

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局



(10) 国际公布号

WO 2021/027481 A1

(43) 国际公布日
2021 年 2 月 18 日 (18.02.2021)

(51) 国际专利分类号:
H04L 12/24 (2006.01) H04L 29/08 (2006.01)

(21) 国际申请号: PCT/CN2020/102302

(22) 国际申请日: 2020 年 7 月 16 日 (16.07.2020)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910741190.5 2019 年 8 月 12 日 (12.08.2019) CN

(71) 申请人: 华为技术有限公司 (HUAWEI TECHNOLOGIES CO., LTD.) [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(72) 发明人: 陶毅(TAO, Yi); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ,

(54) Title: FAULT PROCESSING METHOD, APPARATUS, COMPUTER DEVICE, STORAGE MEDIUM AND STORAGE SYSTEM

(54) 发明名称: 故障处理方法、装置、计算机设备、存储介质及存储系统

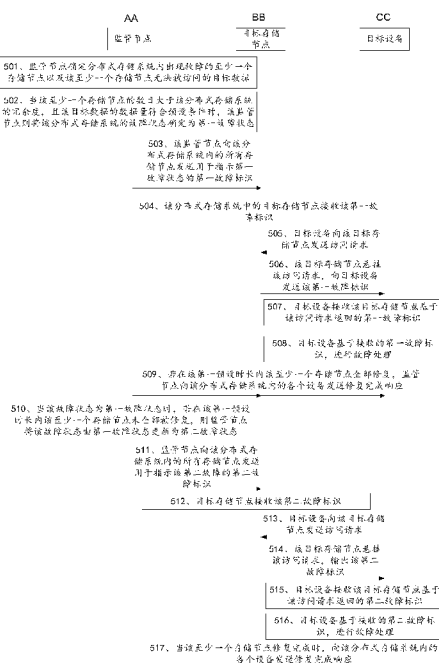


图 5

(57) Abstract: The present invention relates to the technical field of fault processing, and disclosed are a distributed storage system fault processing method, an apparatus, a computer device, a storage medium and a storage system. In the present method, the fault state of the distributed storage system is determined according to at least one storage node that has experienced a fault among multiple storage nodes, and thus there is no need to determine the fault state of the distributed storage system when all storage nodes experience a fault. After the fault state is determined, the fault state can be immediately sent to each storage node within the distributed storage

NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布：

— 包括国际检索报告(条约第21条(3))。

system so that each storage node performs fault processing on the basis of the determined fault state, thus reducing the time used for the distributed storage system to recover to normal.

(57) 摘要：本发明公开了一种分布式存储系统故障处理方法、装置、计算机设备、存储介质及存储系统，属于故障处理技术领域。本方法根据所述多个存储节点中的至少一个出现故障的存储节点，确定所述分布式存储系统的故障状态，从而无需当所有存储节点均故障时，才确定分布式存储系统的故障状态，当确定完故障状态后，可以立即向分布式存储系统内的每个存储节点发送故障状态，以便每个存储节点根据确定的故障状态进行故障处理，从而可以降低分布式存储系统恢复正常所用的时间。

故障处理方法、装置、计算机设备、存储介质及存储系统

技术领域

本发明涉及故障处理技术领域，特别涉及一种分布式存储系统故障处理方法、装置、计算机设备、存储介质及分布式存储系统。

背景技术

随着大数据技术的发展，为了存储更多的数据以及防止数据丢失，分布式存储系统越来越受到企业的青睐，分布式存储系统中的存储节点随着使用时间的增长，不可避免的会出现故障，当分布式存储系统中的存储节点均出现故障时，为了保证故障的存储节点不影响正常的业务，为用户提供业务的计算节点可以对分布式存储系统中的故障，进行故障处理。

其中，故障处理可以是以下过程：在分布式存储系统中，客户端向多个存储节点发送小型计算机系统接口（small computer system interface，SCSI）请求；当分布式存储系统中的所有存储节点均出现故障时，若存储节点出现的是短时间内可修复的故障，则存储节点不会响应 SCSI 请求，当客户端未获取任何存储节点的响应时，则客户端将该分布式存储系统的故障状态确定为全部路径异常（all path down，APD）状态，APD 状态为 VMWare 虚拟机定义的一种存储节点的故障状态，用于表示后端存储节点的所有路径均无法响应主机请求，客户端悬挂未处理的 SCSI 请求，等待技术人员修复分布式存储系统内的故障；当存储节点出现短时间内不可修复的故障时，存储节点向客户端返回存储异常的消息，则客户端接收到该存储异常的消息后，将该分布式存储系统的故障状态确定为永久设备丢失（permanent device lost，PDL）状态，其中，PDL 状态为 VMWare 虚拟机定义的一种存储节点的故障状态，用于表示后端存储节点长期或永久故障，由于存储节点长时间出现故障会导致分布式存储系统内的文件系统损坏，当客户端确定的故障状态为 PDL 状态时，该客户端将文件系统下电，并等待技术人员修复故障。

在上述故障处理过程中，只有当分布式存储系统中所有存储节点均出现故障时，客户端才会执行上述故障处理的过程，但是，当分布式存储系统中的部分存储节点出现故障时，客户端不会执行上述故障处理的过程。然而，在分布式存储系统中部分存储节点出现故障是一种比较常见的现象，一旦部分存储节点出现故障，若不对分布式存储系统中的存储节点进行故障诊断，则客户端并无法确定分布式存储系统中的存储节点是否有故障，从而技术人员不能及时地从客户端获知存储节点出现故障的消息，也就不能立即对出现故障的存储节点进行修复，从而延长了分布式存储系统恢复正常所用的时间。

发明内容

本发明实施例提供了一种分布式存储系统故障处理方法、装置、计算机设备、存储介质及存储系统，能够降低分布式存储系统恢复正常所用的时间。该技术方案如下：

第一方面，提供了一种分布式存储系统故障处理方法，所述分布式存储系统包含多个存储节点；该方法包括：

根据所述多个存储节点中的至少一个出现故障的存储节点，确定所述分布式存储系统的故障状态；所述故障状态用于指示所述至少一个出现故障的存储节点能否在第一预设时长内全部被修复；

向所述多个存储节点中每一个存储节点发送所述故障状态。

基于上述实现方式，根据所述多个存储节点中的至少一个出现故障的存储节点，确定所述分布式存储系统的故障状态，从而无需当所有存储节点均故障时，才确定分布式存储系统的故障状态，当确定完故障状态后，可以立即向分布式存储系统内的每个存储节点发送故障状态，以便每个存储节点根据确定的故障状态进行故障处理，从而可以降低分布式存储系统恢复正常所用的时间。

在一种可能实现方式中，所述方法还包括：

根据所述分布式存储系统的故障状态，进行故障处理。

在一种可能实现方式中，所述根据所述多个存储节点中的至少一个出现故障的存储节点，确定所述分布式存储系统的故障状态包括：

确定所述分布式存储系统内出现故障的至少一个存储节点以及所述至少一个存储节点内无法被访问的目标数据；

根据所述至少一个存储节点以及所述目标数据，确定所述故障状态。

在一种可能实现方式中，所述根据所述至少一个存储节点以及所述目标数据，确定所述故障状态包括：

当所述至少一个存储节点的数目大于所述分布式存储系统的冗余度，且所述目标数据的数据量符合预设条件时，将所述故障状态确定为第一故障状态，所述第一故障状态用于指示所述至少一个存储节点能在所述第一预设时长内全部被修复。

基于上述可能的实现方式，在第一故障状态下，目标设备不会下电文件系统，一旦在第一预设时间内该至少一个存储节点均能被修复，那么也就可以避免下电文件系统，从而可以减少修复文件系统的时间，使得分布式存储系统可以尽快恢复业务，以保证服务质量。

在一种可能实现方式中，所述根据所述至少一个存储节点以及所述目标数据，确定所述故障状态包括：

当所述至少一个存储节点的数目大于所述分布式存储系统的冗余度，且所述目标数据的数据量符合预设条件时，根据所述分布式存储系统的故障场景，确定所述故障状态，所述故障场景用于指示所述至少一个存储节点是否同时出现故障。

在一种可能实现方式中，所述预设条件包括下述任一项：

所述目标数据的数据量与第一预设数据量之间的比值大于预设比值，所述第一预设数据量为所述分布式存储系统存储的所有数据的总数据量；

所述目标数据的数据量大于第二预设数据量。

在一种可能实现方式中，所述根据所述至少一个存储节点以及所述目标数据，确定所述故障状态之后，所述方法还包括：

当所述故障状态为所述第一故障状态时，若在所述第一预设时长内所述至少一个存储

节点未全部被修复，则将所述故障状态由第一故障状态更新为第二故障状态，所述第二故障状态用于指示所述至少一个存储节点不能在所述第一预设时长内全部被修复。

在一种可能实现方式中，所述根据所述分布式存储系统的故障场景，确定所述分布式存储系统的故障状态之前，所述方法还包括：

根据所述至少一个存储节点出现故障的时间，确定所述故障场景。

在一种可能实现方式中，所述根据所述至少一个存储节点出现故障的时间，确定所述故障场景包括：

当所述至少一个存储节点在目标时长内均出现故障时，将所述故障场景确定为第一故障场景，否则，将所述故障场景确定为第二故障场景，所述第一故障场景用于指示所述至少一个存储节点同时出现故障，所述第二故障场景用于指示所述至少一个存储节点出现故障的时间不同。

在一种可能实现方式中，所述根据所述分布式存储系统的故障场景，确定所述分布式存储系统的故障状态包括：

若所述故障场景为第一故障场景，根据所述至少一个存储节点中每一个存储节点的故障类型，确定所述故障状态，所述第一故障场景用于指示所述至少一个存储节点同时出现故障，所述故障类型用于指示一个存储节点的故障能否在第二预设时长内被修复；

若所述故障场景为第二故障场景，根据所述至少一个存储节点中最晚出现故障的第一存储节点的故障类型，确定所述故障状态，所述第二故障场景用于指示所述至少一个存储节点出现故障的时间不同。

在一种可能实现方式中，所述根据所述至少一个存储节点中每一个存储节点的故障类型，确定所述故障状态包括：

当所述至少一个存储节点中每一个存储节点的故障类型均为第一故障类型时，将所述故障状态确定为第一故障状态，所述第一故障类型用于指示一个存储节点的故障能在所述第二预设时长内被修复，所述第一故障状态用于指示所述至少一个存储节点能在所述第一预设时长内全部被修复；

当所述至少一个存储节点中目标个数的存储节点的故障类型为第二故障类型时，若所述目标个数小于或等于所述分布式存储系统的冗余度，将所述故障状态确定为所述第一故障状态，否则，将所述故障状态确定为第二故障状态，所述第二故障类型用于指示一个存储节点的故障不能在所述第二预设时长内被修复，所述第二故障状态用于指示所述至少一个存储节点不能在所述第一预设时长内全部被修复。

在一种可能实现方式中，所述根据所述至少一个存储节点中最晚出现故障的第一存储节点的故障类型，确定所述故障状态包括：

当所述第一存储节点的故障类型为第一故障类型时，将所述故障状态确定为第一故障状态，所述第一故障类型用于指示一个存储节点的故障能在所述第二预设时长内被修复，所述第一故障状态用于指示所述至少一个存储节点能在所述第一预设时长内全部被修复；

当所述第一存储节点的故障类型为第二故障类型时，则将所述故障状态确定为第二故障状态，所述第二故障类型用于指示一个存储节点的故障不能在所述第二预设时长内被修复，所述第二故障状态用于指示所述至少一个存储节点不能在所述第一预设时长内全部被

修复。

在一种可能实现方式中，所述根据所述分布式存储系统的故障场景，确定所述分布式存储系统的故障状态之前，所述方法还包括：

对于所述至少一个存储节点中的任一存储节点，当所述任一存储节点出现预设的网络故障、预设的异常掉电故障、预设的误操作故障、预设的硬件故障或预设的软件故障时，将所述任一存储节点的故障类型确定为第一故障类型，否则，将所述任一存储节点的故障类型确定为第二故障类型，所述第一故障类型用于指示一个存储节点的故障能在所述第二预设时长内被修复，所述第二故障类型用于指示一个存储节点的故障不能在所述第二预设时长内被修复。

在一种可能实现方式中，所述根据所述分布式存储系统的故障状态，进行故障处理之后，所述方法还包括：

当所述至少一个存储节点修复完成时，向所述分布式存储系统内的各个设备发送修复完成响应，所述修复完成响应用于指示所述分布式存储系统内没有故障设备。

第二方面，提供了一种分布式存储系统故障处理方法，所述分布式存储系统包含多个存储节点；该方法包括：

向所述分布式存储系统中的目标存储节点发送访问请求；

接收所述目标存储节点返回的响应；所述响应中包含所述分布式存储系统的故障状态；所述故障状态用于指示至少一个出现故障的存储节点能否在第一预设时长内全部被修复。

在一种可能实现方式中，所述接收所述目标存储节点返回的响应之后，所述方法还包括：

基于所述响应中包含的故障状态，进行故障处理。

在一种可能实现方式中，所述故障状态的故障标识包括第一故障标识或第二故障标识中的任一个，其中，所述第一故障标识用于指示第一故障状态，所述第二故障标识用于指示第二故障状态，所述第一故障状态用于指示至少一个存储节点能在第一预设时长内全部被修复，所述第二故障状态用于指示所述至少一个存储节点不能在所述第一预设时长内全部被修复，所述存储节点为所述分布式存储系统中出现故障的存储节点。

在一种可能实现方式中，所述基于所述响应中包含的故障状态，进行故障处理包括：

当所述访问请求由所述分布式存储系统中的目标客户端基于目标虚拟机发送，且所述目标虚拟机是 VMWare 虚拟机时，若所述故障状态为所述第一故障状态，不向所述目标虚拟机响应所述访问请求，所述第一故障状态用于指示至少一个存储节点能在第一预设时长内全部被修复；

当所述访问请求由所述目标客户端基于目标虚拟机发送，且所述目标虚拟机是 VMWare 虚拟机时，若所述故障状态为所述第二故障状态时，向所述目标虚拟机返回存储异常的消息，所述第二故障状态用于指示所述至少一个存储节点不能在所述第一预设时长内全部被修复。

