(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 106469100 B
(45)授权公告日 2019.04.05

(21)申请号 201510504685.8

(22)申请日 2015.08.17

(65)同一申请的已公布的文献号
申请公布号 CN 106469100 A

(43)申请公布日 2017.03.01

(73)专利权人 华为技术有限公司
地址 518129 广东省深圳市龙岗区坂田华为总部办公楼

(72)发明人 庄仕岳

(74)专利代理机构 深圳市深佳知识产权代理事务所(普通合伙) 44285
代理人 王仲凯

(51)Int.Cl.
G06F 11/14(2006.01)
H04L 29/08(2006.01)

(56)对比文件
CN 101488104 A,2009.07.22,
US 7636724 B2,2009.12.22,
CN 103023968 A,2013.04.03,
CN 103699494 A,2014.04.02,

审查员 田晶

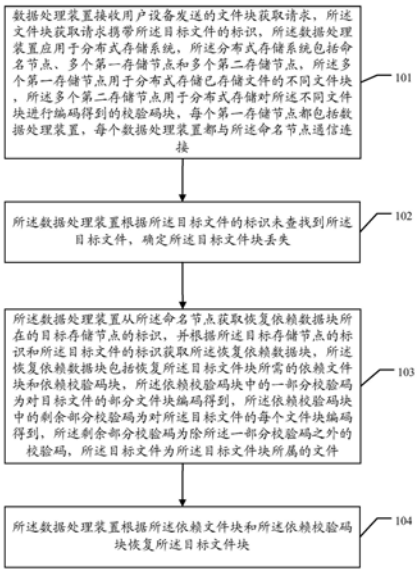
权利要求书4页 说明书21页 附图8页

(54)发明名称

一种数据恢复的方法、存储的方法相应的装置及系统

(57)摘要

本发明公开了一种数据恢复的方法,应用于分布式存储系统,分布式存储系统包括命名节点、多个第一存储节点和多个第二存储节点,每个第一存储节点都包括数据处理装置,每个数据处理装置都与命名节点通信连接,所述方法包括:数据处理装置接收用户设备发送的文件块获取请求,文件块获取请求携带目标文件的标识;根据目标文件的标识确定目标文件块丢失;从命名节点获取恢复依赖数据块所在的目标存储节点的标识,并根据目标存储节点的标识和目标文件的标识获取恢复依赖数据块,恢复目标文件块。本发明实施例提供的数据恢复的方法,降低了存储开销,数据恢复时目标文件块中的一部分只需要依赖部分依赖文件块就可以得到降低了数据恢复时的网络开销。



CN 106469100 B

1.一种数据恢复的方法，其特征在于，所述方法应用于分布式存储系统，所述分布式存储系统包括命名节点、多个第一存储节点和多个第二存储节点，所述多个第一存储节点用于分布式存储已存储文件的不同文件块，所述多个第二存储节点用于分布式存储对所述不同文件块进行编码得到的校验码块，每个第一存储节点都包括数据处理装置，每个数据处理装置都与所述命名节点通信连接，所述方法包括：

所述数据处理装置接收用户设备发送的文件块获取请求，所述文件块获取请求携带目标文件的标识；

所述数据处理装置根据所述目标文件的标识未查找到所述目标文件，确定目标文件块丢失；

所述数据处理装置从所述命名节点获取恢复依赖数据块所在的目标存储节点的标识，并根据所述目标存储节点的标识和所述目标文件的标识获取所述恢复依赖数据块，所述恢复依赖数据块包括恢复所述目标文件块所需的依赖文件块和依赖校验码块，所述依赖校验码块中的一部分校验码为对目标文件的部分文件块编码得到，所述依赖校验码块中的剩余部分校验码为对所述目标文件的每个文件块编码得到，所述剩余部分校验码为除所述一部分校验码之外的校验码，所述目标文件为所述目标文件块所属的文件；

所述数据处理装置根据所述依赖文件块和所述依赖校验码块恢复所述目标文件块。

2.根据权利要求1所述的方法，其特征在于，所述数据处理装置根据所述依赖文件块和所述依赖校验码块恢复所述目标文件块，包括：

所述数据处理装置根据部分字节编码函数、所述依赖文件块和所述依赖校验码块恢复所述目标文件块中的第一字节，所述部分字节编码函数为采用所述目标文件中的部分文件块进行编码得到编码结果的函数；

所述数据处理装置根据全字节编码函数、所述依赖文件块和所述依赖校验码块恢复所述目标文件块中的第二字节，所述全字节编码函数为采用所述目标文件中的每个文件块进行编码得到编码结果的函数。

3.根据权利要求2所述的方法，其特征在于，所述数据处理装置根据部分字节编码函数、所述依赖文件块和所述依赖校验码块恢复所述目标文件块中的第一字节，包括：

所述数据处理装置从第一编码参数所对应的依赖文件块中获取恢复所述第一字节所需的依赖字节，从第一编码结果所对应的依赖校验码块中获取恢复所述第一字节所需的校验码，所述第一编码参数为所述部分字节编码函数中的编码参数，所述第一编码结果为采用所述部分字节编码函数对所述第一编码参数所指示的依赖字节和所述第一字节进行编码所得到的结果；

所述数据处理装置根据恢复所述第一字节所需的依赖字节，对恢复所述第一字节所需的校验码进行解码，得到所述第一字节。

4.根据权利要求2或3所述的方法，其特征在于，所述数据处理装置根据全字节编码函数、所述依赖文件块和所述依赖校验码块恢复所述目标文件块中的第二字节，包括：

所述数据处理装置从第二编码参数所对应的依赖文件块中获取恢复所述第二字节所需的依赖字节，从第二编码结果所对应的依赖校验码块中获取恢复所述第二字节所需的校验码，所述第二编码参数为所述全字节编码函数中的编码参数，所述第二编码结果为采用所述全字节编码函数对所述第二编码参数所指示的依赖字节和所述第二字节进行编码所

得到的结果；

所述数据处理装置根据恢复所述第二字节所需的依赖字节，对恢复所述第二字节所需的校验码进行解码，得到所述第二字节。

5.一种数据存储的方法，其特征在于，所述方法应用于分布式存储系统，所述分布式存储系统包括命名节点、多个第一存储节点和多个第二存储节点，所述多个第一存储节点用于分布式存储待存储文件的不同文件块，所述多个第二存储节点用于分布式存储对所述不同文件块进行编码得到的校验码块，每个第二存储节点都包括数据处理装置，每个数据处理装置都与所述命名节点通信连接，所述方法包括：

所述数据处理装置接收所述命名节点发送的多个目标存储节点的标识和目标文件的标识，所述多个目标存储节点为已存储了所述目标文件的不同文件块的第一存储节点；

所述数据处理装置根据所述目标存储节点的标识和部分字节编码函数对所述目标文件中的部分文件块进行编码，得到第一校验码，所述部分字节编码函数为采用所述目标文件中的部分文件块进行编码得到编码结果的函数；

所述数据处理装置根据所述目标存储节点的标识和全字节编码函数对所述目标文件中的每个文件块进行编码，得到第二校验码，所述全字节编码函数为采用所述目标文件中的每个文件块进行编码得到编码结果的函数；

所述数据处理装置将所述第一校验码和所述第二校验码存储到所述数据处理装置所属的第二存储节点的存储空间中。

6.根据权利要求5所述的方法，其特征在于，所述数据处理装置根据所述目标存储节点的标识和部分字节编码函数对所述目标文件中的部分文件块进行编码，得到第一校验码，包括：

所述数据处理装置从第一编码参数所对应的目标存储节点中获取所述第一编码参数所指示的字节，所述第一编码参数为所述部分字节编码函数中的每个编码参数；

所述数据处理装置根据所述部分字节编码函数对所述第一编码参数所指示的字节进行编码，得到第一校验码。

7.根据权利要求5或6所述的方法，其特征在于，所述数据处理装置根据所述目标存储节点的标识和全字节编码函数对所述目标文件中的每个文件块进行编码，得到第二校验码，包括：

所述数据处理装置从第二编码参数所对应的目标存储节点中获取所述第二编码参数所指示的字节，所述第二编码参数为所述全字节编码函数中的每个编码参数；

所述数据处理装置根据所述全字节编码函数对所述第二编码参数所指示的字节进行编码，得到第二校验码。

8.根据权利要求5或6所述的方法，其特征在于，所述数据处理装置根据所述目标存储节点的标识和部分字节编码函数对所述目标文件中的部分文件块进行编码，得到第一校验码之前，所述方法还包括：

所述数据处理装置根据所述目标存储节点的数量和所述命名节点所指定的校验节点的数量，确定所述部分字节编码函数中第一参数的数量和紧邻的两个校验节点中的部分字节编码函数所包含的相同第一参数的个数，所述紧邻的两个校验节点所包含的部分字节编码函数中第一参数的重叠个数最多。

9.一种数据处理装置，其特征在于，应用于分布式存储系统，所述分布式存储系统包括命名节点、多个第一存储节点和多个第二存储节点，所述多个第一存储节点用于分布式存储已存储文件的不同文件块，所述多个第二存储节点用于分布式存储对所述不同文件块进行编码得到的校验码块，每个第一存储节点都包括所述数据处理装置，每个数据处理装置都与所述命名节点通信连接，所述数据处理装置包括：

接收模块，用于接收用户设备发送的文件块获取请求，所述文件块获取请求携带目标文件的标识；

确定模块，用于根据所述接收模块接收的所述目标文件的标识未查找到所述目标文件，确定目标文件块丢失；

获取模块，用于在所述确定模块确定所述目标文件丢失后，从所述命名节点获取恢复依赖数据块所在的目标存储节点的标识，并根据所述目标存储节点的标识和所述目标文件的标识获取所述恢复依赖数据块，所述恢复依赖数据块包括恢复所述目标文件块所需的依赖文件块和依赖校验码块，所述依赖校验码块中的一部分校验码为对目标文件的部分文件块编码得到，所述依赖校验码块中的剩余部分校验码为对所述目标文件的每个文件块编码得到，所述剩余部分校验码为除所述一部分校验码之外的校验码，所述目标文件为所述目标文件块所属的文件；

恢复模块，用于根据所述获取模块获取的所述依赖文件块和所述依赖校验码块恢复所述目标文件块。

10.根据权利要求9所述的数据处理装置，其特征在于，所述恢复模块包括：

第一恢复单元，用于根据部分字节编码函数、所述依赖文件块和所述依赖校验码块恢复所述目标文件块中的第一字节，所述部分字节编码函数为采用所述目标文件中的部分文件块进行编码得到编码结果的函数；

第二恢复单元，用于根据全字节编码函数、所述依赖文件块和所述依赖校验码块恢复所述目标文件块中的第二字节，所述全字节编码函数为采用所述目标文件中的每个文件块进行编码得到编码结果的函数。

11.根据权利要求10所述的数据处理装置，其特征在于，

所述第一恢复单元，具体用于从第一编码参数所对应的依赖文件块中获取恢复所述第一字节所需的依赖字节，从第一编码结果所对应的依赖校验码块中获取恢复所述第一字节所需的校验码，所述第一编码参数为所述部分字节编码函数中的编码参数，所述第一编码结果为采用所述部分字节编码函数对所述第一编码参数所指示的依赖字节和所述第一字节进行编码所得到的结果；根据恢复所述第一字节所需的依赖字节，对恢复所述第一字节所需的校验码进行解码，得到所述第一字节。

12.根据权利要求10或11所述的数据处理装置，其特征在于，

所述第二恢复单元，具体用于从第二编码参数所对应的依赖文件块中获取恢复所述第二字节所需的依赖字节，从第二编码结果所对应的依赖校验码块中获取恢复所述第二字节所需的校验码，所述第二编码参数为所述全字节编码函数中的编码参数，所述第二编码结果为采用所述全字节编码函数对所述第二编码参数所指示的依赖字节和所述第二字节进行编码所得到的结果；根据恢复所述第二字节所需的依赖字节，对恢复所述第二字节所需的校验码进行解码，得到所述第二字节。

13.一种数据处理装置，其特征在于，应用于分布式存储系统，所述分布式存储系统包括命名节点、多个第一存储节点和多个第二存储节点，所述多个第一存储节点用于分布式存储待存储文件的不同文件块，所述多个第二存储节点用于分布式存储对所述不同文件块进行编码得到的校验码块，每个第二存储节点都包括数据处理装置，每个数据处理装置都与所述命名节点通信连接，所述数据处理装置包括：

接收模块，用于接收所述命名节点发送的多个目标存储节点的标识和目标文件的标识，所述多个目标存储节点为已存储了所述目标文件的不同文件块的第一存储节点；

第一编码模块，用于根据所述接收模块接收的所述目标存储节点的标识和部分字节编码函数对所述目标文件中的部分文件块进行编码，得到第一校验码，所述部分字节编码函数为采用所述目标文件中的部分文件块进行编码得到编码结果的函数；

第二编码模块，用于根据所述接收模块接收的所述目标存储节点的标识和全字节编码函数对所述目标文件中的每个文件块进行编码，得到第二校验码，所述全字节编码函数为采用所述目标文件中的每个文件块进行编码得到编码结果的函数；

存储调度模块，用于将所述第一编码模块编码得到的所述第一校验码和所述第二编码模块编码得到的所述第二校验码存储到所述数据处理装置所属的第二存储节点的存储空间中。

14.根据权利要求13所述的数据处理装置，其特征在于，

所述第一编码模块，具体用于从第一编码参数所对应的目标存储节点中获取所述第一编码参数所指示的字节，所述第一编码参数为所述部分字节编码函数中的每个编码参数；根据所述部分字节编码函数对所述第一编码参数所指示的字节进行编码，得到第一校验码。

15.根据权利要求13或14所述的数据处理装置，其特征在于，

所述第二编码模块，具体用于从第二编码参数所对应的目标存储节点中获取所述第二编码参数所指示的字节，所述第二编码参数为所述全字节编码函数中的每个编码参数；根据所述全字节编码函数对所述第二编码参数所指示的字节进行编码，得到第二校验码。

16.根据权利要求13或14所述的数据处理装置，其特征在于，所述数据处理装置还包括：

确定模块，用于根据所述接收模块接收的所述目标存储节点的数量和所述命名节点所指定的校验节点的数量，确定所述部分字节编码函数中第一参数的数量和紧邻的两个校验节点中的部分字节编码函数所包含的相同第一参数的个数，所述紧邻的两个校验节点所包含的部分字节编码函数中第一参数的重叠个数最多。

17.一种分布式存储系统，其特征在于，包括命名节点、多个第一存储节点和多个第二存储节点，所述多个第一存储节点用于分布式存储已存储文件的不同文件块，所述多个第二存储节点用于分布式存储对所述不同文件块进行编码得到的校验码块，每个第一存储节点都包括第一数据处理装置，每个第二存储节点都包括第二数据处理装置，每个第一数据处理装置和每个第二数据处理装置都与所述命名节点通信连接；

