

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
6 October 2005 (06.10.2005)

PCT

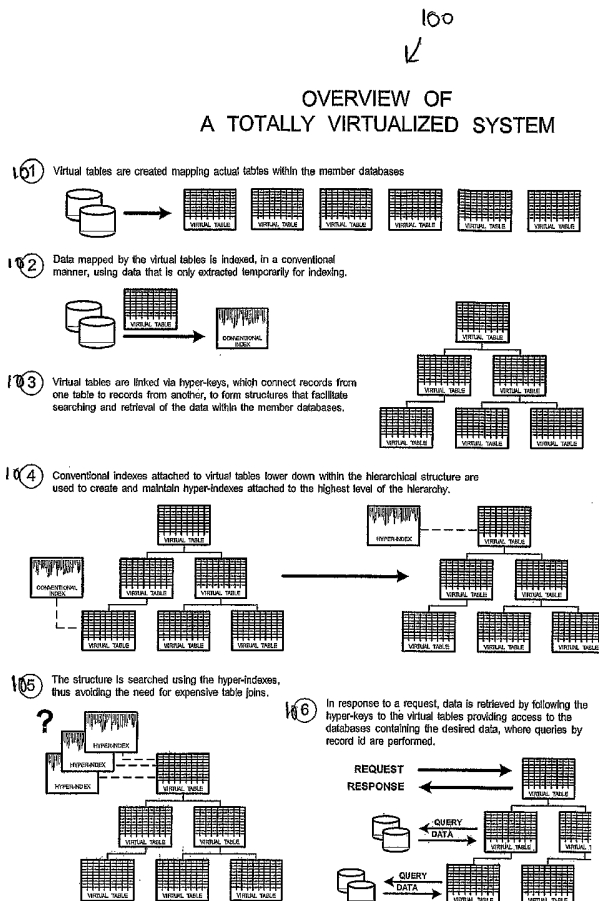
(10) International Publication Number
WO 2005/092032 A2

- (51) International Patent Classification: **Not classified**
- (21) International Application Number: PCT/US2005/009354
- (22) International Filing Date: 22 March 2005 (22.03.2005)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 60/554,935 22 March 2004 (22.03.2004) US
- (71) Applicant (for all designated States except US): **SLIC-CWARE CORPORATION** [US/US]; 11809 Coopers Ct., Reston, Virginia 20191-2314 (US).
- (72) Inventor; and
- (75) Inventor/Applicant (for US only): **FOSTER, Randall, J** [US/US]; 11809 Coopers Ct., Reston, Virginia 20191-2314 (US).
- (74) Agent: **GROSSMAN, Jon, D.**; Dickstein Shapiro Morin & Oshinsky LLP, 2101 L Street, N.W., Washington, DC 20037-1526 (US).
- (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI,

[Continued on next page]

(54) Title: SECURE VIRTUAL DATA WAREHOUSING SYSTEM AND METHOD

OVERVIEW OF A TOTALLY VIRTUALIZED SYSTEM



(57) Abstract: The invention advances the technology of virtual data warehousing: by allowing data to be organized in any of a number of virtual structures without the need for replication; by allowing the data to be accessed through the virtual structure at speeds in excess of those for conventional methods; by allowing data to be searched at significantly higher speeds, using a new indexing technology that enables tables to be indexed by data external to themselves, thus removing the need for complex joins; and by providing a workable security model for compartmenting access to both the data and the indices used to support searching.

WO 2005/092032 A2



FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO,
SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN,
GQ, GW, ML, MR, NE, SN, TD, TG).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Published:

— *without international search report and to be republished upon receipt of that report*

SECURE VIRTUAL DATA WAREHOUSING SYSTEM AND METHOD**CROSS-REFERENCE TO RELATED APPLICATIONS**

[0001] This application claims the benefit of U.S. Provisional Application No. 60/554,935, entitled SECURE VIRTUAL DATA WAREHOUSING, filed March 22, 2004, the entire disclosure of which is incorporated herein by reference.

FIELD OF THE INVENTION

[0002] The invention relates generally to data processing methods and systems. Specifically, the invention relates to efficient and secure retrieval of data stored at a number of remote sites via a virtual data warehouse.

