

# (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 11 月 3 日 (03.11.2016)



(10) 国际公布号  
WO 2016/173428 A1

- (51) 国际专利分类号:  
G06F 17/30 (2006.01)
- (21) 国际申请号: PCT/CN2016/079649
- (22) 国际申请日: 2016 年 4 月 19 日 (19.04.2016)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:  
201510200985.7 2015 年 4 月 26 日 (26.04.2015) CN
- (71) 申请人: 阿里巴巴集团控股有限公司 (ALIBABA GROUP HOLDING LIMITED) [GB/CN]; 开曼群岛英属开曼群岛大开曼资本大厦一座四层 847 号邮箱, Cayman Islands (KY)。
- (72) 发明人: 陈守元 (CHEN, Shouyuan); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。 邓小勇 (DENG, Xiaoyong); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。 何健超 (HE, Jianchao); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。
- (74) 代理人: 北京三友知识产权代理有限公司 (BEIJING SANYOU INTELLECTUAL PROPERTY AGENCY LTD.); 中国北京市金融街 35 号国际企业大厦 A 座 16 层, Beijing 100033 (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)。

[见续页]

(54) Title: DATA TRANSMISSION METHOD AND APPARATUS

(54) 发明名称: 一种数据传输方法及装置

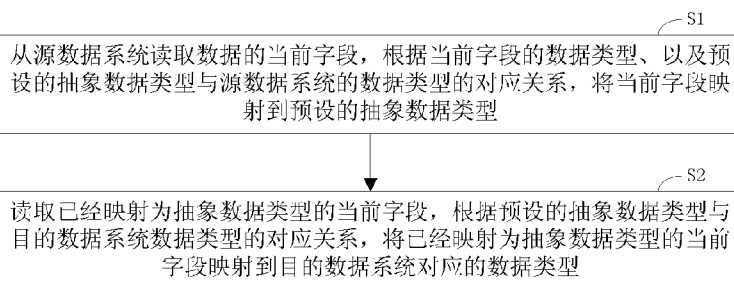


图 3

S1 READ A CURRENT FIELD OF DATA FROM A SOURCE DATA SYSTEM, AND MAP THE CURRENT FIELD TO A PRESET ABSTRACT DATA TYPE ACCORDING TO A DATA TYPE OF THE CURRENT FIELD AND A PRESET CORRESPONDENCE BETWEEN THE ABSTRACT DATA TYPE AND A DATA TYPE OF THE SOURCE DATA SYSTEM

S2 READ THE CURRENT FIELD MAPPED AS THE ABSTRACT DATA TYPE, AND MAP THE CURRENT FIELD MAPPED AS THE ABSTRACT DATA TYPE, TO A DATA TYPE CORRESPONDING TO A DESTINATION DATA SYSTEM ACCORDING TO A PRESET CORRESPONDENCE BETWEEN THE ABSTRACT DATA TYPE AND A DATA TYPE OF THE DESTINATION DATA SYSTEM

(57) Abstract: Disclosed are a data transmission method and apparatus, so as to transmit data between a source data system and a destination data system. The data transmission method comprises: reading a current field of data from a source data system, and mapping the current field to a preset abstract data type according to a data type of the current field and a preset correspondence between the abstract data type and a data type of the source data system; and reading the current field mapped as the abstract data type, and mapping the current field mapped as the abstract data type, to a data type corresponding to a destination data system according to a preset correspondence between the abstract data type and a data type of the destination data system. Correspondingly, the data transmission apparatus of the present invention comprises a read module and a write module. In the method and the apparatus of the present invention, by setting a unified abstract data type, data is transmitted among any heterogeneous data systems, and the phenomenon of data conversion distortion and the situation of data type loss do not exist.

(57) 摘要:

[见续页]

WO 2016/173428 A1

**本国际公布:**

— 包括国际检索报告(条约第 21 条(3))。

---

本发明公开了一种数据传输方法及装置，用于在源数据系统与目的数据系统间传输数据，所述数据传输方法从源数据系统读取数据的当前字段，根据当前字段的数据类型、以及预设的抽象数据类型与源数据系统的数据类型的对应关系，将当前字段映射到预设的抽象数据类型；读取已经映射为抽象数据类型的当前字段，根据预设的抽象数据类型与目的数据系统数据类型的对应关系，将已经映射为抽象数据类型的当前字段映射到目的数据系统对应的数据类型。对应地，本发明的数据传输装置包括读模块和写模块。本发明的方法及装置通过设定统一的抽象数据类型，完成任意异构数据系统之间数据的传输，同时不存在数据转换失真现象和数据类型丢失的情况。

## 一种数据传输方法及装置

本申请要求 2015 年 04 月 26 日递交的申请号为 201510200985.7、发明名称为“一种数据传输方法及装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

### 5 技术领域

本发明属于数据库技术领域，尤其涉及异构数据系统之间数据传输方法及装置。

### 背景技术

在当前大数据时代，越来越多的企业与公司逐渐意识到对大数据的获取、分析、挖掘可以为后期商务决策、市场预估、公司发展等提供极具价值的参考模型。大数据能够发挥的价值基于数据量之大、数据入口之大、数据采集规模之大，能提供充分的数据依据为企业发展服务。然而大数据需要解决的首要问题在于将大量用户、系统生成的数据进行大集中，只有数据大集中才能产生大数据。如图 1 所示，各类在线业务系统（用户中心、商品中心、交易中心）下生成的海量数据存放在各类异构数据系统之中，数据同步需要将各类数据源生成的数据进行传输、同步、汇总到统一的大数据存储/计算体系之中(例如 Hadoop、或者阿里集团的 ODPS)。

