SSOAR

Open Access Repository
www.ssoar.info

Data integration in the distributed information systems

Bisterfeld, O.A

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:
Bisterfeld, O. (2012). Data integration in the distributed information systems. *International Journal of Advanced Studies*, 1, 1-4. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-315608

Nutzungsbedingungen:
Dieser Text wird unter einer Basic Digital Peer Publishing-Lizenz zur Verfügung gestellt. Nähere Auskünfte zu den DiPP-Lizenzen finden Sie hier:
http://www.dipp.nrw.de/lizenzen/dppl/service/dppl/

Terms of use:
This document is made available under a Basic Digital Peer Publishing Licence. For more Information see:
http://www.dipp.nrw.de/lizenzen/dppl/service/dppl/

gesis Leibniz-Institut für Sozialwissenschaften

Mitglied der Leibniz-Gemeinschaft

Data integration in the distributed information systems

Problems and decision methods

Bisterfeld O.A.
Department of Computer Science
Ryazan State University named for S.A. Esenin
Russia
bist19@yandex.ru

Abstract
Problems of data integration in distributed information systems are analyzed. The author offered a method of preliminary correct formation of the integrable structured information resources. This method provides an acceleration of data interchanges between databases.

Keywords: data interchange; data integration; distributed information systems.

I. Introduction

Prospects of researches in the field of integration of information resources are defined by that in a number of specialized systems of different function, in medical researches and in medical attendance, in education formation of the integrable structured information resources are required. In many such systems the technology of relational databases (DB) for accumulation and storage of structured data now is used.

In the distributed information systems, databases, each of which has possibility of independent functioning in each component of system, interact on telecommunication communication links.

Integrable resources shall be brought in a status when all identical fragments of DB of different components of distributed system contain identical data. Periodic information exchanges use for coordination of integrable resources. Markup languages of data: XML, HTML, etc., allowing to structure exchange data arrays, use for the organization of information exchanges.

II. Problems of data integration

The most widespread known methods of data integration in system of interacting DB [1, 2, 3, 4] don't allow to organize correct data arrays.

In technology of data handling by means of relational DB for several last decades qualifiers and data coding systems are widely used. It is the method of integration of data most widespread on practice in the distributed information systems. Qualifiers and data coding systems are prepared and used in advance in interacting DB of systems. For example, in all interacting DB of any system it is accepted to represent a type of the document a code (1 – the outgoing letter; 2 – the incoming letter, 3 – the order etc.). In known methods [1, 2] repeated repetition of the same data records according to the semantic contents is allowed when data arrive from different data sources. Preliminary formation of the centralized identifiers (CI) for qualifiers and coding systems isn't provided. That leads to superfluous procedures in case of exchanges between DB.

There is a method [3], allowing to create in integrable fragments of interacting DB uniform data array without repetitions of records when identical to record arrive from several data sources also. However, formation of such integrable fragments of a DB in separate components of information systems occurs gradually, on a course of reception of exchange data arrays.

Procedures of CI formation implemented by known methods don't provide preliminary mailing CI arrays considerably changed during any moment from service of data identification (SDI). Any DB of system receives CI only on own requests (when in this DB new record will be entered into «an integrable information resource» on own initiative). It can lead to excessive double-sided exchanges from a DB to the SDI and is reverse. Probability of intensive requests from a DB to the SDI very big if on organizational aspects in any period «integrable information resources» considerably extend.

Absence in known methods of "anticipatory" mailing of the changed arrays of CI from the SDI can lead to time delays of readiness of a DB to data output. Such time delays can be unacceptable in the information systems working in real time.

In known methods the preliminary centralized preparation of new CI arrays isn't provided. If such arrays can be in advance prepared and delivered, efficiency of the forthcoming exchanges between a DB will be better.

III. Method of information exchange

The author offered a method of preliminary correct formation of the integrable structured information resources, providing an acceleration of data interchanges between databases [5].