BACKGROUND OF THE INVENTION

[0003] Dealing with data distributed over multiple databases has long been a problem. As accessing the data from multiple databases has become more of a necessity, people have increasingly turned to data warehousing. Data warehousing provides a method in which the data from various databases could be replicated in a manner that allowed for the data to be treated as belonging to a single database. Further, if one had the resources, the new replicated structure was not required to be a relational database, but instead could be implemented as any kind of structure desired.

[0004] Data warehousing provided for a simpler and faster method for extracting results. As new data warehouse techniques and software became available, warehouses were designed in a manner that best fit the type of data requests that were expected. By abandoning the constraints of the relational database, far more efficient designs could be implemented. However, data warehousing is a very costly endeavor, and each reorganization of the data required complete replication – putting limits on how flexible the data warehouse could be. Unfortunately, for

many potential users, the cost of maintaining something the size of a data warehouse was prohibitive. Data warehouses were, and still are, a major undertaking.

[0005] In recent years a new technology, Virtual Data Warehousing, has attempted to allow multiple databases to be interrogated together without the need of constructing and maintaining a massive data warehouse. Instead of replicating all the databases within a single massive database, a virtual data warehouse provides a common access to all the databases, making the collection of independent databases appear as a data warehouse to the user. This greatly reduces infrastructure and maintenance costs; and the data does not have to be replicated. But, virtual data warehousing suffers from terribly poor response times, and major drains on the overall system that cannot be adequately anticipated due to unscheduled massive processing within the member databases.

[0006] An additional problem that has arisen involves dealing with many users attempting to access data from many sources. In this scenario there is a problem properly restricting access to certain data while granting access to other data. Conventional data warehouses require very specific software to be written to address this problem. Out of the box, however, there is no simple solution. Conventional virtual data warehouses rely on the security models of the member databases. But, these models break down as soon as any performance upgrades such as caching or indexing of data is done external to the member databases.

BRIEF SUMMARY OF THE INVENTION

[0007] The present invention overcomes the above-noted deficiencies by building upon current technologies for retrieving data from various sources, including conventional automated database login and retrieval, and remote and local document retrieval. Additionally, the invention builds upon current indexing technology, and is intended to be able to be integrated with any form of data base access, and proprietary or open source indexing methods by a team of appropriate experts.

[0008] The present invention greatly extends the capabilities of a conventional virtual data warehouse. The invention allows the user to benefit from the savings and simplified maintenance of a virtual data warehouse, while at the same time, experiencing the major performance gains, security, and simplicity of access that generally only accompany a very well designed and costly data warehouse. The invention makes use of hyper-indices – indices that are derived from data external to the tables they are attached to, hyper-keys – a means of directly linking records across tables without the use of run-time lookups, and virtual tables, a conventional means of mapping and accessing data from database tables or other sources such as file systems without the need for replication.

[0009] The invention allows the data held within the member databases to be organized, using virtual tables and hyper-keys, in any fashion that best supports user requests without permanently replicating that data. Multiple structures could be maintained within the same Secure Virtual Data Warehouse. Everything from spreadsheet style organization to complex hierarchical structures could be maintained, with the same data being reused over and over again. And no data replication need occur. If a given structure is no longer needed, it is destroyed. The data remains intact within the member databases. If a new structure is needed, it is created.

[0010] In addition, hyper-indices allow records mapped to one virtual table to be indexed by data accessed through a different virtual table, provided that the second virtual table has been linked to the first by hyper-keys – either directly or indirectly. In other words, if an index were placed on the age field in a table called “Personal Information”, and that table was linked to a table called “Executives”, which was linked to a table called “Companies”; a hyper-index could be created, which would allow the table “Companies” to be searched for businesses having executives under the age of 35 directly. The names and ages of those individuals could also be retrieved via the hyper-keys just as if the data belonged to the selected records in the

“Companies” table. Large, complex searches, which would normally require expensive joins and take large amounts of resources and time, are reduced by this invention to simple searches combining multiple indices on the same virtual table, which finish quickly.

[0011] The unscheduled massive processing within the member databases of a conventional virtual data warehouse is replaced by a simple extraction of the final results. And even that can be as simple as extracting fields of data from records selected by record ID – the least impact possible.