所述第一数据处理装置为上述权利要求9-12任一所述的数据处理装置；

所述第二数据处理装置为上述权利要求13-16任一所述的数据处理装置。

一种数据恢复的方法、存储的方法相应的装置及系统

技术领域

[0001] 本发明涉及数据存储技术领域，具体涉及一种数据恢复的方法、存储的方法、相应的装置及系统。

背景技术

[0002] 在大容量分布存储系统中，为提高数据存储的可靠性，可以采用多副本方案，即把磁盘里面的数据复制到多个副本磁盘中，当其中任意一个磁盘失效时，从其他任意一个存活的磁盘中把数据读取出来放入新磁盘中即完成数据恢复，这种技术实现简单，恢复耗时最少，但存储开销很大。

[0003] 为了解决多副本方案存储开销大的问题，出现了纠删码(Reed-Solomon Code，RS)技术，例如：RS(10,4)，即对10个磁盘的数据进行编码，产生的编码结果存放在4个冗余磁盘里面，存储开销为(10+4)/10＝1.4倍，存储开销比多副本方案的明显减少，但是当一个磁盘失效时，需要从10个磁盘读取数据进行解码，才能实现数据的恢复。而多副本方案只需要从1个磁盘读数据即完成恢复，相比之下网络带宽开销增加了10倍，网络带宽开销大则是RS技术的缺点。

[0004] 由此可见，现有技术中的分布式存储方案，要么存储开销大，要么数据恢复时网络开销大。

发明内容

[0005] 为解决现有技术中数据分布式存储系统中数据恢复时网络开销大的问题，本发明实施例提供一种数据恢复的方法，可以在低存储开销的前提下，降低数据恢复时的网络开销。本发明实施例还提供了相应的数据存储的方法、相应的装置及系统。

[0006] 本发明第一方面提供一种数据恢复的方法，所述方法应用于分布式存储系统，所述分布式存储系统包括命名节点、多个第一存储节点和多个第二存储节点，所述多个第一存储节点用于分布式存储已存储文件的不同文件块，所述多个第二存储节点用于分布式存储对所述不同文件块进行编码得到的校验码块，每个第一存储节点都包括数据处理装置，每个数据处理装置都与所述命名节点通信连接，所述方法包括：

[0007] 所述数据处理装置接收用户设备发送的文件块获取请求，所述文件块获取请求携带所述目标文件的标识；

[0008] 所述数据处理装置根据所述目标文件的标识未查找到所述目标文件，确定所述目标文件块丢失；

[0009] 所述数据处理装置从所述命名节点获取恢复依赖数据块所在的目标存储节点的标识，并根据所述目标存储节点的标识和所述目标文件的标识获取所述恢复依赖数据块，所述恢复依赖数据块包括恢复所述目标文件块所需的依赖文件块和依赖校验码块，所述依赖校验码块中的一部分校验码为对目标文件的部分文件块编码得到，所述依赖校验码块中的剩余部分校验码为对所述目标文件的每个文件块编码得到，所述剩余部分校验码为除所

述一部分校验码之外的校验码，所述目标文件为所述目标文件块所属的文件；

[0010] 所述数据处理装置根据所述依赖文件块和所述依赖校验码块恢复所述目标文件块。

[0011] 结合第一方面，在第一种可能的实现方式中，所述数据处理装置根据所述依赖文件块和所述依赖校验码块恢复所述目标文件块，包括：

[0012] 所述数据处理装置根据部分字节编码函数、所述依赖文件块和所述依赖校验码块恢复所述目标文件块中的第一字节，所述部分字节编码函数为采用所述目标文件中的部分文件块进行编码得到编码结果的函数；

[0013] 所述数据处理装置根据全字节编码函数、所述依赖文件块和所述依赖校验码块恢复所述目标文件块中的第二字节，所述全字节编码函数为采用所述目标文件中的每个文件块进行编码得到编码结果的函数。

[0014] 结合第一方面第一种可能的实现方式，在第二种可能的实现方式中，所述数据处理装置根据部分字节编码函数、所述依赖文件块和所述依赖校验码块恢复所述目标文件块中的第一字节，包括：

[0015] 所述数据处理装置从第一编码参数所对应的依赖文件块中获取恢复所述第一字节所需的依赖字节，从第一编码结果所对应的依赖校验码块中获取恢复所述第一字节所需的校验码，所述第一编码参数为所述部分字节编码函数中的编码参数，所述第一编码结果为采用所述部分字节编码函数对所述第一编码参数所指示的依赖字节和所述第一字节进行编码所得到的结果；

[0016] 所述数据处理装置根据恢复所述第一字节所需的依赖字节，对恢复所述第一字节所需的校验码进行解码，得到所述第一字节。

[0017] 结合第一方面第一种或第二种可能的实现方式，在第三种可能的实现方式中，所述数据处理装置根据全字节编码函数、所述依赖文件块和所述依赖校验码块恢复所述目标文件块中的第二字节，包括：

[0018] 所述数据处理装置从第二编码参数所对应的依赖文件块中获取恢复所述第二字节所需的依赖字节，从第二编码结果所对应的依赖校验码块中获取恢复所述第二字节所需的校验码，所述第二编码参数为所述全字节编码函数中的编码参数，所述第二编码结果为采用所述全字节编码函数对所述第二编码参数所指示的依赖字节和所述第二字节进行编码所得到的结果；

[0019] 所述数据处理装置根据恢复所述第二字节所需的依赖字节，对恢复所述第二字节所需的校验码进行解码，得到所述第二字节。

[0020] 本发明第二方面提供一种数据存储的方法，所述方法应用于分布式存储系统，所述分布式存储系统包括命名节点、多个第一存储节点和多个第二存储节点，所述多个第一存储节点用于分布式存储待存储文件的不同文件块，所述多个第二存储节点用于分布式存储对所述不同文件块进行编码得到的校验码块，每个第二存储节点都包括数据处理装置，每个数据处理装置都与所述命名节点通信连接，所述方法包括：

[0021] 所述数据处理装置接收所述命名节点发送的多个目标存储节点的标识和目标文件的标识，所述多个目标存储节点为已存储了所述目标文件的不同文件块的第一存储节点；

[0022] 所述数据处理装置根据所述目标存储节点的标识和部分字节编码函数对所述目标文件中的部分文件块进行编码，得到第一校验码，所述部分字节编码函数为采用所述目标文件中的部分文件块进行编码得到编码结果的函数；

[0023] 所述数据处理装置根据所述目标存储节点的标识和全字节编码函数对所述目标文件中的每个文件块进行编码，得到第二校验码，所述全字节编码函数为采用所述目标文件中的每个文件块进行编码得到编码结果的函数；

[0024] 所述数据处理装置将所述第一校验码和所述第二校验码存储到所述数据处理装置所属的第二存储节点的存储空间中。

[0025] 结合第二方面，在第一种可能的实现方式中，所述数据处理装置根据所述目标存储节点的标识和部分字节编码函数对所述目标文件中的部分文件块进行编码，得到第一校验码，包括：

[0026] 所述数据处理装置从第一编码参数所对应的目标存储节点中获取所述第一编码参数所指示的字节，所述第一编码参数为所述部分字节编码函数中的每个编码参数；

[0027] 所述数据处理装置根据所述部分字节编码函数对所述第一编码参数所指示的字节进行编码，得到第一校验码。

[0028] 结合第二方面或第二方面第一种可能的实现方式，在第二种可能的实现方式中，所述数据处理装置根据所述目标存储节点的标识和全字节编码函数对所述目标文件中的每个文件块进行编码，得到第二校验码，包括：

[0029] 所述数据处理装置从第二编码参数所对应的目标存储节点中获取所述第二编码参数所指示的字节，所述第二编码参数为所述全字节编码函数中的每个编码参数；

[0030] 所述数据处理装置根据所述全字节编码函数对所述第二编码参数所指示的字节进行编码，得到第二校验码。

[0031] 结合第二方面或第二方面第一种可能的实现方式，在第三种可能的实现方式中，所述数据处理装置根据所述目标存储节点的标识和部分字节编码函数对所述目标文件中的部分文件块进行编码，得到第一校验码之前，所述方法还包括：

[0032] 所述数据处理装置根据所述目标存储节点的数量和所述命名节点所指定的校验节点的数量，确定所述部分字节编码函数中第一参数的数量和紧邻的两个校验节点中的部分字节编码函数所包含的相同第一参数的个数，所述紧邻的两个校验节点所包含的部分字节编码函数中第一参数的重叠个数最多。

[0033] 本发明第三方面提供一种数据处理装置，应用于分布式存储系统，所述分布式存储系统包括命名节点、多个第一存储节点和多个第二存储节点，所述多个第一存储节点用于分布式存储已存储文件的不同文件块，所述多个第二存储节点用于分布式存储对所述不同文件块进行编码得到的校验码块，每个第一存储节点都包括所述数据处理装置，每个数据处理装置都与所述命名节点通信连接，所述数据处理装置包括：

[0034] 接收模块，用于接收用户设备发送的文件块获取请求，所述文件块获取请求携带所述目标文件的标识；

[0035] 确定模块，用于根据所述接收模块接收的所述目标文件的标识未查找到所述目标文件，确定所述目标文件块丢失；

[0036] 获取模块，用于在所述确定模块确定所述目标文件丢失后，从所述命名节点获取

恢复依赖数据块所在的目标存储节点的标识，并根据所述目标存储节点的标识和所述目标文件的标识获取所述恢复依赖数据块，所述恢复依赖数据块包括恢复所述目标文件块所需的依赖文件块和依赖校验码块，所述依赖校验码块中的一部分校验码为对目标文件的部分文件块编码得到，所述依赖校验码块中的剩余部分校验码为对所述目标文件的每个文件块编码得到，所述剩余部分校验码为除所述一部分校验码之外的校验码，所述目标文件为所述目标文件块所属的文件；

[0037] 恢复模块，用于根据所述获取模块获取的所述依赖文件块和所述依赖校验码块恢复所述目标文件块。

[0038] 结合第三方面，在第一种可能的实现方式中，所述恢复模块包括：

[0039] 第一恢复单元，用于根据部分字节编码函数、所述依赖文件块和所述依赖校验码块恢复所述目标文件块中的第一字节，所述部分字节编码函数为采用所述目标文件中的部分文件块进行编码得到编码结果的函数；

[0040] 第二恢复单元，用于根据全字节编码函数、所述依赖文件块和所述依赖校验码块恢复所述目标文件块中的第二字节，所述全字节编码函数为采用所述目标文件中的每个文件块进行编码得到编码结果的函数。

[0041] 结合第三方面第一种可能的实现方式，在第二种可能的实现方式中，

[0042] 所述第一恢复单元，具体用于从第一编码参数所对应的依赖文件块中获取恢复所述第一字节所需的依赖字节，从第一编码结果所对应的依赖校验码块中获取恢复所述第一字节所需的校验码，所述第一编码参数为所述部分字节编码函数中的编码参数，所述第一编码结果为采用所述部分字节编码函数对所述第一编码参数所指示的依赖字节和所述第一字节进行编码所得到的结果；根据恢复所述第一字节所需的依赖字节，对恢复所述第一字节所需的校验码进行解码，得到所述第一字节。

[0043] 结合第三方面第一种或第二种可能的实现方式，在第三种可能的实现方式中，

[0044] 所述第二恢复单元，具体用于从第二编码参数所对应的依赖文件块中获取恢复所述第二字节所需的依赖字节，从第二编码结果所对应的依赖校验码块中获取恢复所述第二字节所需的校验码，所述第二编码参数为所述全字节编码函数中的编码参数，所述第二编码结果为采用所述全字节编码函数对所述第二编码参数所指示的依赖字节和所述第二字节进行编码所得到的结果；根据恢复所述第二字节所需的依赖字节，对恢复所述第二字节所需的校验码进行解码，得到所述第二字节。

[0045] 本发明第四方面提供一种数据处理装置，应用于分布式存储系统，所述分布式存储系统包括命名节点、多个第一存储节点和多个第二存储节点，所述多个第一存储节点用于分布式存储待存储文件的不同文件块，所述多个第二存储节点用于分布式存储对所述不同文件块进行编码得到的校验码块，每个第二存储节点都包括数据处理装置，每个数据处理装置都与所述命名节点通信连接，所述数据处理装置包括：

[0046] 接收模块，用于接收所述命名节点发送的多个目标存储节点的标识和目标文件的标识，所述多个目标存储节点为已存储了所述目标文件的不同文件块的第一存储节点；

[0047] 第一编码模块，用于根据所述接收模块接收的所述目标存储节点的标识和部分字节编码函数对所述目标文件中的部分文件块进行编码，得到第一校验码，所述部分字节编码函数为采用所述目标文件中的部分文件块进行编码得到编码结果的函数；

[0048] 第二编码模块，用于根据所述接收模块接收的所述目标存储节点的标识和全字节编码函数对所述目标文件中的每个文件块进行编码，得到第二校验码，所述全字节编码函数为采用所述目标文件中的每个文件块进行编码得到编码结果的函数；

[0049] 存储调度模块，用于将所述第一编码模块编码得到的所述第一校验码和所述第二编码模块编码得到的所述第二校验码存储到所述数据处理装置所属的第二存储节点的存储空间中。

[0050] 结合第四方面，在第一种可能的实现方式中，

[0051] 所述第一编码模块，具体用于从第一编码参数所对应的目标存储节点中获取所述第一编码参数所指示的字节，所述第一编码参数为所述部分字节编码函数中的每个编码参数；根据所述部分字节编码函数对所述第一编码参数所指示的字节进行编码，得到第一校验码。

[0052] 结合第四方面或第四方面第一种可能的实现方式，在第二种可能的实现方式中，

[0053] 所述第二编码模块，具体用于从第二编码参数所对应的目标存储节点中获取所述第二编码参数所指示的字节，所述第二编码参数为所述全字节编码函数中的每个编码参数；根据所述全字节编码函数对所述第二编码参数所指示的字节进行编码，得到第二校验码。

[0054] 结合第四方面或第四方面第一种可能的实现方式，在第三种可能的实现方式中，

[0055] 所述数据处理装置还包括：

[0056] 确定模块，用于根据所述接收模块接收的所述目标存储节点的数量和所述命名节点所指定的校验节点的数量，确定所述部分字节编码函数中第一参数的数量和紧邻的两个校验节点中的部分字节编码函数所包含的相同第一参数的个数，所述紧邻的两个校验节点所包含的部分字节编码函数中第一参数的重叠个数最多。

