

(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
2020 年 9 月 24 日 (24.09.2020)



(10) 国际公布号
WO 2020/186931 A1

(51) 国际专利分类号:
G06F 3/06 (2006.01)

(21) 国际申请号: PCT/CN2020/074191

(22) 国际申请日: 2020 年 2 月 3 日 (03.02.2020)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910197865.4 2019年3月15日 (15.03.2019) CN

(71) 申请人: 北京京东尚科信息技术有限公司 (BEIJING JINGDONG SHANGKE INFORMATION TECHNOLOGY CO., LTD.) [CN/CN]; 中国北京市海淀区知春路 76 号 8 层, Beijing 100086 (CN)。北京京东世纪贸易有限公司 (BEIJING JINGDONG CENTURY TRADING CO., LTD.) [CN/CN]; 中国北京市北京经济技术开发区科创十一街 18 号 C 座 2 层 201 室, Beijing 100176 (CN)。

(72) 发明人: 吕苏娜 (LYU, Suna); 中国北京市海淀区知春路 76 号 8 层, Beijing 100086 (CN)。

(74) 代理人: 中国国际贸易促进委员会专利商标事务所 (CCPIT PATENT AND TRADEMARK LAW OFFICE); 中国北京市西城区阜成门外大街 2 号万通新世界广场 8 层, Beijing 100037 (CN)。

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

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

(54) **Title:** DATA STORAGE MANAGEMENT METHOD AND APPARATUS, AND COMPUTER-READABLE STORAGE MEDIUM

(54) 发明名称: 数据存储管理方法、装置和计算机可读存储介质

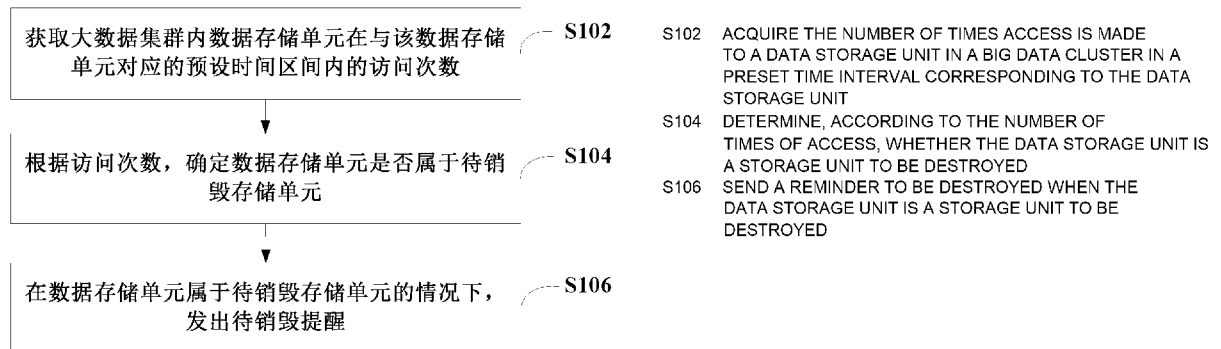


图 1

(57) **Abstract:** Disclosed are a data storage management method and apparatus, and a computer-readable storage medium, relating to the technical field of data storage. The method of the disclosure comprises: acquiring the number of times access is made to a data storage unit in a big data cluster in a corresponding preset time interval; determining, according to the number of times access is made to the data storage unit in the corresponding preset time interval, whether the data storage unit is a storage unit to be destroyed; and sending a reminder to be destroyed when the data storage unit is a storage unit to be destroyed.

(57) **摘要:** 本公开涉及一种数据存储管理方法、装置和计算机可读存储介质, 涉及数据存储技术领域。本公开的方法包括: 获取大数据集群内数据存储单元在对应的预设时间区间内的访问次数; 根据数据存储单元在对应的预设时间区间内的访问次数, 确定数据存储单元是否属于待销毁存储单元; 在数据存储单元属于待销毁存储单元的情况下, 发出待销毁提醒。



WO 2020/186931 A1

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

本国际公布：

— 包括国际检索报告(条约第21条(3))。

数据存储管理方法、装置和计算机可读存储介质

相关申请的交叉引用

本申请是以 CN 申请号为 201910197865.4 申请日为 2019 年 3 月 15 日的申请为基础，并主张其优先权，该 CN 申请的公开内容在此作为整体引入本申请中。

技术领域

本公开涉及数据存储技术领域，特别涉及一种数据存储管理方法、装置和计算机可读存储介质。

背景技术

随着互联网技术的发展，数据呈现爆炸式增长。数据可以存储于大数据集群或者数据仓库中，用户可以进行查询、添加、删除等操作。

目前，随着数据量不断增加，工作人员可能根据实际需求对数据进行筛查，将一些不需要的数据存储单元进行删除。

发明内容

发明人发现：由工作人员进行数据筛查费时费力，可能造成误删除，当开发人员利用某一历史数据进行开发时，该数据有可能因为空间问题会删除，给开发人员开发带来困难。如果不进行存储管理和数据的删除，数据存储单元随着时间的推移越来越胖，数据的查询速度缓慢。

本公开所要解决的一个技术问题是：如何对数据存储进行有效管理，提高数据存储管理的效率。

根据本公开的一些实施例，提供一种数据存储管理方法，包括：获取大数据集群内数据存储单元在与该数据存储单元对应的预设时间区间内的访问次数；数据存储单元包括：数据表或分区文件；根据访问次数，确定数据存储单元是否属于待销毁存储单元；在数据存储单元属于待销毁存储单元的情况下，发出待销毁提醒。

在一些实施例中，获取大数据集群内数据存储单元在与该数据存储单元对应的预设时间区间内的访问次数包括：周期性查询数据存储单元的最后一次访问时间；根据数据存储单元的最后一次访问时间的变化，更新访问次数的记录；根据访问次数的记

录确定访问次数。

在一些实施例中，获取大数据集群内数据存储单元在与该数据存储单元对应的预设时间区间内的访问次数包括：在数据存储单元为数据表的情况下，根据数据表所处的时间阶段，获取数据表在时间阶段对应的预设时间区间内的访问次数；其中，数据表被设置生命周期，生命周期被划分为多个时间阶段。

在一些实施例中，获取大数据集群内数据存储单元在与该数据存储单元对应的预设时间区间内的访问次数包括：在数据存储单元为分区文件的情况下，周期性获取分区文件在当前周期对应的预设时间区间内的访问次数。