在一种可能实现方式中，所述基于所述响应中包含的故障状态，进行故障处理包括：

当所述访问请求由所述分布式存储系统中的目标客户端基于目标虚拟机发送，且所述

目标虚拟机不是 VMWare 虚拟机时，若所述故障状态为所述第一故障状态，向所述目标虚拟机发送重试请求，所述重试请求用于指示重新下发所述访问请求，所述第一故障状态用于指示至少一个存储节点能在第一预设时长内全部被修复；

当所述访问请求由所述目标客户端基于目标虚拟机发送，且所述目标虚拟机不是 VMWare 虚拟机时，若所述故障状态为所述第二故障状态，向所述目标虚拟机返回所述目标虚拟机可识别的目标错误，所述目标错误用于指示存储介质故障，所述第二故障状态用于指示所述至少一个存储节点不能在所述第一预设时长内全部被修复。

基于上述可能的实现方式，提供了不同的故障处理方式，从而使得该本发明实施例所提供的故障处理方式更具有普适性。

在一种可能实现方式中，所述向分布式存储系统内的任一存储节点发送访问请求之前，所述方法还包括：

接收所述分布式存储系统中的目标客户端发送的目标访问请求；

所述向分布式存储系统中的目标存储节点发送访问请求包括：

基于所述目标访问请求，向分布式存储系统中的目标存储节点发送所述访问请求。

在一种可能实现方式中，所述接收所述目标存储节点返回的响应之后，所述方法还包括：

接收目标存储节点返回的修复完成响应，所述修复完成响应用于指示所述分布式存储系统内没有故障设备。

第三方面，提供了一种分布式存储系统，所述分布式存储系统包括监管节点和多个存储节点；

所述监管节点用于：

根据所述多个存储节点中的至少一个出现故障的存储节点确定所述分布式存储系统的故障状态；所述故障状态用于指示所述至少一个出现故障的存储节点能否在第一预设时长内全部被修复；

向所述多个存储节点中每一个存储节点发送所述故障状态；

所述多个存储节点中的每一个存储节点，用于接收所述故障状态。

在一种可能的实现方式中，所述故障状态的故障标识包括第一故障标识和第二故障标识中的任一个，所述第一故障标识用于指示所述第一故障状态，所述第二故障标识用于指示第二故障状态，所述第一故障状态用于指示所述至少一个存储节点能在所述第一预设时长内全部被修复，所述第二故障状态用于指示所述至少一个存储节点不能在所述第一预设时长内全部被修复。

在一种可能的实现方式中，所述多个存储节点中的每一个存储节点，还用于当接收到所述故障状态后，若再接收到所述访问请求，悬挂所述访问请求，基于接收的故障状态，进行故障处理。

第四方面，提供了一种故障处理装置，用于执行上述分布式存储系统故障处理方法。具体地，该故障处理装置包括用于执行上述第一方面或上述第一方面的任一种可选方式提供的故障处理方法的功能模块。

第五方面，提供了一种故障处理装置，用于执行上述分布式存储系统故障处理方法。

具体地，该故障处理装置包括用于执行上述第二方面或上述第二方面的任一种可选方式提供的故障处理方法的功能模块。

第六方面，提供一种计算机设备，该计算机设备包括处理器和存储器，该存储器中存储有至少一条指令，该指令由该处理器加载并执行以实现如上述分布式存储系统故障处理方法所执行的操作。

第七方面，提供一种存储介质，该存储介质中存储有至少一条指令，该指令由处理器加载并执行以实现如上述分布式存储系统故障处理方法所执行的操作。

附图说明

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，对于本领域普通技术人员来讲，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。

图1是本发明实施例提供的一种分布式存储系统的示意图；

图2是本发明实施例提供的一种分布式存储系统的网络环境的示意图；

图3是本发明实施例提供的一种分布式存储系统内各个设备之间的交互示意图

图4是本发明实施例提供的一种计算机设备的结构示意图；

图5是本发明实施例提供的一种分布式存储系统故障处理方法的流程图；

图6是本发明实施例提供的一种分布式存储系统故障处理方法的流程图；

图7是本发明实施例提供的一种故障处理装置的结构示意图；

图8是本发明实施例提供的一种故障处理装置的结构示意图。

具体实施方式

为使本发明的目的、技术方案和优点更加清楚，下面将结合附图对本发明实施方式作进一步地详细描述。

图1是本发明实施例提供的一种分布式存储系统的示意图，参见图1，该分布式存储系统包括至少一个客户端101、多个存储节点102以及监管节点103。其中，客户端101用于为用户提供数据存储以及数据读取的业务，也即是，客户端101可以将用户上传的数据存储在存储节点102中，也可以从存储节点102中读取数据。

存储节点102，用于存储客户端101写入的数据，还用于向客户端101返回数据，返回的数据可以是客户端101请求读取的数据，还可以向客户端101返回监管节点103下发的分布式存储系统的故障状态，以便客户端101根据该分布式存储系统的故障状态，处理出现的故障。

监管节点103，用于监控分布式存储系统中各个存储节点102是否出现故障，当分布式存储系统中出现故障的存储节点的数目高于该分布式存储系统的冗余度时，可能会影响到业务的正常运行，因此，该监管节点103可以根据分布式存储系统中出现故障的存储节点，确定该分布式存储系统的故障状态，并将确定的故障状态下发至所有的存储节点102，存储节点102能够告知客户端101该分布式存储系统的故障状态，从而可以使得客户端101

或存储节点 102 可以根据该分布式存储系统的故障状态, 进行故障处理。

各个存储节点 102 的基板管理控制器 (baseboard management controller, BMC) 可以实时监控各自的存储节点是否出现故障, 当该至少一个存储节点 102 中的任一存储节点的 BMC 监控到该任一存储节点出现故障时, 该 BMC 可以存储该任一存储节点出现故障的原因以及出现故障的时间, 还可以将该任一存储节点出现故障的原因以及出现故障的时间发送给监管节点 103, 以便监管节点 103 可以获知该任一存储节点是否出现故障。

当存储节点的 BMC 不将存储节点出现故障的原因以及出现故障的时间发送给监管节点 103 时, 监管节点 103 可以通过访问各个存储节点 102, 来获知各个存储节点 102 是否存在故障, 若存在故障, 则可以从故障的存储节点获取出现故障的原因以及出现故障的时间。

无论监管节点 103 是接收故障的存储节点发送的出现故障的原因以及出现故障的时间, 还是主动地从故障的存储节点获取发送的出现故障的原因以及出现故障的时间, 该监管节点 103 均可以对故障的存储节点出现故障的原因以及出现故障的时间进行存储, 以便后续可以根据故障的存储节点发送的出现故障的原因以及出现故障的时间, 确定故障类型以及故障场景, 其中, 对故障类型的描述参见下文中的步骤 603, 对故障场景的描述参见下文中的步骤 602。

该监管节点 103 可以对故障的存储节点出现故障的原因以及出现故障的时间进行存储可以包括: 将故障的存储节点发送的出现故障的原因以及出现故障的时间存储在故障表中, 该故障表可以存储编号、存储节点的标识、故障时间以及故障原因, 其中, 编号用于指示第几个出现故障的存储节点, 一个存储节点的标识用于唯一指示一个存储节点, 该标识可以是存储节点的互联网协议地址 (internet protocol address, IP), 还可以存储节点的媒体访问控制地址 (media access control address, MAC), 还可以是该存储节点在该分布式存储系统内的编号, 本发明实施例对该存储节点的标识不做具体限定。另外, 故障时间为存储节点出现故障的时间, 故障原因为存储节点出现故障的原因。

例如表 1 所示的故障表, 从表 1 可知, 当前分布式存储系统内出现故障的存储节点有 2 个, 分别为标识为 X 的存储节点以及标识为 Y 的存储节点, 其中, 标识为 X 的存储节点出现故障的时间为 A, 出现故障的原因为 D; 标识为 Y 的存储节点出现故障的时间为 B, 出现故障的原因为 C。

表 1

编号	存储节点的标识	故障时间	故障原因
01	X	A	D
02	Y	C	E

需要说明的是, 当故障表内的记录的存储节点被修复后, 该监管节点可以在该故障表中, 删除被修复完成的存储节点的相关信息, 从而监管节点可以根据故障表中最后一个存储节点的编号, 确定分布式存储系统内出现故障的存储节点的数目。

当分布式存储系统内的存储节点出现故障时, 可能造成分布式存储系统内的文件系统的损坏, 当文件系统损坏时, 文件系统内的元数据可能出现错误, 由于元数据是指用来描

述一个文件的特征的系统数据，例如访问权限、文件所有者以及文件内的数据块的分布信息等，当元数据出现错误时，元数据所指示的文件中的数据块可能无法被访问。

在一些实施中，当客户端访问任一存储节点内的数据块失败时，该客户端可以将该数据块的数据量以及唯一标识该数据块的数据标识发送给监管节点，该监管节点接收到该数据量以及数据标识后，对接收到的数据量以及数据标识进行存储，具体的可以存储在数据表中，该数据表可以用于存储数据总量、数据标识以及与每个数据标识对应的数据量，其中，数据总量为分布式存储系统内当前不可被访问的所有数据的数据量。

例如表 2 所示的数据表，从表 2 可知，分布式存储系统中当前不能被访问的数据为数据标识 M 所指示的数据块内的数据以及数据标识 N 所指示的数据块内的数据，其中，当前不能被访问的数据的数据总量为 30 千字节 (kilobyte, KB)，当前不能被访问的数据包括数据标识 M 所指示的数据块内 10KB 的数据以及数据标识 N 所指示的数据块内 20KB 的数据。

表 2

数据标识	数据量	数据总量
M	10KB	30KB
N	20KB	

需要说明的是，当客户端再一次访问不能被访问的数据块时，若该客户端可以访问成功，则该客户端向监管节点发送携带该数据块的数据标识的访问数据成功响应，该访问数据成功响应用于指示该数据块内的数据可以被访问，当接收到该访问数据成功响应后，在该数据表中，该监管节点可以删除该数据标识对应的数据量，并更新数据表内的数据总量。例如，该访问数据成功响应携带数据标识 M，则该监管节点删除数据表内与数据标识 M 相关的信息，并将数据总量更新为 20KB。需要说明的是，数据表中还可以存储有与数据标识对应的存储节点的标识，以指示哪个存储节点内的那个数据块不能被访问。

在一些实施例中，由于分布式存储系统负责的业务量比较大，客户端 101 和存储节点 102 的数目可能比较多，为了方便客户端 101 与存储节点 102 之间的数据传输，客户端 101 所在的应用层可以设置有至少一个业务交换机，以客户端 101 与存储节点 102 之间的交互。为了便于存储节点 102 之间的数据传输，可以在存储节点 102 所在的存储层设置至少一个存储交换机，以实现各个存储节点 102 之间的交互。为了便于监管节点 103 与存储节点 102 之间的数据传输，可以设置有监管交换机，以实现监管节点 103 与存储节点 102 之间的交互。

从以上的描述可知，在该分布式存储系统中，除了需要提供业务服务以外，还需要提供监控服务，对于不同的服务可以通过不同的网络来实现。为了实现客户端、存储节点以及监管节点之间的网络连接，该客户端、存储节点以及监管节点中均可以安装有至少一个网口，该至少一个网口可以用于连接不同的网络，不同的网络可以传输不同服务的数据，该至少一个网口可以分别是连接业务网络的业务网口、连接监管网络的监管网口，以及连接 BMC 网络的 BMC 网口。

为了说明分布式存储系统中的网络环境，参见图 2 所示的本发明实施例提供的一种分布式存储系统的网络环境的示意图，在该分布式存储系统中网络可以包括业务网络、监管

网络以及 BMC 网络。

其中，业务网络是存储节点之间用于心跳、数据同步以及镜像时所使用的网络，例如，当将存储节点 1 中存储的数据块 1 同步至存储节点 2 时，存储节点 1 可以通过业务网口，在该业务网络中，向存储节点 2 发送数据块 1，那么存储节点 2 通过自己的业务网口，可以接收到数据块 1，并将数据块 1 存储在存储节点 2 内。

监管网络是监控存储节点是否出现故障以及进行信息查询时所使用的网络，在监管网络中，可以传输监管节点下发的分布式存储系统的故障状态，还可以查询出现故障的存储节点。在一些可能的实施方式中，监管节点可以通过监管网口向存储节点的监管网口，在监管网络中，发送分布式存储系统的故障状态，存储节点可以通过自己的监管网口从监管网络中接收监管节点的故障状态，当存储节点再接收到客户端的业务请求（也即是下文中的 SCSI 请求）后，可以不处理接收的业务请求，直接向客户端返回下发的故障状态，以便客户端可以根据故障状态，进行相应的故障处理。

BMC 网络是管理 BMC 的网络，监管节点通过访问该 BMC 网络的 BMC 网口，可以监控 BMC 的状态，根据监控的 BMC 的状态，可以确定存储节点是否有故障。需要说明的是，BMC 网络为可选的网络，在一些实施方式中，可以不通过 BMC 来监控存储节点的是否有故障，而是可以通过其他方式，来监控存储节点是否有故障，因此，分布式存储系统内还可以不设置该 BMC 网络，直接通过监管网络，来实现监控。

当分布式存储系统内的存储节点均和业务网络、监管网络以及 BMC 网络连接时，监管节点则可以从三网中实时接收存储节点的状态信息，例如是否故障的信息。

为了进一步说明客户端、存储节点以及监管节点之间交互过程，参见图 3 所示的本发明实施例提供的一种分布式存储系统内各个设备之间的交互示意图，在图 3 中，一个存储节点中可以安装有至少一个对象存储（object storage device, OSD）进程、SCSI 处理进程以及节点监控服务（node monitor service, NMS）代理进程。

其中，一个 OSD 进程可以对应存储节点中用于存储数据的一个或多个存储介质，该存储介质可以是硬盘，OSD 进程用于管理对于一个或多个存储介质的访问请求，访问请求用于指示对待处理的数据进行处理，其中，对待处理的数据进行处理可以包括读取所述至少一个或多个存储介质内存储的数据块，待处理的数据块包括待处理的数据，对待处理的数据进行处理可以还包括将待处理的数据写入所述至少一个或多个存储介质，当访问请求可以使用 SCSI 发送时，该访问请求可以视为 SCSI 请求。