[0057] 本发明第五方面提供一种分布式存储系统，包括命名节点、多个第一存储节点和多个第二存储节点，所述多个第一存储节点用于分布式存储已存储文件的不同文件块，所述多个第二存储节点用于分布式存储对所述不同文件块进行编码得到的校验码块，每个第一存储节点都包括第一数据处理装置，每个第二存储节点都包括第二数据处理装置，每个第一数据处理装置和每个第二数据处理装置都与所述命名节点通信连接；

[0058] 所述第一数据处理装置为上述第三方面或第三方面任一实现方式所述的数据处理装置；

[0059] 所述第二数据处理装置为上述第四方面或第四方面任一实现方式所述的数据处理装置。

[0060] 本发明实施例提供的数据恢复的方法，应用于分布式存储系统，所述分布式存储系统包括命名节点、多个第一存储节点和多个第二存储节点，所述多个第一存储节点用于分布式存储已存储文件的不同文件块，所述多个第二存储节点用于分布式存储对所述不同文件块进行编码得到的校验码块，每个第一存储节点都包括数据处理装置，每个数据处理装置都与所述命名节点通信连接，所述方法包括：所述数据处理装置接收用户设备发送的文件块获取请求，所述文件块获取请求携带所述目标文件的标识；所述数据处理装置根据所述目标文件的标识未查找到所述目标文件，确定所述目标文件块丢失；所述数据处理装置从所述命名节点获取恢复依赖数据块所在的存储节点的标识，所述恢复依赖数据块包括

恢复所述目标文件块所需的依赖文件块和依赖校验码块，所述依赖校验码块中的一部分校验码为对目标文件的部分文件块编码得到，所述依赖校验码块中的剩余部分校验码为对所述目标文件的每个文件块编码得到，所述剩余部分校验码为除所述一部分校验码之外的校验码，所述目标文件为所述目标文件块所属的文件；所述数据处理装置根据所述依赖文件块和所述依赖校验码块恢复所述目标文件块。与现有技术中数据无法同时兼顾数据存储开销和数据恢复时的网络开销相比，本发明实施例提供的数据恢复的方法，校验码块是通过部分字节编码和全字节编码结合得到的，降低了存储开销，数据恢复时目标文件块中的一部分只需要依赖部分依赖文件块就可以得到降低了数据恢复时的网络开销。

附图说明

[0061] 为了更清楚地说明本发明实施例中的技术方案，下面将对实施例描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明的一些实施例，对于本领域技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

[0062] 图1是本发明实施例中分布式存储系统的一实施例示意图；

[0063] 图2是本发明实施例中分布式存储系统的另一实施例示意图；

[0064] 图3是本发明实施例中数据存储的方法的一实施例示意图；

[0065] 图4是本发明实施例中一场景示例示意图；

[0066] 图5是本发明实施例中数据恢复的方法的一实施例示意图；

[0067] 图6是本发明实施例中数据恢复的方法的另一实施例示意图；

[0068] 图7是本发明实施例中数据存储的方法的另一实施例示意图；

[0069] 图8是本发明实施例中数据处理装置的一实施例示意图；

[0070] 图9是本发明实施例中数据处理装置的一实施例示意图；

[0071] 图10是本发明实施例中数据处理装置的一实施例示意图；

[0072] 图11是本发明实施例中数据处理装置的一实施例示意图；

[0073] 图12是本发明实施例中数据处理装置的一实施例示意图；

[0074] 图13是本发明实施例中数据处理装置的一实施例示意图；

[0075] 图14是本发明实施例中数据处理装置的一实施例示意图。

具体实施方式

[0076] 本发明实施例提供一种数据恢复的方法，可以在低存储开销的前提下，降低数据恢复时的网络开销。本发明实施例还提供了相应的数据存储的方法、相应的装置及系统。以下分别进行详细说明。

[0077] 下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域技术人员在没有作出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

[0078] 图1为本发明实施例中分布式存储系统的一实施例示意图。

[0079] 如图1所示，分布式存储系统包括命名节点(NameNode)和多个存储节点(Node)，每

个存储节点都与命名节点通信连接。实际上，命名节点与存储节点间可以通过交换机通信连接。

[0080] 图2为本发明实施例中分布式存储系统的另一实施例示意图。

[0081] 如图2所示，多个存储节点被划分到多个机架，同一个机架内的存储节点可以通过1G的交换机通信连接，图2中表示了3个存储节点组成一个机架，它们通过交换机通信连接；机架间则通过更高带宽的交换机通信连接；NameNode节点管理着整个集群的元数据，直接联入上层交换机，NameNode、存储节点、交换机、机架构成了一个分布式存储集群。元数据在本发明实施例中指的是文件中各文件块与存储路径的对应关系。一个文件可以分布式存储在多个存储节点上，例如：5个存储节点上，则该文件有5个文件块，而且，各文件块的数据内容不相同。用户设备对于分布式存储系统的使用包括数据存入和数据读出两个方面，用户在存入或读出数据时，通过网络接入分布式存储系统。

[0082] 需要说明的是，本发明实施例提供的存储节点可以是独立的物理主机，也可以是位于一个或多个物理主机上的虚拟机。

[0083] 下面分别从数据存入和数据读出两个方面介绍本发明实施例中数据存储的过程和数据恢复的过程：

[0084] 需要预先说明的是，本发明实施例中多个包括两个以及两个以上。

[0085] 首先结合图3介绍数据存储的过程：

[0086] 图3为本发明实施例中数据存储的方法的一实施例示意图。

[0087] 本发明实施例的分布式存储系统包括命名节点、多个第一存储节点和多个第二存储节点，所述多个第一存储节点用于分布式存储待存储文件的不同文件块，所述多个第二存储节点用于分布式存储对所述不同文件块进行编码得到的校验码块，每个第二存储节点都包括数据处理装置，每个数据处理装置都与所述命名节点通信连接，如图3所示，第一存储节点有十个，十个第一存储节点的标识分别为N1、N2至N10，第二存储节点有四个，四个第二存储节点的标识分别为N11至N14，当然，图3中只是举例，实际上分布式存储系统中有很多第一存储节点和第二存储节点，每个存储节点都有其对应的标识。当用户设备要将目标文件存储到分布式存储系统时，先向命名节点发送存储请求，命名节点可以根据目标文件的大小、第一存储存储空间的大小等参数，为该目标文件分配第一存储节点，命名节点中维护有每个第一存储节点和每个第二存储节点存储空间的大小，同时，命名节点在为目标文件分配第一存储节点后，还会对应维护其目标文件的标识与分配的每个第一存储节点的标识的对应关系。例如：当为目标文件分配了N1至N10十个第一存储节点后，命名节点就会维护目标文件的标识和N1至N10的对应关系。用户设备接收命名节点发送的文件存储响应，该文件存储响应中携带N1至N10十个第一存储节点的标识，则用户设备将目标文件拆分为十个文件块，分别存储到十个第一存储节点中，当然，用户设备也可以不对目标文件进行拆分，而是轮循存储，十个第一存储节点中的文件块大小可以相同，也可以不相同，本发明实施例中对此不作限定。在目标文件存储到十个第一存储节点后，为了确保数据的可靠性，命名节点会为该目标文件分配第二存储节点，用于存储该目标文件十个文件块，命名节点为该目标函数分配了四个第二存储节点，四个第二存储节点的标识分别为N11至N14，然后，命名节点向N11至N14中的数据处理装置发送十个第一存储节点N1至N10的标识和目标文件的标识。这十个第一存储节点的标识为目标存储节点的标识。数据处理装置在接收到N1至N10

的标识和目标文件的标识后，即获知了要对N1至N10中目标文件的10个文件块进行编码。

[0088] 本发明实施例中，数据处理装置对目标文件中各文件块的编码采用两种编码函数，第一种为部分字节编码函数，第二种为全字节编码函数。部分字节编码函数为采用所述目标文件中的部分文件块进行编码得到编码结果的函数。全字节编码函数为采用所述目标文件中的每个文件块进行编码得到编码结果的函数。

[0089] 所述数据处理装置根据所述目标存储节点的标识和部分字节编码函数对所述目标文件中的部分文件块进行编码，得到第一校验码。

[0090] 所述数据处理装置根据所述目标存储节点的标识和全字节编码函数对所述目标文件中的每个文件块进行编码，得到第二校验码。

[0091] 所述数据处理装置将所述第一校验码和所述第二校验码存储到所述数据处理装置所属的第二存储节点的存储空间中。

[0092] 下面结合实例说明本发明实施例中部分字节编码和全字节编码的过程：

[0093] 本发明实施例应用场景中，用于存储目标文件的第一存储节点有十个，为该目标文件分配的用户存储校验码块的第二存储节点有4个。a1,a2,…,a10分别为N1,N2,…,N10十个第一存储节点上目标文件的不同文件块中的第n个字节，b1,b2,…,b10分别为N1,N2,…,N10十个第一存储节点上目标文件的不同文件块中的第n+1个字节；c11,c21,c31,c41是第二存储节点N11,N12,N13,N14上的第n个字节，它的值通过部分字节编码函数g1,g2,g3,g4计算得到，c12,c22,c32,c42是第二存储节点N11,N12,N13,N14上的第n+1个字节，它的值通过全字节编码函数f1,f2,f3,f4+g5计算得到。

[0094] 数据处理装置对上述应用场景的编码原理可以参阅表1进行理解：

[0095] 表1:应用场景实例的编码原理

[0096]

存储节点标识	第n个字节	第n+1个字节
N1	a_1	b_1
N2	a_2	b_2
…	…	…
N9	a_9	b_9
N10	a_{10}	b_{10}
N11	$c_{11}=g_1(a_1, a_2, a_3, a_4)$	$c_{12}=f_1(b_1, b_2, \ldots, b_9, b_{10})$
N12	$c_{21}=g_2(a_3, a_4, a_5, a_6)$	$c_{22}=f_2(b_1, b_2, \ldots, b_9, b_{10})$
N13	$c_{31}=g_3(a_5, a_6, a_7, a_8)$	$c_{32}=f_3(b_1, b_2, \ldots, b_9, b_{10})$
N14	$c_{41}=g_4(a_7, a_8, a_9, a_{10})$	$c_{42}=f_4(b_1, b_2, \ldots, b_9, b_{10})+g_5(a_9, a_{10}, a_1, a_2)$

[0097] 其中，$g_1(a_1,a_2,a_3,a_4)=B_{11}*a_1+B_{12}*a_2+B_{13}*a_3+B_{14}*a_4$；

[0098] $g_2(a_3,a_4,a_5,a_6)=B_{23}*a_3+B_{24}*a_4+B_{25}*a_5+B_{26}*a_6$；

[0099] $f_1(b_1,b_2,\cdots,b_9,b_{10})=B_{11}*b_1+B_{12}*b_2+\cdots,+B_{19}*b_9+B_{1,10}*b_{10}$；

[0100] $f_2(b_1,b_2,\cdots,b_9,b_{10})=B_{21}*b_1+B_{22}*b_2+\cdots,+B_{29}*b_9+B_{2,10}*b_{10}$；

[0101] 本处只是列举了其中几个函数关系式，其他的函数关系式可以参阅上述关系式进行改写，本处不一一赘述。

[0102] 而且，上述只是举例说明了每个文件块中第n个字节和第n+1个字节的编码原理，实际上，每个文件块中其他字节也可以使用这两个编码函数进行编码。本处不一一列出。

[0103] 本发明对a组采用部分字节编码，只有一部分字节作为它的编码函数的输入参数，对b组采用全字节编码，所有字节都作为它的编码函数的输入参数。

[0104] 本发明实施例采用的编码矩阵为：

[0105] $$\begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1,10} \\ B_{21} & B_{22} & \cdots & B_{2,10} \\ B_{31} & B_{32} & \cdots & B_{3,10} \\ B_{41} & B_{42} & \cdots & B_{4,10} \end{bmatrix}$$

[0106] 本发明实施例中对编码矩阵值不做限定，只要保证逆矩阵存在就可以，各编码函数g1～g5，f1～f4中需要的系数从该矩阵中获取。

[0107] 由上述示例的部分字节编码函数可见，紧邻的两个部分字节编码函数的字节参数有两个字节的重叠。这样，可以保证每个字节都能被编进多个校验码里，即使该目标文件有多个文件块丢失时，也能保证丢失的字节正常的恢复出来。紧邻的两个部分字节编码函数即为g_1与g_2，g_2与g_3，g_3与g_4这种参数重叠数最多的函数。

[0108] 当然，重叠两个字节只是本示例的情况，具体部分字节编码函数中字节参数的数量和紧邻的两个部分编码函数中字节重复的数量可以根据该目标文件的第一存储节点的数量和第二存储节点的数量确定。保证每个字节都被编进至少两个校验码块即可。

[0109] 如图4所示，$a_1,a_2,\cdots,a_9,a_{10}$这十个字节，按照紧邻的两个函数重叠两个字节的情况可以有五个部分字节编码函数。

[0110] 分别为：$g_1(a_1,a_2,a_3,a_4)$，$g_2(a_3,a_4,a_5,a_6)$，(a_5,a_6,a_7,a_8)，$g_4(a_7,a_8,a_9,a_{10})$，(a_9,a_{10},a_1,a_2)，保证了$a_1,a_2,\cdots,a_9,a_{10}$每个字节都被编进了两个校验码块中。实际上，也可以有三个重叠字节，这样，$a_1,a_2,\cdots,a_9,a_{10}$每个字节都被编进3个校验码块中，可靠性更高，但存储开销会增大。

[0111] 下面举例计算本发明上述示例的部分字节编码函数中选择4个字节作为参与编码的字节，重叠长度为两个字节的原因。

[0112] 假设：nodeNum为需要参与部分字节编码的存储节点的数量，len为节点重叠部分的长度，第二存储节点的数量用r表示，第一存储节点的数量用k表示，通过如下公式确定nodeNum和len。

[0113] $r=(k-nodeNum)/(nodeNum-len)+1$；

[0114] 带宽开销减少率$ratio=(k-nodeNum)/2k$；

[0115] 本实施例中，通过多次赋值尝试。在$nodeNum=4,len=2$是最佳的参数组合，增加nodeNum的值会导致网络带宽开销增加，而增加len的值则会导致存储开销增加。因此，确定最终$nodeNum=4,len=2$，通过这两个值可以确定每个部分字节编码函数。

[0116] 在一个数据处理装置对目标文件的不同文件块中的相应字节都编码结束后，就将校验码存储在该数据处理装置所在的第二存储节点的存储空间中。

[0117] 在上述示例中，N11、N12、N13和N14中的数据处理装置都分别按照各自的部分字节编码函数和全字节编码函数进行编码，可以并行编码，提高了编码效率。

[0118] 下面结合图5介绍本发明实施例中数据恢复的过程：

[0119] 图5所示的分布式存储系统包括命名节点、多个第一存储节点和多个第二存储节点，所述多个第一存储节点用于分布式存储已存储文件的不同文件块，所述多个第二存储节点用于分布式存储对所述不同文件块进行编码得到的校验码块，每个第一存储节点都包括数据处理装置，每个数据处理装置都与所述命名节点通信连接。