在一些实施例中，根据访问次数，确定数据存储单元是否属于待销毁存储单元包括：在数据存储单元为数据表的情况下，根据数据表所处的时间阶段获取时间阶段对应的第一访问次数阈值；在访问次数小于或等于第一访问次数阈值的情况下，确定数据表属于待销毁存储单元；其中，不同时间阶段对应的第一访问次数阈值相同或不同。

在一些实施例中，根据访问次数，确定数据存储单元是否属于待销毁存储单元包括：在数据存储单元为分区文件的情况下，在访问次数小于或等于第二访问次数阈值的情况下，确定分区文件属于待销毁存储单元。

在一些实施例中，该方法还包括：根据数据表在时间阶段对应的预设时间区间内的访问次数，将数据表分配至不同的存储设备进行存储。

在一些实施例中，将数据存储单元分配至不同的存储设备进行存储包括：将数据表在时间阶段对应的预设时间区间内的访问次数与时间阶段对应的多个访问次数阈值进行比对，确定数据表的数据热度等级；根据数据表的数据热度等级，将数据表分配至与数据热度等级对应性能的存储设备进行存储。

在一些实施例中，该方法还包括：在数据存储单元不属于待销毁存储单元，且数据存储单元的生命周期达到结束时间点的情况下，重新配置数据存储单元的生命周期。

根据本公开的另一些实施例，提供一种数据存储管理装置，包括：访问次数获取模块，用于获取大数据集群内数据存储单元在与该数据存储单元对应的预设时间区间内的访问次数；数据存储单元包括：数据表或分区文件；状态确定模块，用于根据访问次数，确定数据存储单元是否属于待销毁存储单元；提醒模块，用于在数据存储单元属于待销毁存储单元的情况下，发出待销毁提醒。

在一些实施例中，访问次数获取模块用于周期性查询数据存储单元的最后一次访问时间；根据数据存储单元的最后一次访问时间的变化，更新访问次数的记录；根据

访问次数的记录确定访问次数。

在一些实施例中，访问次数获取模块用于在数据存储单元为数据表的情况下，根据数据表所处的时间阶段，获取数据表在时间阶段对应的预设时间区间内的访问次数；其中，数据表被设置生命周期，生命周期被划分为多个时间阶段。

5 在一些实施例中，访问次数获取模块用于在数据存储单元为分区文件的情况下，周期性获取分区文件在当前周期对应的预设时间区间内的访问次数。

10 在一些实施例中，状态确定模块用于在数据存储单元为数据表的情况下，根据数据表所处的时间阶段获取时间阶段对应的第一访问次数阈值；在访问次数小于或等于第一访问次数阈值的情况下，确定数据表属于待销毁存储单元；其中，不同时间阶段对应的第一访问次数阈值相同或不同。

在一些实施例中，状态确定模块用于在数据存储单元为分区文件的情况下，在访问次数小于或等于第二访问次数阈值的情况下，确定分区文件属于待销毁存储单元。

在一些实施例中，该装置还包括：存储迁移模块，用于根据数据表在时间阶段对应的预设时间区间内的访问次数，将数据表分配至不同的存储设备进行存储。

15 在一些实施例中，存储迁移模块用于将数据表在时间阶段对应的预设时间区间内的访问次数与时间阶段对应的多个访问次数阈值进行比对，确定数据表的数据热度等级；根据数据表的数据热度等级，将数据表分配至与数据热度等级对应性能的存储设备进行存储。

20 在一些实施例中，该装置还包括：重配置模块，用于在数据存储单元不属于待销毁存储单元，且数据存储单元的生命周期达到结束时间点的情况下，重新配置数据存储单元的生命周期。

根据本公开的又一些实施例，提供一种数据存储管理装置，包括：存储器；以及耦接至存储器的处理器，处理器被配置为基于存储在存储器中的指令，执行如前述任意实施例的数据存储管理方法。

25 根据本公开的再一些实施例，提供一种计算机可读存储介质，其上存储有计算机程序，其中，该程序被处理器执行时实现前述任意实施例的数据存储管理方法。

30 本公开中自动检测大数据集群内数据存储单元在该数据存储单元对应的预设时间区间内的访问次数，根据访问次数判断是否可以被销毁，如果数据存储单元可以被销毁，则发出待销毁提醒。本公开的方法能够自动对大数据集群中的数据存储进行有效管理，及时将不再需要的数据存储单元进行销毁，释放存储空间，提高数据的查询

效率和数据存储管理的效率。

通过以下参照附图对本公开的示例性实施例的详细描述，本公开的其它特征及其优点将会变得清楚。

5 附图说明

此处所说明的附图用来提供对本公开的进一步理解，构成本申请的一部分，本公开的示意性实施例及其说明被配置为解释本公开，并不构成对本公开的不当限定。在附图中：

图 1 示出本公开的一些实施例的数据存储管理方法的流程示意图。

10 图 2 示出本公开的另一些实施例的数据存储管理方法的流程示意图。

图 3 示出本公开的一些实施例的数据存储管理装置的结构示意图。

图 4 示出本公开的另一些实施例的数据存储管理装置的结构示意图。

图 5 示出本公开的又一些实施例的数据存储管理装置的结构示意图。

图 6 示出本公开的再一些实施例的数据存储管理装置的结构示意图。

15

具体实施方式

下面将结合本公开实施例中的附图，对本公开实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本公开一部分实施例，而不是全部的实施例。

以下对至少一个示例性实施例的描述实际上仅仅是说明性的，决不作为对本公开及其
20 应用或使用的任何限制。基于本公开中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本公开保护的范围。

本公开提供一种数据存储管理方法，下面结合图 1 进行描述。

图 1 为本公开数据存储管理方法一些实施例的流程图。如图 1 所示，该实施例的方法包括：步骤 S102~S106。

25 在步骤 S102 中，获取大数据集群内数据存储单元在与该数据存储单元对应的预设时间区间内的访问次数。

在一些实施例中，周期性查询数据存储单元的最后一次访问时间；根据数据存储单元的最后一次访问时间的变化，更新访问次数的记录；根据访问次数的记录确定在对应的预设时间区间内的访问次数。例如，在数据存储于 **Hive** 数据仓中的情况下，
30 可以每隔预设周期轮询一次 **Hive metastore**（元数据存储）的数据库（例如数据库为

