁盘上索引节点的另一访问,则该过程循环回535。否则,过程结束。

[0094] 应当理解,图5A的流程图500A中以及图5B的流程图500B示出的过程是在涉及目录时,使用父目录列表来跟踪磁盘上索引节点的访问模式的示例。如前文所述,应用程序可执行一个或多个目录的列出,以确定给定目录中磁盘上索引节点是顺序地列出还是非顺序地列出。因为目录通常存储在磁盘(例如,磁盘150)上,所以每次在给定目录中访问文件时确定目录的数据和/或文件结构可消耗显著的计算资源。然而,因为访问模式跟踪器225和父目录列表是存储器115的部分,并且因为父目录列表保留多个目录的数据和/或文件结构的列出,所以访问模式跟踪器225可仅使用父目录列表来确定给定目录的核心内索引节点是顺序还是非顺序的,而不用每次访问给定目录中的文件时访问磁盘150。

[0095] 图6A是根据一个实施方案的示出了用于处理与预提取索引节点相关的输入/输出(I/O)操作的过程的流程图600A。该过程通过检测为与磁盘上索引节点相关联的元数据块(例如,结构化文件160中的元数据单元)发出的I/O操作开始于605。在610处,该过程确定是否检测到I/O操作。如果未检测到I/O操作,则该过程循环回605。但是如果检测到I/O操作,则在615处,该过程拦截用于访问和/或读取磁盘上索引节点的命令(或调用)。

[0096] 在620处,该过程访问元数据预读操作的元数据预读值,并且在625处,通过将(总计发出的)元数据预读值与要预读的元数据块的总计数目进行比较,来分析元数据预读值。在630处,该过程等待I/O操作完成,并且在635处,发出异步元数据预读指令(例如,使用元数据预读发生器240)。在640处,该过程确定是否存在另一I/O操作。如果存在另一I/O操作,则该过程循环至605。否则,过程结束。

[0097] 图6B是根据一个实施方案的示出了用于处理与预提取索引节点相关的处理I/O操作的过程的流程图。该过程通过确定给定I/O操作是否完成而开始于645。如果I/O操作未完成,则在650处,该过程更新块总计(例如,要预读的元数据块的总计数目或“N”),并前进到625(在图6A中)。然而,如果I/O操作完成,则在660处,该过程生成队列(例如,单独线程),并且在665处,发出元数据预读,该预读包括未包括在异步元数据预读指令中的剩余元数据块(例如,图6A)。在670处,该过程确定是否存在要拦截的另一命令(例如,对读取磁盘上索引节点的调用)。如果存在要拦截的另一命令,则该过程循环回615(在图6A中)。否则,过程结束。

[0098] 应当理解,图6A的流程图600A以及图6B的流程图600B中示出的过程可用于通过将发出的元数据预读值和块总计进行比较,来标识和预提取作为各个不同元数据块的部分的磁盘上索引节点。因为目录列出过程期间索引节点列表的全部或一部分可预提取到存储器中,所以作为元数据预读指令的候选项的磁盘上索引节点可在此类磁盘上未被顺序地访问的情况下被标识。另外,还应当理解,元数据预读检测和异步元数据预读指令的发出是在阻止线程的上下文中执行的,而初始I/O操作正在等待在后台完成。这些方法减少对系统性能的索引节点访问模式检测开销,并且还避免单独线程的创建和计划,这可能推迟后续读取的块的可用性。

[0099] 图7A是根据某些实施方案的示出用于处理磁盘上索引节点的访问以及向全局索引节点列表添加目录的条目的过程的流程图。该过程通过确定是否存在磁盘上索引节点的访问(例如,一个或多个I/O操作引起的磁盘上索引节点的读取命令/调用)开始于705。如果不存在磁盘上索引节点的访问,则该过程循环回705。然而,如果存在磁盘上索引节点的访问,则在710处,该过程标识与磁盘上索引节点相关联的目录(例如,使用父目录索引节点编

号)。在715处,该过程访问全局索引节点列表(例如,全局索引节点列表215和/或父目录列表)。

[0100] 在720处,该过程确定目录是否在全局索引节点列表上(例如,索引节点的父目录索引节点编号是否存在于全局索引节点列表上)。如果目录未存在于全局索引节点列表上,则在725处,该过程向全局索引节点列表添加目录(例如,通过向全局索引节点列表添加索引节点的父目录索引节点编号的条目)。然而,如果目录存在于全局索引节点列表上,则在730处,该过程确认为目录设置了顺序标记,并且在735处,发出元数据预读指令(例如,针对目录中列出的(或作为其部分的)剩余磁盘上索引节点)。在740处,该过程确定是否存在(例如,磁盘上索引节点的)另一访问。如果存在另一访问,则该过程循环回705。否则,过程结束。

[0101] 图7B是根据一个实施方案的示出了用于处理对访问磁盘上索引节点的请求的过程的流程图。该过程通过检测I/O操作开始于745。在750处,该过程拦截用于访问(或读取)磁盘上索引节点(或调用)的命令(或调用)。在755处,该过程确定磁盘上索引节点访问模式(例如,除了别的方法以外,基于块访问、以及顺序或非顺序访问来检测)是否触发、致使或导致磁盘上索引节点的元数据预读。如果磁盘上索引节点访问模式没有触发元数据预读(例如,基于块访问、顺序访问、或非顺序访问),则在760处,该过程等待I/O操作完成。应当指出的是,在图6B的流程图600B中示出了用于等待I/O操作完成的示例性过程。

[0102] 然而,如果磁盘上索引节点访问模式没有触发元数据预读,则在765处,该过程确定块总计(例如,要预读的磁盘上索引节点元数据块的总计数目或“N”)是否小于或等于元数据预读值(例如,发出的元数据预读值280或总计发出的元数据预读值)。如果块总计不小于或等于元数据预读值,则在760处,该过程等待I/O操作完成。然而,如果块总计小于或等于元数据预读值,则在770处,该过程发出异步元数据预读的下个块,并且在775处,确定I/O操作是否完成。如果I/O操作未完成,则在780处,该过程递增计数器(例如,针对发出的元数据预读),并循环回765。然而,如果I/O操作完成,则在785处,该过程生成单独的线程并且发出异步元数据预读的剩余块。在790处,该过程确定是否存在另一I/O操作。如果存在另一I/O操作,则该过程循环回745。否则,过程结束。

[0103] 通常,如果I/O操作被阻止,则线程无法进一步进行。应当理解,因为索引节点访问模式检测和元数据预读生成的方法、系统和过程是在阻止线程的上下文中(并且在初始I/O在后台等待完成时)执行的,所以对系统性能的索引节点访问模式检测开销减少。此外,允许初始I/O操作完成也可避免创建和计划单独线程的成本(这会推迟后续读取操作的数据块的可用性)。

[0104] 另外,应当指出的是,可将其他数据结构诸如属性与节点相关联。与本文所述的索引节点访问模式检测以及发出元数据预读指令相关的方法、系统和过程可预先填充存储器中的这些(其他)数据结构并且可异步初始化各个索引节点锁。因此,应当理解,除了其他功能之外,本文所述的方法、系统和过程能够基于块访问、顺序访问、和非顺序访问来跟踪索引节点的访问模式,并为索引节点发出预读指令。

[0105] 示例性计算环境

[0106] 图8是根据一个实施方案的计算系统800的框图,示出了如何可在软件中实现访问模式跟踪器和元数据预读发生器。计算系统800在广义上表示能够执行计算机可读指令的

任何单处理器或多处理器计算设备或系统。计算系统800的示例包括但不限于以下各种设备中的任何一者或多者：工作站、个人计算机、膝上型电脑、客户端侧终端、服务器、分布式计算系统、手持设备（例如，个人数字助理和移动电话）、网络设备、存储控制器（例如，阵列控制器、磁带驱动器控制器或硬盘驱动器控制器）等。在其最基本的配置下，计算系统800可包括至少一个处理器110和存储器115。通过执行实现计算设备105的软件，计算系统800成为专用计算设备，该计算设备被配置为跟踪索引节点访问模式并为索引节点发出预读指令。

[0107] 处理器110通常表示能够处理数据或解译和执行指令的任何类型或形式的处理单元。在某些实施方案中，处理器110可接收来自软件应用程序或模块的指令。这些指令可使处理器110执行本文描述和/或示出的实施方案中的一个或多个实施方案的功能。例如，处理器110可执行本文所述的操作中的全部或一些和/或可以是用于执行本文所述的操作中的全部或一些操作的装置。处理器110还可执行本文描述和/或示出的任何其他操作、方法或过程和/或可以是用于执行本文描述和/或示出的任何其他操作、方法或过程的装置。