[0120] 在图5所示的示例中，结合图3所示的场景，目标文件存储在N1,N2,…,N10十个第一存储节点上，目标文件的校验码存储在N11、N12、N13和N14四个第二存储节点上。每个第一存储节点上都有一个数据处理装置。实际上第一存储节点和第二存储节点上都有数据处理装置，只是在数据恢复的解码过程中，第二存储节点上的数据处理装置暂时不需要使用。

[0121] 在用户想使用目标文件时，用户设备向命名节点发送目标文件获取请求，目标文件获取请求中携带该目标文件的标识。命名节点根据该目标文件的标识，从该目标文件存储时建立的目标文件的标识与第一存储节点的标识的关联关系中，确定该目标文件存储在N1,N2,…,N10十个第一存储节点上。则命名节点向用户设备返回N1,N2,…,N10这十个第一存储节点的标识。

[0122] 用户设备根据N1,N2,…,N10这十个第一存储节点的标识，向N1,N2,…,N10这十个第一存储节点发送文件块的获取请求，该文件块的获取请求中携带目标文件的标识。

[0123] 如果N1,N2,…,N10上存储的该目标文件的文件块没有丢失，则会分别向用户设备返回相应的文件块，但本发明场景示例中第一存储节点N1上的文件块丢失，则意味着a_1、b_1等存储在第一存储节点N1上的相关字节丢失，需要恢复丢失的文件块后再向用户设备返回恢复的文件块。下面以恢复a_1、b_1为例说明本发明实施例中数据恢复的过程。

[0124] 由图3对应的场景示例可知，a_1是根据部分字节编码函数编码在校验码$c_{11}=g_1(a_1,a_2,a_3,a_4)$中的，所以要恢复$a_1$需要依赖$a_2,a_3,a_4$和$c_{11}$四个值，则可以从N2、N3、N4和N11中分别获取a_2,a_3,a_4和c_{11}这四个值，然后按照编码的逆过程进行解码，即可得到a_1。同理，要恢复b_1可以在获取$b_2,\cdots,b_9,b_{10}$和c_{12}后，通过$c_{12}=f_1(b_1,b_2,\cdots,b_9,b_{10})$函数的逆过程恢复出$b_1$。当然，恢复$b_1$还可以采用其他几个全字节编码函数，但原理是相同的。

[0125] 由以上可知，本发明实施例提供的数据存储的方法，采用部分字节编码和全字节编码混合的方式，提高了字节可靠性的同时，降低了存储空间，提高了编码效率。同时，在数据恢复时，针对部分字节编码函数编码的字节，不需要获取每个字节进行解码，降低了网络数据恢复时的网络开销。

[0126] 参阅图6，本发明实施例提供的数据恢复的方法的一实施例包括：

[0127] 101、数据处理装置接收用户设备发送的文件块获取请求，所述文件块获取请求携带所述目标文件的标识，所述数据处理装置应用于分布式存储系统，所述分布式存储系统包括命名节点、多个第一存储节点和多个第二存储节点，所述多个第一存储节点用于分布式存储已存储文件的不同文件块，所述多个第二存储节点用于分布式存储对所述不同文件块进行编码得到的校验码块，每个第一存储节点都包括数据处理装置，每个数据处理装置

都与所述命名节点通信连接。

[0128] 102、所述数据处理装置根据所述目标文件的标识未查找到所述目标文件，确定所述目标文件块丢失。

[0129] 103、所述数据处理装置从所述命名节点获取恢复依赖数据块所在的目标存储节点的标识，并根据所述目标存储节点的标识和所述目标文件的标识获取所述恢复依赖数据块，所述恢复依赖数据块包括恢复所述目标文件块所需的依赖文件块和依赖校验码块，所述依赖校验码块中的一部分校验码为对目标文件的部分文件块编码得到，所述依赖校验码块中的剩余部分校验码为对所述目标文件的每个文件块编码得到，所述剩余部分校验码为除所述一部分校验码之外的校验码，所述目标文件为所述目标文件块所属的文件。

[0130] 104、所述数据处理装置根据所述依赖文件块和所述依赖校验码块恢复所述目标文件块。

[0131] 与现有技术中数据无法同时兼顾数据存储开销和数据恢复时的网络开销相比，本发明实施例提供的数据恢复的方法，校验码块是通过部分字节编码和全字节编码结合得到的，降低了存储开销，数据恢复时目标文件块中的一部分只需要依赖部分依赖文件块就可以得到降低了数据恢复时的网络开销。

[0132] 可选地，在上述图6对应的实施例的基础上，本发明实施例提供的数据恢复的方法的第一个可选实施例中，所述数据处理装置根据所述依赖文件块和所述依赖校验码块恢复所述目标文件块，可以包括：

[0133] 所述数据处理装置根据部分字节编码函数、所述依赖文件块和所述依赖校验码块恢复所述目标文件块中的第一字节，所述部分字节编码函数为采用所述目标文件中的部分文件块进行编码得到编码结果的函数；

[0134] 所述数据处理装置根据全字节编码函数、所述依赖文件块和所述依赖校验码块恢复所述目标文件块中的第二字节，所述全字节编码函数为采用所述目标文件中的每个文件块进行编码得到编码结果的函数。

[0135] 可选地，在上述图6对应的第一个可选实施例的基础上，本发明实施例提供的数据恢复的方法的第二个可选实施例中，所述数据处理装置根据部分字节编码函数、所述依赖文件块和所述依赖校验码块恢复所述目标文件块中的第一字节，可以包括：

[0136] 所述数据处理装置从第一编码参数所对应的依赖文件块中获取恢复所述第一字节所需的依赖字节，从第一编码结果所对应的依赖校验码块中获取恢复所述第一字节所需的校验码，所述第一编码参数为所述部分字节编码函数中的编码参数，所述第一编码结果为采用所述部分字节编码函数对所述第一编码参数所指示的依赖字节和所述第一字节进行编码所得到的结果；

[0137] 所述数据处理装置根据恢复所述第一字节所需的依赖字节，对恢复所述第一字节所需的校验码进行解码，得到所述第一字节。

[0138] 可选地，在上述图6对应的第一个或第二个可选实施例的基础上，本发明实施例提供的数据恢复的方法的第三个可选实施例中，所述数据处理装置根据全字节编码函数、所述依赖文件块和所述依赖校验码块恢复所述目标文件块中的第二字节，可以包括：

[0139] 所述数据处理装置从第二编码参数所对应的依赖文件块中获取恢复所述第二字节所需的依赖字节，从第二编码结果所对应的依赖校验码块中获取恢复所述第二字节所需

的校验码，所述第二编码参数为所述全字节编码函数中的编码参数，所述第二编码结果为采用所述全字节编码函数对所述第二编码参数所指示的依赖字节和所述第二字节进行编码所得到的结果；

[0140] 所述数据处理装置根据恢复所述第二字节所需的依赖字节，对恢复所述第二字节所需的校验码进行解码，得到所述第二字节。

[0141] 图6对应的实施例或可选实施例中，第一编码参数指的是部分字节编码函数中的参数，例如：图3场景示例中的g_1函数中的a_1,a_2,a_3,a_4，第二编码参数指的是全字节编码函数中的参数，例如：f函数中的b_1,b_2,…,b_9,b_{10}，第一字节可以参阅a组中的第n个字节进行理解，第二字节可以参阅b组中的第n+1个字节进行理解，第一字节总体来说是采用部分字节编码函数的字节，第二字节是采用全字节编码函数的字节。依赖文件块和所述依赖校验码块是恢复丢失文件块所需依赖的存活文件块和校验码块，例如：恢复第一存储节点N1中的目标文件的文件块需要依赖N2、N3和N4中目标文件的相关文件块和N11中目标文件的校验码块。

[0142] 图6对应的实施例或任一可选实施例中可以参阅图1至图5部分的相关描述进行理解，本处不做过多赘述。

[0143] 参阅图7，本发明实施例提供的数据存储的方法的一实施例包括：

[0144] 201、数据处理装置接收所述命名节点发送的多个目标存储节点的标识和目标文件的标识，所述多个目标存储节点为已存储了所述目标文件的不同文件块的第一存储节点，所述数据处理装置应用于分布式存储系统，所述分布式存储系统包括命名节点、多个第一存储节点和多个第二存储节点，所述多个第一存储节点用于分布式存储待存储文件的不同文件块，所述多个第二存储节点用于分布式存储对所述不同文件块进行编码得到的校验码块，每个第二存储节点都包括数据处理装置，每个数据处理装置都与所述命名节点通信连接。

[0145] 202、所述数据处理装置根据所述目标存储节点的标识和部分字节编码函数对所述目标文件中的部分文件块进行编码，得到第一校验码，所述部分字节编码函数为采用所述目标文件中的部分文件块进行编码得到编码结果的函数。

[0146] 203、所述数据处理装置根据所述目标存储节点的标识和全字节编码函数对所述目标文件中的每个文件块进行编码，得到第二校验码，所述全字节编码函数为采用所述目标文件中的每个文件块进行编码得到编码结果的函数。

[0147] 204、所述数据处理装置将所述第一校验码和所述第二校验码存储到所述数据处理装置所属的第二存储节点的存储空间中。

[0148] 与现有技术相比，本发明实施例提供的数据存储的方法，采用部分字节编码和全字节编码混合的方式，提高了字节可靠性的同时，降低了存储空间，提高了编码效率。同时，在数据恢复时，针对部分字节编码函数编码的字节，不需要获取每个字节进行解码，降低了网络数据恢复时的网络开销。

[0149] 可选地，在上述图7对应的实施例的基础上，本发明实施例提供的数据存储的方法的第一个可选实施例中，所述数据处理装置根据所述目标存储节点的标识和部分字节编码函数对所述目标文件中的部分文件块进行编码，得到第一校验码，可以包括：

[0150] 所述数据处理装置从第一编码参数所对应的目标存储节点中获取所述第一编码

参数所指示的字节，所述第一编码参数为所述部分字节编码函数中的每个编码参数；

[0151] 所述数据处理装置根据所述部分字节编码函数对所述第一编码参数所指示的字节进行编码，得到第一校验码。

[0152] 可选地，在上述图7对应的实施例或第一个可选实施例的基础上，本发明实施例提供的数据存储的方法的第二个可选实施例中，所述数据处理装置根据所述目标存储节点的标识和全字节编码函数对所述目标文件中的每个文件块进行编码，得到第二校验码，可以包括：

[0153] 所述数据处理装置从第二编码参数所对应的目标存储节点中获取所述第二编码参数所指示的字节，所述第二编码参数为所述全字节编码函数中的每个编码参数；

[0154] 所述数据处理装置根据所述全字节编码函数对所述第二编码参数所指示的字节进行编码，得到第二校验码。

[0155] 可选地，在上述图7对应的实施例或第一个可选实施例的基础上，本发明实施例提供的数据存储的方法的第三个可选实施例中，所述数据处理装置根据所述目标存储节点的标识和部分字节编码函数对所述目标文件中的部分文件块进行编码，得到第一校验码之前，所述方法还可以包括：

[0156] 所述数据处理装置根据所述目标存储节点的数量和所述命名节点所指定的校验节点的数量，确定所述部分字节编码函数中第一参数的数量和紧邻的两个校验节点中的部分字节编码函数所包含的相同第一参数的个数，所述紧邻的两个校验节点所包含的部分字节编码函数中第一参数的重叠个数最多。

[0157] 图7对应的实施例或可选实施例中，第一编码参数指的是部分字节编码函数中的参数，例如：图3场景示例中的g_1函数中的a_1,a_2,a_3,a_4，第二编码参数指的是全字节编码函数中的参数，例如：f函数中的b_1,b_2,…,b_9,b_{10}，第一校验码可以参阅c_{11}进行理解，第二校验码可以参阅c_{21}进行理解，第一校验码总体来说是采用部分字节编码函数进行编码得到的校验码，第二字节是采用全字节编码函数进行编码得到的校验码。

[0158] 图7对应的实施例或任一可选实施例中可以参阅图1至图5部分的相关描述进行理解，本处不做过多赘述。

[0159] 参阅图8，本发明实施例提供的数据处理装置30的一实施例包括：数据处理装置30应用于分布式存储系统，所述分布式存储系统包括命名节点、多个第一存储节点和多个第二存储节点，所述多个第一存储节点用于分布式存储已存储文件的不同文件块，所述多个第二存储节点用于分布式存储对所述不同文件块进行编码得到的校验码块，每个第一存储节点都包括所述数据处理装置，每个数据处理装置都与所述命名节点通信连接，所述数据处理装置30包括：

[0160] 接收模块301，用于接收用户设备发送的文件块获取请求，所述文件块获取请求携带所述目标文件的标识；

[0161] 确定模块302，用于根据所述接收模块301接收的所述目标文件的标识未查找到所述目标文件，确定所述目标文件块丢失；

[0162] 获取模块303，用于在所述确定模块302确定所述目标文件丢失后，从所述命名节点获取恢复依赖数据块所在的目标存储节点的标识，并根据所述目标存储节点的标识和所述目标文件的标识获取所述恢复依赖数据块，所述恢复依赖数据块包括恢复所述目标文件

块所需的依赖文件块和依赖校验码块，所述依赖校验码块中的一部分校验码为对目标文件的部分文件块编码得到，所述依赖校验码块中的剩余部分校验码为对所述目标文件的每个文件块编码得到，所述剩余部分校验码为除所述一部分校验码之外的校验码，所述目标文件为所述目标文件块所属的文件；

[0163] 恢复模块304，用于根据所述获取模块303获取的所述依赖文件块和所述依赖校验码块恢复所述目标文件块。

[0164] 与现有技术相比，本发明实施例提供的数据处理装置，在数据恢复时，针对部分字节编码函数编码的字节，不需要获取每个字节进行解码，降低了网络数据恢复时的网络开销。

[0165] 可选地，在上述图8对应的实施例的基础上，参阅图9，本发明实施例提供的数据处理装置30的第一个可选实施例中，所述恢复模块304包括：

[0166] 第一恢复单元3041，用于根据部分字节编码函数、所述依赖文件块和所述依赖校验码块恢复所述目标文件块中的第一字节，所述部分字节编码函数为采用所述目标文件中的部分文件块进行编码得到编码结果的函数；

[0167] 第二恢复单元3042，用于根据全字节编码函数、所述依赖文件块和所述依赖校验码块恢复所述目标文件块中的第二字节，所述全字节编码函数为采用所述目标文件中的每个文件块进行编码得到编码结果的函数。

