

(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
2019 年 9 月 6 日 (06.09.2019)



(10) 国际公布号
WO 2019/165671 A1

- (51) 国际专利分类号:
G06F 17/30 (2006.01)
- (21) 国际申请号: PCT/CN2018/081518
- (22) 国际申请日: 2018 年 4 月 2 日 (02.04.2018)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201810161582.X 2018 年 2 月 27 日 (27.02.2018) CN
- (71) 申请人: 平安科技 (深圳) 有限公司 (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN]; 中国广东省深圳市福田区八卦岭八卦三路平安大厦 6 楼, Guangdong 518000 (CN).
- (72) 发明人: 石宇 (SHI, Yu); 中国广东省深圳市福田区八卦岭八卦三路平安大厦 6 楼, Guangdong 518000 (CN)。 李云利 (LI, Yunli); 中国广东省

深圳市福田区八卦岭八卦三路平安大厦 6 楼, Guangdong 518000 (CN)。 张文君 (ZHANG, Wenjun); 中国广东省深圳市福田区八卦岭八卦三路平安大厦 6 楼, Guangdong 518000 (CN)。 张小敏 (ZHANG, Xiaomin); 中国广东省深圳市福田区八卦岭八卦三路平安大厦 6 楼, Guangdong 518000 (CN)。 吴春海 (WU, Chunhai); 中国广东省深圳市福田区八卦岭八卦三路平安大厦 6 楼, Guangdong 518000 (CN)。 李辉 (LI, Hui); 中国广东省深圳市福田区八卦岭八卦三路平安大厦 6 楼, Guangdong 518000 (CN)。 陈晶 (CHEN, Jing); 中国广东省深圳市福田区八卦岭八卦三路平安大厦 6 楼, Guangdong 518000 (CN)。

(74) 代理人: 深圳众鼎专利商标代理事务所 (普通合伙) (ZHONGDING INTELLECTUAL PROPERTY AGENCY); 中国广东省深圳市龙岗区龙城

(54) Title: METHOD FOR RAPIDLY IMPORTING BIG DATA, APPARATUS, TERMINAL DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 大数据快速导入的方法、装置、终端设备及存储介质

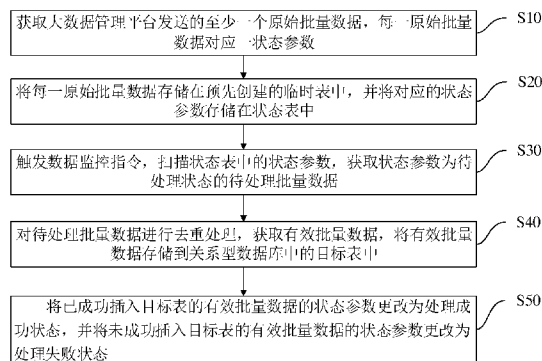


图 1

- S10 Acquire at least one piece of original batch data sent by a big data management platform, each piece of original batch data corresponding to a state parameter
- S20 Store each piece of original batch data into a pre-created temporary table, and store the corresponding state parameter into a state table
- S30 Trigger a data monitoring instruction, scan the state parameter in the state table, and acquire to-be-processed batch data of which the state parameter is a to-be-processed state
- S40 Perform de-duplication processing on the to-be-processed batch data, to acquire effective batch data, and store the effective batch data into a target table in a relational database
- S50 Change the state parameter of the effective batch data which has been successfully inserted into the target table into a process success state, and change the state parameter of the effective batch data which is not successfully inserted into the target table into a process failure state

(57) Abstract: Disclosed in the present application are a method for rapidly importing big data, an apparatus, a terminal device, and a storage medium. The method for rapidly importing big data comprises: acquiring at least one piece of original batch data sent by a big data management platform, each piece of original batch data corresponding to a state parameter; storing each piece of original batch data into a pre-created temporary table, and storing the corresponding state parameter into a state table; triggering a data monitoring instruction, scanning the state parameter in the state table, and acquiring to-be-processed batch data of which the state parameter is in a to-be-processed state; performing de-duplication processing on the to-be-processed batch data, to acquire effective batch data, and storing the effective batch data into a target table in a relational database. The method can improve the efficiency of importing batch data into the relational database.

(57) 摘要: 本申请公开了一种大数据快速导入的方法、装置、终端设备及存储介质。该大数据快速导入的方法包括: 获取大数据管理平台发送的至少一个原始批量数据, 每一原始批量数据对应一状态参数; 将每一原始批量数据存储在预先创建的临时表中, 并将对应的状态参数存储在状态表中; 触发数据监控指令, 扫描状态表中的状态参数, 获取状态参数为待处理状态的待处理批量数据; 对待处理批量数据进行去重处理, 获取有效批量数据, 将有效批量数据存储到关系型数据库中的目标表中。该方法可以提高批量数据导入到关系型数据库的效率。

街道中心城清林路 546 号城投商务中心 4 层/B, Guangdong 518172 (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, 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)。

本国际公布:

— 包括国际检索报告 (条约第 21 条 (3))。

大数据快速导入的方法、装置、终端设备及存储介质

本申请以 2018 年 2 月 27 日提交的申请号为 201810161582.X, 名称为“大数据快速导入的方法、装置、终端设备及存储介质”的中国发明专利申请为基础, 并要求其优先权。

技术领域

本申请涉及大数据处理领域, 尤其涉及一种大数据快速导入的方法、装置、终端设备及存储介质。

背景技术

在互联网和金融行业, 大数据管理平台处理的数据的数据量较大, 其数据占用磁盘空间会达到 500G 甚至 T 级别。为了便于对数据进行分析处理, 需将大数据管理平台中数据量较大的数据(以下简称批量数据)导入到关系型数据库(如 MySQL, Oracle 或 Postgres 等)中。在当前大数据管理平台中, 通常采用 Sqoop 工具将批量数据导入到关系型数据库。其中, Sqoop 工具是一个用来将 Hadoop(即分布式系统基础架构)和关系型数据库中的数据相互转移的工具, 即可以将关系型数据库中的数据导入到 Hadoop 的 HDFS(Hadoop Distributed File System, 分布式文件系统)中, 也可以将 HDFS 中的数据导入到关系型数据库中。当前批量数据导入到关系型数据库的过程中, 经常会出现数据导入效率低的问题。

发明内容

本申请提供一种大数据快速导入的方法、装置、终端设备及存储介质, 以解决当前批量数据导入到关系型数据库的过程经常会出现数据导入效率低的问题。

第一方面, 本申请提供一种大数据快速导入的方法, 包括:

获取大数据管理平台发送的至少一个原始批量数据, 每一原始批量数据对应一状态参数;

将每一所述原始批量数据存储于预先创建的临时表中, 并将对应的状态参数存储于状态表中;

触发数据监控指令, 扫描所述状态表中的状态参数, 获取状态参数为待处理状态的待处理批量数据;

对所述待处理批量数据进行去重处理, 获取有效批量数据, 将所述有效批量数据存储到关系型数据库中的目标表中。

第二方面，本申请提供一种大数据快速导入的装置，包括：

原始批量数据获取模块，用于获取大数据管理平台发送的至少一个原始批量数据，每一原始批量数据对应一状态参数；

第一存储模块，用于将每一所述原始批量数据存储在预先创建的临时表中，并将对应的状态参数存储在状态表中；

待处理批量数据获取模块，用于触发数据监控指令，扫描所述状态表中的状态参数，获取状态参数为待处理状态的待处理批量数据；

第二存储模块，用于对所述待处理批量数据进行去重处理，获取有效批量数据，将所述有效批量数据存储到关系型数据库中的目标表中。

第三方面，本申请提供一种终端设备，包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机可读指令，所述处理器执行所述计算机可读指令时实现如下步骤：