SCSI 处理进程用于从业务网络中获取客户端发送的 SCSI 请求，并转换和分解 SCSI 请求，得到多个 SCSI 子请求，并将多个 SCSI 子请求下发到对应的 OSD 进程。例如，SCSI 请求携带待读取的数据的逻辑区块地址（logical block address, LBA）为 100-200，由于 LBA 100-150 所指示的存储位置在存储节点 1 的存储介质 1 中，LBA 151-200 所指示的存储位置在存储节点 1 的存储介质 2 中，则 SCSI 处理进程可以将 SCSI 请求转换和分解为 2 个 SCSI 子请求，其中，SCSI 子请求 1 用于指示请求读取存储介质 1 中 LBA 100-150 处存储的数据，SCSI 子请求 2 用于指示请求读取存储介质 2 中 LBA 151-200 处存储的数据，从而 SCSI 处理进程可以将 SCSI 子请求 1 发送至于存储介质 1 对应的 OSD 进程，将 SCSI 子请求 2 发送至于存储介质 2 对应的 OSD 进程。

一个NMS代理进程用于接收监管节点下发的分布式存储系统的故障状态,并向一个存储节点的所有OSD进程下发接收的故障状态。例如,客户端通过监管网络向存储节点发送分布式存储系统的故障状态,存储节点内的NMS代理进程从监管网络中获取监管节点发送的故障状态,并将获取的故障状态下发至该存储节点内的各个OSD进程中,当任一OSD进程接收到故障状态后,若再接收到SCSI处理进程发送的SCSI子请求或SCSI请求,则直接向SCSI处理进程发送接收的故障状态,以便安装SCSI处理进程的设备根据接收的故障状态,进行故障处理。

需要说明的是,在一些实施方式中,SCSI处理进程未被安装在存储节点内,而是安装在客户端内,本发明实施例对安装该SCSI处理进程的设备不做具体限定。例如,客户端的SCSI处理进程中的SCSI请求携带待读取的数据的LBA为0-100,由于LBA 0-50所指示的存储位置在存储节点1中,LBA 51-100所指示的存储位置在存储节点2中,则SCSI处理进程可以将SCSI请求转换和分解为2个SCSI子请求,其中,SCSI子请求1用于指示请求读取存储节点1中LBA 0-50处存储的数据,SCSI子请求2用于指示请求读取存储节点2中LBA 51-100处存储的数据,从而SCSI处理进程可以将SCSI子请求1发送至于存储节点1内的OSD进程,将SCSI子请求2发送至于存储节点2内的OSD进程。

客户端、存储节点以及监管节点均可以是计算机设备,为了进一步说明,计算机设备的硬件结构,参见图4所示的本发明实施例提供的一种计算机设备的结构示意图,计算机设备400包括可因配置或性能不同而产生比较大的差异,可以包括一个或一个以上处理器(central processing units, CPU)401和一个或一个以上的存储器402,其中,该存储器402中存储有至少一条指令,该至少一条指令由该处理器401加载并执行以实现下的故障处理方法实施例提供的方法。当然,该计算机设备400还可以具有有线或无线网络接口、键盘以及输入输出接口等部件,以便进行输入输出,该计算机设备400还可以包括其他用于实现设备功能的部件,在此不做赘述。

在示例性实施例中,还提供了一种计算机可读存储介质,例如包括指令的存储器,上述指令可由终端中的处理器执行以完成下述实施例中的故障处理方法。例如,该计算机可读存储介质可以是只读存储器(read-only memory, ROM)、随机存取存储器(random access memory, RAM)、只读光盘(compact disc read-only memory, CD-ROM)、磁带、软盘和光数据存储节点等。

在本发明实施例中,监管节点可以根据出现故障的存储节点,来确定分布式存储系统的故障状态,监管节点并将确定的故障状态下发至分布式存储系统内所有的存储节点;客户端在接收到用户的读请求或者写请求后,可以向分布式存储系统内的存储节点发送SCSI请求,用于完成用户的读请求或者写请求;当任一存储节点接收到SCSI请求后,任一存储节点暂不处理接收的SCSI请求,并向客户端返回接收的故障状态,使得客户端可以基于任一存储节点返回的故障状态,进行故障处理。而在一些实施例中,任一存储节点也可以基于故障状态进行故障处理,在一种可能的实现方式中,当任一存储节点接收到监管节点发送的故障状态时,若再接收到客户端发送的SCSI请求,则任一存储节点可以根据故障状态,进行故障处理。

为了进一步说明上述过程,参见如图5所示的本发明实施例提供的一种分布式存储系

统故障处理方法的流程图，该方法具体包括：

501、监管节点确定分布式存储系统内出现故障的至少一个存储节点以及该至少一个存储节点无法被访问的目标数据。

该目标数据还可以是该分布式存储系统中存储的且不能被访问的数据，该目标数据还可以仅是该至少一个存储节点所存储的无法被访问的数据。本发明实施例以该目标数据是该至少一个存储节点内不能被访问的数据为例进行说明。

该监管节点可以通过查询的方式，确定该至少一个存储节点以及该目标数据，在一种可能的实现方式中，该监管节点可以每经过第八预设时长，查询故障表以及数据表，从该故障表中确定出现故障的存储节点，从数据表中确定无法被访问的数据。该第八预设时长可以是 10 分钟或者 1 小时，本发明实施例对该第八预设时长不做具体限定。

需要说明的是，在前文中对从该故障表中确定出现故障的存储节点的方式以及从数据表中确定无法被访问的数据的方式有描述，在此不做赘述。

当该监管节点执行完本步骤 501 后，还可以根据至少一个存储节点以及所述目标数据，确定故障状态，那么，确定分布式存储系统内出现故障的至少一个存储节点以及该至少一个存储节点无法被访问的目标数据；根据至少一个存储节点以及所述目标数据，确定故障状态的过程，也即是，根据所述多个存储节点中的至少一个出现故障的存储节点，确定所述分布式存储系统的故障状态的过程。其中，根据所述多个存储节点中的至少一个出现故障的存储节点，确定所述分布式存储系统的故障状态的过程可以通过下述步骤 502 所示的过程的来实现。

502、当该至少一个存储节点的数目大于该分布式存储系统的冗余度，且该目标数据的数据量符合预设条件时，该监管节点则将该分布式存储系统的故障状态确定为第一故障状态。

该冗余度为分布式存储系统内存储的数据的冗余度，也即是该分布式存储系统内存储的数据的副本数目，分布式存储系统的故障状态用于指示该至少一个存储节点能否在第一预设时长内全部被修复，故障状态可以包括第一故障状态以及第二故障状态中的任一个，其中，第二故障状态用于指示该至少一个存储节点不能在该第一预设时长内全部被修复。也即是，当该至少一个存储节点可以在第一预设时长内被修复时，可以认为分布式存储系统内的故障可以在短时间内修复，分布式存储系统处于第一故障状态，该第一故障状态也即是节点短时故障（transient node down, TND）状态。当该至少一个存储节点可以在第一预设时长内被修复时，则认为分布式存储系统内的故障不可以在短时间内修复，分布式存储系统处于第二故障状态，第二故障状态也即是节点长期故障（permanent node down, PND）状态。第一预设时长可以是 20 分钟或者 2 两个小时，本发明实施例对该第一预设时长不做具体限定。

当分布式存储系统内存储的数据较多且存储节点的数目也较多时，若分布式存储系统内出现故障的存储节点的数目小于该分布式存储系统的冗余度，该分布式存储系统内的未故障的存储节点可以根据可以被访问的数据，重构故障的存储节点内不能被访问的数据，因此，当分布式存储系统内出现故障的存储节点的数目小于该分布式存储系统的冗余度时，不会影响到该分布式存储系统的正常业务，那么，也就无需修复故障的存储节点。但是，

若出现故障的存储节点的数目大于该分布式存储系统的冗余度，该分布式存储系统内的未出现故障的存储节点无法根据可以被访问的数据，重构故障的存储节点内不能被访问的数据，从而就可能会影响到分布式存储系统内正常的业务。并且考虑到分布式存储系统内存储的数据能否被访问也是影响业务的重要因素，当分布式存储系统内无法被访问的目标数据过多时，对分布式存储系统所提供的业务的影响也就比较大，可能会影响到业务的正常运行，因此，该监管节点可以先确定分布式存储系统的故障状态，以便可以快速根据故障状态，进行故障处理，来最大限度的降低对业务的影响程度。而当该目标数据的较少时，对业务影响也就相对较小，可能不会影响到业务的正常运行，为了可以为用户持续提供业务，该监管节点可以暂不进行故障处理。

监管节点可以通过预设条件，确定目标数据的数据量是否能够影响到业务的正常运行，该预设条件可以包括下述任一项：目标数据的数据量与第一预设数据量之间的比值大于预设比值，该第一预设数据量为该分布式存储系统存储的所有数据的总数据量；该目标数据的数据量大于第二预设数据量。当目标数据的数据量与第一预设数据量之间的比值大于预设比值时，或者当该目标数据的数据量大于第二预设数据量时，说明目标数据的数据量比较大，也就可能会影响到业务的正常运行，从而当目标数据的数据量符合预设条件时，该监管节点可以先将分布式存储系统的故障状态设置为第一故障状态，以便可以快速根据故障状态，进行故障处理，以降低对业务的影响程度。若超过第一预设时间，该至少一个存储节点未全部修复完成，则可以再将该故障状态更新为第二故障状态。需要说明的是，该预设比值可以是 0.4、0.5 或者是 0.6，本发明实施例对该预设比值、第一预设数据量以及第二预设数据量不做具体限定。

为了能够将该至少一个存储节点的数目与该分布式存储系统的冗余度相比较，该监管节点可以查询该监管节点所存储的故障表，从故障表中确定该至少一个存储节点的数目。为了能够确定目标数据的数据量是否符合预设条件，该监管节点可以查询所存储的数据表，从数据表中确定该分布式存储系统内目前无法访问的目标数据的数据量。例如，监管节点通过查询表 2 可以确定目标数据包括数据标识 M 所指示的数据块内 10KB 的数据以及数据标识 N 所指示的数据块内 20KB 的数据。

需要说明的是，在前文中介绍故障表时，对从故障表中确定在该分布式存储系统内出现故障的存储节点的数目的过程进行了叙述，在此，本发明实施例对从故障表中确定在该分布式存储系统内出现故障的存储节点的数目的过程不做赘述。

需要说明的是，本步骤 502 所示的过程也即是根据该至少一个存储节点以及所述目标数据，确定该分布式存储系统的故障状态的过程。

503、该监管节点向该分布式存储系统内的所有存储节点发送用于指示第一故障状态的第一故障标识。

故障状态的故障标识包括第一故障标识和第二故障标识中的任一个，其中，第一故障标识用于指示第一故障状态，第二故障标识用于指示第二故障状态，该第一故障标识和第二故障标识可以不同，例如，第一故障标识可以是 s，第二故障标识可以是 t，本发明实施例对第一故障标识和第二故障标识的表示方式不做具体限定。

该监管节点可以向每个存储节点的 NMS 代理进程发送该第一故障标识，从而实现向所

有存储节点发送该第一故障标识，以告知所有存储节点分布式存储系统目前的故障状态为第一故障状态。

需要说明的是，本步骤 503 所示的过程也即是监管节点向分布式存储系统包含的多个存储节点中每一个存储节点发送所述故障状态的过程。

需要说明的是，在一些实施例中，监管节点在确定完故障状态后，还可以根据所述存储系统的故障状态，进行故障处理，本发明实施例对监管节点进行故障处理的过程不做具体限定。

504、该分布式存储系统中的目标存储节点接收该第一故障标识。

该目标存储节点为该分布式存储系统中的任一存储节点，该目标存储节点内的每个 OSD 进程可以从该目标存储节点的 NMS 代理进程获取该第一故障标识，从而该目标存储节点的每个 OSD 进程可以获取该第一故障标识。需要说明的是，该分布式存储系统内的每一个存储节点都可以执行本步骤 504，对于故障的存储节点可能能够接收到该第一故障标识，也可能接收不到该第一故障标识。

505、目标设备向该目标存储节点发送访问请求，

该访问请求用于指示读取该目标存储节点所存储的数据或者向该目标存储节点写入数据。该目标设备为安装有 SCSI 处理进程的设备，可以是该目标客户端，还可以是目标存储节点，其中，目标客户端为该分布式存储系统中的任一客户端。本步骤 506 可以由目标设备内的 SCSI 处理进程来实现。

在本步骤之前 505 之前，分布式存储系统内的目标客户端可以向目标设备发送目标访问请求，该目标访问请求用于执行对第一目标数据进行处理，其中第一目标数据包括该访问请求所指示的数据，该目标访问请求可以携带目标存储地址，该目标存储地址可以是第一目标数据的存储地址。由目标客户端可以安装的目标虚拟机发送目标访问请求来。具体地，该目标虚拟机可以向该目标设备内的 SCSI 处理进程发送该目标访问请求。目标虚拟机向 SCSI 处理进程发送目标访问请求可以由用户的动作来触发，例如，当用户在客户端的界面内输入待读取的数据的存储地址，并点击读取按钮时，触发客户端内的目标虚拟机向 SCSI 处理进程发送目标访问请求，以请求读取到用户输入的存储地址处存储的数据。

然后，该目标设备接收分布式存储系统内的目标客户端发送的目标访问请求，本步骤 506 可以通下述方式来实现：目标设备基于该目标访问请求，向分布式存储系统内的目标存储节点发送访问请求。具体地，SCSI 处理进程接收到该目标访问请求后，根据目标地址，对该目标访问请求进行转化和分解，得到多个访问请求，每个访问请求可以携带该目标地址中的部分地址，这部分地址可以是该目标存储节点内的任一 OSD 进程所管理的存储介质中的偏移地址，从而 SCSI 处理进程向目标存储节点内的 OSD 进程发送对应的访问请求，将访问请求转换访问请求的过程也即是前述的转换 SCSI 请求的过程。

506、当该分布式存储系统内的目标存储节点接收到该第一故障标识后，若该目标存储节点再接收到访问请求，该目标存储节点悬挂该访问请求，向目标设备发送该第一故障标识。

本步骤 506 可以由该目标存储节点内接收该访问请求的 OSD 进程来实现。当该目标存储节点接收到该第一故障标识后，说明该目标存储节点已经知道该分布式存储系统中有故

障的存储节点，且目前的故障状态为第一故障状态，由于目前分布式存储系统内的存储节点出现故障，则该目标存储节点可以将该访问请求悬挂，暂不处理该访问请求，等待故障的存储节点被自动修复或者手动修复。