[0168] 可选地，在上述图9对应的实施例的基础上，本发明实施例提供的数据处理装置30的第二个可选实施例中，

[0169] 所述第一恢复单元3041，具体用于从第一编码参数所对应的依赖文件块中获取恢复所述第一字节所需的依赖字节，从第一编码结果所对应的依赖校验码块中获取恢复所述第一字节所需的校验码，所述第一编码参数为所述部分字节编码函数中的编码参数，所述第一编码结果为采用所述部分字节编码函数对所述第一编码参数所指示的依赖字节和所述第一字节进行编码所得到的结果；根据恢复所述第一字节所需的依赖字节，对恢复所述第一字节所需的校验码进行解码，得到所述第一字节。

[0170] 可选地，在上述图9对应的实施例的基础上，本发明实施例提供的数据处理装置30的第三个可选实施例中，

[0171] 所述第二恢复单元3042，具体用于从第二编码参数所对应的依赖文件块中获取恢复所述第二字节所需的依赖字节，从第二编码结果所对应的依赖校验码块中获取恢复所述第二字节所需的校验码，所述第二编码参数为所述全字节编码函数中的编码参数，所述第二编码结果为采用所述全字节编码函数对所述第二编码参数所指示的依赖字节和所述第二字节进行编码所得到的结果；根据恢复所述第二字节所需的依赖字节，对恢复所述第二字节所需的校验码进行解码，得到所述第二字节。

[0172] 图8或图9对应的实施例或可选实施例中，第一编码参数指的是部分字节编码函数中的参数，例如：图3场景示例中的g_1函数中的a_1，a_2，a_3，a_4，第二编码参数指的是全字节编码函数中的参数，例如：f函数中的b_1，b_2，…，b_9，b_{10}，第一字节可以参阅a组中的第n个字节进行理解，第二字节可以参阅b组中的第n+1个字节进行理解，第一字节总体来说是采用部分字节编码函数的字节，第二字节是采用全字节编码函数的字节。依赖文件块和所述依赖校验码块是恢复丢失文件块所需依赖的存活文件块和校验码块，例如：恢复第一存储节点N1中

的目标文件的文件块需要依赖N2、N3和N4中目标文件的相关文件块和N11中目标文件的校验码块。

[0173] 图8或图9对应的实施例或任一可选实施例中可以参阅图1至图6部分的相关描述进行理解，本处不做过多赘述。

[0174] 参阅图10，本发明实施例提供的数据处理装置40的一实施例包括：数据处理装置40应用于分布式存储系统，所述分布式存储系统包括命名节点、多个第一存储节点和多个第二存储节点，所述多个第一存储节点用于分布式存储待存储文件的不同文件块，所述多个第二存储节点用于分布式存储对所述不同文件块进行编码得到的校验码块，每个第二存储节点都包括数据处理装置，每个数据处理装置都与所述命名节点通信连接，所述数据处理装置包括：

[0175] 接收模块401，用于接收所述命名节点发送的多个目标存储节点的标识和目标文件的标识，所述多个目标存储节点为已存储了所述目标文件的不同文件块的第一存储节点；

[0176] 第一编码模块402，用于根据所述接收模块401接收的所述目标存储节点的标识和部分字节编码函数对所述目标文件中的部分文件块进行编码，得到第一校验码，所述部分字节编码函数为采用所述目标文件中的部分文件块进行编码得到编码结果的函数；

[0177] 第二编码模块403，用于根据所述接收模块401接收的所述目标存储节点的标识和全字节编码函数对所述目标文件中的每个文件块进行编码，得到第二校验码，所述全字节编码函数为采用所述目标文件中的每个文件块进行编码得到编码结果的函数；

[0178] 存储调度模块404，用于将所述第一编码模块402编码得到的所述第一校验码和所述第二编码模块403编码得到的所述第二校验码存储到所述数据处理装置所属的第二存储节点的存储空间中。

[0179] 与现有技术相比，本发明实施例提供的数据处理装置，采用部分字节编码和全字节编码混合的方式，提高了字节可靠性的同时，降低了存储空间，提高了编码效率。同时，在数据恢复时，针对部分字节编码函数编码的字节，不需要获取每个字节进行解码，降低了网络数据恢复时的网络开销。

[0180] 可选地，在上述图10对应的实施例的基础上，本发明实施例提供的数据处理装置40的第一个可选实施例中，

[0181] 所述第一编码模块402，具体用于从第一编码参数所对应的目标存储节点中获取所述第一编码参数所指示的字节，所述第一编码参数为所述部分字节编码函数中的每个编码参数；根据所述部分字节编码函数对所述第一编码参数所指示的字节进行编码，得到第一校验码。

[0182] 可选地，在上述图10对应的实施例或第一个可选实施例的基础上，本发明实施例提供的数据处理装置40的第二个可选实施例中，

[0183] 所述第二编码模块，具体用于从第二编码参数所对应的目标存储节点中获取所述第二编码参数所指示的字节，所述第二编码参数为所述全字节编码函数中的每个编码参数；根据所述全字节编码函数对所述第二编码参数所指示的字节进行编码，得到第二校验码。

[0184] 可选地，在上述图10对应的实施例或第一个可选实施例的基础上，参阅图11，本发

明实施例提供的数据处理装置40的第三个可选实施例中，所述数据处理装置40还包括：

[0185] 确定模块405，用于根据所述接收模块401接收的所述目标存储节点的数量和所述命名节点所指定的校验节点的数量，确定所述部分字节编码函数中第一参数的数量和紧邻的两个校验节点中的部分字节编码函数所包含的相同第一参数的个数，所述紧邻的两个校验节点所包含的部分字节编码函数中第一参数的重叠个数最多。

[0186] 图10和图11对应的实施例或可选实施例中，第一编码参数指的是部分字节编码函数中的参数，例如：图3场景示例中的g_1函数中的a_1,a_2,a_3,a_4，第二编码参数指的是全字节编码函数中的参数，例如：f函数中的b_1,b_2,…,b_9,b_{10}，第一校验码可以参阅c_{11}进行理解，第二校验码可以参阅c_{21}进行理解，第一校验码总体来说是采用部分字节编码函数进行编码得到的校验码，第二字节是采用全字节编码函数进行编码得到的校验码。

[0187] 图10和图11对应的实施例或任一可选实施例中可以参阅图1至图5、图7部分的相关描述进行理解，本处不做过多赘述。

[0188] 在上述数据处理装置的多个实施例中，应当理解的是，在一种实现方式下，接收模块、获取模块可以是由输入/输出I/O设备(比如网卡)来实现，确定模块、恢复模块、第一编码模块、第二编码模块、存储调度模块可以由处理器执行存储器中的程序或指令来实现的(换言之，即由处理器以及与所述处理器耦合的存储器中的特殊指令相互配合来实现)；在另一种实现方式下接收模块、获取模块可以是由输入/输出I/O设备(比如网卡)来实现，确定模块、恢复模块、第一编码模块、第二编码模块、存储调度模块也可以分别通过专有电路来实现，具体实现方式参见现有技术，这里不再赘述；在再一种实现方式下，接收模块、获取模块可以是由输入/输出I/O设备(比如网卡)来实现，确定模块、恢复模块、第一编码模块、第二编码模块、存储调度模块也可以通过现场可编程门阵列(FPGA，Field－Programmable Gate Array)来实现，具体实现方式参见现有技术，这里不再赘述，本发明包括但不限于前述实现方式，应当理解的是，只要按照本发明的思想实现的方案，都落入本发明实施例所保护的范围。

[0189] 本实施例提供了一种数据处理装置的硬件结构，参见图12所示，一种数据处理装置的硬件结构可以包括：

[0190] 收发器件、软件器件以及硬件器件三部分；

[0191] 收发器件为用于完成包收发的硬件电路；

[0192] 硬件器件也可称“硬件处理模块”，或者更简单的，也可简称为“硬件”，硬件器件主要包括基于FPGA、ASIC之类专用硬件电路(也会配合其他配套器件，如存储器)来实现某些特定功能的硬件电路，其处理速度相比通用处理器往往要快很多，但功能一经定制，便很难更改，因此，实现起来并不灵活，通常用来处理一些固定的功能。需要说明的是，硬件器件在实际应用中，也可以包括MCU(微处理器，如单片机)、或者CPU等处理器，但这些处理器的主要功能并不是完成大数据的处理，而主要用于进行一些控制，在这种应用场景下，由这些器件搭配的系统为硬件器件。

[0193] 软件器件(或者也简单“软件”)主要包括通用的处理器(例如CPU)及其一些配套的器件(如内存、硬盘等存储设备)，可以通过编程来让处理器具备相应的处理功能，用软件来实现时，可以根据业务灵活配置，但往往速度相比硬件器件来说要慢。软件处理完后，可以通过硬件器件将处理完的数据通过收发器件进行发送，也可以通过一个与收发器件相连的

接口向收发器件发送处理完的数据。

[0194] 本实施例中，收发器件用于接收目标文件的标识、目标存储节点的标识等。

[0195] 硬件器件及软件器件的其他功能在前述实施例中已经详细论述，这里不再赘述。

[0196] 下面结合附图就接收模块、获取模块可以是由输入/输出I/O设备(比如网卡)来实现，确定模块、恢复模块、第一编码模块、第二编码模块、存储调度模块可以是可以由处理器执行存储器中的程序或指令来实现的技术方案来做详细的介绍：

[0197] 图13是本发明实施例提供的数据处理装置30的结构示意图。数据处理装置30应用于分布式存储系统，所述分布式存储系统包括命名节点、多个第一存储节点和多个第二存储节点，所述多个第一存储节点用于分布式存储已存储文件的不同文件块，所述多个第二存储节点用于分布式存储对所述不同文件块进行编码得到的校验码块，每个第一存储节点都包括数据处理装置，每个数据处理装置都与所述命名节点通信连接。所述数据处理装置30包括处理器310、存储器350和输入/输出I/O设备330，存储器350可以包括只读存储器和随机存取存储器，并向处理器310提供操作指令和数据。存储器350的一部分还可以包括非易失性随机存取存储器(NVRAM)。

[0198] 在一些实施方式中，存储器350存储了如下的元素，可执行模块或者数据结构，或者他们的子集，或者他们的扩展集:

[0199] 当数据处理装置30为源设备时：

[0200] 在本发明实施例中，通过调用存储器350存储的操作指令(该操作指令可存储在操作系统中)，

[0201] 通过I/O设备330接收用户设备发送的文件块获取请求，所述文件块获取请求携带所述目标文件的标识；

[0202] 通过I/O设备330从所述命名节点获取恢复依赖数据块所在的目标存储节点的标识，并根据所述目标存储节点的标识和所述目标文件的标识获取所述恢复依赖数据块，所述恢复依赖数据块包括恢复所述目标文件块所需的依赖文件块和依赖校验码块，所述依赖校验码块中的一部分校验码为对目标文件的部分文件块编码得到，所述依赖校验码块中的剩余部分校验码为对所述目标文件的每个文件块编码得到，所述剩余部分校验码为除所述一部分校验码之外的校验码，所述目标文件为所述目标文件块所属的文件；

[0203] 根据所述依赖文件块和所述依赖校验码块恢复所述目标文件块。

[0204] 与现有技术中数据无法同时兼顾数据存储开销和数据恢复时的网络开销相比，本发明实施例提供的数据处理装置，校验码块是通过部分字节编码和全字节编码结合得到的，降低了存储开销，数据恢复时目标文件块中的一部分只需要依赖部分依赖文件块就可以得到降低了数据恢复时的网络开销。

[0205] 处理器310控制数据处理装置30的操作，处理器310还可以称为CPU(Central Processing Unit，中央处理单元)。存储器350可以包括只读存储器和随机存取存储器，并向处理器310提供指令和数据。存储器350的一部分还可以包括非易失性随机存取存储器(NVRAM)。具体的应用中数据处理装置30的各个组件通过总线系统320耦合在一起，其中总线系统320除包括数据总线之外，还可以包括电源总线、控制总线和状态信号总线等。但是为了清楚说明起见，在图中将各种总线都标为总线系统320。

[0206] 上述本发明实施例揭示的方法可以应用于处理器310中，或者由处理器310实现。

处理器310可能是一种集成电路芯片，具有信号的处理能力。在实现过程中，上述方法的各步骤可以通过处理器310中的硬件的集成逻辑电路或者软件形式的指令完成。上述的处理器310可以是通用处理器、数字信号处理器(DSP)、专用集成电路(ASIC)、现成可编程门阵列(FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。可以实现或者执行本发明实施例中的公开的各方法、步骤及逻辑框图。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。结合本发明实施例所公开的方法的步骤可以直接体现为硬件译码处理器执行完成，或者用译码处理器中的硬件及软件模块组合执行完成。软件模块可以位于随机存储器，闪存、只读存储器，可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器350，处理器310读取存储器350中的信息，结合其硬件完成上述方法的步骤。

[0207] 可选地，处理器310具体用于：

[0208] 根据部分字节编码函数、所述依赖文件块和所述依赖校验码块恢复所述目标文件块中的第一字节，所述部分字节编码函数为采用所述目标文件中的部分文件块进行编码得到编码结果的函数；

[0209] 根据全字节编码函数、所述依赖文件块和所述依赖校验码块恢复所述目标文件块中的第二字节，所述全字节编码函数为采用所述目标文件中的每个文件块进行编码得到编码结果的函数。

[0210] 可选地，处理器310具体用于：

[0211] 从第一编码参数所对应的依赖文件块中获取恢复所述第一字节所需的依赖字节，从第一编码结果所对应的依赖校验码块中获取恢复所述第一字节所需的校验码，所述第一编码参数为所述部分字节编码函数中的编码参数，所述第一编码结果为采用所述部分字节编码函数对所述第一编码参数所指示的依赖字节和所述第一字节进行编码所得到的结果；

[0212] 根据恢复所述第一字节所需的依赖字节，对恢复所述第一字节所需的校验码进行解码，得到所述第一字节。

[0213] 可选地，处理器310具体用于：

[0214] 从第二编码参数所对应的依赖文件块中获取恢复所述第二字节所需的依赖字节，从第二编码结果所对应的依赖校验码块中获取恢复所述第二字节所需的校验码，所述第二编码参数为所述全字节编码函数中的编码参数，所述第二编码结果为采用所述全字节编码函数对所述第二编码参数所指示的依赖字节和所述第二字节进行编码所得到的结果；

[0215] 根据恢复所述第二字节所需的依赖字节，对恢复所述第二字节所需的校验码进行解码，得到所述第二字节。

[0216] 图13对应的实施例或可选实施例中，第一编码参数指的是部分字节编码函数中的参数，例如：图3场景示例中的g_1函数中的a_1,a_2,a_3,a_4，第二编码参数指的是全字节编码函数中的参数，例如：f函数中的b_1,b_2,…,b_9,b_{10}，第一字节可以参阅a组中的第n个字节进行理解，第二字节可以参阅b组中的第n+1个字节进行理解，第一字节总体来说是采用部分字节编码函数的字节，第二字节是采用全字节编码函数的字节。依赖文件块和所述依赖校验码块是恢复丢失文件块所需依赖的存活文件块和校验码块，例如：恢复第一存储节点N1中的目标文件的文件块需要依赖N2、N3和N4中目标文件的相关文件块和N11中目标文件的校验码块。