Mysql)，若 TBLS 表中的 LAST_ACCESS_TIME（最后一次访问时间）变化，根据 LAST_ACCESS_TIME 的改变记录访问次数。访问次数可以每隔一定时间进行清零，重新计算。可用 Hive 中的命令 desc extended table_name 进行查看，并可根据 dfs.namenode.access.time.precision 参数设置查看的粒度，即设置预设周期的长度。

例如，确定数据存储单元所处时间阶段，获取与该时间阶段对应的预设时间区间内的访问次数。在一些实施例，大数据集群或数据仓库中的数据存储单元为数据表，可以对数据表设置生命周期。生命周期的设置可以在数据表创建时进行设置。数据表的生命周期可以表示该数据表的由创建到销毁的时间段，生命周期的时间长度可以根据实际业务的需求进行设置，不同数据表的生命周期可以不同，例如，生命周期设置为 3 年或 2 个月等。

生命周期可以划分为多个时间阶段，例如，生命周期划分为第一时间阶段、第二时间阶段、第三时间阶段、第四时间阶段等。不同时间阶段对应的预设时间区间的时间长度可以相同或不同，不同时间阶段对应的预设时间区间可以不重叠。不同时间阶段可以根据数据表的数据热度的变化阶段而设置，例如，可以将数据表的数据热度划分为不同的数据热度等级，例如，数据热度等级包括：在线热数据、在线温数据、离线冷数据和待销毁数据等。

可以统计业务对应的预设数量的历史数据表的数据热度情况，根据这些历史数据表处于在线热数据的状态持续的时间长度和时间范围，可以设置第一时间阶段（例如，第一时间阶段为数据表建立开始到第三个月）；根据这些历史数据表处于在线温数据的状态持续的时间长度和时间范围，可以设置第二时间阶段（例如，第二时间阶段为数据表建立第四个月到一年）；以此类推，第三时间阶段对应离线冷数据（例如，第三时间阶段为数据表建立第二年开始到第二年结束），第四时间阶段对应待销毁数据（例如，第三时间阶段为数据表建立第三年开始到生命周期结束）。不是每个数据表都会经历不同的数据热度等级，时间阶段可以是基于相同业务类型的多数数据表的统计结果而设置。

在一些实施例，在数据存储单元为数据表的情况下，根据数据表所处的时间阶段，获取数据表在该时间阶段对应的预设时间区间内的访问次数。例如，数据表当前处于第三时间阶段，则获取数据表建立第二年开始到第二年结束这一时间段内的访问次数。

在一些实施例，数据表可以划分为分区文件进行存储，即数据存储单元可以是

分区文件。在数据存储单元为分区文件的情况下，周期性获取分区文件在当前周期对应的预设时间区间内的访问次数，当前周期即为数据存储单元所处时间阶段。例如，每隔 1 个月获取分区文件在包括当前周期的最近两年内的访问次数。分区文件也可以参考数据表的划分时间阶段的，不同时间阶段对应不同预设时间区间的方式，统计访问次数。

在步骤 S104 中，根据访问次数，确定数据存储单元是否属于待销毁存储单元。

不同数据存储单元统计访问次数的时间区间不同，确定是否属于待销毁存储单元的方式也可以不同。在一些实施例中，在数据存储单元为数据表的情况下，根据数据表所处的时间阶段获取时间阶段对应的第一访问次数阈值；在访问次数小于或等于第一访问次数阈值的情况下，确定数据表属于待销毁存储单元。不同时间阶段对应的第一访问次数阈值相同或不同。例如，第一时间阶段对应的第一访问次数阈值设置为负数，这样数据表在第一时间阶段中则不会被确定为待销毁存储单元，在第三时间阶段或第四时间阶段可以将第一访问次数阈值设置为 0，当数据表在第三时间阶段或第四时间阶段访问次数为 0 的情况下，则会被确定为待销毁存储单元。

在一些实施例中，在数据存储单元为分区文件的情况下，在访问次数小于或等于第二访问次数阈值的情况下，确定分区文件属于待销毁存储单元。例如，分区文件在两年访问次数等于 0，则分区文件属于待销毁存储单元。

在一些实施例中，除了上述对数据存储单元设置生命周期划分时间阶段的方案，数据存储单元还可以被设置不同的统计访问次数的时间阶段，不同的时间阶段对应不同的预设时间区间，例如，第五时间阶段为从当前时间开始的最近三个月，第六时间阶段为最近一年间的时间段除去第一时间阶段，第七时间阶段为最近两年等。不同访问次数统计阶段对应的预设时间区间可以重叠。不同的访问次数统计阶段可以对应不同的第三访问次数阈值，在访问次数小于或等于第三访问次数阈值的情况下，确定数据表属于待销毁存储单元。

在一些实施例中，数据存储单元在与该数据存储单元对应的预设时间区间内的访问次数小于或等于第四访问次数阈值的情况下，确定数据存储单元属于待销毁存储单元。无论数据存储单元属于哪个阶段，只要数据存储单元在与该数据存储单元对应预设时间区间内（例如最近两年）的访问次数小于或等于第四访问次数阈值，则确定数据存储单元属于待销毁存储单元。

在步骤 S106 中，在数据存储单元属于待销毁存储单元的情况下，发出待销毁提

醒。

发出待销毁提醒以便工作人员获知有待销毁存储单元，并且工作人员可以根据业务需求再次确认是否将待销毁存储单元进行销毁。可以在操作界面中显示待销毁存储单元，或者以邮件、短信等形式向工作人员发出待销毁提醒。

5 在一些实施例中，在数据存储单元不属于待销毁存储单元，且数据存储单元的生命周期达到结束时间点的情况下，重新配置数据存储单元的生命周期。如果工作人员确定不对待销毁存储单元进行销毁，则重新配置数据存储单元的生命周期。重新配置的生命周期可以与原来的生命周期不同。例如，对于上述数据存储单元每次前一生命周期结束后可以按照一定步长缩短生命周期的长度作为下一次的周期的长度。

10 上述实施例的方法中自动检测大数据集群内数据存储单元在与该数据存储单元对应的预设时间区间内的访问次数，根据访问次数判断是否可以被销毁，如果数据存储单元可以被销毁，则发出待销毁提醒。上述实施例的方法能够自动对大数据集群中的数据存储进行有效管理，及时将不再需要的数据存储单元进行销毁，释放存储空间，提高数据的查询效率和数据存储管理的效率。

