

## (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

(43) 国际公布日  
2018 年 1 月 18 日 (18.01.2018)



(10) 国际公布号  
**WO 2018/010104 A1**

(51) 国际专利分类号:  
**G06F 9/50** (2006.01)

(ZHONG, Yong); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(21) 国际申请号: PCT/CN2016/089836

(22) 国际申请日: 2016 年 7 月 12 日 (12.07.2016)

(25) 申请语言: 中文

(26) 公布语言: 中文

(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: 李亿(LI, Yi); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。  
李靖(LI, Jing); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 钟颢

(74) 代理人: 北京中博世达专利商标代理有限公司 (BEIJING ZBSD PATENT&TRADEMARK AGENT LTD.); 中国北京市海淀区交大东路31号11号楼8层, Beijing 100044 (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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, 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, SM, ST, SV, SY, TH,

(54) Title: METHOD, DEVICE AND SYSTEM FOR CONTROLLING IO BANDWIDTH AND PROCESSING IO ACCESS REQUEST

(54) 发明名称: 控制IO带宽和处理IO访问请求的方法、装置及系统

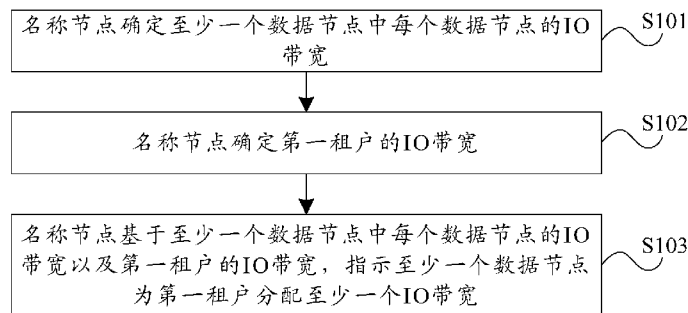


图 4

S101 A name node determines an IO bandwidth of each data node of at least one data node

S102 The name node determines an IO bandwidth of a first tenant

S103 The name node instructs, on the basis of the IO bandwidth of each data node of the at least one data node and the IO bandwidth of the first tenant, the at least one data node to allocate at least one IO bandwidth to the first tenant

(57) Abstract: The present invention relates to the technical field of storage. An embodiment of the present invention provides a method, device and system for controlling an IO bandwidth and processing an IO access request, the IO bandwidth being allocated in a more flexible way. The method for controlling an IO bandwidth is applied to a distributed file system comprising a name node and at least one data node. The method comprises: the name node determining an IO bandwidth of each data node of the at least one data node and an IO bandwidth of a first tenant; and the name node instructing, on the basis of the IO bandwidth of each data node of the at least one data node and the IO bandwidth of the first tenant, the at least one data node to allocate at least one IO bandwidth to the first tenant, wherein the at least one IO bandwidth is in one-to-one correspondence with the at least one data node, and each of the IO bandwidths of the at least one IO bandwidth is greater than 0 and less than or equal to the IO bandwidth of a corresponding data node.

TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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, 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)。

根据细则4.17的声明:

- 关于申请人有权申请并被授予专利 (细则4.17(ii))

本国际公布:

- 包括国际检索报告 (条约第21条(3))。

---

(57) 摘要: 本发明实施例提供一种控制IO带宽和处理IO访问请求的方法、装置及系统, 涉及存储技术领域, 分配IO带宽的方式比较灵活。该控制IO带宽的方法应用于包括名称节点和至少一个数据节点的分布式文件系统。该方法包括: 名称节点确定该至少一个数据节点中每个数据节点的IO带宽和第一租户的IO带宽; 名称节点基于该至少一个数据节点中每个数据节点的IO带宽以及第一租户的IO带宽, 指示该至少一个数据节点为第一租户分配该至少一个IO带宽, 其中, 该至少一个IO带宽与该至少一个数据节点一一对应, 该至少一个IO带宽中的每个IO带宽大于0, 且小于或者等于对应的数据节点的IO带宽。

## 控制 IO 带宽和处理 IO 访问请求的方法、装置及系统

### 技术领域

本发明涉及存储技术领域，尤其涉及一种控制输入输出（input output, 简称 IO）带宽和处理 IO 访问请求的方法、装置及系统。

### 背景技术

随着存储技术的不断发展，分布式文件系统的应用越来越广泛。如图 1 所示，为一种应用访问分布式文件系统的示意图。

通常，当某个应用，例如应用 1 访问分布式文件系统时，该分布式文件系统可根据应用 1 的 IO 访问请求 1 为应用 1 分配 IO 带宽 1。当另一个应用，例如应用 2 访问分布式文件系统时，该分布式文件系统可根据应用 2 的 IO 访问请求 2 为应用 2 重新分配新的 IO 带宽 2，该 IO 带宽 2 和该 IO 带宽 1 为不同的 IO 带宽。

在上述技术方案中，分布式文件系统需要为不同的应用分配不同的 IO 带宽，IO 带宽的分配方式不够灵活。

### 发明内容

本申请提供一种控制 IO 带宽和处理 IO 访问请求的方法、装置及系统，分配 IO 带宽的方式比较灵活。

为达到上述目的，本申请采用如下技术方案：

第一方面，提供一种应用于分布式文件系统的控制 IO 带宽的方法。该分布式文件系统包括名称节点和所述至少一个数据节点。

该方法包括：名称节点确定所述至少一个数据节点中每个数据节点的 IO 带宽和第一租户的 IO 带宽；并基于所述至少一个数据节点中每个数据节点的 IO 带宽以及第一租户的 IO 带宽，指示所述至少一个数据节点为第一租户分配至少一个 IO 带宽。

其中，上述至少一个 IO 带宽与所述至少一个数据节点一一对

应，所述至少一个 IO 带宽中的每个 IO 带宽大于 0，并且小于或者等于对应的数据节点的 IO 带宽。

通过上述方法，本申请可以实现为租户分配 IO 带宽，分配 IO 带宽的方式比较灵活。

该租户的 IO 带宽为至少一个用户共享，所述至少一个用户中的每个用户对应至少一个应用，因此属于该租户的至少一个用户对应的多个应用可以共享为该租户分配的 IO 带宽，即使在不同时刻该多个应用可能不同，但该多个应用仍共享为该租户分配的 IO 带宽，因此能够提高分布式文件系统中 IO 带宽的利用率。

在第一方面的第一种可选的实现方式中，上述至少一个 IO 带宽的总和小于或者等于第一租户的 IO 带宽；或者，上述至少一个 IO 带宽的总和小于或者等于第一租户的 IO 带宽与额外带宽之和。

通过上述第一方面的第一种可选的实现方式，名称节点可以指示数据节点为第一租户分配合适的 IO 带宽，例如根据第一租户的优先级为第一租户分配 IO 带宽，从而尽可能地满足第一租户的 IO 带宽需求。

在第一方面的第二种可选的实现方式中，所述至少一个数据节点包括第一数据节点，第一数据节点在名称节点的指示下为第一租户分配的 IO 带宽是第一 IO 带宽，第一 IO 带宽是所述至少一个 IO 带宽中与第一数据节点对应的 IO 带宽。

上述名称节点基于所述至少一个数据节点中每个数据节点的 IO 带宽以及第一租户的 IO 带宽，指示所述至少一个数据节点为第一租户分配至少一个 IO 带宽之后，该方法还包括：名称节点确定第一租户分别在所述至少一个 IO 带宽中使用的 IO 带宽为至少一个使用带宽，所述至少一个使用带宽与所述至少一个 IO 带宽一一对应；且名称节点根据所述至少一个使用带宽，指示所述至少一个数据节点分别调整所述至少一个 IO 带宽中对应的 IO 带宽，或者指示第一数据节点调整第一 IO 带宽。

由于名称节点为第一租户分配所述至少一个 IO 带宽之后，第一

租户可能并不会将这所述至少一个 IO 带宽全部使用完,或者所述至少一个 IO 带宽可能并不能满足第一租户所需的 IO 带宽,或者第一租户使用所述至少一个 IO 带宽的差别较大(例如所述至少一个 IO 带宽中部分个数的 IO 带宽被第一租户全部使用,但所述至少一个 IO 带宽中其他部分个数的 IO 带宽被第一租户使用较少),因此名称节点需要根据第一租户在所述至少一个 IO 带宽中的所述至少一个使用带宽,指示相应的数据节点(可以是所述至少一个数据节点包括的一个或多个数据节点)调整为第一租户分配的 IO 带宽,如此名称节点能够合理地第一租户分配 IO 带宽。

在第一方面的第三种可选的实现方式中,上述名称节点基于所述至少一个数据节点中每个数据节点的 IO 带宽以及第一租户的 IO 带宽,指示所述至少一个数据节点为第一租户分配所述至少一个 IO 带宽之后,该方法还包括:名称节点确定第一租户分别在所述至少一个 IO 带宽中使用的 IO 带宽为至少一个使用带宽,所述至少一个使用带宽与所述至少一个 IO 带宽一一对应;且当所述至少一个 IO 带宽之和与所述至少一个使用带宽之和的差值小于或者等于第一阈值时,名称节点指示第二数据节点为第一租户分配 IO 带宽,所述至少一个数据节点不包括第二数据节点。

名称节点确定所述至少一个 IO 带宽之和与第一租户分别在所述至少一个 IO 带宽中使用的所述至少一个使用带宽之和的差值小于或者等于第一阈值时,表示第一租户实际使用的 IO 带宽已经非常接近所述至少一个数据节点为其分配的 IO 带宽(即第一租户基本全部使用了所述至少一个数据节点为其分配的 IO 带宽),此时为了优先保证第一租户的业务,使得第一租户具有充足的 IO 带宽,名称节点还可以指示分布式存储系统中除上述至少一个数据节点之外的其他数据节点(例如第二数据节点)为第一租户分配 IO 带宽,即名称节点指示增加为第一租户分配的 IO 带宽。

在第一方面的第四种可选的实现方式中,上述名称节点基于所述至少一个数据节点中每个数据节点的 IO 带宽以及第一租户的 IO

带宽,指示所述至少一个数据节点为第一租户分配至少一个 IO 带宽之后,该方法还包括:名称节点确定第一租户分别在所述至少一个 IO 带宽中使用的 IO 带宽为至少一个使用带宽,所述至少一个使用带宽与所述至少一个 IO 带宽一一对应;且当所述至少一个 IO 带宽之和与所述至少一个使用带宽之和的差值大于第一阈值时,名称节点指示所述至少一个数据节点中的第三数据节点将第三数据节点为第一租户分配的 IO 带宽调整为 0。

名称节点确定所述至少一个 IO 带宽之和与第一租户分别在所述至少一个 IO 带宽中使用的所述至少一个使用带宽之和的差值大于第一阈值时,表示第一租户实际使用的 IO 带宽远小于所述至少一个数据节点为其分配的 IO 带宽(即第一租户可能只使用了所述至少一个数据节点为其分配的 IO 带宽中的一部分带宽),此时为了节省 IO 带宽(或者将第一租户没有使用的 IO 带宽给其他租户使用),名称节点还可以指示所述至少一个数据节点中的第三数据节点将第三数据节点为第一租户分配的 IO 带宽调整为 0,即名称节点指示减少为第一租户分配的 IO 带宽。

在第一方面的第五种可选的实现方式中,上述至少一个数据节点包括第一数据节点,第一数据节点在名称节点的指示下为第一租户分配的 IO 带宽是第一 IO 带宽,第一 IO 带宽是所述至少一个 IO 带宽中与第一数据节点对应的 IO 带宽。

上述名称节点基于所述至少一个数据节点中每个数据节点的 IO 带宽以及第一租户的 IO 带宽,指示所述至少一个数据节点为第一租户分配所述至少一个 IO 带宽之后,该方法还包括:名称节点确定第一租户使用的第一 IO 带宽中的 IO 带宽;且根据第一租户使用的第一 IO 带宽中的 IO 带宽,指示第一数据节点调整第一 IO 带宽。

通过上述第一方面的第五种可选的实现方式,名称节点可以以数据节点为单位,指示相应的数据节点(例如第一数据节点)分别在数据节点内部调整数据节点为第一租户分配的 IO 带宽。如此,名称节点可以根据第一租户在各数据节点实际使用的 IO 带宽,指示对

应数据节点为第一租户提供更加合适的 IO 带宽,另外还可以尽可能地满足第一租户的 IO 带宽需求。

在第一方面的第六种可选的实现方式中, m 个租户包括第一租户, m 为正整数。

上述名称节点确定第一租户的 IO 带宽的方法包括:名称节点根据 m 个租户中每个租户的权重,从所述至少一个数据节点中每个数据节点的 IO 带宽的总和,确定第一租户的 IO 带宽。

其中,名称节点可以通过公式一确定第一租户的 IO 带宽。

上述公式一为:第一租户的 IO 带宽=第一租户的权重/由所述至少一个数据节点分配 IO 带宽的所有租户的权重之和\*所述至少一个数据节点中每个数据节点的 IO 带宽的总和。

由于名称节点通过公式一为租户分配 IO 带宽时是按照每个租户的权重分配的,因此为每个租户分配的 IO 带宽可以尽量满足每个租户对 IO 带宽的需求,并且均衡地满足每个租户对 IO 带宽的需求。

第二方面,提供又一种控制 IO 带宽的方法,该方法应用于分布式文件系统,该分布式文件系统包括至少一个数据节点,所述至少一个数据节点包括第一数据节点。该方法包括:第一数据节点确定所述至少一个数据节点中每个数据节点的 IO 带宽和第一租户的 IO 带宽,并根据所述至少一个数据节点中每个数据节点的 IO 带宽和第一租户的 IO 带宽,为第一租户分配第一 IO 带宽。第一 IO 带宽大于 0,且小于或等于第一数据节点的 IO 带宽。

使用上述方法提供的技术方案,本申请可以实现为租户分配 IO 带宽,分配 IO 带宽的方式比较灵活。

该租户的 IO 带宽为至少一个用户共享,所述至少一个用户中的每个用户对应至少一个应用,因此属于该租户的至少一个用户对应的多个应用可以共享为该租户分配的 IO 带宽,即使在不同时刻该多个应用可能不同,但该多个应用仍共享为该租户分配的 IO 带宽,因此能够提高分布式文件系统中 IO 带宽的利用率。

第二方面所示的控制 IO 带宽的方法与上述第一方面所示的控

制 IO 带宽的方法相比，由于第二方面所示的控制 IO 带宽的方法可以由各个数据节点自行为租户分配 IO 带宽，而无需在名称节点的指示下为租户分配 IO 带宽，因此可以减轻名称节点的负担，节省名称节点的资源，降低名称节点的开销。

进一步地，由于各个数据节点无需在名称节点的指示下为租户分配 IO 带宽，因此各个数据节点可以实时地自行为租户在本数据节点上分配 IO 带宽，从而还可以提高在租户级别控制 IO 带宽的实时性。

在第二方面的第一种可选的实现方式中，上述第一 IO 带宽小于或等于第一租户的 IO 带宽；或者，上述第一 IO 带宽小于或等于第一租户的 IO 带宽与额外带宽之和。

通过上述第二方面的第一种可选的实现方式，第一数据节点可以为第一租户分配合适的 IO 带宽，例如根据第一租户的优先级为第一租户分配 IO 带宽，从而尽可能地满足第一租户的 IO 带宽需求。

在第二方面的第二种可选的实现方式中，在第一数据节点为第一租户分配第一 IO 带宽之后，该方法还包括：第一数据节点确定第一租户分别在所述至少一个数据节点中使用的 IO 带宽为至少一个使用带宽；且第一数据节点根据所述至少一个使用带宽，调整第一 IO 带宽。

由于第一数据节点为第一租户分配第一 IO 带宽之后，第一租户可能并不会将第一 IO 带宽全部使用完，或者第一 IO 带宽可能并不能满足第一租户所需的 IO 带宽，或者第一租户使用第一 IO 带宽的差别较大（例如第一 IO 带宽中部分 IO 带宽被第一租户全部使用，但第一 IO 带宽中其他部分 IO 带宽被第一租户使用较少），因此第一数据节点需要根据第一租户的实际使用带宽（即所述至少一个使用带宽），调整第一数据节点为第一租户分配的第一 IO 带宽。如此第一数据节点能够合理地第一租户分配 IO 带宽。

在第二方面的第三种可选的实现方式中，在第一数据节点为第一租户分配第一 IO 带宽之后，该方法还包括：第一数据节点确定第

一租户使用的第一 IO 带宽中的 IO 带宽；且第一数据节点根据第一租户使用的第一 IO 带宽中的 IO 带宽，调整第一 IO 带宽。

第一数据节点为第一租户分配第一 IO 带宽之后，第一数据节点可以根据第一租户在第一 IO 带宽中实际使用的 IO 带宽，调整第一 IO 带宽。如此，第一数据节点能够为第一租户提供合适的 IO 带宽，另外还可以尽可能地满足第一租户的 IO 带宽需求。

在第二方面的第四种可选的实现方式中，上述第一数据节点确定第一租户的 IO 带宽的方法包括：第一数据节点根据 m 个租户中每个租户的权重，从所述至少一个数据节点中每个数据节点的 IO 带宽的总和，确定第一租户的 IO 带宽。

对于第二方面第四种可选的实现方式的具体描述和技术效果可以参见上述第一方面的第六种可选的实现方式中的相关描述，此处不再赘述。

第三方面，提供一种处理 IO 访问请求的方法，该方法应用于分布式文件系统，该分布式文件系统包括至少一个数据节点，所述至少一个数据节点包括第一数据节点。该方法包括：第一数据节点接收租户访问分布式文件系统的 IO 访问请求；并根据该 IO 访问请求，确定该 IO 访问请求对应的租户；以及根据第一数据节点已为该租户分配的 IO 带宽，确定是否延迟响应该 IO 访问请求。

本申请提供的处理 IO 访问请求的方法应用于分布式文件系统，该分布式文件系统包括至少一个数据节点，所述至少一个数据节点包括第一数据节点。该方法中第一数据节点通过接收租户访问分布式文件系统的 IO 访问请求，并根据该 IO 访问请求确定该 IO 访问请求对应的租户，然后再根据第一数据节点已为该租户分配的 IO 带宽，确定是否延迟响应该 IO 访问请求。本申请可以实现在数据节点上为租户分配 IO 带宽，进而数据节点可以根据为租户分配的 IO 带宽确定是否延迟响应与租户对应的 IO 访问请求（即本申请一定会响应 IO 访问请求，可以为立即响应或者延迟响应），因此可以提高响应 IO 访问请求的成功率。

在第三方面的第一种可选的实现方式中，上述第一数据节点根据第一数据节点已为租户分配的 IO 带宽，确定是否延迟响应 IO 访问请求的方法包括：第一数据节点确定第一数据节点已为租户分配的 IO 带宽与租户在第一数据节点使用的 IO 带宽的差值；且在确定该差值大于第三阈值时，响应该 IO 访问请求。

当第一数据节点确定第一数据节点已为租户分配的 IO 带宽与该租户在第一数据节点使用的 IO 带宽的差值大于第三阈值时，表示第一数据节点已为租户分配的 IO 带宽中还有部分空闲 IO 带宽，即第一数据节点有充足的 IO 带宽响应该租户对分布式文件系统的 IO 访问请求，那么第一数据节点可以立即响应该 IO 访问请求。

在第三方面的第二种可选的实现方式中，上述第一数据节点根据第一数据节点已为租户分配的 IO 带宽，确定是否延迟响应 IO 访问请求的方法包括：第一数据节点确定第一数据节点已为租户分配的 IO 带宽与租户在第一数据节点使用的 IO 带宽的差值；且在确定该差值小于或者等于第三阈值时，延迟响应该 IO 访问请求。

当第一数据节点确定第一数据节点已为租户分配的 IO 带宽与该租户在第一数据节点使用的 IO 带宽的差值小于或者等于第三阈值时，表示第一数据节点已为该租户分配的 IO 带宽已经基本没有空闲 IO 带宽，即第一数据节点暂时没有充足的 IO 带宽响应该租户对分布式文件系统的 IO 访问请求，那么第一数据节点可以延迟响应该 IO 访问请求。

在第三方面的第三种可选的实现方式中，上述第一数据节点为租户分配 IO 带宽之前，该方法还包括：第一数据节点接收名称节点的指示。

上述第一数据节点为租户分配 IO 带宽的方法包括：第一数据节点根据名称节点的指示，为租户分配 IO 带宽。

由于可以由名称节点统一指示各个数据节点为租户分配 IO 带宽，因此能够合理地为租户分配 IO 带宽。

在第三方面的第四种可选的实现方式中，上述第一数据节点根

据 IO 访问请求，确定 IO 访问请求对应的租户的方法包括：第一数据节点根据 IO 访问请求携带的租户标识，确定 IO 访问请求对应的租户。

在第三方面的第五种可选的实现方式中，分布式文件系统管理的属于租户的文件所具有的文件属性中包含租户标识；具体地，分布式文件系统管理有属于租户的文件，其中，该文件的文件属性包含该租户的租户标识。一种场景是，分布式文件系统管理有不同租户的文件，可以根据文件的文件属性包含的租户标识确定该文件所属的租户。

