(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 June 2017 (08.06.2017)

WIPO | PCT

(10) International Publication Number
WO 2017/092447 A1

(51) International Patent Classification:
G06F 17/30 (2006.01)

(21) International Application Number:
PCT/CN2016/097717

(22) International Filing Date:
31 August 2016 (31.08.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
201510889904.9 4 December 2015 (04.12.2015) CN

(71) Applicants: **JIANGXI ELECTRIC POWER CORPORATION INFORMATION AND COMMUNICATIONS BRANCH OF STATE GRID** [CN/CN]; No.666, Hubin East Rd., Qingshanhu District, Nanchang, Jiangxi 330077 (CN). **STATE GRID CORPORATION OF CHINA** [CN/CN]; No. 86, West Chang'an Street, Xicheng District, Beijing 100031 (CN).

(72) Inventors: **FU, Pingping**; No.666, Hubin East Rd., Qingshanhu District, Nanchang, Jiangxi 330077 (CN). **MA, Yong**; No.666, Hubin East Rd., Qingshanhu District, Nanchang, Jiangxi 330077 (CN). **HAO, Hanyong**; No.86, West Chang'an Street, Xicheng District, Beijing 100031 (CN). **CHEN, Yanqing**; No.2, Chuangxin Rd., Gaoxin District, Nanchang, Jiangxi 330096 (CN). **CHEN, Xuelian**; No.666, Hubin East Rd., Qingshanhu District, Nanchang, Jiangxi 330077 (CN).

(74) Agent: **LEADER PATENT & TRADEMARK FIRM**; 8F-6, Bldg. A, Winland International Center, No.32 Xizhimen North Street, Haidian District, Beijing 100082 (CN).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: 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) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (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,

[Continued on next page]

(54) Title: METHOD AND APPARATUS FOR DATA QUALITY MANAGEMENT AND CONTROL

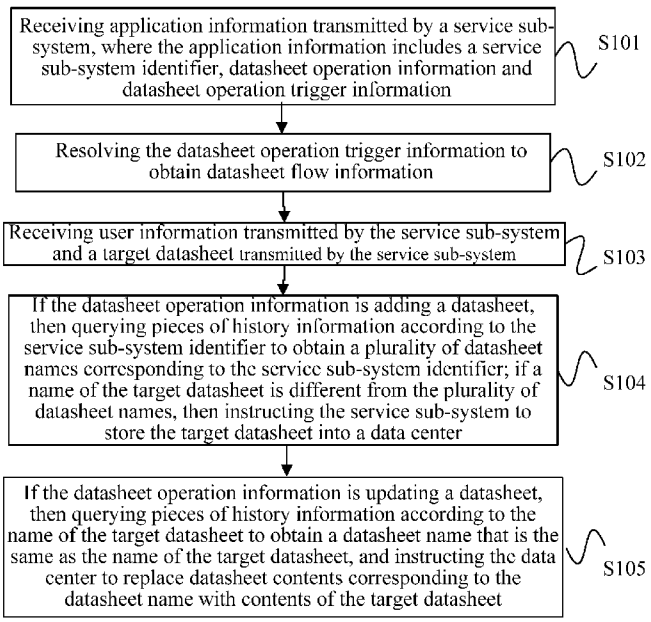


FIG. 1

(57) Abstract: A method and an apparatus for data quality management and control are provided. The method includes: receiving application information transmitted by a service sub-system, where the application information includes a service sub-system identifier, datasheet operation information and datasheet operation trigger information (S101); resolving datasheet operation trigger information to obtain datasheet flow information (S102); receiving user information transmitted by the service sub-system and a target datasheet transmitted by the service sub-system (S103); if a name of the target datasheet is different from a plurality of datasheet names corresponding to the service sub-system identifier, then instructing the service sub-system to store the target datasheet into a data center (S104); if the datasheet operation information is updating a datasheet, instructing the data center to replace datasheet contents corresponding to the datasheet name with contents of the target datasheet (S105).

WO 2017/092447 A1

SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

METHOD AND APPARATUS FOR DATA QUALITY MANAGEMENT AND CONTROL

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to Chinese Patent Application No. 201510889904.9, filed on December 04, 2015, the content of which is hereby incorporated by reference in its entirety.

TECHNICAL FIELD

[0002] Embodiments of the present invention relate to computer technologies and, in particular, to a method and an apparatus for data quality management and control.

BACKGROUND

[0003] With rapid development of enterprise informatization, a great number of data of a service application system accesses a data center, which results in rapid growth for databases and datasheets in the data center and a huge number of data resources.

[0004] In the prior art, the service application system includes a plurality of sub-systems. There is a data transmission interface between each sub-system and the data center, through which a sub-system stores data generated by itself into a tablespace allocated by the data center for the sub-system.

[0005] However, as growth rates of the databases and the datasheets become faster, and the amount of data stored by the data center becomes larger, the datasheets in the data center have the following four problems: 1) one tablespace may store datasheets generated by a plurality of sub-systems, and the datasheets are stored in the tablespace disorderly, resulting in difficulty in searching a datasheet; 2) same datasheets may be stored in a plurality of tablespaces, resulting in greater redundancy for the datasheets; 3) after a datasheet is updated,

both an old version and a new version exist, resulting in inconsistency of datasheets; 4) datasheet flow is not clear.

SUMMARY

[0006] Embodiments of the present invention provide a method and an apparatus for data quality management and control, which aim at lowering difficulty in searching a datasheet, reducing redundancy of the datasheet, improving consistency of datasheets and guaranteeing a clear flow of the datasheet to achieve management and control of data quality.