[0217] 图13对应的实施例或任一可选实施例中可以参阅图1至图5、图6、图8、图9部分的相关描述进行理解，本处不做过多赘述。

[0218] 图14是本发明实施例提供的数据处理装置40的结构示意图。数据处理装置40应用于分布式存储系统，所述分布式存储系统包括命名节点、多个第一存储节点和多个第二存储节点，所述多个第一存储节点用于分布式存储待存储文件的不同文件块，所述多个第二存储节点用于分布式存储对所述不同文件块进行编码得到的校验码块，每个第二存储节点都包括数据处理装置，每个数据处理装置都与所述命名节点通信连接。所述数据处理装置40包括处理器410、存储器450和输入/输出I/O设备430，存储器450可以包括只读存储器和随机存取存储器，并向处理器410提供操作指令和数据。存储器450的一部分还可以包括非易失性随机存取存储器(NVRAM)。

[0219] 在一些实施方式中，存储器450存储了如下的元素，可执行模块或者数据结构，或者他们的子集，或者他们的扩展集:

[0220] 当数据处理装置40为源设备时：

[0221] 在本发明实施例中，通过调用存储器450存储的操作指令(该操作指令可存储在操作系统中)，

[0222] 通过I/O设备430接收所述命名节点发送的多个目标存储节点的标识和目标文件的标识，所述多个目标存储节点为已存储了所述目标文件的不同文件块的第一存储节点；

[0223] 根据所述目标存储节点的标识和部分字节编码函数对所述目标文件中的部分文件块进行编码，得到第一校验码，所述部分字节编码函数为采用所述目标文件中的部分文件块进行编码得到编码结果的函数；

[0224] 根据所述目标存储节点的标识和全字节编码函数对所述目标文件中的每个文件块进行编码，得到第二校验码，所述全字节编码函数为采用所述目标文件中的每个文件块进行编码得到编码结果的函数；

[0225] 将所述第一校验码和所述第二校验码存储到所述数据处理装置所属的第二存储节点的存储空间中。

[0226] 与现有技术相比，本发明实施例提供的数据存储的方法，采用部分字节编码和全字节编码混合的方式，提高了字节可靠性的同时，降低了存储空间，提高了编码效率。同时，在数据恢复时，针对部分字节编码函数编码的字节，不需要获取每个字节进行解码，降低了网络数据恢复时的网络开销。

[0227] 处理器410控制数据处理装置40的操作，处理器410还可以称为CPU(Central Processing Unit，中央处理单元)。存储器450可以包括只读存储器和随机存取存储器，并向处理器410提供指令和数据。存储器450的一部分还可以包括非易失性随机存取存储器(NVRAM)。具体的应用中数据处理装置40的各个组件通过总线系统420耦合在一起，其中总线系统420除包括数据总线之外，还可以包括电源总线、控制总线和状态信号总线等。但是为了清楚说明起见，在图中将各种总线都标为总线系统420。

[0228] 上述本发明实施例揭示的方法可以应用于处理器410中，或者由处理器410实现。处理器410可能是一种集成电路芯片，具有信号的处理能力。在实现过程中，上述方法的各步骤可以通过处理器410中的硬件的集成逻辑电路或者软件形式的指令完成。上述的处理器410可以是通用处理器、数字信号处理器(DSP)、专用集成电路(ASIC)、现成可编程门阵列

(FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。可以实现或者执行本发明实施例中的公开的各方法、步骤及逻辑框图。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。结合本发明实施例所公开的方法的步骤可以直接体现为硬件译码处理器执行完成，或者用译码处理器中的硬件及软件模块组合执行完成。软件模块可以位于随机存储器，闪存、只读存储器，可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器450，处理器410读取存储器450中的信息，结合其硬件完成上述方法的步骤。

[0229] 可选地，处理器410具体用于：

[0230] 从第一编码参数所对应的目标存储节点中获取所述第一编码参数所指示的字节，所述第一编码参数为所述部分字节编码函数中的每个编码参数；

[0231] 根据所述部分字节编码函数对所述第一编码参数所指示的字节进行编码，得到第一校验码。

[0232] 可选地，处理器410具体用于：

[0233] 从第二编码参数所对应的目标存储节点中获取所述第二编码参数所指示的字节，所述第二编码参数为所述全字节编码函数中的每个编码参数；

[0234] 根据所述全字节编码函数对所述第二编码参数所指示的字节进行编码，得到第二校验码。

[0235] 可选地，处理器410具体用于：

[0236] 根据所述目标存储节点的数量和所述命名节点所指定的校验节点的数量，确定所述部分字节编码函数中第一参数的数量和紧邻的两个校验节点中的部分字节编码函数所包含的相同第一参数的个数，所述紧邻的两个校验节点所包含的部分字节编码函数中第一参数的重叠个数最多。

[0237] 图13对应的实施例或可选实施例中，第一编码参数指的是部分字节编码函数中的参数，例如：图3场景示例中的g_1函数中的a_1,a_2,a_3,a_4，第二编码参数指的是全字节编码函数中的参数，例如：f函数中的b_1,b_2,…,b_9,b_{10}，第一校验码可以参阅c_{11}进行理解，第二校验码可以参阅c_{21}进行理解，第一校验码总体来说是采用部分字节编码函数进行编码得到的校验码，第二字节是采用全字节编码函数进行编码得到的校验码。

[0238] 图13对应的实施例或任一可选实施例中可以参阅图1至图5、图7、图10、图11部分的相关描述进行理解，本处不做过多赘述。

[0239] 本发明实施例提供的分布式存储系统，包括命名节点、多个第一存储节点和多个第二存储节点，所述多个第一存储节点用于分布式存储已存储文件的不同文件块，所述多个第二存储节点用于分布式存储对所述不同文件块进行编码得到的校验码块，每个第一存储节点都包括第一数据处理装置，每个第二存储节点都包括第二数据处理装置，每个第一数据处理装置和每个第二数据处理装置都与所述命名节点通信连接；

[0240] 第一数据处理装置可以参阅图3部分的描述进行理解，第二数据处理装置可以参阅图5部分的描述进行理解，本处不再重复赘述。

[0241] 与现有技术中数据无法同时兼顾数据存储开销和数据恢复时的网络开销相比，本发明实施例提供的分布式存储系统，校验码块是通过部分字节编码和全字节编码结合得到的，降低了存储开销，数据恢复时目标文件块中的一部分只需要依赖部分依赖文件块就可

以得到降低了数据恢复时的网络开销。

[0242] 本领域普通技术人员可以理解上述实施例的各种方法中的全部或部分步骤是可以通过程序来指令相关的硬件来完成，该程序可以存储于一计算机可读存储介质中，存储介质可以包括:ROM、RAM、磁盘或光盘等。

[0243] 以上对本发明实施例所提供的数据存储的方法、恢复的方法、相应装置以及系统进行了详细介绍，本文中应用了具体个例对本发明的原理及实施方式进行了阐述，以上实施例的说明只是用于帮助理解本发明的方法及其核心思想;同时，对于本领域的一般技术人员，依据本发明的思想，在具体实施方式及应用范围上均会有改变之处，综上所述，本说明书内容不应理解为对本发明的限制。

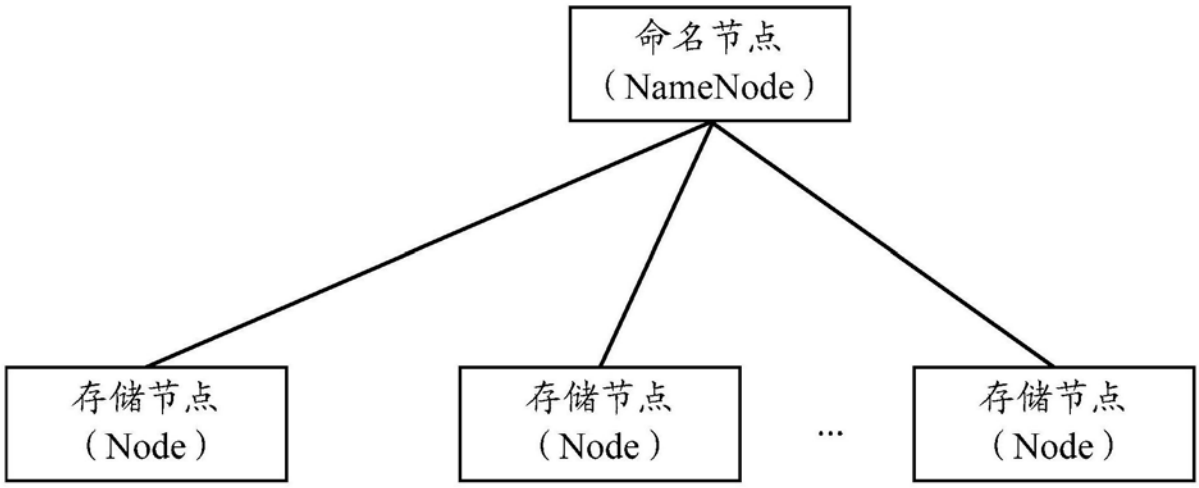

命名节点
（NameNode）
存储节点
（Node）
存储节点
（Node）
...
存储节点
（Node）

图1

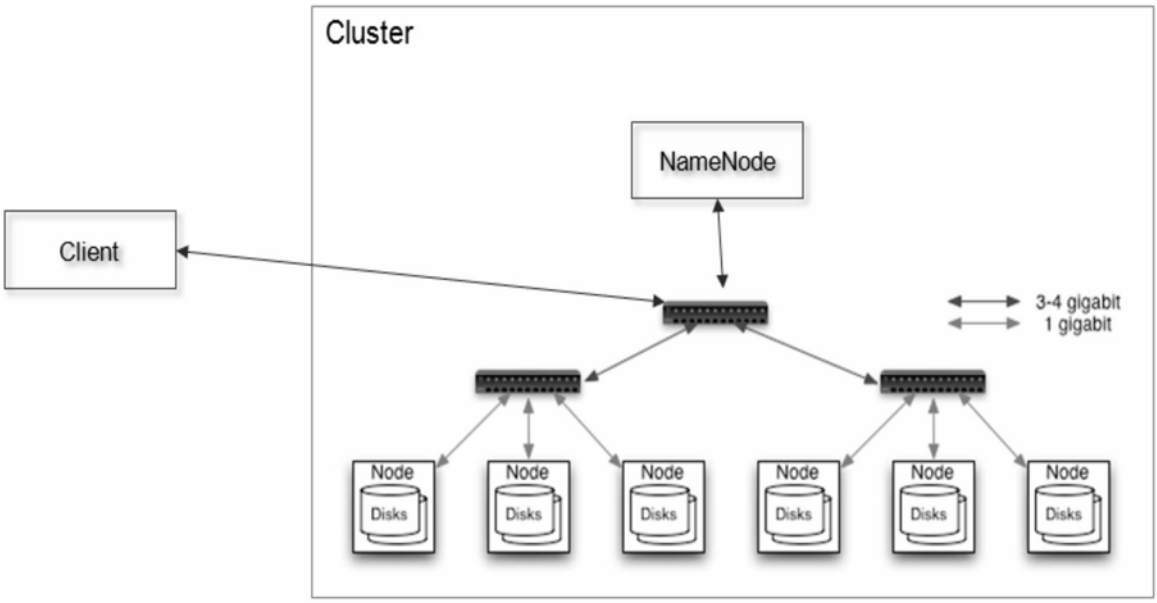

Cluster
NameNode
Client
3-4 gigabit
1 gigabit
Node
Disks
Node
Disks
Node
Disks
Node
Disks
Node
Disks
Node
Disks

图2

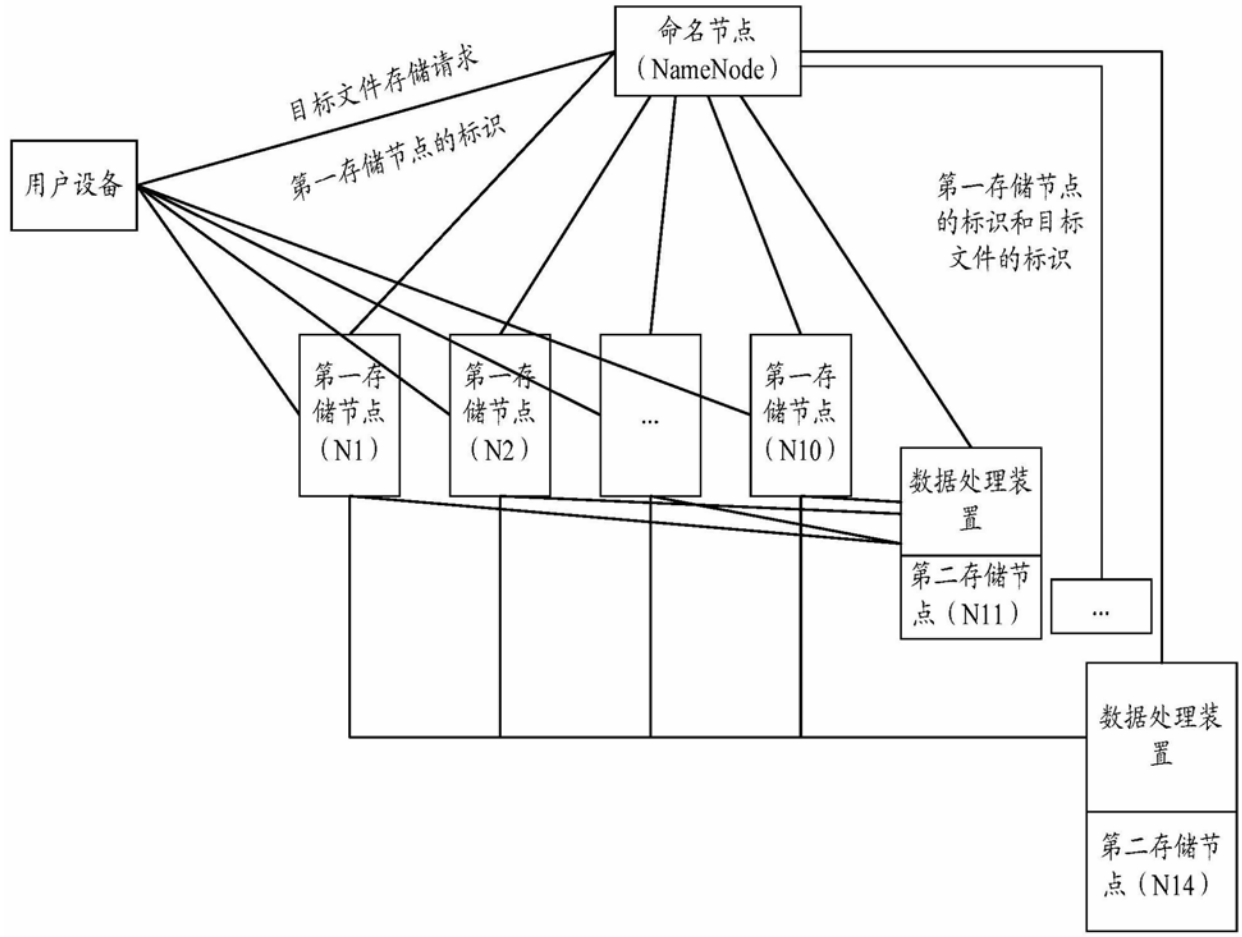

命名节点
（NameNode）
目标文件存储请求
第一存储节点的标识
用户设备
第一存储节点的标识和目标文件的标识
第一存储节点（N1）
第一存储节点（N2）
...
第一存储节点（N10）
数据处理装置
第二存储节点（N11）
...
数据处理装置
第二存储节点（N14）

