

(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
2017年3月30日 (30.03.2017)

(10) 国际公布号
WO 2017/049965 A1

- (51) 国际专利分类号: G06F 17/30 (2006.01) G06F 11/14 (2006.01)
- (21) 国际申请号: PCT/CN2016/085608
- (22) 国际申请日: 2016年6月13日 (13.06.2016)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权: 201510620975.9 2015年9月25日 (25.09.2015) 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: 房浩 (FANG, Hao); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 毕杰山 (BI, Jieshan); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 钟超强 (ZHONG, Chaoqiang); 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, 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, 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))

[见续页]

(54) Title: DATA BACKUP METHOD AND DATA PROCESSING SYSTEM

(54) 发明名称: 数据备份的方法和数据处理系统

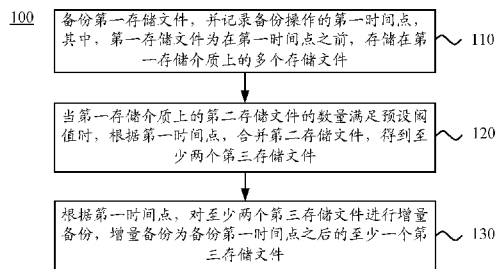


图 2

- 110 Back up a first storage file, and record a first time point of the backup operation, wherein the first storage file is a plurality of storage files stored on a first storage medium before the first time point
- 120 When the number of second storage files on the first storage medium satisfies a preset threshold, on the basis of the first time point, combine the second storage files to obtain at least two third storage files
- 130 On the basis of the first time point perform incremental backup on the at least two third storage files, the incremental backup backing up at least one third storage file after the first time point

(57) Abstract: A data backup method and data processing system, the method (100) comprising: backing up a first storage file, and recording a first time point of the backup operation, wherein the first storage file is a plurality of storage files stored on a first storage medium before the first time point (110); when the number of second storage files on the first storage medium satisfies a preset threshold, on the basis of the first time point, combining the second storage files to obtain at least two third storage files (120); and on the basis of the first time point performing incremental backup on the at least two third storage files, the incremental backup backing up at least one third storage file after the first time point (130). Thus, the data backup method, through recording a data backup time point and determining backup incremental data, avoids an incremental data storage process in which an entire data list is read, reducing incremental data backup time, reducing network and magnetic disk resource consumption by incremental data backup.

(57) 摘要:

[见续页]



WO 2017/049965 A1



-
- 关于申请人有权要求在先申请的优先权(细则 4.17(iii))
- 本国际公布:**
- 包括国际检索报告(条约第 21 条(3))。

一种数据备份的方法 and 数据处理系统, 该方法 (100) 包括: 备份第一存储文件, 并记录备份操作的第一时间点, 其中, 第一存储文件为在第一时间点之前, 存储在所述第一存储介质上的多个存储文件 (110); 当所述第一存储介质上的第二存储文件的数量满足预设阈值时, 根据第一时间点, 合并第二存储文件, 得到至少两个第三存储文件 (120); 根据第一时间点, 对至少两个第三存储文件进行增量备份, 增量备份为备份第一时间点之后的至少一个第三存储文件 (130)。因此, 该数据备份的方法, 通过记录数据备份的时间点, 确定备份增量数据, 避免了增量数据存储读取全表数据的过程, 缩短了增量数据备份的时间, 减少了增量数据备份对网络和磁盘资源的消耗。

数据备份的方法和数据处理系统

技术领域

本发明涉及数据库领域，尤其涉及一种数据备份的方法和数据处理系统。

5

背景技术

分布式数据库在写数据时，数据会存储到用户表的一个或多个子表中，当满足一定的条件时，将数据进行备份，以保证数据的可靠性。

但目前分布式数据库将数据备份消耗的时间长，且备份时对网络和磁盘资源的消耗较大。

10

发明内容

本申请提供一种数据备份的方法和数据处理系统，通过记录数据备份的时间点，确定备份增量数据，缩短了增量数据备份的时间，减少了增量数据备份对网络和磁盘资源的消耗。

15

一方面，本申请提供了一种数据备份的方法，该方法包括：

在第一时间点备份存储在第一存储介质上的第一存储文件，并记录第一时间点；当第一存储介质上的第二存储文件数量达到预设阈值时，以记录的第一时间点分界点，将第二存储文件合并为至少两个第三存储文件，其中，第二存储包括第一存储文件和第一时间点之后到当前时间点新生成的存储文件；对第一时间点之后的至少一个第三存储文件进行增量备份。

20

在一个可能的设计中，对第一时间点之后的至少一个第三存储文件进行增量备份，具体包括：获取合并后的至少两个第三存储文件的时间戳范围；根据每个第三存储文件的时间戳范围，以第一时间点为分界点，确定第一时间点之后的第三存储文件，并进行增量备份。

25

在一个可能的设计中，根据预设规则，备份第一存储文件。

在另一个可能的设计中，获取至少两个第三存储文件的时间戳范围，包括：从每个第三存储文件的元数据块中获取时间戳范围。

在又一个可能的设计中，在所述根据所述第一时间点，对至少两个第三存储文件进行增量备份之后，方法还包括：记录增量备份操作的第二时间点，以
5 用于根据第二时间点合并或备份新生成的存储文件。

一方面，本申请提供了一种数据处理系统，该系统包括：

备份模块，用于在第一时间点备份存储在第一存储介质上的第一存储文件，并记录备份操作的第一时间点；合并模块，用于当第一存储介质上的第二存储文件的数量满足预设阈值时，以记录的第一时间点分界点，将第二存储文件合
10 并为至少两个第三存储文件，其中，第二存储包括第一存储文件和第一时间点之后到当前时间点新生成的存储文件；备份模块还用于，备份第一时间点之后的至少一个第三存储文件。