获取大数据管理平台发送的至少一个原始批量数据，每一原始批量数据对应一状态参数；

将每一所述原始批量数据存储在预先创建的临时表中，并将对应的状态参数存储在状态表中；

触发数据监控指令，扫描所述状态表中的状态参数，获取状态参数为待处理状态的待处理批量数据；

对所述待处理批量数据进行去重处理，获取有效批量数据，将所述有效批量数据存储到关系型数据库中的目标表中。

第四方面，本申请提供一个或多个存储有计算机可读指令的非易失性可读存储介质，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器实现如下步骤：

获取大数据管理平台发送的至少一个原始批量数据，每一原始批量数据对应一状态参数；

将每一所述原始批量数据存储在预先创建的临时表中，并将对应的状态参数存储在状态表中；

触发数据监控指令，扫描所述状态表中的状态参数，获取状态参数为待处理状态的待处理批量数据；

对所述待处理批量数据进行去重处理，获取有效批量数据，将所述有效批量数据存储

到关系型数据库中的目标表中。

本申请的一个或多个实施例的细节在下面的附图及描述中提出。本申请的其他特征和优点将从说明书、附图以及权利要求书变得明显。

附图说明

为了更清楚地说明本申请的技术方案，下面将对本申请的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申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

图 1 是实施例 1 中大数据快速导入的方法的一流程图；

图 2 是图 1 中步骤 S20 的一具体流程图；

图 3 是图 2 中步骤 S22 的一具体流程图；

图 4 是图 1 中步骤 S30 的一具体流程图；

图 5 是实施例 2 中大数据快速导入的装置的一原理框图；

图 6 是实施例 4 中提供的终端设备的示意图。

具体实施方式

下面将结合本申请中的附图，对本申请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申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

实施例 1

图 1 示出本实施例中大数据快速导入的方法的流程图。该大数据快速导入的方法应用数据库管理系统中，该数据库管理系统与大数据管理平台相连，用于将大数据管理平台的数据量较大的大型数据（即批量数据）快速导入到关系型数据库（如 MySQL , Oracle 或 Postgres 等传统的关系型数据），以提高批量数据导入的效率。本实施例中，以 Oracle 数据库管理系统为例进行说明，用于将大数据管理平台中的批量数据快速导入到 Oracle 数据库中。如图 1 所示，该大数据快速导入的方法包括如下步骤：

S10：获取大数据管理平台发送的至少一个原始批量数据，每一原始批量数据对应一状态参数。

原始批量数据是指大数据管理平台预先存储但未上传给关系型数据库的批量数据。其中，批量数据是指数据量较大的数据。具体地，该原始批量数据具体可以是不同的业务系统（如大金融资产业务系统或大医疗健康业务系统等）发送给大数据管理平台的业务数据。

该业务数据包括但不限于本实施例提及的大金融资产业务系统提供的开户信息数据和大医疗健康业务系统的用户注册信息数据。

由于大数据管理平台存储着大量的来自不同业务系统发送的业务数据，业务系统在发送业务数据时，都是批量发送的，数据量较大，因此，可以每一业务系统批量发送的业务数据定义为原始批量数据。大数据管理平台存储不同业务数据发送的原始批量数据，并将该原始批量数据作为一个整体导入到关系型数据库。

每一原始批量数据都有一个对应的状态参数。状态参数是指描述每一原始批量数据对应的处理状态的参数，该处理状态包括待处理状态、处理中状态、处理成功状态和处理失败状态四种状态参数。本实施例中，为了方便辨识不同的处理状态，使用状态参数表示对应的处理状态，具体为：0 对应待处理状态，1 对应处理中状态，2 对应处理成功状态，3 对应处理失败状态。具体地，大数据管理平台存储的每一原始批量数据对应的状态参数为待处理，即对应 0 这一状态参数。

由于四种处理状态中，只有待处理状态对应的原始批量数据是未经处理的，需上传到 Oracle 数据库管理系统进行处理，因此，大数据管理平台仅将存储的状态参数为 0 和对应的原始批量数据采用 Sqoop 工具发送给 Oracle 数据库管理系统。只将处于待处理状态的原始批量数据发送给 Oracle 数据库管理系统，方便有效辨识待处理状态的原始批量数据，有效减少 Oracle 数据库管理系统的数据处理量，提高数据处理效率。相应地，状态参数不为 0 的状态参数和对应的原始批量数据不会发送给 Oracle 数据库管理系统，以避免对原始批量数据进行重复处理，节省处理时间，以提高处理效率。

S20：将每一原始批量数据存储预先创建的临时表中，并将对应的状态参数存储在状态表中。

临时表是指 Oracle 数据库管理系统临时创建的，用来临时存储大数据管理平台发送的原始批量数据的数据表。状态表是指用于记录存储在 Oracle 临时表中的原始批量数据对应的处理状态的表格。

Oracle 数据库管理系统在获取到大数据管理平台发送的状态参数和对应的原始批量数据后，将原始批量数据存储到预先创建的临时表中，对应的状态参数存储到预先创建的状态表中。其中，一个原始批量数据对应一个状态表。可以理解地，Oracle 数据库管理系统在获取到原始批量数据时，其对应的处理状态为待处理状态，设置状态表中的状态参数为 0；在开始对原始批量数据进行处理时，其对应的处理状态为处理中状态，使得状态表中的状态参数更改为 1；并可依据其处理成功与否，分别确定其对应的属于处理成功状态

还是处理失败状态，以确定状态表中的状态参数为 2 或 3。本实施例中的处理是指将原始批量数据导入到关系型数据库的处理过程。

进一步地，当状态表中的状态参数为 2 时，即原始批量数据导出从临时表中导入到关系型数据库的处理过程成功时，临时表会将对应的原始批量数据释放，以节省临时表的存储空间。临时表起到了临时存储的作用。

S30：触发数据监控指令，扫描状态表中的状态参数，获取状态参数为待处理状态的待处理批量数据。

数据监控指令是指触发进行数据监控操作的指令。待处理批量数据是指状态参数为待处理状态对应的原始批量数据。

当 Oracle 数据库管理系统将获取到的原始批量数据和对应的状态参数存储到预先创建的临时表中和状态表后，在特定触发事件发生时，Oracle 数据库管理系统会触发数据监控指令，启动数据监控程序，扫描状态表中存储的所有状态参数，当状态参数为 0 时，获取该状态参数对应的待处理批量数据。通过数据监控程序，只获取状态参数为 0 对应的待处理批量数据，方便 Oracle 数据库管理系统有针对性地对待处理批量数据执行步骤 S40，减少了批量数据的处理数量，以提高处理效率。本实施例中的特定触发事件是指在预先设置的触发时间内，触发数据监控指令执行数据监控操作的事件，当系统当前时间为该触发时间时，则形成这一特定触发事件。