上述第一数据节点根据 IO 访问请求，确定 IO 访问请求对应的租户的方法包括：第一数据节点根据 IO 访问请求指定的文件，获取 IO 访问请求指定的文件所具有的文件属性；且根据获取的文件属性包含的租户标识，确定 IO 访问请求对应的租户。

通过上述两种方法，第一数据节点接收到 IO 访问请求后，可以准确地确定出与该 IO 访问请求对应的租户，从而再根据第一数据节点已为该租户分配的 IO 带宽，确定是否延迟响应该 IO 访问请求。

在第三方面的第六种可选的实现方式中，该方法还包括：第一数据节点基于第一数据节点为租户分配的 IO 带宽，根据使用该租户的至少一个用户中每个用户的权重，为至少一个用户中的每个用户分配 IO 带宽。

其中，第一数据节点基于第一数据节点为租户分配的 IO 带宽，通过公式二为使用该租户的至少一个用户中的每个用户分配 IO 带宽。

上述公式二为：用户的 IO 带宽=该用户的权重/使用该租户的所有用户的权重之和\*第一数据节点为租户分配的 IO 带宽。

由于第一数据节点通过公式二为属于租户的每个用户分配 IO 带宽时是按照每个用户的权重分配的，因此为每个用户分配的 IO 带宽可以尽量满足每个用户对 IO 带宽的需求，并且均衡地满足每个用户对 IO 带宽的需求。

在第三方面的第七种可选的实现方式中，该方法还包括：第一数据节点基于第一数据节点为启动至少一个客户端的第一用户分配的 IO 带宽，根据至少一个客户端中每个客户端的权重，为至少一个客户端中的每个客户端分配 IO 带宽。

其中，第一数据节点基于第一数据节点为启动至少一个客户端的第一用户分配的 IO 带宽，通过公式三为由第一用户启动的至少一个客户端中的每个客户端分配 IO 带宽。

上述公式三可以为：客户端的 IO 带宽=该客户端的权重/由第一用户启动的所有客户端的权重之和\*第一数据节点为第一用户分配的 IO 带宽。

由于第一数据节点通过公式三为由一个用户登录的每个客户端分配 IO 带宽时是按照每个客户端的权重分配的，因此为每个客户端分配的 IO 带宽可以尽量满足每个客户端对 IO 带宽的需求，并且均衡地满足每个客户端对 IO 带宽的需求。

第四方面，提供一种名称节点，该名称节点应用于分布式文件系统，该分布式文件系统包括该名称节点和至少一个数据节点，该名称节点包括确定单元和指示单元。

其中，确定单元用于确定所述至少一个数据节点中每个数据节点的 IO 带宽和第一租户的 IO 带宽；指示单元用于基于确定单元确定的所述至少一个数据节点中每个数据节点的 IO 带宽以及第一租户的 IO 带宽，指示所述至少一个数据节点为第一租户分配至少一个 IO 带宽。所述至少一个 IO 带宽与所述至少一个数据节点一一对应，所述至少一个 IO 带宽中的每个 IO 带宽小于或者等于对应的数据节点的 IO 带宽，所述至少一个 IO 带宽中的每个 IO 带宽大于 0。

本申请提供的名称节点，可以为租户分配 IO 带宽，分配 IO 带宽的方式比较灵活。

该租户的 IO 带宽为至少一个用户共享，所述至少一个用户中的每个用户对应至少一个应用，因此属于该租户的至少一个用户对应的多个应用可以共享为该租户分配的 IO 带宽，即使在不同时刻该多

个应用可能不同，但该多个应用仍共享为该租户分配的 IO 带宽，因此能够提高分布式文件系统中 IO 带宽的利用率。

需要说明的是，本申请提供的名称节点包括但不限于上述第四方面中的确定单元和指示单元，并且上述第四方面中的确定单元和指示单元具有的功能包括但不限于上述描述的功能。应知，该名称节点可以包括用于执行上述第一方面以及第一方面的各种可选的实现方式所述的控制 IO 带宽的方法的单元和/或模块，名称节点包括确定单元和指示单元仅仅是一种举例，这些单元和/或模块是为了执行上述第一方面以及第一方面的各种可选的实现方式所述的控制 IO 带宽的方法，而对名称节点进行的逻辑上的划分。

上述第四方面各种可选的实现方式的技术效果的描述具体可参见上述对第一方面各种相应可选的实现方式的技术效果的相关描述，此处不再赘述。

第五方面，提供一种数据节点，该数据节点应用于分布式文件系统，该分布式文件系统包括至少一个数据节点，所述至少一个数据节点包括该数据节点。该数据节点包括确定单元和分配单元。

其中，确定单元用于确定所述至少一个数据节点中每个数据节点的 IO 带宽和第一租户的 IO 带宽；分配单元用于根据确定单元确定的所述至少一个数据节点中每个数据节点的 IO 带宽和第一租户的 IO 带宽，为第一租户分配第一 IO 带宽。第一 IO 带宽小于或等于第一数据节点的 IO 带宽，第一 IO 带宽大于 0。

本申请提供的数据节点，可以为第一租户分配第一 IO 带宽，分配 IO 带宽的方式比较灵活。

该第一租户的第一 IO 带宽为至少一个用户共享，所述至少一个用户中的每个用户对应至少一个应用，因此属于该第一租户的至少一个用户对应的多个应用可以共享为该第一租户分配的第一 IO 带宽，即使在不同时刻该多个应用可能不同，但该多个应用仍共享为该第一租户分配的第一 IO 带宽，因此能够提高该数据节点的 IO 带宽的利用率。

需要说明的是，本申请提供的数据节点包括但不限于上述第五方面中的确定单元和分配单元，并且上述第五方面中的确定单元和分配单元具有的功能包括但不限于上述描述的功能。应知，该数据节点可以包括用于执行上述第二方面以及第二方面的各种可选的实现方式所述的控制 IO 带宽的方法的单元和/或模块，该数据节点包括确定单元和分配单元仅仅是一种举例，这些单元和/或模块是为了执行上述第二方面以及第二方面的各种可选的实现方式所述的控制 IO 带宽的方法，而对数据节点进行的逻辑上的划分。

上述第五方面各种可选的实现方式的技术效果的描述具体可参见上述对第二方面各种相应可选的实现方式的技术效果的相关描述，此处不再赘述。

第六方面，提供又一种数据节点，该数据节点应用于分布式文件系统，该分布式文件系统包括至少一个数据节点，所述至少一个数据节点包括该数据节点。该数据节点包括接收单元、确定单元和分配单元。

其中，接收单元用于接收租户访问分布式文件系统的 IO 访问请求；确定单元用于根据接收单元接收的该 IO 访问请求，确定该 IO 访问请求对应的租户，并根据分配单元已为租户分配的 IO 带宽，确定是否延迟响应该 IO 访问请求。

本申请提供的数据节点应用于分布式文件系统，该分布式文件系统包括至少一个数据节点，所述至少一个数据节点包括该数据节点。该数据节点通过接收租户访问分布式文件系统的 IO 访问请求，并根据该 IO 访问请求确定该 IO 访问请求对应的租户，然后再根据该数据节点已为该租户分配的 IO 带宽，确定是否延迟响应该 IO 访问请求。申请可以实现在数据节点上为租户分配 IO 带宽，进而数据节点可以根据为租户分配的 IO 带宽确定是否延迟响应与租户对应的 IO 访问请求（即本发明实施例一定会响应 IO 访问请求，可以为立即响应或者延迟响应），因此可以提高响应 IO 访问请求的成功率。

需要说明的是，本申请提供的数据节点包括但不限于上述第六

方面中的接收单元、确定单元和分配单元，并且上述第六方面中的接收单元、确定单元和分配单元具有的功能包括但不限于上述描述的功能。应知，该数据节点可以包括用于执行上述第三方面以及第三方面的各种可选的实现方式所述的处理 IO 访问请求的方法的单元和/或模块，该数据节点包括接收单元、确定单元和分配单元仅仅是一种举例，这些单元和/或模块是为了执行上述第三方面以及第三方面的各种可选的实现方式所述的处理 IO 访问请求的方法，而对数据节点进行的逻辑上的划分。

上述第六方面各种可选的实现方式的技术效果的描述具体可参见上述对第三方面各种相应可选的实现方式的技术效果的相关描述，此处不再赘述。

第七方面，提供一种名称节点，该名称节点应用于分布式文件系统，该分布式文件系统包括该名称节点和至少一个数据节点。该名称节点包括至少一个处理器、接口电路、存储器和系统总线。

其中，存储器用于存储计算机执行指令，至少一个处理器、接口电路和存储器通过系统总线相互连接，当该名称节点运行时，至少一个处理器执行存储器存储的计算机执行指令，以使该名称节点执行上述第一方面或者第一方面的各种可选的实现方式所示的控制 IO 带宽的方法。

第八方面，提供一种计算机可读存储介质，计算机可读存储介质中存储有计算机执行指令，当名称节点的至少一个处理器执行该计算机执行指令时，名称节点执行上述第一方面或者第一方面的各种可选的实现方式所示的控制 IO 带宽的方法。

另外，第八方面还提供一种计算机程序产品，该计算机程序产品包括计算机执行指令，该计算机执行指令存储在计算机可读存储介质中。名称节点的至少一个处理器可以从计算机可读存储介质读取该计算机执行指令，并在名称节点执行该计算机执行指令时实施上述第一方面或者第一方面的各种可选的实现方式所示的控制 IO 带宽的方法。

上述第七方面和第八方面的技术效果的描述具体可参见上述对第一方面及其各种可选的实现方式的技术效果的相关描述，此处不再赘述。

第九方面，提供一种数据节点，该数据节点应用于分布式文件系统，该分布式文件系统包括至少一个数据节点，至少一个数据节点包括该数据节点。该数据节点包括至少一个处理器、接口电路、存储器和系统总线。

其中，存储器用于存储计算机执行指令，至少一个处理器、接口电路和存储器通过系统总线相互连接，当该数据节点运行时，至少一个处理器执行存储器存储的计算机执行指令，以使该数据节点执行上述第二方面或者第二方面的各种可选的实现方式所示的控制 IO 带宽的方法。

第十方面，提供一种计算机可读存储介质，计算机可读存储介质中存储有计算机执行指令，当数据节点的至少一个处理器执行该计算机执行指令时，数据节点执行上述第二方面或者第二方面的各种可选的实现方式所示的控制 IO 带宽的方法。

另外，第十方面还提供一种计算机程序产品，该计算机程序产品包括计算机执行指令，该计算机执行指令存储在计算机可读存储介质中。数据节点的至少一个处理器可以从计算机可读存储介质读取该计算机执行指令，至少一个处理器执行该计算机执行指令使得数据节点实施上述第二方面或者第二方面的各种可选的实现方式所示的控制 IO 带宽的方法。

上述第九方面和第十方面的技术效果的描述具体可参见上述对第二方面及其各种可选的实现方式的技术效果的相关描述，此处不再赘述。

第十一方面，提供又一种数据节点，该数据节点应用于分布式文件系统，该分布式文件系统包括至少一个数据节点，至少一个数据节点包括该数据节点。该数据节点包括至少一个处理器、接口电路、存储器和系统总线。

其中，存储器用于存储计算机执行指令，至少一个处理器、接口电路和存储器通过系统总线相互连接，当该数据节点运行时，至少一个处理器执行存储器存储的计算机执行指令，以使该数据节点执行上述第三方面或者第三方面的各种可选的实现方式所示的处理 IO 访问请求的方法。

第十二方面，提供一种计算机可读存储介质，计算机可读存储介质中存储有计算机执行指令，当数据节点的至少一个处理器执行该计算机执行指令时，数据节点执行上述第三方面或者第三方面的各种可选的实现方式所示的处理 IO 访问请求的方法。

另外，第十二方面还提供一种计算机程序产品，该计算机程序产品包括计算机执行指令，该计算机执行指令存储在计算机可读存储介质中。数据节点的至少一个处理器可以从计算机可读存储介质读取该计算机执行指令，至少一个处理器执行该计算机执行指令使得数据节点实施上述第三方面或者第三方面的各种可选的实现方式所示的处理 IO 访问请求的方法。

上述第十一方面和第十二方面的技术效果的描述具体可参见上述对第三方面及其各种可选的实现方式的技术效果的相关描述，此处不再赘述。

第十三方面，提供一种分布式文件系统，该分布式文件系统包括名称节点和至少一个数据节点，该名称节点为上述第四方面或第七方面所述的名节点。

或者，该分布式文件系统包括至少一个数据节点，该至少一个数据节点包括第一数据节点，该第一数据节点为上述第五方面或第九方面所述的数据节点。

或者，该分布式文件系统包括至少一个数据节点，该至少一个数据节点包括第一数据节点，该第一数据节点为上述第六方面或第十一方面所述的数据节点。

上述第十三方面的技术效果的描述具体可参见上述对第四方面、第七方面、第五方面、第九方面、第六方面或第十一方面的技

术效果的相关描述，此处不再赘述。

## 附图说明

为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术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。

图 1 为现有技术提供的访问分布式文件系统的架构示意图；

图 2 为本发明实施例提供的分布式文件系统的架构示意图；

图 3 为本发明实施例提供的访问分布式文件系统的架构示意图；

图 4 为本发明实施例提供的一种控制 IO 带宽的方法示意图一；

图 5 为本发明实施例提供的一种控制 IO 带宽的方法示意图二；

图 6 为本发明实施例提供的一种控制 IO 带宽的方法示意图三；

图 7 为本发明实施例提供的一种控制 IO 带宽的方法示意图四；

图 8 为本发明实施例提供的一种控制 IO 带宽的方法示意图五；

图 9 为本发明实施例提供的又一种控制 IO 带宽的方法示意图；

图 10 为本发明实施例提供的处理 IO 访问请求的方法示意图一；

图 11 为本发明实施例提供的处理 IO 访问请求的方法示意图二；

图 12 为本发明实施例提供的处理 IO 访问请求的方法示意图三；

图 13 为本发明实施例提供的名称节点的结构示意图；

图 14 为本发明实施例提供的一种数据节点的结构示意图；

图 15 为本发明实施例提供的又一种数据节点的结构示意图；

图 16 为本发明实施例提供的名称节点的硬件示意图；

图 17 为本发明实施例提供的一种数据节点的硬件示意图；

图 18 为本发明实施例提供的又一种数据节点的硬件示意图。

## 具体实施方式

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

方案进行详细地描述，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。

本发明的说明书和权利要求书中的术语“第一”、“第二”和“第三”等是用于区别不同的对象，而不是用于描述对象的特定顺序。例如，第一数据节点、第二数据节点和第三数据节点等是用于区别不同的数据节点，而不是用于描述数据节点的特定顺序。

在本发明实施例中，“示例性的”或者“例如”等词用于表示作例子、例证或说明。本发明实施例中被描述为“示例性的”或者“例如”的任何实施例或设计方案不应被解释为比其它实施例或设计方案更优选或更具优势。确切而言，使用“示例性的”或者“例如”等词旨在以具体方式呈现相关概念。

此外，本发明的描述中所提到的术语“包括”和“具有”以及它们的任何变形，意图在于覆盖不排他的包含。例如包含了一系列步骤或单元的过程、方法、系统、产品或设备没有限定于已列出的步骤或单元，而是可选地还包括其他没有列出的步骤或单元，或可选地还包括对于这些过程、方法、产品或设备固有的其它步骤或单元。

租户：是指与分布式文件系统中的资源相关的、逻辑上的概念。一个租户表示在分布式文件系统中拥有多少资源。例如，一个租户表示在分布式文件系统中拥有多少 IO 带宽等。

用户：是指登录客户端的账号。

应用：是指由用户登录的客户端对应的应用。例如，由用户登录的数据库客户端对应的数据库应用、由用户登录的网页客户端对应的网页应用等。由用户登录的客户端对应的应用可以通过租户访问分布式文件系统。

一个用户可以登录至少一个客户端，即一个用户可以对应至少一个应用；至少一个用户属于一个租户，即至少一个用户可以共享该租户的 IO 带宽；或者说属于一个租户的至少一个用户对应的多个应用可以共享该租户的 IO 带宽。

本发明实施例从租户角度分配 IO 带宽，提供一种控制 IO 带宽

和处理 IO 访问请求的方法，该方法应用于分布式文件系统中。

一方面，本发明实施例提供一种控制 IO 带宽的方法，该方法应用于分布式文件系统，该分布式文件系统包括名称节点和至少一个数据节点。该方法中名称节点通过确定该至少一个数据节点中每个数据节点的 IO 带宽和第一租户的 IO 带宽，并根据该至少一个数据节点中每个数据节点的 IO 带宽和第一租户的 IO 带宽，指示该至少一个数据节点为第一租户分配该至少一个 IO 带宽。如此，本发明实施例可以实现为租户分配 IO 带宽，分配 IO 带宽的方式比较灵活。

另外，现有技术中，每个新应用访问分布式文件系统时，分布式文件系统都会为新应用分配新的 IO 带宽，导致分布式文件系统中的 IO 带宽的利用率较低。但本发明实施例为租户分配 IO 带宽后，属于该租户的至少一个用户对应的多个应用可以共享为该租户分配的 IO 带宽，即使在不同时刻该多个应用可能不同，例如增加了新应用，但这些应用仍共享为该租户分配的 IO 带宽，即不会如现有技术一样对新增的应用另外分配 IO 带宽，因此能够提高分布式文件系统中 IO 带宽的利用率。

又一方面，本发明实施例提供又一种控制 IO 带宽的方法，该方法应用于分布式文件系统，该分布式文件系统包括至少一个数据节点，该至少一个数据节点包括第一数据节点。该方法中第一数据节点通过确定该至少一个数据节点中每个数据节点的 IO 带宽和第一租户的 IO 带宽，并根据该至少一个数据节点中每个数据节点的 IO 带宽和第一租户的 IO 带宽，为第一租户分配第一 IO 带宽。如此，本发明实施例可以实现为租户分配 IO 带宽，分配 IO 带宽的方式比较灵活。本发明实施例可以实现为租户分配 IO 带宽，属于该租户的至少一个用户对应的多个应用可以共享为该租户分配的 IO 带宽，即使在不同时刻该多个应用可能不同，例如增加了新应用，但这些应用仍共享为该租户分配的 IO 带宽，即不会如现有技术一样对新增的应用另外分配 IO 带宽，因此能够提高分布式文件系统中 IO 带宽的利用率。

又一方面，本发明实施例提供一种处理 IO 访问请求的方法，该方法应用的分布式文件系统包括至少一个数据节点，该至少一个数据节点包括第一数据节点。本发明实施例提供的处理 IO 访问请求的方法，通过第一数据节点接收租户访问分布式文件系统的 IO 访问请求，并根据该 IO 访问请求确定该 IO 访问请求对应的租户，然后再根据第一数据节点已为该租户分配的 IO 带宽，确定是否延迟响应该 IO 访问请求。如此，本发明实施例可以实现在数据节点上为租户分配 IO 带宽，进而数据节点可以根据为租户分配的 IO 带宽确定是否延迟响应与租户对应的 IO 访问请求（即本申请一定会响应 IO 访问请求，可以为立即响应或者延迟响应），因此可以提高响应 IO 访问请求的成功率。

本发明实施例提供的控制 IO 带宽的方法和处理 IO 访问请求的方法可以应用于分布式文件系统。如图 2 所示，为本发明实施例提供的一种分布式文件系统的架构示意图。在图 2 中，分布式文件系统包括名称节点（英文：name node）和与该名称节点连接的至少一个数据节点（英文：data node）。

可选的，如图 2 所示的分布式文件系统还可以包括 IO 记录模块和租户管理模块，IO 记录模块和租户管理模块可以设置在名称节点和/或至少一个数据节点上。

结合图 2，如图 3 所示，为应用访问分布式文件系统的架构示意图。图 3 中，分布式文件系统至少一个（例如为  $n$  个， $n$  为正整数）数据节点分别为数据节点 1、……、数据节点  $x$ 、……、数据节点  $n$ 。实际存储系统中，一个数据节点为至少一个租户分配 IO 带宽，一个租户的 IO 带宽分配给属于该租户的至少一个用户使用，图 3 示意为租户 2 的 IO 带宽被分配给属于租户 2 的用户 1 和用户 2 使用，一个用户可能会登录至少一个客户端，图 3 中客户端 1 和客户端 2 均是用户 1 登录的客户端。在图 3 中，每个应用分别对应一个客户端，例如应用 1 对应客户端 1，应用 2 对应客户端 2，应用 3 对应客户端 2，应用 1、应用 2 和应用 3 均可访问分布式文件系统；