因此数据同步将直接面临各类数据类型不一、存取接口不一、事务模型不一、流量控制模型不一，等各类问题。同时，随着对接的数据存储日益增长，数据同步将面临几何级的数据传输通道增加，给后期数据同步能力扩展造成了巨大的开发、运维、技术支持成本。

20 对于一个复杂的数据同步网络，如图 2 所示，在网络中存在很多不同的数据源时，如同时存在 Hadoop、Mysql、Oracle、Sqlserver、ODPS 等数据源，将这些数据源进行数据同步比较困难。Sqoop 是 Hadoop 开源社区提供的基于 Hadoop MapReduce 运行框架，将 Hadoop 和 RDBMS 间数据同步的工具，Sqoop 在解决异构数据传输过程中，仅仅面对 Hadoop/HDFS 和 RDBMS 两类存储系统的交互，使用了最简单的 CSV 文本作为传输交换协议。以 Mysql 传输数据到 HDFS 为例：

25 用户使用 Sqoop Client 启动 Sqoop，Sqoop 在 Hadoop 上启动 MapReduce 的 Job，该 Job 实际运行的是 Sqoop 的数据传输任务。Sqoop 在 MapReduce 的 Map 阶段启动数据通道，将 Mysql 数据进行抽取同时转为 CSV 格式作为中间格式。Sqoop 将上述 CSV 格式数据直接写入 Hadoop/HDFS 文件系统中。然而该方法仅仅支持 Hadoop 和 RDBMS 间数

据同步，存在如下的技术问题：

a. 数据传输易失真：以开源社区的 Sqoop 为例，数据全部使用 CSV 文本格式进行保存，存在将其他非 String 类型转为 String 类型的过程，该过程容易导致数据转换失真。

5 b. 部分类型无法支持：以开源社区的 Sqoop 为例，由于文本化的 CSV 格式决定了不能转换为 String 格式的数据必定是无法支持的，导致同步产品功能存在缺陷，例如二进制数据无法通过该方式进行数据传输。

c. 数据类型丢失：一旦其他类型转为 String，类型完全丢失，对于类型存在强要求的传输领域无法适配，例如数据同步用户要求源端为整形数据，到目标端一定全部为整形数据，否则失败。但由于类型丢失，该需求无法完成。

10 d. 数据类型转换行为未定义，使用各个数据源的转换方式，没有统一的转换行为定义，导致数据同步对接不同数据系统使用不同的转换方式，转换行为随数据系统变化而变化，甚至随数据系统特定版本变化而变化，导致数据系统升级易导致数据同步数据变化诱发数据质量问题。

### 发明内容

15 本发明的目的就是提供一种数据传输方法及装置，解决现有技术仅支持特定的数据系统间的数据传输，并且存在数据失真、数据类型丢失等技术问题。

为了实现上述目的，本发明技术方案如下：

一种数据传输方法，用于在源数据系统与目的数据系统间传输数据，其特征在于，所述数据传输方法包括步骤：

20 从源数据系统读取数据的当前字段，根据当前字段的数据类型、以及预设的抽象数据类型与源数据系统的数据类型的对应关系，将当前字段映射到预设的抽象数据类型；

读取已经映射为抽象数据类型的当前字段，根据预设的抽象数据类型与目的数据系统数据类型的对应关系，将已经映射为抽象数据类型的当前字段映射到目的数据系统对应的数据类型。

25 其中，所述预设的抽象数据类型包括 Long、Double、String、Date、Bool 和 Bytes，对应不同的源数据系统和目的数据系统，分别预设有抽象数据类型与源数据系统的数据类型的映射关系表、抽象数据类型与目的数据系统的数据类型的映射关系表；

进一步地，所述根据当前字段的数据类型、以及预设的抽象数据类型与源数据系统的数据类型的对应关系，将当前字段映射到预设的抽象数据类型，包括步骤：

30 判断当前字段的数据类型，查找抽象数据类型与源数据系统的数据类型的映射关系

表，找到直接对应的抽象数据类型；

将当前字段的数据类型映射到所找到的直接对应的抽象数据类型；

同理，所述根据预设的抽象数据类型与目的数据系统数据类型的对应关系，将已经映射为抽象数据类型的当前字段映射到目的数据系统对应的数据类型，包括步骤：

- 5       根据当前字段的抽象数据类型，查找抽象数据类型与目的数据系统的数据类型的映射关系表，找到直接对应的目的数据系统的数据类型；

将当前字段的抽象数据类型映射到直接对应的目的数据系统的数据类型。

本发明所述 Long 支持无限长度，所述 Double 支持无限精度，所述 String 支持无限字符长度。

- 10       进一步地，所述数据传输方法，还包括步骤：

当读取的抽象数据类型当前字段的数据类型在目的数据系统中没有可以直接映射的对应数据类型时，将当前字段对应的抽象数据类型转换为与目的数据系统数据类型有直接映射关系的抽象数据类型。