[0007] An aspect of an embodiment of the present invention provides a method for data quality management and control, including:

[0008] receiving application information transmitted by a service sub-system, where the application information includes a service sub-system identifier, datasheet operation information and datasheet operation trigger information;

[0009] resolving the datasheet operation trigger information to obtain datasheet flow information;

[0010] receiving user information transmitted by the service sub-system and a target datasheet transmitted by the service sub-system;

[0011] if the datasheet operation information is adding a datasheet, then querying pieces of history information according to the service sub-system identifier to obtain a plurality of datasheet names corresponding to the service sub-system identifier; if a name of the target datasheet is different from the plurality of datasheet names, then instructing the service sub-system to store the target datasheet into a data center;

[0012] if the datasheet operation information is updating a datasheet, then querying pieces of history information according to the name of the target datasheet to obtain a datasheet name that is the same as the name of the target datasheet, and instructing the data center to replace datasheet contents corresponding to the datasheet name with contents of the target datasheet;

[0013] where the pieces of history information at least comprise the service sub-system identifier and the datasheet name.

[0014] Another aspect of an embodiment of the present invention provides an apparatus for data quality management and control, including:

[0015] a receiving module, configured to receive application information transmitted by a service sub-system, where the application information includes a service sub-system identifier, datasheet operation information and datasheet operation trigger information; receive user information transmitted by the service sub-system and a target datasheet transmitted by the service sub-system;

[0016] a resolving module, configured to resolve the datasheet operation trigger information to obtain datasheet flow information;

[0017] a querying module, configured to: if the datasheet operation information is adding a datasheet, then query pieces of history information according to the service sub-system identifier to obtain a plurality of datasheet names corresponding to the service sub-system identifier; if the datasheet operation information is updating a datasheet, then query pieces of history information according to a name of the target datasheet to obtain a datasheet name that is the same as the name of the target datasheet; and

[0018] a transmitting module, configured to: if the name of the target datasheet is different from the plurality of datasheet names, then instruct the service sub-system to store the target datasheet into a data center; and instruct the data center to replace datasheet contents corresponding to the datasheet name with contents of the target datasheet;

[0019] where the pieces of history information at least comprise the service sub-system identifier and the datasheet name.

[0020] According to a method and an apparatus for data quality management and control provided in embodiments of the present invention, a target datasheet may be found through a datasheet name in history information stored by a system for data quality management and control; redundancy of datasheets is reduced by querying pieces of history information to obtain a plurality of datasheet names corresponding to a service sub-system identifier, and determining whether a name of the target datasheet is the same as the plurality of datasheet names, instructing a service sub-system to store the target datasheet into a data center only when the name of the target datasheet is different from the plurality of datasheet names; consistency of the datasheets is improved by querying pieces of history information to obtain

a datasheet name that is the same as the name of the target datasheet, and replacing datasheet contents corresponding to the datasheet name with contents of the target datasheet; and flow of the datasheet is guaranteed to be clear by resolving datasheet operation trigger information to obtain datasheet flow information; thus, management and control of data quality are achieved.

BRIEF DESCRIPTION OF DRAWINGS

[0021] FIG. 1 is a flow chart of a method for data quality management and control according to an embodiment of the present invention;

[0022] FIG. 2 is a network structural diagram applicable to a method for data quality management and control according to an embodiment of the present invention;

[0023] FIG. 3 is a structural diagram of an apparatus for data quality management and control according to an embodiment of the present invention;

[0024] FIG. 4 is a structural diagram of an apparatus for data quality management and control according to another embodiment of the present invention.

DESCRIPTION OF EMBODIMENTS

[0025] FIG. 1 is a flow chart of a method for data quality management and control according to an embodiment of the present invention; FIG. 2 is a network structural diagram applicable to a method for data quality management and control according to an embodiment of the present invention. An embodiment of the present invention provides a method for data quality management and control, targeting at several problems occurring in data of a data center due to faster growth rates of databases and datasheets and larger amount of data stored by the data center, where the method has the following specific steps.

[0026] Step S101, receiving application information transmitted by a service sub-system, where the application information includes a service sub-system identifier, datasheet operation information and datasheet operation trigger information;

[0027] An executive agent of embodiments of the present invention is a system for data quality management and control, as shown in FIG. 2, the system for data quality management

and control may perform information interactions with a plurality of service sub-systems in a service system, and may also perform information interactions with a data center. The data center is used to store a database and/or a datasheet generated by the service sub-system. The service system includes a plurality of service sub-systems, such as a production management system, a marketing business system, a human resource system, an electricity trading system, a financial system, etc.; each service sub-system has a plurality of users correspondingly, and each user may generate a database or a datasheet in the service sub-system; when a user generates a database or a datasheet in the service sub-system, the service sub-system transmits application information to the system for data quality management and control, where the application information includes a service sub-system identifier, datasheet operation information and datasheet operation trigger information; one database includes a plurality of datasheets, and the service sub-system identifier uniquely identifies one service sub-system; the datasheet operation information indicates an operation performed by a user for a database or a datasheet generated by itself, such as introducing a datasheet, updating a datasheet, deleting a datasheet, etc.; and the datasheet operation trigger information indicates a reason why the user performs the above-described operation for the database or the datasheet.

[0028] Step S102, resolving the datasheet operation trigger information to obtain datasheet flow information;

[0029] The system for data quality management and control resolves the reason why the user performs the above-described operation for the database or the datasheet so as to obtain datasheet flow information, where the datasheet flow information specifically indicates a datasheet generated by the user is viewed, called or stored by which service sub-system, that is, the datasheet flow information may view, call or store the service sub-system identifier of the datasheet.

