

(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
2016年10月27日 (27.10.2016)



(10) 国际公布号
WO 2016/169237 A1

- (51) 国际专利分类号:
G06F 17/30 (2006.01)
- (21) 国际申请号: PCT/CN2015/092759
- (22) 国际申请日: 2015年10月23日 (23.10.2015)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201510198455.3 2015年4月23日 (23.04.2015) CN
- (71) 申请人: 中兴通讯股份有限公司 (ZTE CORPORATION) [CN/CN]; 中国广东省深圳市南山区高新技术产业园科技南路中兴通讯大厦, Guangdong 518057 (CN)。
- (72) 发明人: 韩焯 (HAN, Ye); 中国广东省深圳市南山区高新技术产业园科技南路中兴通讯大厦, Guangdong 518057 (CN)。
- (74) 代理人: 北京康信知识产权代理有限责任公司 (KANGXIN PARTNERS,P.C.); 中国北京市海淀区知春路甲48号盈都大厦A座16层, Beijing 100098 (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)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

(54) Title: DATA PROCESSING METHOD AND DEVICE

(54) 发明名称: 数据处理方法及装置

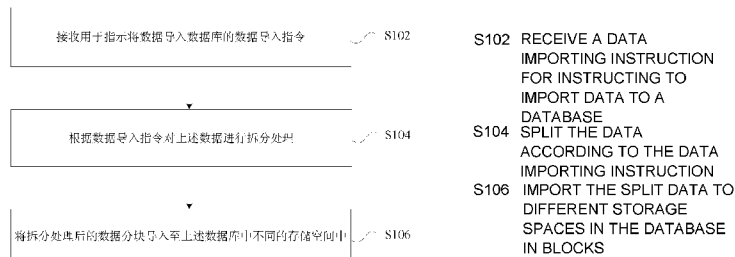
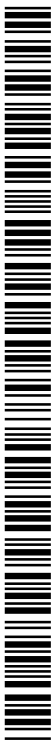


图1

(57) Abstract: Provided are a data processing method and device. The method comprises: receiving a data importing instruction for instructing a data import to a database; splitting the data according to the data importing instruction; and importing the split data to different storage spaces in the database in blocks. The present invention addresses the existing problem in the related art of low data importing efficiency, thus improving data importing efficiency.

(57) 摘要: 本发明提供了一种数据处理方法及装置, 其中, 该方法包括: 接收用于指示将数据导入数据库的数据导入指令; 根据数据导入指令对上述数据进行拆分处理; 将拆分处理后的数据分块导入至上述数据库中不同的存储空间中。通过本发明, 解决了相关技术中存在的数据库导入效率低的问题, 进而达到了提高数据库导入效率的效果。



WO 2016/169237 A1

数据处理方法及装置

技术领域

本发明涉及通信领域，具体而言，涉及一种数据处理方法及装置。

背景技术

随着科技的发展，数据库在人们的生活中起着越来越重要的作用。在当前的信息化社会中，充分有效地管理和利用各类资源，是进行科学研究和决策管理的前提条件，数据库技术是管理信息系统、办公自动化系统、决策支持系统等各类信息系统的核心部分，是进行科学研究和决策管理的重要技术手段。

传统的数据库系统一般是通过高端设备，例如小型机或者高端存储来保证数据库完整性，或者通过增加内存中央处理器（Central Processing Unit，简称为 CPU）来提高数据库处理能力。但是这种集中式的数据库架构越来越不适合海量数据库处理，数据导入效率低，而且也得付出高额的费用。

针对相关技术中存在的数据库导入效率低的问题，目前尚未提出有效的解决方案。

发明内容

本发明提供了一种数据处理方法及装置，以至少解决相关技术中存在的数据库导入效率低的问题。

根据本发明的一个方面，提供了一种数据处理方法，包括：接收用于指示将数据导入数据库的数据导入指令；根据所述数据导入指令对所述数据进行拆分处理；将拆分处理后的数据分块导入至所述数据库中不同的存储空间中。

可选地，根据所述数据导入指令对所述数据进行拆分处理包括：根据所述数据导入指令确定表的表结构和所述数据在所述表上的数据分布信息；根据所述表结构、所述数据分布信息和所述数据导入指令中携带的所述数据的描述符信息识别所述数据中每个数据行字段；根据识别的所述数据中每个数据行字段对所述数据进行拆分处理。

可选地，根据所述数据导入指令对所述数据进行拆分处理包括：判断所述数据是否满足拆分规则；在判断结果为是的情况下，对所述数据进行拆分处理；在判断结果为否的情况下，对所述数据进行修正处理；对修正处理后的数据进行拆分处理。

可选地，将拆分处理后的数据分块导入至所述数据库中不同的存储空间中包括：下载拆分处理后的数据；将下载的所述拆分处理后的数据分块导入至所述数据库中不同的存储空间中。

可选地，在将拆分处理后的数据分块导入至所述数据库中不同的存储空间中之后，还包括：删除下载的所述拆分处理后的数据。

可选地，在将拆分处理后的数据分块导入至所述数据库中不同的存储空间中之后，还包括：汇总对拆分处理后的数据进行导入处理后的导入结果；反馈所述导入结果。

根据本发明的另一方面，提供了一种数据处理装置，包括：接收模块，设置为接收用于指示将数据导入数据库的数据导入指令；处理模块，设置为根据所述数据导入指令对所述数据进行拆分处理；导入模块，设置为将拆分处理后的数据分块导入至所述数据库中不同的存储空间中。