- 15       通过本步骤可以有效降低抽象数据类型与源数据系统和目的数据系统数据类型之间映射关系表的复杂度。

进一步地，所述将当前字段对应的抽象数据类型转换为与目的数据系统数据类型有直接映射关系的抽象数据类型，具体转换规则包括：

- 1) 、Long 转换为 Date，将 Long 数值视作 Unix 时间戳转换为 Date，精度为毫秒；  
2) 、Long 转换为 Bool，0 视为 false，否则为 true；  
20       3) 、String 转换为 Long，整数型字符串支持转换为 Long，否则报错，不支持科学计数法字符串；  
4) 、String 转换为 Double，浮点型字符串支持转换为 Long，否则报错，不支持科学计数法；  
5) 、String 转换为 Date，根据自定义时间格式满足用户时间格式时转换，否则报错；  
25       6) 、String 转换为 Bool，“true” 转为 true，“false” 转为 false，其他报错；  
7) 、Date 转为 Long，将 Date 转为 Unix 时间戳；  
8) 、Date 转换为 String，根据自定义时间格式转换为字符串；  
9) 、Bool 转换为 Long，true 转为 1，false 转为 0；  
10) 、Bytes 转换为 String，根据用户指定的字符集编码转为 String。

- 30       本发明还提出了一种数据传输装置，用于在源数据系统与目的数据系统间传输数据，

所述数据传输装置包括：

读模块，用于从源数据系统读取数据的当前字段，根据当前字段的数据类型、以及预设的抽象数据类型与源数据系统的数据类型的对应关系，将当前字段映射到预设的抽象数据类型；

5 写模块，用于读取已经映射为抽象数据类型的当前字段，根据预设的抽象数据类型与目的数据系统数据类型的对应关系，将已经映射为抽象数据类型的当前字段映射到目的数据系统对应的数据类型。

进一步地，所述数据传输装置还设置有配置模块，用于预设抽象数据类型，所述预设的抽象数据类型包括 Long、Double、String、Date、Bool 和 Bytes，对应不同的源数据  
10 系统和目的数据系统，分别预设抽象数据类型与源数据系统的数据类型的映射关系表、抽象数据类型与目的数据系统的数据类型的映射关系表；

则，所述读模块在将当前字段映射到预设的抽象数据类型时，执行如下操作：

判断当前字段的数据类型，查找抽象数据类型与源数据系统的数据类型的映射关系表，找到直接对应的抽象数据类型；

15 将当前字段的数据类型映射到所找到的直接对应的抽象数据类型；

同理，所述写模块在将已经映射为抽象数据类型的当前字段映射到目的数据系统对应的数据类型，执行如下操作：

根据当前字段的抽象数据类型，查找抽象数据类型与目的数据系统的数据类型的映射关系表，找到直接对应的目的数据系统的数据类型；

20 将当前字段的抽象数据类型映射到直接对应的目的数据系统的数据类型。

进一步地，所述写模块还用于在读取的抽象数据类型当前字段的数据类型在目的数据系统中没有可以直接映射的对应数据类型时，将当前字段对应的抽象数据类型转换为与目的数据系统数据类型有直接映射关系的抽象数据类型。

本发明提出的数据传输方法及装置，通过设定统一的抽象数据类型，完成任意异构  
25 数据系统之间数据的传输，同时不存在数据转换失真现象；采用了最大精度/长度保存中间数据，避免数据精度丢失情况；支持二进制类型数据传输，其他更多类型也可以通过扩展抽象数据类型来支持，即使出现新开发的数据库也能同样很好的支持；并且由于使用抽象数据类型来进行传输，避免数据类型丢失情况。

#### 附图说明

30 图 1 为大数据存储/计算系统数据传输示意图；

图 2 为复杂的数据传输网络示意图；  
图 3 为本发明数据系统之间数据传输方法流程图；  
图 4 为本发明不同数据系统间数据传输示意图；  
图 5 为本发明数据传输装置结构示意图。

5 具体实施方式

下面结合附图和实施例对本发明技术方案做进一步详细说明，以下实施例不构成对本发明的限定。

如图 3 所示，本发明实施例一种数据传输方法包括如下步骤：

步骤 S1、从源数据系统读取数据的当前字段，根据当前字段的数据类型、以及预设的抽象数据类型与源数据系统的数据类型的对应关系，将当前字段映射到预设的抽象数据类型。

本实施例预设了能适配绝大部分主流数据类型的抽象数据类型，具体如表 1 所示：

表 1 本实施例预设的抽象数据类型

| 抽象数据类型        | 备注说明                                      |
|---------------|-------------------------------------------|
| Long（整形类型）    | 无限长度，支持所有定点数，包括 Int、Short、Long、BigInteger |
| Double（浮点类型）  | 无限精度，支持所有浮点数，包括 Float、Double、BigDecimal   |
| String（字符串类型） | 无限字符长度，支持字符串类型，使用通用字符集(Unicode)           |
| Date（时间类型）    | 支持时间类型，包括日期、时间、日期时间类型                     |
| Bool（布尔类型）    | 支持布尔值                                     |
| Bytes（字节类型）   | 支持二进制，例如存放诸如 MP3 等非结构化数据                  |

如表 1 所示，本实施例预设的抽象数据类型覆盖范围广，能够对接绝大部分数据系统支持的主流数据类型，从而能够将不同数据系统的数据类型映射到预设的抽象数据类型上，或从预设的抽象数据类型转换为目的数据系统需要的数据类型。