S40：对待处理批量数据进行去重处理，获取有效批量数据，将有效批量数据存储到关系型数据库中的目标表中。

有效批量数据是指对待处理批量数据进行去重处理后得到的批量数据。本实施例中，获取到的待处理批量数据后，Oracle 数据库管理系统会基于指定列对获取的待处理批量数据进行去重处理。该指定列是指预先指定的包含数据存储时间的数据列。在开始对待处理批量数据进行去重处理时，需修改该待处理批量数据对应的状态表中的状态参数为 1，表示该待处理批量数据处于处理中状态，以避免重复对该待处理批量数据进行处理，节省处理时间，提高处理效率。

当状态表中的状态参数改为 1 时，Oracle 数据库管理系统会将修改后的状态参数返回给大数据管理平台对应的原始批量数据对应的状态参数，以避免大数据管理平台重复将这相同的原始批量数据发送给 Oracle 数据库管理系统，从而导致重复处理。

在对待处理批量数据进行去重处理过程中，若待处理批量数据存在至少两个相同的数据时，则基于数据对应的存储时间只保留最新时间对应的数据，将该最新时间之前存储的

数据进行清除处理，以节省数据的存储空间。其中，最新时间是指与系统当前时间最接近的时间。将进行去重处理后得到的有效批量数据存储到对应的关系型数据库中，从而提高批量数据导入到关系型数据的效率。

如有效批量数据是保险业务系统发送的某个保险产品对应的投保人的个人信息数据，投保人的个人信息中包括但不限于投保人 ID、投保人姓名、预留手机号和邮箱地址，若一个投保人 ID 对应了至少两个预留手机号或者邮箱地址时，则基于该投保人的个人信息的存储时间，删除最新时间之前的该投保人的个人信息，只保留最新时间对应的个人信息。将最新时间对应的个人信息作为有效批量数据存储到对应的关系型数据库中，方便业务人员获取对应的投保人的个人信息数据，提高了投保人的个人信息数据的准确性。

目标表是指关系型数据库中预先建立的数据表，该目标表是用于存储有效批量数据的数据表。Oracle 数据库管理系统对待处理批量数据进行去重处理，将获取的有效批量数据存储到关系型数据库中的目标表中，便于不同的业务系统在需要获取目标表中的数据时，直接调用目标表。

步骤 S10-S40，大数据管理平台将获取到的每一原始批量数据和对应的状态参数分别存储在 Oracle 数据库管理系统预先创建的临时表和状态表中，Oracle 数据库管理系统触发数据监控指令，扫描状态表中的状态参数，获取状态参数为待处理状态对应的待处理批量数据，对待处理批量数据进行去重处理后将获取的有效批量数据存储到关系型数据库创建的目标表中。通过该方法将原始批量数据存储存储在临时表中，对临时表中的原始批量数据进行处理后，将有效批量数据存储到关系型数据库的目标表中，克服了从大数据平台将原始批量数据导入到关系型数据库的目标表中时，由于数据类型不同导致数据导入的成功率不高的问题，从而也提高了原始批量数据导入到关系型数据库的目标表中的效率。

在一具体实施方式中，如图 2 所示，步骤 S20，将每一原始批量数据存储存储在预先创建的临时表中，具体包括如下步骤：

S21：采用映射归约函数将每一原始批量数据拆分成 N 个子批量数据， $N \geq 2$ 。

由于大数据管理平台导入到临时表中的数据量较大，为了加快原始批量数据导入到临时表中的速度，缩短导入时间，需要使用多个映射归约（Map-Reduce）进程对原始批量数据进行分布式导出。映射归约（Map-Reduce）是一个移动式的基于键值对（key-value）形成的分布式计算框架，由 Map（映射）函数和 Reduce（归约）函数组成。Map（映射）函数根据用户自定义的数据量将一个原始批量数据分解成 N （ $N \geq 2$ ）子批量数据，然后将子批量数据中的键值对映射成新的键值对，一个子批量数据中包括至少一个新的键值对。

每个子批量数据的大小是固定的，而 map（映射）函数对应的执行操作也是相同的，因此，并行执行 N 个子批量数据的操作时不会出现明显的时间不均衡，并行效率最高。

如大数据管理平台将保险业务系统关于投保人的个人信息对应的原始批量数据导入到临时表中时，Map（映射）函数先根据用户自定义的条件将原始批量数据分解成 N ($N \geq 2$) 子批量数据。然后，将子批量数据中的键值对映射成新的键值对，原始批量数据中键值对为：投保人 ID 作为键（key），与投保人 ID 对应的投保业务、投保人姓名、预留手机号和邮箱地址作为值（value），即原始批量数据中的投保人 ID 与该投保人 ID 对应的投保业务、投保人姓名、预留手机号和邮箱地址形成了一种映射关系，组成了一个键值对（key-value）。新的键值对是指以投保业务作为键（key），与投保业务对应的投保人 ID、投保人姓名、预留手机号和邮箱地址作为值（value），即投保业务与该投保业务对应的投保人 ID、投保人姓名、预留手机号和邮箱地址形成了一种新的映射关系，组成了一个新的键值对（key-value）。由于一个投保人 ID 可能会对应多个投保业务，当新的键为投保业务时，一个子批量数据中会包含至少一个投保业务，即一个子批量数据中会出现至少一个新的键。

S22：采用 M 个映射归约函数对 N 个子批量数据进行导出处理，以将导出的子批量数据存储到临时表中， $M \geq N$ ，其中 N、M 为正整数。

将步骤 S21 获取的每一子批量数据发送给对应的 Reduce（归约）函数。一般地，由于 Reduce（归约）函数在对数据进行汇总时，会不可避免地出现汇总失败的问题，因此 Reduce（归约）函数不会少于子批量数据的数量，即 $M \geq N$ ，以便于采用其他子批量数据的 Reduce（归约）函数对处理失败的子批量数据进行处理，提高处理效率。其中，N、M 为正整数。Reduce（归约）函数在 map（映射）函数获取至少一个子批量数据时，会基于新的键对每一子批量数据进行汇总处理，即将同一个新的键（key）对应的值（value）汇总到一起，形成基于新的键确定的数据表，将该数据表根据用户自定义的数据量将对应的子批量数据导入到临时表中。采用映射归约函数中的 M 个归约函数对 N 个子批量数据进行导出处理，存储到临时表中，可以提高数据导出的速度，节省时间。

本实施例中，为了保证步骤 S20 的实现，临时表不设置主键和唯一索引。其中，主键可以是一个列或多列的组合，其值能唯一地标识表中的每一行，通过它可强制表的实体完整性。主键主要用于记录表与表之间的关联，以及用于记录本表的数据修改与删除。唯一索引是对数据库表中一列或多列的值进行排序的一种结构。临时表不设置主键，可以保证原始批量数据中的一分子批量数据先成功导入到临时表中后，不会影响还没有成功导入

到临时表中的其他部分子批量数据。若临时表中设置主键，则在一部分子批量数据导入到目标中后，目标表检测到设置的主键中已经有对应的数据存在，就会使 Oracle 数据库管理系统触发主键不唯一的系统错误，导致整个 Mapreduce 进程终止，从而导致数据导出失败。临时表不设置唯一索引，可以依据不同索引对子批量数据进行逐条的扫描，导入到临时表中，保证了数据导入的全面性。

如一个子批量数据包括两个新的键，投保业务 A 和投保业务 B，将每一子批量数据发送给对应的 Reduce（归约）函数后，Reduce（归约）函数会将每一子批量数据中属于同一个键的值归约到一起，即将投保业务 A 对应的值归约到一起，投保业务 B 对应的值归约到一起，形成投保业务 A 和投保业务 B 对应的数据表。将形成的投保业务 A 和投保业务 B 对应的数据表导入到临时表中进行存储。