应知，应用 1、应用 2 和应用 3 可以为相同应用，或者应用 1、应用 2 和应用 3 中的其中两个应用为相同应用，或者应用 1、应用 2 和应用 3 均为不同应用。图 3 中的其他数据节点均与数据节点 x 类似，此处不再赘述。

上述图 2 和图 3 仅是示例性的对本发明实施例提供的分布式文件系统的架构以及应用访问分布式文件系统的架构进行说明。具体的，分布式文件系统以及应用访问分布式文件系统的具体实现可以根据实际使用需求确定，本发明不作限定。

本发明实施例提供一种控制 IO 带宽的方法，该方法应用于分布式文件系统，该分布式文件系统可以为上述如图 2 或图 3 所示的分布式文件系统。该分布式文件系统包括名称节点和至少一个数据节点。如图 4 所示，该方法包括：

S101、名称节点确定该至少一个数据节点中每个数据节点的 IO 带宽。

上述每个数据节点的 IO 带宽为每个数据节点可分配的 IO 带宽。换句话说，数据节点的 IO 带宽是指该数据节点总共具有的 IO 带宽，该 IO 带宽可供该数据节点分配。

S102、名称节点确定第一租户的 IO 带宽。

第一租户的 IO 带宽为第一租户所需的 IO 带宽。

举例说明，在第一租户向名称节点请求 IO 带宽时，名称节点可以直接设定第一租户的 IO 带宽，例如根据经验值设定或者根据第一租户请求的 IO 带宽的值设定。

举例说明，在多个租户向名称节点请求 IO 带宽时，名称节点会综合考虑各个租户的优先级和各个租户请求的 IO 带宽，调整各个租户所需的 IO 带宽，例如调整第一租户所需的 IO 带宽。

S103、名称节点基于该至少一个数据节点中每个数据节点的 IO 带宽以及第一租户的 IO 带宽，指示该至少一个数据节点为第一租户分配至少一个 IO 带宽。

其中，上述至少一个 IO 带宽与该至少一个数据节点一一对应，该至少一个 IO 带宽中的每个 IO 带宽均小于或者等于对应的数据节点的 IO 带宽，该至少一个 IO 带宽中的每个 IO 带宽均大于 0。具体地，该至少一个数据节点在名称节点的指示下分别为第一租户分配一个 IO 带宽；并且，每个数据节点为第一租户分配的 IO 带宽，大于 0，小于该个数据节点的 IO 带宽。

示例性的，假设该至少一个数据节点为 3 个数据节点，即分布式文件系统包括 3 个数据节点，分别为数据节点 1、数据节点 2 和数据节点 3。名称节点确定这 3 个数据节点的 IO 带宽和第一租户的 IO 带宽之后，名称节点可基于这 3 个数据节点中每个数据节点的 IO 带宽和第一租户的 IO 带宽，指示这 3 个数据节点分别为第一租户分配 IO 带宽，以得到 3 个 IO 带宽。例如，名称节点可以基于这 3 个数据节点中每个数据节点的 IO 带宽和第一租户的 IO 带宽，指示数据节点 1 为第一租户分配 IO 带宽 1，指示数据节点 2 为第一租户分配 IO 带宽 2，指示数据节点 3 为第一租户分配 IO 带宽 3。

本发明实施例提供的控制 IO 带宽的方法可以实现为租户分配 IO 带宽，分配 IO 带宽的方式比较灵活。

该租户的 IO 带宽为至少一个用户共享，该至少一个用户中的每个用户对应至少一个应用，因此属于该租户的至少一个用户对应的多个应用可以共享为该租户分配的 IO 带宽，即使在不同时刻该多个应用可能不同，但该多个应用仍共享为该租户分配的 IO 带宽，因此能够提高分布式文件系统中 IO 带宽的利用率。

可选的，本发明实施例中，上述至少一个 IO 带宽的总和小于或者等于第一租户的 IO 带宽；或者，该至少一个 IO 带宽的总和小于或者等于第一租户的 IO 带宽与额外带宽之和。

举例说明，第一租户所需的 IO 带宽可能比较大，而名称节点可能无法为第一租户分配充足的 IO 带宽，因此上述至少一个 IO 带宽的总和小于或者等于第一租户所需的 IO 带宽，即第一租户的 IO 带宽。

举例说明，第一租户的业务优先级可能比较高，因此名称节点需要优先保证第一租户所需的 IO 带宽，因此名称节点可以在为第一租户分配第一租户所需的 IO 带宽的基础上，还可以为第一租户分配一个额外带宽，以优先保证第一租户所需的 IO 带宽。因此，上述至少一个 IO 带宽的总和可能大于或者等于第一租户所需的 IO 带宽（即第一租户的 IO 带宽），并且小于或者等于第一租户的 IO 带宽与额外带宽之和。其中，额外带宽可以为大于 0 的带宽。

名称节点可以指示数据节点为第一租户分配合适的 IO 带宽，从而尽可能地满足第一租户的 IO 带宽需求。

可选的，本发明实施例提供的控制 IO 带宽的方法中，该至少一个数据节点包括第一数据节点，第一数据节点在名称节点的指示下为第一租户分配的 IO 带宽是第一 IO 带宽，第一 IO 带宽是该至少一个 IO 带宽中与第一数据节点对应的 IO 带宽。

结合图 4，如图 5 所示，在上述 S103 之后，本发明实施例提供的控制 IO 带宽的方法还可以包括：

S104、名称节点确定第一租户分别在至少一个 IO 带宽中使用的 IO 带宽为至少一个使用带宽。

上述至少一个使用带宽与该至少一个 IO 带宽一一对应。

对于该至少一个数据节点中的每个数据节点，该数据节点在名称节点的指示下为第一租户分配一个 IO 带宽，该数据节点可以使用该 IO 带宽响应第一租户的 IO 访问请求；但对于某个时间点，响应第一租户的 IO 访问请求实际使用的 IO 带宽（即使用带宽）是固定的，其可能小于该数据节点为第一租户分配的 IO 带宽，也可能等于该数据节点为第一租户分配的 IO 带宽。以此类推，在该时间点，第一租户在每个数据节点上分别具有一个使用带宽。

示例性的，假设该至少一个数据节点为 3 个数据节点，该至少一个 IO 带宽为 IO 带宽 1、IO 带宽 2 和 IO 带宽 3，那么名称节点确定第一租户分别在这 3 个 IO 带宽中使用的 IO 带宽为 3 个使用带宽。例如，在 IO 带宽 1 中使用的 IO 带宽为使用带宽 1，在 IO 带宽 2 中

使用的 IO 带宽为使用带宽 2，在 IO 带宽 3 中使用的 IO 带宽为使用带宽 3。其中，第一租户在 IO 带宽中的使用带宽小于或者等于分配该 IO 带宽的数据节点的 IO 带宽。

S105、名称节点根据该至少一个使用带宽，指示该至少一个数据节点分别调整该至少一个 IO 带宽中对应的 IO 带宽，或者指示第一数据节点调整第一 IO 带宽。

由于名称节点为第一租户分配该至少一个 IO 带宽之后，第一租户可能并不会将该至少一个 IO 带宽全部使用完，或者该至少一个 IO 带宽可能并不能满足第一租户所需的 IO 带宽，因此名称节点需要根据第一租户在该至少一个 IO 带宽中的该至少一个使用带宽，指示相应的数据节点调整为第一租户分配的 IO 带宽。如此名称节点能够合理地第一租户分配 IO 带宽。

可选的，名称节点可以根据第一租户在该至少一个 IO 带宽中的该至少一个使用带宽，指示该至少一个数据节点分别调整该至少一个 IO 带宽中对应的 IO 带宽。或者，名称节点可以根据第一租户在该至少一个 IO 带宽中的该至少一个使用带宽，指示第一数据节点调整第一 IO 带宽（即第一数据节点为第一租户分配的 IO 带宽）。

本实施例中，至少一个数据节点包括的第一数据节点可以是一个或多个，第一数据节点的具体个数可以根据实际使用需求确定，本发明不作限定。采用上述方法，名称节点可以根据第一租户在该至少一个 IO 带宽中的该至少一个使用带宽，指示该至少一个数据节点中的至少一个数据节点分别调整其为第一租户分配的 IO 带宽。例如，名称节点可以通过执行下述（1）、（2）和（3）中的至少一项调整为第一租户分配的 IO 带宽。

（1）指示数据节点 1 调整数据节点 1 为第一租户分配的 IO 带宽 1。

（2）指示数据节点 2 调整数据节点 2 为第一租户分配的 IO 带宽 2。

（3）指示数据节点 3 调整数据节点 3 为第一租户分配的 IO 带

宽 3。

通过上述方法，在名称节点指示该至少一个数据节点为第一租户分配该至少一个 IO 带宽之后，名称节点可以根据第一租户在该至少一个 IO 带宽中的该至少一个使用带宽，指示相应的数据节点调整其为第一租户分配的 IO 带宽，例如，如果第一租户在该至少一个 IO 带宽中的该至少一个使用带宽比较大（如该至少一个 IO 带宽总和与该至少一个使用带宽总和的差值小于或者等于某个阈值），则名称节点指示相应的数据节点增大其为第一租户分配的 IO 带宽；如果第一租户在该至少一个 IO 带宽中的该至少一个使用带宽比较小（如该至少一个 IO 带宽总和与该至少一个使用带宽总和的差值大于某个阈值），则名称节点指示相应的数据节点减小其为第一租户分配的 IO 带宽。如此，名称节点能够为第一租户提供合适的 IO 带宽，另外还可以尽可能地满足第一租户的 IO 带宽需求。

可选的，结合图 4，如图 6 所示，本发明实施例提供的控制 IO 带宽的方法中，在上述 S103 之后，该方法还可以包括：

S106、名称节点确定第一租户分别在至少一个 IO 带宽中使用的 IO 带宽为至少一个使用带宽。

上述至少一个使用带宽与至少一个 IO 带宽一一对应。

对于该至少一个使用带宽的描述具体可以参见上述 S104 中对该至少一个使用带宽的相关描述，此处不再赘述。

S107、当该至少一个 IO 带宽之和与该至少一个使用带宽之和的差值小于或者等于第一阈值时，名称节点指示第二数据节点为第一租户分配 IO 带宽，该至少一个数据节点不包括第二数据节点。

名称节点指示该至少一个数据节点为第一租户分配该至少一个 IO 带宽之后，名称节点确定第一租户分别在至少一个 IO 带宽中使用的该至少一个使用带宽，当该至少一个 IO 带宽之和与该至少一个使用带宽之和的差值小于或者等于第一阈值时，表示第一租户实际使用的 IO 带宽已经非常接近该至少一个数据节点为其分配的 IO 带宽（即第一租户基本全部使用了该至少一个数据节点为其分配的

IO 带宽), 此时为了优先保证第一租户的业务, 使得第一租户具有充足的 IO 带宽, 名称节点还可以指示分布式存储系统中除上述至少一个数据节点之外的其他数据节点 (例如第二数据节点) 为第一租户分配 IO 带宽, 即名称节点指示增加为第一租户分配的 IO 带宽。

举例来说, 假设上述 3 个 IO 带宽中, IO 带宽 1 为 20Mbit/s (比特每秒), IO 带宽 2 为 30Mbit/s, IO 带宽 3 为 50Mbit/s; 上述 3 个 IO 带宽中的 3 个使用带宽中, 使用带宽 1 为 20Mbit/s, 使用带宽 2 为 30Mbit/s, 使用带宽 3 为 40Mbit/s。如果第一阈值为 20Mbit/s, 那么由于 3 个 IO 带宽之和与 3 个使用带宽之和的差值小于第一阈值, 因此名称节点可以指示分布式存储系统中除上述至少一个数据节点之外的第二数据节点为第一租户分配 IO 带宽。

上述第二数据节点可以为一个数据节点, 也可以为多个数据节点。具体的可以根据实际使用需求确定, 本发明不作限定。

上述图 2 只是以分布式存储系统中包括至少一个数据节点为例进行示例性的说明, 实际应用中, 分布式存储系统中还可以包括更多的数据节点, 本发明不作具体限定。

可选的, 结合图 4, 如图 7 所示, 本发明实施例提供的控制 IO 带宽的方法中, 在上述 S103 之后, 该方法还可以包括:

S108、名称节点确定第一租户分别在至少一个 IO 带宽中使用的 IO 带宽为至少一个使用带宽。

该至少一个使用带宽与该至少一个 IO 带宽一一对应。

对于该至少一个使用带宽的描述具体可以参见上述 S104 中对该至少一个使用带宽的相关描述, 此处不再赘述。

S109、当该至少一个 IO 带宽之和与该至少一个使用带宽之和的差值大于第一阈值时, 名称节点指示第三数据节点将第三数据节点为第一租户分配的 IO 带宽调整为 0, 该至少一个数据节点包括第三数据节点。

名称节点指示该至少一个数据节点为第一租户分配该至少一个 IO 带宽之后, 名称节点确定第一租户分别在至少一个 IO 带宽中

使用的该至少一个使用带宽,当该至少一个 IO 带宽之和与该至少一个使用带宽之和的差值大于第一阈值时,表示第一租户实际使用的 IO 带宽远小于该至少一个数据节点为其分配的 IO 带宽(即第一租户可能只使用了该至少一个数据节点为其分配的 IO 带宽中的一部分带宽),此时为了节省 IO 带宽(或者将第一租户没有使用的 IO 带宽给其他租户使用),名称节点还可以指示该至少一个数据节点中的第三数据节点将第三数据节点为第一租户分配的 IO 带宽调整为 0,即名称节点指示减少为第一租户分配的 IO 带宽。

举例来说,假设上述 3 个 IO 带宽中,IO 带宽 1 为 20Mbit/s,IO 带宽 2 为 30Mbit/s,IO 带宽 3 为 50Mbit/s;上述 3 个 IO 带宽中的 3 个使用带宽中,使用带宽 1 为 20Mbit/s,使用带宽 2 为 20Mbit/s,使用带宽 3 为 0Mbit/s。如果第一阈值为 20Mbit/s,那么由于 3 个 IO 带宽之和与 3 个使用带宽之和的差值大于第一阈值,因此名称节点可以指示该至少一个数据节点中的第三数据节点(例如为第一租户分配 IO 带宽 3 的数据节点),将第三数据节点为第一租户分配的 IO 带宽调整为 0。

上述第三数据节点可以为一个数据节点,也可以为多个数据节点。具体的可以根据实际使用需求确定,本发明不作限定。

可选的,本发明实施例提供的控制 IO 带宽的方法中,该至少一个数据节点包括第一数据节点,第一数据节点在名称节点的指示下为第一租户分配的 IO 带宽是第一 IO 带宽,第一 IO 带宽是该至少一个 IO 带宽中与第一数据节点对应的 IO 带宽。

结合图 4,如图 8 所示,在上述 S103 之后,该方法还可以包括:  
S110、名称节点确定第一租户使用的第一 IO 带宽中的 IO 带宽。  
S111、名称节点根据第一租户使用的第一 IO 带宽中的 IO 带宽,指示第一数据节点调整第一 IO 带宽。

对于第一租户使用的第一 IO 带宽中的 IO 带宽(即第一租户在第一 IO 带宽中使用的 IO 带宽)的描述具体可以参见上述 S104 中对该至少一个使用带宽的相关描述,此处不再赘述。

名称节点指示该至少一个数据节点为第一租户分配该至少一个 IO 带宽之后, 名称节点可以确定第一租户使用的第一 IO 带宽(即该至少一个数据节点中第一数据节点为第一租户分配的 IO 带宽)中的 IO 带宽, 然后名称节点再根据第一租户使用的第一 IO 带宽中的 IO 带宽, 指示第一数据节点调整第一 IO 带宽。例如, 如果第一租户使用的第一 IO 带宽中的 IO 带宽较大(比如大于某个阈值), 则名称节点指示第一数据节点增大第一 IO 带宽; 如果第一租户使用的第一 IO 带宽中的 IO 带宽较小(比如小于或者等于某个阈值), 则名称节点指示第一数据节点减小第一 IO 带宽。如此, 名称节点能够为第一租户提供合适的 IO 带宽, 另外还可以尽可能地满足第一租户的 IO 带宽需求。

通过上述如图 8 所示的方法, 名称节点可以以数据节点为单位, 指示相应的数据节点(例如第一数据节点)分别在数据节点内部调整数据节点为第一租户分配的 IO 带宽。如此, 名称节点可以根据第一租户在各数据节点实际使用的 IO 带宽, 指示对应数据节点为第一租户提供更加合适的 IO 带宽, 另外还可以尽可能地满足第一租户的 IO 带宽需求。

可选的, 本发明实施例提供的控制 IO 带宽的方法中, 第一租户为  $m$  个租户中的一个租户(即  $m$  个租户包括第一租户),  $m$  个租户均为由名称节点分配 IO 带宽的租户,  $m$  为正整数。上述 S102 具体可以包括:

S102a、名称节点根据  $m$  个租户中每个租户的权重, 从该至少一个数据节点中每个数据节点的 IO 带宽的总和, 确定第一租户的 IO 带宽。

上述每个租户的权重具体可以根据每个租户的业务优先级和/或该租户的业务所需的 IO 带宽来确定。具体的, 租户的业务优先级越高, 租户的权重越大; 租户的业务所需的 IO 带宽越大, 租户的权重越大。可选的, 当根据租户的业务优先级和租户的业务所需的 IO 带宽确定租户的权重时, 可以优先根据租户的业务优先级确定一个

权重（例如权重 1），然后再在该权重的基础上，进一步根据租户的业务所需的 IO 带宽对该权重进行相应的调整，并将调整后的权重（例如权重 2）确定为租户的权重。如此，确定的租户的权重比较准确。

可选的，名称节点可以通过公式一确定第一租户的 IO 带宽。

上述公式一可以为：第一租户的 IO 带宽=第一租户的权重/由该至少一个数据节点分配 IO 带宽的所有租户的权重之和\*该至少一个数据节点中每个数据节点的 IO 带宽的总和。

示例性的，假设  $m=3$ ，即由该至少一个数据节点分配 IO 带宽的租户有 3 个，分别为租户 1、租户 2 和租户 3；租户 1 的权重为 1，租户 2 的权重为 2，租户 3 的权重为 3；该至少一个数据节点中每个数据节点的 IO 带宽的总和为  $M$ ；则：

租户 1 的 IO 带宽=租户 1 的权重/（租户 1 的权重+租户 2 的权重+租户 3 的权重）\*该至少一个数据节点中每个数据节点的 IO 带宽的总和= $1/(1+2+3)*M=M/6$ ；

租户 2 的 IO 带宽=租户 2 的权重/（租户 1 的权重+租户 2 的权重+租户 3 的权重）\*该至少一个数据节点中每个数据节点的 IO 带宽的总和= $2/(1+2+3)*M=M/3$ ；

租户 3 的 IO 带宽=租户 3 的权重/（租户 1 的权重+租户 2 的权重+租户 3 的权重）\*该至少一个数据节点中每个数据节点的 IO 带宽的总和= $3/(1+2+3)*M=M/2$ 。

由于名称节点通过公式一为租户分配 IO 带宽时是按照每个租户的权重分配的，因此为每个租户分配的 IO 带宽可以尽量满足每个租户对 IO 带宽的需求，并且均衡地满足每个租户对 IO 带宽的需求。

本发明实施例提供又一种控制 IO 带宽的方法，该方法应用于分布式文件系统，该分布式文件系统可以为上述如图 2 或图 3 所示的分布式文件系统。该分布式文件系统包括至少一个数据节点，该至少一个数据节点包括第一数据节点。如图 9 所示，该方法包括：

S201、第一数据节点确定该至少一个数据节点中每个数据节点的 IO 带宽。

上述每个数据节点的 IO 带宽为每个数据节点可分配的 IO 带宽。换句话说，数据节点的 IO 带宽是指该数据节点总共具有的 IO 带宽，该 IO 带宽可供该数据节点分配。

S202、第一数据节点确定第一租户的 IO 带宽。

第一租户的 IO 带宽为第一租户所需的 IO 带宽。

举例说明，在第一租户向第一数据节点请求 IO 带宽时，第一数据节点直接设定第一租户的 IO 带宽，例如根据经验值设定或者根据第一租户请求的 IO 带宽的值设定。

举例说明，在多个租户向第一数据节点请求 IO 带宽时，第一数据节点会综合考虑各个租户的优先级和各个租户请求的 IO 带宽，并调整各个租户所需的 IO 带宽，例如调整第一租户所需的 IO 带宽。

S203、第一数据节点根据该至少一个数据节点中每个数据节点的 IO 带宽和第一租户的 IO 带宽，为第一租户分配第一 IO 带宽。

其中，上述第一 IO 带宽小于或等于第一数据节点的 IO 带宽，第一 IO 带宽大于 0。具体地，第一数据节点为第一租户分配第一 IO 带宽；并且，该第一 IO 带宽大于 0，小于第一数据节点的 IO 带宽。