表 2 Mysql 各种数据类型与本实施例抽象数据类型的映射关系

| 抽象数据类型 | Mysql 插件对应类型                                              |
|--------|-----------------------------------------------------------|
| Long   | int, tinyint, smallint, mediumint, int, bigint            |
| Double | float, double, decimal                                    |
| String | varchar, char, tinytext, text, mediumtext, longtext, year |
| Date   | date, datetime, timestamp, time                           |
| Bool   | bit, bool                                                 |
| Bytes  | tinyblob, mediumblob, blob, longblob, varbinary           |

以 Mysql 为例，具有如表 2 所示的列表，详细设置了 Mysql 各种数据类型与本实施

例抽象数据类型的映射关系。

对于从 Mysql 抽取数据的逻辑，该逻辑的过程如下：

读取源数据系统的当前字段，判断该字段的数据类型，如果是 int、tinyint、smallint、mediumint、int、bigint 等数据类型，就转换为抽象数据类型 Long，如果是 float、double、decimal 等数据类型，就转换为抽象数据类型 Double，其他根据表 2 以此类推。

容易理解的是，对于 Hadoop、Oracle、Sqlserver、ODPS 等数据系统，同样设置有如表 2 类似的列表，便于根据设置的列表完成源数据系统与本实施例抽象数据类型的转换，或完成本实施例抽象数据类型与目的数据系统数据类型的转换。在本实施例中不一一列举其他数据系统数据类型与本实施例抽象数据类型的对应关系。

进一步地，表 1 也只预设了目前绝大多数主流数据系统数据类型与本实施例抽象数据类型的映射关系，本技术领域人员应该知道，对于未列出的映射关系，可以在实际的应用中进行扩展设置。

例如其他更多类型可以通过扩展抽象数据类型来支持，即使出现新开发的数据库也能同样很好的支持，例如为支持一种新数据系统新定义的 JSON 字段类型，同样扩展抽象数据类型的种类，即可实现 JSON 类型对接。

步骤 S2、读取已经映射为抽象数据类型的当前字段，根据预设的抽象数据类型与目的数据系统数据类型的对应关系，将已经映射为抽象数据类型的当前字段映射到目的数据系统对应的数据类型。

从源数据系统读取的当前字段映射为本实施例统一的抽象数据类型后，可以直接存储在内存之中，便于进行读取并转换为目的数据系统对应的数据类型，也可以存储在专门的存储设备中，本发明并不限于具体的实现形式。

而将当前字段写入到目的数据系统中时，抽象数据类型与目的数据系统的数据类型对应关系同样如表 1 所示，如果目的数据系统是 Mysql，则采用表 2 的对应关系进行转换，表 2 是表 1 的一种具体实施例。

容易理解的是，所有对接的数据系统的数据抽取、写入均使用表 1 的数据模型进行传递数据，将不同的数据源支持的不同数据类型统一抽象到同一套数据模型中，保证所有的数据系统使用同一套协议进行通信，使得任意异构数据存储均可以实现互联互通。图 4 示出了几个不同数据系统间按照本发明数据传输方法进行传输的示意图。

通过上述步骤就实现了抽取数据到写入数据的数据传输阶段全部使用抽象数据类型进行传输。同时为避免类似 CSV 格式数据类型传输过程丢失情况，本实施例抽象数据类

型采用了强数据类型(数据类型之间建立对应关系),而非类似 CSV 的弱数据类型,不会导致数据失真。而现有技术对于数据传输处理各个数据类型的转换交由各个数据系统自行处理,以 Sqoop 为例, Sqoop 将数据全部使用文本化的 CSV 格式,这样在数据传输到目的数据系统过程将文本化的数据转换到目的数据格式依靠目的数据系统自行转换,转换过程不统一且不可靠。

由于现有技术中存在非常多种类的数据系统,各公司的数据系统数据类型不一,有些源数据系统的数据类型在目的数据系统中没有同类的数据类型。例如:源数据系统 Mysql 中的 Date,在目的数据系统如阿里巴巴的 OTS 中就没有同类的数据类型,但是 OTS 中有 bigint,用户需要用 bigint 来与 Mysql 中的 Date 数据类型对应。

为了解决上述问题,本实施例采用先将 Mysql 中的 Date 映射为抽象数据类型的 Date,然后将抽象数据类型 Date 的数据转换为抽象数据类型 Long,而抽象数据类型 Long 与 bigint 存在直接的映射关系,这样就很方便地将抽象数据类型 Long 映射到目的数据系统 OTS 中的 bigint。

这样做的好处是,在设定目的数据系统数据类型与抽象数据类型的映射关系时,只需要设定 Long 与 bigint 的映射关系,而不要设置 Date 与 bigint 的映射关系。否则每一种抽象数据类型都对应非常复杂的其他数据系统的数据类型,将导致数据类型间的映射关系混乱,表 2 的内容将非常复杂。

因此,本实施例的数据传输方法还包括步骤:

当读取的抽象数据类型当前字段的抽象数据类型在目的数据系统中没有可以直接映射的对应数据类型时,将当前字段对应的抽象数据类型转换为与目的数据系统数据类型有直接映射关系的抽象数据类型。

即将抽象数据类型 Date 向转换为抽象数据类型 Long,由于抽象数据类型 Long 与目的数据系统的 bigint 有直接映射关系,从而在写入目的数据系统时,可以从抽象数据类型 Long 映射到 bigint。