在一具体实施方式中，如图 3 所示，步骤 S22，采用 M 个映射归约函数对 N 个子批量数据进行导出处理，以将导出的子批量数据存储存储在临时表中，具体包括如下步骤：

S221：采用 N 个映射归约函数分别对 N 个子批量数据进行导出处理，获取每一子批量数据对应的导出处理结果。

具体地，在 M 个映射归约函数选出 N 个映射归约函数对 N 个子批量数据进行一对一导出处理，即每一个归约函数对应一个子批量数据，使得 N 个子批量数据采用并行导出方式进行处理，以提高数据导出的效率。每一个映射归约函数对应的导出处理结果包括导出成功和导出失败。

S222：若导出处理结果为导出成功时，将导出的子批量数据存储存储在临时表中。

S223：若导出处理结果为导出失败时，则将导出处理结果为导出失败的子批量数据分配给第 N+1 个映射归约函数进行导出处理，直至所有导出处理结果均为导出成功时结束。

具体地，在导出过程中，当导出结果为导出失败时，映射归约函数会启动预先设置的重试机制。该重试机制是指在导出结果为导出失败时，将对应的子批量数据放在第 N+1 个映射归约函数中执行的机制。该大数据快速导入的方法中设置重试机制，可以保证在某一子批量数据的导出处理结果为导出失败时，服务器自动将导出失败的子批量数据放到另外一个空闲的映射归约进程中进行导出处理，提高了数据的导出成功率。

在一具体实施方式中，如图 4 所示，步骤 S30，触发数据监控指令，扫描状态表中的状态参数，具体包括如下步骤：

S31：获取定时触发指令，定时触发指令包括进程 ID 和触发时间。

定时触发指令是指在 oracle 数据库管理系统预先设置的定时触发数据监控一个触发

指令，该定时触发指令包括进程 ID 和触发时间。进程 ID 是指对原始批量数据进行导出处理的数据监控进程的标识，该进程 ID 可用于唯一识别对原始批量数据进行导出处理的数据监控进程。触发时间是指预先设置好的一个时间，当系统当前时间到达该预先设置好的时间时，则触发对应的定时触发指令中进程 ID 对应的数据监控进程。

oracle 数据库管理系统预先设置一个定时触发指令，该定时触发指令在 oracle 数据库管理系统的系统当前时间与预设的触发时间相同时，会触发 oracle 数据库管理系统设置的数据监控指令，基于数据监控指令，执行步骤 S32。

S32：若系统当前时间为触发时间时，触发与进程 ID 相对应的数据监控进程扫描状态表中的状态参数。

获取定时触发指令后，oracle 数据库管理系统会判断该系统当前时间是否与预先设置的触发时间相同。若该当前时间与预先设置的触发时间相同，则触发与进程 ID 相对应的数据监控进程扫描状态表中的状态参数。当扫描到状态参数为 0 时，则对该状态表对应的原始批量数据进行导出处理，即同时对该处理状态为待处理状态的原始批量数据进行导出处理。

在一具体实施方式中，步骤 S40，将有效批量数据存储到关系型数据库中的目标表中，具体包括如下步骤：将有效批量数据采用归并排序算法进行归并处理，插入到关系型数据库创建的目标表中。

归并排序算法是建立在归并操作上的一种有效的排序算法，也称为 merge 算法，归并排序算法是指将获取到的有效批量数据和目标表中存储的数据进行匹配查询，匹配成功时则执行更新操作，匹配不成功时则执行插入操作。本实施例中仅采用插入操作对有效批量数据进行插入，即将目标表中不存在的有效批量数据插入到目标表中。

具体地，oracle 数据库管理系统获取的有效批量数据和存储在关系型数据库中的目标表进行匹配查询，将匹配不成功的有效批量数据插入到目标表中，即将有效批量数据中存在的部分数据，但目标表中不存在的数据插入到目标表中。使用归并排序算法将目标表中不存在的有效批量数据插入到目标表中，仅需要做一次全表扫描就可以完成将目标表中不存在的有效批量数据全部插入到目标表中，提高了数据插入的效率。

在一具体实施方式中，在步骤 S40，获取有效批量数据，将有效批量数据存储到关系型数据库中的目标表中的步骤之后，该大数据快速导入的方法，还包括：

S50：将已成功插入目标表的有效批量数据的状态参数更改为处理成功状态，并将未成功插入目标表的有效批量数据的状态参数更改为处理失败状态。

具体地，有效批量数据成功插入到目标表中后，有效批量数据对应的状态表中的状态参数为 2，表示处理成功状态，有效批量数据没有成功插入到目标表中时，有效批量数据对应的状态表中的状态参数为 3，表示处理失败状态。Oracle 数据库管理系统基于状态参数，可以辨识有效批量数据插入目标表的处理情况，方便 Oracle 数据库管理系统对处理失败的有效批量数据及时作出重新插入的反映。

有效批量数据没有成功插入到目标表中的原因包括，临时表太大、插入超时或者网络等原因。若有效批量数据没有成功插入到目标表的原因为临时表有效批量数据太大或者插入超时，则表示插入失败的原因是临时表中的数据量过大，不利于进行插入操作，应该将该临时表的有效批量数据进行拆分，分为多个子有效批量数据插入到目标表中；若插入失败的原因是网络原因，则刷新该临时表的插入进程，将有效批量数据重新插入到目标表中。

本申请提供的大数据快速导入的方法，在大数据管理平台将获取到的每一原始批量数据和对应的状态参数导入到 Oracle 数据库管理系统中的目标表中时，在 Oracle 数据库管理系统预先创建临时表和状态表，采用映射规约函数将原始批量数据拆分成 N 个子批量数据，进行导出处理，提高数据导出效率。Oracle 数据库管理系统触发数据监控指令，扫描状态表中的状态参数，获取状态参数为待处理状态对应的待处理批量数据，在将待处理批量数据进行去重处理，以避免重复对该待处理批量数据进行处理，节省处理时间，提高处理效率。然后，采用归并排序算法将有效批量数据和目标表中的数据进行匹配，对匹配不成功的数据插入到目标表中，仅需要做一次全表扫描就可以完成将目标表中不存在的有效批量数据全部插入到目标表中，提高了数据插入执行的效率。

应理解，上述实施例中各步骤的序号的大小并不意味着执行顺序的先后，各过程的执行顺序应以其功能和内在逻辑确定，而不对本申请的实施过程构成任何限定。

实施例 2

图 2 示出与实施例 1 中大数据快速导入的方法一一对应的大数据快速导入的装置的原理框图。如图 2 所示，该大数据快速导入的装置包括原始批量数据获取模块 10、第一存储模块 20、待处理批量数据获取模块 30 和第二存储模块 40。其中，原始批量数据获取模块 10、第一存储模块 20、待处理批量数据获取模块 30、第二存储模块 40 和状态参数更改模块 50 的实现功能与实施例中大数据快速导入的方法对应的步骤一一对应，为避免赘述，本实施例不一一详述。