本实施例中，该至少一个数据节点可以包括一个或多个第一数据节点，或者第一数据节点可以为该至少一个数据节点中的任一数据节点。

示例性的，假设该至少一个数据节点为 3 个数据节点，即分布式文件系统包括 3 个数据节点，分别为数据节点 1、数据节点 2 和数据节点 3。以第一数据节点为数据节点 1 为例，数据节点 1 确定这 3 个数据节点的 IO 带宽和第一租户的 IO 带宽之后，数据节点 1 可基于这 3 个数据节点中每个数据节点的 IO 带宽和第一租户的 IO 带宽，为第一租户分配第一 IO 带宽。例如，数据节点 1 可以基于这 3 个数据节点中每个数据节点的 IO 带宽和第一租户的 IO 带宽，为第一租户分配 IO 带宽 1。

本发明实施例提供的控制 IO 带宽的方法可以实现为租户分配 IO 带宽，分配 IO 带宽的方式比较灵活。

该租户的 IO 带宽为至少一个用户共享，该至少一个用户中的每个用户对应至少一个应用，因此属于该租户的至少一个用户对应的多个应用可以共享为该租户分配的 IO 带宽，即使在不同时刻该多个应用可能不同，但该多个应用仍共享为该租户分配的 IO 带宽，因此能够提高分布式文件系统中 IO 带宽的利用率。

本实施例中如图 9 所示的控制 IO 带宽的方法与上述如图 4 所示的控制 IO 带宽的方法相比，由于上述如图 9 所示的控制 IO 带宽的方法可以由各个数据节点自行为租户分配 IO 带宽，而无需在名称节点的指示下为租户分配 IO 带宽，因此分配 IO 带宽的过程不需要名称节点的参与，可以减轻名称节点的负担，节省名称节点的资源，降低名称节点的开销。

进一步地，由于各个数据节点无需在名称节点的指示下为租户分配 IO 带宽，因此各个数据节点可以实时地自行为租户分配 IO 带宽，从而还可以提高在租户级别控制 IO 带宽的实时性。

可选的，本发明实施例中，上述第一 IO 带宽小于或等于第一租户的 IO 带宽；或者，第一 IO 带宽小于或等于第一租户的 IO 带宽与额外带宽之和。

举例说明，第一租户所需的 IO 带宽可能比较大，而第一数据节点可能无法为第一租户分配充足的 IO 带宽，因此上述第一 IO 带宽小于或者等于第一租户所需的 IO 带宽，即第一租户的 IO 带宽。

举例说明，第一租户的业务优先级可能比较高，因此第一数据节点需要优先保证第一租户所需的 IO 带宽，因此第一数据节点可以在为第一租户分配第一租户所需的 IO 带宽的基础上，还可以为第一租户分配一个额外带宽，以优先保证第一租户所需的 IO 带宽。因此，上述第一 IO 带宽可以大于或者等于第一租户所需的 IO 带宽（即第一租户的 IO 带宽），并且小于或者等于第一租户的 IO 带宽与额外带宽之和。其中，额外带宽可以为大于 0 的带宽。

第一数据节点可以为第一租户分配合适的 IO 带宽,从而尽可能地满足第一租户的 IO 带宽需求。

可选的,结合图 9,在上述 S203 之后,本发明实施例提供的控制 IO 带宽的方法还可以包括:

S204、第一数据节点确定第一租户分别在至少一个数据节点中使用的 IO 带宽为至少一个使用带宽。

上述至少一个使用带宽与该至少一个数据节点一一对应。

示例性的,假设该至少一个数据节点为 3 个数据节点,分别为数据节点 1、数据节点 2 和数据节点 3,以第一数据节点为数据节点 1 为例,第一数据节点从这三个数据节点(数据节点 1、数据节点 2 和数据节点 3)获得第一租户在这 3 个数据节点中分别使用的 IO 带宽,并将获得的 3 个 IO 带宽作为 3 个使用带宽;其中,在数据节点 1 中使用的 IO 带宽为使用带宽 1,在数据节点 2 中使用的 IO 带宽为使用带宽 2,在数据节点 3 中使用的 IO 带宽为使用带宽 3。

第一租户在数据节点中的使用带宽小于或者等于该数据节点的 IO 带宽。

S205、第一数据节点根据该至少一个使用带宽,调整第一 IO 带宽。

举例说明,第一数据节点为第一租户分配第一 IO 带宽之后,第一租户可能并不会将第一 IO 带宽全部使用完,或者第一 IO 带宽可能并不能满足第一租户所需的 IO 带宽,或者第一租户使用第一 IO 带宽的差别较大(例如第一 IO 带宽中部分 IO 带宽被第一租户全部使用,但第一 IO 带宽中其他部分 IO 带宽被第一租户使用较少),因此第一数据节点需要根据第一租户的实际使用带宽(即至少一个使用带宽),调整第一数据节点为第一租户分配的第一 IO 带宽。如此第一数据节点能够合理地第一租户分配 IO 带宽。

可选的,结合图 9,在上述 S203 之后,本发明实施例提供的控制 IO 带宽的方法还可以包括:

S206、第一数据节点确定第一租户使用的第一 IO 带宽中的 IO

带宽。

S207、第一数据节点根据第一租户使用的第一 IO 带宽中的 IO 带宽，调整第一 IO 带宽。

对于第一租户使用的第一 IO 带宽中的 IO 带宽（即第一租户在第一 IO 带宽中使用的 IO 带宽）的描述具体可以参见上述 S204 中对至少一个使用带宽的相关描述，此处不再赘述。

第一数据节点为第一租户分配第一 IO 带宽之后，第一数据节点可以根据第一租户在第一 IO 带宽中实际使用的 IO 带宽，调整第一 IO 带宽。例如，如果第一租户在第一 IO 带宽中实际使用的 IO 带宽比较大（如第一 IO 带宽与第一租户在第一 IO 带宽中实际使用的 IO 带宽的差值小于或者等于某个阈值），则第一数据节点增大第一 IO 带宽；如果第一租户在第一 IO 带宽中实际使用的 IO 带宽比较小（如第一 IO 带宽与第一租户在第一 IO 带宽中实际使用的 IO 带宽的差值大于某个阈值），则第一数据节点减小第一 IO 带宽。如此，第一数据节点能够为第一租户提供合适的 IO 带宽，另外还可以尽可能地满足第一租户的 IO 带宽需求。

可选的，本发明实施例提供的控制 IO 带宽的方法中，第一租户为  $m$  个租户中的一个租户（即  $m$  个租户包括第一租户）， $m$  个租户均为由至少一个数据节点（包括第一数据节点）分配 IO 带宽的租户， $m$  为正整数。结合图 9，上述 S202 具体可以包括：

S202a、第一数据节点根据  $m$  个租户中每个租户的权重，从该至少一个数据节点中每个数据节点的 IO 带宽的总和，确定第一租户的 IO 带宽。

对于 S202a 的描述具体可以参见上述实施例中对 S102a 的相关描述，此处不再赘述。

由于第一数据节点为租户分配 IO 带宽时是按照每个租户的权重分配的，因此为每个租户分配的 IO 带宽可以尽量满足每个租户对 IO 带宽的需求，并且均衡地满足每个租户对 IO 带宽的需求。

本发明实施例提供一种处理 IO 访问请求的方法,该方法应用于分布式文件系统,该分布式文件系统可以为上述如图 2 或图 3 所示的分布式文件系统。该分布式文件系统包括至少一个数据节点,该至少一个数据节点包括第一数据节点。如图 10 所示,该方法包括:

S301、第一数据节点接收 IO 访问请求,该 IO 访问请求是租户访问分布式文件系统的请求。

S302、第一数据节点根据该 IO 访问请求,确定该 IO 访问请求对应的租户。

S303、第一数据节点根据第一数据节点已为该租户分配的 IO 带宽,确定是否延迟响应该 IO 访问请求。

本发明实施例中,第一数据节点接收到租户访问分布式文件系统的 IO 访问请求之后,第一数据节点首先根据该 IO 访问请求,确定与该 IO 访问请求对应的租户(即是哪一个租户访问分布式文件系统),然后第一数据节点再根据第一数据节点已为该租户分配的 IO 带宽,确定是否延迟响应该 IO 访问请求。例如,如果第一数据节点为该租户分配的 IO 带宽已经被该租户使用了很大一部分,即第一数据节点没有充足的 IO 带宽响应该 IO 访问请求(租户的空闲 IO 带宽小于租户响应该 IO 访问请求所需的 IO 带宽),那么第一数据节点可以延迟响应该 IO 访问请求;如果第一数据节点为该租户分配的 IO 带宽只被该租户使用了很少的一部分,即第一数据节点有充足的 IO 带宽响应该 IO 访问请求(租户的空闲 IO 带宽大于或者等于租户响应该 IO 访问请求所需的 IO 带宽),那么第一数据节点可以立即响应该 IO 访问请求。具体的,将在下述实施例中进行详细地说明,此处不再赘述。

上述第一数据节点立即响应 IO 访问请求可以理解为:第一数据节点确定可以响应 IO 访问请求之后,第一数据节点直接响应该 IO 访问请求。

本发明实施例提供的处理 IO 访问请求的方法应用于分布式文件系统,该分布式文件系统包括至少一个数据节点,该至少一个数

据节点包括第一数据节点。该方法中第一数据节点通过接收租户访问分布式文件系统的 IO 访问请求，并根据该 IO 访问请求确定该 IO 访问请求对应的租户，然后再根据第一数据节点已为该租户分配的 IO 带宽，确定是否延迟响应该 IO 访问请求。本发明实施例可以实现在数据节点上为租户分配 IO 带宽，进而数据节点可以根据为租户分配的 IO 带宽确定是否延迟响应与租户对应的 IO 访问请求（即本发明实施例一定会响应 IO 访问请求，可以为立即响应或者延迟响应），因此可以提高响应 IO 访问请求的成功率。

可选的，结合图 10，如图 11 所示，上述 S303 具体可以包括：

S303a、第一数据节点确定第一数据节点已为租户分配的 IO 带宽与该租户在第一数据节点使用的 IO 带宽的差值。

上述差值可以表示该租户的空闲 IO 带宽。

S303b、第一数据节点在确定该差值大于第三阈值时，响应该 IO 访问请求。

当第一数据节点确定第一数据节点已为租户分配的 IO 带宽与该租户在第一数据节点使用的 IO 带宽的差值大于第三阈值时，表示第一数据节点已为租户分配的 IO 带宽中还有部分空闲 IO 带宽，即第一数据节点有充足的 IO 带宽响应该租户对分布式文件系统的 IO 访问请求，那么第一数据节点可以立即响应该 IO 访问请求。

可选的，结合图 10，如图 12 所示，上述 S303 具体可以包括：

S303c、第一数据节点确定第一数据节点已为租户分配的 IO 带宽与该租户在第一数据节点使用的 IO 带宽的差值。

S303d、第一数据节点在确定该差值小于或者等于第三阈值时，延迟响应该 IO 访问请求。

当第一数据节点确定第一数据节点已为租户分配的 IO 带宽与该租户在第一数据节点使用的 IO 带宽的差值小于或者等于第三阈值时，表示第一数据节点已为该租户分配的 IO 带宽已经基本没有空闲 IO 带宽，即第一数据节点暂时没有充足的 IO 带宽响应该租户对分布式文件系统的 IO 访问请求，那么第一数据节点可以延迟响应该

IO 访问请求，即等到第一数据节点为该租户分配的 IO 带宽中剩余有足够的 IO 带宽响应该 IO 访问请求时，第一数据节点再响应该 IO 访问请求。

可选的，本发明实施例提供的处理 IO 访问请求的方法中，上述第一数据节点为租户分配 IO 带宽可以通过下述方法实现：

S304、第一数据节点接收名称节点的指示。

S305、第一数据节点根据名称节点的指示，为租户分配 IO 带宽。

第一数据节点可以在名称节点的指示下为租户分配 IO 带宽。具体的，名称节点可以确定第一数据节点需要为该租户分配多少 IO 带宽，然后名称节点将该 IO 带宽指示给第一数据节点，由第一数据节点为该租户分配名称节点指示的 IO 带宽。如此，可以由名称节点统一指示各个数据节点为租户分配 IO 带宽，从而能够合理地为租户分配 IO 带宽。

可选的，结合图 10，本发明实施例中，上述 S302 具体可以包括：

S302a、第一数据节点根据 IO 访问请求携带的租户标识，确定该 IO 访问请求对应的租户。

本发明实施例中，由于租户访问分布式文件系统的 IO 访问请求中可以携带租户标识；具体地，分布式文件系统管理有属于租户的文件，其中，该文件的文件属性包含该租户的租户标识。一种场景是，分布式文件系统管理有不同租户的文件，可以根据文件的文件属性包含的租户标识确定该文件所属的租户。

第一数据节点接收到 IO 访问请求之后，可以根据该 IO 访问请求中携带的租户标识，确定与该 IO 访问请求对应的租户。

可选的，结合图 10，本发明实施例中，分布式文件系统管理的属于租户的文件所具有的文件属性中包含租户标识。上述 S302 具体可以包括：

S302b、第一数据节点根据 IO 访问请求指定的文件，获取该 IO

访问请求指定的文件所具有的文件属性。

S302c、第一数据节点根据获取的文件属性包含的租户标识，确定该 IO 访问请求对应的租户。

本发明实施例中，由于分布式文件系统管理的属于租户的文件所具有的文件属性中包含租户标识，因此第一数据节点接收到 IO 访问请求之后，第一数据节点首先可以根据该 IO 访问请求指定的文件，获取该 IO 访问请求指定的文件所具有的文件属性，然后第一数据节点再根据该文件属性中包含的租户标识，确定与该 IO 访问请求对应的租户。

通过上述两种方法，第一数据节点接收到 IO 访问请求后，可以准确地确定出与该 IO 访问请求对应的租户，从而再根据第一数据节点已为该租户分配的 IO 带宽，确定是否延迟响应该 IO 访问请求。

可选的，本发明实施例提供的分布式文件系统中，一个租户的 IO 带宽分配给属于该租户的至少一个用户使用，一个用户可能会登录至少一个客户端（即一个用户可能对应多个应用）。第一数据节点为租户分配 IO 带宽之后，第一数据节点还可以进一步为属于该租户的至少一个用户分配 IO 带宽，以及为由至少一个用户登录的至少一个客户端分配 IO 带宽。如此，第一数据节点可以依次按照“租户—>用户—>客户端”的形式最终为每个客户端分配 IO 带宽。

下面对第一数据节点为属于该租户的至少一个用户分配 IO 带宽进行示例性的说明。

第一数据节点基于第一数据节点为租户分配的 IO 带宽，根据至少一个用户中每个用户的权重，为至少一个用户中的每个用户分配 IO 带宽。

上述每个用户的权重具体可以根据每个用户的业务优先级和/或该用户的业务所需的 IO 带宽来确定。具体的，用户的业务优先级越高，用户的权重越大；用户的业务所需的 IO 带宽越大，用户的权重越大。可选的，当根据用户的业务优先级和用户的业务所需的 IO 带宽确定用户的权重时，可以优先根据用户的业务优先级确定一个

权重（例如权重 3），然后再在该权重的基础上，进一步根据用户的业务所需的 IO 带宽对该权重进行相应的调整，并将调整后的权重（例如权重 4）确定为用户的权重。如此，确定的用户的权重比较准确。

第一数据节点基于第一数据节点为租户分配的 IO 带宽，根据至少一个用户中每个用户的权重，为至少一个用户中的每个用户分配 IO 带宽的方式可以为：

第一数据节点基于第一数据节点为租户分配的 IO 带宽，通过公式二为使用该租户的至少一个用户中的每个用户分配 IO 带宽。

其中，上述公式二可以为：用户的 IO 带宽=该用户的权重/使用该租户的所有用户的权重之和\*第一数据节点为租户分配的 IO 带宽。

示例性的，假设使用某租户的用户有 3 个，分别为用户 1、用户 2 和用户 3；用户 1 的权重为 1，用户 2 的权重为 2，用户 3 的权重为 3；第一数据节点为该租户分配的 IO 带宽为 N；则：

用户 1 的 IO 带宽=用户 1 的权重/（用户 1 的权重+用户 2 的权重+用户 3 的权重）\*第一数据节点为租户分配的 IO 带宽=1/（1+2+3）\*N=N/6；

用户 2 的 IO 带宽=用户 2 的权重/（用户 1 的权重+用户 2 的权重+用户 3 的权重）\*第一数据节点为租户分配的 IO 带宽=2/（1+2+3）\*N=N/3；

用户 3 的 IO 带宽=用户 3 的权重/（用户 1 的权重+用户 2 的权重+用户 3 的权重）\*第一数据节点为租户分配的 IO 带宽=3/（1+2+3）\*N=N/2。

本发明实施例中，由于第一数据节点通过公式二为属于租户的每个用户分配 IO 带宽时是按照每个用户的权重分配的，因此为每个用户分配的 IO 带宽可以尽量满足每个用户对 IO 带宽的需求，并且均衡地满足每个用户对 IO 带宽的需求。

下面对第一数据节点为属于该租户的第一用户登录的至少一个

客户端分配 IO 带宽进行示例性的说明。其中，第一用户是指属于该租户的至少一个用户中的任一用户。

第一数据节点基于第一数据节点为第一用户分配的 IO 带宽，根据至少一个客户端中每个客户端的权重，为至少一个客户端中的每个客户端分配 IO 带宽。

上述每个客户端的权重具体可以根据每个客户端的业务优先级和/或该客户端的业务所需的 IO 带宽来确定。具体的，客户端的业务优先级越高，客户端的权重越大；客户端的业务所需的 IO 带宽越大，客户端的权重越大。可选的，当根据客户端的业务优先级和客户端的业务所需的 IO 带宽确定客户端的权重时，可以优先根据客户端的业务优先级确定一个权重（例如权重 5），然后再在该权重的基础上，进一步根据客户端的业务所需的 IO 带宽对该权重进行相应的调整，并将调整后的权重（例如权重 6）确定为客户端的权重。如此，确定的客户端的权重比较准确。

第一数据节点基于第一数据节点为第一用户分配的 IO 带宽，根据至少一个客户端中每个客户端的权重，为至少一个客户端中的每个客户端分配 IO 带宽的方式可以为：

第一数据节点基于第一数据节点为第一用户分配的 IO 带宽，通过公式三为由第一用户启动的至少一个客户端中的每个客户端分配 IO 带宽。

其中，上述公式三可以为：客户端的 IO 带宽=该客户端的权重/由第一用户启动的所有客户端的权重之和\*第一数据节点为第一用户分配的 IO 带宽。

示例性的，假设第一用户启动的客户端有 3 个，分别为客户端 1、客户端 2 和客户端 3；客户端 1 的权重为 1，客户端 2 的权重为 2，客户端 3 的权重为 3；第一数据节点为第一用户分配的 IO 带宽为 A；则：

客户端 1 的 IO 带宽=客户端 1 的权重/（客户端 1 的权重+客户端 2 的权重+客户端 3 的权重）\*第一数据节点为第一用户分配的 IO

带宽 =  $1 / (1 + 2 + 3) * A = A/6$ ;

客户端 2 的 IO 带宽 = 客户端 2 的权重 / (客户端 1 的权重 + 客户端 2 的权重 + 客户端 3 的权重) \* 第一数据节点为第一用户分配的 IO 带宽 =  $2 / (1 + 2 + 3) * A = A/3$ ;

客户端 3 的 IO 带宽 = 客户端 3 的权重 / (客户端 1 的权重 + 客户端 2 的权重 + 客户端 3 的权重) \* 第一数据节点为第一用户分配的 IO 带宽 =  $3 / (1 + 2 + 3) * A = A/2$ 。

本发明实施例中，由于第一数据节点通过公式三为由一个用户登录的每个客户端分配 IO 带宽时是按照每个客户端的权重分配的，因此为每个客户端分配的 IO 带宽可以尽量满足每个客户端对 IO 带宽的需求，并且均衡地满足每个客户端对 IO 带宽的需求。

如图 13 所示，本发明实施例提供一种名称节点，该名称节点应用于分布式文件系统，该分布式文件系统包括该名称节点和至少一个数据节点，该名称节点用于执行以上如图 4 至图 8 任一所示的控制 IO 带宽的方法中的名称节点所执行的步骤。该名称节点可以包括相应步骤所对应的单元和/或模块。示例性的，以该名称节点包括相应步骤所对应的单元为例，该名称节点包括确定单元 10 和指示单元 11。

其中，确定单元 10，用于确定该至少一个数据节点中每个数据节点的 IO 带宽和第一租户的 IO 带宽；指示单元 11，用于基于确定单元 10 确定的该至少一个数据节点中每个数据节点的 IO 带宽以及第一租户的 IO 带宽，指示该至少一个数据节点为第一租户分配至少一个 IO 带宽。该至少一个 IO 带宽与该至少一个数据节点一一对应，该至少一个 IO 带宽中的每个 IO 带宽小于或者等于对应的数据节点的 IO 带宽，该至少一个 IO 带宽中的每个 IO 带宽大于 0。