[0108] 存储器115通常表示能够存储数据和/或其他计算机可读指令的任何类型或形式的易失性或非易失性存储设备或介质。示例包括但不限于随机存取存储器（RAM）、只读存储器（ROM）、闪存存储器或任何其他合适的存储器设备。尽管并非必需，但是在某些实施方案中，计算系统800可包括易失性存储器单元和非易失性存储设备两者。在一个示例中，可将实现访问模式跟踪器和元数据预读发生器的程序指令载入存储器115。

[0109] 在某些实施方案中，除处理器110和/或存储器115之外，计算系统800还可包括一个或多个部件或元件。例如，如图8所示，计算系统800可包括存储器控制器820、输入/输出（I/O）控制器835和通信接口845，它们中的每一者都可经由通信基础设施805互连。通信基础设施805通常表示能够促进计算设备的一个或多个部件之间的通信的任何类型或形式的基础设施。通信基础设施805的示例包括但不限于通信总线（诸如工业标准架构（ISA）、外围部件互连（PCI）、PCI Express（PCIe）或类似的总线）和网络。

[0110] 存储器控制器820通常表示能够处理存储器或数据或者控制计算系统800的一个或多个部件之间的通信的任何类型/形式的设备。在某些实施方案中，存储器控制器820可经由通信基础设施805控制处理器110、存储器115和I/O控制器835之间的通信。在某些实施方案中，存储器控制器820可单独或结合其他元件来执行本文描述和/或示出的操作或特征中的一个或多个和/或可以是用于单独或与其他元件组合来执行本文描述和/或示出的操作或特征中的一者或多者的装置。

[0111] I/O控制器835通常表示能够协调和/或控制一个或多个计算设备诸如计算设备105的输入功能和输出功能的任何类型或形式的模块。例如，在某些实施方案中，I/O控制器835可控制或促进计算系统800的一个或多个元件之间的数据传送，所述元件诸如处理器110、存储器115、通信接口845、显示适配器815、输入接口825和存储接口840。

[0112] 通信接口845在广义上表示能够促进计算系统800与一个或多个其他设备之间的通信的任何类型或形式的通信设备或适配器。通信接口845可促进计算系统800与包括附加计算系统的私有或公共网络之间的通信。通信接口845的示例包括但不限于有线网络接口（诸如网络接口卡）、无线网络接口（诸如无线网络接口卡）、调制解调器以及任何其他合适的接口。通信接口845可经由与网络诸如互联网的直接链路提供到远程服务器的直接连接，

并且还可通过例如局域网例如以太网网络、个人区域网、电话或电缆网络、蜂窝电话连接、卫星数据连接或任何其他合适的连接来间接提供此类连接。

[0113] 通信接口845还可表示主机适配器,该主机适配器被配置为经由外部总线或通信信道来促进计算系统800与一个或多个附加网络或存储设备之间的通信。主机适配器的示例包括:小型计算机系统接口(SCSI)主机适配器、通用串行总线(USB)主机适配器、电气与电子工程师协会(IEEE) 1394主机适配器、串行高级技术附件(SATA)、串行SCSI(SAS)和外部SATA(eSATA)主机适配器、高级技术附件(ATA)和并行ATA(PATA)主机适配器、光纤信道接口适配器、以太网适配器等。通信接口845还可允许计算系统800执行分布式或远程计算(例如,通过从远程设备接收指令/将指令发送到远程设备,以供执行)。

[0114] 如图8所示,计算系统800还可包括至少一个显示设备810,该至少一个显示设备经由显示适配器815耦接到通信基础设施805。显示设备810通常表示能够可视地显示由显示适配器815转发的信息的任何类型或形式的设备。类似地,显示适配器815通常表示被配置为转发来自通信基础设施805(或来自帧缓冲器,如本领域所已知)的图形、文本和其他数据以在显示设备810上显示的任何类型或形式的设备。计算系统800还可包括经由输入接口825耦接到通信基础设施805的至少一个输入设备830。输入设备830通常表示能够向计算系统800提供由计算机或人生成的输入的任何类型或形式的输入设备。输入设备830的示例包括键盘、指向设备、语音识别设备或任何其他输入设备。

[0115] 计算系统800还可包括存储设备850(例如,磁盘150),该存储设备经由存储接口840耦接到通信基础设施805。存储设备850通常表示能够存储数据和/或其他计算机可读指令的任何类型或形式的存储设备或介质。例如,存储设备850可包括磁盘驱动器(例如,所谓的硬盘驱动器)、软盘驱动器、磁带驱动器、光盘驱动器、闪存驱动器等。存储接口840通常表示用于在存储设备850与计算系统800的其他部件之间传送和/或传输数据的任何类型或形式的接口或设备。存储设备850可被配置为对被配置为存储计算机软件、数据或其他计算机可读信息的可移除存储单元执行读取和/或写入。合适的可移除存储单元的示例包括软盘、磁带、光盘、闪存存储设备等。存储设备850还可包括用于允许将计算机软件、数据或其他计算机可读指令加载到计算系统800中的其他类似结构或设备。例如,存储设备850可被配置为读取和写入软件、数据或其他计算机可读信息。存储设备850还可以是计算系统800的一部分,或者可以是通过其他接口系统访问的独立设备。

[0116] 可以将许多其他设备或子系统连接到计算系统800。相反地,图8所示的部件和设备不必都存在以实践本文所述和/或示出的实施方案。上文提及的设备和子系统也可以按照与图8所示不同的方式互连。计算系统800也可采用任何数量的软件配置、固件配置和/或硬件配置。例如,本文公开的实施方案中的一个或多个实施方案可被编码为计算机可读存储介质上的计算机程序(也被称为计算机软件、软件应用程序、计算机可读指令或计算机控制逻辑)。计算机可读存储介质的示例包括磁存储介质(例如,硬盘驱动器和软盘)、光学存储介质(例如,CD-ROM或DVD-ROM)、电子存储介质(例如,固态驱动器和闪存介质)等。此类计算机程序也可被传输至计算系统800,以经由网络诸如互联网存储在存储器中或存储在载体介质上。

[0117] 可以将包含计算机程序的计算机可读介质加载到计算系统800中。然后,可将存储在计算机可读介质上的计算机程序的全部或一部分存储在存储器860和/或存储设备850的

各个部分中。当由处理器110执行时,加载到计算系统800中的计算机程序可使处理器110执行本文描述和/或示出的实施方案中的一个或多个实施方案的功能和/或可使处理器成为用于执行本文描述和/或示出的实施方案中的一个或多个实施方案的功能的装置。除此之外或另选地,可在固件和/或硬件中实现本文描述和/或示出的实施方案中的一个或多个实施方案。例如,计算系统800可被配置为适于实现本文公开的实施方案中的一个或多个实施方案的专用集成电路(ASIC)。

[0118] 示例性联网环境

[0119] 图9是根据一个实施方案的联网系统900的框图,示出了各种设备可如何经由网络进行通信。在某些实施方案中,除了其他的之外,附网存储(NAS)设备可被配置为使用各种协议诸如网络文件系统(NFS)、服务器消息块(SMB)或通用互联网文件系统(CIFS)等来与计算设备105和存储系统145进行通信。

[0120] 网络185通常表示能够促进计算设备105和存储系统145之间的通信的任何类型或形式的计算机网络或体系结构。在某些实施方案中,通信接口,诸如图8中的通信接口845,可用于提供计算设备105、存储系统145和网络155之间的连接。应当注意,本文描述和/或示出的实施方案并非限于互联网或任何特定的基于网络的环境。例如,网络185可以是存储区域网络(SAN)。计算设备105和存储系统145可集成或分离。如果是分离的,例如,计算设备105和存储系统145可通过本地连接(例如,使用Bluetooth™、外围部件互连(PCI)、小型计算机系统接口(SCSI)等)耦接,或经由一个或多个网络诸如互联网、LAN或SAN耦接。

[0121] 在一个实施方案中,所公开的实施方案中的一个或多个的全部或一部分可编码成计算机程序并加载到计算设备105、索引节点访问模式跟踪和元数据预读指令发出系统910、索引节点访问模式跟踪系统940、和/或元数据预读指令生成系统950上并由它们执行。本文所公开的实施方案中的一个或多个的全部或一部分还可以被编码为计算机程序,存储在计算设备105、索引节点访问模式跟踪和元数据预读指令发出系统910、和/或索引节点访问模式跟踪系统940上,并且通过网络185分配。

[0122] 在一些示例中,计算设备105的全部或一部分可表示基于云计算或网络的环境的部分。云计算环境可经由互联网提供各种服务和应用程序。这些基于云的服务(例如,软件即服务、平台即服务、基础设施即服务等)可通过网页浏览器或其他远程接口访问。本文所述的各种功能可通过远程桌面环境或任何其他基于云的计算环境提供。

[0123] 此外,本文所述的部件中的一个或多个部件可将数据、物理设备和/或物理设备的表示从一种形式转换为另一种形式。例如,计算设备105和/或索引节点访问模式跟踪和元数据预读指令发出系统910可转变计算设备105的行为以便致使计算设备105和/或索引节点访问模式跟踪和元数据预读指令发出系统910跟踪索引节点的访问模式并发出预读指令。