可选地，所述处理模块包括：确定单元，设置为根据所述数据导入指令确定表的表结构和所述数据在所述表上的数据分布信息；识别单元，设置为根据所述表结构、所述数据分布信息和所述数据导入指令中携带的所述数据的描述符信息识别所述数据中每个数据行字段；第一处理单元，设置为根据识别的所述数据中每个数据行字段对所述数据进行拆分处理。

可选地，所述处理模块包括：判断单元，设置为判断所述数据是否满足拆分规则；第二处理单元，设置为在所述判断单元的判断结果为是的情况下，对所述数据进行拆分处理；修正单元，设置为在所述判断单元的判断结果为否的情况下，对所述数据进行修正处理；第三处理单元，设置为对修正处理后的数据进行拆分处理。

可选地，所述导入模块包括：下载单元，设置为下载拆分处理后的数据；导入单元，设置为将下载的所述拆分处理后的数据分块导入至所述数据库中不同的存储空间中。

可选地，所述装置还包括：删除模块，设置为删除下载的所述拆分处理后的数据。

可选地，所述装置还包括：汇总模块，设置为汇总对拆分处理后的数据进行导入处理后的导入结果；反馈模块，设置为反馈所述导入结果。

通过本发明，采用接收用于指示将数据导入数据库的数据导入指令；根据所述数据导入指令对所述数据进行拆分处理；将拆分处理后的数据分块导入至所述数据库中不同的存储空间中，解决了相关技术中存在的数据导入效率低的问题，进而达到了提高数据导入效率的效果。

附图说明

此处所说明的附图用来提供对本发明的进一步理解，构成本申请的一部分，本发明的示意性实施例及其说明用于解释本发明，并不构成对本发明的不当限定。在附图中：

图 1 是根据本发明实施例的数据处理方法的流程图；

图 2 是根据本发明实施例的数据处理装置的结构框图；

图 3 是根据本发明实施例的数据处理装置中处理模块 24 的第一种结构框图；

图 4 是根据本发明实施例的数据处理装置中处理模块 24 的第二种结构框图；

图 5 是根据本发明实施例的数据处理装置中导入模块 26 的结构框图；

图 6 是根据本发明实施例的数据处理装置的第一种优选结构框图；

图 7 是根据本发明实施例的数据处理装置的第二种优选结构框图；

图 8 是根据本发明实施例的导入系统结构框图；

图 9 是根据本发明实施例的数据导入处理流程图。

具体实施方式

下文中将参考附图并结合实施例来详细说明本发明。需要说明的是，在不冲突的情况下，本申请中的实施例及实施例中的特征可以相互组合。

需要说明的是，本发明的说明书和权利要求书及上述附图中的术语“第一”、“第二”等是用于区别类似的对象，而不必用于描述特定的顺序或先后次序。

在本实施例中提供了一种数据处理方法，图 1 是根据本发明实施例的数据处理方法的流程图，如图 1 所示，该流程包括如下步骤：

步骤 S102，接收用于指示将数据导入数据库的数据导入指令；

步骤 S104，根据数据导入指令对上述数据进行拆分处理；

步骤 S106，将拆分处理后的数据分块导入至上述数据库中不同的存储空间中。

通过上述步骤，在执行将数据导入到数据库中的处理时，首先对数据进行拆分处理，然后将拆分处理后的数据分块导入到数据库不同的存储空间中，并且对数据分块导入时，可以并行执行，提高导入效率。从而解决了相关技术中存在的数据库导入效率低的问题，进而达到了提高数据库导入效率的效果。上述的数据库可以称之为分布式数据库系统，该中数据库系统通过采用普通廉价的设备构建出高可用性和高扩展的集群，从而摆脱了大型设备的依赖。一个好的分布式数据库架构可以比较方便达到高可用性，可以达到向外扩展的能力。其中大数据量的导入导出功能是分布式数据库中较为关键的技术。

在对数据执行拆分处理时，可以有多种拆分方式，在一个可选的实施例中，可以依据用于拆分的表对上述数据进行拆分，其中，根据上述数据导入指令对数据进行拆分处理包括：根据上述数据导入指令确定表的表结构和数据在该表上的数据分布信息；根据上述表结构、数据分布信息和数据导入指令中携带的数据的描述符信息识别数据中每个数据行字段；根据识别的数据中每个数据行字段对数据进行拆分处理。其中，接收到数据导入指令后也可以首先去确定该数据导入指令的合法性，然后获取导入的目的库表的表结构信息和分布策略信息，进而读取数据文件，根据表结构信息和分布策略对待导入数据文件进行拆分，拆分成多个底层数据库（即，上述的存储空间）存储相对应的多个小文件并传送到各目的底层数据库的指

定目录下，然后通过集群管理模块下发指令到各底层数据库执行对应文件的导入。

能对数据进行拆分的前提是，该数据需要满足预定的拆分规则，但是也会存在数据不满足拆分规则的情况，该情况下，就需要对数据进行修正，以使该数据满足拆分规则。在一个可选的实施例中，根据上述数据导入指令对数据进行拆分处理包括：判断上述数据是否满足拆分规则；在判断结果为是的情况下，对数据进行拆分处理；在判断结果为否的情况下，对上述数据进行修正处理；对修正处理后的数据进行拆分处理。其中，在对数据进行修正处理时，可以有多种修正方式，可以由管理员，即人为地去进行修正；也可以在无需人工干预的情况下，由执行拆分处理的模块获取其他模块根据某些修正规则去进行修正；当然可以由人工和相应地模块相互配合去进行修正，等等。采用该方法可以及时获知需要导入数据库的数据是否能够拆分，从而进一步提高拆分效率。并且，对于需要导入的数据存在错误的情况，可以将错误行数据提取出来，保证导入数据的正确性。