本实施例针对不同抽象数据类型之间的相互转换,内置不同的规则,避免跨类型数据转换行为未定义。表 3 示出了一种典型的抽象数据类型之间的转换关系:

其中, Y 表示自动类型转换或同类型转换, N 表示不能跨类型转换,数字代表有条件类型转换:

[1] Long 转换为 Date,将 Long 数值视作 Unix 时间戳转换为 Date,精度为毫秒。

[2] Long 转换为 Bool, 0 视为 false, 否则为 true。

[3] String 转换为 Long，整数型字符串支持转换为 Long，否则报错。不支持科学计数法字符串。

[4] String 转换为 Double，浮点型字符串支持转换为 Long，否则报错。不支持科学计数法。

5 [5] String 转换为 Date，用户自定义时间格式(例如 YYYY-MM-DD)，满足用户时间格式可以转换，否则报错。

[6] String 转换为 Bool，“true”转为 true，“false”转为 false，其他报错。

[7] Date 转为 Long，将 Date 转为 Unix 时间戳。

10 [8] Date 转换为 String，根据用户自定义时间格式(例如 YYYY-MM-DD)转换为字符串。

[9] Bool 转换为 Long，true 转为 1，false 转为 0。

[10] Bytes 转换为 String，根据用户指定的字符集编码转为 String。

表 3 一种典型的抽象数据类型之间的转换关系

| 类型<br>转换 | Long | Double | String | Date | Bool | Byte |
|----------|------|--------|--------|------|------|------|
| Long     | Y    | Y      | Y      | [1]  | [2]  | N    |
| Double   | Y    | Y      | Y      | N    | N    | N    |
| String   | [3]  | [4]    | Y      | [5]  | [6]  | N    |
| Date     | [7]  | N      | [8]    | Y    | N    | N    |
| Bool     | [9]  | N      | Y      | N    | Y    | N    |
| Bytes    | N    | N      | [10]   | N    | N    | Y    |

从而无论源数据系统、目的数据系统是何种数据库，都能根据该数据库中字段的数据类型进行相应的映射，转换为统一的数据类型体系。由于数据没有发生跨类型数据转换，不存在数据转换失真现象；同时统一的数据类型体系采用了最大精度/长度保存中间数据，避免数据精度丢失情况；本实施例预设的抽象数据类型还支持二进制类型数据传输。并且由于使用抽象数据类型来进行传输，例如整形数据同样使用整形数据类型进行传输，避免数据类型丢失情况。

20 本实施例还提出了基于上述传输方法的装置，如图 5 所示，包括：

读模块，用于从源数据系统读取数据的当前字段，根据当前字段的数据类型、以及预设的抽象数据类型与源数据系统的数据类型的对应关系，将当前字段映射到预设的抽象数据类型；

写模块，用于读取已经映射为抽象数据类型的当前字段，根据预设的抽象数据类型与目的数据系统数据类型的对应关系，将已经映射为抽象数据类型的当前字段映射到目的数据系统对应的数据类型。

本实施例中，对应不同的源数据系统和目的数据系统，分别预设有抽象数据类型与源数据系统的数据类型的映射关系表、抽象数据类型与目的数据系统的数据类型的映射关系表。映射关系表可以预先设置好，直接存储在读模块和写模块中，也可以另外设置专门的配置模块，用于预设抽象数据类型及映射关系表并存储映射关系表。专门设置配置模块便于进行扩展，根据具体的数据系统进行调整，实现起来更加灵活。

本实施例读模块在将当前字段映射到预设的抽象数据类型时，执行如下操作：

10 判断当前字段的数据类型，查找抽象数据类型与源数据系统的数据类型的映射关系表，找到直接对应的抽象数据类型；

将当前字段的数据类型映射到所找到的直接对应的抽象数据类型；

同理，写模块在将已经映射为抽象数据类型的当前字段映射到目的数据系统对应的数据类型，执行如下操作：

15 根据当前字段的抽象数据类型，查找抽象数据类型与目的数据系统的数据类型的映射关系表，找到直接对应的目的数据系统的数据类型；

将当前字段的抽象数据类型映射到直接对应的目的数据系统的数据类型。

进一步地，写模块还用于在读取的抽象数据类型当前字段的数据类型在目的数据系统中没有可以直接映射的对应数据类型时，将当前字段对应的抽象数据类型转换为与目的数据系统数据类型有直接映射关系的抽象数据类型。

20

以上实施例仅用以说明本发明的技术方案而非对其进行限制，在不背离本发明精神及其实质的情况下，熟悉本领域的技术人员当可根据本发明作出各种相应的改变和变形，但这些相应的改变和变形都应属于本发明所附的权利要求的保护范围。

## 权 利 要 求 书

1、一种数据传输方法，用于在源数据系统与目的数据系统间传输数据，其特征在于，所述数据传输方法包括步骤：

从源数据系统读取数据的当前字段，根据当前字段的数据类型、以及预设的抽象数据类型与源数据系统的数据类型的对应关系，将当前字段映射到预设的抽象数据类型；

读取已经映射为抽象数据类型的当前字段，根据预设的抽象数据类型与目的数据系统数据类型的对应关系，将已经映射为抽象数据类型的当前字段映射到目的数据系统对应的数据类型。