可选的，上述至少一个 IO 带宽的总和小于或者等于第一租户的 IO 带宽；或者，上述至少一个 IO 带宽的总和小于或者等于第一租户的 IO 带宽与额外带宽之和。

可选的，上述至少一个数据节点包括第一数据节点，第一数据节点在指示单元 11 的指示下为第一租户分配的 IO 带宽是第一 IO 带宽，第一 IO 带宽是该至少一个 IO 带宽中与第一数据节点对应的 IO 带宽。

确定单元 10，还用于在指示单元 11 指示该至少一个数据节点为第一租户分配该至少一个 IO 带宽之后，确定第一租户分别在至少一个 IO 带宽中使用的 IO 带宽为至少一个使用带宽，该至少一个使用带宽与该至少一个 IO 带宽一一对应；指示单元 11，还用于根据确定单元 10 确定的该至少一个使用带宽，指示该至少一个数据节点分别调整该至少一个 IO 带宽中对应的 IO 带宽，或者指示第一数据节点调整第一 IO 带宽。

可选的，确定单元 10，还用于在指示单元 11 指示该至少一个数据节点为第一租户分配该至少一个 IO 带宽之后，确定第一租户分别在至少一个 IO 带宽中使用的 IO 带宽为至少一个使用带宽，该至少一个使用带宽与该至少一个 IO 带宽一一对应；指示单元 11，还用于当该至少一个 IO 带宽之和与确定单元 10 确定的该至少一个使用带宽之和的差值小于或者等于第一阈值时，指示第二数据节点为第一租户分配 IO 带宽，该至少一个数据节点不包括第二数据节点。

可选的，确定单元 10，还用于在指示单元 11 指示该至少一个数据节点为第一租户分配该至少一个 IO 带宽之后，确定第一租户分别在至少一个 IO 带宽中使用的 IO 带宽为至少一个使用带宽，该至少一个使用带宽与该至少一个 IO 带宽一一对应；指示单元 11，还用于当该至少一个 IO 带宽之和与确定单元 10 确定的该至少一个使用带宽之和的差值大于第一阈值时，指示第三数据节点将第三数据节点为第一租户分配的 IO 带宽调整为 0，该至少一个数据节点包括第三数据节点。

可选的，该至少一个数据节点包括第一数据节点，第一数据节点在指示单元 11 的指示下为第一租户分配的 IO 带宽是第一 IO 带

宽，第一 IO 带宽是该至少一个 IO 带宽中与第一数据节点对应的 IO 带宽。

确定单元 10，还用于在指示单元 11 指示该至少一个数据节点为第一租户分配该至少一个 IO 带宽之后，确定第一租户使用的第一 IO 带宽中的 IO 带宽；指示单元 11，还用于根据确定单元 10 确定的第一租户使用的第一 IO 带宽中的 IO 带宽，指示第一数据节点调整第一 IO 带宽。

可选的，m 个租户包括第一租户，m 为正整数。

确定单元，具体用于根据 m 个租户中每个租户的权重，从该至少一个数据节点中每个数据节点的 IO 带宽的总和，确定第一租户的 IO 带宽。

确定单元 10 和指示单元 11 的一种实现方式，可以是名称节点中的至少一个处理器执行存储器中存储的计算机执行指令实现。

可以理解，本实施例的名称节点可对应于上述如图 4 至图 8 任一所示的控制 IO 带宽的方法中的名称节点，并且本实施例的名称节点中的各个单元和/或模块的划分等均是实现了实现如图 4 至图 8 任一所示的方法流程，为了避免重复，在此不再赘述各个单元和/或模块的功能实现。

本发明实施例提供的名称节点应用于分布式文件系统，该分布式文件系统包括该名称节点和至少一个数据节点。该名称节点通过确定该至少一个数据节点中每个数据节点的 IO 带宽和第一租户的 IO 带宽，并根据该至少一个数据节点中每个数据节点的 IO 带宽和第一租户的 IO 带宽，指示该至少一个数据节点为第一租户分配该至少一个 IO 带宽。本发明实施例提供的名称节点，可以指示数据节点为租户分配 IO 带宽，分配 IO 带宽的方式比较灵活；进而，该租户的 IO 带宽为至少一个用户共享，至少一个用户中的每个用户对应至少一个应用，因此属于该租户的多个应用可以共享为该租户分配的 IO 带宽，即使在不同时刻该多个应用可能不同，但该多个应用仍共享为该租户分配的 IO 带宽，因此能够提高分布式文件系统中 IO 带

宽的利用率。

如图 14 所示,本发明实施例提供一种数据节点,该数据节点应用于分布式文件系统,该分布式文件系统包括至少一个数据节点,本发明实施例提供的该数据节点可以是该至少一个数据节点中的任意数据节点。该数据节点用于执行以上如图 9 所示的控制 IO 带宽的方法中的第一数据节点所执行的步骤。该数据节点可以包括相应步骤所对应的单元和/或模块。示例性的,以该数据节点包括相应步骤所对应的单元为例,该数据节点包括确定单元 20 和分配单元 21。

其中,确定单元 20,用于确定该至少一个数据节点中每个数据节点的 IO 带宽和第一租户的 IO 带宽;分配单元 21,用于根据确定单元 20 确定的该至少一个数据节点中每个数据节点的 IO 带宽和第一租户的 IO 带宽,为第一租户分配第一 IO 带宽。第一 IO 带宽小于或等于第一数据节点的 IO 带宽,第一 IO 带宽大于 0。

可选的,确定单元 20,还用于在分配单元 21 为第一租户分配第一 IO 带宽之后,确定第一租户分别在该至少一个数据节点中使用的 IO 带宽为该至少一个使用带宽;分配单元 21,还用于根据确定单元 20 确定的该至少一个使用带宽,调整第一 IO 带宽。

可选的,确定单元 20,还用于在分配单元 21 为第一租户分配第一 IO 带宽之后,确定第一租户使用的第一 IO 带宽中的 IO 带宽;分配单元 21,还用于根据确定单元 20 确定的第一租户使用的第一 IO 带宽中的 IO 带宽,调整第一 IO 带宽。

可选的,第一租户为 m 个租户中的一个租户(即 m 个租户包括第一租户),m 为正整数。

确定单元 20,具体用于根据 m 个租户中每个租户的权重,从该至少一个数据节点中每个数据节点的 IO 带宽的总和,确定第一租户的 IO 带宽。

确定单元 20 和分配单元 21 的一种实现方式,可以是数据节点中的至少一个处理器执行存储器中存储的计算机执行指令实现。

可以理解，本实施例的数据节点可对应于上述如图 9 所示的控制 IO 带宽的方法中的第一数据节点，并且本实施例的数据节点中的各个单元和/或模块的划分等均是实现了如图 9 所示的方法流程，为了避免重复，在此不再赘述各个单元和/或模块的功能实现。

本发明实施例提供的数据节点应用于分布式文件系统，该分布式文件系统包括至少一个数据节点，该至少一个数据节点包括该数据节点。该数据节点通过确定该至少一个数据节点中每个数据节点的 IO 带宽和第一租户的 IO 带宽，并根据该至少一个数据节点中每个数据节点的 IO 带宽和第一租户的 IO 带宽，为第一租户分配第一 IO 带宽。本发明实施例提供的数据节点，可以实现为租户分配 IO 带宽，分配 IO 带宽的方式比较灵活；进而，该租户的 IO 带宽为至少一个用户共享，至少一个用户中的每个用户对应至少一个应用，因此属于该租户的多个应用可以共享为该租户分配的 IO 带宽，即使在不同时刻该多个应用可能不同，但该多个应用仍共享为该租户分配的 IO 带宽，因此能够提高分布式文件系统中 IO 带宽的利用率。

如图 15 所示，本发明实施例提供一种数据节点，该数据节点应用于分布式文件系统，该分布式文件系统包括至少一个数据节点，至少一个数据节点包括该数据节点，该数据节点用于执行以上如图 10 至图 12 任一所示的处理 IO 访问请求的方法中的数据节点所执行的步骤。该数据节点可以包括相应步骤所对应的单元和/或模块。示例性的，以该数据节点包括相应步骤所对应的单元为例，该数据节点包括接收单元 30、确定单元 31 和分配单元 32。

其中，接收单元 30，用于接收 IO 访问请求，该 IO 访问请求是租户访问分布式文件系统的请求；确定单元 31，用于根据接收单元 30 接收的该 IO 访问请求，确定该 IO 访问请求对应的租户；并根据分配单元 32 已为租户分配的 IO 带宽，确定是否延迟响应该 IO 访问请求。

可选的，确定单元 31，具体用于确定分配单元 32 已为租户分

配的 IO 带宽与租户在该数据节点使用的 IO 带宽的差值；并在确定该差值大于第三阈值时，响应该 IO 访问请求。

可选的，确定单元 31，具体用于确定分配单元 32 已为租户分配的 IO 带宽与租户在该数据节点使用的 IO 带宽的差值；并在确定该差值小于或者等于第三阈值时，延迟响应该 IO 访问请求。

可选的，接收单元 30，还用于在分配单元 32 为租户分配 IO 带宽之前，接收名称节点的指示；分配单元 32，用于根据接收单元 30 接收的名称节点的指示，为租户分配 IO 带宽。

可选的，确定单元 31，具体用于根据接收单元 30 接收的 IO 访问请求携带的租户标识，确定该 IO 访问请求对应的租户。

可选的，分布式文件系统管理的属于租户的文件所具有的文件属性中包含租户标识。

确定单元 31，具体用于根据接收单元 30 接收的 IO 访问请求指定的文件，获取该 IO 访问请求指定的文件所具有的文件属性；并根据获取的文件属性包含的租户标识，确定该 IO 访问请求对应的租户。

接收单元 30 可以通过数据节点中的接口电路实现；确定单元 31 和分配单元 32 的一种实现方式，可以是数据节点中的至少一个处理器执行存储器中存储的计算机执行指令实现。

可以理解，本实施例的数据节点可对应于上述如图 10 至图 12 任一所示的处理 IO 访问请求的方法中的第一数据节点，并且本实施例的数据节点中的各个单元和/或模块的划分等均是实现了如图 10 至图 12 任一所示的方法流程，为了避免重复，在此不再赘述各个单元和/或模块的功能实现。

本发明实施例提供的数据节点应用于分布式文件系统，该分布式文件系统包括至少一个数据节点，至少一个数据节点包括该数据节点。该数据节点通过接收租户访问分布式文件系统的 IO 访问请求，并根据该 IO 访问请求确定该 IO 访问请求对应的租户，然后再根据该数据节点已为该租户分配的 IO 带宽，确定是否延迟响应该 IO

访问请求。本发明实施例可以实现在数据节点上为租户分配 IO 带宽,进而数据节点可以根据为租户分配的 IO 带宽确定是否延迟响应与租户对应的 IO 访问请求(即本申请一定会响应 IO 访问请求,可以为立即响应或者延迟响应),因此可以提高响应 IO 访问请求的成功率。

如图 16 所述,本发明实施例提供一种名称节点,该名称节点应用于分布式文件系统,该分布式文件系统包括该名称节点和至少一个数据节点。该名称节点包括至少一个处理器 40、接口电路 41、存储器 42 和系统总线 43。

其中,存储器 42 用于存储计算机执行指令,至少一个处理器 40、接口电路 41 和存储器 42 通过系统总线 43 相互连接,当该名称节点运行时,至少一个处理器 40 执行存储器 42 存储的计算机执行指令,以使该名称节点执行如图 4 至图 8 任一所示的控制 IO 带宽的方法。具体的控制 IO 带宽的方法可参见上述如图 4 至图 8 任一所示的实施例中的相关描述,此处不再赘述。

处理器 40 具体可以为中央处理器(英文:central processing unit,CPU)。处理器 40 还可以为其他通用处理器、数字信号处理器(英文:digital signal processing,简称 DSP)、专用集成电路(英文:application specific integrated circuit,简称 ASIC)、现场可编程门阵列(英文:field-programmable gate array,简称 FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。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。

处理器 40 还可以为专用处理器,该专用处理器可以包括具有名称节点其他专用处理功能的芯片。

存储器 42 可以包括易失性存储器(英文:volatile memory),例如随机存取存储器(英文:random-access memory,RAM);存储器 42 也可以包括非易失性存储器(英文:non-volatile memory),例如只读

存储器（英文：read-only memory, ROM），快闪存储器（英文：flash memory），硬盘（英文：hard disk drive, HDD）或固态硬盘（英文：solid-state drive, SSD）；存储器 42 还可以包括上述种类的存储器的组合。

系统总线 43 可以包括数据总线、电源总线、控制总线和信号状态总线等。本实施例中为了清楚说明，在图 16 中将各种总线都示意作为系统总线 43。

接口电路 41 具体可以是名称节点上的 IO 接口、网络接口或者通信接口。至少一个处理器 40 通过接口电路 41 与其他设备，例如数据节点之间进行数据的交互。

在具体实现过程中，上述如图 4 至图 8 任一所示的方法流程中的各步骤均可以通过硬件形式的处理器 40 执行存储器 42 中存储的软件形式的计算机执行指令实现。为避免重复，此处不再赘述。

本发明实施例提供一种计算机可读存储介质，计算机可读存储介质中存储有计算机执行指令，当名称节点的至少一个处理器执行该计算机执行指令时，名称节点执行上述如图 4 至图 8 任一所示的控制 IO 带宽的方法。具体的控制 IO 带宽的方法可参见上述如图 4 至图 8 任一所示的实施例中的相关描述，此处不再赘述。

本发明实施例提供一种计算机程序产品，该计算机程序产品包括计算机执行指令，该计算机执行指令存储在计算机可读存储介质中。名称节点的至少一个处理器可以从计算机可读存储介质读取该计算机执行指令，至少一个处理器执行该计算机执行指令使得名称节点实施上述如图 4 至图 8 任一所示的控制 IO 带宽的方法。

上述的计算机可读存储介质可以为上述的存储器 42。

本发明实施例提供的名称节点应用于分布式文件系统，该分布式文件系统包括该名称节点和至少一个数据节点。该名称节点通过确定该至少一个数据节点中每个数据节点的 IO 带宽和第一租户的 IO 带宽，并根据该至少一个数据节点中每个数据节点的 IO 带宽和第一租户的 IO 带宽，指示该至少一个数据节点为第一租户分配至少

一个 IO 带宽。本发明实施例提供的名称节点，可以指示数据节点为租户分配 IO 带宽，分配 IO 带宽的方式比较灵活；进而，该租户的 IO 带宽为至少一个用户共享，至少一个用户中的每个用户对应至少一个应用，因此属于该租户的多个应用可以共享为该租户分配的 IO 带宽，即使在不同时刻该多个应用可能不同，但该多个应用仍共享为该租户分配的 IO 带宽，因此能够提高分布式文件系统中 IO 带宽的利用率。

如图 17 所示，本发明实施例提供一种数据节点，该数据节点应用于分布式文件系统，该分布式文件系统包括至少一个数据节点，该至少一个数据节点包括该数据节点。该数据节点包括至少一个处理器 50、接口电路 51、存储器 52 和系统总线 53。

其中，存储器 52 用于存储计算机执行指令，至少一个处理器 50、接口电路 51 和存储器 52 通过系统总线 53 相互连接，当该数据节点运行时，至少一个处理器 50 执行存储器 52 存储的计算机执行指令，以使该数据节点执行如图 9 所示的控制 IO 带宽的方法。具体的控制 IO 带宽的方法可参见上述如图 9 所示的实施例中的相关描述，此处不再赘述。

处理器 50 具体可以为 CPU。处理器 50 还可以为其他通用处理器、DSP、ASIC、FPGA 或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。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。

处理器 50 还可以为专用处理器，该专用处理器可以包括具有数据节点其他专用处理功能的芯片。

存储器 52 可以包括易失性存储器，例如 RAM；存储器 52 也可以包括非易失性存储器，例如 ROM，快闪存储器，HDD 或 SSD；存储器 52 还可以包括上述种类的存储器的组合。

系统总线 53 可以包括数据总线、电源总线、控制总线和信号状态总线等。本实施例中为了清楚说明，在图 17 中将各种总线都示意

为系统总线 53。

接口电路 51 具体可以是数据节点上的 IO 接口、网络接口或者通信接口。至少一个处理器 50 通过接口电路 51 与其他设备，例如名称节点之间进行数据的交互。

在具体实现过程中，上述如图 9 所示的方法流程中的各步骤均可以通过硬件形式的处理器 50 执行存储器 52 中存储的软件形式的计算机执行指令实现。为避免重复，此处不再赘述。

本发明实施例提供一种计算机可读存储介质，计算机可读存储介质中存储有计算机执行指令，当数据节点的至少一个处理器执行该计算机执行指令时，数据节点执行上述如图 9 所示的控制 IO 带宽的方法。具体的控制 IO 带宽的方法可参见上述如图 9 所示的实施例中的相关描述，此处不再赘述。

本发明实施例提供一种计算机程序产品，该计算机程序产品包括计算机执行指令，该计算机执行指令存储在计算机可读存储介质中。数据节点中的至少一个处理器可以从计算机可读存储介质读取该计算机执行指令，至少一个处理器执行该计算机执行指令使得数据节点实施如图 9 所示的控制 IO 带宽的方法。

上述计算机可读存储介质可以为上述的存储器 52。

本发明实施例提供的数据节点应用于分布式文件系统，该分布式文件系统包括至少一个数据节点，该至少一个数据节点包括该数据节点。该数据节点通过确定该至少一个数据节点中每个数据节点的 IO 带宽和第一租户的 IO 带宽，并根据该至少一个数据节点中每个数据节点的 IO 带宽和第一租户的 IO 带宽，为第一租户分配第一 IO 带宽。本发明实施例提供的数据节点，可以实现为租户分配 IO 带宽，分配 IO 带宽的方式比较灵活；进而，该租户的 IO 带宽为至少一个用户共享，至少一个用户中的每个用户对应至少一个应用，因此属于该租户的多个应用可以共享为该租户分配的 IO 带宽，即使在不同时刻该多个应用可能不同，但该多个应用仍共享为该租户分配的 IO 带宽，因此能够提高分布式文件系统中 IO 带宽的利用率。

如图 18 所示, 本发明实施例提供一种数据节点, 该数据节点应用于分布式文件系统, 该分布式文件系统包括至少一个数据节点, 该至少一个数据节点包括该数据节点。该数据节点包括至少一个处理器 60、接口电路 61、存储器 62 和系统总线 63。

其中, 存储器 62 用于存储计算机执行指令, 至少一个处理器 60、接口电路 61 和存储器 62 通过系统总线 63 相互连接, 当该数据节点运行时, 至少一个处理器 60 执行存储器 62 存储的计算机执行指令, 以使该数据节点执行如图 10 至图 12 任一所示的处理 IO 访问请求的方法。具体的处理 IO 访问请求的方法可参见上述如图 10 至图 12 任一所示的实施例中的相关描述, 此处不再赘述。

处理器 60 具体可以为 CPU。处理器 60 还可以为其他通用处理器、DSP、ASIC、FPGA 或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。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。

处理器 60 还可以为专用处理器, 该专用处理器可以包括具有数据节点其他专用处理功能的芯片。

存储器 62 可以包括易失性存储器, 例如 RAM; 存储器 62 也可以包括非易失性存储器, 例如 ROM, 快闪存储器, HDD 或 SSD; 存储器 62 还可以包括上述种类的存储器的组合。

系统总线 63 可以包括数据总线、电源总线、控制总线和信号状态总线等。本实施例中为了清楚说明, 在图 18 中将各种总线都示意作为系统总线 63。

接口电路 61 具体可以是数据节点上的 IO 接口、网络接口或者通信接口。至少一个处理器 60 通过接口电路 61 与其他设备, 例如名称节点之间进行数据的交互。

在具体实现过程中, 上述如图 10 至图 12 任一所示的方法流程中的各步骤均可以通过硬件形式的处理器 60 执行存储器 62 中存储的软件形式的计算机执行指令实现。为避免重复, 此处不再赘述。

本发明实施例提供一种计算机可读存储介质，计算机可读存储介质中存储有计算机执行指令，当数据节点的至少一个处理器执行该计算机执行指令时，数据节点执行上述如图 10 至图 12 任一所示的处理 IO 访问请求的方法。具体的处理 IO 访问请求的方法可参见上述如图 10 至图 12 任一所示的实施例中的相关描述，此处不再赘述。

本发明实施例提供一种计算机程序产品，该计算机程序产品包括计算机执行指令，该计算机执行指令存储在计算机可读存储介质中。数据节点的至少一个处理器可以从计算机可读存储介质读取该计算机执行指令，至少一个处理器执行该计算机执行指令使得数据节点实施上述如图 10 至图 12 任一所示的处理 IO 访问请求的方法。