[0030] Step S103, receiving user information transmitted by the service sub-system and a target datasheet transmitted by the service sub-system;

[0031] The system for data quality management and control reviews the application information transmitted by the service sub-system, after the review passes, the service sub-system transmits user information and a target datasheet to the system for data quality

management and control; the target datasheet is a datasheet generated by a user corresponding to the user information in the service sub-system.

[0032] Step S104, if the datasheet operation information is adding a datasheet, then querying pieces of history information according to the service sub-system identifier to obtain a plurality of datasheet names corresponding to the service sub-system identifier; if a name of the target datasheet is different from the plurality of datasheet names, then instructing the service sub-system to store the target datasheet into a data center;

[0033] If the target datasheet is a newly-added datasheet, that is, the target datasheet has not been stored in the data center, then the system for data quality management and control queries pieces of history information according to the service sub-system identifier to obtain a plurality of datasheet names corresponding to the service sub-system identifier; the system for data quality management and control is pre-stored with pieces of history information, each piece of history information at least includes the service sub-system identifier and the datasheet name; the system for data quality management and control determines whether a name of the target datasheet is the same as each datasheet name in the plurality of datasheet names respectively, if the name of the target datasheet is different from the plurality of datasheet names, then the system for data quality management and control instructs the service sub-system to store the target datasheet into the data center; if the name of the target datasheet is the same as any one of the plurality of datasheet names, then transmit warning information to the service sub-system, and notify the user of a repetition of the name of the target datasheet to avoid a repetition of datasheets in the data center.

[0034] Step S105, if the datasheet operation information is updating a datasheet, then querying pieces of history information according to the name of the target datasheet to obtain a datasheet name that is the same as the name of the target datasheet, and instructing the data center to replace datasheet contents corresponding to the datasheet name with contents of the target datasheet;

[0035] where the pieces of history information at least comprise the service sub-system identifier and the datasheet name.

[0036] If the target datasheet is an updated datasheet, that is, the target datasheet has been stored in the data center, the data center is stored with a datasheet having a name that is the

same as that of the target datasheet, and history information pre-stored by the system for data quality management and control also includes the datasheet name, then the system for data quality management and control queries pieces of history information according to the name of the target datasheet to obtain a datasheet name that is the same as the name of the target datasheet, and instructs the data center to replace datasheet contents corresponding to the datasheet name with contents of the target datasheet, thereby avoiding a phenomenon where there is inconsistency of datasheets stored in the data center.

[0037] According to the embodiment of the present invention, a target datasheet may be found through a datasheet name in history information stored by a system for data quality management and control; redundancy of datasheets is reduced by querying pieces of history information to obtain a plurality of datasheet names corresponding to a service sub-system identifier, and determining whether a name of the target datasheet is the same as the plurality of datasheet names, instructing a service sub-system to store the target datasheet into a data center only when the name of the target datasheet is different from the plurality of datasheet names; consistency of the datasheets is improved by querying pieces of history information to obtain a datasheet name that is the same as the name of the target datasheet, and replacing datasheet contents corresponding to the datasheet name with contents of the target datasheet; and flow of the datasheet is guaranteed to be clear by resolving datasheet operation trigger information to obtain datasheet flow information; thus, management and control of data quality are achieved.

[0038] Based on the above embodiment, after the instructing the service sub-system to store the target datasheet into the data center, further including: receiving a name of a tablespace storing the target datasheet as transmitted by the data center and an interface mode between the service sub-system and the data center transmitted by the data center; and generating current record information, where the current record information includes the service sub-system identifier, the name of the tablespace, the name of the target datasheet, the user information, the datasheet flow information, operation authorization information of the target datasheet and the interface mode.

[0039] After the system for data quality management and control instructs the service sub-system to store the target datasheet into the data center, the service sub-subsystem

transmits the target datasheet to the data center through an interface mode between the service sub-system and the data center, where the interface mode specifically includes any one of data warehousing technologies (Extract Transform Load, ETL for short), OGG, DBlink and Webservice. The data center may open up a tablespace for several service sub-systems, preferably, the data center opens up a tablespace for each service sub-system, the data center stores into the tablespace a target datasheet transmitted by the service sub-system through the interface mode, and transmits a name of the tablespace and the interface mode between the service sub-system and the data center to the system for data quality management and control.

[0040] The system for data quality management and control generates current record information according to the application information, the user information and the target datasheet which are transmitted by the service sub-system, the name of the tablespace and the interface mode which are transmitted by the data center as described in the above embodiment, where the current record information includes the service sub-system identifier, the name of the tablespace, the name of the target datasheet, the user information, the datasheet flow information, operation authorization information of the target datasheet and the interface mode, and the operation authorization information of the target datasheet is an authority set by the system for data quality management and control for the user for operating the target datasheet. When the system for data quality management and control receives application information transmitted by another service sub-system, the current record information then becomes history information.

[0041] After the receiving the user information transmitted by the service sub-system and the target datasheet transmitted by the service sub-system, further including: determining whether data in the target datasheet is in conformity with a preset data rule; if the data in the target datasheet is not in conformity with the preset data rule, then transmitting warning information to the service sub-system to enable a user to modify a data format in the target datasheet.