[0124] 尽管已结合若干实施方案描述了本公开,但是本公开并非旨在限于本文阐述的具体形式。相反地,本公开旨在覆盖可被合理地包括在由所附权利要求书限定的本公开的范围内的此类替代物、修改形式和等同物。

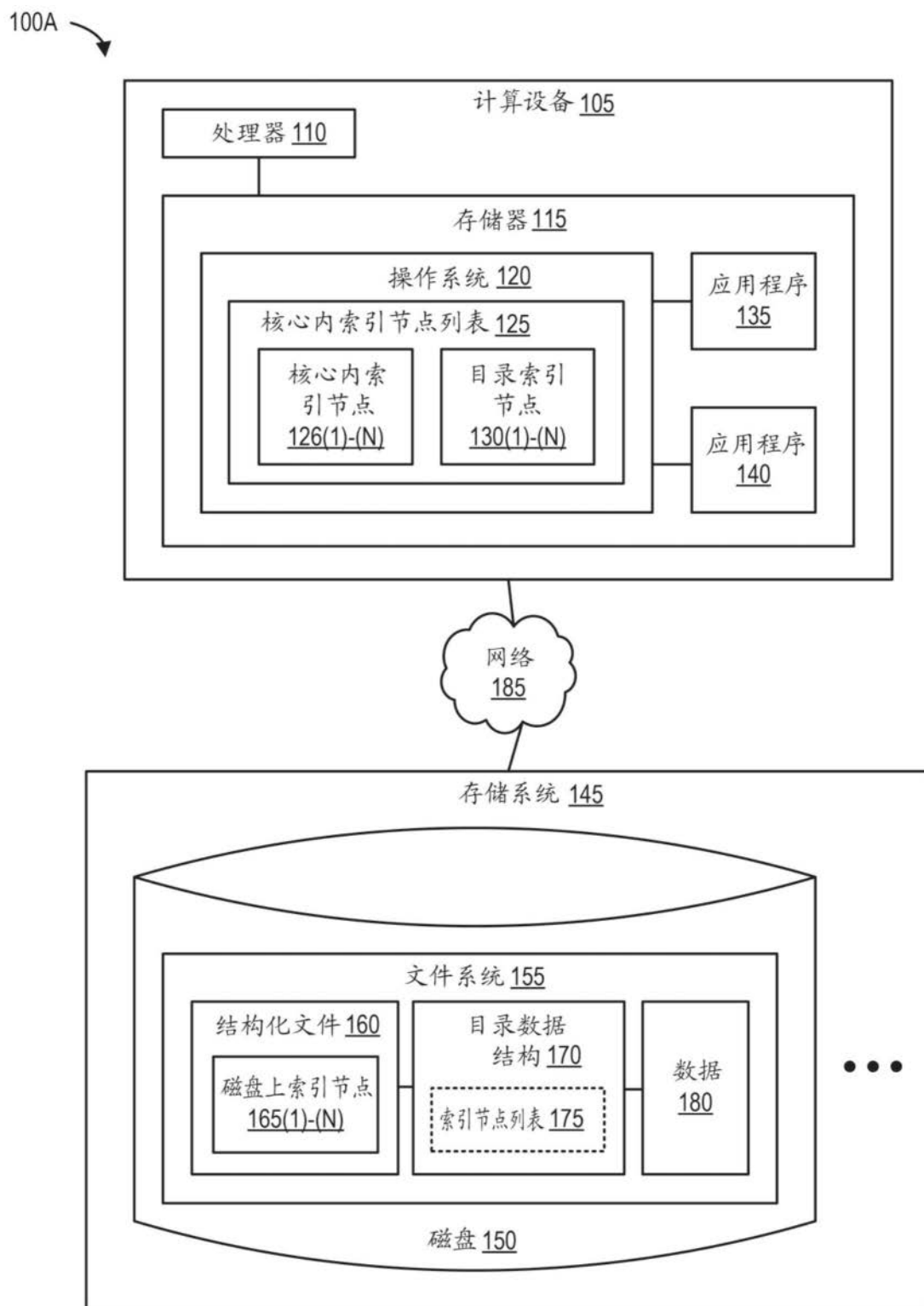


图1A

100B

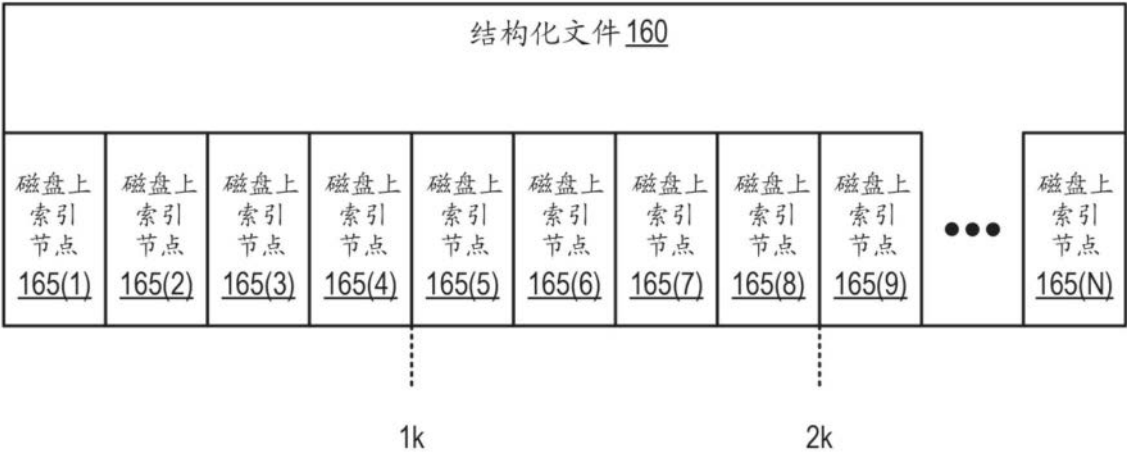


图1B

100C

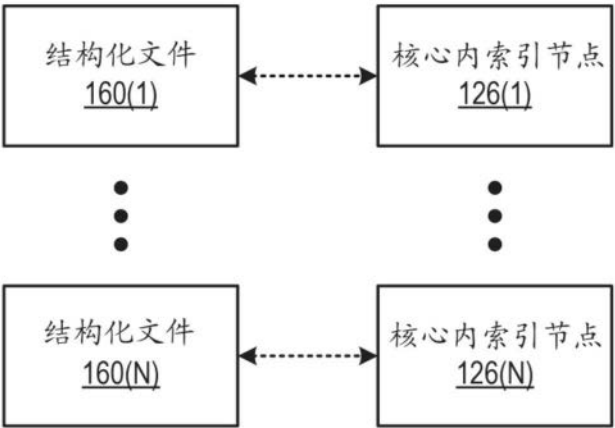


图1C

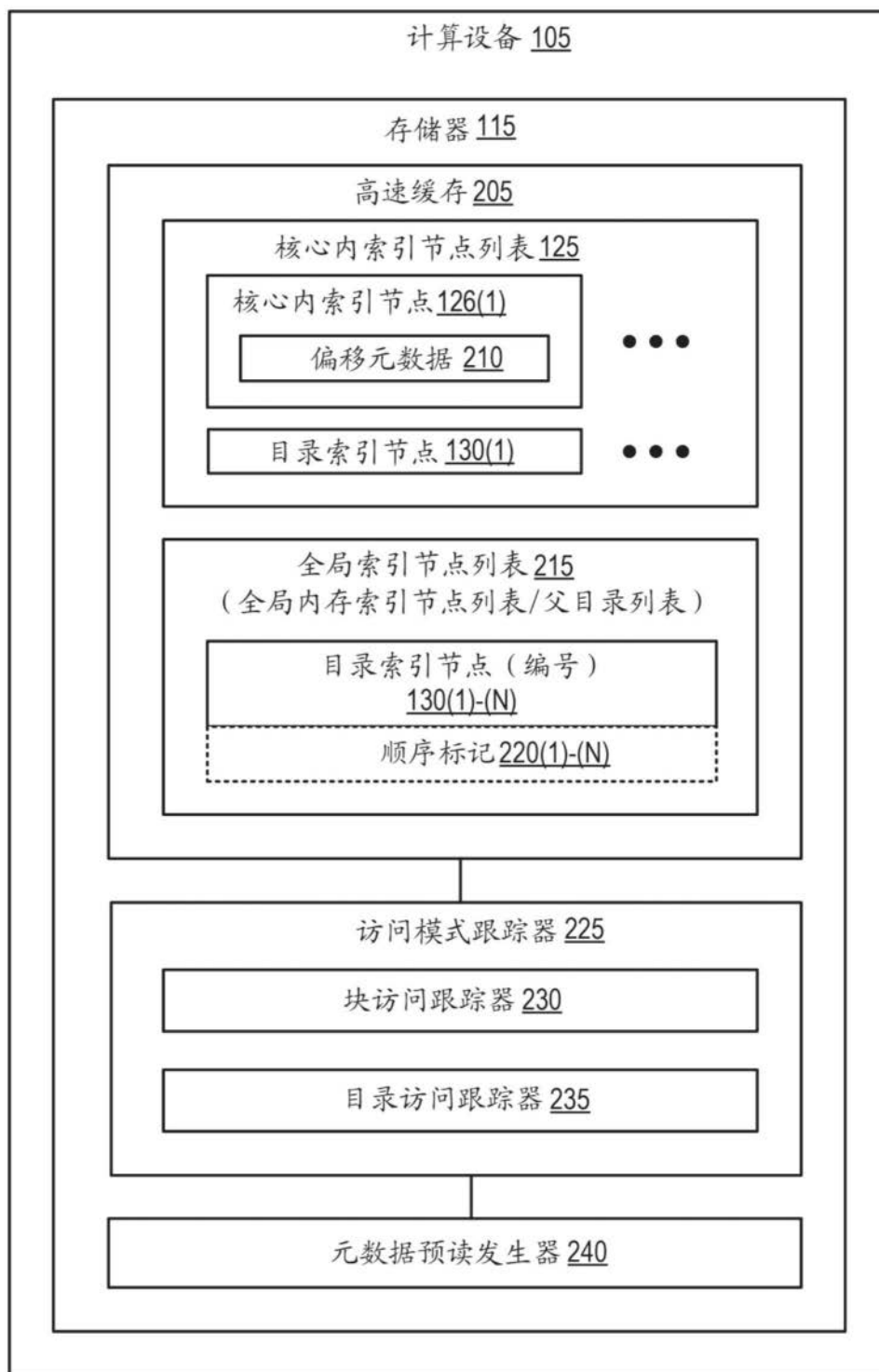
200A 

图2A

200B



图2B

200C

全局索引节点列表 <u>215</u> (针对目录 <u>265(1)-(N)</u>)	
目录索引节点 字段 <u>255</u>	顺序标记 字段 <u>260</u>
目录索引节点 <u>130(1)</u>	1
目录索引节点 <u>130(2)</u>	0
添加?	设置?
目录索引节点 <u>130(4)</u>	1
目录索引节点 <u>130(5)</u>	0

图2C

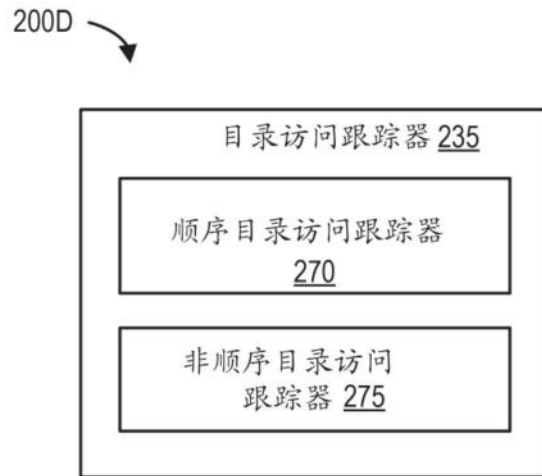


图2D

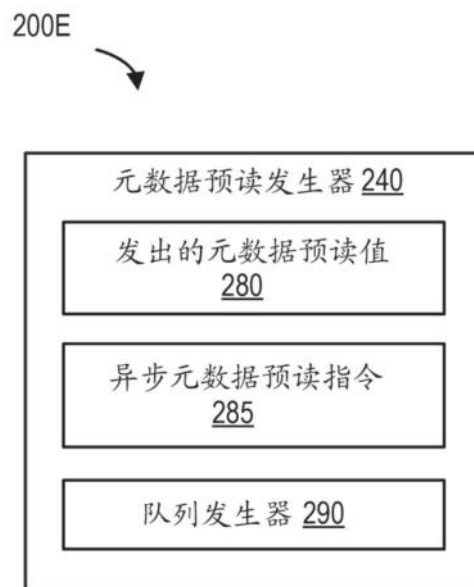


图2E

顺序索引节点结构
300A

目录 <u>265(1)</u>	
文件名称字段 <u>305</u>	索引节点编号字段 <u>310</u>