在一个可能的设计中，数据处理系统还包括：获取模块，用于获取合并后的至少两个第三存储文件的时间戳范围；确定模块，用于每个第三存储文件的
15 时间戳范围，以第一时间点为分界点，确定第一时间点之后的第三存储文件，备份模块对确定模块确定的第一时间点后的第三存储文件进行增量备份。

在一个可能的设计中，备份模块具体用于，根据预设条件，备份第一存储文件。

在另一个可能的设计中，获取模块具体用于，从每个第三存储文件的元数
20 据块中获取时间戳范围。

在又一个可能的设计中，备份模块还用于，记录增量备份操作的第二时间点，以用于根据第二时间点合并或备份新生成的存储文件。

一方面，本申请提供了一种数据处理系统，该数据处理系统包括处理器和存储器，处理器用于根据存储器存储的预设规则备份存储在存储器中的第一存
25 储文件，并记录进行备份操作的第一时间点；当存储器中的第二存储文件数量超过预设阈值时，以记录的第一时间点为分界点，将第二存储文件合并为至少

两个第三存储文件，其中，第二存储包括第一存储文件和第一时间点之后到当前时间点新生成的存储文件；以记录的第一时间点为分界点，备份第一时间点之后的第三存储文件。

5 在一个可能的设计中，处理器获取合并后的至少两个第三存储文件的时间戳范围；并以第一时间点为分界点，确定第一时间点之后的第三存储文件，并进行增量备份。

在另一个可能的设计中，处理器从每个第三存储文件的元数据块中获取时间戳范围。

10 在又一个可能的设计中，处理器记录增量备份操作的第二时间点，以用于根据第二时间点合并或备份新生成的存储文件。

本申请提供的数据备份的方法和数据处理系统，通过记录数据备份的时间点，确定备份增量数据，避免了增量数据存储读取全表数据的过程，缩短了增量数据备份的时间，减少了增量数据备份对网络和磁盘资源的消耗。

15 附图说明

图 1 为一个用户表中子表的示意图；

图 2 为本发明实施例提供的一种分布式数据库数据备份的方法的流程图；

图 3 为本发明实施例提供的另一种分布式数据库数据备份的方法的流程图；

图 4 为本发明实施例提供的一种数据处理系统的示意图；

20 图 5 为本发明实施例提供的另一数据处理系统的示意图；

图 6 为本发明实施例提供的又一数据处理系统的示意图。

具体实施方式

25 本发明实施例适用于以键-值 (Key Value) 类型的存储方式进行存储的数据库，即，将用户需要存储的数据存放在 Value 部分，构建与 Value 对应的 Key，当用户需要查找 Value 时，通过 Key 索引查找对应的 Value，其中，数据的存储

是按照 Key 的字典顺序自然排序的。

当用户需要在分布式数据库中存储数据时，创建相应的用户表 (Table)。一个 Table 中可以存储大量的数据，为了方便管理和维护，可以按照记录的 Key 值得字典顺序切割生成一个或多个子表 (Region)。每个 Region 是一个 Key 区间数据的集合，也可以说是连续行数据的集合。每个 KeyValue 所对应的 Key 只包括在一个 Region 区间，也就是说多个 Region 之间没有交集，如图 1 所示，图 1 为一个用户表中子表的示意图，其中，图 1 中 Region-1 中的 Key-A 代表着在字典顺序上一个无穷小的值，Region-M 中的 Key-Y 代表着在字典顺序上一个无穷大的值。

10 当分布式数据库在写数据时，数据可以同时写入 WAL 和 Region 的内存中。当 Region 内存中的数据满足预设的内存阈值时，将 Region 中的数据固化到磁盘上，同时在磁盘上生成存储文件 Hfile，Hfile 文件的元数据块中存储有该 Region 中数据记录的时间戳范围。

15 为避免影响数据库读数据的性能，当磁盘上存储文件 Hfile 的数量满足预设的数量阈值时，对磁盘上多个存储文件 Hfile 进行合并 (Compaction)，得到至少一个存储文件 Hfile。

为防止系统出现操作失误或者系统故障导致的数据丢失，需要将系统中的全部数据或者部分数据集合从应用主机的硬盘或者阵列复制到本地或者其他的安全存储介质中。

20 下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本发明的一部分实施例，而不是全部实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动的前提下所获得的所有其他实施例，都应属于本发明保护的范围。

25 为便于说明，结合图 2 对本发明实施例的技术方案进行详细的说明。应理解，这仅是用以说明本发明实施例的技术方案的一个例子，并不对本发明实施例构成任何的限定。

图 2 为本发明实施例提供的一种数据备份的方法的流程图，如图 2 所示，该方法 100 的执行主体为数据处理系统，该方法 100 包括：

110，备份第一存储文件，并记录备份操作的第一时间点，其中，第一存储文件为在第一时间点之前，存储在第一存储介质上的多个存储文件；

5 120，当第一存储介质上的第二存储文件的数量满足预设阈值时，根据第一时间点，合并第二存储文件，得到至少两个第三存储文件；

130，根据第一时间点，对至少两个第三存储文件进行增量备份，增量备份为备份第一时间点之后的至少一个第三存储文件。

具体的，为保证数据库中的数据安全，将数据库中的全部数据或者部分数
10 据进行备份。数据处理系统可以包括多个终端，每个终端包括第一存储介质。

在 110 中，第一存储介质可以具体为磁盘。第一存储文件可以为一个或者多个终端的磁盘上的全部或者部分存储文件。在第一时间点，备份磁盘上的全部或者部分存储文件，记录当前备份文件的时间点。

可选地，在本发明实施例中，可以根据预设规则，备份第一存储文件。该
15 预设规则可以为根据用户需求制定的数据备份的条件。例如，用户需要在每月的第一天进行一次全量备份，每天进行一次增量备份，该预设规则可以通过计时器或者其他实现用户需求的逻辑方法实现规则的制定。通过预设规则进行数据备份，既方便了用户的使用，又可以保证数据的可靠性。