[0042] The system for data quality management and control is pre-stored with a data rule, data in target datasheets transmitted by each service sub-system should be in conformity with the data rule, that is, unified standardization is performed through the data rule to the data in the target datasheets transmitted by each service sub-system, so that the target datasheets

transmitted by each service sub-system may be identified by other service sub-systems, specifically, after receiving the user information transmitted by the service sub-system and the target datasheet transmitted by the service sub-system, the system for data quality management and control determines whether data in the target datasheet is in conformity with a preset data rule; if the data in the target datasheet is not in conformity with the preset data rule, then transmit warning information to the service sub-system to enable a user to modify a data format in the target datasheet.

[0043] According to the embodiment of the present invention, unified standardization is performed to data in target datasheets transmitted by each service sub-system through a data rule pre-stored in the system for data quality management and control, which ensures that the target datasheets transmitted by each service sub-system can be identified by other service sub-systems.

[0044] Based on the above-described embodiment, after the generating the current record information, further including: analyzing an importance degree of each datasheet according to the current record information and the pieces of history information, the more the datasheet flow information corresponding to the datasheet name, the higher the importance degree of the datasheet.

[0045] Each time a piece of current record information is generated, the system for data quality management and control puts the current record information into a record information library, an importance degree of each datasheet may be obtained by analyzing all pieces of record information in the record information library, and the importance degree of each datasheet may be measured through quantity of datasheet flow information corresponding to counted datasheet names, for instance, if a name of a target datasheet occurs in pieces of record information, it is obtained from analysis of the pieces of record information that the target datasheet is viewed, called or stored by a plurality of service sub-systems, then it indicates that the target datasheet has higher importance degree, and the quantity of the datasheet flow information corresponding to the datasheet names is proportional to the importance degree of the datasheet.

[0046] The method for data quality management and control also includes: setting a datasheet collecting rule, and collecting a plurality of datasheets from the data center

according to the datasheet collecting rule; determining whether names of any two datasheets in the plurality of the datasheets are the same, if the names of the two datasheets are the same, then determining whether contents of the two datasheets are the same; if the contents of the two datasheets are the same, then transmitting a first deleting command to the data center to enable the data center to delete any one of the two datasheets; if the contents of the two datasheets are different, then acquiring timestamps of the two datasheets from the data center, and transmitting a second deleting command to the data center to enable the data center to delete a datasheet having a smaller timestamp in the two datasheets.

[0047] The system for data quality management and control may also set a datasheet collecting rule, where the datasheet collecting rule includes periodic collecting, event triggered collecting, datasheet feature-based collecting, preset time point collecting, etc. The system for data quality management and control collects a plurality of datasheets from the data center according to the datasheet collecting rule, and determines whether names of any two datasheets in the plurality of datasheets are the same, if the names of the two datasheets are the same, then determine whether contents of the two datasheets are the same; if the contents of the two datasheets are the same, then transmit a first deleting command to the data center to enable the data center to delete any one of the two datasheets, thereby further reducing the redundancy of the datasheets in the data center; if the contents of the two datasheets are different, then acquire timestamps of the two datasheets from the data center, and transmit a second deleting command to the data center to enable the data center to delete a datasheet having a smaller timestamp in the two datasheets, thereby further increasing the consistency of the datasheets in the data center.

[0048] According to the embodiment of the present invention, through the system for data quality management and control, a plurality of datasheets in the data center is collected, and names and contents of any two datasheets in the plurality of datasheets are determined; the redundancy of the datasheets in the data center is further reduced by deleting any one of two datasheets having same names and same contents; and meanwhile the consistency of the datasheets in the data center is further improved by deleting a datasheet having a smaller timestamp in the two datasheets having the same names but different contents.

[0049] FIG. 3 is a structural diagram of an apparatus for data quality management and control according to an embodiment of the present invention. The apparatus for data quality management and control as shown in FIG. 3 is specifically the system for data quality management and control as shown in FIG. 2. The apparatus for data quality management and control according to the embodiment of the present invention may perform processing flows according to embodiments of the method for data quality management and control, as shown in FIG. 3, the apparatus 30 for data quality management and control includes: a receiving module 31, a resolving module 32, a querying module 33 and a transmitting module 34. The receiving module 31 is configured to receive application information transmitted by a service sub-system, where the application information includes a service sub-system identifier, datasheet operation information and datasheet operation trigger information; and receive user information transmitted by the service sub-system and a target datasheet transmitted by the service sub-system. The resolving module 32 is configured to resolve the datasheet operation trigger information to obtain datasheet flow information. The querying module 33 is configured to: if the datasheet operation information is adding a datasheet, then query pieces of history information according to the service sub-system identifier to obtain a plurality of datasheet names corresponding to the service sub-system identifier; if the datasheet operation information is updating a datasheet, then query pieces of history information according to a name of the target datasheet to obtain a datasheet name that is the same as the name of the target datasheet. The transmitting module 34 is configured to: if the name of the target datasheet is different from the plurality of datasheet names, then instruct the service sub-system to store the target datasheet into a data center; and instruct the data center to replace datasheet contents corresponding to the datasheet name with contents of the target datasheet, where the pieces of history information at least include the service sub-system identifier and the datasheet name.

[0050] According to the embodiment of the present invention, a target datasheet may be found through a datasheet name in history information stored by a system for data quality management and control; redundancy of datasheets is reduced by querying pieces of history information to obtain a plurality of datasheet names corresponding to a service sub-system identifier, and determining whether a name of the target datasheet is the same as the plurality

of datasheet names, instructing a service sub-system to store the target datasheet into a data center only when the name of the target datasheet is different from the plurality of datasheet names; consistency of the datasheets is improved by querying pieces of history information to obtain a datasheet name that is the same as the name of the target datasheet, and replacing datasheet contents corresponding to the datasheet name with contents of the target datasheet; and flow of the datasheet is guaranteed to be clear by resolving datasheet operation trigger information to obtain datasheet flow information; thus, management and control of data quality are achieved.