文件 <u>315(1)</u>	磁盘上索引节点 <u>165(4)</u>
文件 <u>315(2)</u>	磁盘上索引节点 <u>165(5)</u>
文件 <u>315(3)</u>	磁盘上索引节点 <u>165(6)</u>
文件 <u>315(4)</u>	磁盘上索引节点 <u>165(7)</u>
文件 <u>315(5)</u>	磁盘上索引节点 <u>165(8)</u>
文件 <u>315(6)</u>	磁盘上索引节点 <u>165(9)</u>

图3A

非顺序索引节点结构
300B

目录 <u>265(2)</u>	
文件名称字段 <u>305</u>	索引节点编号字段 <u>310</u>
文件 <u>315(1)</u>	磁盘上索引节点 <u>165(4)</u>
文件 <u>315(2)</u>	磁盘上索引节点 <u>165(9)</u>
文件 <u>315(3)</u>	磁盘上索引节点 <u>165(15)</u>
文件 <u>315(4)</u>	磁盘上索引节点 <u>165(11)</u>
文件 <u>315(5)</u>	磁盘上索引节点 <u>165(19)</u>
文件 <u>315(6)</u>	磁盘上索引节点 <u>165(6)</u>

图3B

300C

高速缓存 <u>205</u>				
父目录列表				
目录索引节点 <u>130(1)</u>	目录索引节点 <u>130(2)</u>	添加?	目录索引节点 <u>130(4)</u>	目录索引节点 <u>130(5)</u>
1	0	设置?	1	0