15 为了进一步提高数据查询的效率，可以针对不同数据热度的数据存储单元进行分别存储，下面结合图 2 进行描述。

图 2 为本公开数据存储管理方法另一些实施例的流程图。如图 2 所示，该实施例的方法包括：步骤 S202~S204。

20 在步骤 S202 中，根据数据表所处的时间阶段，获取数据表在该时间阶段对应的预设时间区间内的访问次数。

在步骤 S204 中，根据数据表在该时间阶段对应的预设时间区间内的访问次数，将数据表分配至不同的存储设备进行存储。

25 例如，参考前述实施例，数据表被设置生命周期并划分不同的时间阶段。在一些实施例中，将数据表在时间阶段对应的预设时间区间内的访问次数与时间阶段对应的多个访问次数阈值进行比对，确定数据表的数据热度等级；根据数据表的数据热度等级，将数据表分配至与数据热度等级对应性能的存储设备进行存储。

30 一个时间阶段对应多个访问次数阈值，不同的访问次数阈值对应不同的数据热度等级。例如，第一时间阶段对应的阈值包括 100,50,30 等，如果数据表在第一时间阶段的访问次数超过 100，确定数据表的数据热度等级为最高等级，为在线热数据。如果数据表在第一时间阶段的访问次数小于 100 大于 50，确定数据表的数据热度等级为

第二热度等级，为在线温数据，以此类推。不同时间阶段的访问次数阈值设置可以不同。

数据表的数据热度等级越高，被分配的存储设备（例如，机架）的性能越好（例如，处理效率高，存储空间大）。将不同的数据表根据访问情况分配至不同的存储设备，可以使访问次数多的数据表的查询和访问效率提高，提升用户体验。

例如，参考前述实施例，数据存储单元还可以被设置不同的统计访问次数的时间阶段。分别统计不同时间阶段的访问次数，将数据存储单元在各个时间阶段对应的预设时间区间内的访问次数与该时间阶段对应的多个访问次数阈值进行比对，确定数据存储单元的数据热度等级；根据数据存储单元的数据热度等级，将数据表分配至与数据热度等级对应性能的存储设备进行存储。

上述实施例的方法中根据数据存储单元的数据热度等级将数据存储单元分配至不同性能的存储设备进行存储，随着数据生命周期数据在不同的存储设备间流转，使得访问次数高的数据存储单元能够被高性能的存储设备进行处理，提高数据访问和查询的效率，提升用户体验。

本公开还提供一种数据存储管理装置，下面结合图 3 进行描述。

图 3 为本公开数据存储管理装置的一些实施例的结构图。如图 3 所示，该实施例的装置 30 包括：访问次数获取模块 302，状态确定模块 304，提醒模块 306。

访问次数获取模块 302，用于获取大数据集群内数据存储单元在与该数据存储单元对应的预设时间区间内的访问次数；数据存储单元包括：数据表或分区文件。

在一些实施例中，访问次数获取模块 302 用周期性查询数据存储单元的最后一次访问时间；根据数据存储单元的最后一次访问时间的变化，更新访问次数的记录；根据访问次数的记录确定访问次数。

在一些实施例中，访问次数获取模块 302 用于在数据存储单元为数据表的情况下，根据数据表所处的时间阶段，获取数据表在时间阶段对应的预设时间区间内的访问次数；其中，数据表被设置生命周期，生命周期被划分为多个时间阶段。

在一些实施例中，访问次数获取模块 302 用于在数据存储单元为分区文件的情况下，周期性获取分区文件在当前周期对应的预设时间区间内的访问次数。

状态确定模块 304，用于根据访问次数，确定数据存储单元是否属于待销毁存储单元。

在一些实施例中，状态确定模块 304 用于在数据存储单元为数据表的情况下，根

据数据表所处的时间阶段获取时间阶段对应的第一访问次数阈值；在访问次数小于或等于第一访问次数阈值的情况下，确定数据表属于待销毁存储单元；其中，不同时间阶段对应的第一访问次数阈值相同或不同。

5 在一些实施例中，状态确定模块 304 用于在数据存储单元为分区文件的情况下，在访问次数小于或等于第二访问次数阈值的情况下，确定分区文件属于待销毁存储单元。

提醒模块 306，用于在数据存储单元属于待销毁存储单元的情况下，发出待销毁提醒。

下面结合图 4 描述本公开数据存储管理装置的另一些实施例。

10 图 4 为本公开数据存储管理装置的一些实施例的结构图。如图 4 所示，该实施例的装置 40 包括：访问次数获取模块 402，状态确定模块 404，提醒模块 406 分别与访问次数获取模块 302，状态确定模块 304，提醒模块 306 的功能相同或相似；装置 40 还包括：存储迁移模块 408。

15 存储迁移模块 408 用于根据数据表在时间阶段对应的预设时间区间内的访问次数，将数据表分配至不同的存储设备进行存储。

在一些实施例中，存储迁移模块 408 用于将数据表在时间阶段对应的预设时间区间内的访问次数与时间阶段对应的多个访问次数阈值进行比对，确定数据表的数据热度等级；根据数据表的数据热度等级，将数据表分配至与数据热度等级对应性能的存储设备进行存储。

20 在一些实施例中，装置 40 还包括：重配置模块 410，用于在数据存储单元不属于待销毁存储单元，且数据存储单元的生命周期达到结束时间点的情况下，重新配置数据存储单元的生命周期。

本公开的实施例中的数据存储管理装置可各由各种计算设备或计算机系统来实现，下面结合图 5 以及图 6 进行描述。

25 图 5 为本公开数据存储管理装置的一些实施例的结构图。如图 5 所示，该实施例的装置 50 包括：存储器 510 以及耦接至该存储器 510 的处理器 520，处理器 520 被配置为基于存储在存储器 510 中的指令，执行本公开中任意一些实施例中的数据存储管理方法。

30 其中，存储器 510 例如可以包括系统存储器、固定非易失性存储介质等。系统存储器例如存储有操作系统、应用程序、引导装载程序（Boot Loader）、数据库以及其