[0051] FIG. 4 is a structural diagram of an apparatus for data quality management and control according to another embodiment of the present invention. Based on the above-described embodiment, the receiving module 31 is further configured to, after the transmitting module 34 instructs the service sub-system to store the target datasheet into the data center, receive a name of a tablespace storing the target datasheet as transmitted by the data center and an interface mode between the service sub-system and the data center transmitted by the data center; the apparatus 30 for data quality management and control also includes a recording module 35. The recording module 35 is configured to generate current record information, where the current record information includes the service sub-system identifier, the name of the tablespace, the name of the target datasheet, the user information, the datasheet flow information, operation authorization information of the target datasheet and the interface mode.

[0052] The apparatus 30 for data quality management and control also includes a determining module 36. The determining module 36 is configured to: after the receiving module 31 receives the user information transmitted by the service sub-system and the target datasheet transmitted by the service sub-system, determine whether data in the target datasheet is in conformity with a preset data rule; the transmitting module 34 is further configured to: if the data in the target datasheet is not in conformity with the preset data rule, then transmit warning information to the service sub-system to enable a user to modify a data format in the target datasheet.

[0053] The determining module 36 is further configured to analyze an importance degree of each datasheet according to the current record information and the pieces of history

information, the more the datasheet flow information corresponding to the datasheet name, the higher the importance degree of the datasheet.

[0054] The apparatus 30 for data quality management and control also includes a collecting module 37. The collecting module 37 is configured to set a datasheet collecting rule, and collect a plurality of datasheets from the data center according to the datasheet collecting rule; the determining module 36 is further configured to determine whether names of any two datasheets in the plurality of the datasheets are the same, if the names of the two datasheets are the same, then determine whether contents of the two datasheets are the same; the transmitting module 34 is further configured to: if the contents of the two datasheets are the same, then transmit a first deleting command to the data center to enable the data center to delete any one of the two datasheets; if the contents of the two datasheets are different, then acquire timestamps of the two datasheets from the data center, and transmit a second deleting command to the data center to enable the data center to delete a datasheet having a smaller timestamp in the two datasheets.

[0055] The apparatus for data quality management and control according to the embodiment of the present invention may be specifically used to perform the method embodiments provided in FIG. 1 above, and specific functions will not be repeated herein.

[0056] According to the embodiment of the present invention, unified standardization is performed to data in target datasheets transmitted by each service sub-system through a data rule pre-stored in the system for data quality management and control, which ensures that the target datasheets transmitted by each service sub-system can be identified by other service sub-systems; through the system for data quality management and control, a plurality of datasheets in the data center is collected, and names and contents of any two datasheets in the plurality of datasheets are determined; the redundancy of the datasheets in the data center is further reduced by deleting any one of two datasheets having same names and same contents; and meanwhile the consistency of the datasheets in the data center is further improved by deleting a datasheet having a smaller timestamp in the two datasheets having the same names but different contents.

[0057] In conclusion, according to embodiments of the present invention, a target datasheet may be found through a datasheet name in history information stored by a system

for data quality management and control; redundancy of datasheets is reduced by querying pieces of history information to obtain a plurality of datasheet names corresponding to a service sub-system identifier, and determining whether a name of the target datasheet is the same as the plurality of datasheet names, instructing a service sub-system to store the target datasheet into a data center only when the name of the target datasheet is different from the plurality of datasheet names; consistency of the datasheets is improved by querying pieces of history information to obtain a datasheet name that is the same as the name of the target datasheet, and replacing datasheet contents corresponding to the datasheet name with contents of the target datasheet; and flow of the datasheet is guaranteed to be clear by resolving datasheet operation trigger information to obtain datasheet flow information; thus, management and control of data quality are achieved; unified standardization is performed to data in target datasheets transmitted by each service sub-system through a data rule pre-stored in the system for data quality management and control, which ensures that the target datasheets transmitted by each service sub-system can be identified by other service sub-systems; through the system for data quality management and control, a plurality of datasheets in the data center is collected, and names and contents of any two datasheets in the plurality of datasheets are determined; the redundancy of the datasheets in the data center is further reduced by deleting any one of two datasheets having same names and same contents; and meanwhile the consistency of the datasheets in the data center is further improved by deleting a datasheet having a smaller timestamp in the two datasheets having the same names but different contents.

[0058] In embodiments provided in the present invention, it should be understood that the disclosed apparatus and method may be implemented in other manners. For example, the above-described apparatus embodiments are merely exemplary. For example, the unit division is merely logical function division and there may be other divisions during actual implementation. For example, a plurality of units or components may be combined or integrated into another system, or some features may be ignored or not performed. In addition, the displayed or discussed mutual couplings or direct couplings or communication connections may be implemented through some interfaces. The indirect couplings or

communication connections between the apparatuses or units may be implemented in electronic, mechanical, or other forms.

[0059] The units described as separate parts may or may not be physically separate, and parts displayed as units may or may not be physical units, that is, may be located in one position, or may be distributed on a plurality of network units. A part or all of the units may be selected according to actual needs to achieve the objectives of the solutions of the embodiments.

[0060] In addition, functional units in the embodiments of the present invention may be integrated into one processing unit, or each of the units may exist alone physically, or two or more units are integrated into one unit. The integrated unit may be implemented in a form of hardware, or may be implemented in a form of a hardware plus software functional unit.