原始批量数据获取模块 10，用于获取大数据管理平台发送的至少一个原始批量数据，每一原始批量数据对应一状态参数。

第一存储模块 20，用于将每一原始批量数据存储在预先创建的临时表中，并将对应的状态参数存储在状态表中。

待处理批量数据获取模块 30，用于触发数据监控指令，扫描状态表中的状态参数，获取状态参数为待处理状态的待处理批量数据。

第二存储模块 40，用于对待处理批量数据进行去重处理，获取有效批量数据，将有效批量数据存储到关系型数据库中的目标表中。

优选地，临时表不设置主键和唯一索引。

第一存储模块 20 包括：子批量数据获取单元 21 和子批量数据存储单元 22。

子批量数据获取单元 21，用于采用映射归约函数将每一原始批量数据拆分成 N 个子批量数据， $N \geq 2$ 。

子批量数据存储单元 22，用于采用 M 个映射归约函数对 N 个子批量数据进行导出处理，以将导出的子批量数据存储存储在临时表中， $M \geq N$ ，其中 N 、 M 为正整数。

优选地，子批量数据存储单元 22 包括导出处理结果获取单元 221、第一导出结果处理单元 222 和第二导出结果处理单元 223。

导出处理结果获取单元 221，用于采用 N 个映射归约函数分别对 N 个子批量数据进行导出处理，获取每一子批量数据对应的导出处理结果。

第一导出结果处理单元 222，用于若导出处理结果为导出成功时，将导出的子批量数据存储存储在临时表中。

第二导出结果处理单元 223，用于若导出处理结果为导出失败时，则将导出处理结果为导出失败的子批量数据分配给第 $N+1$ 个映射归约函数进行导出处理，直至所有导出处理结果均为导出成功时结束。

优选地，待处理批量数据获取模块 30 包括定时触发指令获取单元 31 和数据监控进程扫描单元 32。

定时触发指令获取单元 31，用于获取定时触发指令，定时触发指令包括进程 ID 和触发时间。

数据监控进程扫描单元 32，用于若系统当前时间为触发时间时，触发与进程 ID 相对应的数据监控进程扫描状态表中的状态参数。

优选地，第二存储模块 40，用于将有效批量数据采用归并排序算法进行归并处理，插入到关系型数据库创建的目标表中。

优选地，该大数据快速导入的装置，还包括状态参数更改模块 50。

状态参数更改模块 50，用于将已成功插入目标表的有效批量数据的状态参数更改为已成功处理状态，并将未成功插入目标表的有效批量数据的状态参数更改为未成功处理状态。

实施例 3

本实施例提供一个或多个存储有计算机可读指令的非易失性可读存储介质。该一个或多个存储有计算机可读指令的非易失性可读存储介质中，计算机可读指令被一个或多个处理器执行时，使得一个或多个处理器执行实施例 1 中大数据快速导入的方法，为避免重复，这里不再赘述。或者，该计算机可读指令被处理器执行时实现实施例 2 中大数据快速导入的装置中各模块/单元的功能，为避免重复，这里不再赘述。

可以理解地，一个或多个存储有计算机可读指令的非易失性可读存储介质可以包括：能够携带所述计算机可读指令代码的任何实体或装置、记录介质、U 盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器（ROM，Read-Only Memory）、随机存取存储器（RAM，Random Access Memory）、电载波信号和电信信号等。

实施例 4

图 6 是本申请一实施例提供的终端设备的示意图。如图 6 所示，该实施例的终端设备 60 包括：处理器 61、存储器 62 以及存储在存储器 62 中并可在处理器 61 上运行的计算机可读指令 63。处理器 61 执行计算机可读指令 63 时实现上述实施例 1 中大数据快速导入的方法的步骤，例如图 1 所示的步骤 S10 至 S50。或者，处理器 61 执行计算机可读指令 63 时实现上述实施例 2 中大数据快速导入的装置中各模块/单元的功能，例如图 5 所示原始批量数据获取模块 10、第一存储模块 20、待处理批量数据获取模块 30、第二存储模块 40 和状态参数更改模块 50 的功能。

所属领域的技术人员可以清楚地了解到，为了描述的方便和简洁，仅以上述各功能单元、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单元、模块完成，即将所述装置的内部结构划分成不同的功能单元或模块，以完成以上描述的全部或者部分功能。

以上所述实施例仅用以说明本申请的技术方案，而非对其限制；尽管参照前述实施例对本申请进行了详细的说明，本领域的普通技术人员应当理解：其依然可以对前述各实施例所记载的技术方案进行修改，或者对其中部分技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本申请各实施例技术方案的精神和范围，均应包含在本申请的保护范围之内。

权 利 要 求 书

1. 一种大数据快速导入的方法，其特征在于，包括：

获取大数据管理平台发送的至少一个原始批量数据，每一原始批量数据对应一状态参数；

将每一所述原始批量数据存储在预先创建的临时表中，并将对应的状态参数存储在状态表中；

触发数据监控指令，扫描所述状态表中的状态参数，获取状态参数为待处理状态的待处理批量数据；

对所述待处理批量数据进行去重处理，获取有效批量数据，将所述有效批量数据存储到关系型数据库中的目标表中。

2. 如权利要求 1 所述的大数据快速导入的方法，其特征在于，所述临时表不设置主键和唯一索引；

所述将每一所述原始批量数据存储在预先创建的临时表中，包括：

采用映射归约函数将每一所述原始批量数据拆分成 N 个子批量数据， $N \geq 2$ ；

采用 M 个映射归约函数对 N 个所述子批量数据进行导出处理，以将导出的子批量数据存储于所述临时表中， $M \geq N$ ，其中 N 、 M 为正整数。

3. 如权利要求 2 所述的大数据快速导入的方法，其特征在于，所述采用 M 个映射归约函数对 N 个所述子批量数据进行导出处理，以将导出的子批量数据存储于所述临时表中，包括：

采用 N 个映射归约函数分别对 N 个所述子批量数据进行导出处理，获取每一所述子批量数据对应的导出处理结果；

若所述导出处理结果为导出成功时，将导出的子批量数据存储于所述临时表中；

若所述导出处理结果为导出失败时，则将导出处理结果为导出失败的所述子批量数据分配给第 $N+1$ 个映射归约函数进行导出处理，直至所有导出处理结果均为导出成功时结束。

4. 如权利要求 1 所述的大数据快速导入的方法，其特征在于，所述触发数据监控指令，扫描所述状态表中的状态参数，包括：

获取定时触发指令，所述定时触发指令包括进程 ID 和触发时间；

若系统当前时间为所述触发时间时，触发与所述进程 ID 相对应的数据监控进程扫描所述状态表中的状态参数。

5. 如权利要求 1 所述的大数据快速导入的方法, 其特征在于, 所述将所述有效批量数据存储到关系型数据库中的目标表中, 包括;

将所述有效批量数据采用归并排序算法进行归并处理, 插入到关系型数据库创建的目标表中。

6. 如权利要求 1 所述的大数据快速导入的方法, 其特征在于, 在所述获取有效批量数据, 将所述有效批量数据存储到关系型数据库中的目标表中的步骤之后, 所述大数据快速导入的方法, 还包括:

将已成功插入所述目标表的有效批量数据的状态参数更改为已成功处理状态, 并将未成功插入所述目标表的有效批量数据的状态参数更改为处理失败状态。

7. 一种大数据快速导入的装置, 其特征在于, 包括:

原始批量数据获取模块, 用于获取大数据管理平台发送的至少一个原始批量数据, 每一原始批量数据对应一状态参数;

第一存储模块, 用于将每一所述原始批量数据存储预先创建的临时表中, 并将对应的状态参数存储在状态表中;