应理解，在本发明实施例中，备份第一存储文件也可以为人为的主观行事。

20 在 120 中，第二存储文件为当前时间点第一存储介质上存储的存储文件，第二存储文件中包括备份过的第一存储文件和第一时间点到当前时间点之间新生成的存储文件。第三存储文件为将第二存储文件合并后生成的新的存储文件。当磁盘上的存储文件数量达到预设的数量阈值时，根据第一时间点，将第一时间点之前的存储文件和第一时间点之后的存储文件分别合成为至少一个存储文
25 件。

在 130 中，将第一时间点之后的存储文件进行增量备份，读取所有存储文

件的时间戳范围，查询第一时间点之后的存储文件，根据查询结果增量备份存储文件中数据。

可选地，在本发明实施例中，获取合并后的至少两个第三存储文件的时间戳范围，可以从每个第三存储文件的元数据块中获取时间戳范围。

5 在本发明实施例中，在根据所述第一时间点，对至少两个第三存储文件进行增量备份之后，该方法 100 还包括：记录增量备份操作的第二时间点，以用于根据所述第二时间点合并或备份新生成的存储文件。

具体的，数据处理系统在对数据进行增量备份之后，记录当前增量备份操作的时间点，以便于在该时间点之后的时间段内，磁盘上的存储文件数量达到
10 预设阈值时，数据处理系统根据记录的该时间点合并存储文件，并将对合并后的增量数据进行备份。

采用本发明实施例提供的分布式数据库数据备份的方法，通过记录上一次备份的时间点，确定增量备份的文件，进行增量备份，避免了增量数据备份读取全表数据的过程，缩短了增量数据备份的时间，减少了增量数据备份对网络
15 和磁盘资源的消耗。

优选地，在本发明另一实施例中，如图 3 所示，根据第一时间点，对至少两个第三存储文件进行增量备份，包括：

131，获取至少两个第三存储文件的时间戳范围；

132，根据至少两个第三存储文件的时间戳范围和所述第一时间点，确定增
20 量备份的至少一个第三存储文件；

133，备份至少一个第三存储文件。

具体的，数据处理系统读取合并后的所有存储文件的时间戳范围，根据每个存储文件的时间戳范围，确定属于时间戳范围在第一时间点之后的第三存储文件，备份确定的第一时间点之后的第三存储文件。

25 采用本发明实施例提供的分布式数据库数据备份的方法，通过记录上一次备份的时间点，确定增量备份的文件，进行增量备份，避免了增量数据备份读

取全表数据的过程，缩短了增量数据备份的时间，减少了增量数据备份对网络和磁盘资源的消耗。

为便于说明，以第一存储介质为磁盘，磁盘上的存储文件满足 5 个存储文件时执行合并操作流程为例，对本发明实施例的技术方案进行详细的说明。应理解，这仅是用以说明本发明实施例的技术方案的一个例子，并不对本发明实施例构成任何的限定。

在时间点 T1 之前备份磁盘上存储的 3 个存储文件，分别为 Hfile1、Hfile2 和 Hfile3，终端并记录备份操作的时间点 T1；在时间点 T1 到时间点 T2 的时间内，磁盘上新生成 2 个存储文件 Hfile4 和 Hfile5，此时，磁盘上的存储文件满足预设的数量阈值，根据记录的时间点 T1，将时间点 T1 之前的 3 个存储文件 Hfile1、Hfile2 和 Hfile3 合并为一个存储文件 Hfile6，将时间点 T1 之后的 2 个存储文件合并为一个存储文件 Hfile7；在时间点 T3 时，终端读取合并后的 2 个存储文件 Hfile6 和 Hfile7 的时间戳范围，确定时间点 T1 后的存储文件 Hfile7，备份确定的存储文件 Hfile7，该存储文件中的数据为时间点 T1 后的增量数据，该备份过程称为增量备份。

应理解，在本发明实施例中，时间点 T1、时间点 T2 和时间点 T3 是按照时间顺序排列的三个时间点。

在本发明实施例中，为存储文件 Hfile 以数字 1-7 的数字表示不同的存储文件，只是为了清楚描述本发明实施例的技术方案，在本发明实施例中，存储文件的区分依据每个存储文件的元数据块内容进行区分。

还应理解，在本发明实施例中，依第一存储介质上的存储文件数量满足预设阈值，执行存储文件的合并流程，只是为了详细描述本发明实施例的技术方案，在本发明实施例中，触发第一存储进行存储文件合并的条件还可以采用其他条件，在本发明的实施例中不作任何限制。

上文中结合图 1 至图 3，详细描述了根据本发明实施例的数据备份的方法，下面将结合图 4 至图 6，详细描述根据本发明实施例的数据处理系统。

图 4 为本发明实施例提供的一种数据处理系统的示意性框图。如图 5 所示的数据处理系统 400 包括：备份模块 410 和合并模块 420。

5 备份模块 410，用于备份第一存储文件，并记录备份操作的第一时间点，其中，第一存储文件为在第一时间点之前，存储在第一存储介质上的多个存储文件；

合并模块 420，用于当第一存储介质上的第二存储文件的数量满足预设阈值时，根据第一时间点，合并第二存储文件，得到至少两个第三存储文件；

10 备份模块 410 还用于，根据第一时间点，对至少两个第三存储文件进行增量备份，增量备份为备份第一时间点之后的至少一个第三存储文件。

具体的，第一存储介质可以具体为磁盘。第一存储文件可以为一个或者多个终端的磁盘上的全部或者部分存储文件。在第一时间点，备份模块 410 备份磁盘上的全部或者部分存储文件，记录当前备份文件的时间点。