为了使得目标设备也能获知分布式存储系统的故障状态，则该目标存储节点在接收到目标设备发送的访问请求时，该目标存储节点可以将该第一故障标识输出给目标设备。具体地，该目标存储节点内的任一 OSD 进程接收到目标设备 SCSI 处理进程发送的访问请求后，任一 OSD 进程向该 SCSI 处理进程发送该第一故障标识。当然，在一些实施例中，监管节点确定完该分布式存储系统的故障状态后，也可以将故障状态直接发送给目标设备，以便目标设备可以获取分布式存储系统的故障状态，也就无需通过存储节点向目标设备发送故障状态。

需要说明的是，本步骤 506 所示的过程也即是当该分布式存储系统中的目标存储节点接收到该故障标识后，若该目标存储节点再接收到访问请求时，输出故障标识的过程。

507、目标设备接收该目标存储节点基于该访问请求返回的第一故障标识。

本步骤 507 以由该目标设备内的 SCSI 处理进程来实现，本步骤 507 所示的过程也即是接收该目标存储节点基于该访问请求返回的故障标识的过程。当任一 OSD 进程向该 SCSI 处理进程发送该第一故障状态，该 SCSI 处理进程可以接收该第一故障标识。需要说明的是，本步骤 507 所示的过程，也即是，接收所述目标存储节点返回的响应；所述响应中包含所述分布式存储系统的故障状态；所述故障状态用于指示至少一个出现故障的存储节点能否在第一预设时长内全部被修复的过程，其中，所述响应中包含所述分布式存储系统的故障状态也即是故障标识。

508、目标设备基于接收的第一故障标识，进行故障处理。

本步骤 508 可以由目标设备内安装的 SCSI 处理进程来执行，当该 SCSI 处理进程接收到该第一故障标识后，该 SCSI 处理进程可以将基于第一故障标识，进行故障处理，以便目标客户端内能够进行相应地处理。

在一些实施例中，目标客户端中与该 SCSI 处理进程对接目标虚拟机可能是 VMWare 虚拟机，也可能不是 VMWare 虚拟机，由于不同的虚拟机目标设备进行故障处理的方式不同，因此，本步骤 508 可以通过下述方式 1-2 中的任一方式实现。

方式 1、当该访问请求由该分布式存储系统中的目标客户端基于目标虚拟机发送，且该目标虚拟机是 VMWare 虚拟机时，若该故障状态为该第一故障状态，SCSI 处理进程不向该目标虚拟机响应该访问请求。

当 SCSI 处理进程接收到第一故障标识时，说明分布式存储系统处于第一故障状态，对于 VMWare 虚拟机而言，第一故障状态也即是 ADP 状态。且由于 SCSI 处理进程接收的访问请求为该目标虚拟机发送的，那么，为了使得 VMWare 虚拟机可以感知到分布式存储系统处于自己定义的 APD 状态，SCSI 处理进程不响应目标虚拟机，又因为分布式存储系统内所有的 SCSI 处理进程只要向 OSD 进程下发访问请求，均会接收到第一故障标识，而处理访问请求的每个 SCSI 处理进程均不会响应目标虚拟机，从而可以模拟分布式存储系统内所有链路无响应 (DOWN)，对于目标虚拟机而言，没有接收到 SCSI 处理进程的响应，目标虚拟机就会持续发送访问请求，从而即使分布式存储系统内的存储节点没有全部故障时，

也可以根据 VMWare 虚拟机定义的故障状态进行故障处理。

方式 2、当该访问请求由该目标客户端基于目标虚拟机发送，且该目标虚拟机不是 VMWare 虚拟机时，若该故障状态为该第一故障状态，则 SCSI 处理进程向该目标虚拟机发送重试请求，该重试请求用于指示重新下发该访问请求。

重试请求中携带的检测关键字可以是 Unit Attention (0x6) 错误码，Unit Attention (0x6) 错误码可以指示存储节点内的存储介质或链路状态发生了变化，也即是出现了故障，该目标虚拟机接收到该重试请求后，目标虚拟机向该 SCSI 处理进程重新下发一个访问请求，用于实现上述方式 1 中的处理方式。

对于不同的虚拟机，本发明实施例提供了不同的故障处理方式，从而使得该本发明实施例所提供的故障处理方式更具有普适性。

需要说明的是，本步骤 508 所示的过程也即是基于响应中包含的故障状态，进行故障处理的过程。

509、若在该第一预设时长内该至少一个存储节点全部修复，监管节点向该分布式存储系统内的各个设备发送修复完成响应，该修复完成响应用于指示该分布式存储系统内没有故障设备。

该各个设备包括存储节点以及客户端。当在该第一预设时长内该至少一个存储节点全部修复时，说明此时分布式存储系统内的没有故障设备，若该监管节点存储了该用于标识第一故障状态的第一故障标识，该监管节点可以删除该第一故障标识，并向该分布式存储系统内的各个设备发送修复完成响应，已告知各个设备该分布式存储系统内没有故障设备，可以正常工作，那么当各个设备接收到该修复完成响应后，删除之前接收的故障标识，并可以开始正常工作。需要说明的是，步骤 509 所示的过程也即是当该至少一个存储节点修复完成时，向该分布式存储系统内的各个设备发送修复完成响应的过程。

在现有技术中，当客户端未获取任何存储节点的响应时，直接认为分布式存储系统的故障状态为 APD 状态，当存储节点明确返回存储异常的消息时才会认为分布式存储系统的故障状态确定为 PDL 状态，若存储节点出现长期故障时，可能无法向客户端返回存储异常的消息，从而客户端可能将故障状态确定为 APD 状态，因此，现有技术中确定分布式存储系统的故障状态并不精确，且若将 PDL 状态误认为 APD 状态，将导致业务侧修复工作无法开展，最终反而延长故障修复时长。而在本发明提供的实施例中，分布式存储系统内的每个存储节点均知道分布式存储系统的故障状态，对于未出现故障的存储节点而言，可以基于 SCSI 请求向目标设备返回故障状态，从而目标设备可以明确的知道分布式存储系统的故障状态，进而可以提高目标设备确定故障状态的精度。

510、当该故障状态为第一故障状态时，若在该第一预设时长内该至少一个存储节点未全部被修复，则监管节点将该故障状态由第一故障状态更新为第二故障状态。

由于第一预设时长仅是一个预设的时长，在该第一预设时长内，该至少一个存储节点可能无法被全部修复，那么，当该至少一个存储节点可能未全部被修复时，就可能需要更长的时间修复还未修复的存储节点。由于修复还未修复的存储节点所用的时长不确定，可能会比较久，因此，该监管节点可以直接将该故障状态由第一故障状态更新为第二故障状态。

511、当将该故障状态由第一故障状态更新为该第二故障状态时，监管节点向该分布式存储系统内的所有存储节点发送用于指示该第二故障的第二故障标识。

监管节点向所有存储节点发送第二故障标识的方式与步骤 503 中所有存储节点发送第一故障标识的方式同理，在此，本发明实施例对该本步骤 511 不做赘述。本步骤 511 所示的过程也即是向该多个存储节点中每一个存储节点发送所述故障状态的过程。

512、目标存储节点接收该第二故障标识。

目标存储节点接收该第二故障标识与步骤 504 中接收第一故障标识的方式同理，在此，本发明实施例对该本步骤 512 不做赘述。

513、目标设备向该目标存储节点发送访问请求。

目标设备向该目标存储节点发送访问请求的方式在步骤 505 中有相关描述，在此，本发明实施例对该本步骤 513 不做赘述。

514、当该分布式存储系统内的目标存储节点接收到该第二故障标识时，若该目标存储节点再接收到访问请求，该目标存储节点悬挂该访问请求，输出该第二故障标识。

该任一存储节点悬挂访问请求以及输出该第二故障标识方式与步骤 506 中该目标存储节点悬挂访问请求以及输出该第一故障标识的方式同理，在此，本发明实施例对该本步骤 514 不做赘述。

515、目标设备接收该目标存储节点基于该访问请求返回的第二故障标识。

目标设备接收该第二故障标识的方式与步骤 507 中接收第一故障标识的方式同理，在此，本发明实施例对此不做赘述。

516、目标设备基于接收的第二故障标识，进行故障处理。

本步骤 516 可以由目标设备的 SCSI 处理进程来执行，在一些实施例中，客户端中与该 SCSI 处理进程对接目标虚拟机可能是 VMWare 虚拟机，也可能不是 VMWare 虚拟机，由于不同的虚拟机目标设备进行故障处理的方式不同，因此，本步骤 516 可以通过下述方式 3-4 中的任一方式实现。

方式 3、当该访问请求由该目标客户端基于目标虚拟机发送，且该目标虚拟机是 VMWare 虚拟机时，若该故障状态为该第二故障状态，SCSI 处理进程向 VMWare 虚拟机返回存储异常的消息。

当该故障标识第二故障标识时，说明至少一个存储节点在第一预设时间内不能全部被修复，需要更长的时间，目标虚拟机可以进行 PDL 状态下的故障处理，为了目标虚拟机可以感知到 PDL 状态，存储异常的消息可以携带 VMWare 虚拟机自定义的 SK 0x0，ASC&ASCQ 0x0200 或 SK 0x5，ASC&ASCQ 0x2500 等 SCSI 错误，SCSI 错误可以指示分布式存储系统内的状态为 PDL 状态，从而目标虚拟机接收到该存储异常的消息就可以感知到 PDL 状态，那么该目标虚拟机可以将该分布式存储系统内的文件系统下电，等待技术人员修复分布式存储系统内的故障，或者按照用户自定义的故障处理方式选择较优故障处理方式来处理故障的存储节点。例如，将故障的存储节点下电。

需要说明的是，由于分布式存储系统内的存储节点出现不能短时修复的故障时，可能导致文件系统异常，为了保证分布式存储系统内的存储节点的故障被修复后，文件系统能够被正常使用，则需要先将文件系统下电。当修复完成时，在将文件系统进行上电，并修

复文件系统。

方式 4、当该访问请求由该目标客户端基于目标虚拟机发送，且该目标虚拟机不是 VMWare 虚拟机时，若该故障状态为该第二故障状态，该目标设备向该目标虚拟机返回该目标虚拟机可识别的目标错误，该目标错误用于指示存储介质故障。

该目标错误可以是 Sense key 0x3 错误，也即是存储介质错误 (Medium Error)，一般的虚拟机均可以识别，当目标虚拟机接收到该目标错误，说明此时分布式存储系统的状态为第二故障状态，那么该目标设备可以将分布式文件系统下电，等待技术人员修复分布式存储系统内的故障，或者按照用户自定义的故障处理方式选择较优故障处理方式来处理故障的存储节点。

需要说明的是，本步骤 516 所示的过程也即是基于接收的故障标识，目标设备进行故障处理的过程，也即是基于所述响应中包含的故障状态，进行故障处理的过程。

517、当该至少一个存储节点修复完成时，向该分布式存储系统内的各个设备发送修复完成响应。

本步骤 517 与步骤 509 同理，在此本发明实施例对本步骤 517 不做赘述。需要说明的是，该每个存储节点的故障可以是被自身所修复，还可以被技术人员修复，本发明实施例对存储节点的修复方式不做具体限定。

需要说明的是，当客户端接收到该修复完成响应时，若文件系统已经下电，则该客户端对该文件系统上电，并修复该文件系统。由于文件系统内存储有大量的元数据，当对该文件系统进行修复时，需要扫描该文件系统内的所有元数据，并修改扫描到的错误元数据，一般该修复文件系统的过程需要消耗部分时间，在第一故障状态下，客户端不会下电文件系统，一旦在第一预设时间内该至少一个存储节点均能被修复，那么也就可以避免下电文件系统，从而可以减少修复文件系统的时间，使得分布式存储系统可以尽快恢复业务，以保证服务质量。

本发明实施例所示的方法，根据所述多个存储节点中的至少一个出现故障的存储节点，确定所述分布式存储系统的故障状态，从而无需当所有存储节点均故障时，才确定分布式存储系统的故障状态，当确定完故障状态后，可以立即向分布式存储系统内的每个存储节点发送故障状态，以便每个存储节点根据确定的故障状态进行故障处理，从而可以降低分布式存储系统恢复正常所用的时间。并且，对于不同的虚拟机，本发明实施例提供了不同的故障处理方式，从而使得该本发明实施例所提供的故障处理方式更具有普适性。并且，分布式存储系统内的每个存储节点均知道分布式存储系统的故障状态，对于未出现故障的存储节点而言，可以基于访问请求向目标设备返回故障状态，从而目标设备可以明确的知道分布式存储系统的故障状态，进而可以提高目标设备确定故障状态的精度。并且，在第一故障状态下，目标设备不会下电文件系统，一旦在第一预设时间内该至少一个存储节点均能被修复，那么也就可以避免下电文件系统，且当该至少一个存储节点恢复后，文件系统以及业务可以立即恢复，从而可以减少修复文件系统的时间，使得分布式存储系统可以尽快恢复业务，以保证服务质量。

由于存储节点出现的故障可能在短时间内可修复，还可能需要长时间修复，且每个存

储节点的修复时间，都可以影响到整个分布式存储系统的修复时间，因此，在一些实施例中，还可以根据每个存储节点的故障类型，来确定分布式存储系统的故障状态。为了进一步说明此过程，参见图 6 所示的本发明实施例所提供的一种分布式存储系统故障处理方法的流程图，该方法的流程可以包括以下步骤。

601、监管节点确定分布式存储系统内出现故障的至少一个存储节点以及该至少一个存储节点无法被访问的目标数据。

本步骤 601 与步骤 501 同理，在此本发明实施例对步骤 601 不做赘述。

602、监管节点根据该至少一个存储节点出现故障的时间，确定该分布式存储系统的故障场景，该故障场景用于指示该至少一个存储节点是否同时出现故障。

该故障场景可以包括第一故障场景和第二故障场景中的任一个，其中，该第一故障场景用于指示该至少一个存储节点同时出现故障，该第二故障场景用于指示该至少一个存储节点出现故障的时间不同。

该监管节点可以从存储的故障表中确定每一个存储节点出现故障的时间，从而该监管节点就可以根据至少一个存储节点出现故障的时间是否相同，确定故障场景。

在一种可能的实现方式中，当该至少一个存储节点在目标时长内均出现故障时，该监管节点将该故障场景确定为第一故障场景，否则，将该故障场景确定为第二故障场景。需要说明的是本步骤 602 所示的过程也即是根据所述至少一个存储节点出现故障的时间，确定所述故障场景的过程。