待处理批量数据获取模块, 用于触发数据监控指令, 扫描所述状态表中的状态参数, 获取状态参数为待处理状态的待处理批量数据;

第二存储模块, 用于对所述待处理批量数据进行去重处理, 获取有效批量数据, 将所述有效批量数据存储到关系型数据库中的目标表中。

8. 如权利要求 7 所述的大数据快速导入的装置, 其特征在于, 所述临时表不设置主键和唯一索引;

所述第一存储模块包括:

子批量数据获取单元, 用于采用映射归约函数将每一所述原始批量数据拆分成 N 个子批量数据, $N \geq 2$;

子批量数据存储单元, 用于采用 M 个映射归约函数对 N 个所述子批量数据进行导出处理, 以将导出的子批量数据存储所述临时表中, $M \geq N$, 其中 N 、 M 为正整数。

9. 一种终端设备, 包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机可读指令, 其特征在于, 所述处理器执行所述计算机可读指令时实现如下步骤:

获取大数据管理平台发送的至少一个原始批量数据, 每一原始批量数据对应一状态参数;

将每一所述原始批量数据存储在预先创建的临时表中，并将对应的状态参数存储在状态表中；

触发数据监控指令，扫描所述状态表中的状态参数，获取状态参数为待处理状态的待处理批量数据；

对所述待处理批量数据进行去重处理，获取有效批量数据，将所述有效批量数据存储到关系型数据库中的目标表中。

10. 如权利要求 9 所述的终端设备，其特征在于，所述临时表不设置主键和唯一索引；

所述将每一所述原始批量数据存储在预先创建的临时表中，包括：

采用映射归约函数将每一所述原始批量数据拆分成 N 个子批量数据， $N \geq 2$ ；

采用 M 个映射归约函数对 N 个所述子批量数据进行导出处理，以将导出的子批量数据存储于所述临时表中， $M \geq N$ ，其中 N 、 M 为正整数。

11. 如权利要求 10 所述的终端设备，其特征在于，所述采用 M 个映射归约函数对 N 个所述子批量数据进行导出处理，以将导出的子批量数据存储于所述临时表中，包括：

采用 N 个映射归约函数分别对 N 个所述子批量数据进行导出处理，获取每一所述子批量数据对应的导出处理结果；

若所述导出处理结果为导出成功时，将导出的子批量数据存储于所述临时表中；

若所述导出处理结果为导出失败时，则将导出处理结果为导出失败的所述子批量数据分配给第 $N+1$ 个映射归约函数进行导出处理，直至所有导出处理结果均为导出成功时结束。

12. 如权利要求 9 所述的终端设备，其特征在于，所述触发数据监控指令，扫描所述状态表中的状态参数，包括：

获取定时触发指令，所述定时触发指令包括进程 ID 和触发时间；

若系统当前时间为所述触发时间时，触发与所述进程 ID 相对应的数据监控进程扫描所述状态表中的状态参数。

13. 如权利要求 9 所述的终端设备，其特征在于，所述将所述有效批量数据存储到关系型数据库中的目标表中，包括：

将所述有效批量数据采用归并排序算法进行归并处理，插入到关系型数据库创建的目标表中。

14. 如权利要求 9 所述的终端设备，其特征在于，在所述获取有效批量数据，将所述有效批量数据存储到关系型数据库中的目标表中的步骤之后，所述处理器执行所述计算机可读指令时还实现如下步骤：

将已成功插入所述目标表的有效批量数据的状态参数更改为已成功处理状态，并将未成功插入所述目标表的有效批量数据的状态参数更改为处理失败状态。

15. 一个或多个存储有计算机可读指令的非易失性可读存储介质，其特征在于，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器实现如下步骤：

获取大数据管理平台发送的至少一个原始批量数据，每一原始批量数据对应一状态参数；

将每一所述原始批量数据存储于预先创建的临时表中，并将对应的状态参数存储于状态表中；

触发数据监控指令，扫描所述状态表中的状态参数，获取状态参数为待处理状态的待处理批量数据；

对所述待处理批量数据进行去重处理，获取有效批量数据，将所述有效批量数据存储到关系型数据库中的目标表中。

16. 如权利要求 15 所述的非易失性可读存储介质，其特征在于，所述临时表不设置主键和唯一索引；

所述将每一所述原始批量数据存储于预先创建的临时表中，包括：

采用映射归约函数将每一所述原始批量数据拆分成 N 个子批量数据， $N \geq 2$ ；

采用 M 个映射归约函数对 N 个所述子批量数据进行导出处理，以将导出的子批量数据存储于所述临时表中， $M \geq N$ ，其中 N 、 M 为正整数。

17. 如权利要求 16 所述的非易失性可读存储介质，其特征在于，所述采用 M 个映射归约函数对 N 个所述子批量数据进行导出处理，以将导出的子批量数据存储于所述临时表中，包括：

采用 N 个映射归约函数分别对 N 个所述子批量数据进行导出处理，获取每一所述子批量数据对应的导出处理结果；

若所述导出处理结果为导出成功时，将导出的子批量数据存储于所述临时表中；

若所述导出处理结果为导出失败时，则将导出处理结果为导出失败的所述子批量数据分配给第 $N+1$ 个映射归约函数进行导出处理，直至所有导出处理结果均为导出成功时结束。