[0061] The integrated unit implemented in the form of the software functional unit as described above may be stored in a computer-readable storage medium. The software functional unit described above is stored in a storage medium, and includes several instructions capable of enabling a computer device (which may be a personal computer, a server, a network device, or the like) or a processor to perform a part of the steps of the methods described in the embodiments of the present invention. The foregoing storage medium includes: any medium that can store program codes, such as a USB flash disk, a mobile hard disk, a read-only memory (Read-Only Memory, ROM), a random access memory (Random Access Memory, RAM), a magnetic disk, or an optical disc.

[0062] Persons of ordinary skill in the art may clearly understand that, for the purpose of convenient and brief description, the division of the above functional modules is merely described for exemplary purposes, during practical use, the above functions may be assigned to different functional modules as required, that is, an internal structure of the apparatus is divided into different functional modules to perform all or a part of functions as described above. For a specific operating process of the above apparatus, reference may be made to corresponding procedures in the foregoing method embodiments, which will not be repeated herein.

[0063] Finally, it should be noted that the foregoing embodiments are merely intended for describing technical solutions of the present invention rather than limiting the present

invention. Although the present invention is described in detail with reference to the foregoing embodiments, persons of ordinary skill in the art should understand that they may still make modifications to the technical solutions described in the foregoing embodiments, or make equivalent replacements to some or all technical features therein; however, these modifications or replacements do not make the essence of corresponding technical solutions depart from the scope of the technical solutions in the embodiments of the present invention.

WHAT IS CLAIMED IS:

1. A method for data quality management and control, comprising:

receiving application information transmitted by a service sub-system, wherein the application information comprises a service sub-system identifier, datasheet operation information and datasheet operation trigger information;

resolving the datasheet operation trigger information to obtain datasheet flow information;

receiving user information transmitted by the service sub-system and a target datasheet transmitted by the service sub-system;

if the datasheet operation information is adding a datasheet, then querying pieces of history information according to the service sub-system identifier to obtain a plurality of datasheet names corresponding to the service sub-system identifier; if a name of the target datasheet is different from the plurality of datasheet names, then instructing the service sub-system to store the target datasheet into a data center;

if the datasheet operation information is updating a datasheet, then querying pieces of history information according to the name of the target datasheet to obtain a datasheet name that is the same as the name of the target datasheet, and instructing the data center to replace datasheet contents corresponding to the datasheet name with contents of the target datasheet;

wherein the pieces of history information at least comprise the service sub-system identifier and the datasheet name.

2. The method according to claim 1, after the instructing the service sub-system to store the target datasheet into the data center, further comprising:

receiving a name of a tablespace storing the target datasheet as transmitted by the data center and an interface mode between the service sub-system and the data center transmitted by the data center; and

generating current record information, wherein the current record information comprises the service sub-system identifier, the name of the tablespace, the name of the target datasheet,

the user information, the datasheet flow information, operation authorization information of the target datasheet and the interface mode.

3. The method according to claim 2, after the receiving the user information transmitted by the service sub-system and the target datasheet transmitted by the service sub-system, further comprising:

determining whether data in the target datasheet is in conformity with a preset data rule;

if the data in the target datasheet is not in conformity with the preset data rule, then transmitting warning information to the service sub-system to enable a user to modify a data format in the target datasheet.

4. The method according to claim 3, after the generating the current record information, further comprising:

analyzing an importance degree of each datasheet according to the current record information and the pieces of history information, the more the datasheet flow information corresponding to the datasheet name, the higher the importance degree of the datasheet.

5. The method according to claim 4, further comprising:

setting a datasheet collecting rule, and collecting a plurality of datasheets from the data center according to the datasheet collecting rule;

determining whether names of any two datasheets in the plurality of the datasheets are the same, if the names of the two datasheets are the same, then determining whether contents of the two datasheets are the same;

if the contents of the two datasheets are the same, then transmitting a first deleting command to the data center to enable the data center to delete any one of the two datasheets;

if the contents of the two datasheets are different, then acquiring timestamps of the two datasheets from the data center, and transmitting a second deleting command to the data center to enable the data center to delete a datasheet having a smaller timestamp in the two datasheets.

6. An apparatus for data quality management and control, comprising:

a receiving module, configured to receive application information transmitted by a service sub-system, wherein the application information comprises a service sub-system identifier, datasheet operation information and datasheet operation trigger information;

receive user information transmitted by the service sub-system and a target datasheet transmitted by the service sub-system;

a resolving module, configured to resolve the datasheet operation trigger information to obtain datasheet flow information;

a querying module, configured to: if the datasheet operation information is adding a datasheet, then query pieces of history information according to the service sub-system identifier to obtain a plurality of datasheet names corresponding to the service sub-system identifier; if the datasheet operation information is updating a datasheet, then query pieces of history information according to a name of the target datasheet to obtain a datasheet name that is the same as the name of the target datasheet; and

a transmitting module, configured to: if the name of the target datasheet is different from the plurality of datasheet names, then instruct the service sub-system to store the target datasheet into a data center; and instruct the data center to replace datasheet contents corresponding to the datasheet name with contents of the target datasheet;

wherein the pieces of history information at least comprise the service sub-system identifier and the datasheet name.

7. The apparatus for data quality management and control according to claim 6, wherein the receiving module is further configured to, after the transmitting module instructs the service sub-system to store the target datasheet into the data center, receive a name of a tablespace storing the target datasheet as transmitted by the data center and an interface mode between the service sub-system and the data center transmitted by the data center;

the apparatus for data quality management and control further comprises a recording module configured to generate current record information, wherein the current record information comprises the service sub-system identifier, the name of the tablespace, the name of the target datasheet, the user information, the datasheet flow information, operation authorization information of the target datasheet and the interface mode.