15 可选地，在本发明实施例中，备份模块 410 可以根据预设规则，备份第一存储文件。该预设规则可以为根据用户需求制定的数据备份的条件。通过预设规则进行数据备份，既方便了用户的使用，又可以保证数据的可靠性。

应理解，在本发明实施例中，备份第一存储文件也可以为人为的主观行事。

20 第二存储文件为当前时间点第一存储介质上存储的存储文件，第二存储文件中包括备份过的第一存储文件和第一时间点到当前时间点之间新生成的存储文件。。第三存储文件为将第二存储文件合并后生成的新的存储文件。当磁盘上的存储文件数量达到预设的数量阈值时，合并模块 420 根据第一时间点，将第一时间点之前的存储文件和第一时间点之后的存储文件分别合成为至少一个存储文件。

25 备份模块 410 将第一时间点之后的存储文件进行增量备份，读取所有存储文件的时间戳范围，查询第一时间点之后的存储文件，根据查询结果增量备份

存储文件中数据。

采用本发明实施例提供的分布式数据库数据备份的方法，通过记录上一次备份的时间点，确定增量备份的文件，进行增量备份，避免了增量数据备份读取全表数据的过程，缩短了增量数据备份的时间，减少了增量数据备份对网络和磁盘资源的消耗。

优选地，作为本发明的另一实施例，如图 5 所示，数据处理系统还包括获取模块 430 和确定模块 440，

获取模块 430，用于获取至少两个第三存储文件的时间戳范围；

确定模块 440，用于根据至少两个所述第三存储文件的时间戳范围和第一时间点，确定增量备份的至少一个第三存储文件；

备份模块 410 用于备份至少一个所述第三存储文件。

具体的，获取模块 430 读取合并后的所有存储文件的时间戳范围，根据每个存储文件的时间戳范围，确定模块 440 确定属于时间戳范围在第一时间点之后的第三存储文件，备份模块 410 备份确定的第一时间点之后的第三存储文件。

可选地，在本发明实施例中，获取模块 430 可以从每个第三存储文件的元数据块中获取时间戳范围。

在本发明实施例中，备份模块 410 还用于，记录增量备份操作的第二时间点，以用于根据所述第二时间点合并或备份新生成的存储文件。

具体的，备份模块 410 对数据进行增量备份之后，记录当前增量备份操作的时间点，以便于在该时间点之后的时间段内，磁盘上的存储文件数量达到预设阈值时，合并模块 420 根据记录的该时间点合并存储文件，备份模块 410 对该时间点之后的合并后的存储文件进行增量备份。

采用本发明实施例提供的分布式数据库数据备份的方法，通过记录上一次备份的时间点，确定增量备份的文件，进行增量备份，避免了增量数据备份读取全表数据的过程，缩短了增量数据备份的时间，减少了增量数据备份对网络和磁盘资源的消耗。

需要说明的是，在本发明实施例中，根据本发明实施例的数据处理系统对应根据本发明实施例的方法 100 的执行主体，并且数据处理系统中的各个模块的上述和其他操作和/或功能分别为了实现图 2 和图 3 中的各个方法的相应流程，为了简洁，在这里不再赘述。

5

图 6 为本发明实施例提供的又一数据处理系统的示意图，如图 6 所示，该数据处理系统 600 包括：处理器 610 和存储器 620 和通信总线系统 630，其中，处理器 610 和存储器 620 通过通信总线系统 630 相连接。其中，

处理器 610，用于备份第一存储文件，并记录备份操作的第一时间点，其中，
10 第一存储文件为在第一时间点之前，存储在第一存储介质上的多个存储文件；

处理器 610，用于当第一存储介质上的第二存储文件的数量满足预设阈值时，根据第一时间点，合并第二存储文件，得到至少两个第三存储文件；

处理器 610 还用于，根据第一时间点，对至少两个第三存储文件进行增量备份，增量备份为备份第一时间点之后的至少一个第三存储文件。

15 具体的，第一存储介质可以具体为磁盘。第一存储文件可以为一个或者多个终端的磁盘上的全部或者部分存储文件。在第一时间点，处理器 610 备份磁盘上的全部或者部分存储文件，记录当前备份文件的时间点。

在本发明实施例中，处理器 610 可以根据预设规则备份第一存储文件，该预设规则可以为根据用户的需求设定的数据备份的条件。在本发明实施例中，
20 处理器 610 也可以通过用户个人的需求，主观执行数据备份时下发的指令进行第一存储文件的备份。

第二存储文件为当前时间点第一存储介质上存储的存储文件，第二存储文件可以包括备份过的第一存储文件和第一时间点到当前时间点新生成的存储文件。第三存储文件为将第二存储文件合并后生成的新的存储文件。当磁盘上的
25 存储文件数量达到预设的数量阈值时，处理器 610 根据第一时间点，将第一时间点之前的存储文件和第一时间点之后的存储文件分别合成为至少一个存储文

件。

处理器 610 将第一时间点之后的存储文件进行增量备份，读取所有存储文件的时间戳范围，查询第一时间点之后的存储文件，根据查询结果增量备份存储文件中数据。

5 采用本发明实施例提供的分布式数据库数据备份的方法，通过记录上一次备份的时间点，确定增量备份的文件，进行增量备份，避免了增量数据备份读取全表数据的过程，缩短了增量数据备份的时间，减少了增量数据备份对网络和磁盘资源的消耗。

10 应理解，在本发明实施例中，该处理器 610 可以是中央处理单元（Central Processing Unit, CPU），该处理器 610 还可以是其他通用处理器、数字信号处理器（DSP）、专用集成电路（ASIC）、现成可编程门阵列（FPGA）或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。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。

15 该总线系统 630 除包括数据总线之外，还可以包括电源总线、控制总线 and 状态信号总线等。但是为了清楚说明起见，在图中将各种总线都标为总线系统 630。