The offered method provides:

- the centralized "preliminary" generation of arrays of CI and their mailing to DB regarding qualifiers and data coding systems;
- the centralized "preliminary" mailing of arrays of CI in cases of their changes;
- the centralized "preliminary" generation of additional arrays of CI and their mailing to DB in case of plan extensions of systems «integrable information resources».

Description of information exchange's method

The conceptual diagrams representing information models, are created as in a known method. Models describe

types of information objects (entities) with sets of their properties (attributes) and the relations in between, used in case of information exchanges between DB. In the analysis of such diagrams, information objects which can be duplicated by transmissions will be revealed. Such objects will be marked by labels. Unique keys will be appropriated to types of information objects (entities), types of attributes, types of the relations.

Electronic analogs of "application-oriented diagrams of exchanges" create in the centralized SDI and deliver to DB for storage. Arrays of CI, concerning qualifiers and data coding systems, also shall be prepared in SDI and are delivered to DB.

In the course of operation primary entering new records about the objects (entities) having labels in application-oriented diagrams of exchanges, into any DB is accompanied by initialization of procedure of request of the unique key of CI centralized for an information system. The unique key CI can be created by generation of increasing sequence of integral numbers. The generated key of CI and entity with a set of its attributes shall be written in SDI.CI shall be sent to the DB which has requested CI. CI and own unique key of object register in a DB.

In the course of operation, in case of changes in arrays of CI (or as option, in case of accumulation of a certain volume of such changes) SDI creates mailing to all DB system of additional arrays of CI (with information on attributes of the appropriate information objects). In the DB of the information system which has received a new array CI, if necessary, also additional procedure of preliminary data entry about the appropriate information objects of «an integrable information resource» is carried out. Besides, additional arrays of CI in SDI can beforehand be created and delivered on a DB in case of plan extensions of «integrable information resources».

During information exchanges in a DB (transferring data) the exchange array is created. By means of unique own keys of records or types of records (if uniqueness of keys is provided only within copies of objects of one type) CI are selected. CI will be added in record of exchange array.

In a DB (accepting data) the analysis of exchange array and detection of the transferred objects having CI are carried out. Verification of presence in a DB of the same "own objects" is executed.

To provide data storage in a DB, change of records of exchange array is necessary. Readdressing of links to "own objects" shall be executed. Data on the objects which are not available in a DB will be entered first of all. Received "own keys" are remembered together with keys from an exchange array ("own keys" are prepared for changeover by them of keys in an exchange array). Then carry out the second change of an exchange array - readdressing of links to "own keys". The relations and the object properties readdressed on "own copies" and "own keys", shall be saved in a DB in the last queue.

Such approach allows to realize correct compression of information when entering data from an exchange array (data on objects already available in a DB and their properties aren't entered). Time for support of readiness of a DB to transmission of exchange data considerably decreases. Combination of data from several DB (transferring data) in logically connected information array is carried out. There isn't need of the systematic solution of questions about incorrect reproduction identical information objects arriving from several DB in DB (accepting data). As a result, maintenance of an information system becomes cheaper.

The structure of an information system and system for implementation of an offered exchange method is given on a figure 1.

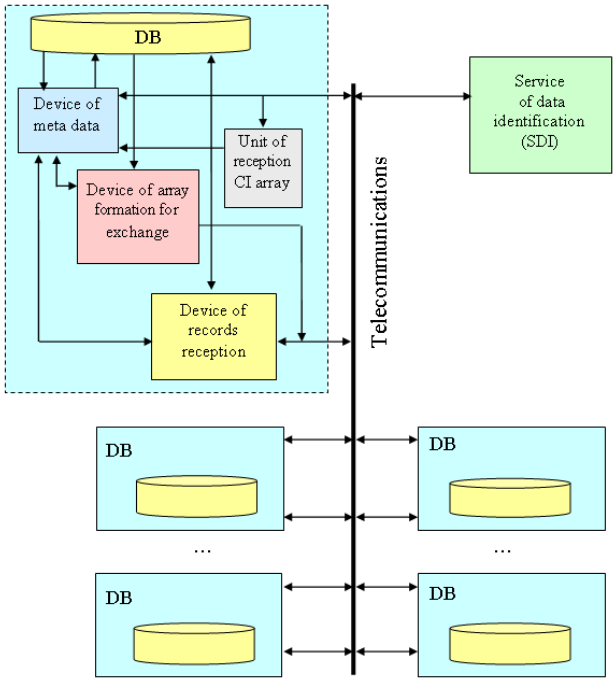

Figure 1. The structure of an information system and system for implementation of an offered exchange method