2、根据权利要求 1 所述的数据传输方法，其特征在于，所述预设的抽象数据类型包括 Long、Double、String、Date、Bool 和 Bytes，对应不同的源数据系统和目的数据系统，分别预设抽象数据类型与源数据系统的数据类型的映射关系表、抽象数据类型与目的数据系统的数据类型的映射关系表；

所述根据当前字段的数据类型、以及预设的抽象数据类型与源数据系统的数据类型的对应关系，将当前字段映射到预设的抽象数据类型，包括步骤：

判断当前字段的数据类型，查找抽象数据类型与源数据系统的数据类型的映射关系表，找到直接对应的抽象数据类型；

将当前字段的数据类型映射到所找到的直接对应的抽象数据类型；

所述根据预设的抽象数据类型与目的数据系统数据类型的对应关系，将已经映射为抽象数据类型的当前字段映射到目的数据系统对应的数据类型，包括步骤：

根据当前字段的抽象数据类型，查找抽象数据类型与目的数据系统的数据类型的映射关系表，找到直接对应的目的数据系统的数据类型；

将当前字段的抽象数据类型映射到直接对应的目的数据系统的数据类型。

3、根据权利要求 2 所述的数据传输方法，其特征在于，所述 Long 支持无限长度，所述 Double 支持无限精度，所述 String 支持无限字符长度。

4、根据权利要求 2 所述的数据传输方法，其特征在于，所述数据传输方法，还包括步骤：

当读取的抽象数据类型当前字段的数据类型在目的数据系统中没有可以直接映射的对应数据类型时，将当前字段对应的抽象数据类型转换为与目的数据系统数据类型有直接映射关系的抽象数据类型。

5、根据权利要求 4 所述的数据传输方法，其特征在于，所述将当前字段对应的抽

象数据类型转换为与目的数据系统数据类型有直接映射关系的抽象数据类型，具体转换规则包括：

1)、Long 转换为 Date，将 Long 数值视作 Unix 时间戳转换为 Date，精度为毫秒；

2)、Long 转换为 Bool，0 视为 false，否则为 true；

5 3)、String 转换为 Long，整数型字符串支持转换为 Long，否则报错，不支持科学计数法字符串；

4)、String 转换为 Double，浮点型字符串支持转换为 Long，否则报错，不支持科学计数法；

10 5)、String 转换为 Date，根据自定义时间格式满足用户时间格式时转换，否则报错；

6)、String 转换为 Bool，“true”转为 true，“false”转为 false，其他报错；

7)、Date 转为 Long，将 Date 转为 Unix 时间戳；

8)、Date 转换为 String，根据自定义时间格式转换为字符串；

9)、Bool 转换为 Long，true 转为 1，false 转为 0；

15 10)、Bytes 转换为 String，根据用户指定的字符集编码转为 String。

6、一种数据传输装置，用于在源数据系统与目的数据系统间传输数据，其特征在于，所述数据传输装置包括：

20 读模块，用于从源数据系统读取数据的当前字段，根据当前字段的数据类型、以及预设的抽象数据类型与源数据系统的数据类型的对应关系，将当前字段映射到预设的抽象数据类型；

写模块，用于读取已经映射为抽象数据类型的当前字段，根据预设的抽象数据类型与目的数据系统数据类型的对应关系，将已经映射为抽象数据类型的当前字段映射到目的数据系统对应的数据类型。

25 7、根据权利要求 6 所述的数据传输装置，其特征在于，所述数据传输装置还设置有配置模块，用于预设抽象数据类型，所述预设的抽象数据类型包括 Long、Double、String、Date、Bool 和 Bytes，对应不同的源数据系统和目的数据系统，分别预设有抽象数据类型与源数据系统的数据类型的映射关系表、抽象数据类型与目的数据系统的数据类型的映射关系表；

所述读模块在将当前字段映射到预设的抽象数据类型时，执行如下操作：

30 判断当前字段的数据类型，查找抽象数据类型与源数据系统的数据类型的映射关系

表，找到直接对应的抽象数据类型；

将当前字段的数据类型映射到所找到的直接对应的抽象数据类型；

所述写模块在将已经映射为抽象数据类型的当前字段映射到目的数据系统对应的数据类型，执行如下操作：

- 5       根据当前字段的抽象数据类型，查找抽象数据类型与目的数据系统的数据类型的映射关系表，找到直接对应的目的数据系统的数据类型；

将当前字段的抽象数据类型映射到直接对应的目的数据系统的数据类型。

8、根据权利要求 7 所述的数据传输装置，其特征在于，所述 Long 支持无限长度，所述 Double 支持无限精度，所述 String 支持无限字符长度。

- 10       9、根据权利要求 7 所述的数据传输装置，其特征在于，所述写模块还用于在读取的抽象数据类型当前字段的数据类型在目的数据系统中没有可以直接映射的对应数据类型时，将当前字段对应的抽象数据类型转换为与目的数据系统数据类型有直接映射关系的抽象数据类型。

- 15       10、根据权利要求 9 所述的数据传输装置，其特征在于，所述将当前字段对应的抽象数据类型转换为与目的数据系统数据类型有直接映射关系的抽象数据类型，具体转换规则包括：