他程序等。

图 6 为本公开数据存储管理装置的另一一些实施例的结构图。如图 6 所示，该实施例的装置 60 包括：存储器 610 以及处理器 620，分别与存储器 510 以及处理器 520 类似。还可以包括输入输出接口 630、网络接口 640、存储接口 650 等。这些接口 630，
5 640，650 以及存储器 610 和处理器 620 之间例如可以通过总线 660 连接。其中，输入输出接口 630 为显示器、鼠标、键盘、触摸屏等输入输出设备提供连接接口。网络接口 640 为各种联网设备提供连接接口，例如可以连接到数据库服务器或者云端存储服务器等。存储接口 650 为 SD 卡、U 盘等外置存储设备提供连接接口。

本领域内的技术人员应当明白，本公开的实施例可提供为方法、系统、或计算机
10 程序产品。因此，本公开可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且，本公开可采用在一个或多个其中包含有计算机可用程序代码的计算机可用非瞬时性存储介质（包括但不限于磁盘存储器、CD-ROM、光学存储器等）上实施的计算机程序产品的形式。

本公开是参照根据本公开实施例的方法、设备（系统）、和计算机程序产品的流程图和 / 或方框图来描述的。应理解为可由计算机程序指令实现流程图和 / 或方框图
15 中的每一流程和 / 或方框、以及流程图和 / 或方框图中的流程和 / 或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器，使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生被配置为实现在流程图一个流程或多个流程和 / 或方框图一个
20 个方框或多个方框中指定的功能的装置。

这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中，使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品，该指令装置实现在流程图一个流程或多个流程和 / 或方框图一个
方框或多个方框中指定的功能。

25 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上，使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理，从而在计算机或其他可编程设备上执行的指令提供被配置为实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能的步骤。

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

权 利 要 求

1. 一种数据存储管理方法，包括：

获取大数据集群内数据存储单元在与所述数据存储单元对应的预设时间区间内的访问次数；所述数据存储单元包括：数据表或分区文件；

根据所述访问次数，确定所述数据存储单元是否属于待销毁存储单元；

在所述数据存储单元属于待销毁存储单元的情况下，发出待销毁提醒。

2. 根据权利要求 1 所述的数据存储管理方法，其中，

所述获取大数据集群内数据存储单元在与所述数据存储单元对应的预设时间区间内的访问次数包括：

周期性查询所述数据存储单元的最后一次访问时间；

根据所述数据存储单元的最后一次访问时间的变化，更新访问次数的记录；

根据访问次数的记录确定所述访问次数。

3. 根据权利要求 1 所述的数据存储管理方法，其中，

所述获取大数据集群内数据存储单元在与所述数据存储单元对应的预设时间区间内的访问次数包括：

在所述数据存储单元为数据表的情况下，根据所述数据表所处的时间阶段，获取所述数据表在所述时间阶段对应的预设时间区间内的访问次数；

其中，所述数据表被设置生命周期，所述生命周期被划分为多个时间阶段。

4. 根据权利要求 1 所述的数据存储管理方法，其中，

所述获取大数据集群内数据存储单元在与所述数据存储单元对应的预设时间区间内的访问次数包括：

在所述数据存储单元为分区文件的情况下，周期性获取所述分区文件在当前周期对应的预设时间区间内的访问次数。

5. 根据权利要求 3 所述的数据存储管理方法，其中，

所述根据所述访问次数，确定所述数据存储单元是否属于待销毁存储单元包括：

在所述数据存储单元为数据表的情况下，根据所述数据表所处的时间阶段获取所述时间阶段对应的第一访问次数阈值；在所述访问次数小于或等于所述第一访问次数阈值的情况下，确定所述数据表属于待销毁存储单元；其中，不同时间阶段对应的第一访问次数阈值相同或不同。

6. 根据权利要求 4 所述的数据存储管理方法，其中，

所述根据访问次数，确定所述数据存储单元是否属于待销毁存储单元包括：

在所述数据存储单元为分区文件的情况下，在所述访问次数小于或等于所述第二访问次数阈值的情况下，确定所述分区文件属于待销毁存储单元。

7. 根据权利要求 3 所述的数据存储管理方法，还包括：

根据所述数据表在所述时间阶段对应的预设时间区间内的访问次数，将所述数据表分配至不同的存储设备进行存储。

8. 根据权利要求 7 所述的数据存储管理方法，其中，

所述将所述数据存储单元分配至不同的存储设备进行存储包括：

将所述数据表在所述时间阶段对应的预设时间区间内的访问次数与所述时间阶段对应的多个访问次数阈值进行比对，确定所述数据表的数据热度等级；

根据所述数据表的数据热度等级，将所述数据表分配至与所述数据热度等级对应性能的存储设备进行存储。

9. 根据权利要求 1-8 任一项所述的数据存储管理方法，还包括：

在所述数据存储单元不属于待销毁存储单元，且所述数据存储单元的生命周期达到结束时间点的情况下，重新配置所述数据存储单元的生命周期。

10. 一种数据存储管理装置，包括：

访问次数获取模块，用于获取大数据集群内数据存储单元在与所述数据存储单元对应的预设时间区间内的访问次数；所述数据存储单元包括：数据表或分区文件；

状态确定模块，用于根据所述访问次数，确定所述数据存储单元是否属于待销毁存储单元；

提醒模块，用于在所述数据存储单元属于待销毁存储单元的情况下，发出待销毁提醒。

11. 根据权利要求 10 所述的数据存储管理装置，其中，

所述访问次数获取模块用于周期性查询所述数据存储单元的最后一次访问时间；根据所述数据存储单元的最后一次访问时间的变化，更新访问次数的记录；根据访问次数的记录确定所述访问次数。

12. 根据权利要求 10 所述的数据存储管理装置，其中，

所述访问次数获取模块用于在所述数据存储单元为数据表的情况下，根据所述数据表所处的时间阶段，获取所述数据表在所述时间阶段对应的预设时间区间内的访问次数；

其中，所述数据表被设置生命周期，所述生命周期被划分为多个时间阶段。

13. 根据权利要求 10 所述的数据存储管理装置，其中，

所述访问次数获取模块用于在所述数据存储单元为分区文件的情况下，周期性获取所述分区文件在当前周期对应的预设时间区间内的访问次数。

14. 根据权利要求 12 所述的数据存储管理装置，其中，

所述状态确定模块用于在所述数据存储单元为数据表的情况下，根据所述数据表所处的时间阶段获取所述时间阶段对应的第一访问次数阈值；在所述访问次数小于或等于所述第一访问次数阈值的情况下，确定所述数据表属于待销毁存储单元；其中，不同时间阶段对应的第一访问次数阈值相同或不同。