图3C

400A

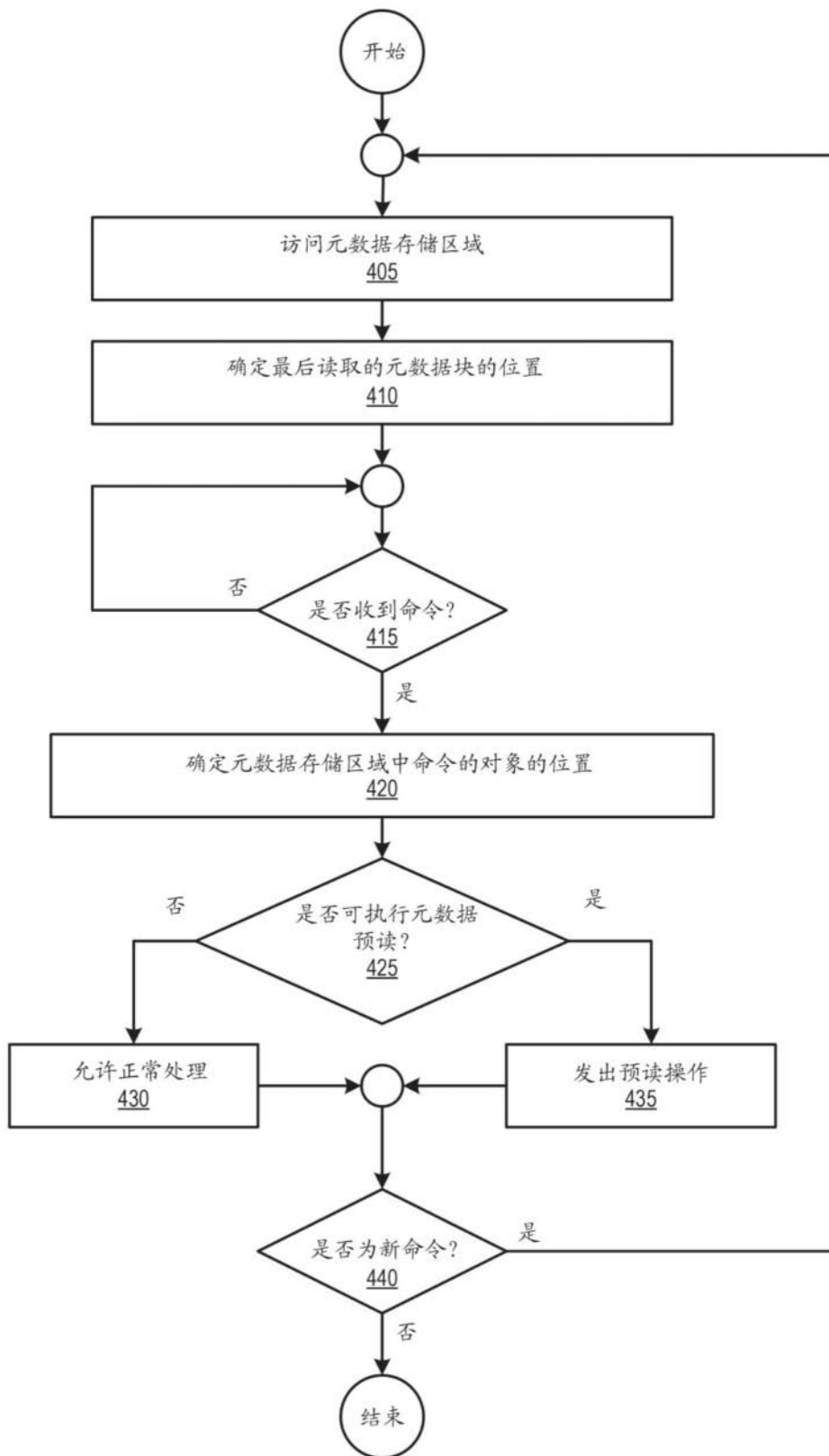


图4A

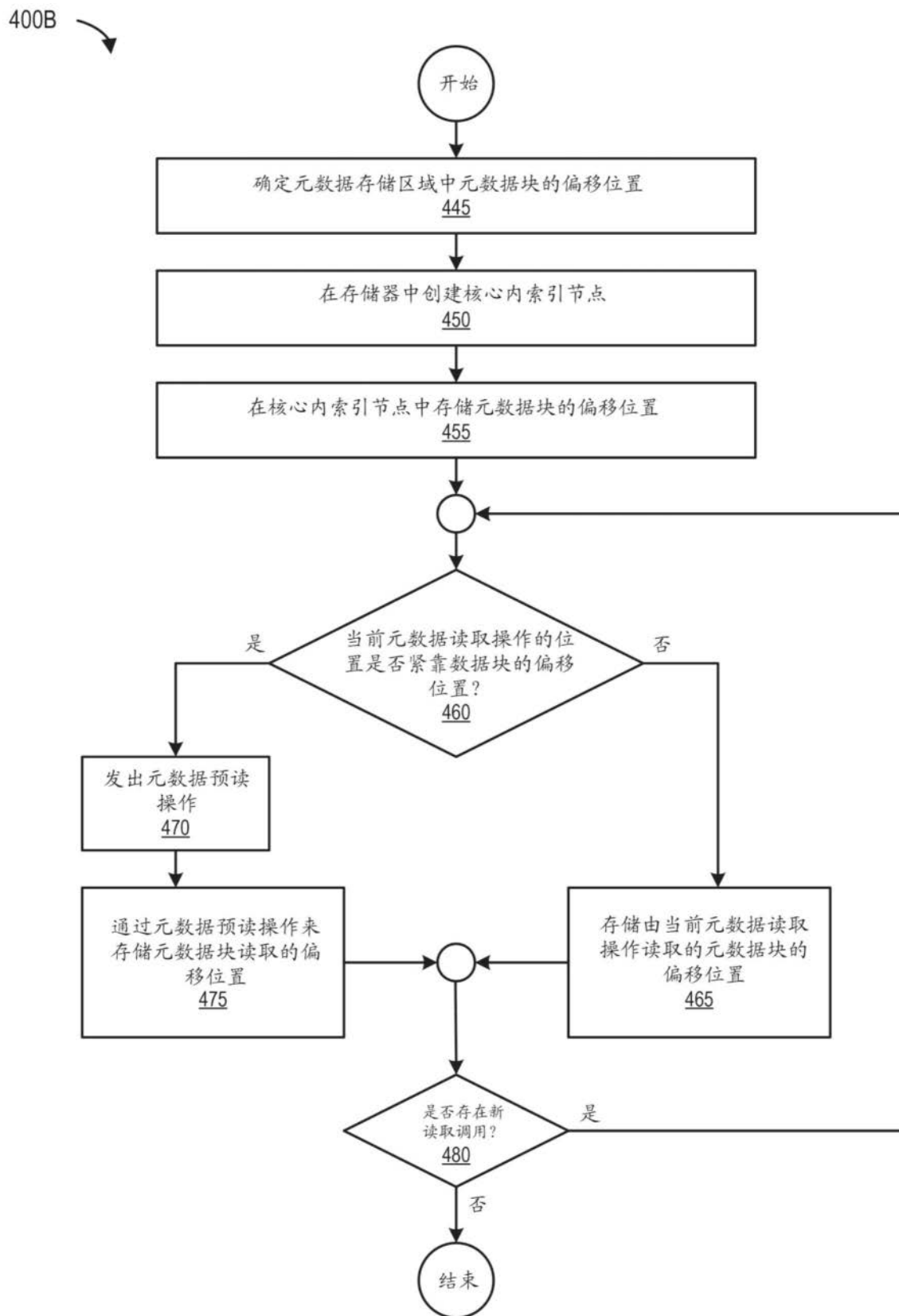


图4B

500A