在一个可选的实施例中，在将拆分处理后的数据分块导入至数据库中不同的存储空间中时，可以先下载拆分处理后的数据；再将下载的拆分处理后的数据分块导入至数据库中不同的存储空间中。

当将数据导入到数据库之后，下载的数据可以不用继续保留，在一个可选的实施例中，在将拆分处理后的数据分块导入至数据库中不同的存储空间之后，还包括：删除下载的拆分处理后的数据。从而实现清楚垃圾数据文件的目的，减少内存的占用。从而可以使得数据库存储更多的数据。

在将拆分处理后的数据分块导入至数据库中不同的存储空间之后，还可以反馈导入结果，在一个可选的实施例中，还包括：汇总对拆分处理后的数据进行导入处理后的导入结果；反馈上述导入结果。从而可以使得用户清楚的确定导入结果。

通过以上的实施方式的描述，本领域的技术人员可以清楚地了解到根据上述实施例的方法可借助软件加必需的通用硬件平台的方式来实现，当然也可以通过硬件，但很多情况下前者是更佳的实施方式。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（如ROM/RAM、磁碟、光盘）中，包括若干指令用以使得一台终端设备（可以是手机，计算机，服务器，或者网络设备等等）执行本发明各个实施例的方法。

在本实施例中还提供了一种数据处理装置，该装置用于实现上述实施例及优选实施方式，已经进行过说明的不再赘述。如以下所使用的，术语“模块”可以实现预定功能的软件和/或硬件的组合。尽管以下实施例所描述的装置较佳地以软件来实现，但是硬件，或者软件和硬件的组合的实现也是可能并被构想的。

图2是根据本发明实施例的数据处理装置的结构框图，如图2所示，该装置包括接收模块22、处理模块24和导入模块26，下面对该装置进行说明。

接收模块22，设置为接收用于指示将数据导入数据库的数据导入指令；处理模块24，连

接至上述接收模块 22，设置为根据上述数据导入指令对数据进行拆分处理；导入模块 26，连接至上述处理模块 24，设置为将拆分处理后的数据分块导入至数据库中不同的存储空间中。

图 3 是根据本发明实施例的数据处理装置中处理模块 24 的第一种结构框图，如图 3 所示，该处理模块 24 包括确定单元 32、识别模块 34 和第一处理单元 36，下面对该装置进行说明。

确定单元 32，设置为根据数据导入指令确定表的表结构和数据在表上的数据分布信息；识别单元 34，连接至上述确定单元 32，设置为根据表结构、数据分布信息和数据导入指令中携带的数据的描述符信息识别数据中每个数据行字段；第一处理单元 36，连接至上述识别单元 34，设置为根据识别的数据中每个数据行字段对数据进行拆分处理。

图 4 是根据本发明实施例的数据处理装置中处理模块 24 的第二种结构框图，如图 4 所示，该处理模块 24 包括判断单元 42、第二处理单元 44、修正单元 46 和第三处理单元 48，下面对该处理模块 24 进行说明。

判断单元 42，设置为判断上述数据是否满足拆分规则；第二处理单元 44，连接至上述判断单元，设置为在判断单元 42 的判断结果为是的情况下，对上述数据进行拆分处理；修正单元 46，连接至上述判断单元 42，设置为在判断单元 42 的判断结果为否的情况下，对上述数据进行修正处理；第三处理单元 48，连接至上述修正单元 46，设置为对修正处理后的数据进行拆分处理。

图 5 是根据本发明实施例的数据处理装置中导入模块 26 的结构框图，如图 5 所示，该导入模块 26 包括下载单元 52 和导入单元 54，下面对该装置进行说明。

下载单元 52，设置为下载拆分处理后的数据；导入单元 54，连接至上述下载单元 52，设置为将下载的拆分处理后的数据分块导入至数据库中不同的存储空间中。

图 6 是根据本发明实施例的数据处理装置的第一种优选结构框图，如图 6 所示，该装置除包括图 5 所示的所有模块外，还包括删除模块 62，下面对该装置进行说明。

删除模块 62，连接至上述导入模块 26，设置为删除下载的拆分处理后的数据。

图 7 是根据本发明实施例的数据处理装置的第二种优选结构框图，如图 7 所示，该装置除包括图 2 所示的所有模块外，还包括汇总模块 72 和反馈模块 74，下面对该装置进行说明。

汇总模块 72，连接至上述导入模块 26，设置为汇总对拆分处理后的数据进行导入处理后的导入结果；反馈模块 74，连接至上述汇总模块 72，设置为反馈上述导入结果。

下面结合具体的实施例对本发明继续进行说明。

从前述可以看出，相关技术中已有的方案都是针对传统的单个数据库进行，无需考虑表的分布结构以及系统架构，效率较低。而发明实施例中的方案是基于分布式数据库系统，满足数据库的原子性/一致性/隔离性/耐久性（Atomicity/Consistency/Isolation/Durability，简称为 ACID）特性，且可以并发执行。采用 shell 脚本进行导入导出，具有高度实时性，移植性和可行性，极高的用户体验，是对现有技术的一次重大革新。

图 8 是根据本发明实施例的导入系统结构框图，如图 8 所示，包括数据导入客户端模块 82（该模块可以位于外部系统和数据导入服务端模块之间，也可以位于外部系统中，该模块未在图 8 中画出）、数据导入服务端模块 84（对应于图 8 中的下载服务器 84，同上述的接收模块 22、处理模块 24 和导入模块 26）、元数据中心模块 86（对应于图 8 中的元数据服务器 84）、集群管理中心模块 88（对应于图 8 中的集群管理器 88，同上述的汇总模块 72 和反馈模块 74）、数据库代理模块 810（同上述的删除模块 62）和数据库模块 812，下面对各模块进行说明。