IV. UNITS DEVICE OF SYSTEM

The structure of the SDI device is given on a figure 2.

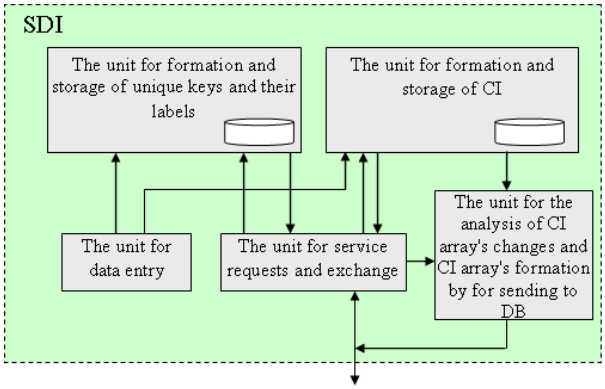

Figure 2. The structure of the SDI device

The structure of the device of meta data of information system's databases is given on a figure 3.

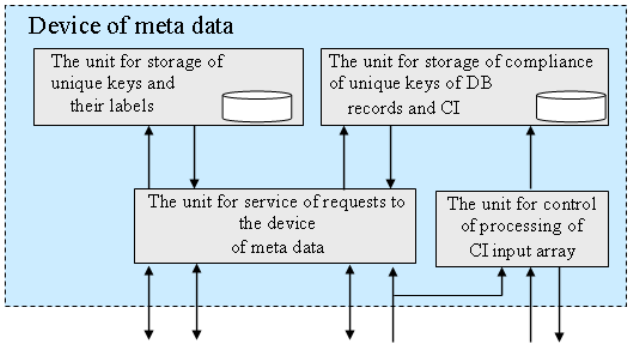

Figure 3. The structure of the device of meta data of information system's databases

The unit for storage of unique keys and their labels, the unit for storage of compliance of unique keys of DB (figure 2), the unit for storage of unique keys and their labels, the unit for storage of compliance of unique keys of DB (figure 3) can be executed in the form of relational databases.

The structure of the device of array formation for exchange is given on a figure 4.

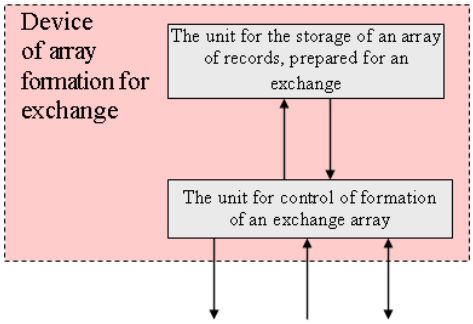

Figure 4. The structure of the device of array formation for exchange

The structure of the device of of records reception is given on a figure 5.

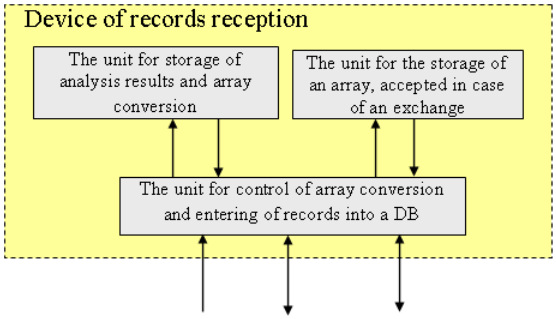

Figure 5. The structure of the device of records reception

The unit for the storage of an array of records, prepared for an exchange (figure 4), the unit for the storage of an array, accepted in case of an exchange, the unit for storage of analysis results and array conversion (figure 5) are intended for temporal data storage and can be executed in the form of storage devices (or in the form of arrays in case of implementation by a program method).