图3

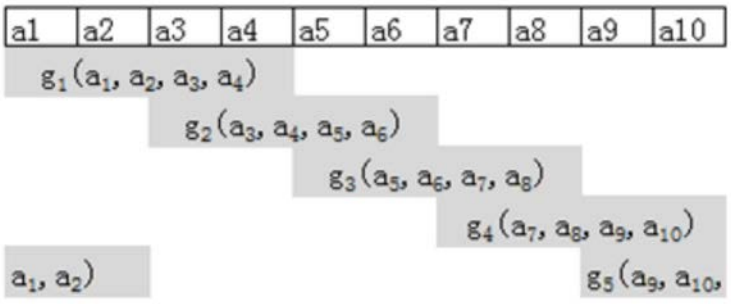

a1
a2
a3
a4
a5
a6
a7
a8
a9
a10
$g_1(a_1, a_2, a_3, a_4)$
$g_2(a_3, a_4, a_5, a_6)$
$g_3(a_5, a_6, a_7, a_8)$
$g_4(a_7, a_8, a_9, a_{10})$
$a_1, a_2)$
$g_5(a_9, a_{10},$

图4

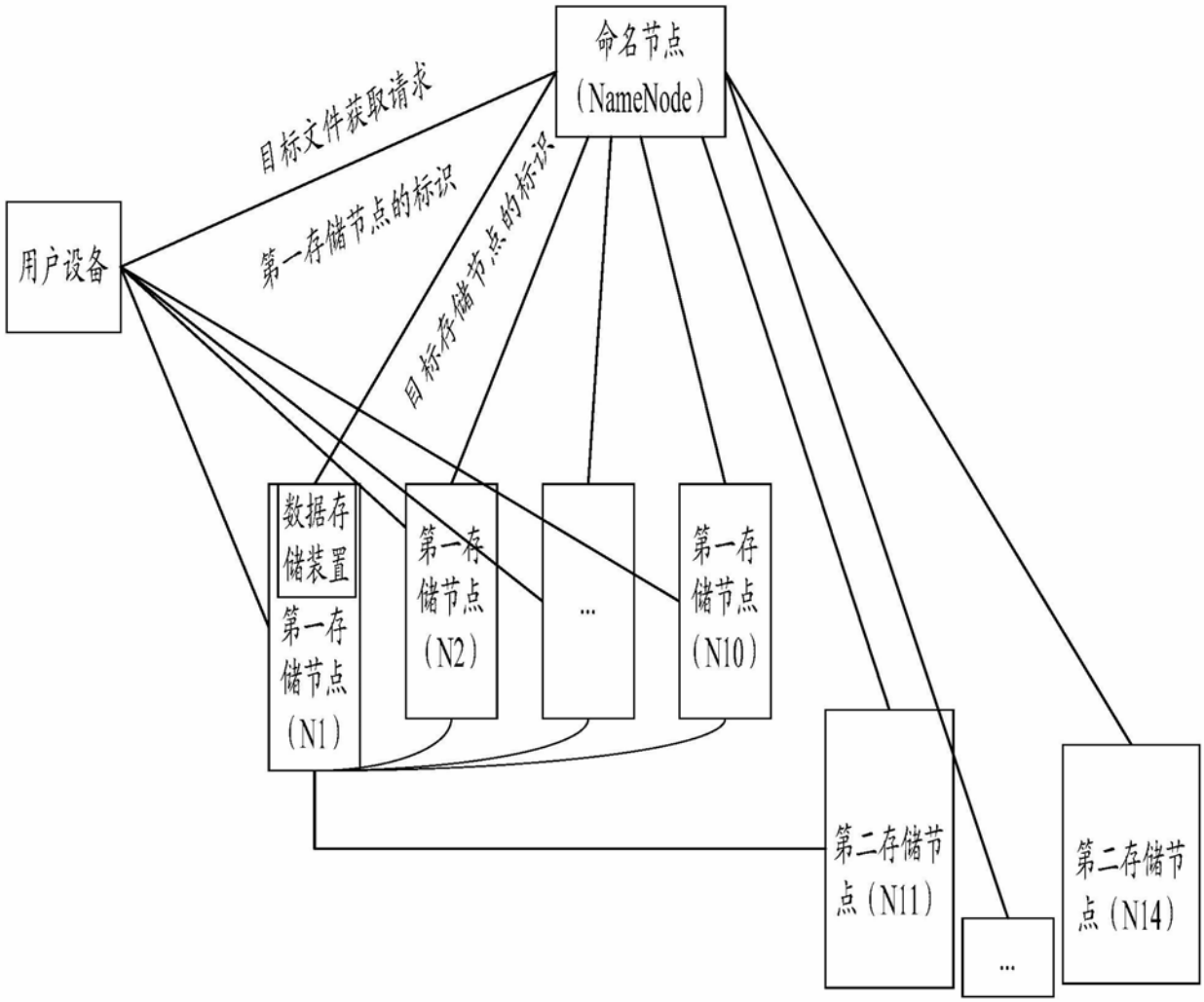

命名节点
(NameNode)
目标文件获取请求
第一存储节点的标识
目标存储节点的标识
用户设备
数据存储装置
第一存储节点(N1)
第一存储节点(N2)
...
第一存储节点(N10)
第二存储节点(N11)
...
第二存储节点(N14)

图5

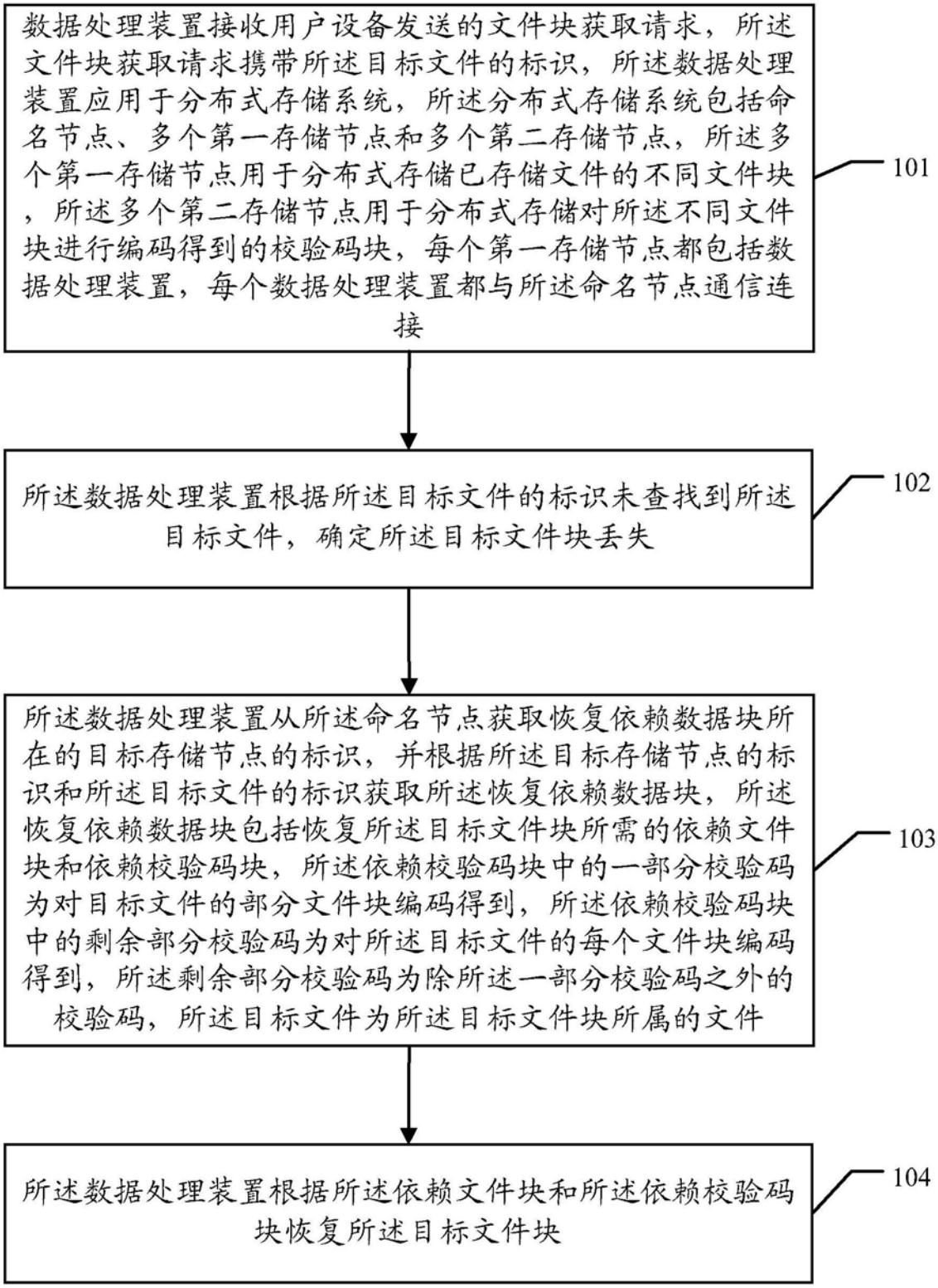

数据处理装置接收用户设备发送的文件块获取请求，所述文件块获取请求携带所述目标文件的标识，所述数据处理装置应用于分布式存储系统，所述分布式存储系统包括命名节点、多个第一存储节点和多个第二存储节点，所述多个第一存储节点用于分布式存储已存储文件的不同文件块，所述多个第二存储节点用于分布式存储对所述不同文件块进行编码得到的校验码块，每个第一存储节点都包括数据处理装置，每个数据处理装置都与所述命名节点通信连接
101
所述数据处理装置根据所述目标文件的标识未查找到所述目标文件，确定所述目标文件块丢失
102
所述数据处理装置从所述命名节点获取恢复依赖数据块所在的目标存储节点的标识，并根据所述目标存储节点的标识和所述目标文件的标识获取所述恢复依赖数据块，所述恢复依赖数据块包括恢复所述目标文件块所需的依赖文件块和依赖校验码块，所述依赖校验码块中的一部分校验码为对目标文件的部分文件块编码得到，所述依赖校验码块中的剩余部分校验码为对所述目标文件的每个文件块编码得到，所述剩余部分校验码为除所述一部分校验码之外的校验码，所述目标文件为所述目标文件块所属的文件
103
所述数据处理装置根据所述依赖文件块和所述依赖校验码块恢复所述目标文件块
104

图6

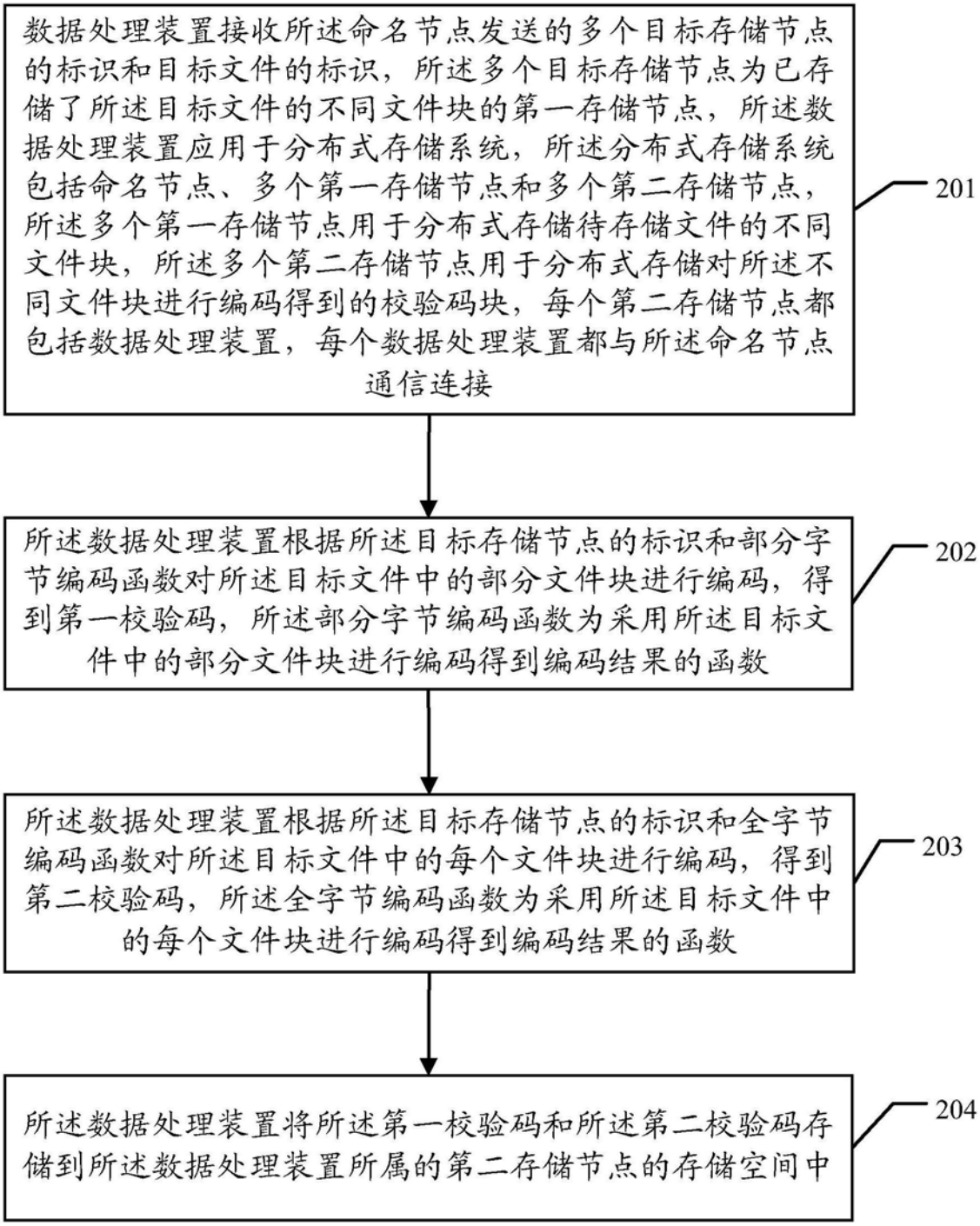

数据处理装置接收所述命名节点发送的多个目标存储节点的标识和目标文件的标识，所述多个目标存储节点为已存储了所述目标文件的不同文件块的第一存储节点，所述数据处理装置应用于分布式存储系统，所述分布式存储系统包括命名节点、多个第一存储节点和多个第二存储节点，所述多个第一存储节点用于分布式存储待存储文件的不同文件块，所述多个第二存储节点用于分布式存储对所述不同文件块进行编码得到的校验码块，每个第二存储节点都包括数据处理装置，每个数据处理装置都与所述命名节点通信连接
201
所述数据处理装置根据所述目标存储节点的标识和部分字节编码函数对所述目标文件中的部分文件块进行编码，得到第一校验码，所述部分字节编码函数为采用所述目标文件中的部分文件块进行编码得到编码结果的函数
202
所述数据处理装置根据所述目标存储节点的标识和全字节编码函数对所述目标文件中的每个文件块进行编码，得到第二校验码，所述全字节编码函数为采用所述目标文件中的每个文件块进行编码得到编码结果的函数
203
所述数据处理装置将所述第一校验码和所述第二校验码存储到所述数据处理装置所属的第二存储节点的存储空间中
204