数据导入客户端模块 82 (LoadClient) 主要面向用户，用户通过该模块发起导入导出命令。

数据导入服务端模块 84 (LoadServer) 设置为接受客户端发送的导入导出命令，根据数据分布策略对数据文件进行拆分和合并，和其他模块进行交互，协调整个导入导出流程。

元数据中心模块 86 设置为保存和管理整个分布式数据库系统所有元数据信息。

集群管理中心模块 88 主要负责各个数据库集群 (DBCluster) 的监控、管理和维护。

数据库代理模块 810，该模块为数据库节点管理监控模块，它负责实时监控其所管辖的 DB 节点的运行状态是否正常，定期收集运行统计信息。

数据库模块 812 为底层模块，保存所有数据。

利用图 8 中所示的导入系统结构框图可以实现的核心算法如下：

对于导入流程：

数据导入导出服务端模块 84 根据集群 ID、数据库名和表名去元数据中心模块 86 查询该表的元数据信息，用于获取表结构定义和数据分布信息；

数据导入服务端模块 84 使用获取的上述信息（加上数据文件描述符信息）来识别数据文件 (datafilename) 中的每个数据行字段，进行数据文件拆分；

数据导入服务端模块 84 请求集群管理中心模块 88 去通知各个数据库代理模块 810 去下载所管辖 DBGroup 的拆分文件；

数据导入服务端模块 84 在各个数据库代理模块 810 下载成功之后再请求集群管理中心模块 88 去通知各个数据库代理模块 810 执行真正的加载数据文件 LOAD DATA INFILE 命令；

LOAD DATA INFILE 命令执行成功后，数据导入导出服务端模块 84 再请求集群管理中心 88 去通知各个数据库代理模块 810 删除垃圾数据文件（这里的垃圾数据文件可以是已经加载之后的下载的数据）；

数据导入服务端模块 84 汇总结果并通知数据导入导出客户端模块 82。

图 9 是根据本发明实施例的数据导入处理流程图，如图 9 所示，该流程包括如下步骤：

步骤 S902，数据导入客户端模块 82 向数据导入服务端模块 84 发送导入数据请求；

步骤 S904, 数据导入服务端模块 84 根据集群 ID、数据库名和表名向元数据中心模块 86 发送查询数据库元数据请求, 该请求用于查询该表的元数据信息;

步骤 S906, 元数据中心模块 86 返回表结构定义和数据分布信息, 包括该表各个字段类型和长度, 以及分发键和分布在哪些 DBGroup 上;

步骤 S908, 数据导入服务端模块 84 使用元数据中心模块 88 返回的信息 (此外, 再加上数据文件描述符信息) 来识别数据文件 (datafilename) 中的每个数据行字段, 进行数据文件拆分, 拆分过程中若发现数据错误, 如类型不符合表定义, 将错误数据挑选出来, 放入错误文件中;

步骤 S910, 数据导入服务端模块 84 请求集群管理中心模块 88 去通知各个数据库代理模块 810 去下载所管辖 DBGroup 的拆分文件;

步骤 S912, 集群管理中心模块 88 通知各个数据库代理模块 810 去下载所管辖 DBGroup 的拆分文件;

步骤 S914, 各个数据库代理模块 810 通知 ftp 服务连接数据导入服务端模块 84 所在服务器, 下载对应的拆分文件, 各个数据库代理模块 810 下载对应的拆分文件成功后, 向集群管理中心模块 88 返回成功响应;

步骤 S916, 集群管理中心模块 88 汇总下载结果;

步骤 S918, 集群管理中心模块 88 收到所有数据库代理模块 810 成功响应后, 向数据导入服务端模块 84 返回成功响应;

步骤 S920, 数据导入服务端模块 84 下载成功之后再请求集群管理中心模块 88 去通知各个数据库代理模块 810 执行真正的 LOAD DATA INFILE 命令;

步骤 S922, 集群管理中心模块 88 通知数据库代理模块 810 去执行真正的 LOAD DATA INFILE 命令;

步骤 S924, 各个数据库代理模块 810 连接管理的数据库模块, 执行真正的 LOAD DATA INFILE 命令; 各个数据库代理模块 810 执行真正的 LOAD DATA INFILE 命令成功后, 向集群管理中心模块 88 返回成功响应;

步骤 S926, 集群管理中心模块 88 收到所有数据库代理模块 810 成功响应后, 向数据导入服务端模块 84 返回成功响应; LOAD DATA INFILE 命令执行成功后, 数据导入服务端模块 84 再请求集群管理中心模块 88 去通知各个数据库代理模块删除垃圾数据文件; 数据导入服务端模块 84 汇总结果并通知数据导入客户端模块 82。

上述实施例中的方案是基于分布式数据库系统提出的, 可以导入 Mysql 数据库支持的所有数据类型, 当然也可以支持其他类型的数量。在分布式数据库系统上应用本发明实施例中的方案, 可以提高 2 到 3 倍的并发量, 均衡负载, 且保证导入导出数据的正确性, 系统强壮性较好。

需要说明的是，上述各个模块是可以通过软件或硬件来实现的，对于后者，可以通过以下方式实现，但不限于此：上述模块均位于同一处理器中；或者，上述模块分别位于多个处理器中。