18. 如权利要求 15 所述的非易失性可读存储介质，其特征在于，所述触发数据监控指令，扫描所述状态表中的状态参数，包括：

获取定时触发指令，所述定时触发指令包括进程 ID 和触发时间；

若系统当前时间为所述触发时间时，触发与所述进程 ID 相对应的数据监控进程扫描

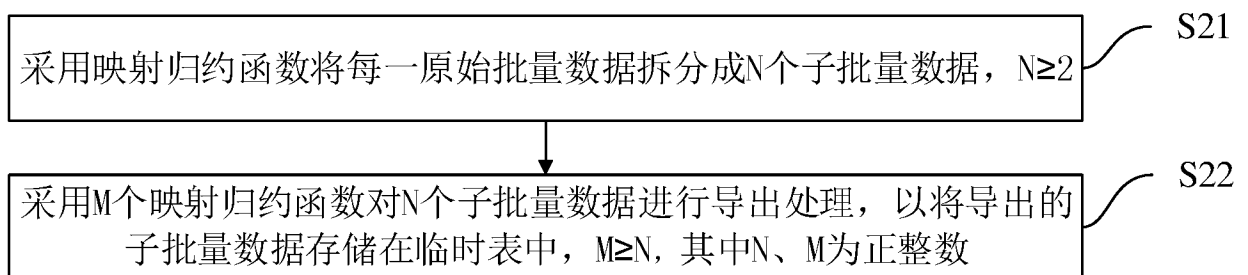
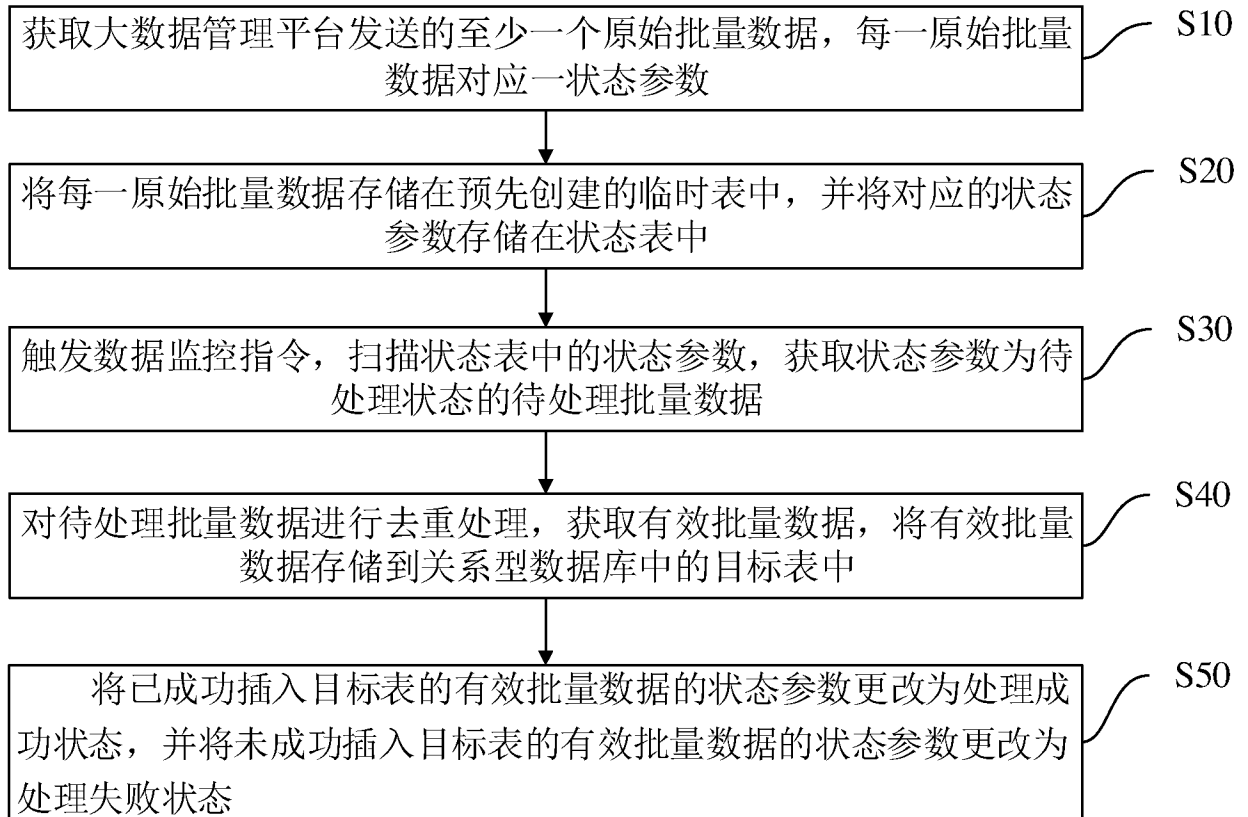
所述状态表中的状态参数。

19. 如权利要求 15 所述的非易失性可读存储介质，其特征在于，所述将所述有效批量数据存储到关系型数据库中的目标表中，包括：

将所述有效批量数据采用归并排序算法进行归并处理，插入到关系型数据库创建的目标表中。

20. 如权利要求 15 所述的非易失性可读存储介质，其特征在于，在所述获取有效批量数据，将所述有效批量数据存储到关系型数据库中的目标表中的步骤之后，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器还实现如下步骤：