上述计算机可读存储介质可以为上述的存储器 62。

本发明实施例提供的数据节点应用于分布式文件系统，该分布式文件系统包括至少一个数据节点，该至少一个数据节点包括该数据节点。该数据节点通过接收租户访问分布式文件系统的 IO 访问请求，并根据该 IO 访问请求确定该 IO 访问请求对应的租户，然后再根据该数据节点已为该租户分配的 IO 带宽，确定是否延迟响应该 IO 访问请求。本发明实施例可以实现在数据节点上为租户分配 IO 带宽，进而数据节点可以根据为租户分配的 IO 带宽确定是否延迟响应与租户对应的 IO 访问请求（即本申请一定会响应 IO 访问请求，可以为立即响应或者延迟响应），因此可以提高响应 IO 访问请求的成功率。

本发明实施例提供一种分布式文件系统，该分布式文件系统包括名称节点和至少一个数据节点，该名称节点为上述如图 13 或图 16 所示的名称节点。

或者，该分布式文件系统包括至少一个数据节点，该至少一个数据节点包括第一数据节点，该第一数据节点为上述如图 14 或图 17 所示的数据节点。

或者，该分布式文件系统包括至少一个数据节点，该至少一个数据节点包括第一数据节点，该第一数据节点为上述如图 15 或图 18 所示的数据节点。

本发明实施例提供的分布式文件系统可以为上述如图 2 或图 3 所示的分布式文件系统。具体的，可以参见上述实施例中对如图 2 或图 3 所示的分布式文件系统的相关描述，此处不再赘述。

本发明实施例提供的分布式文件系统包括名称节点和至少一个数据节点。该名称节点通过确定该至少一个数据节点中每个数据节点的 IO 带宽和第一租户的 IO 带宽，并根据该至少一个数据节点中每个数据节点的 IO 带宽和第一租户的 IO 带宽，指示该至少一个数据节点为第一租户分配该至少一个 IO 带宽。本发明实施例提供的名称节点，可以指示数据节点为租户分配 IO 带宽，分配 IO 带宽的方式比较灵活；进而，该租户的 IO 带宽为至少一个用户共享，至少一个用户中的每个用户对应至少一个应用，因此属于该租户的多个应用可以共享为该租户分配的 IO 带宽，即使在不同时刻该多个应用可能不同，但该多个应用仍共享为该租户分配的 IO 带宽，因此能够提高分布式文件系统中 IO 带宽的利用率。

本发明实施例提供的分布式文件系统包括至少一个数据节点，该至少一个数据节点包括第一数据节点。第一数据节点通过确定该至少一个数据节点中每个数据节点的 IO 带宽和第一租户的 IO 带宽，并根据该至少一个数据节点中每个数据节点的 IO 带宽和第一租户的 IO 带宽，为第一租户分配第一 IO 带宽。本发明实施例提供的数据节点，可以实现为租户分配 IO 带宽，分配 IO 带宽的方式比较灵活；进而，该租户的 IO 带宽为至少一个用户共享，至少一个用户中的每个用户对应至少一个应用，因此属于该租户的多个应用可以共享为该租户分配的 IO 带宽，即使在不同时刻该多个应用可能不同，但该多个应用仍共享为该租户分配的 IO 带宽，因此能够提高分布式文件系统中 IO 带宽的利用率。

本发明实施例提供的分布式文件系统包括至少一个数据节点，

该至少一个数据节点包括第一数据节点。第一数据节点通过接收租户访问分布式文件系统的 IO 访问请求，并根据该 IO 访问请求确定该 IO 访问请求对应的租户，然后再根据第一数据节点已为该租户分配的 IO 带宽，确定是否延迟响应该 IO 访问请求。本发明实施例可以实现在数据节点上为租户分配 IO 带宽，进而数据节点可以根据为租户分配的 IO 带宽确定是否延迟响应与租户对应的 IO 访问请求(即本申请一定会响应 IO 访问请求，可以为立即响应或者延迟响应)，因此可以提高响应 IO 访问请求的成功率。

通过以上的实施方式的描述，所属领域的技术人员可以清楚地了解到，为描述的方便和简洁，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。上述描述的系统，装置和单元的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

在本申请所提供的几个实施例中，应该理解到，所揭露的系统，装置和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的，例如，所述模块或单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口，装置或单元的间接耦合或通信连接，可以是电性，机械或其它的形式。

所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

另外，在本发明各个实施例中的各功能单元可以集成在一个处

理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现，也可以采用软件功能单元的形式实现。

所述集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用，可以存储在一个计算机可读取存储介质中。基于这样的理解，本发明的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储介质中，包括若干指令用以使得一台计算机设备（可以是个人计算机，服务器，或者网络设备）或处理器执行本发明各个实施例所述方法的全部或部分步骤。而前述的存储介质包括：快闪存储器、移动硬盘、只读存储器、随机存取存储器、磁碟或者光盘等各种可以存储程序代码的介质。

以上所述，仅为本发明的具体实施方式，但本发明的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本发明揭露的技术范围内，可轻易想到变化或替换，都应涵盖在本发明的保护范围之内。因此，本发明的保护范围应所述以权利要求的保护范围为准。

## 权 利 要 求 书

1、一种控制输入输出 IO 带宽的方法，其特征在于，所述方法应用于分布式文件系统，所述分布式文件系统包括名称节点和至少一个数据节点；所述方法包括：

所述名称节点确定所述至少一个数据节点中每个数据节点的 IO 带宽；

所述名称节点确定第一租户的 IO 带宽；

所述名称节点基于所述至少一个数据节点中每个数据节点的 IO 带宽以及所述第一租户的 IO 带宽，指示所述至少一个数据节点为所述第一租户分配至少一个 IO 带宽，其中，所述至少一个 IO 带宽与所述至少一个数据节点一一对应，所述至少一个 IO 带宽中的每个 IO 带宽小于或者等于对应的数据节点的 IO 带宽，所述至少一个 IO 带宽中的每个 IO 带宽大于 0。

2、根据权利要求 1 所述的方法，其特征在于，

所述至少一个 IO 带宽的总和小于或者等于所述第一租户的 IO 带宽；或者，

所述至少一个 IO 带宽的总和小于或者等于所述第一租户的 IO 带宽与额外带宽之和。

3、根据权利要求 1 或 2 所述的方法，其特征在于，所述至少一个数据节点包括第一数据节点，所述第一数据节点在所述名称节点的指示下为所述第一租户分配的 IO 带宽是第一 IO 带宽，所述第一 IO 带宽是所述至少一个 IO 带宽中与所述第一数据节点对应的 IO 带宽；

所述名称节点基于所述至少一个数据节点中每个数据节点的 IO 带宽以及所述第一租户的 IO 带宽，指示所述至少一个数据节点为所述第一租户分配至少一个 IO 带宽之后，所述方法还包括：

所述名称节点确定所述第一租户分别在所述至少一个 IO 带宽中使用的 IO 带宽为至少一个使用带宽，所述至少一个使用带宽与所述至少一个 IO 带宽一一对应；

所述名称节点根据所述至少一个使用带宽，指示所述至少一个数

据节点分别调整所述至少一个 IO 带宽中对应的 IO 带宽,或者指示所述第一数据节点调整所述第一 IO 带宽。

4、根据权利要求 1 至 3 任一项所述的方法,其特征在于,所述名称节点基于所述至少一个数据节点中每个数据节点的 IO 带宽以及所述第一租户的 IO 带宽,指示所述至少一个数据节点为所述第一租户分配至少一个 IO 带宽之后,所述方法还包括:

所述名称节点确定所述第一租户分别在所述至少一个 IO 带宽中使用的 IO 带宽为至少一个使用带宽,所述至少一个使用带宽与所述至少一个 IO 带宽一一对应;

当所述至少一个 IO 带宽之和与所述至少一个使用带宽之和的差值小于或者等于第一阈值时,所述名称节点指示第二数据节点为所述第一租户分配 IO 带宽,所述至少一个数据节点不包括所述第二数据节点。

5、根据权利要求 1 至 4 任一项所述的方法,其特征在于,所述名称节点基于所述至少一个数据节点中每个数据节点的 IO 带宽以及所述第一租户的 IO 带宽,指示所述至少一个数据节点为所述第一租户分配至少一个 IO 带宽之后,所述方法还包括:

所述名称节点确定所述第一租户分别在所述至少一个 IO 带宽中使用的 IO 带宽为至少一个使用带宽,所述至少一个使用带宽与所述至少一个 IO 带宽一一对应;

当所述至少一个 IO 带宽之和与所述至少一个使用带宽之和的差值大于第一阈值时,所述名称节点指示第三数据节点将所述第三数据节点为所述第一租户分配的 IO 带宽调整为 0,所述至少一个数据节点包括所述第三数据节点。

6、根据权利要求 1 至 5 任一项所述的方法,其特征在于,所述至少一个数据节点包括第一数据节点,所述第一数据节点在所述名称节点的指示下为所述第一租户分配的 IO 带宽是第一 IO 带宽,所述第一 IO 带宽是所述至少一个 IO 带宽中与所述第一数据节点对应的 IO 带宽;

所述名称节点基于所述至少一个数据节点中每个数据节点的 IO 带宽以及所述第一租户的 IO 带宽，指示所述至少一个数据节点为所述第一租户分配至少一个 IO 带宽之后，所述方法还包括：

所述名称节点确定所述第一租户使用的所述第一 IO 带宽中的 IO 带宽；

所述名称节点根据所述第一租户使用的所述第一 IO 带宽中的 IO 带宽，指示所述第一数据节点调整所述第一 IO 带宽。

7、根据权利要求 1 至 6 任一项所述的方法，其特征在于， $m$  个租户包括所述第一租户， $m$  为正整数；

所述名称节点确定第一租户的 IO 带宽，包括：

所述名称节点根据所述  $m$  个租户中每个租户的权重，从所述至少一个数据节点中每个数据节点的 IO 带宽的总和，确定所述第一租户的 IO 带宽。

8、一种处理输入输出 IO 访问请求的方法，其特征在于，所述方法应用于分布式文件系统，所述分布式文件系统包括至少一个数据节点，所述至少一个数据节点包括第一数据节点；所述方法包括：

所述第一数据节点接收 IO 访问请求，所述 IO 访问请求是租户访问所述分布式文件系统的请求；

所述第一数据节点根据所述 IO 访问请求，确定所述 IO 访问请求对应的所述租户；

所述第一数据节点根据所述第一数据节点已为所述租户分配的 IO 带宽，确定是否延迟响应所述 IO 访问请求。

9、根据权利要求 8 所述的方法，其特征在于，所述第一数据节点根据所述第一数据节点已为所述租户分配的 IO 带宽，确定是否延迟响应所述 IO 访问请求，包括：

所述第一数据节点确定所述第一数据节点已为所述租户分配的 IO 带宽与所述租户在所述第一数据节点使用的 IO 带宽的差值；

所述第一数据节点在确定所述差值大于第三阈值时，响应所述 IO 访问请求。

10、根据权利要求 8 或 9 所述的方法，其特征在于，所述第一数据节点根据所述第一数据节点已为所述租户分配的 IO 带宽，确定是否延迟响应所述 IO 访问请求，包括：

所述第一数据节点确定所述第一数据节点已为所述租户分配的 IO 带宽与所述租户在所述第一数据节点使用的 IO 带宽的差值；

所述第一数据节点在确定所述差值小于或者等于第三阈值时，延迟响应所述 IO 访问请求。

11、根据权利要求 8 至 10 任一项所述的方法，其特征在于，所述第一数据节点为所述租户分配 IO 带宽之前，所述方法还包括：

所述第一数据节点接收所述名称节点的指示；

所述第一数据节点为所述租户分配 IO 带宽，包括：

所述第一数据节点根据所述名称节点的指示，为所述租户分配 IO 带宽。

12、根据权利要求 8 至 11 任一项所述的方法，其特征在于，所述第一数据节点根据所述 IO 访问请求，确定所述 IO 访问请求对应的所述租户，包括：

所述第一数据节点根据所述 IO 访问请求携带的租户标识，确定所述 IO 访问请求对应的所述租户。

13、根据权利要求 8 至 11 任一项所述的方法，其特征在于，所述分布式文件系统管理的属于所述租户的文件所具有的文件属性中包含租户标识；

所述第一数据节点根据所述 IO 访问请求，确定所述 IO 访问请求对应的所述租户，包括：

所述第一数据节点根据所述 IO 访问请求指定的文件，获取所述 IO 访问请求指定的文件所具有的文件属性；

所述第一数据节点根据获取的文件属性包含的所述租户标识，确定所述 IO 访问请求对应的所述租户。

14、一种名称节点，其特征在于，所述名称节点应用于分布式文

件系统，所述分布式文件系统包括所述名称节点和至少一个数据节点，所述名称节点包括确定单元和指示单元：

所述确定单元，用于确定所述至少一个数据节点中每个数据节点的 IO 带宽和第一租户的 IO 带宽；

所述指示单元，用于基于所述确定单元确定的所述至少一个数据节点中每个数据节点的 IO 带宽以及所述第一租户的 IO 带宽，指示所述至少一个数据节点为所述第一租户分配至少一个 IO 带宽，其中，所述至少一个 IO 带宽与所述至少一个数据节点一一对应，所述至少一个 IO 带宽中的每个 IO 带宽小于或者等于对应的数据节点的 IO 带宽，所述至少一个 IO 带宽中的每个 IO 带宽大于 0。

15、根据权利要求 14 所述的名称节点，其特征在于，

所述至少一个 IO 带宽的总和小于或者等于所述第一租户的 IO 带宽；或者，

所述至少一个 IO 带宽的总和小于或者等于所述第一租户的 IO 带宽与额外带宽之和。

16、根据权利要求 14 或 15 所述的名称节点，其特征在于，所述至少一个数据节点包括第一数据节点，所述第一数据节点在所述指示单元的指示下为所述第一租户分配的 IO 带宽是第一 IO 带宽，所述第一 IO 带宽是所述至少一个 IO 带宽中与所述第一数据节点对应的 IO 带宽；

所述确定单元，还用于在所述指示单元指示所述至少一个数据节点为所述第一租户分配至少一个 IO 带宽之后，确定所述第一租户分别在所述至少一个 IO 带宽中使用的 IO 带宽为至少一个使用带宽，所述至少一个使用带宽与所述至少一个 IO 带宽一一对应；

所述指示单元，还用于根据所述确定单元确定的所述至少一个使用带宽，指示所述至少一个数据节点分别调整所述至少一个 IO 带宽中对应的 IO 带宽，或者指示所述第一数据节点调整所述第一 IO 带宽。

17、根据权利要求 14 至 16 任一项所述的名称节点，其特征在于，所述确定单元，还用于在所述指示单元指示所述至少一个数据节

点为所述第一租户分配至少一个 IO 带宽之后, 确定所述第一租户分别在所述至少一个 IO 带宽中使用的 IO 带宽为至少一个使用带宽, 所述至少一个使用带宽与所述至少一个 IO 带宽一一对应;

所述指示单元, 还用于当所述至少一个 IO 带宽之和与所述确定单元确定的所述至少一个使用带宽之和的差值小于或者等于第一阈值时, 指示第二数据节点为所述第一租户分配 IO 带宽, 所述至少一个数据节点不包括所述第二数据节点。

18、根据权利要求 14 至 17 任一项所述的名称节点, 其特征在于,

所述确定单元, 还用于在所述指示单元指示所述至少一个数据节点为所述第一租户分配至少一个 IO 带宽之后, 确定所述第一租户分别在所述至少一个 IO 带宽中使用的 IO 带宽为至少一个使用带宽, 所述至少一个使用带宽与所述至少一个 IO 带宽一一对应;

所述指示单元, 还用于当所述至少一个 IO 带宽之和与所述确定单元确定的所述至少一个使用带宽之和的差值大于第一阈值时, 指示第三数据节点将所述第三数据节点为所述第一租户分配的 IO 带宽调整为 0, 所述至少一个数据节点包括所述第三数据节点。

19、根据权利要求 14 至 18 任一项所述的名称节点, 其特征在于, 所述至少一个数据节点包括第一数据节点, 所述第一数据节点在所述指示单元的指示下为所述第一租户分配的 IO 带宽是第一 IO 带宽, 所述第一 IO 带宽是所述至少一个 IO 带宽中与所述第一数据节点对应的 IO 带宽;

所述确定单元, 还用于在所述指示单元指示所述至少一个数据节点为所述第一租户分配至少一个 IO 带宽之后, 确定所述第一租户使用的所述第一 IO 带宽中的 IO 带宽;

所述指示单元, 还用于根据所述确定单元确定的所述第一租户使用的所述第一 IO 带宽中的 IO 带宽, 指示所述第一数据节点调整所述第一 IO 带宽。

20、根据权利要求 14 至 19 任一项所述的名称节点, 其特征在于, m 个租户包括所述第一租户, m 为正整数;

所述确定单元，具体用于根据所述  $m$  个租户中每个租户的权重，从所述至少一个数据节点中每个数据节点的 IO 带宽的总和，确定所述第一租户的 IO 带宽。

21、一种数据节点，其特征在于，所述数据节点应用于分布式文件系统，所述分布式文件系统包括至少一个数据节点，所述至少一个数据节点包括所述数据节点；所述数据节点包括接收单元、确定单元和分配单元：

所述接收单元，用于接收 IO 访问请求，所述 IO 访问请求是租户访问所述分布式文件系统的请求；

所述确定单元，用于根据所述接收单元接收的所述 IO 访问请求，确定所述 IO 访问请求对应的所述租户；并根据所述分配单元已为所述租户分配的 IO 带宽，确定是否延迟响应所述 IO 访问请求。

22、根据权利要求 21 所述的数据节点，其特征在于，

所述确定单元，具体用于确定所述分配单元已为所述租户分配的 IO 带宽与所述租户在所述数据节点使用的 IO 带宽的差值；并在确定所述差值大于第三阈值时，响应所述 IO 访问请求。

23、根据权利要求 21 或 22 所述的数据节点，其特征在于，

所述确定单元，具体用于确定所述分配单元已为所述租户分配的 IO 带宽与所述租户在所述数据节点使用的 IO 带宽的差值；并在确定所述差值小于或者等于第三阈值时，延迟响应所述 IO 访问请求。

24、根据权利要求 21 至 23 任一项所述的数据节点，其特征在于，

所述接收单元，还用于在所述分配单元为所述租户分配 IO 带宽之前，接收所述名称节点的指示；

所述分配单元，用于根据所述接收单元接收的所述名称节点的指示，为所述租户分配 IO 带宽。

25、根据权利要求 21 至 24 任一项所述的数据节点，其特征在于，

所述确定单元，具体用于根据所述接收单元接收的所述 IO 访问请求携带的租户标识，确定所述 IO 访问请求对应的所述租户。

26、根据权利要求 21 至 24 任一项所述的数据节点，其特征在于，

所述分布式文件系统管理的属于所述租户的文件所具有的文件属性中包含租户标识；

所述确定单元，具体用于根据所述接收单元接收的所述 IO 访问请求指定的文件，获取所述 IO 访问请求指定的文件所具有的文件属性；并根据获取的文件属性包含的所述租户标识，确定所述 IO 访问请求对应的所述租户。

27、一种名称节点，其特征在于，所述名称节点应用于分布式文件系统，所述分布式文件系统包括所述名称节点和至少一个数据节点，所述名称节点包括至少一个处理器、存储器和系统总线；

所述存储器用于存储计算机执行指令，所述至少一个处理器和所述存储器通过所述系统总线相互连接，当所述名称节点运行时，所述至少一个处理器执行所述存储器存储的所述计算机执行指令，以使所述名称节点执行如权利要求 1 至 7 任一项所述的控制输入输出 IO 带宽的方法。

28、一种数据节点，其特征在于，所述数据节点应用于分布式文件系统，所述分布式文件系统包括至少一个数据节点，所述至少一个数据节点包括所述数据节点；所述数据节点包括至少一个处理器、存储器和系统总线；

所述存储器用于存储计算机执行指令，所述至少一个处理器和所述存储器通过所述系统总线相互连接，当所述数据节点运行时，所述至少一个处理器执行所述存储器存储的所述计算机执行指令，以使所述数据节点执行如权利要求 8 至 13 任一项所述的处理输入输出 IO 访问请求的方法。

29、一种分布式文件系统，其特征在于，所述分布式文件系统包括名称节点和至少一个数据节点；其中，所述名称节点为如权利要求 14 至 20 任一项或 27 所述的名称节点。

30、一种分布式文件系统，其特征在于，所述分布式文件系统包括至少一个数据节点；其中，所述至少一个数据节点包括第一数据节点，所述第一数据节点为如权利要求 21 至 26 任一项或 28 所述的数