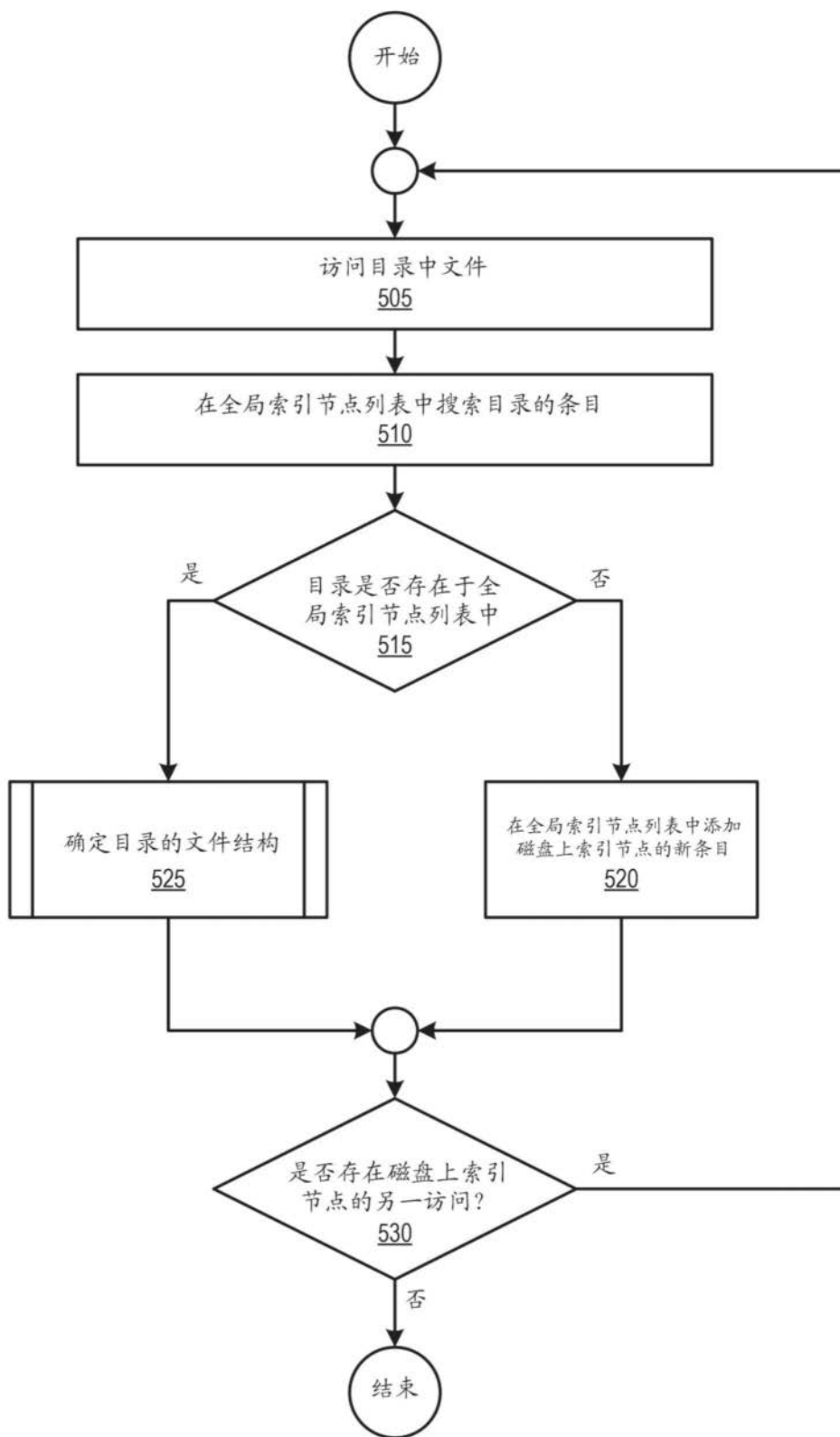


图5A

500B

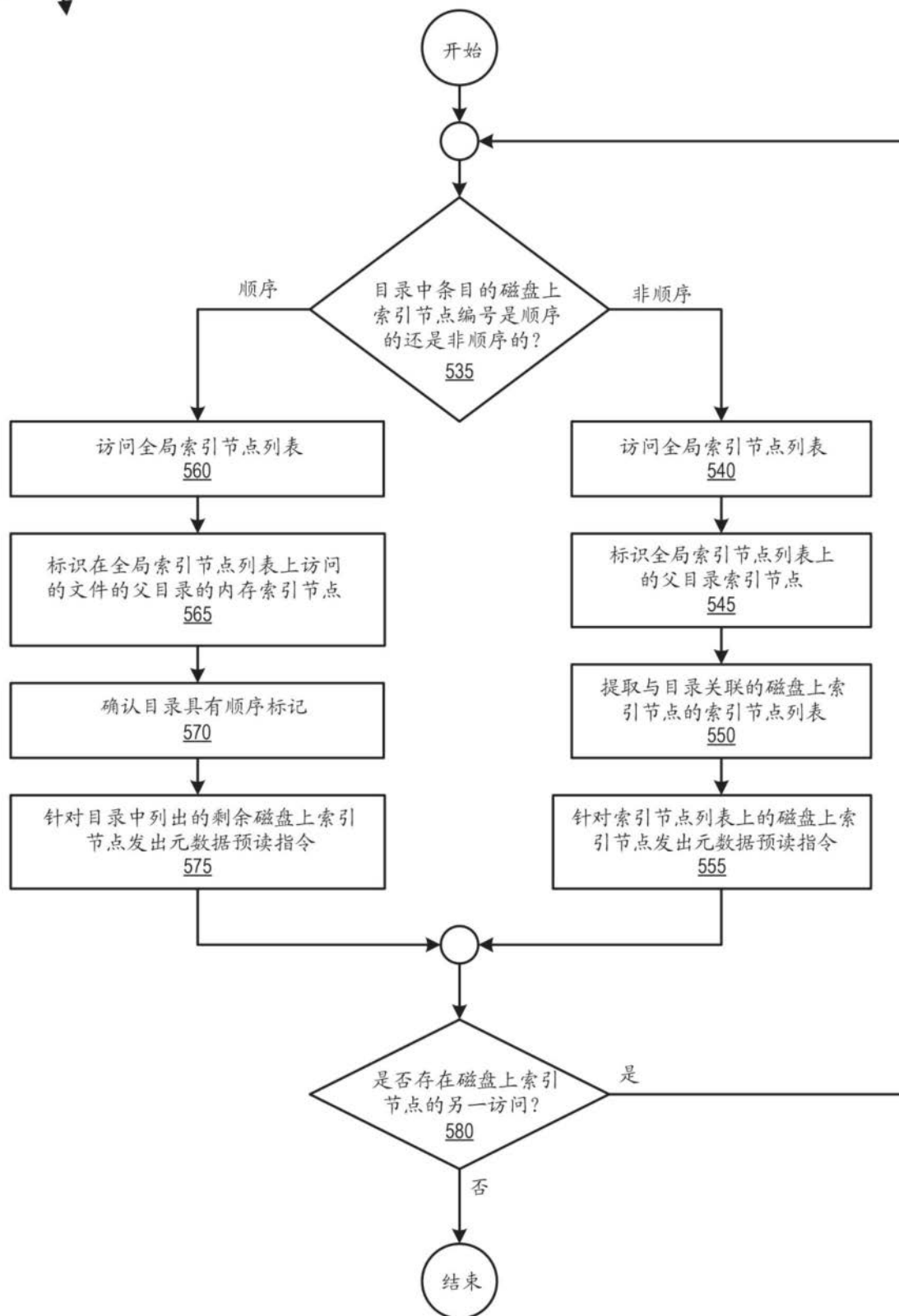


图5B

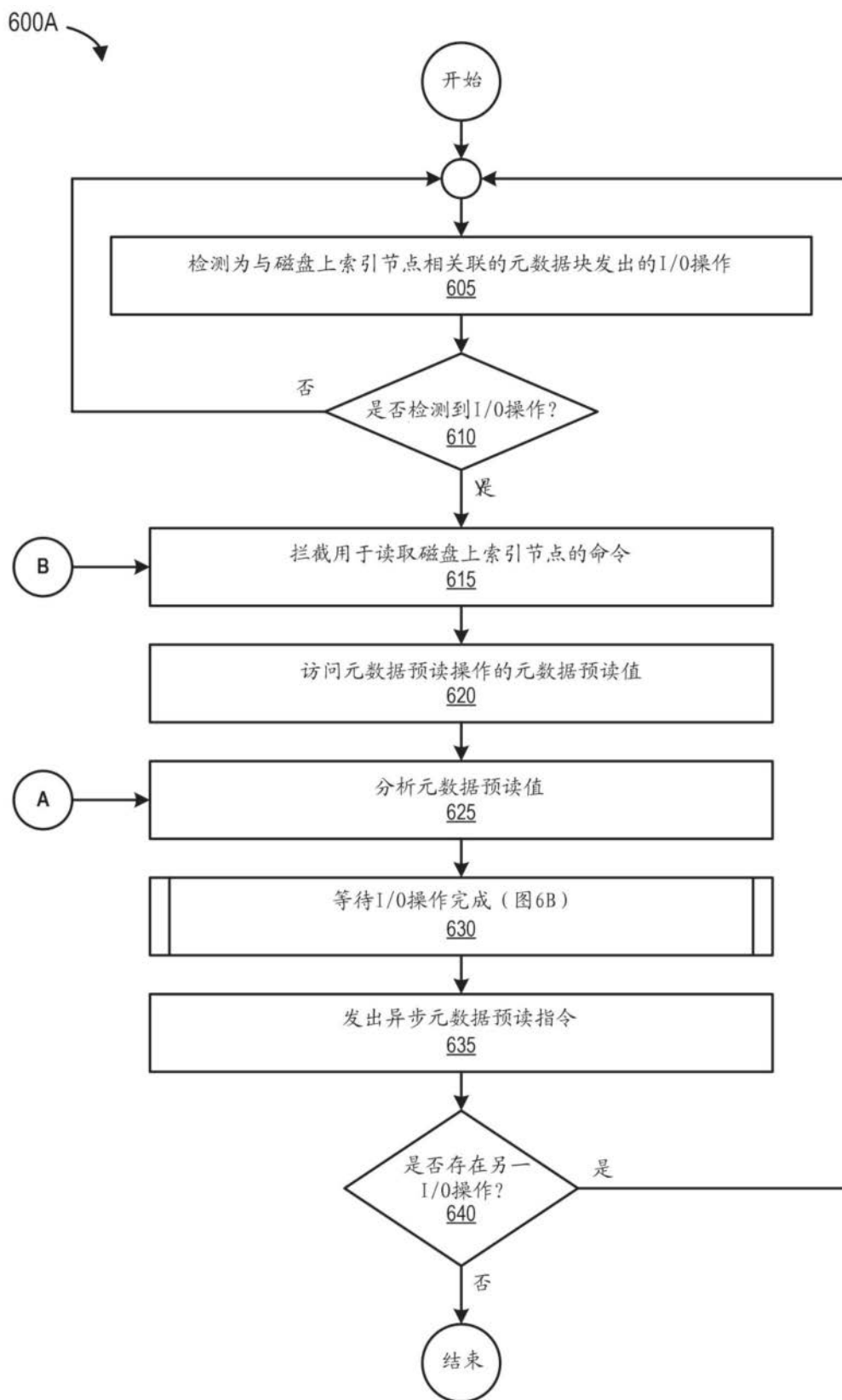