20 在实现过程中，上述方法的各步骤可以通过处理器 610 中的硬件的集成电路或者软件形式的指令完成。结合本发明实施例所公开的方法的步骤可以直接体现为硬件处理器执行完成，或者用处理器中的硬件及软件模块组合执行完成。软件模块可以位于随机存储器，闪存、只读存储器，可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器，处理器 610 读取存储器中的信息，结合其硬件完成上述方法的步骤。为避免重复，这里不再详细描述。

25 存储器，该存储器可以包括只读存储器和随机存取存储器，并向处理器 610 提供指令和数据。存储器的一部分还可以包括非易失性随机存取存储器。例如，存储器还可以存储设备类型的信息。

处理器 610, 用于获取至少两个第三存储文件的时间戳范围;

处理器 610, 用于根据至少两个所述第三存储文件的时间戳范围和第一时间点, 确定增量备份的至少一个第三存储文件;

处理器 610 用于备份至少一个所述第三存储文件。

- 5 具体的, 处理器 610 读取合并后的所有存储文件的时间戳范围, 根据每个存储文件元数据块中的时间戳范围, 处理器 610 确定属于时间戳范围在第一时间点之后的第三存储文件, 处理器 610 备份确定的第一时间点之后的第三存储文件。

需要说明的是, 在本发明实施例中, 根据本发明实施例的数据处理系统
10 700 对应根据本发明实施例的数据处理系统 400, 并且数据处理系统 700 中的各个模块的上述和其他操作和/或功能分别为了实现图 2 和图 3 中的各个方法的相应流程, 为了简洁, 在这里不再赘述。

专业人员应该还可以进一步意识到, 结合本文中公开的实施例描述的各示例的单元及算法步骤, 能够以电子硬件、计算机软件或者二者的结合来实现,
15 为了清楚地说明硬件和软件的可互换性, 在上述说明中已经按照功能一般性地描述了各示例的组成及步骤。这些功能究竟以硬件还是软件方式来执行, 取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能, 但是这种实现不应认为超出本发明的范围。

20 结合本文中公开的实施例描述的方法或算法的步骤可以用硬件、处理器执行的软件模块, 或者二者的结合来实施。软件模块可以置于随机存储器(RAM)、内存、只读存储器(ROM)、电可编程 ROM、电可擦除可编程 ROM、寄存器、硬盘、可移动磁盘、CD-ROM、或技术领域内所公知的任意其它形式的存储介质中。

以上所述的具体实施方式, 对本发明的目的、技术方案和有益效果进行了
25 进一步详细说明, 所应理解的是, 以上所述仅为本发明的具体实施方式而已, 并不用于限定本发明的保护范围, 凡在本发明的精神和原则之内, 所做的任何修改、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

1、一种数据备份的方法，其特征在于，所述方法包括：

备份第一存储文件，并记录备份操作的第一时间点，其中，所述第一存储文件为在所述第一时间点之前，存储在所述第一存储介质上的多个存储文件；

5 当所述第一存储介质上的第二存储文件的数量满足预设阈值时，根据所述第一时间点，合并所述第二存储文件，得到至少两个第三存储文件；

根据所述第一时间点，对所述至少两个第三存储文件进行增量备份，所述增量备份为备份所述第一时间点之后的至少一个所述第三存储文件。

2、根据权利要求1所述的方法，其特征在于，所述根据所述第一时间点，
10 对所述至少两个第三存储文件进行增量备份，包括：

获取至少两个所述第三存储文件的时间戳范围；

根据至少两个所述第三存储文件的时间戳范围和所述第一时间点，确定增量备份的至少一个所述第三存储文件；

备份至少一个所述第三存储文件。

3、根据权利要求1所述的方法，其特征在于，所述备份第一存储文件，包
15 括：

根据预设规则，备份所述第一存储文件。

4、根据权利要求2所述的方法，其特征在于，所述获取至少两个所述第三
20 存储文件的时间戳范围，包括：

从每个所述第三存储文件的元数据块中获取时间戳范围。

5、根据权利要求1至4任一项所述的方法，其特征在于，在所述根据所述第一
25 时间点，对所述至少两个第三存储文件进行增量备份之后，所述方法还包括：记录增量备份操作的第二时间点，以用于根据所述第二时间点合并或备份新生成的存储文件。

6、一种数据处理系统，其特征在于，所述数据处理系统包括：

备份模块，用于备份第一存储文件，并记录备份操作的第一时间点，其中，

所述第一存储文件为在所述第一时间点之前，存储在所述第一存储介质上的多个存储文件；

合并模块，用于当所述第一存储介质上的第二存储文件的数量满足预设阈值时，根据所述第一时间点，合并所述第二存储文件，得到至少两个第三存储文件；

所述备份模块还用于，根据所述第一时间点，对所述至少两个第三存储文件进行增量备份，所述增量备份为备份所述第一时间点之后的至少一个所述第三存储文件。

7、根据权利 6 所述的数据处理系统，其特征在于，所述数据处理系统还包括获取模块和确定模块，

所述获取模块，用于获取至少两个所述第三存储文件的时间戳范围；

所述确定模块，用于根据至少两个所述第三存储文件的时间戳范围和所述第一时间点，确定增量备份的至少一个所述第三存储文件；

所述备份模块用于备份至少一个所述第三存储文件。

8、根据权利要求 6 所述的数据处理系统，其特征在于，所述备份模块具体用于，根据预设条件，备份所述第一存储文件。

9、根据权利要求 7 所述的数据处理系统，其特征在于，所述获取模块具体用于，从每个所述第三存储文件的元数据块中获取时间戳范围。

10、根据权利要求 6 至 9 任一项所述的数据处理系统，其特征在于，所述备份模块还用于，记录增量备份操作的第二时间点，以用于根据所述第二时间点合并或备份新生成的存储文件。