据节点。

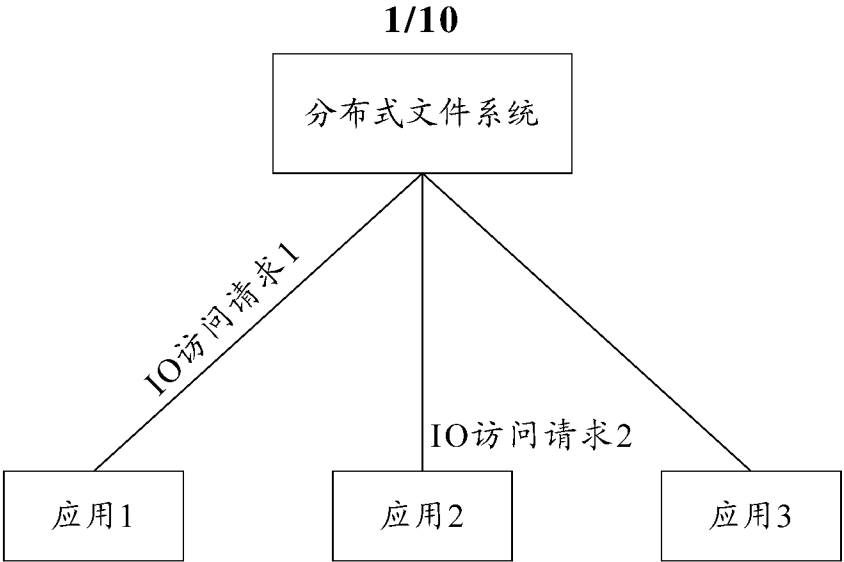


图 1

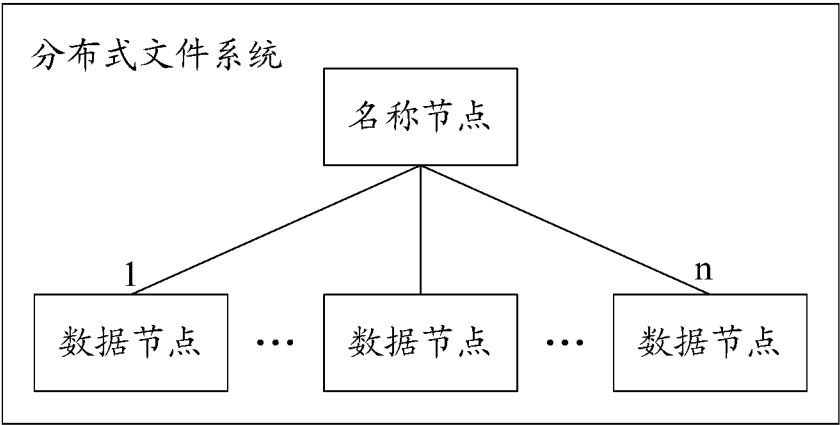


图 2

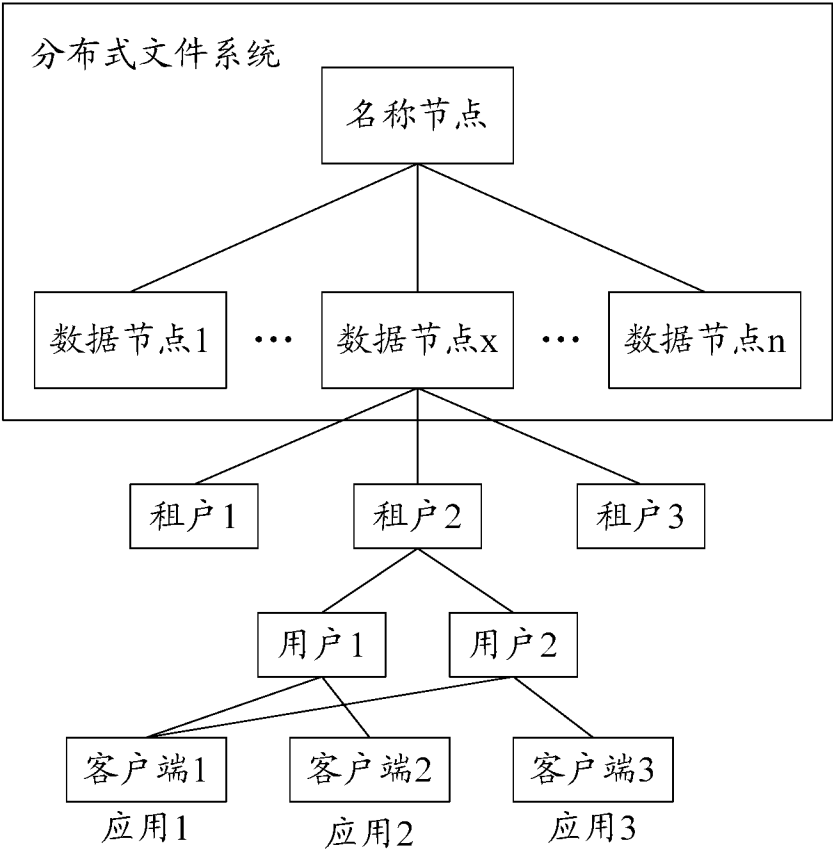


图 3

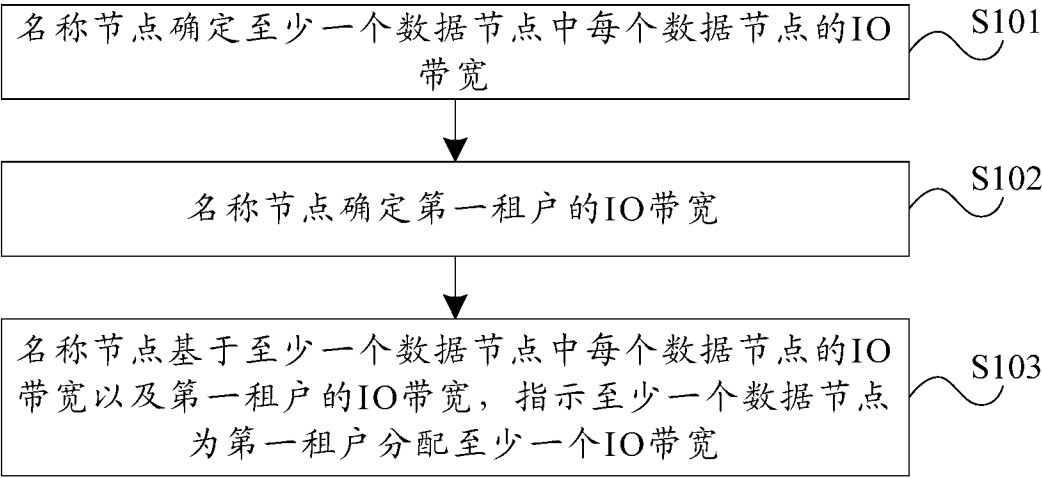


图 4

3/10

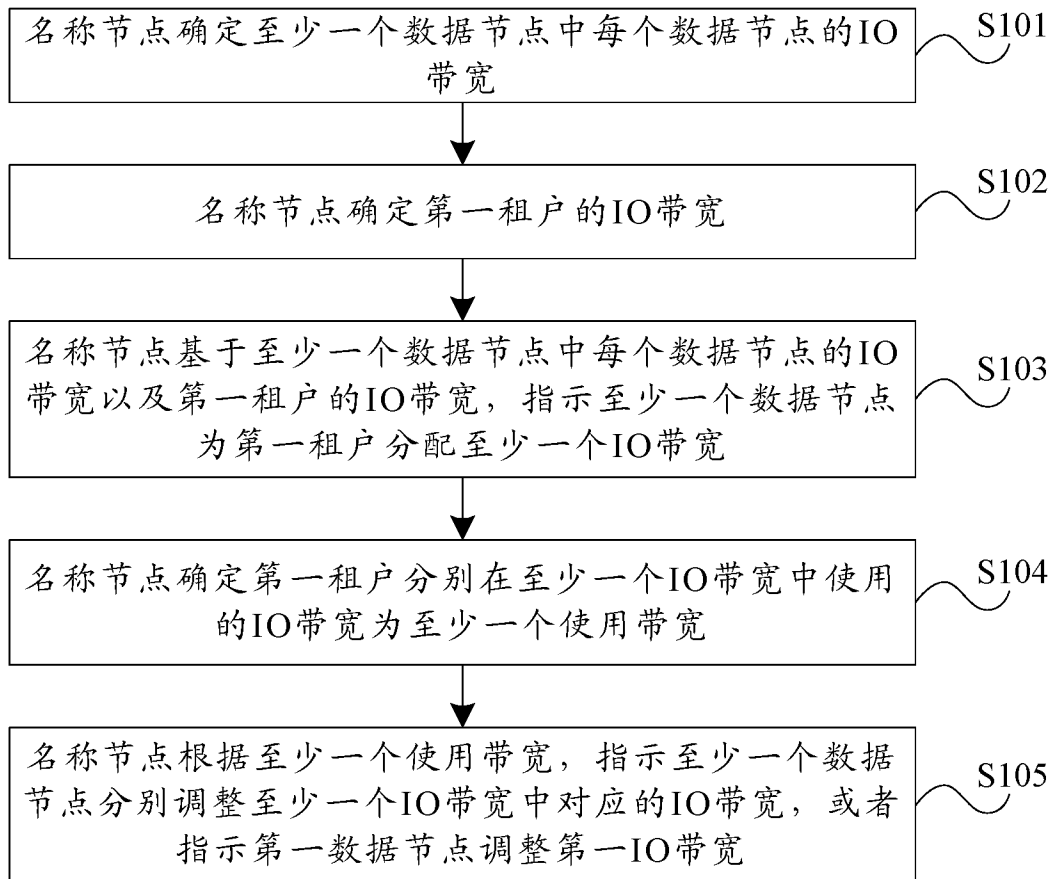


图 5

## 4/10

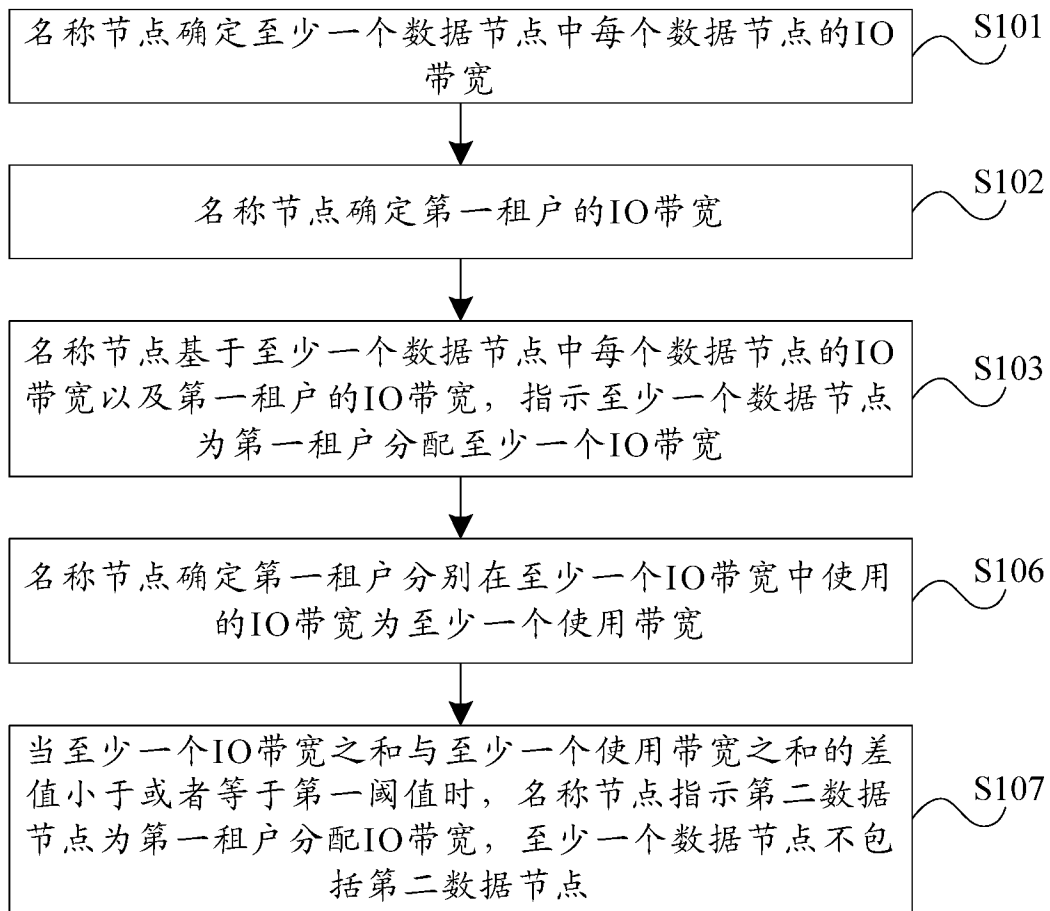


图 6

## 5/10

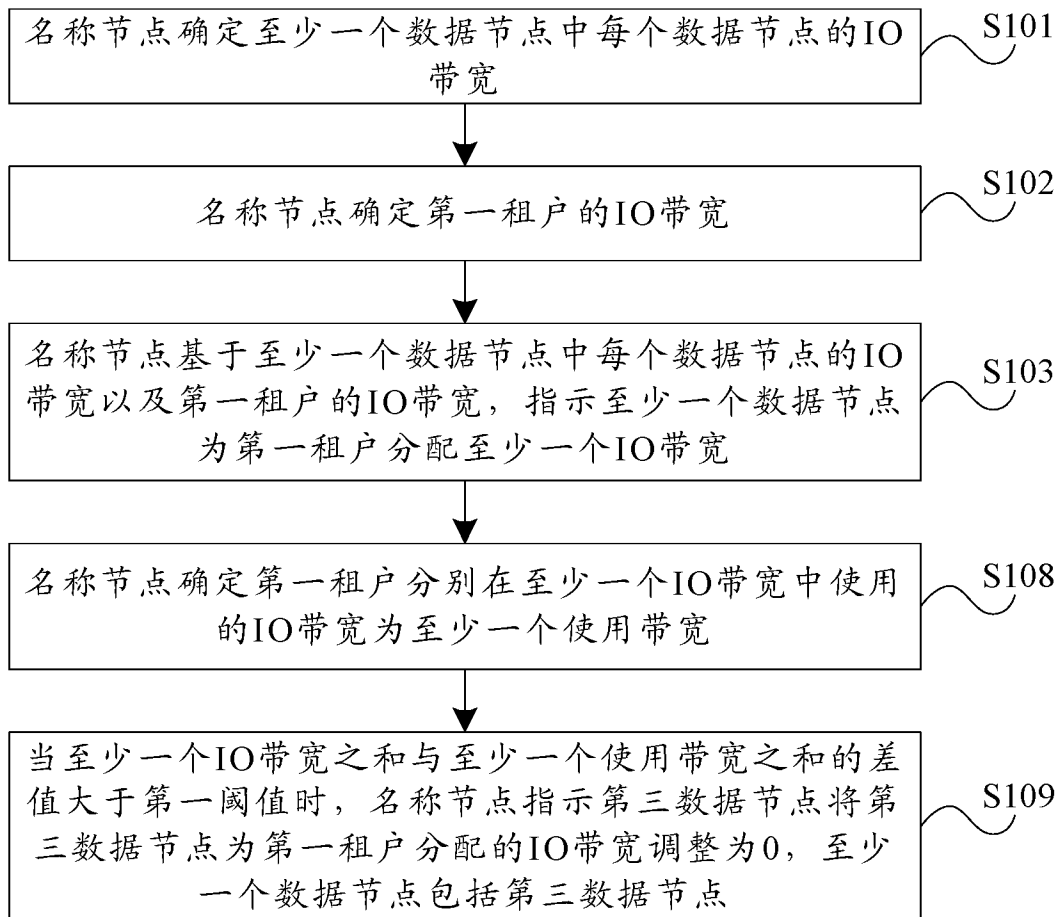


图 7

## 6/10

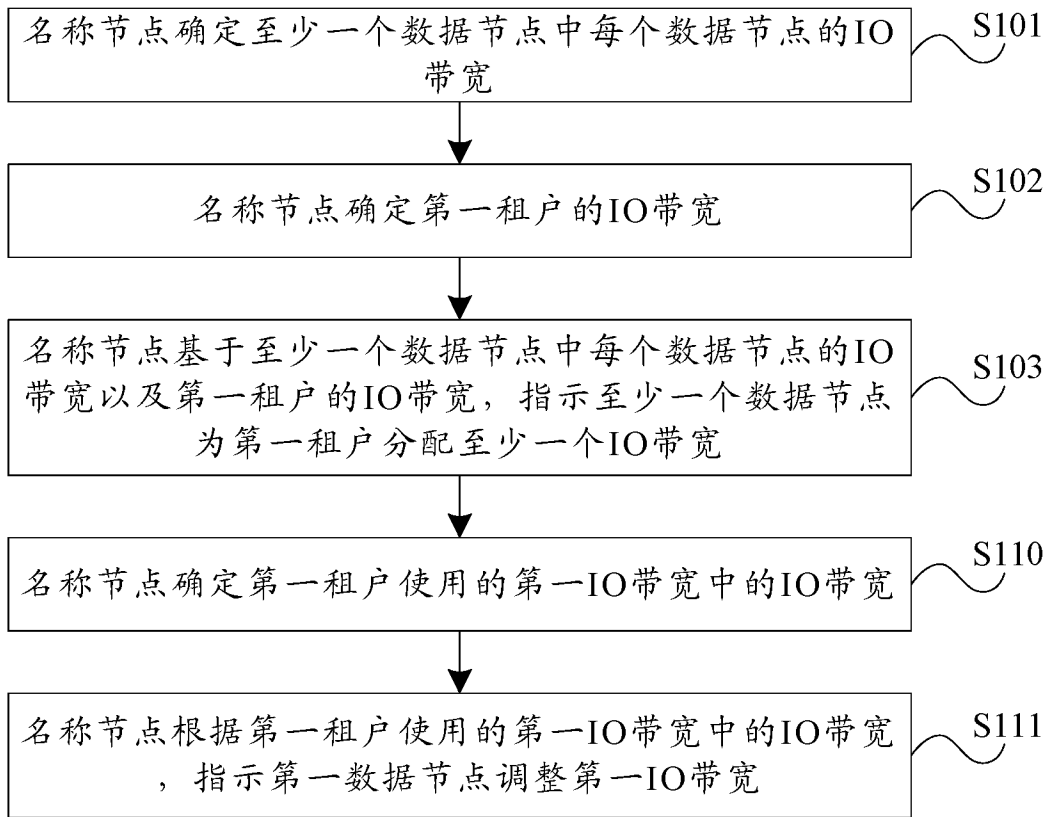


图 8

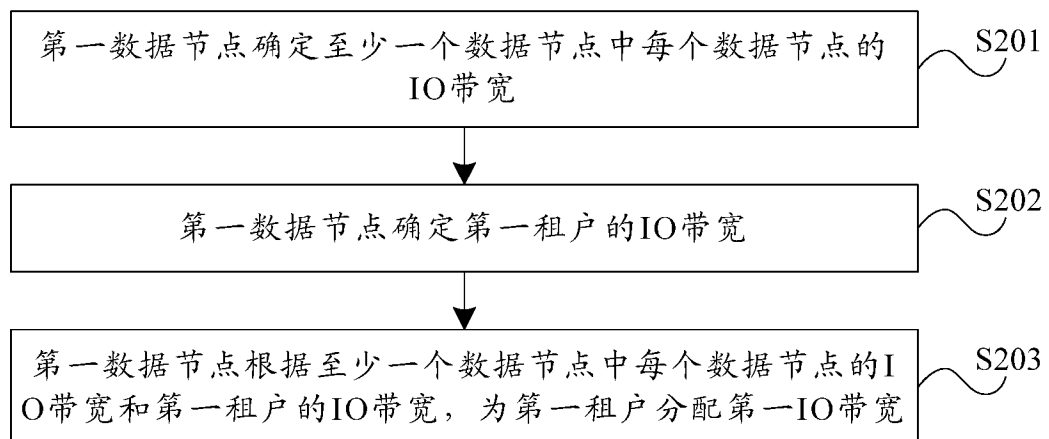


图 9

7/10

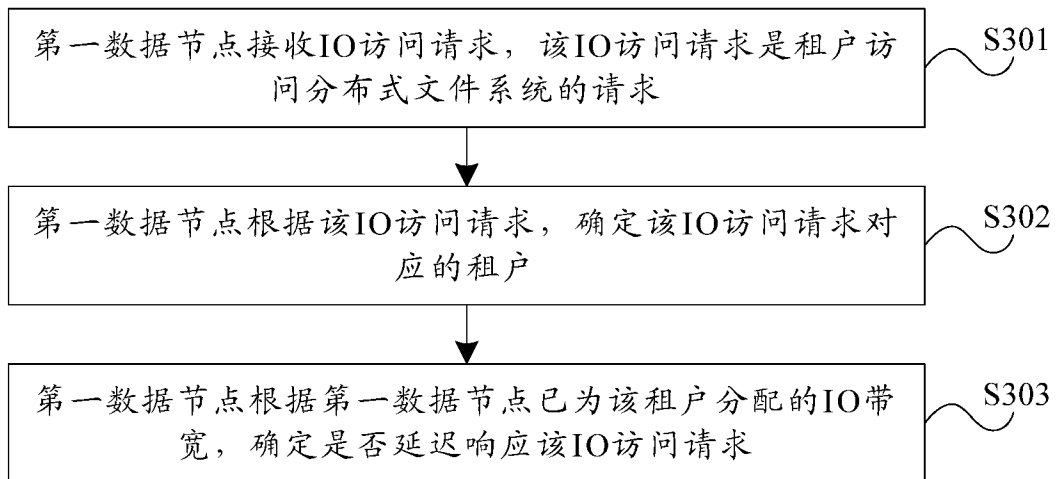


图 10

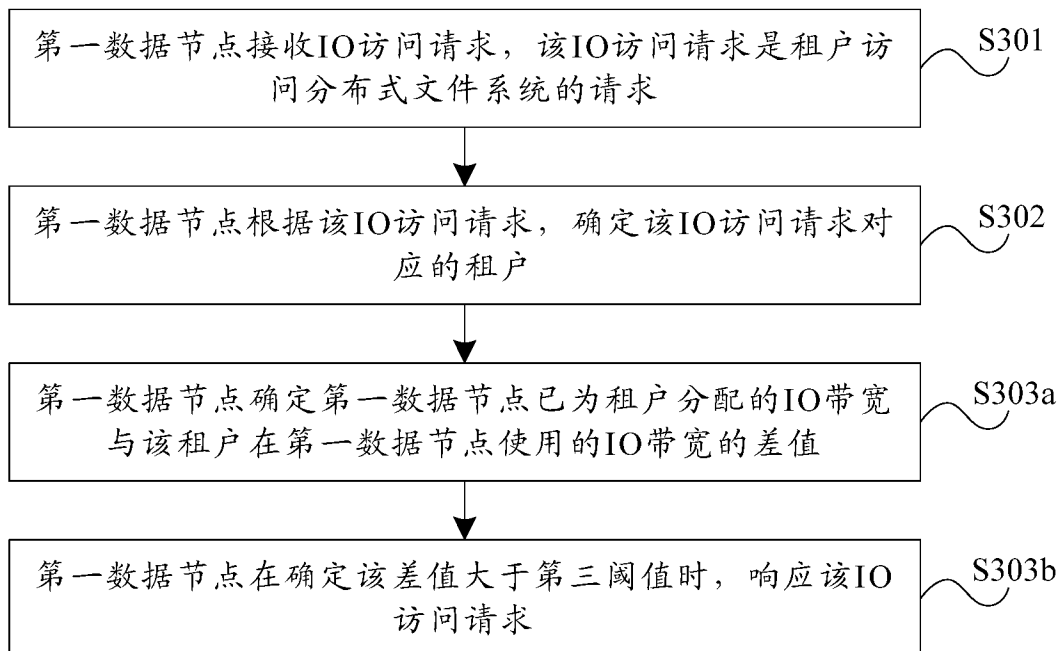


图 11

## 8/10

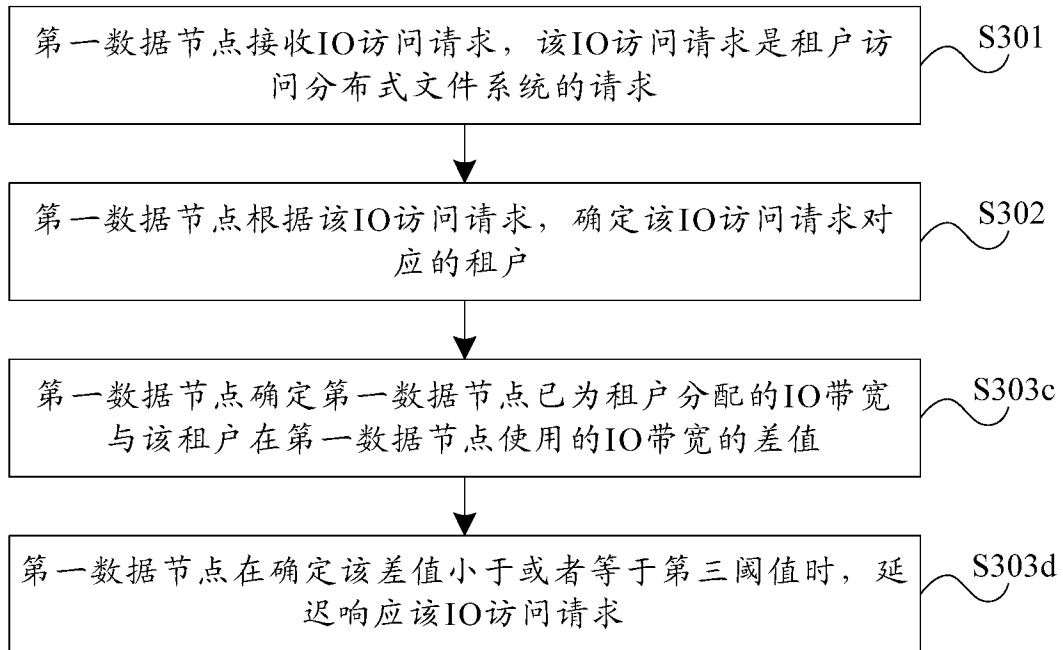


图 12

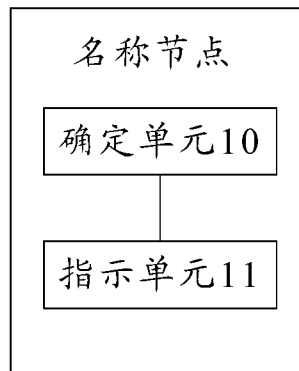


图 13

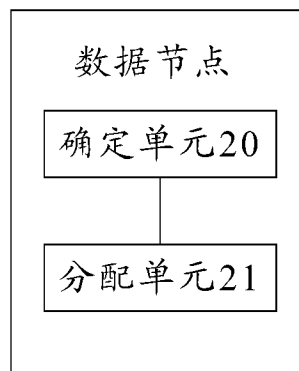


图 14

9/10

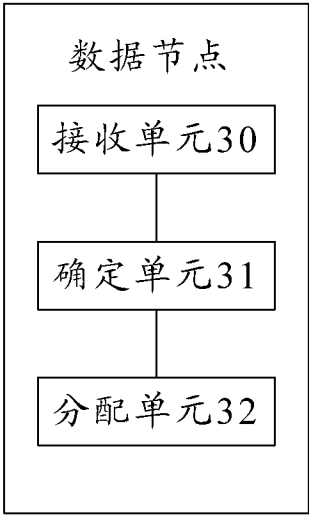


图 15

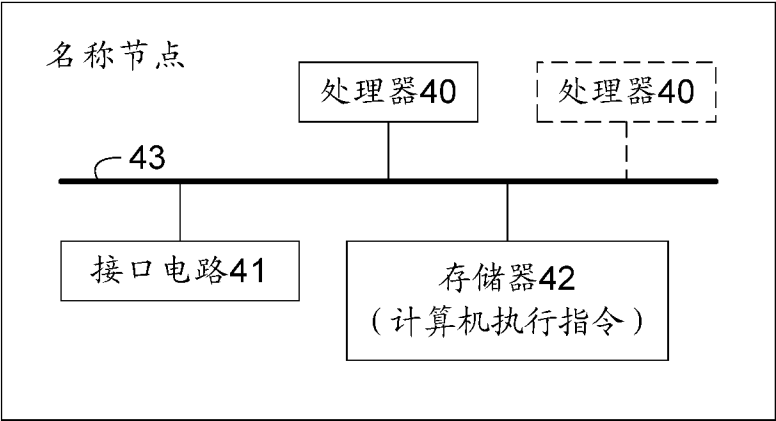


图 16

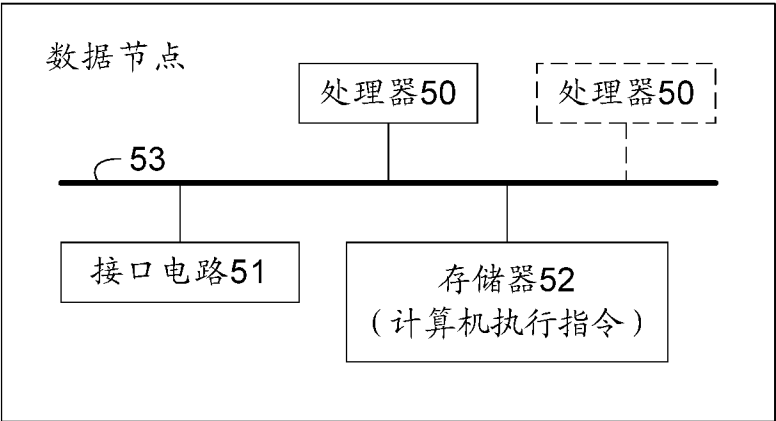


图 17

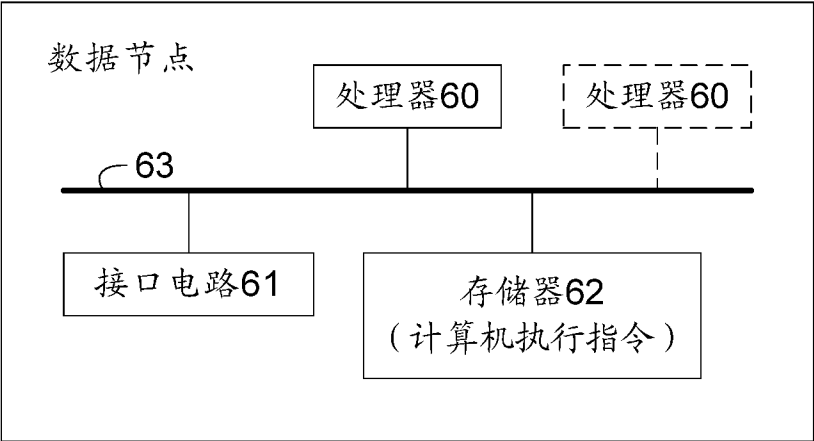


图 18

# INTERNATIONAL SEARCH REPORT

International application No.

**PCT/CN2016/089836**

## A. CLASSIFICATION OF SUBJECT MATTER

G06F 9/50 (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)

G06F 9/-

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)

CNXTX, CNABS, CNKI: bandwidth, tenant, rent, data node, name node, delay, assign

WPI, EPODOC: tenant, datanode, namenode, delay, assign, bandwidth

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                              | Relevant to claim No. |
|-----------|---------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| X         | CN 103729250 A (INTERNATIONAL BUSINESS MACHINES CORPORATION), 16 April 2014 (16.04.2014), description, paragraphs [0002]-[0032] | 1-30                  |
| A         | CN 102646121 A (WUHAN UNIVERSITY), 22 August 2012 (22.08.2012), the whole document                                              | 1-30                  |
| A         | CN 102857558 A (GUANGDONG POLYTECHNIC OF SCIENCE AND TECHNOLOGY), 02 January 2013 (02.01.2013), the whole document              | 1-30                  |
| A         | CN 103782556 A (HUAWEI TECHNOLOGIES CO., LTD.), 07 May 2014 (07.05.2014), the whole document                                    | 1-30                  |

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

|                                                                                                                                                                         |                                                                                                                                                                                                                                                  |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| * Special categories of cited documents:                                                                                                                                | “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                                              |
| “A” document defining the general state of the art which is not considered to be of particular relevance                                                                | “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                                                                     |
| “E” earlier application or patent but published on or after the international filing date                                                                               | “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 |
| “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) | “&” document member of the same patent family                                                                                                                                                                                                    |
| “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                                                                  |                                                                                                                                                                                                                                                  |

|                                                                                                                                                                                                                    |                                                                                       |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|
| Date of the actual completion of the international search<br>09 April 2017 (09.04.2017)                                                                                                                            | Date of mailing of the international search report<br><b>02 May 2017 (02.05.2017)</b> |
| Name and mailing address of the ISA/CN:<br>State Intellectual Property Office of the P. R. China<br>No. 6, Xitucheng Road, Jimenqiao<br>Haidian District, Beijing 100088, China<br>Facsimile No.: (86-10) 62019451 | Authorized officer<br><b>HANG, Xuemeng</b><br>Telephone No.: (86-10) <b>62089103</b>  |

**INTERNATIONAL SEARCH REPORT**  
Information on patent family members

International application No.

**PCT/CN2016/089836**

| Patent Documents referred<br>in the Report | Publication Date | Patent Family    | Publication Date |
|--------------------------------------------|------------------|------------------|------------------|
| CN 103729250 A                             | 16 April 2014    | US 2014108639 A1 | 17 April 2014    |
|                                            |                  | US 9253053 B2    | 02 February 2016 |
| CN 102646121 A                             | 22 August 2012   | None             |                  |
| CN 102857558 A                             | 02 January 2013  | CN 102857558 B   | 25 November 2015 |
| CN 103782556 A                             | 07 May 2014      | WO 2015018035 A1 | 12 February 2015 |
|                                            |                  | EP 3024184 A1    | 25 May 2016      |
|                                            |                  | CN 103782556 B   | 07 December 2016 |
|                                            |                  | EP 3024184 A4    | 27 July 2016     |

| <b>A. 主题的分类</b><br>G06F 9/50 (2006.01) i<br><br>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>B. 检索领域</b><br>检索的最低限度文献 (标明分类系统和分类号)<br>G06F 9/-<br><br>包含在检索领域中的除最低限度文献以外的检索文献<br><br>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))<br>CNTXT, CNABS, CNKI: 带宽, 租户, 租用, 数据节点, 名称节点, 延迟, 分配 WPI, EPDOC: tenant, datanode, namenode, delay, assign, bandwidth                                                                                                                                                                                                                                                                                                                                                      |                                                                                                                                                                                     |                                              |                                                                                                                                                                                                            |                                                                                                                                                                                     |         |   |                                                                               |      |   |                                                           |      |   |                                                                |      |   |                                                              |      |
| <b>C. 相关文件</b> <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>X</td> <td>CN 103729250 A (国际商业机器公司) 2014年 4月 16日 (2014 - 04 - 16)<br/>说明书第[0002]-[0032]段</td> <td>1-30</td> </tr> <tr> <td>A</td> <td>CN 102646121 A (武汉大学) 2012年 8月 22日 (2012 - 08 - 22)<br/>全文</td> <td>1-30</td> </tr> <tr> <td>A</td> <td>CN 102857558 A (广东科学技术职业学院) 2013年 1月 2日 (2013 - 01 - 02)<br/>全文</td> <td>1-30</td> </tr> <tr> <td>A</td> <td>CN 103782556 A (华为技术有限公司) 2014年 5月 7日 (2014 - 05 - 07)<br/>全文</td> <td>1-30</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 | X | CN 103729250 A (国际商业机器公司) 2014年 4月 16日 (2014 - 04 - 16)<br>说明书第[0002]-[0032]段 | 1-30 | A | CN 102646121 A (武汉大学) 2012年 8月 22日 (2012 - 08 - 22)<br>全文 | 1-30 | A | CN 102857558 A (广东科学技术职业学院) 2013年 1月 2日 (2013 - 01 - 02)<br>全文 | 1-30 | A | CN 103782556 A (华为技术有限公司) 2014年 5月 7日 (2014 - 05 - 07)<br>全文 | 1-30 |
| 类 型*                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | 引用文件, 必要时, 指明相关段落                                                                                                                                                                   | 相关的权利要求                                      |                                                                                                                                                                                                            |                                                                                                                                                                                     |         |   |                                                                               |      |   |                                                           |      |   |                                                                |      |   |                                                              |      |