本发明的实施例还提供了一种存储介质。可选地，在本实施例中，上述存储介质可以被设置为存储用于执行以下步骤的程序代码：

- S1，接收用于指示将数据导入数据库的数据导入指令；
- S2，根据数据导入指令对上述数据进行拆分处理；
- S3，将拆分处理后的数据分块导入至上述数据库中不同的存储空间中。

可选地，在本实施例中，上述存储介质可以包括但不限于：U 盘、只读存储器（Read-Only Memory，简称为 ROM）、随机存取存储器（Random Access Memory，简称为 RAM）、移动硬盘、磁碟或者光盘等各种可以存储程序代码的介质。

显然，本领域的技术人员应该明白，上述的本发明的各模块或各步骤可以用通用的计算装置来实现，它们可以集中在单个的计算装置上，或者分布在多个计算装置所组成的网络上，可选地，它们可以用计算装置可执行的程序代码来实现，从而，可以将它们存储在存储装置中由计算装置来执行，并且在某些情况下，可以以不同于此处的顺序执行所示出或描述的步骤，或者将它们分别制作成各个集成电路模块，或者将它们中的多个模块或步骤制作成单个集成电路模块来实现。这样，本发明不限制于任何特定的硬件和软件结合。

以上所述仅为本发明的优选实施例而已，并不用于限制本发明，对于本领域的技术人员来说，本发明可以有各种更改和变化。凡在本发明的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

工业实用性

如上所述，本发明实施例提供了一种数据处理方法及装置具有以下有益效果：解决了相关技术中存在的数据导入效率低的问题，进而达到了提高数据导入效率的效果。

权利要求书

1. 一种数据处理方法，包括：
 - 接收用于指示将数据导入数据库的数据导入指令；
 - 根据所述数据导入指令对所述数据进行拆分处理；
 - 将拆分处理后的数据分块导入至所述数据库中不同的存储空间中。
2. 根据权利要求 1 所述的方法，其中，根据所述数据导入指令对所述数据进行拆分处理包括：
 - 根据所述数据导入指令确定表的表结构和所述数据在所述表上的数据分布信息；
 - 根据所述表结构、所述数据分布信息和所述数据导入指令中携带的所述数据的描述符信息识别所述数据中每个数据行字段；
 - 根据识别的所述数据中每个数据行字段对所述数据进行拆分处理。
3. 根据权利要求 1 所述的方法，其中，根据所述数据导入指令对所述数据进行拆分处理包括：
 - 判断所述数据是否满足拆分规则；
 - 在判断结果为是的情况下，对所述数据进行拆分处理；
 - 在判断结果为否的情况下，对所述数据进行修正处理；对修正处理后的数据进行拆分处理。
4. 根据权利要求 1 所述的方法，其中，将拆分处理后的数据分块导入至所述数据库中不同的存储空间中包括：
 - 下载拆分处理后的数据；
 - 将下载的所述拆分处理后的数据分块导入至所述数据库中不同的存储空间中。
5. 根据权利要求 4 所述的方法，其中，在将拆分处理后的数据分块导入至所述数据库中不同的存储空间中之后，还包括：
 - 删除下载的所述拆分处理后的数据。
6. 根据权利要求 1 所述的方法，其中，在将拆分处理后的数据分块导入至所述数据库中不同的存储空间中之后，还包括：
 - 汇总对拆分处理后的数据进行导入处理后的导入结果；
 - 反馈所述导入结果。
7. 一种数据处理装置，包括：