603、对于该至少一个存储节点中的任一存储节点，当该任一存储节点出现预设的网络故障、预设的异常掉电故障、预设的误操作故障、预设的硬件故障或预设的软件故障时，该监管节点将该任一存储节点的故障类型确定为第一故障类型，否则，将该任一存储节点的故障类型确定为第二故障类型。

存储节点的故障类型用于表示一个存储节点的故障能否在第二预设时长内被修复，故障类型可以包括第一故障类型和第二故障类型中的任一个，第二故障类型用于指示一个存储节点的故障不能在该第二预设时长内被修复，该第二预设时长可以小于或等于第一预设时长，本发明实施例对该第二预设时长不做具体限定。

其中，该预设的网络故障可以包括下述第 1.1-1.7 项中的任一项：

第 1.1 项、该任一存储节点的业务网口无法访问，该任一存储节点的监管网口能够访问，该业务网口是存储节点之间用于心跳、数据同步以及镜像时所使用的业务网络的网口，该监管网口为监控存储节点是否出现故障以及进行信息查询时所使用的监管网络的网口。

该监管节点可以通过监管节点的业务网口向任一存储节点的业务网口发送因特网包探索器 (ping) 请求，该 ping 请求用于请求建立连接，若可以连接成功，则认为该任一存储节点的业务网能够口访问，反之，则认为该任一存储节点的业务网口无法访问。同理，该监管节点可以通过监管节点的监管网口向任一存储节点的监管网口发送 ping 请求，若可以连接成功，则认为该任一存储节点的监管网口可以访问，否则认为该任一存储节点的监管网口无法访问。

若监管节点通过业务网口无法访问该任一存储节点，说明任一存储节点出现了网络故障，但是通过监管网口能够访问该任一存储节点，说明该任一存储节点出现的故障可以在

短时间内修复，则该任一存储节点出现了预设的网络故障。

第 1.2 项、当该业务网络和该监管网络为同一个目标网络时，该任一存储节点在该目标网络内传输的数据包出现第一预设数目的丢包或第二预设数目的畸形包，且该任一存储节点的业务网口、监管网口以及基板管理控制器 BMC 网口均不可访问，BMC 网口为管理 BMC 的 BMC 网络的网口。

该监管节点可以通过监管节点的 BMC 网口向任一存储节点的 BMC 网口发送 ping 请求，若可以连接成功，则认为该任一存储节点的 BMC 网口可以访问，否则，认为该任一存储节点的 BMC 网口不可以访问。

在交付阶段，技术人员可以将业务网络和监管网络配置为同一个网络，也即是目标网络，当该业务网络和该监管网络为同一个目标网络时，如果该任一存储节点在该目标网络内传输的数据包出现第一预设数目的丢包或第二预设数目的畸形包，说明该任一存储节点出现了网络故障，若该监管节点无法访问该任一存储节点，说明该任一存储节点出现的故障可以在短时间内修复，则该任一存储节点出现了预设的网络故障。

第 1.3 项、当该业务网络和该监管网络为同一个目标网络时，该任一存储节点在该目标网络内传输的数据包出现大于第一预设数目的丢包或大于第二预设数目的畸形包，且该任一存储节点在目标网络内传输数据的时延大于第三预设时长。

需要说明的是，当监管节点在向该任一存储节点发送 ping 请求时，若连接不成功，该任一存储节点会向该监管节点发送连接失败响应，该连接失败响应用于指示连接失败，且该连接失败响应内可以携带时延信息，该时延信息用于指示该任一存储节点在目标网络内传输数据的时延，从而该监管节点可以判断时延信息所指示的时延是否大于第三预设时长。

第 1.4 项、当该业务网络和该监管网络为同一个目标网络时，该目标网络中出现该任一存储节点的优先级的流量控制 PFC 报文大于第三预设数目，且该任一存储节点不可访问。

监管节点可以检测目标网络内的各个优先级的流量控制（priority-based flow control, PFC）报文，从而可以确定该任一存储节点发送的 PFC 报文的数目是否大于第三预设数目，该监管节点就可以确定该任一存储节点是否符合预设条件。本发明实施例对该第三预设数目不做具体限定。

第 1.5 项、当该业务网络和该监管网络为同一个目标网络时，该目标网络中出现该任一存储节点发送第三预设数目的优先级的流量控制 PFC 报文，且该任一存储节点在该目标网络内传输数据的时延大于第四预设时长。

需要说明的是，本发明实施例对该四预设时长不做具体限定。

第 1.6 项、当该业务网络和该监管网络为同一个目标网络时，该目标网络中出现任一存储节点导致的广播风暴，且该任一存储节点的业务网口、监管网口以及 BMC 网口均不可访问。

需要说明的是，当任一存储节点在该目标网络内发送大量的广播包，则该目标网络中可能出现广播风暴。

第 1.7 项、当该业务网络和该监管网络为同一个目标网络时，该目标网络中出现该任一存储节点导致的广播风暴，且该任一存储节点在该目标网络内的时延大于第五预设时长。

需要说明的是，本发明实施例对该五预设时长不做具体限定。

该预设的异常掉电故障可以包括下述第 2.1-2.2 项中的任一项：

第 2.1 项、机框内的所有存储节点的业务网口、监管网口以及 BMC 网口均不可访问，该机框包括所述任一存储节点。

一个机框内可以包括至少一个存储节点，当所有的存储节点业务网口、监管网口以及 BMC 网口均不可访问时，可以认为机框内所有的存储节点均被下电，那么，若该任一存储节点在该机框内，说明该任一存储节点也被下电，只要给机框上电，就可以修复该任一存储节点的故障，则认为该任一存储节点出现了预设的异常掉电故障。

第 2.2 项、在第七预设时长内，第一目标个数的存储节点的业务网口、监管网口以及 BMC 网口均不可访问，该第一目标个数的存储节点包括该任一存储节点。

当在第七预设时长内，第一目标个数的存储节点的业务网口、监管网口以及 BMC 网口均不可访问时，第一目标个数的存储节点均可以认为出现了预设的异常掉电故障。需要说明的是，本发明实施例对该第七预设时长不做具体限定。

该预设的误操作故障可以包括：该任一存储节点被主动下电。例如，当用户点击任一存储节点的关机按钮或者重启按钮时，存储节点认为被主动下电，并将主动下电的信息发送给监管节点，从而监管节点确定该任一存储节点出现了预设的误操作故障。

该预设的硬件故障包括：任一存储节点异常退出，该任一存储节点的 BMC 网口能够访问，且该任一存储节点存在松动的部件。

当任一存储节点异常退出时，可以向监管节点发送异常退出的信息，以表示自己已经异常退出。由于异常退出可能是内部的部件松动导致。该任一部件松动可以是内存条以及卡条等，对于松动的部件通过插拔的方式可以立即恢复，也就是出现的是短时故障。需要说明的是，当该任一存储节点检测到任一部件连接不良时，说明该任一存储节点存在松动的部件，则该任一存储节点可以向监管节点发送松动部件的信息，以便监管节点可以根据松动部件的信息，确定该任一存储节点出现了预设的误操作故障。

该预设的软件故障可以包括下述第 3.1-3.3 项中的任一项：

第 3.1 项、该任一存储节点的操作系统异常导致该任一存储节点异常复位。

当该任一存储节点的内存不足时，导致任一存储节点的操作系统无法继续运行，需要复位，或者是看门狗触发的该任一存储节点异常复位等，当该任一存储节点出现异常复位时，该任一存储节点可以向监管节点发送异常复位的消息，从而该监管节点可以获知该任一存储节点异常复位，则说明该任一存储节点出现了预设的软件故障。

第 3.2 项、该任一存储节点的软件异常导致该任一存储节点的目标进程退出。

该目标进程可以是 OSD 进程，当该任一存储节点出现异常复位时，该任一存储节点可以向监管节点发送目标进程退出的消息，从而该监管节点可以获知该任一存储节点的目标进程退出，则说明该任一存储节点出现了预设的软件故障。

第 3.3 项、该任一存储节点的软件异常导致该任一存储节点的操作系统复位。

由于软件异常，该任一存储节点操作系统出现复位时，该任一存储节点可以向监管节点发送操作系统复位的消息，从而该监管节点可以获知该任一存储节点的操作系统复位，则说明该任一存储节点出现了预设的软件故障。

需要说明是，当该至少一个存储节点中的每个存储节点出现故障时，监管节点就可以

通过本步骤 603 判断每个存储节点的故障类型是第一故障类型，还是第二故障类型，并将每个存储节点的故障类型存储在故障表中，以便需要监管节点需要任一存储节点的故障类型时，可以直接从故障表中获取。

需要说明的是，本步骤 603 中所体现的多种故障类型判别方法，可以精确地确定每个存储节点的故障类型，进而根据每个存储节点的故障类型确，可以更加精确的确定分布式存储系统的故障状态。

604、当该至少一个存储节点的数目大于该分布式存储系统的冗余度，且该目标数据的数据量符合预设条件时，若该故障场景为该第一故障场景，该监管节点则根据该至少一个存储节点中每一个存储节点的故障类型，确定该故障状态。

由于第一故障场景用于表示该至少一个存储节点同时出现，则该监管节点可以根据每一个存储节点，来确定该分布式存储系统的故障状态。

在一种可能的实现方式中，监管节点则根据该至少一个存储节点中每一个存储节点的故障类型，确定该故障状态可以包括：当该至少一个存储节点中每一个存储节点的故障类型均为第一故障类型时，该监管节点将该故障状态确定为该第一故障状态，该第一故障状态用于指示该至少一个存储节点能在所述第一预设时长内全部被修复；当该至少一个存储节点中目标个数的存储节点的故障类型为第二故障类型时，若该目标个数小于或者等于该分布式存储系统的冗余度时，该监管节点将该故障状态确定为该第一故障状态，否则，将该故障状态确定为该第二故障状态，该第二故障状态用于指示所述至少一个存储节点不能在所述第一预设时长内全部被修复。

当该至少一个存储节点中每一个存储节点的故障类型均为第一故障类型时，可以认为分布式存储系统内的故障可以在短时间内被修复，则可以将该故障状态确定为第一故障状态。虽然目标个数的存储节点的故障类型为第二故障类型，但是当目标个数小于或等于该分布式存储系统的冗余度，说明目标个数的存储节点对分布式存储系统的影响不大，则可以将该故障状态确定为第一故障状态。一旦目标个数大于该分布式存储系统的冗余度，说明目标个数的存储节点对分布式存储系统的影响较大，则可以将该故障状态确定为第二故障状态，以便迅速修复故障。

需要说明的是，对于该至少一个存储节点的数目是否大于该分布式存储系统的冗余度，且该目标数据的数据量是否符合预设条件的描述，在前文中有体现，对此本发明实施例不做赘述。

需要说明的是，当故障场景为第一故障场景时，监管节点可以通过步骤 603 中预设的网络故障，预设的异常掉电故障或预设的误操作故障中的任一预设故障，判断每个存储节点的故障类型。

605、当该至少一个存储节点的数目大于该分布式存储系统的冗余度，且该目标数据的数据量符合预设条件时，若该故障场景为第二故障场景，监管节点则根据该至少一个存储节点中最后一个出现故障的第一存储节点的故障类型，确定该故障状态。

由于第二故障场景用于表示该至少一个存储节点出现故障的时间不同，则该监管节点可以根据该至少与一个存储节点中最后一个出现故障的存储节点的故障类型，来确定该分布式存储系统的故障状态，其中，最后一个出现故障的存储节点也即是第一存储节点。

在一种可能的实现方式中，当该第一存储节点的故障类型为第一故障类型时，监管节点则将该故障状态确定为第一故障状态，该第一故障状态用于指示该至少一个存储节点能在该第一预设时长内全部被修复；当该第一存储节点的故障类型为所述第二故障类型时，监管节点则将该故障状态确定为第二故障状态，该第二故障状态用于指示该至少一个存储节点不能在所述第一预设时长内全部被修复。

需要说明的是，本步骤 604 和 605 所示的过程也即是根据所述分布式存储系统的故障场景，确定所述分布式存储系统的故障状态的过程。

需要说明的是，当故障场景为第二故障场景时，监管节点仅需要确定第一存储节点的故障类型即可，无需确定所以存储节点的故障类型，可以通过步骤 603 中预设的网络故障，预设的异常掉电故障、预设的误操作故障、预设的硬件故障或预设的软件故障中的任一预设故障，来判断第一存储节点的故障类型。

606、该监管节点向该分布式存储系统内的所有存储节点发送用于指示故障状态的故障标识。

当该故障状态为第一故障状态时，该监管节点向该分布式存储系统内的所有存储节点发送第一故障标识，当该故障状态为第二故障状态时，向该监管节点向该分布式存储系统内的所有存储节点发送第二故障标识，具体执行过程与步骤 503 同理，在此不做赘述。需要说明的是，本步骤 606 所示的过程也即是向所述多个存储节点中每一个存储节点发送所述故障状态的过程。

607、该分布式存储系统内的目标存储节点接收故障标识。

当该故障标识为第一故障标识时，目标存储节点接收到第一故障标识，当该故障标识为第二故障标识时，目标存储节点接收到第二故障标识，具体执行过程与步骤 504 同理，在此不做赘述。

608、目标设备向该目标存储节点发送访问请求。

本步骤 608 与步骤 505 所示的过程同理，本发明实施例对本步骤 608 不做赘述。

609、当该分布式存储系统内的目标存储节点接收到该故障标识后，若该目标存储节点再接收到访问请求，该目标存储节点悬挂该访问请求，输出该故障标识。

当该故障标识为第一故障标识时，目标存储节点向目标设备输出第一故障标识，当该故障标识为第二故障标识时，目标存储节点向目标设备输出第二故障标识，具体执行过程与步骤 506 同理，在此不做赘述。

610、目标设备接收该目标存储节点基于该访问请求返回的故障标识。

本步骤 610 与步骤 505 所示的过程同理，本发明实施例对本步骤 610 不做赘述。需要说明的是，本步骤 610 所示的过程也即是接收所述目标存储节点返回的响应；所述响应中包含所述分布式存储系统的故障状态；所述故障状态用于指示至少一个出现故障的存储节点能否在第一预设时长内全部被修复的过程。其中，所述响应中包含的故障状态也即是故障标识。