图7

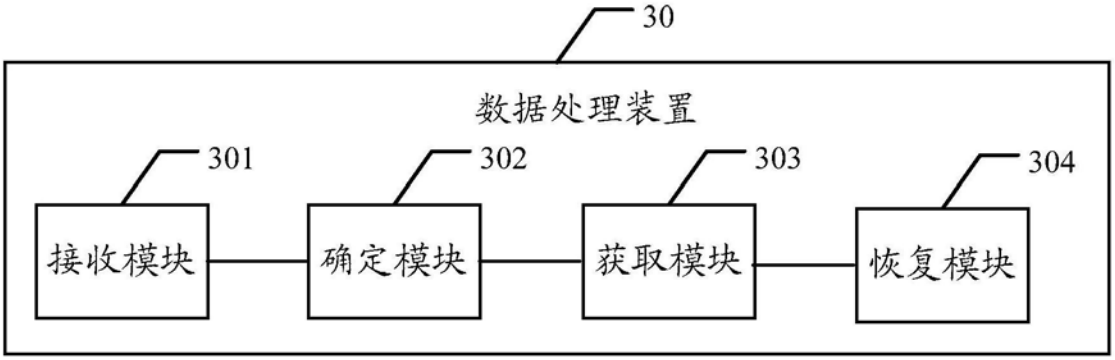

30
数据处理装置
301
接收模块
302
确定模块
303
获取模块
304
恢复模块

图8

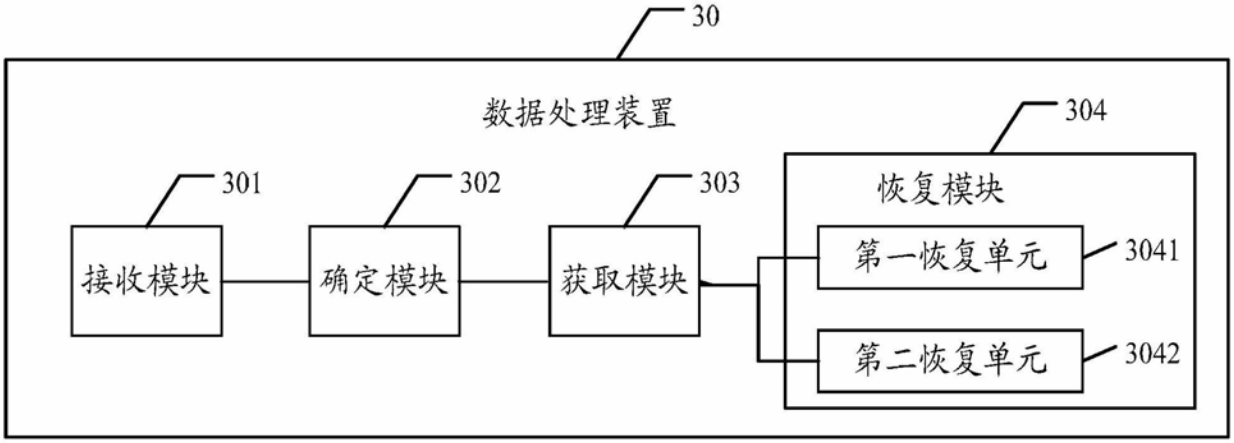

30
数据处理装置
304
恢复模块
301
接收模块
302
确定模块
303
获取模块
第一恢复单元
3041
第二恢复单元
3042

图9

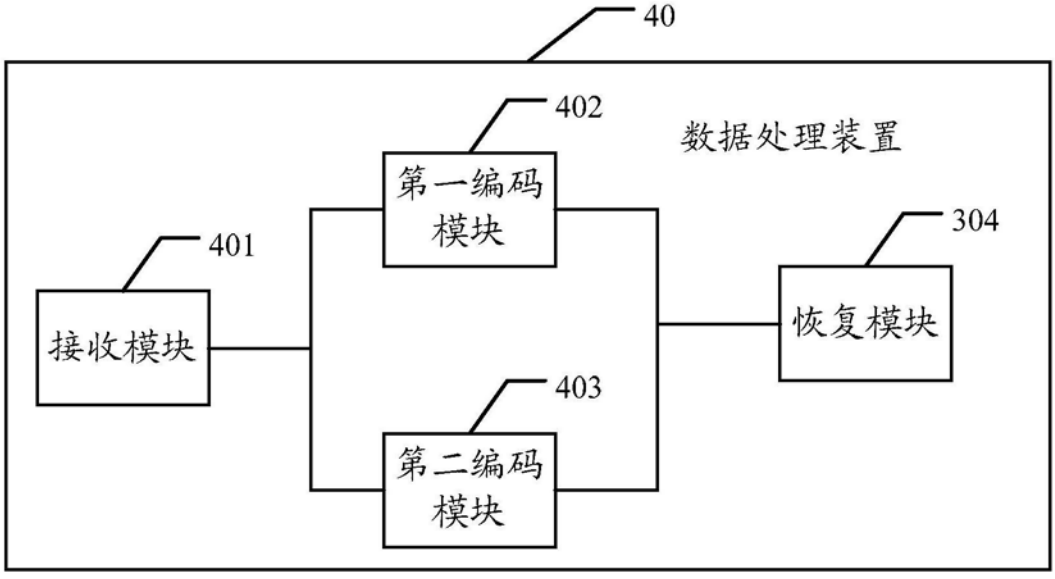

40
402
数据处理装置
第一编码模块
304
401
接收模块
恢复模块
403
第二编码模块

图10

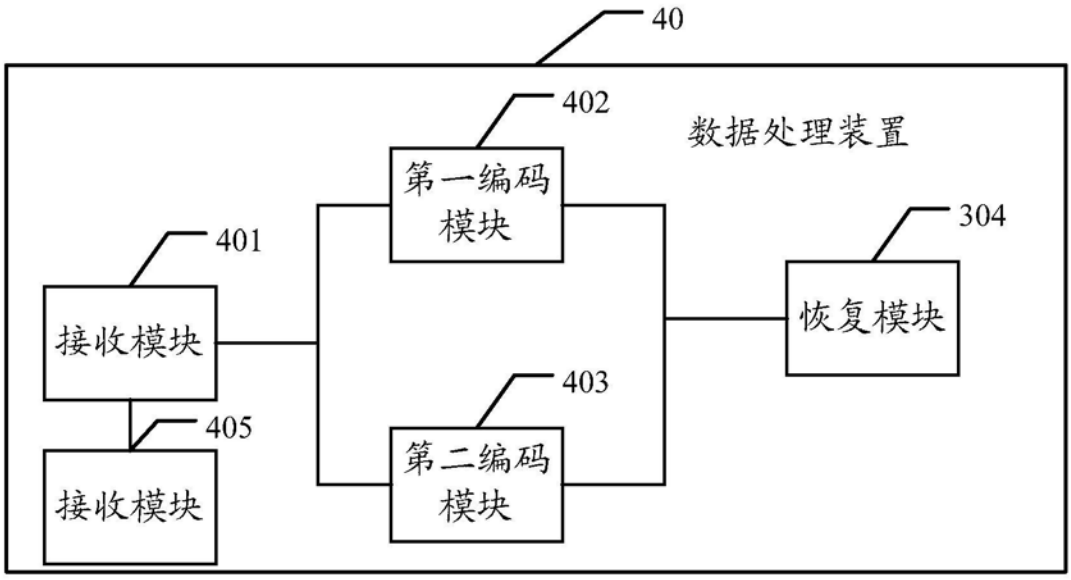

40
数据处理装置
402
第一编码模块
401
接收模块
405
接收模块
403
第二编码模块
304
恢复模块

图11

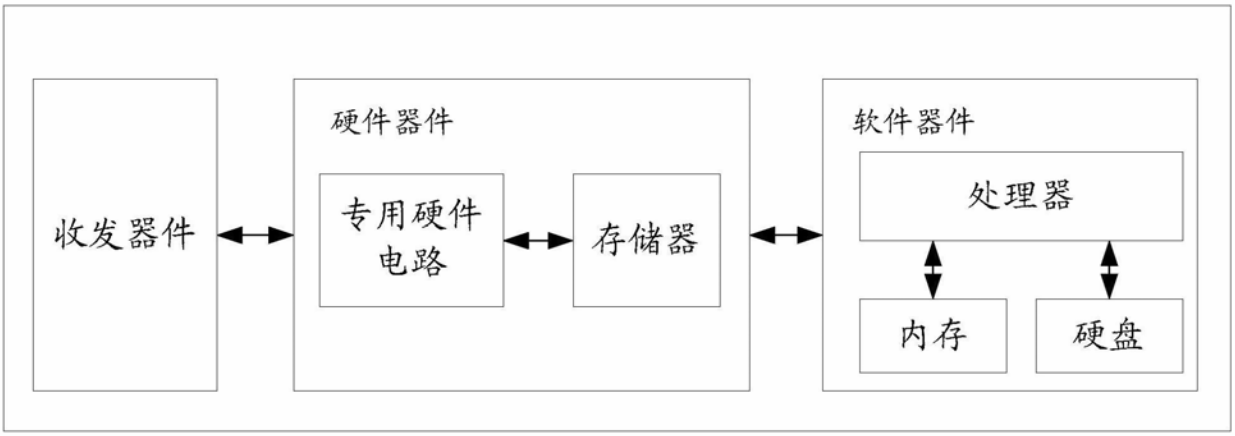

收发器件
硬件器件
专用硬件电路
存储器
软件器件
处理器
内存
硬盘

图12

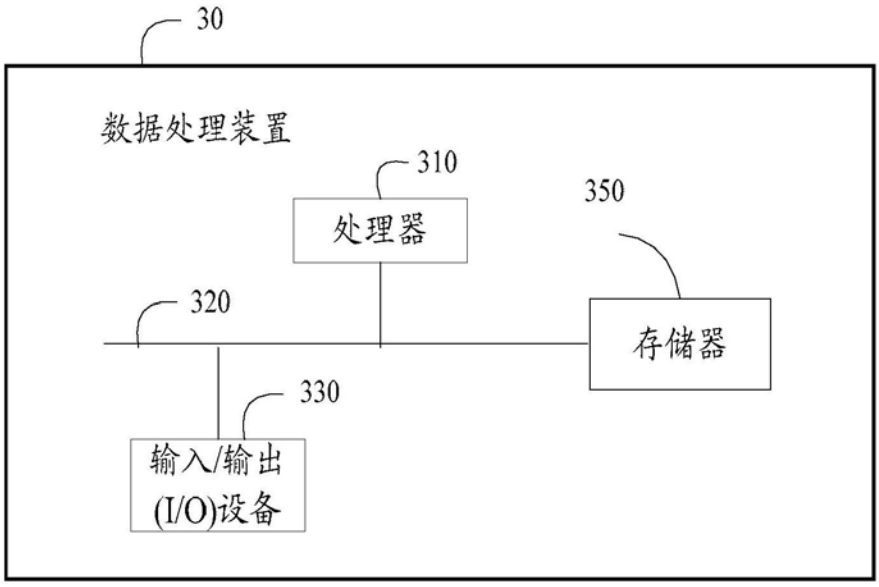

30
数据处理装置
310
处理器
350
320
存储器
330
输入/输出
(I/O)设备

图13

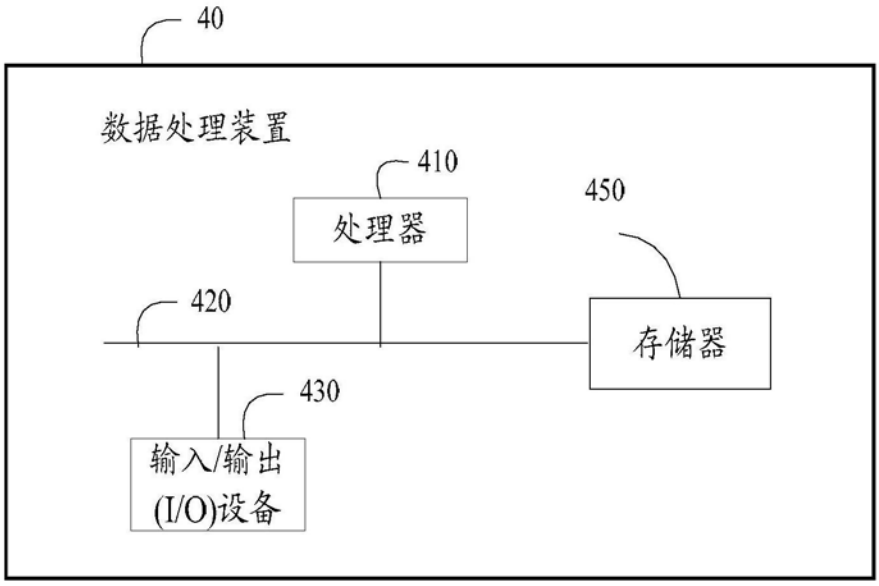

40
数据处理装置
410
处理器
450
420
存储器
430
输入/输出
(I/O)设备

图14