接收模块，设置为接收用于指示将数据导入数据库的数据导入指令；

处理模块，设置为根据所述数据导入指令对所述数据进行拆分处理；

导入模块，设置为将拆分处理后的数据分块导入至所述数据库中不同的存储空间中。

8. 根据权利要求 7 所述的装置，其中，所述处理模块包括：

确定单元，设置为根据所述数据导入指令确定表的表结构和所述数据在所述表上的数据分布信息；

识别单元，设置为根据所述表结构、所述数据分布信息和所述数据导入指令中携带的所述数据的描述符信息识别所述数据中每个数据行字段；

第一处理单元，设置为根据识别的所述数据中每个数据行字段对所述数据进行拆分处理。

9. 根据权利要求 7 所述的装置，其中，所述处理模块包括：

判断单元，设置为判断所述数据是否满足拆分规则；

第二处理单元，设置为在所述判断单元的判断结果为是的情况下，对所述数据进行拆分处理；

修正单元，设置为在所述判断单元的判断结果为否的情况下，对所述数据进行修正处理；

第三处理单元，设置为对修正处理后的数据进行拆分处理。

10. 根据权利要求 7 所述的装置，其中，所述导入模块包括：

下载单元，设置为下载拆分处理后的数据；

导入单元，设置为将下载的所述拆分处理后的数据分块导入至所述数据库中不同的存储空间中。

11. 根据权利要求 10 所述的装置，其中，还包括：

删除模块，设置为删除下载的所述拆分处理后的数据。

12. 根据权利要求 7 所述的装置，其中，还包括：

汇总模块，设置为汇总对拆分处理后的数据进行导入处理后的导入结果；

反馈模块，设置为反馈所述导入结果。

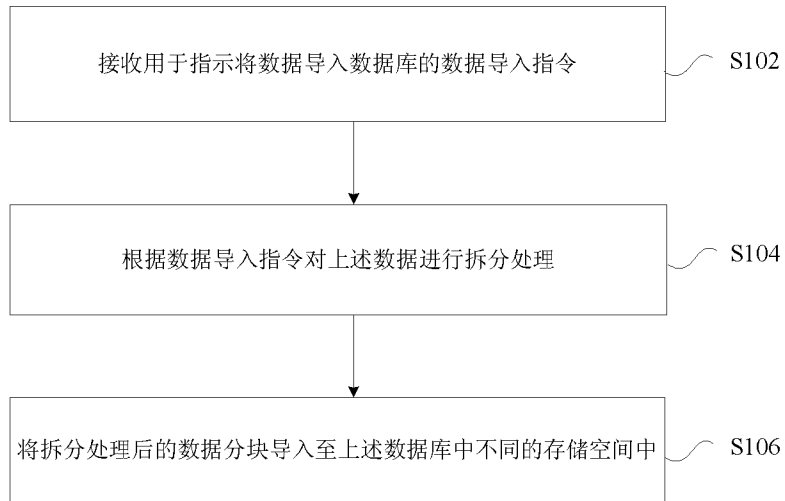


图 1



图 2

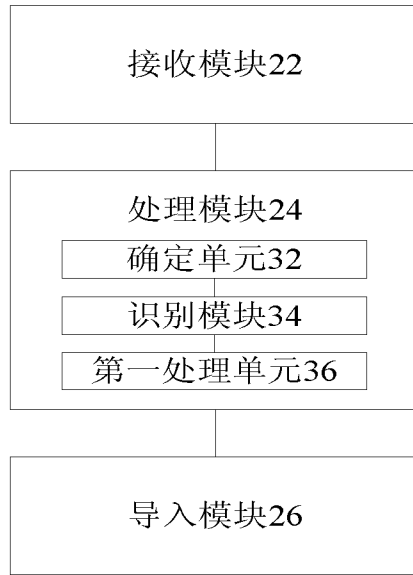


图 3



图 4



图 5

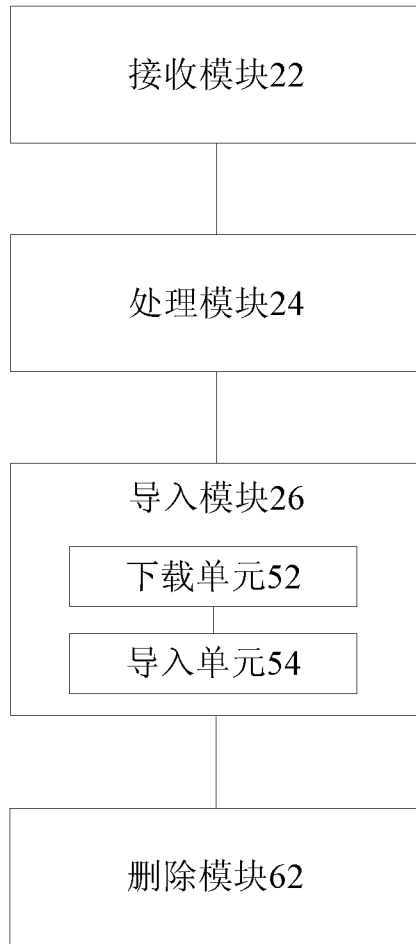


图 6



图 7

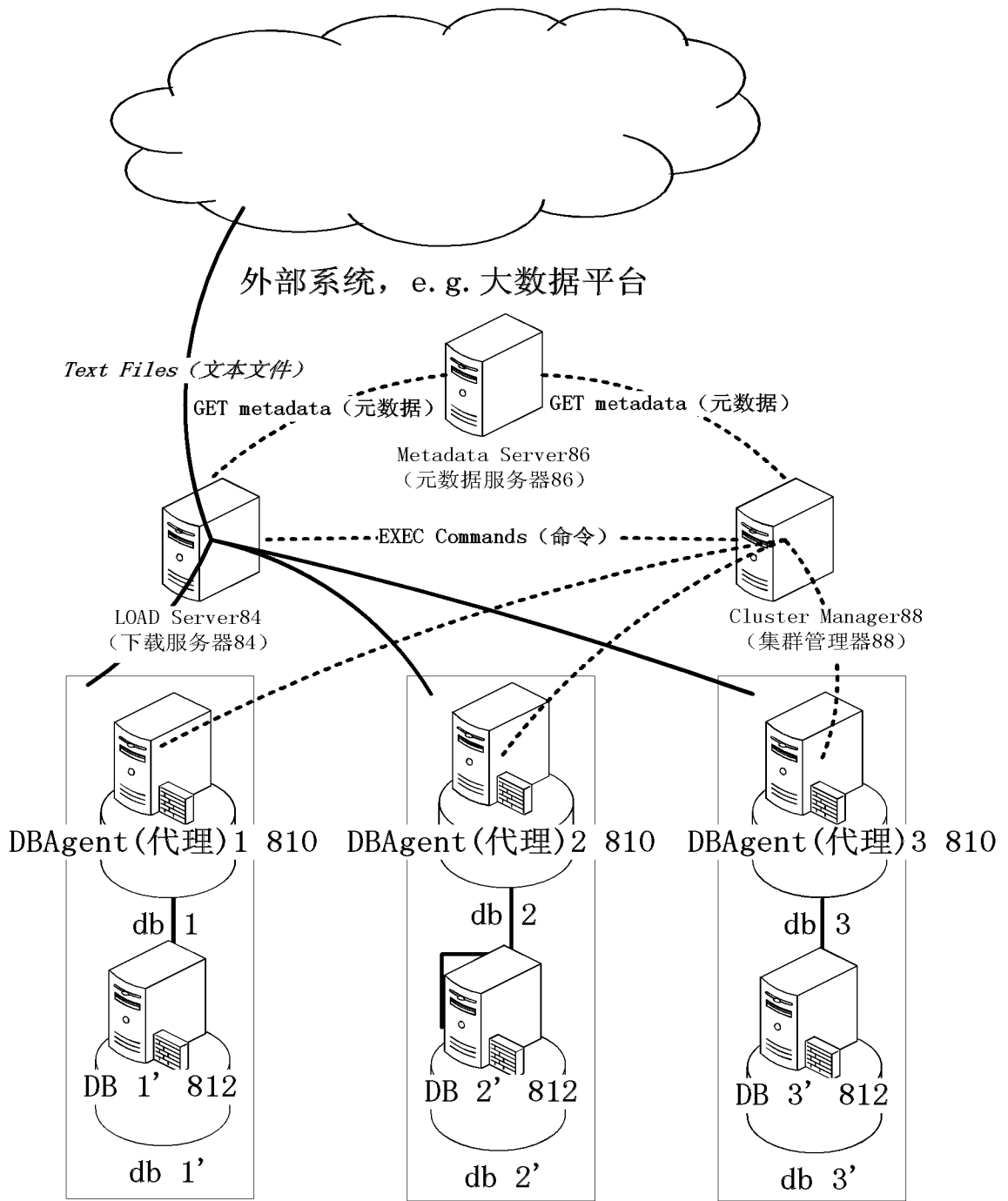


图 8

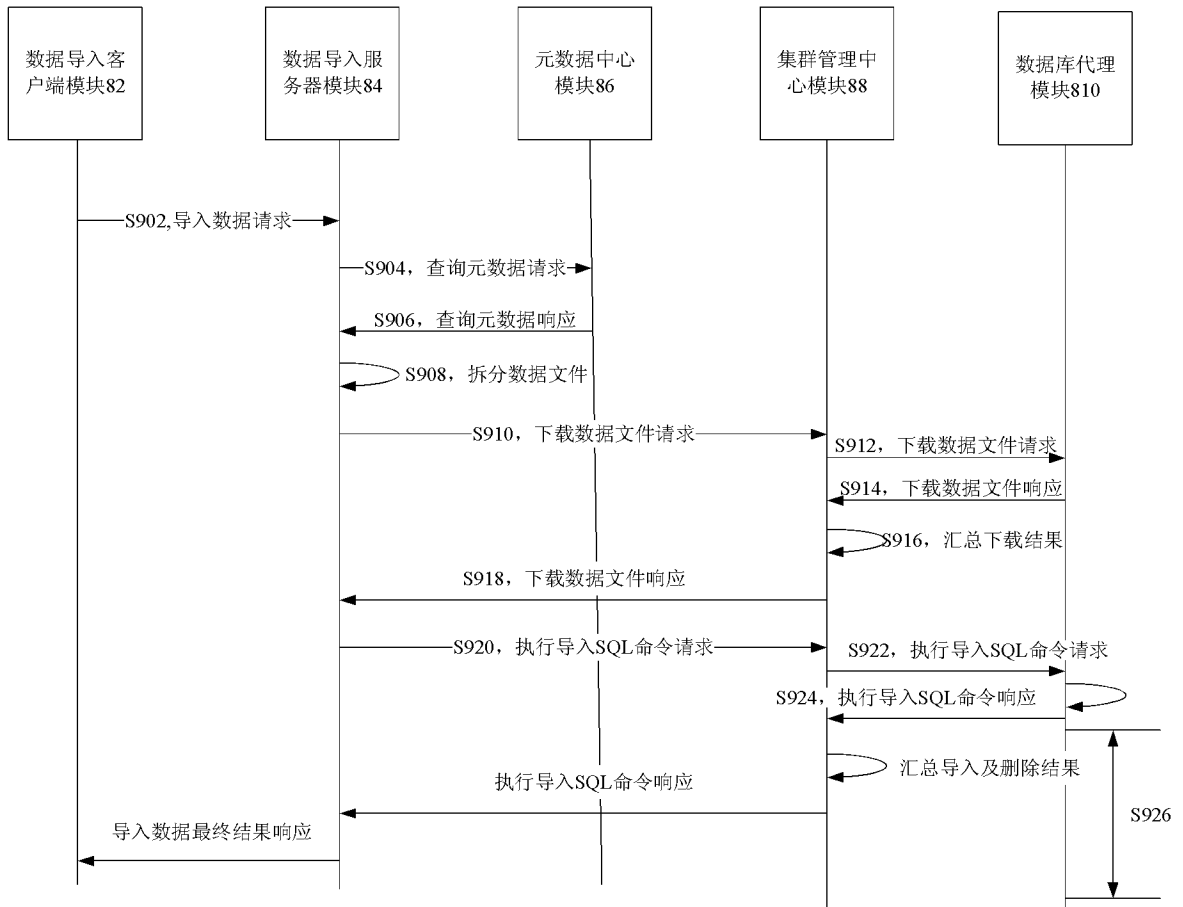


图 9

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2015/092759

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 17/-

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)

CPRSABS; CNTXT; VEN; CNKI: slice, block, segment, separate, concurrent, distributed, import+, data, divid???, group+, split+, partition, sub+, database, distribute, parallel, table

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 103077183 A (BEIJING PEZY DATA TECHNOLOGY CO., LTD.), 01 May 2013 (01.05.2013), description, paragraphs 2-89, and figures 1-5	1-12
X	CN 102750368 A (TIANJIN SHENZHOU GENERAL DATA TECHNOLOGY CO., LTD.), 24 October 2012 (24.10.2012), description, paragraphs 2-56, and figures 1-4	1-12
X	CN 103473334 A (ZHEJIANG SUPCON TECHNOLOGY CO., LTD.), 25 December 2013 (25.12.2013), description, paragraphs 103-283	1-12
A	CN 102906751 A (HUAWEI TECHNOLOGIES CO., LTD.), 30 January 2013 (30.01.2013), the whole document	1-12
A	US 2005055351 A1 (ORACLE INT CORP.), 10 March 2005 (10.03.2005), the whole document	1-12