图6A

600B

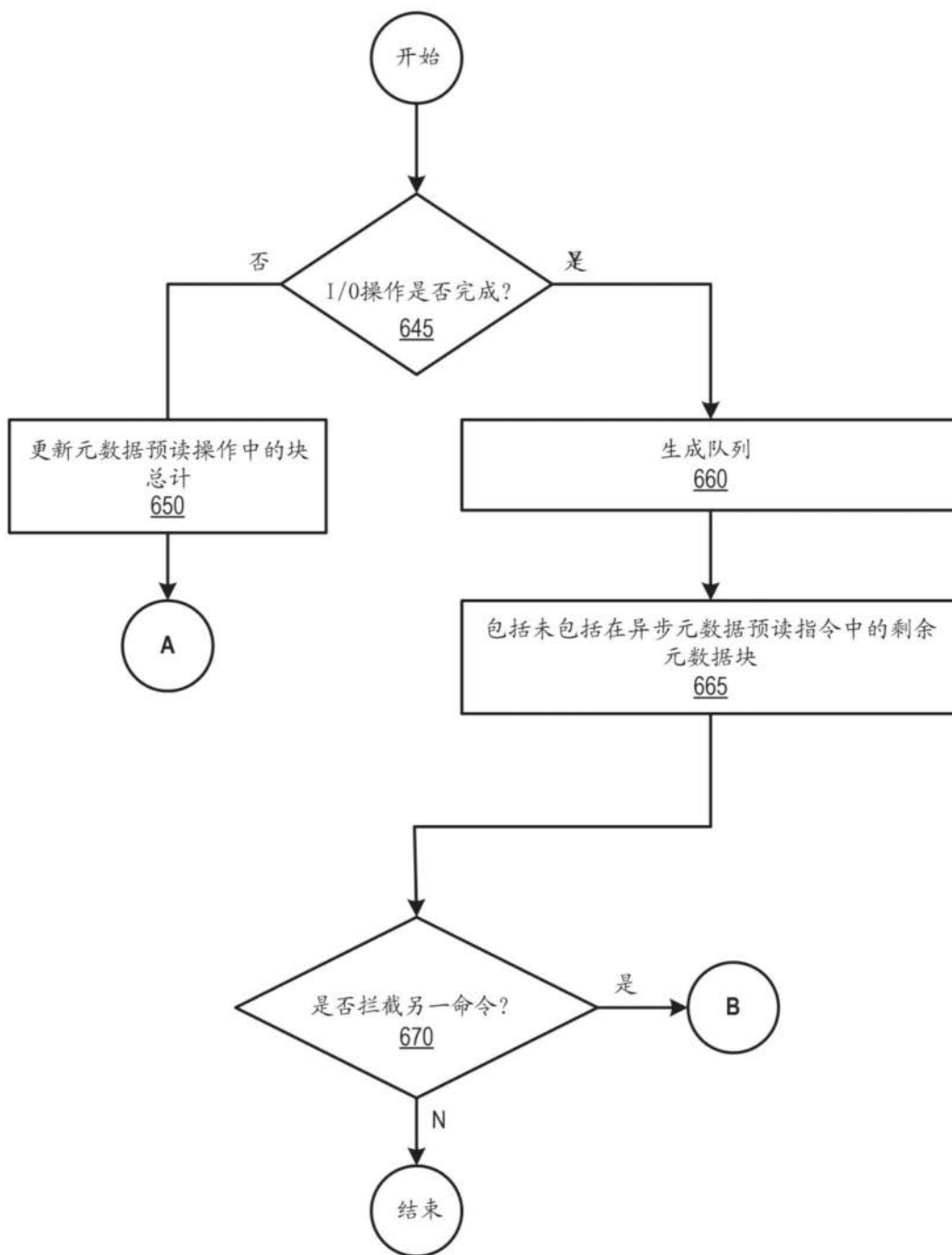


图6B

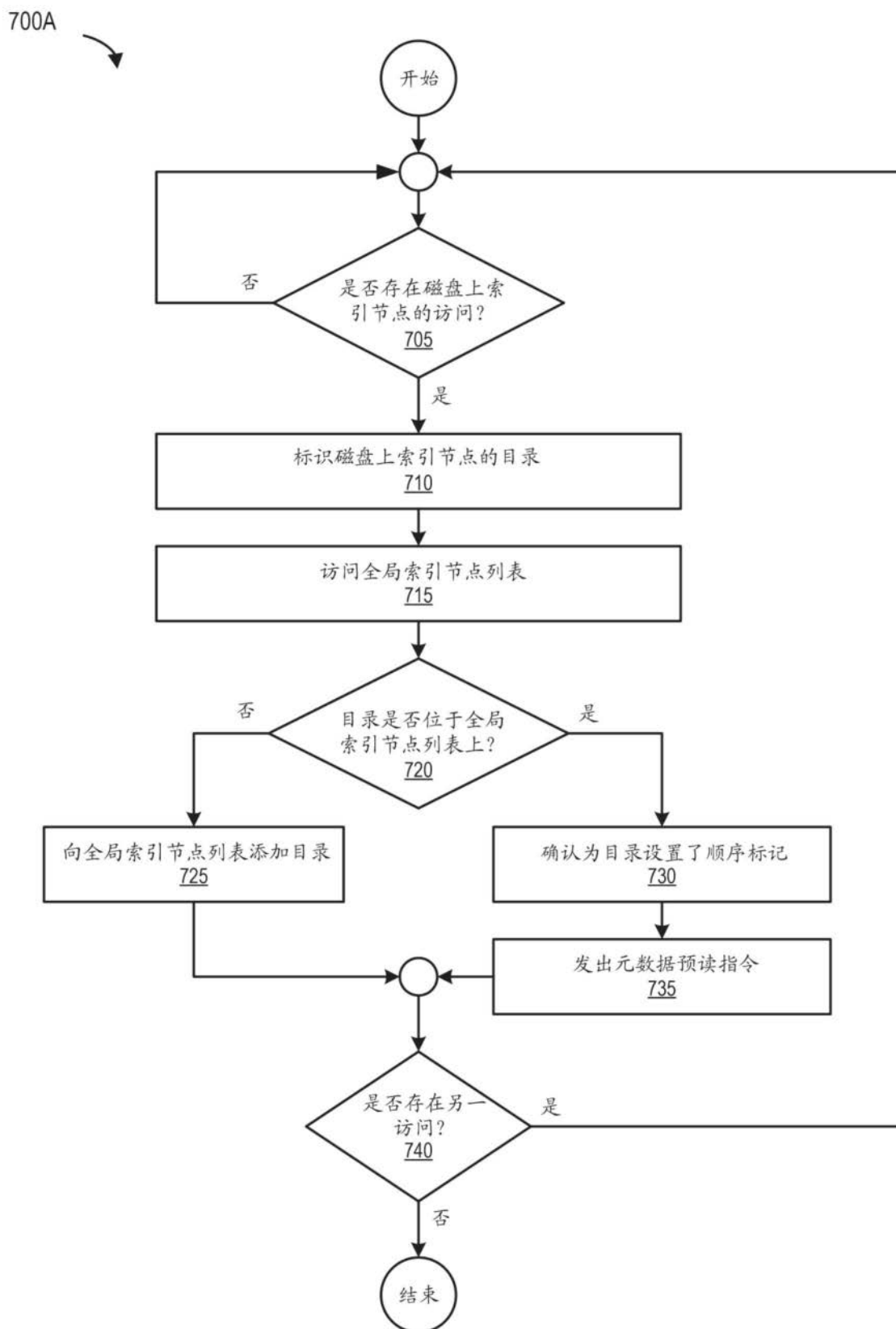


图7A