15. 根据权利要求 13 所述的数据存储管理装置，其中，

所述状态确定模块用于在所述数据存储单元为分区文件的情况下，在所述访问次数小于或等于所述第二访问次数阈值的情况下，确定所述分区文件属于待销毁存储单元。

16. 根据权利要求 12 所述的数据存储管理装置，还包括：

存储迁移模块，用于根据所述数据表在所述时间阶段对应的预设时间区间内的访问次数，将所述数据表分配至不同的存储设备进行存储。

17. 根据权利要求 16 所述的数据存储管理装置，其中，

所述存储迁移模块用于将所述数据表在所述时间阶段对应的预设时间区间内的访问次数与所述时间阶段对应的多个访问次数阈值进行比对，确定所述数据表的数据热度等级；根据所述数据表的数据热度等级，将所述数据表分配至与所述数据热度等级对应性能的存储设备进行存储。

18. 根据权利要求 10-17 任一项所述的数据存储管理装置，还包括：

重配置模块，用于在所述数据存储单元不属于待销毁存储单元，且所述数据存储单元的生命周期达到结束时间点的情况下，重新配置所述数据存储单元的生命周期。

19. 一种数据存储管理装置，包括：

存储器；以及

耦接至所述存储器的处理器，所述处理器被配置为基于存储在所述存储器中的指令，执行如权利要求 1-9 任一项所述的数据存储管理方法。

20. 一种计算机可读存储介质，其上存储有计算机程序，其中，该程序被处理器执行时实现权利要求 1-9 任一项所述方法的步骤。

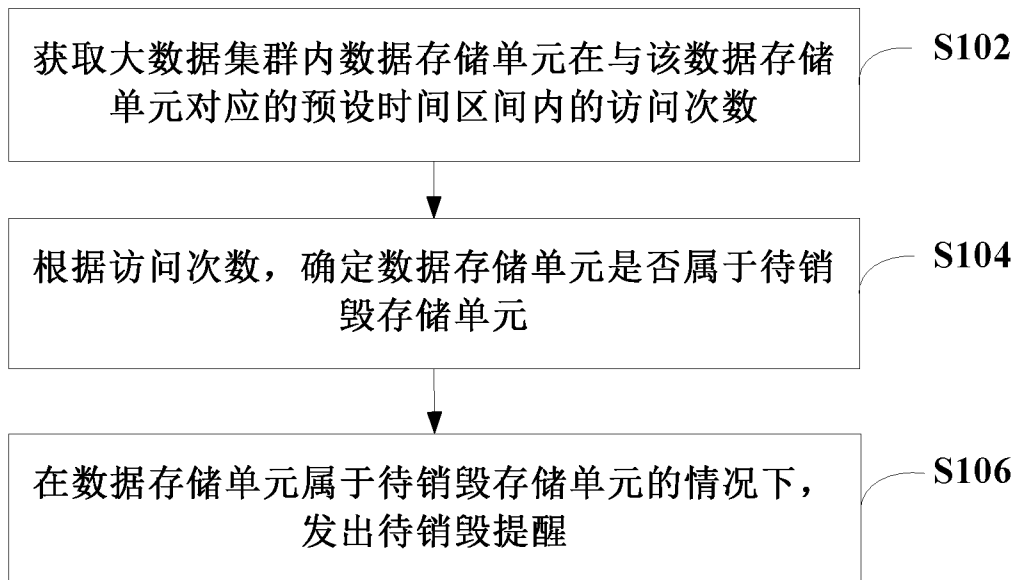


图 1

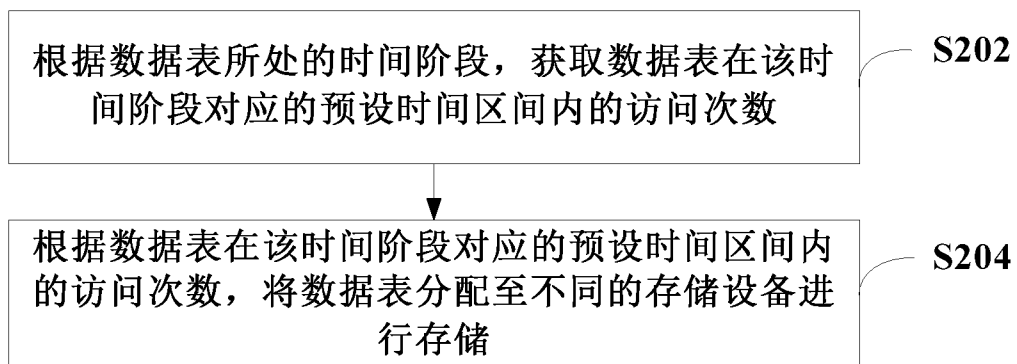


图 2

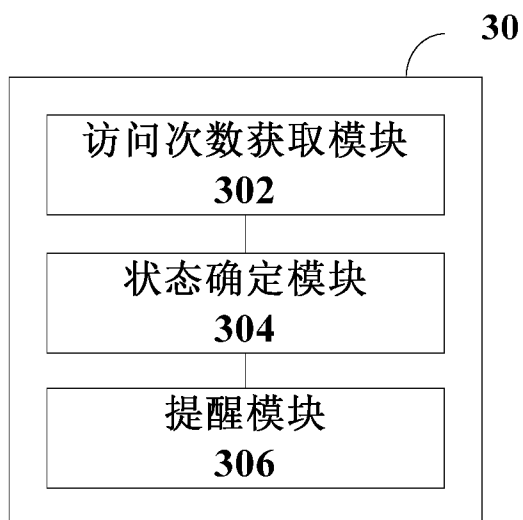


图 3

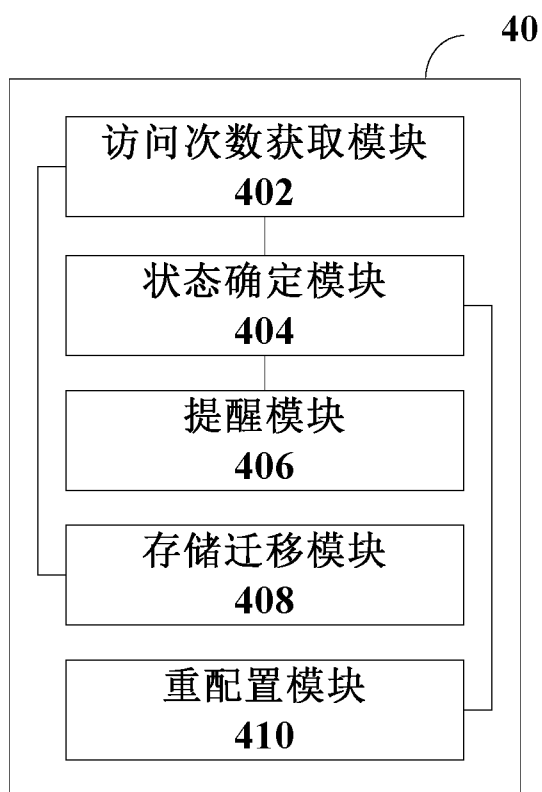


图 4

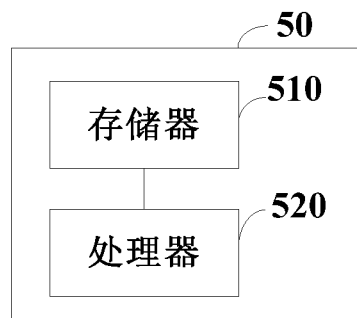


图 5

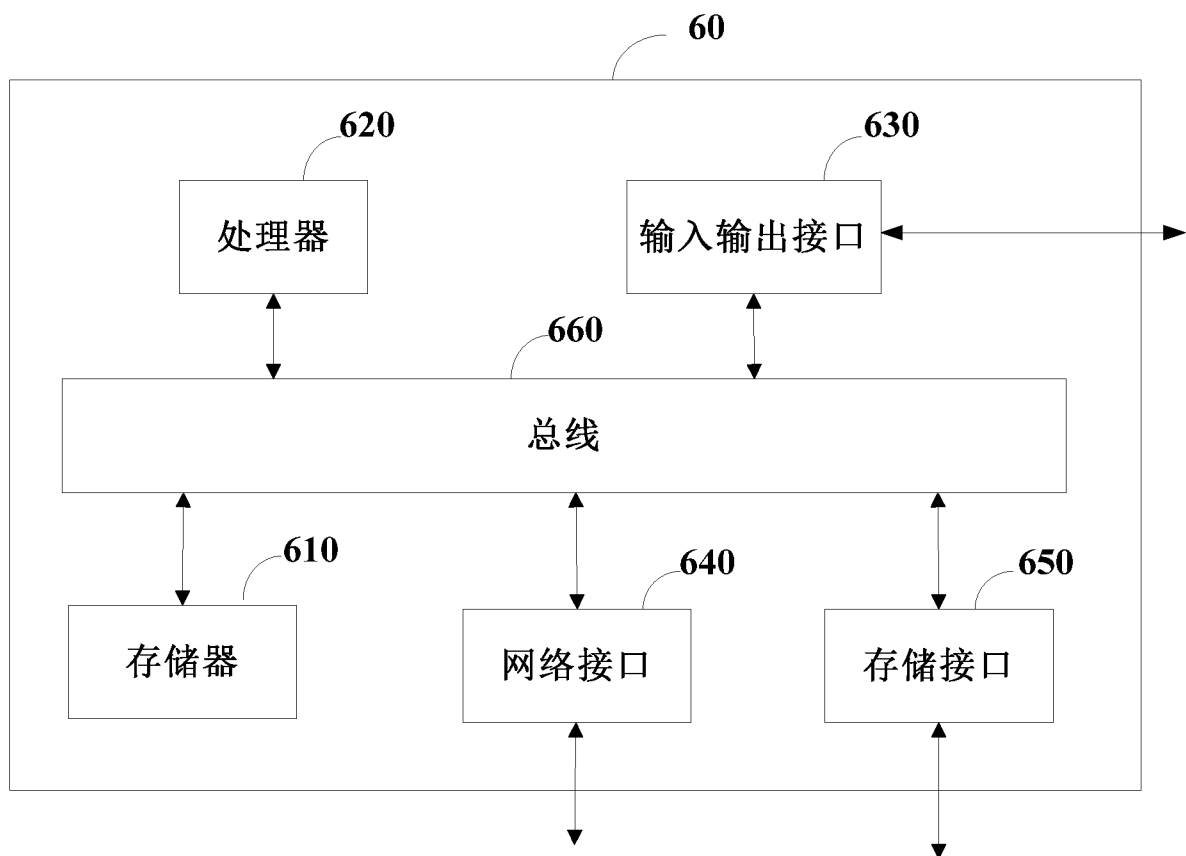


图 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/074191

A. CLASSIFICATION OF SUBJECT MATTER

G06F 3/06(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)

G06F3/-