将已成功插入所述目标表的有效批量数据的状态参数更改为已成功处理状态，并将未成功插入所述目标表的有效批量数据的状态参数更改为处理失败状态。



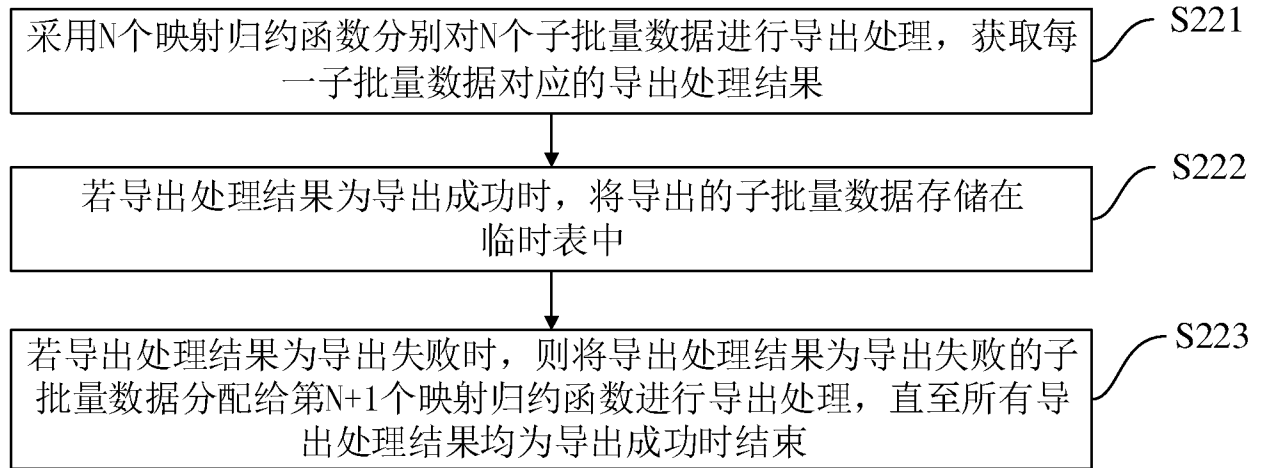


图 3

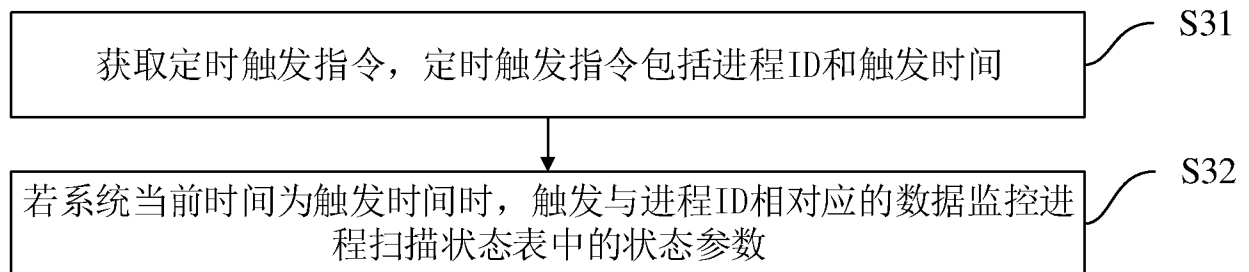


图 4

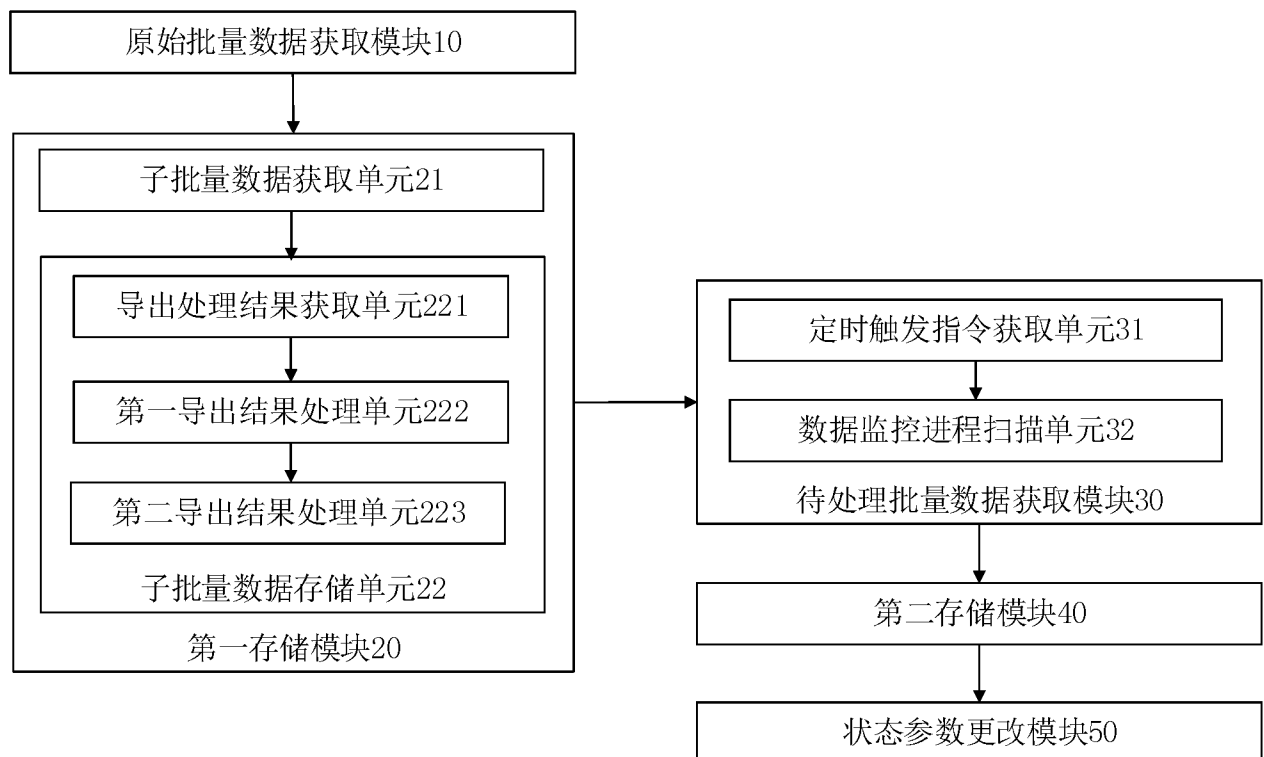


图 5

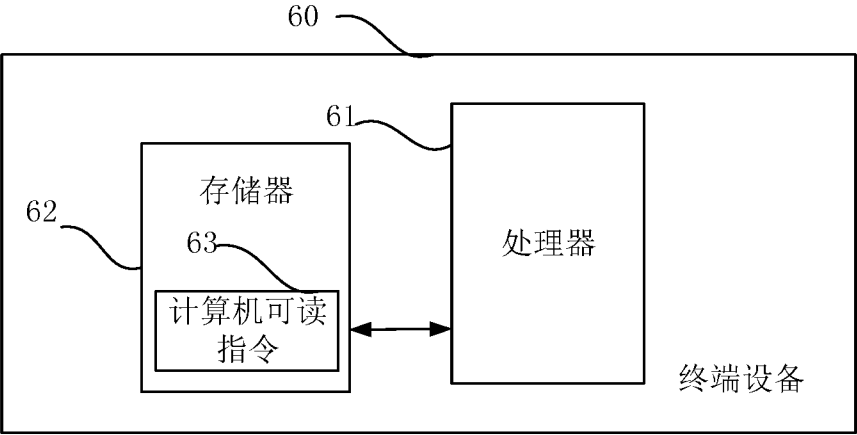


图 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/081518

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/30(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)

CNPAT, CNKI, EPODOC, WPI, IEEE: 大数据, 导入, 检测, 监测, 状态, 待处理, 去重, 存储, 关系型数据库, data, detect, state, lay in, repeat, store, MySQL, Oracle, postgres, Sqoop

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 106649455 A (BEIJING RUN TECHNOLOGY CO., LTD.) 10 May 2017 (2017-05-10) description, paragraphs 0003, 0047-0054 and 0094-0098	1-20
A	CN 104424229 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 18 March 2015 (2015-03-18) entire document	1-20
A	CN 106874334 A (ALIBABA GROUP HOLDING LIMITED) 20 June 2017 (2017-06-20) entire document	1-20
A	US 2004078398 A1 (TAIWAN SEMICONDUCTOR MANUFACTURING CO., LTD.) 22 April 2004 (2004-04-22) 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

30 September 2018

Date of mailing of the international search report

01 November 2018

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

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2018/081518

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	106649455	A	10 May 2017	None			
CN	104424229	A	18 March 2015	None			
CN	106874334	A	20 June 2017	None			
US	2004078398	A1	22 April 2004	US	7200625	B2	03 April 2007

国际检索报告

国际申请号

PCT/CN2018/081518

A. 主题的分类 G06F 17/30 (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) G06F 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用)) CNPAT, CNKI, EPDOC, WPI, IEEE: 大数据, 导入, 检测, 监测, 状态, 待处理, 去重, 存储, 关系型数据库, data, detect, state, lay in, repeat, store, MySQL, Oracle, postgres, Sqoop																	
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 106649455 A (北京锐安科技有限公司) 2017年 5月 10日 (2017 - 05 - 10) 说明书第0003, 0047-0054, 0094-0098段</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 104424229 A (腾讯科技深圳有限公司) 2015年 3月 18日 (2015 - 03 - 18) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 106874334 A (阿里巴巴集团控股有限公司) 2017年 6月 20日 (2017 - 06 - 20) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>US 2004078398 A1 (TAIWAN SEMICONDUCTOR MANUFACTURING CO., LTD.) 2004年 4月 22日 (2004 - 04 - 22) 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 106649455 A (北京锐安科技有限公司) 2017年 5月 10日 (2017 - 05 - 10) 说明书第0003, 0047-0054, 0094-0098段	1-20	A	CN 104424229 A (腾讯科技深圳有限公司) 2015年 3月 18日 (2015 - 03 - 18) 全文	1-20	A	CN 106874334 A (阿里巴巴集团控股有限公司) 2017年 6月 20日 (2017 - 06 - 20) 全文	1-20	A	US 2004078398 A1 (TAIWAN SEMICONDUCTOR MANUFACTURING CO., LTD.) 2004年 4月 22日 (2004 - 04 - 22) 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 106649455 A (北京锐安科技有限公司) 2017年 5月 10日 (2017 - 05 - 10) 说明书第0003, 0047-0054, 0094-0098段	1-20															
A	CN 104424229 A (腾讯科技深圳有限公司) 2015年 3月 18日 (2015 - 03 - 18) 全文	1-20															
A	CN 106874334 A (阿里巴巴集团控股有限公司) 2017年 6月 20日 (2017 - 06 - 20) 全文	1-20															
A	US 2004078398 A1 (TAIWAN SEMICONDUCTOR MANUFACTURING CO., LTD.) 2004年 4月 22日 (2004 - 04 - 22) 全文	1-20															
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																	
* 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 2018年 9月 30日		国际检索报告邮寄日期 2018年 11月 1日															
ISA/CN的名称和邮寄地址 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 传真号 (86-10) 62019451		受权官员 奚惠宁 电话号码 86-10-53961813															

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/081518

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	106649455	A	2017年 5月 10日	无	
CN	104424229	A	2015年 3月 18日	无	
CN	106874334	A	2017年 6月 20日	无	
US	2004078398	A1	2004年 4月 22日	US 7200625 B2	2007年 4月 3日