说明书附图

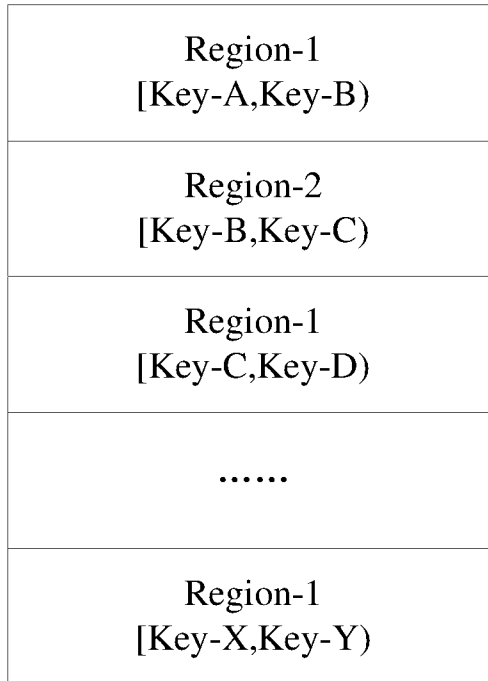


图 1

100

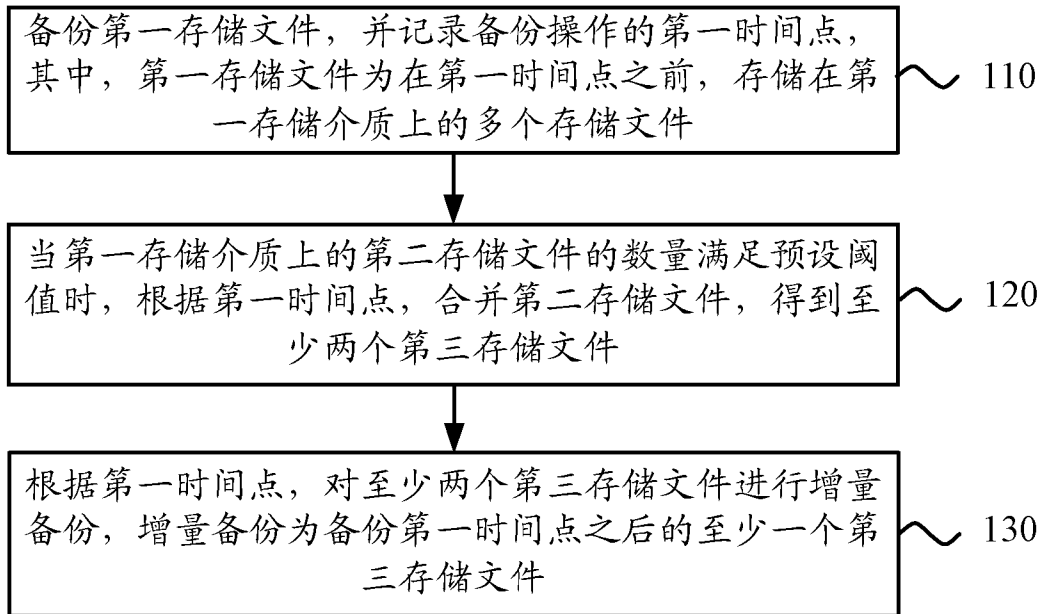


图 2

100

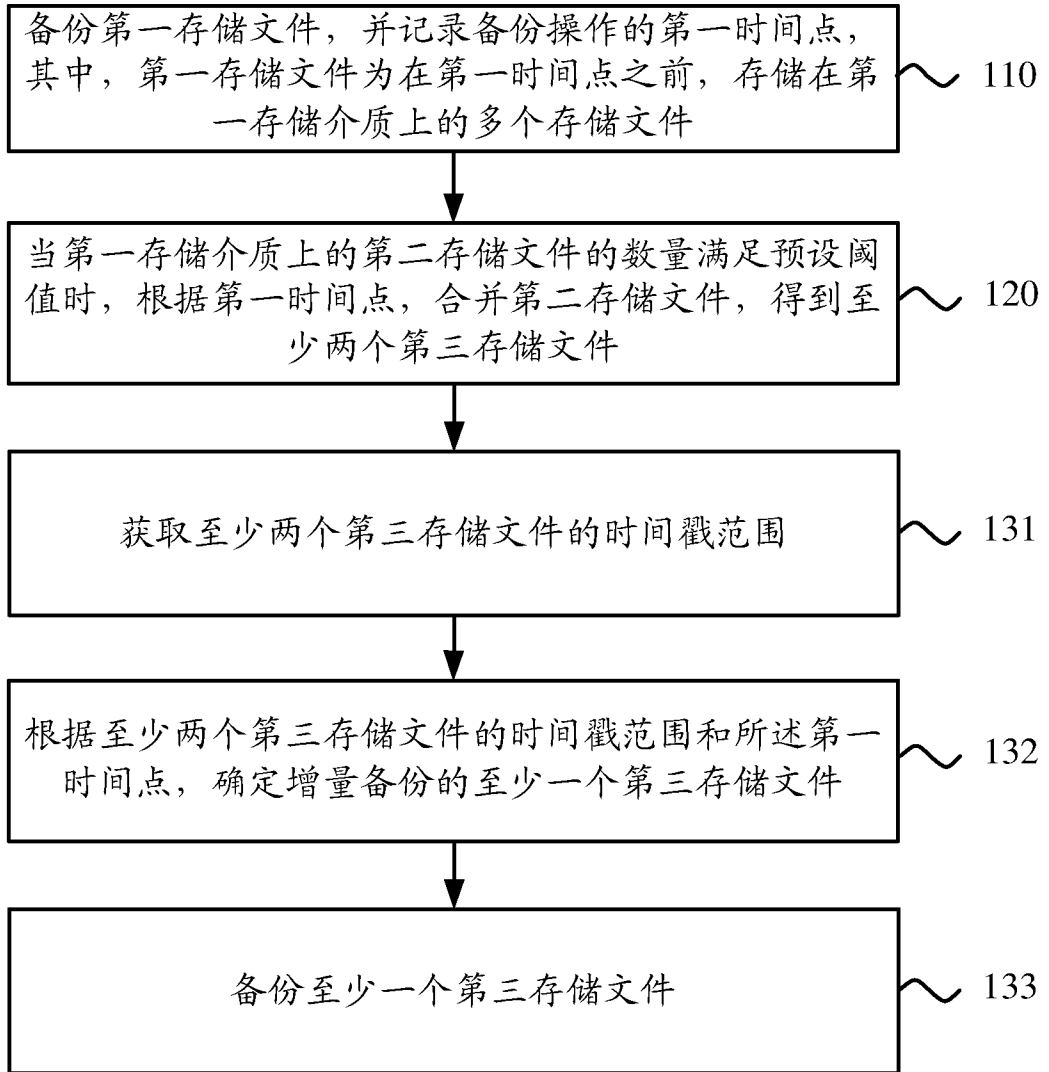


图 3

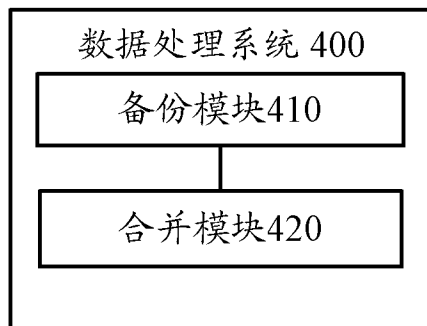


图 4

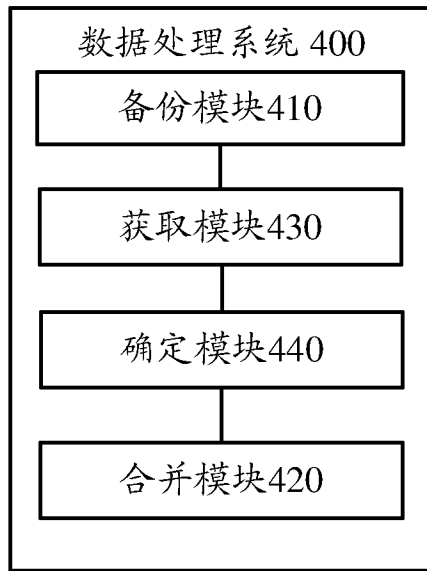


图 5

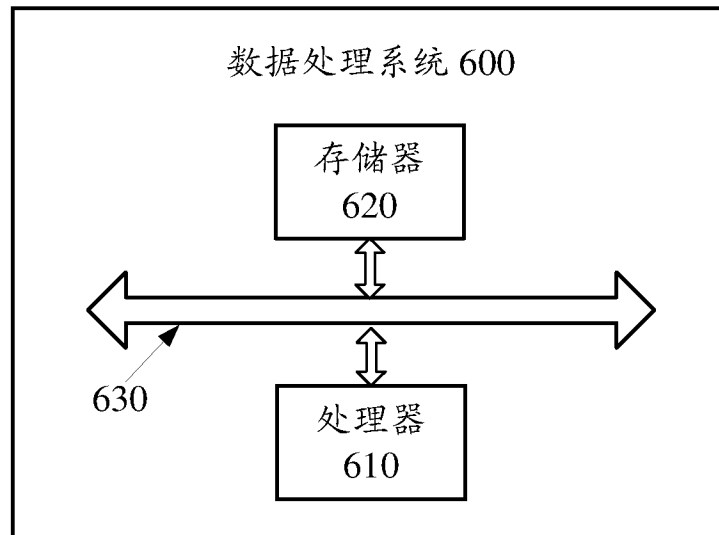


图 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2016/085608

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/30 (2006.01) i; G06F 11/14 (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

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)

WPI; EPODOC; CNPAT; CNKI: difference, data, backup, storage, stamp, threshold value, incremental, time

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 102999400 A (CLOUD COMPUTING BRANCH OF CHINA TELECOM CORP. LTD.), 27 March 2013 (27.03.2013), description, paragraphs [0047], [0051]-[0058] and [0066]-[0072], and figures 1-2	1, 3, 5, 6, 8, 10
Y	CN 102999400 A (CLOUD COMPUTING BRANCH OF CHINA TELECOM CORP. LTD.), 27 March 2013 (27.03.2013), description, paragraphs [0047], [0051]-[0058] and [0066]-[0072], and figures 1-2	2, 4, 5, 7, 9, 10
Y	CN 102541940 A (ASAT CHINA TECHNOLOGY INC.), 04 July 2012 (04.07.2012), description, paragraphs [0038]-[0042], and figure 2	2, 4, 5, 7, 9, 10
PX	CN 105243109 A (HANGZHOU HUAWEI DIGITAL TECHNOLOGY CO., LTD.), 13 January 2016 (13.01.2016), claims 1-10	1-10
A	US 7809691 B1 (SYMANTEC OPERATING CORPORATION), 05 October 2010 (05.10.2010), the whole document	1-10

Further documents are listed in the continuation of Box C.

See patent family annex.