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

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

CNABS; CNTXT; VEN; USTXT; WOTXT; TWABS; EPTXT; CNKI: 集群, 数据库, 文件, 数据表, 访问, 读, 次数, 频率, 删除, 销毁, 提醒, 生命周期, 阶段, 时期, 区间, cluster, data base, file, table, access, read, times, frequency, delete, destroy, alarm, life cycle, stage, period, range

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 103607312 A (GUANGZHOU CHINADUO NETWORK TECHNOLOGY CO., LTD.) 26 February 2014 (2014-02-26) description, paragraphs 86-131	1-2, 9-11, 18-20
Y	CN 103607312 A (GUANGZHOU CHINADUO NETWORK TECHNOLOGY CO., LTD.) 26 February 2014 (2014-02-26) description, paragraphs 86-131	3-8, 12-17
Y	CN 107168654 A (HUAZHONG UNIVERSITY OF SCIENCE AND TECHNOLOGY) 15 September 2017 (2017-09-15) description, paragraphs 45-53	3-8, 12-17
X	CN 104715020 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 17 June 2015 (2015-06-17) description, paragraphs 34-80	1-2, 9-11, 18-20
A	CN 101777028 A (BEIJING PKUNITY MICROSYSTEMS TECHNOLOGY CO., LTD.) 14 July 2010 (2010-07-14) entire document	1-20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