- 1) 、Long 转换为 Date，将 Long 数值视作 Unix 时间戳转换为 Date，精度为毫秒；
- 2) 、Long 转换为 Bool，0 视为 false，否则为 true；
- 3) 、String 转换为 Long，整数型字符串支持转换为 Long，否则报错，不支持科学
- 20   计数法字符串；
- 4) 、String 转换为 Double，浮点型字符串支持转换为 Long，否则报错，不支持科学计数法；
- 5) 、String 转换为 Date，根据自定义时间格式满足用户时间格式时转换，否则报错；
- 25   6) 、String 转换为 Bool，“true”转为 true，“false”转为 false，其他报错；
- 7) 、Date 转为 Long，将 Date 转为 Unix 时间戳；
- 8) 、Date 转换为 String，根据自定义时间格式转换为字符串；
- 9) 、Bool 转换为 Long，true 转为 1，false 转为 0；
- 10) 、Bytes 转换为 String，根据用户指定的字符集编码转为 String。

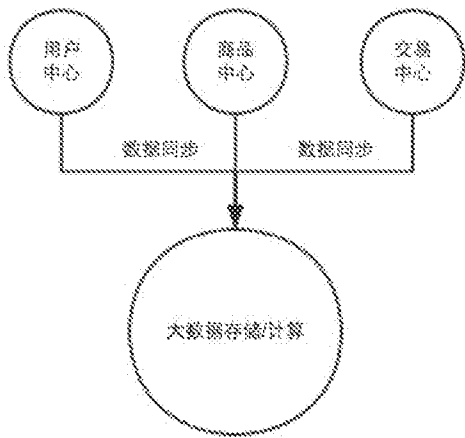


图 1

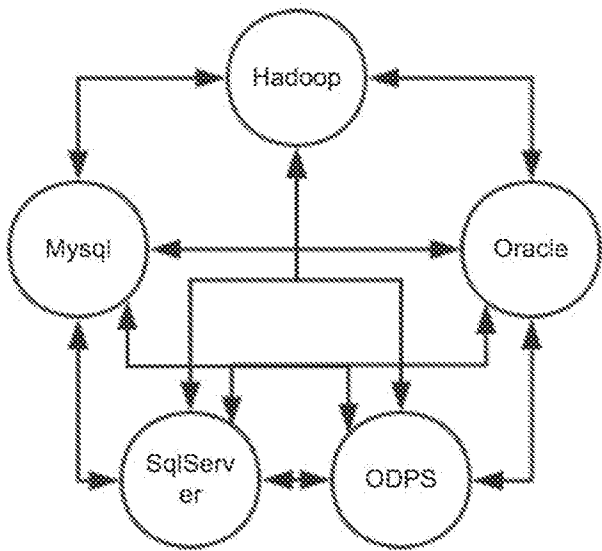


图 2

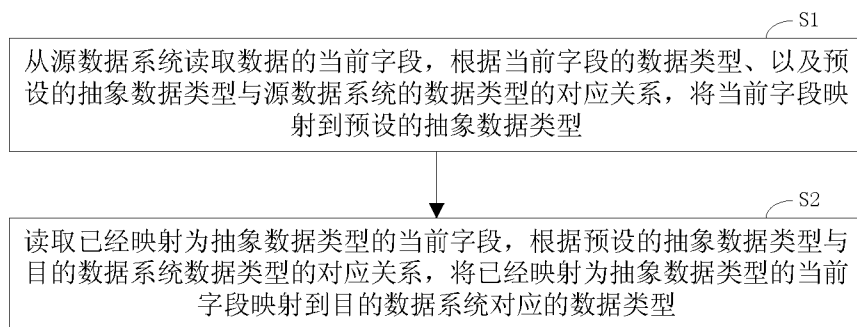


图 3

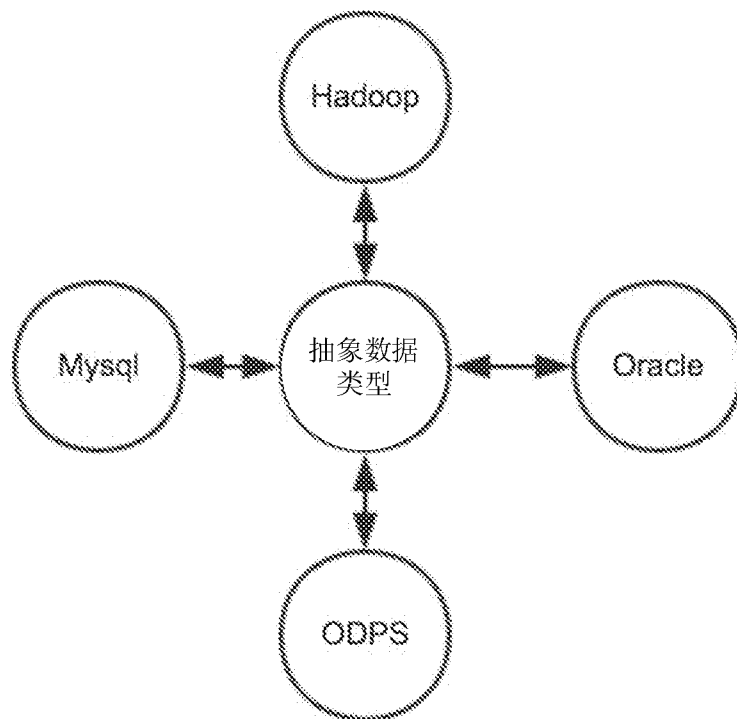


图 4

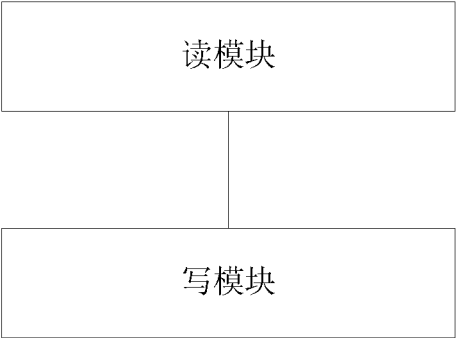


图 5

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/CN2016/079649

## 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 17/30; G06F 9

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

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

CPRSABS, CNTXT, DWPI: heterogeneous, database?, type, source, target, convert+, transform+, mid+, inter+, abstract