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

06 January 2016 (06.01.2016)

Date of mailing of the international search report

20 January 2016 (20.01.2016)

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

XU, Lin

Telephone No.: (86-10) **62089885**

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2015/092759

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 103077183 A	01 May 2013	None	
CN 102750368 A	24 October 2012	CN 102750368 B	26 March 2014
CN 103473334 A	25 December 2013	None	
CN 102906751 A	30 January 2013	WO 2014015488 A1	30 January 2014
		CN 102906751 B	02 December 2015
US 2005055351 A1	10 March 2005	US 8341120 B2	25 December 2012

国际检索报告

国际申请号

PCT/CN2015/092759

<p>A. 主题的分类</p> <p>G06F 17/30(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>G06F17/-</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>CPRSABS;CNTXT;VEN;CNKI:数据, 导入, 分拆, 拆分, 分片, 分块, 切分, 分割, 分组, 划分, 分区, 分开, 数据库, 并行, 并发, 分布式, 表, import+, data, divid???, group+, split+, partition, sub+, database, distribute, parallel, table</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 103077183 A (北京普泽天玑数据技术有限公司) 2013年 5月 1日 (2013 - 05 - 01) 说明书第2-89段, 图1-5</td> <td>1-12</td> </tr> <tr> <td>X</td> <td>CN 102750368 A (天津神舟通用数据技术有限公司) 2012年 10月 24日 (2012 - 10 - 24) 说明书第2-56段, 图1-4</td> <td>1-12</td> </tr> <tr> <td>X</td> <td>CN 103473334 A (浙江中控技术股份有限公司) 2013年 12月 25日 (2013 - 12 - 25) 说明书第103-283段</td> <td>1-12</td> </tr> <tr> <td>A</td> <td>CN 102906751 A (华为技术有限公司) 2013年 1月 30日 (2013 - 01 - 30) 全文</td> <td>1-12</td> </tr> <tr> <td>A</td> <td>US 2005055351 A1 (ORACLE INT CORP) 2005年 3月 10日 (2005 - 03 - 10) 全文</td> <td>1-12</td> </tr> </tbody> </table> <p><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。</p> <p>* 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</p> <table border="1"> <tr> <td>国际检索实际完成的日期</td> <td>国际检索报告邮寄日期</td> </tr> <tr> <td>2016年 1月 6日</td> <td>2016年 1月 20日</td> </tr> <tr> <td>ISA/CN的名称和邮寄地址</td> <td>受权官员</td> </tr> <tr> <td>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</td> <td>徐琳</td> </tr> <tr> <td>传真号 (86-10)62019451</td> <td>电话号码 (86-10)62089885</td> </tr> </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 103077183 A (北京普泽天玑数据技术有限公司) 2013年 5月 1日 (2013 - 05 - 01) 说明书第2-89段, 图1-5	1-12	X	CN 102750368 A (天津神舟通用数据技术有限公司) 2012年 10月 24日 (2012 - 10 - 24) 说明书第2-56段, 图1-4	1-12	X	CN 103473334 A (浙江中控技术股份有限公司) 2013年 12月 25日 (2013 - 12 - 25) 说明书第103-283段	1-12	A	CN 102906751 A (华为技术有限公司) 2013年 1月 30日 (2013 - 01 - 30) 全文	1-12	A	US 2005055351 A1 (ORACLE INT CORP) 2005年 3月 10日 (2005 - 03 - 10) 全文	1-12	国际检索实际完成的日期	国际检索报告邮寄日期	2016年 1月 6日	2016年 1月 20日	ISA/CN的名称和邮寄地址	受权官员	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	徐琳	传真号 (86-10)62019451	电话号码 (86-10)62089885
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求																												
X	CN 103077183 A (北京普泽天玑数据技术有限公司) 2013年 5月 1日 (2013 - 05 - 01) 说明书第2-89段, 图1-5	1-12																												
X	CN 102750368 A (天津神舟通用数据技术有限公司) 2012年 10月 24日 (2012 - 10 - 24) 说明书第2-56段, 图1-4	1-12																												
X	CN 103473334 A (浙江中控技术股份有限公司) 2013年 12月 25日 (2013 - 12 - 25) 说明书第103-283段	1-12																												
A	CN 102906751 A (华为技术有限公司) 2013年 1月 30日 (2013 - 01 - 30) 全文	1-12																												
A	US 2005055351 A1 (ORACLE INT CORP) 2005年 3月 10日 (2005 - 03 - 10) 全文	1-12																												
国际检索实际完成的日期	国际检索报告邮寄日期																													
2016年 1月 6日	2016年 1月 20日																													
ISA/CN的名称和邮寄地址	受权官员																													
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	徐琳																													
传真号 (86-10)62019451	电话号码 (86-10)62089885																													

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2015/092759

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	103077183	A	2013年 5月 1日	无			
CN	102750368	A	2012年 10月 24日	CN	102750368	B	2014年 3月 26日
CN	103473334	A	2013年 12月 25日	无			
CN	102906751	A	2013年 1月 30日	WO	2014015488	A1	2014年 1月 30日
				CN	102906751	B	2015年 12月 2日
US	2005055351	A1	2005年 3月 10日	US	8341120	B2	2012年 12月 25日

表 PCT/ISA/210 (同族专利附件) (2009年7月)