V. DESCRIPTION OF DEVICE OPERATION

For implementation of centralized CI generation the device of service of data identification (SDI) is switched on in offered system (figure 1). By means of the unit for data entry (figure 2) in the unit for formation and storage of unique keys for types of data records and signs of attending need, electronic analogs of application-oriented diagrams of exchanges between databases are created with specifying of their labels and with the qualifier. The first table for support of an offered exchange method has three columns in electronic analog. In it those information objects which in the course of exchanges can be duplicated by transmissions from different DB shall be marked. Such marks are entered into the third column of the table (a logical mark "True" - if duplicating transmissions are possible). When entering new records into the unit for formation and storage of unique keys the unique code for types of data records and their labels is generated, it registers in the first column of the table. This code is the centralized key for identification of object types (CI). Contents of the unit for formation and storage of unique keys for types of data records and their labels repeats in units for storage of unique keys for types of data records and their labels of all DB of an information system by means of the unit for service requests and exchange (figure 2), SDI communication link with the device of meta data of DB and the unit for service of requests to the device of meta data (figure 3). Units for storage of unique keys and their labels of all DB have the same structure of tables of storage, as the SDI appropriate unit (figure 2), but keys in these units only repeat keys from the SDI unit and independently aren't created.

Also records for qualifiers and data coding systems are created in the unit for formation and storage of unique keys by means of the unit for data entry (figure 2). The unit for the analysis of CI array's changes and CI array's formation creates arrays of qualifiers and data coding systems and transfers on communication link in system's databases. Unit of reception CI array accepts an array from the channel. The array is transferred in the device of meta data of a DB. The unit for control of processing of CI input array writes records of qualifiers and coding system (with informative attributes) in a DB. "Own keys" received in a DB return to the unit for control of processing of CI input array.

In the course of operation of an information system when entering any new records into relational databases, the unit for service of requests to the device of meta data selects a sign from the unit for storage of unique keys and their labels (figure 3). If this sign matters "True", the unit for service of requests enters new record in the table of storage of the unit for storage of compliance of unique keys of DB records and CI. This table contains data on existence of the objects, necessary for an exchange method. At first the key of a database will be written in the first column of this table, and the unique key of object will be written in the second column. The unit for service of requests to the device of meta data creates and sends request for receiving CI to SDI.

In case of reception of request for receiving CI, the The unit for service requests (figure 2) selects necessary records from the table of storage of the unit for formation and storage of unique keys. Comparing with the appropriate values of properties will be carried out further. In case of the coincidence, CI, written in the first column of the table, will be given out as result of request. In case of absence of coincidence, the message to the SDI manager will be output.

After an exception of possible uncoordinated actions of users of an information system during input of values (for example if semantics of values of properties is identical) the decision on creation new CI will be accepted in an interactive mode. New record in the table of storage of the unit for formation and storage of unique keys will be created. Created CI will be given out as result of DB request.

When receiving result of request the unit for service of requests to the device of meta data writes CI into the table of storage of the unit for storage of compliance of unique keys of DB records and CI. CI will be written in the third column of the table containing data on existence of objects.

Formation of an array of records for an exchange occurs as follows. The unit for control of formation of an exchange array writes an array into the unit for the storage of an array of records, prepared for an exchange (figure 4). The unit for control selects from the device of meta data CI being stored in it and writes CI into the fourth column of the table.