611、目标设备基于接收的故障标识，进行故障处理。

当该故障标识为第一故障标识时，目标设备基于接收的第一故障标识，进行故障处理，具体执行过程与步骤 508 所示的过程同理。当该故障标识为第二故障标识时，目标设备基

于接收的第二故障标识，进行故障处理，具体执行过程与步骤 516 所示的过程同理，在此，本发明实施例对本步骤 611 不做赘述。

需要说明的是，分布式存储系统内的每个存储节点均知道分布式存储系统的故障状态，对于未出现故障的存储节点而言，可以基于访问请求向目标设备返回故障状态，从而目标设备可以明确的知道分布式存储系统的故障状态，进而可以提高目标设备确定故障状态的精度。

需要说明的是，对于不同的虚拟机，本发明实施例提供了不同的故障处理方式，从而使得该本发明实施例所提供的故障处理方式更具有普适性。

需要说明的是，本步骤 611 所示的过程也即是基于所述响应中包含的故障状态，进行故障处理的过程。

612、当该至少一个存储节点修复完成时，监管节点向该分布式存储系统内的各个设备发送修复完成响应。

本步骤 612 与步骤 509 同理，在此本发明实施例对本步骤 612 不做赘述。需要说明的是，当该故障状态为第一故障状态时，若在该第一预设时长内该至少一个存储节点全部被修复完整，则可以直接执行本步骤 612，若在该第一预设时长内该至少一个存储节点未全部修复，则监管节点将该故障状态由第一故障状态更新为第二故障状态，并跳转执行步骤 606。需要说明的是，在第一故障状态下，客户端不会下电文件系统，一旦在第一预设时间内该至少一个存储节点均能被修复，那么也就可以避免下电文件系统，从而可以减少修复文件系统的时间，使得分布式存储系统可以尽快恢复业务，以保证服务质量。

本发明实施例所示的方法，根据所述多个存储节点中的至少一个出现故障的存储节点，确定所述分布式存储系统的故障状态，从而无需当所有存储节点均故障时，才确定分布式存储系统的故障状态，当确定完故障状态后，可以立即向分布式存储系统内的每个存储节点发送故障状态，以便每个存储节点根据确定的故障状态进行故障处理，从而可以降低分布式存储系统恢复正常所用的时间。并且，对于不同的虚拟机，本发明实施例提供了不同的故障处理方式，从而使得该本发明实施例所提供的故障处理方式更具有普适性。并且，分布式存储系统内的每个存储节点均知道分布式存储系统的故障状态，对于未出现故障的存储节点而言，可以基于访问请求向目标设备返回故障状态，从而目标设备可以明确的知道分布式存储系统的故障状态，进而可以提高目标设备确定故障状态的精度。并且，在第一故障状态下，目标设备不会下电文件系统，一旦在第一预设时间内该至少一个存储节点均能被修复，那么也就可以避免下电文件系统，从而可以减少修复文件系统的时间，使得分布式存储系统可以尽快恢复业务，以保证服务质量。并且，本步骤 603 中所体现的多种故障类型判别方法，可以精确地确定每个存储节点的故障类型，进而根据每个存储节点的故障类型确，可以更加精确的确定分布式存储系统的故障状态。

图 7 是本发明实施例提供的一种故障处理装置的结构示意图，应用于分布式存储系统，所述分布式存储系统包含多个存储节点，该装置包括：

确定模块 701，用于根据所述多个存储节点中的至少一个出现故障的存储节点，确定所述分布式存储系统的故障状态；所述故障状态用于指示所述至少一个出现故障的存储节

点能否在第一预设时长内全部被修复；

发送模块 702，用于向所述多个存储节点中每一个存储节点发送所述故障状态。

可选地，所述装置还包括：

处理模块，用于向所述多个存储节点中每一个存储节点发送所述故障状态。

可选地，所述确定模块 701，所述确定模块包括：

第一确定单元，用于执行上述步骤 501；

第二确定单元，用于根据所述至少一个存储节点以及所述目标数据，确定所述故障状态。

可选地，所述第二确定单元用于：

当所述至少一个存储节点的数目大于所述分布式存储系统的冗余度，且所述目标数据的数据量符合预设条件时，将所述故障状态确定为第一故障状态，所述第一故障状态用于指示所述至少一个存储节点能在所述第一预设时长内全部被修复。

可选地，所述第二确定单元用于：

当所述至少一个存储节点的数目大于所述分布式存储系统的冗余度，且所述目标数据的数据量符合预设条件时，根据所述分布式存储系统的故障场景，确定所述故障状态，所述故障场景用于指示所述至少一个存储节点是否同时出现故障。

可选地，所述预设条件包括下述任一项：

所述目标数据的数据量与第一预设数据量之间的比值大于预设比值，所述第一预设数据量为所述分布式存储系统存储的所有数据的总数据量；

所述目标数据的数据量大于第二预设数据量。

可选地，所述装置还包括：

更新模块，用于当所述故障状态为所述第一故障状态时，若在所述第一预设时长内所述至少一个存储节点未全部被修复，则将所述故障状态由第一故障状态更新为第二故障状态，所述第二故障状态用于指示所述至少一个存储节点不能在所述第一预设时长内全部被修复。

可选地，所述第二确定单元，用于执行上述步骤 602。

可选地，所述第二确定单元，用于当所述至少一个存储节点在目标时长内均出现故障时，将所述故障场景确定为第一故障场景，否则，将所述故障场景确定为第二故障场景，所述第一故障场景用于指示所述至少一个存储节点同时出现故障，所述第二故障场景用于指示所述至少一个存储节点出现故障的时间不同。

可选地，所述第二确定单元包括：

第一确定子单元，用于执行上述步骤 604；

第二确定子单元，用于执行上述步骤 605。

可选地，所述第一确定子单元用于：

当所述至少一个存储节点中每一个存储节点的故障类型均为第一故障类型时，将所述故障状态确定为第一故障状态，所述第一故障类型用于指示一个存储节点的故障能在所述第二预设时长内被修复，所述第一故障状态用于指示所述至少一个存储节点能在所述第一预设时长内全部被修复；

当所述至少一个存储节点中目标个数的存储节点的故障类型为第二故障类型时，若所述目标个数小于或等于所述分布式存储系统的冗余度，将所述故障状态确定为所述第一故障状态，否则，将所述故障状态确定为第二故障状态，所述第二故障类型用于指示一个存储节点的故障不能在所述第二预设时长内被修复，所述第二故障状态用于指示所述至少一个存储节点不能在所述第一预设时长内全部被修复。

可选地，所述第二确定子单元用于：

当所述第一存储节点的故障类型为第一故障类型时，将所述故障状态确定为第一故障状态，所述第一故障类型用于指示一个存储节点的故障能在所述第二预设时长内被修复，所述第一故障状态用于指示所述至少一个存储节点能在所述第一预设时长内全部被修复；

当所述第一存储节点的故障类型为第二故障类型时，则将所述故障状态确定为第二故障状态，所述第二故障类型用于指示一个存储节点的故障不能在所述第二预设时长内被修复，所述第二故障状态用于指示所述至少一个存储节点不能在所述第一预设时长内全部被修复。

可选地，所述确定模块 701，还用于执行步骤 603。

可选地，所述发送模块 702，还用于执行上述步骤 509。

图 8 是本发明实施例提供的一种故障处理装置的结构示意图，所述分布式存储系统包含多个存储节点；该装置包括：

发送模块 801，用于执行上述步骤 608；

接收模块 802，用于接收所述目标存储节点返回的响应；所述响应中包含所述分布式存储系统的故障状态；所述故障状态用于指示至少一个出现故障的存储节点能否在第一预设时长内全部被修复。

可选地，该装置还包括：

处理模块，用于基于所述响应中包含的故障状态，进行故障处理。

可选地，所述故障状态的故障标识包括第一故障标识或第二故障标识中的任一个，其中，所述第一故障标识用于指示第一故障状态，所述第二故障标识用于指示第二故障状态，所述第一故障状态用于指示至少一个存储节点能在第一预设时长内全部被修复，所述第二故障状态用于指示所述至少一个存储节点不能在所述第一预设时长内全部被修复，所述存储节点为所述分布式存储系统中出现故障的存储节点。

可选地，所述处理模块用于：

当所述访问请求由所述分布式存储系统中的目标客户端基于目标虚拟机发送，且所述目标虚拟机是 VMWare 虚拟机时，若所述故障状态为所述第一故障状态，不向所述目标虚拟机响应所述访问请求，所述第一故障状态用于指示至少一个存储节点能在第一预设时长内全部被修复；

当所述访问请求由所述目标客户端基于目标虚拟机发送，且所述目标虚拟机是 VMWare 虚拟机时，若所述故障状态为所述第二故障状态时，向所述目标虚拟机返回存储异常的消息，所述第二故障状态用于指示所述至少一个存储节点不能在所述第一预设时长内全部被修复。

可选地，所述处理模块用于：

当所述目标访问请求由所述分布式存储系统中的目标客户端基于目标虚拟机发送，且所述目标虚拟机不是 VMWare 虚拟机时，若所述故障状态为所述第一故障状态，向所述目标虚拟机发送重试请求，所述重试请求用于指示重新下发所述访问请求，所述第一故障状态用于指示至少一个存储节点能在第一预设时长内全部被修复；

当所述目标访问请求由所述目标客户端基于目标虚拟机发送，且所述目标虚拟机不是 VMWare 虚拟机时，若所述故障状态为所述第二故障状态，向所述目标虚拟机返回所述目标虚拟机可识别的目标错误，所述目标错误用于指示存储介质故障，所述第二故障状态用于指示所述至少一个存储节点不能在所述第一预设时长内全部被修复。

可选地，所述接收模块 802，还用于接收所述分布式存储系统中的目标客户端发送的目标访问请求，所述目标访问请求用于指示对第一目标数据进行处理，所述第一目标数据包括所述目标数据；

所述发送模块 801，用于基于所述目标访问请求，向分布式存储系统内的目标存储节点发送所述访问请求。

可选地，接收模块 802，用于接收目标存储节点返回的修复完成响应，所述修复完成响应用于指示所述分布式存储系统内没有故障设备。

本发明实施例还提供一种分布式存储系统，所述分布式存储系统包括监管节点和多个存储节点；

所述监管节点用于：

根据所述多个存储节点中的至少一个出现故障的存储节点确定所述分布式存储系统的故障状态；所述故障状态用于指示所述至少一个出现故障的存储节点能否在第一预设时长内全部被修复；

向所述多个存储节点中每一个存储节点发送所述故障状态；

所述多个存储节点中的每一个存储节点，用于接收所述故障状态。

可选地，所述故障状态的故障标识包括第一故障标识和第二故障标识中的任一个，所述第一故障标识用于指示所述第一故障状态，所述第二故障标识用于指示第二故障状态，所述第一故障状态用于指示所述至少一个存储节点能在所述第一预设时长内全部被修复，所述第二故障状态用于指示所述至少一个存储节点不能在所述第一预设时长内全部被修复。

可选地，所述多个存储节点中的每一个存储节点，还用于当接收到所述故障标识后，若再接收到所述访问请求，悬挂所述访问请求，基于接收的故障状态，进行故障处理。

需要说明的上述提供的分布式存储系统内的各个设备均可以是实施例 5 和 6 中的设备。

上述所有可选技术方案，可以采用任意结合形成本公开的可选实施例，在此不再一一赘述。

需要说明的是：上述实施例提供的故障处理装置在处理故障时，仅以上述各功能模块的划分进行举例说明，实际应用中，可以根据需要而将上述功能分配由不同的功能模块完成，即将装置的内部结构划分成不同的功能模块，以完成以上描述的全部或者部分功能。

另外，上述实施例提供的分布式存储系统故障处理方法实施例属于同一构思，其具体实现过程详见方法实施例，这里不再赘述。

本领域普通技术人员可以理解实现上述实施例的全部或部分步骤可以通过硬件来完成，也可以通过程序来指令相关的硬件完成，所述的程序可以存储于一种计算机可读存储介质中，上述提到的存储介质可以是只读存储器，磁盘或光盘等。

以上所述仅为本发明的较佳实施例，并不用以限制本发明，凡在本发明的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

权 利 要 求 书

1、一种分布式存储系统故障处理方法，其特征在于，所述分布式存储系统包含多个存储节点；所述方法包括：

根据所述多个存储节点中的至少一个出现故障的存储节点，确定所述分布式存储系统的故障状态；所述故障状态用于指示所述至少一个出现故障的存储节点能否在第一预设时长内全部被修复；

向所述多个存储节点中每一个存储节点发送所述故障状态。

2、根据权利要求1所述的方法，其特征在于，所述方法还包括：

根据所述存储系统的故障状态，进行故障处理。

3、根据权利要求1所述的方法，其特征在于，所述根据所述多个存储节点中的至少一个出现故障的存储节点，确定所述分布式存储系统的故障状态包括：

确定所述分布式存储系统内出现故障的至少一个存储节点以及所述至少一个存储节点内无法被访问的目标数据；

根据所述至少一个存储节点以及所述目标数据，确定所述故障状态。

4、根据权利要求3所述的方法，其特征在于，所述根据所述至少一个存储节点以及所述目标数据，确定所述故障状态包括：

当所述至少一个存储节点的数目大于所述分布式存储系统的冗余度，且所述目标数据的数据量符合预设条件时，将所述故障状态确定为第一故障状态，所述第一故障状态用于指示所述至少一个存储节点能在所述第一预设时长内全部被修复。

5、根据权利要求3所述的方法，其特征在于，所述根据所述至少一个存储节点以及所述目标数据，确定所述故障状态包括：

当所述至少一个存储节点的数目大于所述分布式存储系统的冗余度，且所述目标数据的数据量符合预设条件时，根据所述分布式存储系统的故障场景，确定所述故障状态，所述故障场景用于指示所述至少一个存储节点是否同时出现故障。

6、一种分布式存储系统故障处理方法，其特征在于，所述分布式存储系统包含多个存储节点；所述方法包括：

向所述分布式存储系统中的目标存储节点发送访问请求；

接收所述目标存储节点返回的响应；所述响应中包含所述分布式存储系统的故障状态；所述故障状态用于指示至少一个出现故障的存储节点能否在第一预设时长内全部被修复。

7、根据权利要求6所述的方法，其特征在于，所述接收所述目标存储节点返回的响应之后，所述方法还包括：

基于所述响应中包含的故障状态，进行故障处理。

8、根据权利要求 7 所述的方法，其特征在于，所述基于所述响应中包含的故障状态，进行故障处理包括：

当所述访问请求由所述分布式存储系统中的目标客户端基于目标虚拟机发送，且所述目标虚拟机是 VMWare 虚拟机时，若所述故障状态为所述第一故障状态，不向所述目标虚拟机响应所述访问请求，所述第一故障状态用于指示至少一个存储节点能在第一预设时长内全部被修复；