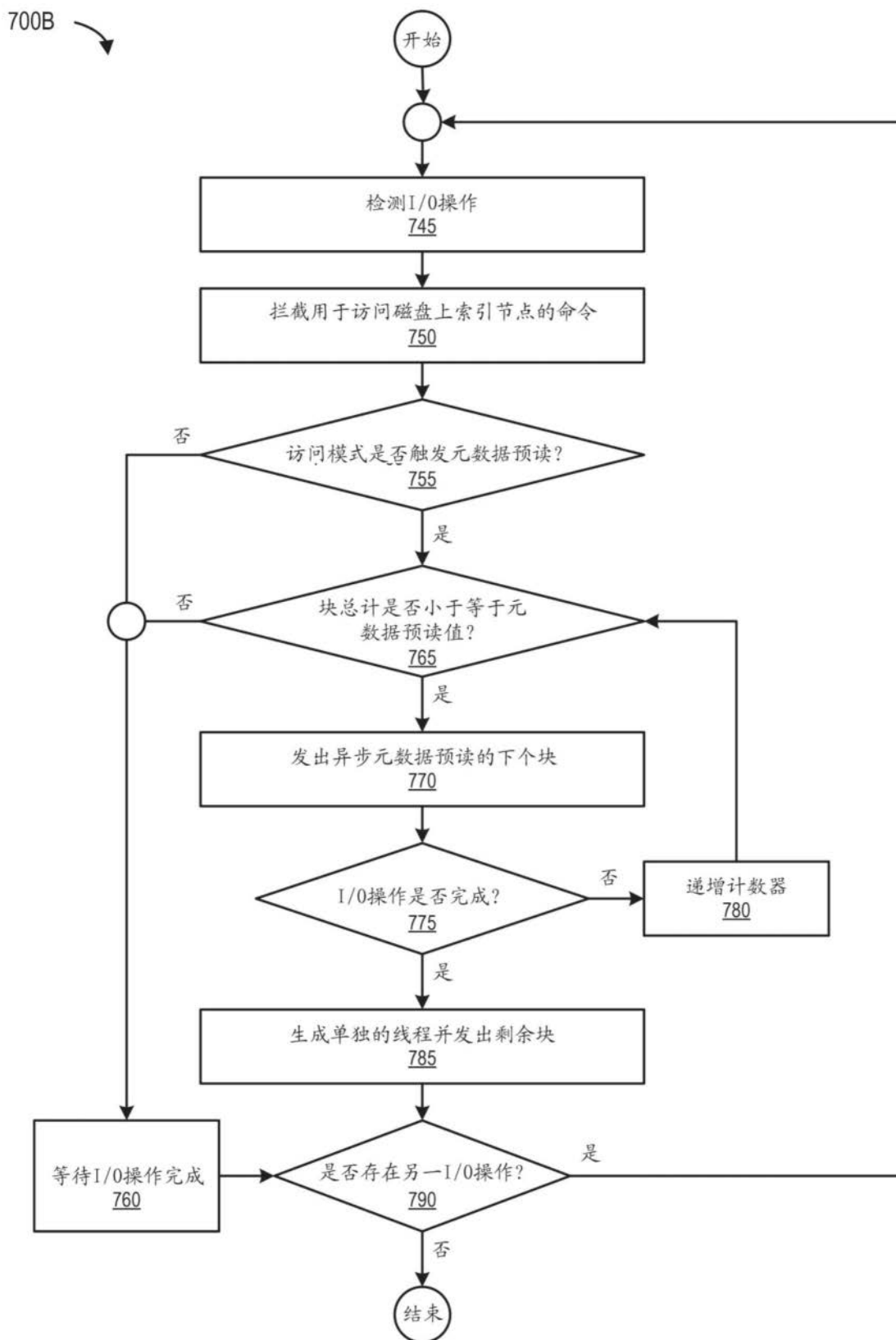


图7B

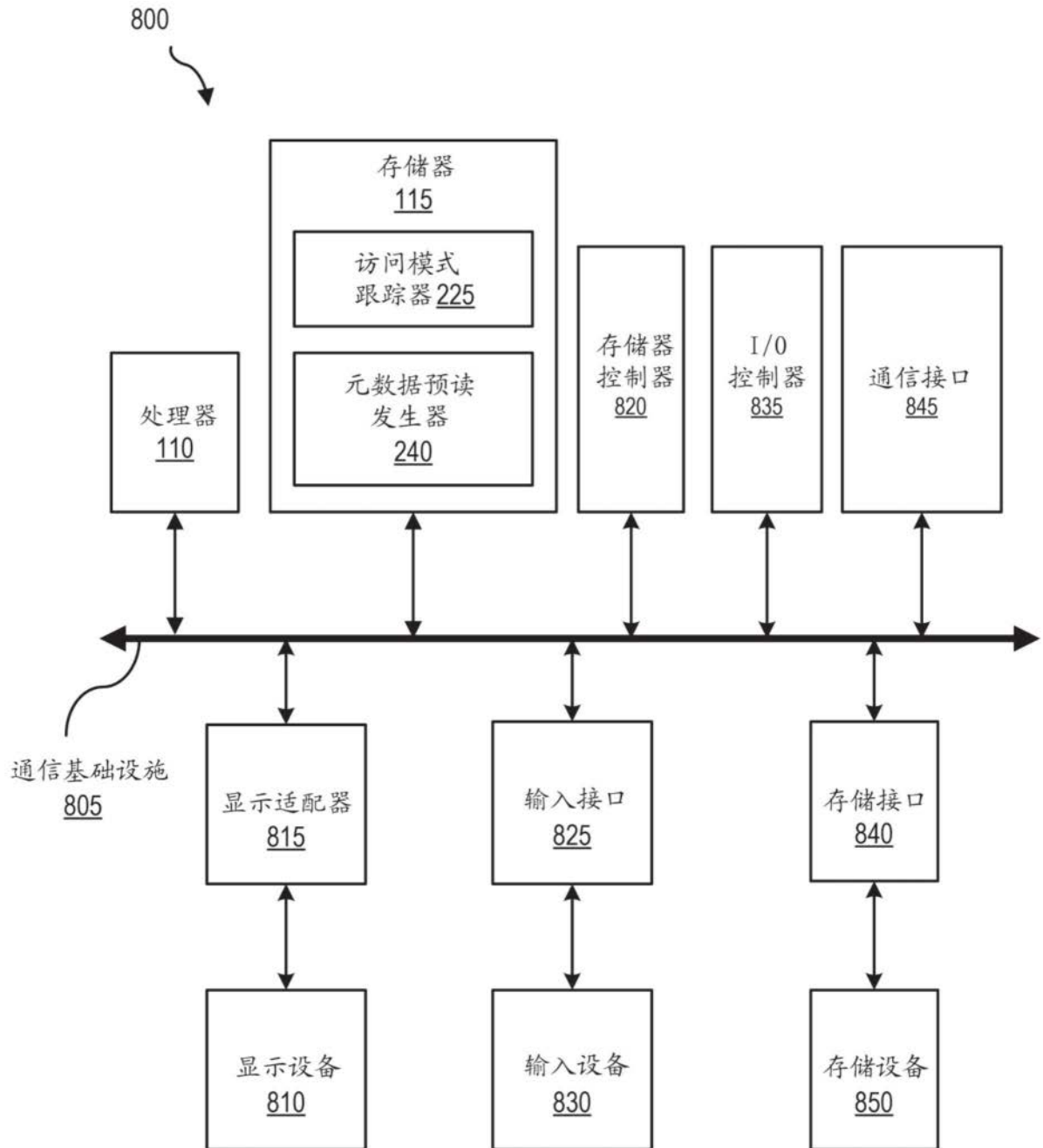


图8

900

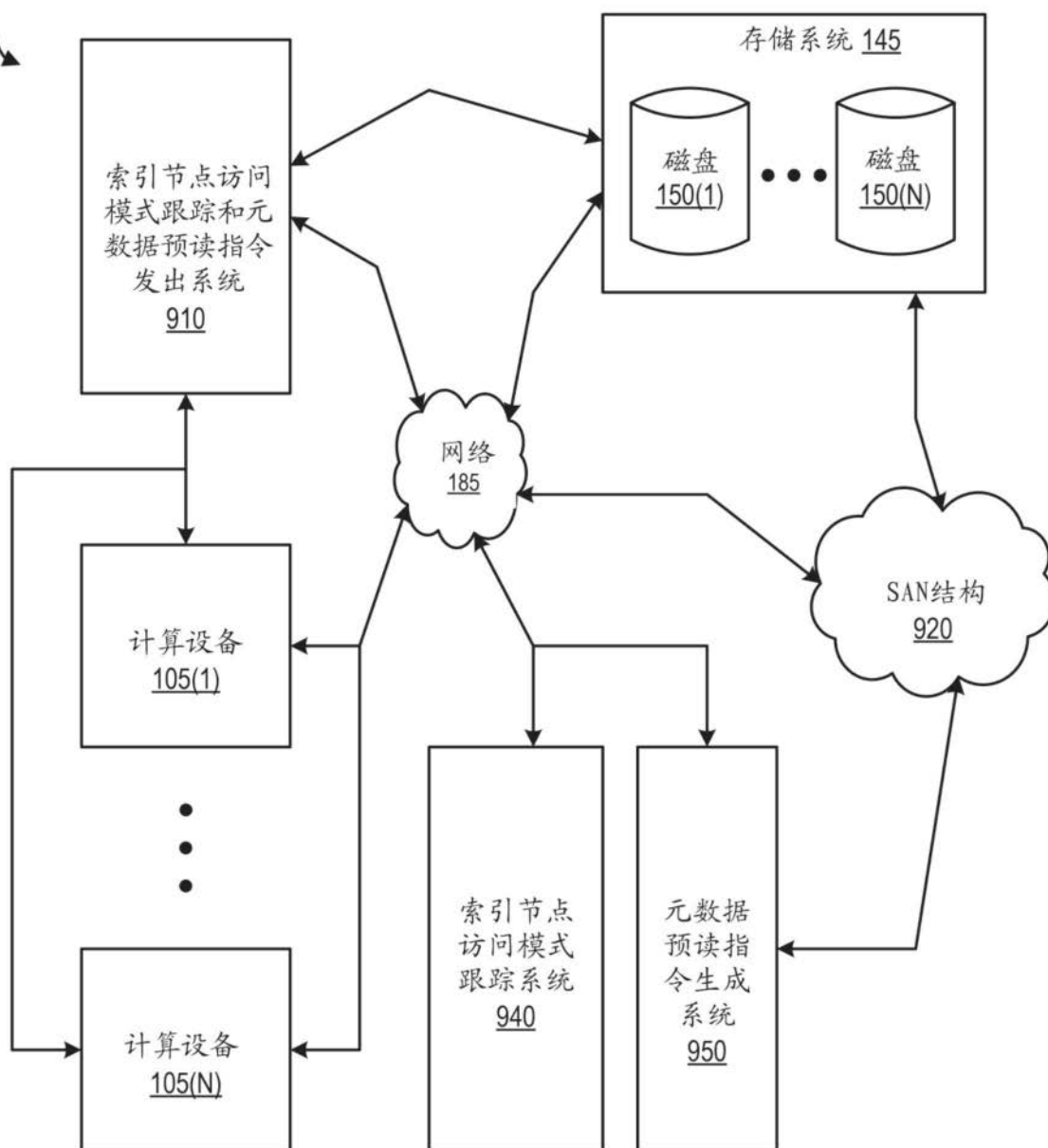


图9