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

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

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

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

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

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

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

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

“&” document member of the same patent family

Date of the actual completion of the international search

26 April 2020

Date of mailing of the international search report

08 May 2020

Name and mailing address of the ISA/CN

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

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/074191**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 107346321 A (ALIBABA GROUP HOLDING LIMITED) 14 November 2017 (2017-11-14) entire document	1-20
A	US 2010161780 A1 (ELECTRONICS AND TELECOMMUNICATIONS RESEARCH INSTITUTE) 24 June 2010 (2010-06-24) entire document	1-20

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2020/074191

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	103607312	A	26 February 2014	CN	103607312	B	15 February 2017
CN	107168654	A	15 September 2017	CN	107168654	B	13 August 2019
CN	104715020	A	17 June 2015	CN	104715020	B	29 December 2017
CN	101777028	A	14 July 2010	None			
CN	107346321	A	14 November 2017	None			
US	2010161780	A1	24 June 2010	KR	20100072770	A	01 July 2010
				US	8126997	B2	28 February 2012
				KR	101189766	B1	10 October 2012

国际检索报告

国际申请号

PCT/CN2020/074191

A. 主题的分类 G06F 3/06 (2006.01) i 按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类																										
B. 检索领域 检索的最低限度文献 (标明分类系统和分类号) G06F3/- 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用)) CNABS; CNTXT; VEN; USTXT; WOTXT; TWABS; EPTXT; CNKI: 集群, 数据库, 文件, 数据表, 访问, 读, 次数, 频率, 删除, 销毁, 提醒, 生命周期, 阶段, 时期, 区间, cluster, data base, file, table, access, read, times, frequency, delete, destroy, alarm, life cycle, stage, period, range																										
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>CN 103607312 A (广州华多网络科技有限公司) 2014年 2月 26日 (2014 - 02 - 26) 说明书第86-131段</td> <td>1-2, 9-11, 18-20</td> </tr> <tr> <td>Y</td> <td>CN 103607312 A (广州华多网络科技有限公司) 2014年 2月 26日 (2014 - 02 - 26) 说明书第86-131段</td> <td>3-8, 12-17</td> </tr> <tr> <td>Y</td> <td>CN 107168654 A (华中科技大学) 2017年 9月 15日 (2017 - 09 - 15) 说明书第45-53段</td> <td>3-8, 12-17</td> </tr> <tr> <td>X</td> <td>CN 104715020 A (腾讯科技深圳有限公司) 2015年 6月 17日 (2015 - 06 - 17) 说明书第34-80段</td> <td>1-2, 9-11, 18-20</td> </tr> <tr> <td>A</td> <td>CN 101777028 A (北京北大众志微系统科技有限责任公司) 2010年 7月 14日 (2010 - 07 - 14) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 107346321 A (阿里巴巴集团控股有限公司) 2017年 11月 14日 (2017 - 11 - 14) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>US 2010161780 A1 (ELECTRONICS AND TELECOMMUNICATIONS RESEARCH INSTITUTE) 2010年 6月 24日 (2010 - 06 - 24) 全文</td> <td>1-20</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 103607312 A (广州华多网络科技有限公司) 2014年 2月 26日 (2014 - 02 - 26) 说明书第86-131段	1-2, 9-11, 18-20	Y	CN 103607312 A (广州华多网络科技有限公司) 2014年 2月 26日 (2014 - 02 - 26) 说明书第86-131段	3-8, 12-17	Y	CN 107168654 A (华中科技大学) 2017年 9月 15日 (2017 - 09 - 15) 说明书第45-53段	3-8, 12-17	X	CN 104715020 A (腾讯科技深圳有限公司) 2015年 6月 17日 (2015 - 06 - 17) 说明书第34-80段	1-2, 9-11, 18-20	A	CN 101777028 A (北京北大众志微系统科技有限责任公司) 2010年 7月 14日 (2010 - 07 - 14) 全文	1-20	A	CN 107346321 A (阿里巴巴集团控股有限公司) 2017年 11月 14日 (2017 - 11 - 14) 全文	1-20	A	US 2010161780 A1 (ELECTRONICS AND TELECOMMUNICATIONS RESEARCH INSTITUTE) 2010年 6月 24日 (2010 - 06 - 24) 全文	1-20
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																								
X	CN 103607312 A (广州华多网络科技有限公司) 2014年 2月 26日 (2014 - 02 - 26) 说明书第86-131段	1-2, 9-11, 18-20																								
Y	CN 103607312 A (广州华多网络科技有限公司) 2014年 2月 26日 (2014 - 02 - 26) 说明书第86-131段	3-8, 12-17																								
Y	CN 107168654 A (华中科技大学) 2017年 9月 15日 (2017 - 09 - 15) 说明书第45-53段	3-8, 12-17																								
X	CN 104715020 A (腾讯科技深圳有限公司) 2015年 6月 17日 (2015 - 06 - 17) 说明书第34-80段	1-2, 9-11, 18-20																								
A	CN 101777028 A (北京北大众志微系统科技有限责任公司) 2010年 7月 14日 (2010 - 07 - 14) 全文	1-20																								
A	CN 107346321 A (阿里巴巴集团控股有限公司) 2017年 11月 14日 (2017 - 11 - 14) 全文	1-20																								
A	US 2010161780 A1 (ELECTRONICS AND TELECOMMUNICATIONS RESEARCH INSTITUTE) 2010年 6月 24日 (2010 - 06 - 24) 全文	1-20																								
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																										
<table border="0"> <tr> <td> * 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 </td> <td> “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件 </td> </tr> </table>			* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																						
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																									
国际检索实际完成的日期	国际检索报告邮寄日期																									
2020年 4月 26日	2020年 5月 8日																									
ISA/CN的名称和邮寄地址	受权官员																									
中国国家知识产权局 (ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088	李艳军																									
传真号 (86-10)62019451	电话号码 86-(20)-28958353																									

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2020/074191

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	103607312	A	2014年 2月 26日	CN 103607312 B	2017年 2月 15日
CN	107168654	A	2017年 9月 15日	CN 107168654 B	2019年 8月 13日
CN	104715020	A	2015年 6月 17日	CN 104715020 B	2017年 12月 29日
CN	101777028	A	2010年 7月 14日	无	
CN	107346321	A	2017年 11月 14日	无	
US	2010161780	A1	2010年 6月 24日	KR 20100072770 A	2010年 7月 1日
				US 8126997 B2	2012年 2月 28日
				KR 101189766 B1	2012年 10月 10日