当所述访问请求由所述目标客户端基于目标虚拟机发送，且所述目标虚拟机是 VMWare 虚拟机时，若所述故障状态为所述第二故障状态时，向所述目标虚拟机返回存储异常的消息，所述第二故障状态用于指示所述至少一个存储节点不能在所述第一预设时长内全部被修复。

9、根据权利要求 7 所述的方法，其特征在于，所述基于所述响应中包含的故障状态，进行故障处理包括：

当所述访问请求由所述分布式存储系统中的目标客户端基于目标虚拟机发送，且所述目标虚拟机不是 VMWare 虚拟机时，若所述故障状态为所述第一故障状态，向所述目标虚拟机发送重试请求，所述重试请求用于指示重新下发所述访问请求，所述第一故障状态用于指示至少一个存储节点能在第一预设时长内全部被修复；

当所述访问请求由所述目标客户端基于目标虚拟机发送，且所述目标虚拟机不是 VMWare 虚拟机时，若所述故障状态为所述第二故障状态，向所述目标虚拟机返回所述目标虚拟机可识别的目标错误，所述目标错误用于指示存储介质故障，所述第二故障状态用于指示所述至少一个存储节点不能在所述第一预设时长内全部被修复。

10、一种分布式存储系统，其特征在于，所述分布式存储系统包括监管节点和多个存储节点；

所述监管节点用于：

根据所述多个存储节点中的至少一个出现故障的存储节点确定所述分布式存储系统的故障状态；所述故障状态用于指示所述至少一个出现故障的存储节点能否在第一预设时长内全部被修复；

向所述多个存储节点中每一个存储节点发送所述故障状态；

所述多个存储节点中的每一个存储节点，用于接收所述故障状态。

11、根据权利要求 10 所述的系统，其特征在于，所述多个存储节点中的每一个存储节点，还用于当接收到所述故障标识后，若再接收到所述访问请求，悬挂所述访问请求，基于接收的故障状态，进行故障处理。

12、一种故障处理装置，其特征在于，应用于分布式存储系统，所述分布式存储系统

包含多个存储节点，所述装置包括：

确定模块，用于根据所述多个存储节点中的至少一个出现故障的存储节点，确定所述分布式存储系统的故障状态；所述故障状态用于指示所述至少一个出现故障的存储节点能否在第一预设时长内全部被修复；

发送模块，用于向所述多个存储节点中每一个存储节点发送所述故障状态。

13、根据权利要求 12 所述的装置，其特征在于，所述装置还包括：

处理模块，用于向所述多个存储节点中每一个存储节点发送所述故障状态。

14、根据权利要求 12 所述的装置，其特征在于，所述确定模块包括：

第一确定单元，用于确定所述分布式存储系统内出现故障的至少一个存储节点以及所述至少一个存储节点内无法被访问的目标数据；

第二确定单元，用于根据所述至少一个存储节点以及所述目标数据，确定所述故障状态。

15、根据权利要求 14 所述的装置，其特征在于，所述第二确定单元用于：

当所述至少一个存储节点的数目大于所述分布式存储系统的冗余度，且所述目标数据的数据量符合预设条件时，将所述故障状态确定为第一故障状态，所述第一故障状态用于指示所述至少一个存储节点能在所述第一预设时长内全部被修复。

16、根据权利要求 14 所述的装置，其特征在于，所述第二确定单元用于：

当所述至少一个存储节点的数目大于所述分布式存储系统的冗余度，且所述目标数据的数据量符合预设条件时，根据所述分布式存储系统的故障场景，确定所述故障状态，所述故障场景用于指示所述至少一个存储节点是否同时出现故障。

17、一种故障处理装置，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含多个存储节点，所述装置包括：

发送模块，用于向所述分布式存储系统中的目标存储节点发送访问请求，所述分布式存储系统包含多个存储节点；

接收模块，用于接收所述目标存储节点返回的响应；所述响应中包含所述分布式存储系统的故障状态；所述故障状态用于指示至少一个出现故障的存储节点能否在第一预设时长内全部被修复。

18、根据权利要求 17 所述的装置，其特征在于，所述装置还包括：

处理模块，用于基于所述响应中包含的故障状态，进行故障处理。

19、根据权利要求 18 所述的装置，其特征在于，所述处理模块用于：

当所述访问请求由所述分布式存储系统中的目标客户端基于目标虚拟机发送，且所述

目标虚拟机是 VMWare 虚拟机时，若所述故障状态为所述第一故障状态，不向所述目标虚拟机响应所述访问请求，所述第一故障状态用于指示至少一个存储节点能在第一预设时长内全部被修复；

当所述访问请求由所述目标客户端基于目标虚拟机发送，且所述目标虚拟机是 VMWare 虚拟机时，若所述故障状态为所述第二故障状态时，向所述目标虚拟机返回存储异常的消息，所述第二故障状态用于指示所述至少一个存储节点不能在所述第一预设时长内全部被修复。

20、根据权利要求 18 所述的装置，其特征在于，所述处理模块用于：

当所述目标访问请求由所述分布式存储系统中的目标客户端基于目标虚拟机发送，且所述目标虚拟机不是 VMWare 虚拟机时，若所述故障状态为所述第一故障状态，向所述目标虚拟机发送重试请求，所述重试请求用于指示重新下发所述访问请求，所述第一故障状态用于指示至少一个存储节点能在第一预设时长内全部被修复；

当所述目标访问请求由所述目标客户端基于目标虚拟机发送，且所述目标虚拟机不是 VMWare 虚拟机时，若所述故障状态为所述第二故障状态，向所述目标虚拟机返回所述目标虚拟机可识别的目标错误，所述目标错误用于指示存储介质故障，所述第二故障状态用于指示所述至少一个存储节点不能在所述第一预设时长内全部被修复。

21、一种计算机设备，其特征在于，所述计算机设备包括处理器和存储器，所述存储器中存储有至少一条指令，所述指令由所述处理器加载并执行以实现如权利要求 1 至权利要求 9 任一项所述的分布式存储系统故障处理方法所执行的操作。

22、一种存储介质，其特征在于，所述存储介质中存储有至少一条指令，所述指令由处理器加载并执行以实现如权利要求 1 至权利要求 9 任一项所述的分布式存储系统故障处理方法所执行的操作。