8. The apparatus for data quality management and control according to claim 7, further comprising:

a determining module, configured to: after the receiving module receives the user information transmitted by the service sub-system and the target datasheet transmitted by the

service sub-system, determine whether data in the target datasheet is in conformity with a preset data rule;

the transmitting module is further configured to: if the data in the target datasheet is not in conformity with the preset data rule, then transmit warning information to the service sub-system to enable a user to modify a data format in the target datasheet.

9. The apparatus for data quality management and control according to claim 8, wherein, the determining module is further configured to analyze an importance degree of each datasheet according to the current record information and the pieces of history information, the more the datasheet flow information corresponding to the datasheet name, the higher the importance degree of the datasheet.

10. The apparatus for data quality management and control according to claim 9, further comprising:

a collecting module, configured to set a datasheet collecting rule, and collect a plurality of datasheets from the data center according to the datasheet collecting rule;

the determining module is further configured to determine whether names of any two datasheets in the plurality of the datasheets are the same, if the names of the two datasheets are the same, then determine whether contents of the two datasheets are the same;

the transmitting module is further configured to: if the contents of the two datasheets are the same, then transmit a first deleting command to the data center to enable the data center to delete any one of the two datasheets; if the contents of the two datasheets are different, then acquire timestamps of the two datasheets from the data center, and transmit a second deleting command to the data center to enable the data center to delete a datasheet having a smaller timestamp in the two datasheets.

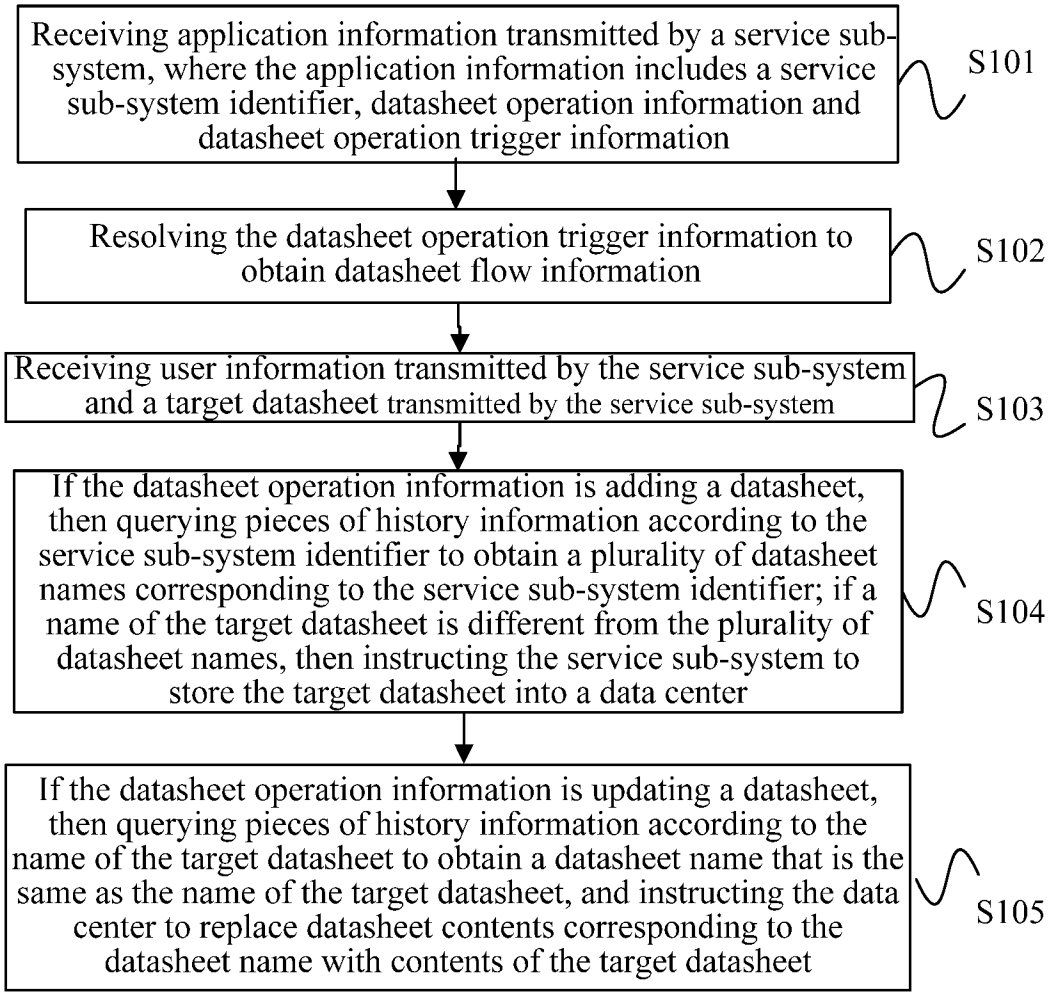

Receiving application information transmitted by a service sub-system, where the application information includes a service sub-system identifier, datasheet operation information and datasheet operation trigger information
S101
Resolving the datasheet operation trigger information to obtain datasheet flow information
S102
Receiving user information transmitted by the service sub-system and a target datasheet transmitted by the service sub-system
S103
If the datasheet operation information is adding a datasheet, then querying pieces of history information according to the service sub-system identifier to obtain a plurality of datasheet names corresponding to the service sub-system identifier; if a name of the target datasheet is different from the plurality of datasheet names, then instructing the service sub-system to store the target datasheet into a data center
S104
If the datasheet operation information is updating a datasheet, then querying pieces of history information according to the name of the target datasheet to obtain a datasheet name that is the same as the name of the target datasheet, and instructing the data center to replace datasheet contents corresponding to the datasheet name with contents of the target datasheet
S105