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                                             | Relevant to claim No. |
|-----------|------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| X         | CN 104376062 A (CHINA NONFERROUS METALS CHANGSHA INVESTI) 25 February 2015 (25.02.2015) description, paragraphs [0009]-[0098], figures 1 and 3 | 1-10                  |
| A         | CN 101364240 A (HANGZHOU H3C TECHNOLOGIES CO., LTD.) 11 February 2009 (11.02.2009) the whole document                                          | 1-10                  |
| A         | US 2013103705 A1 (SAP AG) 25 April 2013 (25.04.2013) the whole document                                                                        | 1-10                  |

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

☒ See patent family annex.

|                                                                                                                                                                         |                                                                                                                                                                                                                                                  |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| * Special categories of cited documents:                                                                                                                                | “T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention                                              |
| “A” document defining the general state of the art which is not considered to be of particular relevance                                                                | “X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone                                                                     |
| “E” earlier application or patent but published on or after the international filing date                                                                               | “Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art |
| “L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) | “&” document member of the same patent family                                                                                                                                                                                                    |
| “O” document referring to an oral disclosure, use, exhibition or other means                                                                                            |                                                                                                                                                                                                                                                  |
| “P” document published prior to the international filing date but later than the priority date claimed                                                                  |                                                                                                                                                                                                                                                  |

|                                                                                                                                                                                                               |                                                                      |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------|
| Date of the actual completion of the international search<br>11 July 2016                                                                                                                                     | Date of mailing of the international search report<br>25 July 2016   |
| Name and mailing address of the ISA<br>State Intellectual Property Office of the P. R. China<br>No. 6, Xitucheng Road, Jimenqiao<br>Haidian District, Beijing 100088, China<br>Facsimile No. (86-10) 62019451 | Authorized officer<br>YAN, Shiyong<br>Telephone No. (86-10) 62411767 |

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

International application No.  
PCT/CN2016/079649

| Patent Documents referred<br>in the Report | Publication Date | Patent Family    | Publication Date |
|--------------------------------------------|------------------|------------------|------------------|
| CN 104376062 A                             | 25 February 2015 | None             |                  |
| CN 101364240 A                             | 11 February 2009 | CN 101364240 B   | 21 July 2010     |
| US 2013103705 A1                           | 25 April 2013    | US 2007203923 A1 | 30 August 2007   |
|                                            |                  | US 8307012 B2    | 06 November 2012 |
|                                            |                  | US 8924415 B2    | 30 December 2014 |

## 国际检索报告

国际申请号

PCT/CN2016/079649

## 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 17/30, G06F 9

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))

CPRSABS;CNTXT;DWPI:异构, 数据库, 类型, 源, 目的, 转换, 中间, 抽象, heterogeneous, database?, type, source, target, convert+, transform+, mid+, inter+, abstract

## C. 相关文件

| 类 型* | 引用文件, 必要时, 指明相关段落                                                                        | 相关的权利要求 |
|------|------------------------------------------------------------------------------------------|---------|
| X    | CN 104376062 A (中国有色金属长沙勘察设计院有限公司) 2015年 2月 25日<br>(2015 - 02 - 25)<br>说明书第9-98段, 图1, 图3 | 1-10    |
| A    | CN 101364240 A (杭州华三通信技术有限公司) 2009年 2月 11日 (2009 - 02 - 11)<br>全文                        | 1-10    |
| A    | US 2013103705 A1 (SAP AG) 2013年 4月 25日 (2013 - 04 - 25)<br>全文                            | 1-10    |

☐ 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

“&amp;” 同族专利的文件

国际检索实际完成的日期

2016年 7月 11日

国际检索报告邮寄日期

2016年 7月 25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

授权官员

颜世莹

传真号 (86-10)62019451

电话号码 (86-10)62411767

国际检索报告  
关于同族专利的信息

|                   |
|-------------------|
| 国际申请号             |
| PCT/CN2016/079649 |

| 检索报告引用的专利文件 |            |    | 公布日<br>(年/月/日) | 同族专利 |            |    | 公布日<br>(年/月/日) |
|-------------|------------|----|----------------|------|------------|----|----------------|
| CN          | 104376062  | A  | 2015年 2月 25日   | 无    |            |    |                |
| CN          | 101364240  | A  | 2009年 2月 11日   | CN   | 101364240  | B  | 2010年 7月 21日   |
| US          | 2013103705 | A1 | 2013年 4月 25日   | US   | 2007203923 | A1 | 2007年 8月 30日   |
|             |            |    |                | US   | 8307012    | B2 | 2012年 11月 6日   |
|             |            |    |                | US   | 8924415    | B2 | 2014年 12月 30日  |

表 PCT/ISA/210 (同族专利附件) (2009年7月)