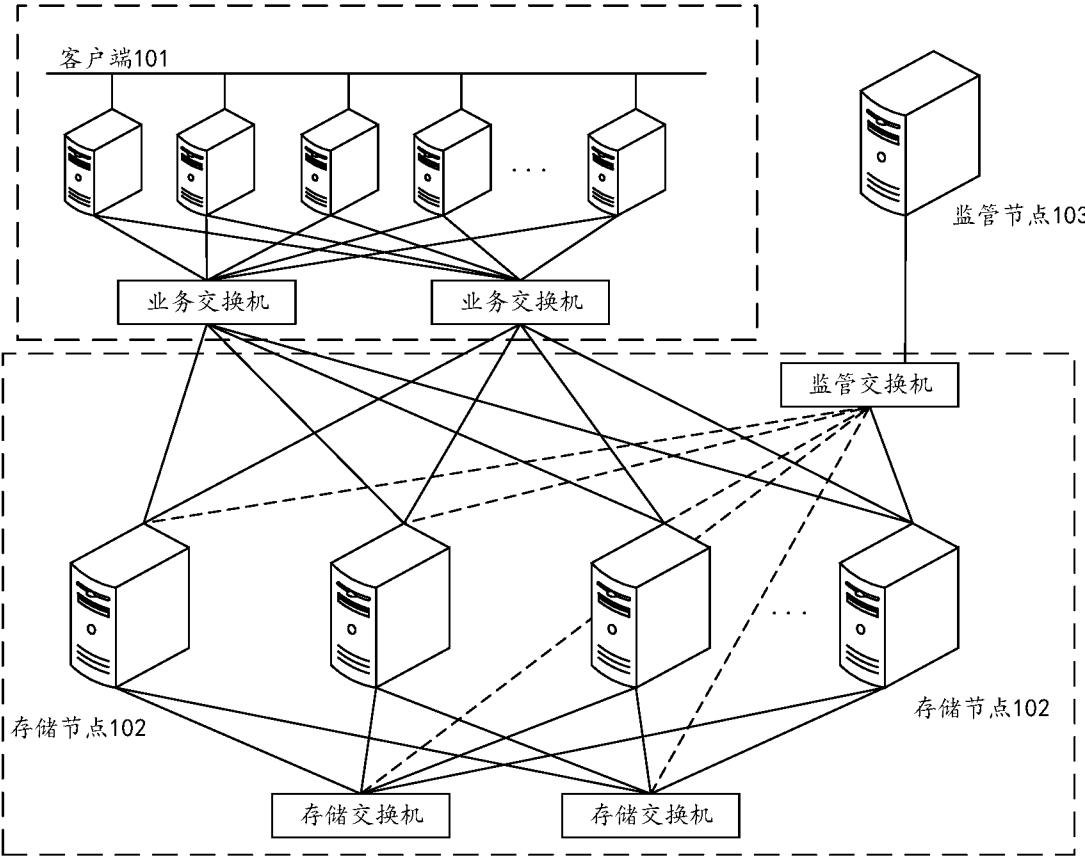


图 1

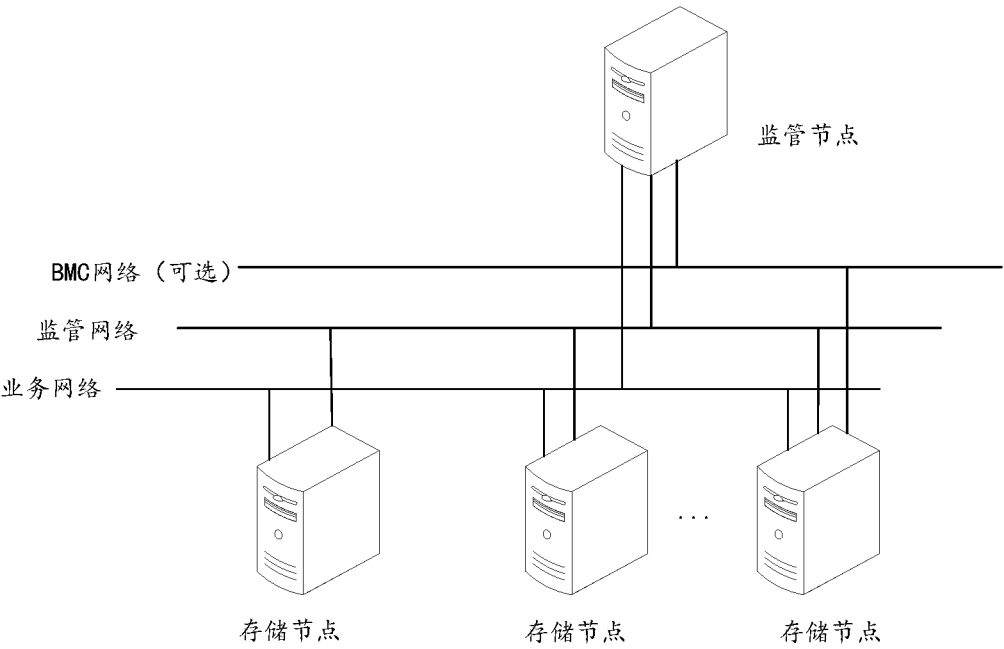


图 2

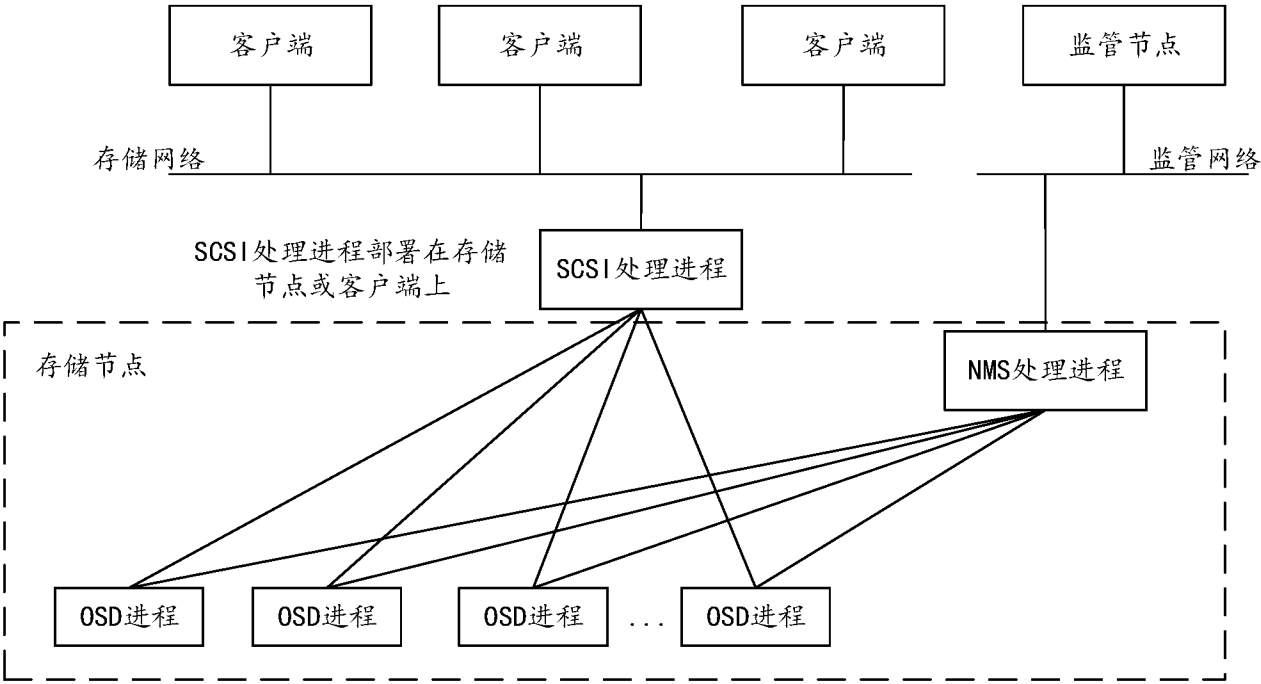


图 3

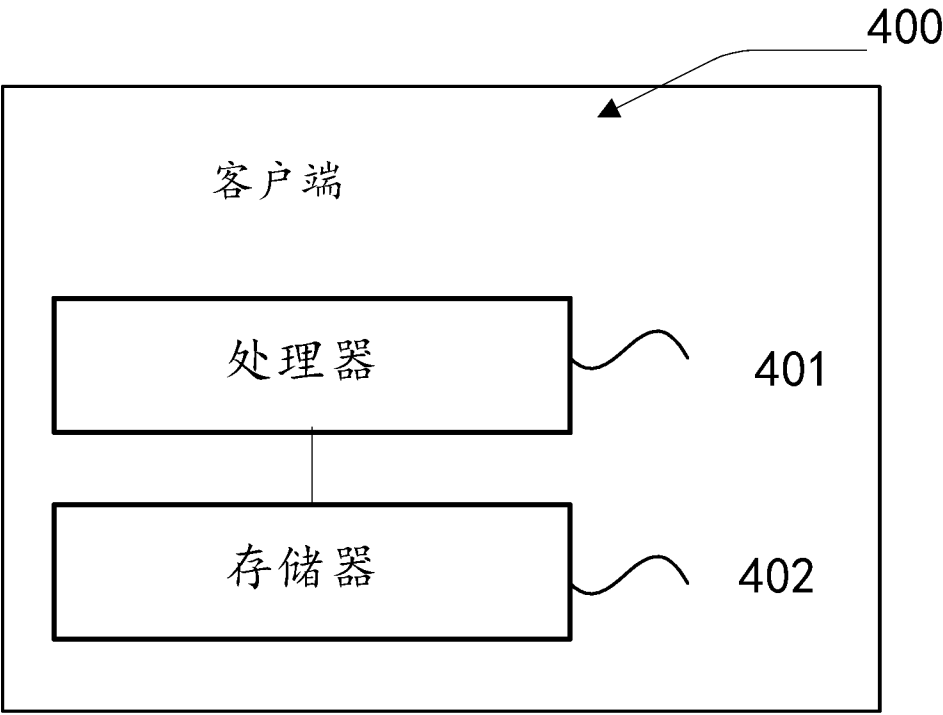


图 4

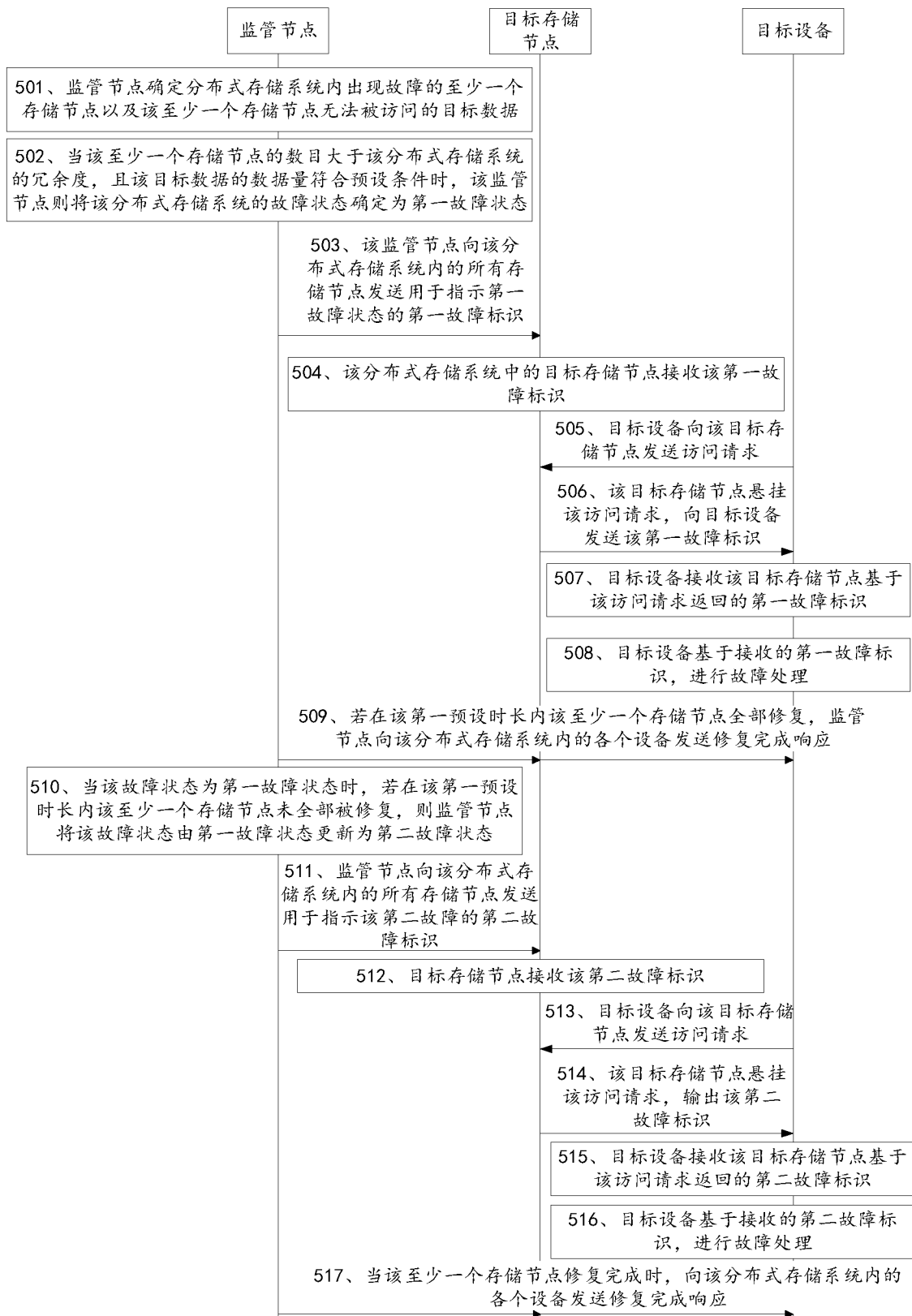


图 5



图 6

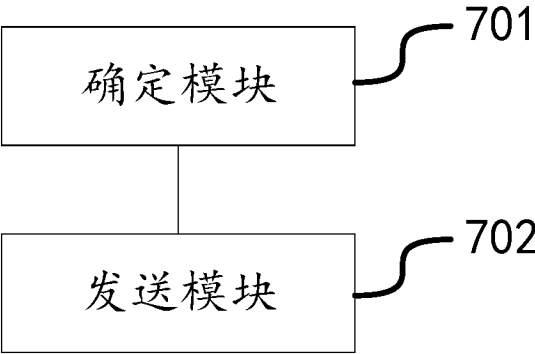


图 7

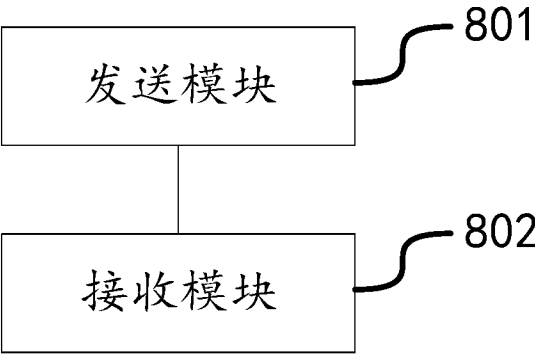


图 8

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/102302

A. CLASSIFICATION OF SUBJECT MATTER

H04L 12/24(2006.01)i; H04L 29/08(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

H04L; G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS; CNTXT; CNKI; VEN; USTXT; EPTXT; WOTXT; JPTXT: 分布式存储, 分布式, 存储, 节点, 故障, 修复, 恢复, 时间, TND, PND, APD, TND, distributed storage, node, fault, failure, repair, recover, time

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 110535692 A (HUAWEI TECHNOLOGIES CO., LTD.) 03 December 2019 (2019-12-03) claims 1-22	1-22
X	CN 104935481 A (HUAZHONG UNIVERSITY OF SCIENCE AND TECHNOLOGY) 23 September 2015 (2015-09-23) description paragraphs [0045]-[0056] and figures 1-4	1-2, 6-13, 17-22
Y	CN 104935481 A (HUAZHONG UNIVERSITY OF SCIENCE AND TECHNOLOGY) 23 September 2015 (2015-09-23) description paragraphs [0045]-[0056] and figures 1-4	3-5, 14-16
Y	CN 103092712 A (ALIBABA GROUP HOLDING LIMITED) 08 May 2013 (2013-05-08) description paragraphs [0023]-[0062], [0109]	3-5, 14-16
A	CN 109831342 A (JIANGSU HUIZHIDA INFORMATION TECHNOLOGY CO., LTD.) 31 May 2019 (2019-05-31) entire document	1-22
A	US 2019004845 A1 (VMWARE INC.) 03 January 2019 (2019-01-03) entire document	1-22

☐ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

09 September 2020

Date of mailing of the international search report

24 September 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2020/102302

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	110535692	A	03 December 2019	None			
CN	104935481	A	23 September 2015	CN	104935481	B	09 March 2018
CN	103092712	A	08 May 2013	HK	1180068	A1	13 January 2017
				CN	103092712	B	30 March 2016
				HK	1180068	A0	11 October 2013
CN	109831342	A	31 May 2019	None			
US	2019004845	A1	03 January 2019	US	10691479	B2	23 June 2020
				IN	201741022624	A	04 January 2019

国际检索报告

国际申请号

PCT/CN2020/102302

A. 主题的分类 H04L 12/24(2006.01)i; H04L 29/08(2006.01)i 按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类																							
B. 检索领域 检索的最低限度文献(标明分类系统和分类号) H04L; G06F 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用)) CNABS;CNTXT;CNKI;VEN;USTXT;EPTXT;WOTXT;JPTXT: 分布式存储, 分布式, 存储, 节点, 故障, 修复, 恢复, 时间, TND, PND, APD, TND, distributed storage, node, fault, failure, repair, recover, time																							
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>PX</td> <td>CN 110535692 A (华为技术有限公司) 2019年 12月 3日 (2019 - 12 - 03) 权利要求1-22</td> <td>1-22</td> </tr> <tr> <td>X</td> <td>CN 104935481 A (华中科技大学) 2015年 9月 23日 (2015 - 09 - 23) 说明书第[0045]-[0056]段及图1-4</td> <td>1-2、6-13、17-22</td> </tr> <tr> <td>Y</td> <td>CN 104935481 A (华中科技大学) 2015年 9月 23日 (2015 - 09 - 23) 说明书第[0045]-[0056]段及图1-4</td> <td>3-5、14-16</td> </tr> <tr> <td>Y</td> <td>CN 103092712 A (阿里巴巴集团控股有限公司) 2013年 5月 8日 (2013 - 05 - 08) 说明书第[0023]-[0062]、[0109]段</td> <td>3-5、14-16</td> </tr> <tr> <td>A</td> <td>CN 109831342 A (江苏汇智达信息科技有限公司) 2019年 5月 31日 (2019 - 05 - 31) 全文</td> <td>1-22</td> </tr> <tr> <td>A</td> <td>US 2019004845 A1 (VMWARE INC) 2019年 1月 3日 (2019 - 01 - 03) 全文</td> <td>1-22</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	PX	CN 110535692 A (华为技术有限公司) 2019年 12月 3日 (2019 - 12 - 03) 权利要求1-22	1-22	X	CN 104935481 A (华中科技大学) 2015年 9月 23日 (2015 - 09 - 23) 说明书第[0045]-[0056]段及图1-4	1-2、6-13、17-22	Y	CN 104935481 A (华中科技大学) 2015年 9月 23日 (2015 - 09 - 23) 说明书第[0045]-[0056]段及图1-4	3-5、14-16	Y	CN 103092712 A (阿里巴巴集团控股有限公司) 2013年 5月 8日 (2013 - 05 - 08) 说明书第[0023]-[0062]、[0109]段	3-5、14-16	A	CN 109831342 A (江苏汇智达信息科技有限公司) 2019年 5月 31日 (2019 - 05 - 31) 全文	1-22	A	US 2019004845 A1 (VMWARE INC) 2019年 1月 3日 (2019 - 01 - 03) 全文	1-22
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																					
PX	CN 110535692 A (华为技术有限公司) 2019年 12月 3日 (2019 - 12 - 03) 权利要求1-22	1-22																					
X	CN 104935481 A (华中科技大学) 2015年 9月 23日 (2015 - 09 - 23) 说明书第[0045]-[0056]段及图1-4	1-2、6-13、17-22																					
Y	CN 104935481 A (华中科技大学) 2015年 9月 23日 (2015 - 09 - 23) 说明书第[0045]-[0056]段及图1-4	3-5、14-16																					
Y	CN 103092712 A (阿里巴巴集团控股有限公司) 2013年 5月 8日 (2013 - 05 - 08) 说明书第[0023]-[0062]、[0109]段	3-5、14-16																					
A	CN 109831342 A (江苏汇智达信息科技有限公司) 2019年 5月 31日 (2019 - 05 - 31) 全文	1-22																					
A	US 2019004845 A1 (VMWARE INC) 2019年 1月 3日 (2019 - 01 - 03) 全文	1-22																					
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																							
<table border="0"> <tr> <td> * 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 </td> <td> “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件 </td> </tr> </table>			* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																			
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																						
国际检索实际完成的日期 2020年 9月 9日		国际检索报告邮寄日期 2020年 9月 24日																					
ISA/CN的名称和邮寄地址 中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		受权官员 马晓晓 电话号码 86-(20)-28950427																					

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2020/102302

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	110535692	A	2019年 12月 3日	无	
CN	104935481	A	2015年 9月 23日	CN	104935481 B 2018年 3月 9日
CN	103092712	A	2013年 5月 8日	HK	1180068 A1 2017年 1月 13日
				CN	103092712 B 2016年 3月 30日
				HK	1180068 A0 2013年 10月 11日
CN	109831342	A	2019年 5月 31日	无	
US	2019004845	A1	2019年 1月 3日	US	10691479 B2 2020年 6月 23日
				IN	201741022624 A 2019年 1月 4日