FIG. 1

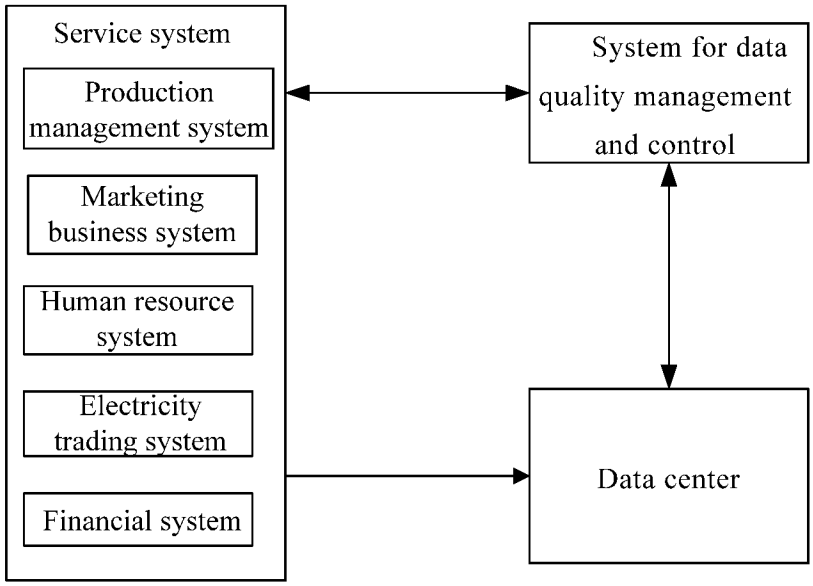

Service system
Production management system
Marketing business system
Human resource system
Electricity trading system
Financial system
System for data quality management and control
Data center

FIG. 2

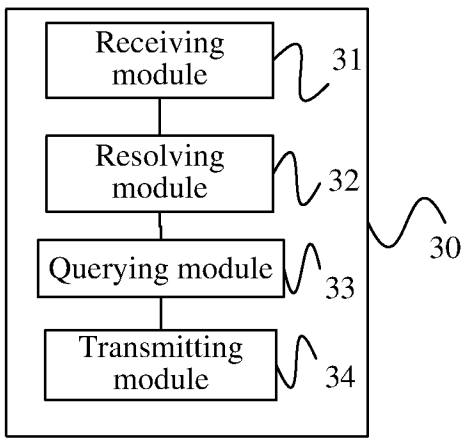

Receiving module
31
Resolving module
32
Querying module
33
Transmitting module
34
30

FIG. 3

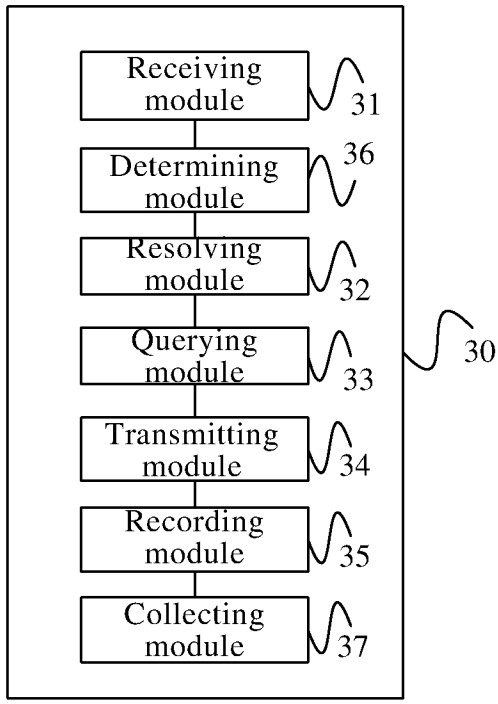

Receiving module
31
Determining module
36
Resolving module
32
Querying module
33
Transmitting module
34
Recording module
35
Collecting module
37
30

FIG. 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2016/097717

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,IEEE,GOOGLE,WPI,EPODOC,CNKI:table,name,duplicate,unique,control,add,update,database,service,identifier,id, business

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 105512283 A (JIANGXI ELECTRIC POWER CORPORATION INFORMATION AND COMMUNICATIONS BRANCH OF STATE GRID ET AL.) 20 April 2016 (2016-04-20) claims 1-10	1-10
X	CN 102202073 A (H3C TECHNOLOGIES CO., LTD.) 28 September 2011 (2011-09-28) description, paragraphs 0076-0085, figures 1-4	1-10
A	CN 103617255 A (STATE GRID CORPORATION OF CHINA ET AL.) 05 March 2014 (2014-03-05) the whole document	1-10
A	US 2015339329 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 26 November 2015 (2015-11-26) the whole document	1-10

☐ 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	Date of mailing of the international search report
15 November 2016	**05 December 2016**

Name and mailing address of the ISA/CN	Authorized officer
STATE INTELLECTUAL PROPERTY OFFICE OF THE P.R.CHINA **6, Xitucheng Rd., Jimen Bridge, Haidian District, Beijing 100088** **China**	**WANG,Ping**
Facsimile No. **(86-10)62019451**	Telephone No. **(86-10)62413666**

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2016/097717

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 105512283 A	20 April 2016	None	
CN 102202073 A	28 September 2011	None	
CN 103617255 A	05 March 2014	None	
US 2015339329 A1	26 November 2015	US 2014280248 A1	18 September 2014