<p>* Special categories of cited documents:</p> <p>“A” document defining the general state of the art which is not considered to be of particular relevance</p> <p>“E” earlier application or patent but published on or after the international filing date</p> <p>“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)</p> <p>“O” document referring to an oral disclosure, use, exhibition or other means</p> <p>“P” document published prior to the international filing date but later than the priority date claimed</p>	<p>“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</p> <p>“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</p> <p>“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</p> <p>“&” document member of the same patent family</p>
---	---

Date of the actual completion of the international search

17 August 2016 (17.08.2016)

Date of mailing of the international search report

26 August 2016 (26.08.2016)

Name and mailing address of the ISA/CN:
 State Intellectual Property Office of the P. R. China
 No. 6, Xitucheng Road, Jimenqiao
 Haidian District, Beijing 100088, China
 Facsimile No.: (86-10) 62019451

Authorized officer

TENG, Mu

Telephone No.: (86-10) **62414007**

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2016/085608

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 102999400 A	27 March 2013	None	
CN 102541940 A	04 July 2012	None	
CN 105243109 A	13 January 2016	None	
US 7809691 B1	05 October 2010	None	

国际检索报告

国际申请号

PCT/CN2016/085608

<p>A. 主题的分类</p> <p>G06F 17/30(2006.01)i; G06F 11/14(2006.01)i</p> <p>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</p>																				
<p>B. 检索领域</p> <p>检索的最低限度文献(标明分类系统和分类号)</p> <p>G06F</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>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))</p> <p>WPI; EPODOC; CNPAT; CNKI; 增量, 备份, 差分, 数据, 存储, 时间, 阈值, 阈值, data, backup, storage, stamp, threshold value, incremental, time</p>																				
<p>C. 相关文件</p> <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 102999400 A (中国电信股份有限公司云计算分公司) 2013年 3月 27日 (2013 - 03 - 27) 说明书[0047]段、[0051]段-[0058] 段、[0066]段-[0072]段、附图1-2</td> <td>1, 3, 5, 6, 8, 10</td> </tr> <tr> <td>Y</td> <td>CN 102999400 A (中国电信股份有限公司云计算分公司) 2013年 3月 27日 (2013 - 03 - 27) 说明书[0047]段、[0051]段-[0058] 段、[0066]段-[0072]段、附图1-2</td> <td>2, 4, 5, 7, 9, 10</td> </tr> <tr> <td>Y</td> <td>CN 102541940 A (上海可鲁系统软件有限公司) 2012年 7月 4日 (2012 - 07 - 04) 说明书[0038]段-[0042]段、附图2</td> <td>2, 4, 5, 7, 9, 10</td> </tr> <tr> <td>PX</td> <td>CN 105243109 A (杭州华为数字技术有限公司) 2016年 1月 13日 (2016 - 01 - 13) 权利要求1-10</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>US 7809691 B1 (SYMANTEC OPERATING CORPORATION) 2010年 10月 5日 (2010 - 10 - 05) 全文</td> <td>1-10</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 102999400 A (中国电信股份有限公司云计算分公司) 2013年 3月 27日 (2013 - 03 - 27) 说明书[0047]段、[0051]段-[0058] 段、[0066]段-[0072]段、附图1-2	1, 3, 5, 6, 8, 10	Y	CN 102999400 A (中国电信股份有限公司云计算分公司) 2013年 3月 27日 (2013 - 03 - 27) 说明书[0047]段、[0051]段-[0058] 段、[0066]段-[0072]段、附图1-2	2, 4, 5, 7, 9, 10	Y	CN 102541940 A (上海可鲁系统软件有限公司) 2012年 7月 4日 (2012 - 07 - 04) 说明书[0038]段-[0042]段、附图2	2, 4, 5, 7, 9, 10	PX	CN 105243109 A (杭州华为数字技术有限公司) 2016年 1月 13日 (2016 - 01 - 13) 权利要求1-10	1-10	A	US 7809691 B1 (SYMANTEC OPERATING CORPORATION) 2010年 10月 5日 (2010 - 10 - 05) 全文	1-10
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求																		
X	CN 102999400 A (中国电信股份有限公司云计算分公司) 2013年 3月 27日 (2013 - 03 - 27) 说明书[0047]段、[0051]段-[0058] 段、[0066]段-[0072]段、附图1-2	1, 3, 5, 6, 8, 10																		
Y	CN 102999400 A (中国电信股份有限公司云计算分公司) 2013年 3月 27日 (2013 - 03 - 27) 说明书[0047]段、[0051]段-[0058] 段、[0066]段-[0072]段、附图1-2	2, 4, 5, 7, 9, 10																		
Y	CN 102541940 A (上海可鲁系统软件有限公司) 2012年 7月 4日 (2012 - 07 - 04) 说明书[0038]段-[0042]段、附图2	2, 4, 5, 7, 9, 10																		
PX	CN 105243109 A (杭州华为数字技术有限公司) 2016年 1月 13日 (2016 - 01 - 13) 权利要求1-10	1-10																		
A	US 7809691 B1 (SYMANTEC OPERATING CORPORATION) 2010年 10月 5日 (2010 - 10 - 05) 全文	1-10																		
<p><input type="checkbox"/> 其余文件在C栏的续页中列出。</p> <p><input checked="" type="checkbox"/> 见同族专利附件。</p>																				
<p>* 引用文件的具体类型:</p> <p>“A” 认为不特别相关的表示了现有技术一般状态的文件</p> <p>“E” 在国际申请日的当天或之后公布的在先申请或专利</p> <p>“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)</p> <p>“O” 涉及口头公开、使用、展览或其他方式公开的文件</p> <p>“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</p> <p>“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</p> <p>“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</p> <p>“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</p> <p>“&” 同族专利的文件</p>																				
<p>国际检索实际完成的日期</p> <p>2016年 8月 17日</p>	<p>国际检索报告邮寄日期</p> <p>2016年 8月 26日</p>																			
<p>ISA/CN的名称和邮寄地址</p> <p>中华人民共和国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088</p> <p>传真号 (86-10)62019451</p>	<p>受权官员</p> <p>滕牧</p> <p>电话号码 (86-10)62414007</p>																			

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2016/085608

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	102999400	A	2013年 3月 27日	无	
CN	102541940	A	2012年 7月 4日	无	
CN	105243109	A	2016年 1月 13日	无	
US	7809691	B1	2010年 10月 5日	无	

表 PCT/ISA/210 (同族专利附件) (2009年7月)