In case of reception of an exchange array the unit for control of array conversion and entering of records into a DB writes an array into the unit for the storage of an array, accepted in case of an exchange (figure 5). At the first stage the unit for control of array conversion and entering of records into a DB reveals by means of requests in the device of meta data the appropriate keys of "own objects" and enters it and signs into the table of of the unit for storage of analysis results and array conversion. In the absence of object with CI in own DB or if the object can't have CI, the unit for control of array conversion and entering of records into a DB enters similar objects in own DB, receives "own keys" and also enters "own keys" and signs into the table of the unit for storage of analysis results and array conversion. At the second stage of conversion of an exchange array the unit for control of array conversion for each object carries out a viewing cycle of an exchange array and readdressing of links. New records keys and the appropriate signs will be written into the table of the unit for storage of analysis results and array conversion. Thus the properties belonging to "own objects", will be marked with a label "don't write". At the last stage the transformed records of properties and the relations will be written into own database. The new data array will be connected by such procedures with existing in a database, the correct array will be as a result received.

Anticipatory mailing can be carried out in the following cases:

- in case of appearance of new records CI;
- in case of "accumulation" of some volume of changes;
- in pauses when the information system is free from execution of target tasks.

The unit for the analysis of CI array's changes (figure 2) in an automatic mode in interaction with the unit for service of requests of databases and an exchange initiates mailing into DB of the accumulated array of "new" CI (with unique keys and informative attributes of information objects corresponding to it).

In the course of an information system's functioning according to plans of extension of integrable information resources by means of the unit for data entry (figure 2) new data arrays can be entered into the unit for formation and storage of unique keys. In this case new arrays of CI, attributes of information objects will be created. By means of the unit for the analysis of CI array's changes, addition of integrable information resources will be transferred into information system's DB.

ACKNOWLEDGMENT

As a result of application of the expected invention, in comparison with a known method, the primary effect consisting in increase of the functional capabilities of an information system is reached:

- identification in a DB of qualifiers system and system of coding by the centralized identifiers (CI) ceased to be "step-by-step", it provides an acceleration of data interchanges between a DB;
- "anticipatory" mailing of accumulated changes (additions) of CI tables reduced intensity of exchanges between SDI and DB of information system , it provides an acceleration of data interchanges between DB in case of execution by an information system of target tasks;
- "anticipatory" mailing of data, plan adding, in integrable information resources reduced intensity of exchanges between SDI and DB of information system, it provides an acceleration of data interchanges between DB in case of execution by an information system of new target tasks.

The offered approach will allow to gain the considerable economic effect:

- it is not necessary to solve the questions connected to incorrect placement in a DB of qualifiers and coding systems, therefore maintenance costs of an information system will decrease;
- logically connected information arrays from the data received from several DB, will be created automatically in case of extension of integrable information resources, it also will reduce maintenance costs of an information system;
- reduction of expenses by maintenance will be also at the expense of automation and a correctness of addition of integrable information resources in databases on any including network and double-sided topology of exchanges.

REFERENCES

[1] Devlin B. Data warehouse: from architecture to implementation. Addison Wesley Longman, Inc. (1997). ISBN 0-201-96425-2.

[2] IBM. «Business Intelligence Architecture on S/390. Presentation Guide». SG24-5747-00, IBM Corporation (2000).

[3] Vezenov V.I., Presnjakov A.N. Organizacija informacionnyh obmenov i integracija dannyh v raspredelennyh integrirovannyh avtomatizirovannyh sistemah upravlenija tehnologicheskimi processami. [The organization of information exchanges and data integration in the distributed integrated automated process control systems]. – RU: Cifrovaja obrabotka signalov, №3, 2010, pp. 9-12.

[4] Vezenov V.I., Novikov Ju.A., Presnjakov A.N., Svetnikov O.G., Hlebnikov N.Ju. Sposob informacionnogo obmena mezhdu bazami dannyh informacionnyh sistem i sistema dlja ego osuwestvlenija. [Method of information exchange between databases of information systems, and system for its implementation]. R.F. Patent RU 2351010 C1, 2007.

[5] Bisterfeld O.A. Sposob informacionnogo obmena mezhdu bazami dannyh informacionnyh sistem i sistema dlja ego osuwestvlenija. [Method of information exchange between databases of information systems, and system for its implementation]. R.F. Patent RU 2447495 C1, 2011.