| X                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | CN 103729250 A (国际商业机器公司) 2014年 4月 16日 (2014 - 04 - 16)<br>说明书第[0002]-[0032]段                                                                                                       | 1-30                                         |                                                                                                                                                                                                            |                                                                                                                                                                                     |         |   |                                                                               |      |   |                                                           |      |   |                                                                |      |   |                                                              |      |
| A                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | CN 102646121 A (武汉大学) 2012年 8月 22日 (2012 - 08 - 22)<br>全文                                                                                                                           | 1-30                                         |                                                                                                                                                                                                            |                                                                                                                                                                                     |         |   |                                                                               |      |   |                                                           |      |   |                                                                |      |   |                                                              |      |
| A                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | CN 102857558 A (广东科学技术职业学院) 2013年 1月 2日 (2013 - 01 - 02)<br>全文                                                                                                                      | 1-30                                         |                                                                                                                                                                                                            |                                                                                                                                                                                     |         |   |                                                                               |      |   |                                                           |      |   |                                                                |      |   |                                                              |      |
| A                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | CN 103782556 A (华为技术有限公司) 2014年 5月 7日 (2014 - 05 - 07)<br>全文                                                                                                                        | 1-30                                         |                                                                                                                                                                                                            |                                                                                                                                                                                     |         |   |                                                                               |      |   |                                                           |      |   |                                                                |      |   |                                                              |      |
| <input type="checkbox"/> 其余文件在C栏的续页中列出。 <input checked="" type="checkbox"/> 见同族专利附件。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                                                                                                                                                                                     |                                              |                                                                                                                                                                                                            |                                                                                                                                                                                     |         |   |                                                                               |      |   |                                                           |      |   |                                                                |      |   |                                                              |      |
| <table border="0"> <tr> <td>           * 引用文件的具体类型:<br/>           “A” 认为不特别相关的表示了现有技术一般状态的文件<br/>           “E” 在国际申请日的当天或之后公布的在先申请或专利<br/>           “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)<br/>           “O” 涉及口头公开、使用、展览或其他方式公开的文件<br/>           “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<br/>           “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<br/>           “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<br/>           “&amp;” 同族专利的文件         </td> </tr> </table>           |                                                                                                                                                                                     |                                              | * 引用文件的具体类型:<br>“A” 认为不特别相关的表示了现有技术一般状态的文件<br>“E” 在国际申请日的当天或之后公布的在先申请或专利<br>“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)<br>“O” 涉及口头公开、使用、展览或其他方式公开的文件<br>“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<br>“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<br>“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<br>“&” 同族专利的文件 |         |   |                                                                               |      |   |                                                           |      |   |                                                                |      |   |                                                              |      |
| * 引用文件的具体类型:<br>“A” 认为不特别相关的表示了现有技术一般状态的文件<br>“E” 在国际申请日的当天或之后公布的在先申请或专利<br>“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)<br>“O” 涉及口头公开、使用、展览或其他方式公开的文件<br>“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<br>“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<br>“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<br>“&” 同族专利的文件 |                                              |                                                                                                                                                                                                            |                                                                                                                                                                                     |         |   |                                                                               |      |   |                                                           |      |   |                                                                |      |   |                                                              |      |
| 国际检索实际完成的日期<br><br>2017年 4月 9日                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                                                                                                                                                     | 国际检索报告邮寄日期<br><br>2017年 5月 2日                |                                                                                                                                                                                                            |                                                                                                                                                                                     |         |   |                                                                               |      |   |                                                           |      |   |                                                                |      |   |                                                              |      |
| ISA/CN的名称和邮寄地址<br><br>中华人民共和国国家知识产权局 (ISA/CN)<br>中国北京市海淀区蓟门桥西土城路6号 100088<br><br>传真号 (86-10) 62019451                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                                                                                                                                                                                     | 受权官员<br><br>杭雪蒙<br><br>电话号码 (86-10) 62089103 |                                                                                                                                                                                                            |                                                                                                                                                                                     |         |   |                                                                               |      |   |                                                           |      |   |                                                                |      |   |                                                              |      |

国际检索报告  
关于同族专利的信息

|                   |
|-------------------|
| 国际申请号             |
| PCT/CN2016/089836 |

| 检索报告引用的专利文件 |           |   | 公布日<br>(年/月/日) | 同族专利 |            |    | 公布日<br>(年/月/日) |
|-------------|-----------|---|----------------|------|------------|----|----------------|
| CN          | 103729250 | A | 2014年 4月 16日   | US   | 2014108639 | A1 | 2014年 4月 17日   |
|             |           |   |                | US   | 9253053    | B2 | 2016年 2月 2日    |
| CN          | 102646121 | A | 2012年 8月 22日   | 无    |            |    |                |
| CN          | 102857558 | A | 2013年 1月 2日    | CN   | 102857558  | B  | 2015年 11月 25日  |
| CN          | 103782556 | A | 2014年 5月 7日    | WO   | 2015018035 | A1 | 2015年 2月 12日   |
|             |           |   |                | EP   | 3024184    | A1 | 2016年 5月 25日   |
|             |           |   |                | CN   | 103782556  | B  | 2016年 12月 7日   |
|             |           |   |                | EP   | 3024184    | A4 | 2016年 7月 27日   |

表 PCT/ISA/210 (同族专利附件) (2009年7月)