[0012] Finally, the problem of securing access to the individual data sources is accomplished by attaching a bit map of user groups, which are authorized to have access, to each virtual table. Each virtual table is, in turn, mapped to a data source sharing the same access restrictions. Access to data via a path of hyper-keys is restricted by each of the tables containing a hyper-key from the path. Access to a conventional index is restricted by the virtual table it is applied to. Access to a hyper-index is restricted by the virtual table it is applied to, the virtual table from which the data was indexed, and each table containing a hyper-key in the path between the two virtual tables.

BRIEF DESCRIPTION OF THE DRAWINGS

[0013] FIG. 1 is an overview flow chart of the virtualized system forming the present invention;

[0014] FIG. 2 is a flow chart that depicts the steps involved in registering a table and the associated primary hyper-key;

[0015] FIG. 3 is a flow chart that depicts the steps involved in registering a secondary hyper-key;

[0016] FIG. 4 is a flow chart showing the steps involved in registering a hyper-key link;

[0017] FIG. 5 is a flow chart illustrating the steps involved in registering a conventional index so that it may be used by the system to create one or more hyper indexes;

[0018] Figure 6. is a flow chart illustrating the steps involved in building a hyper-index descriptor;

[0019] Figures 7A-7B are flow charts illustrating respectively the steps for retrieving an external ID for a component that has been registered and for creating a hyperlink path as part of a hyper-index descriptor;

[0020] Figure 8. Is a flow chart that illustrates the steps involved in registering a hyper-index;

[0021] Figure 9. is a flow chart that shows the steps involved in updating a hyper-index; and

[0022] Figures 10-A and 10-B are respectively flow chart diagrams for filling an update request tree and for updating the individual hyper-index data sections.

DETAILED DESCRIPTION OF THE INVENTION

[0023] Overview. This invention is further evolves the virtual data warehouse. A virtual data warehouse can be thought of as a collection of virtual tables. Each virtual table is representative of a table within an actual database, and may actually contain records which correspond to records within the actual table, and may even contain some cached fields. Accessing the virtual table generally causes the system to access the actual table within the database. More efficient virtual data warehouses may actually, cache and even index some of the data from the member databases. This invention takes this caching and indexing in a different direction allowing virtual data warehouses to achieve a whole new level of flexibility, efficiency, and security.

[0024] This invention allows the data from the member databases to be reorganized in any number of ways by simply linking virtual tables together. Virtual tables are linked together via hyper-keys, which individually link one or more records within a virtual table to one or more records within a different virtual table. Hyper-keys are similar to keys within a conventional database, but have some important differences. A hyper-key can best be thought of as a direct

link from a record in one virtual table to one or more records in another table. Hyper-keys, are of two types: foreign hyper-keys which are used to access records from a different virtual table; and local hyper-keys which are exported to allow other virtual tables to access records from the virtual table they reside in. Further, two types of local hyper-keys exist: primary hyper-keys, which have a one-to-one correlation of values to records and can be used to identify a single record; and secondary hyper-keys, which can have a one-to-many correlation of values to records. Exactly one primary hyper-key is assigned to each virtual table. For that reason, a primary hyper-key can always be used to locate an individual record. Finally, a field containing a foreign hyper-key can be exported as a link to allow a way for one virtual table to link through a second virtual table to a third virtual table, and so on.

[0025] Unlike conventional database keys, which can be used to access records from any number of tables, each foreign hyper-key must be identified with a very specific virtual table and the field where the matching hyper-key exists. This allows the hyper-keys to be resolved as soon as data is entered into the system, and to be replaced by hyper-IDs within the data. A hyper-ID, which is either a record ID (in the case of a primary hyper-key) or a virtual link to a list of record IDs (in the case of a secondary hyper-key) provides for very fast processing of any indirection. For simplicity, we will refer to the hyper-ID associated with the primary hyper-key as the record ID. There is no reason why conventional database keys and hyper-keys cannot exist within the same Secure Virtual Data Warehouse. In order to extend the capabilities of hyper-keys and allow for a truly hierarchical structure, the invention uses special virtual tables that have no corresponding actual tables, but only contain hyper-keys. These special virtual tables can be used as nodes within the desired structure. By combining these special virtual tables with virtual tables that actually map data, various desirable data structures can be obtained.

[0026] To provide for a more efficient means of access while still keeping the original data normalized and within the original data structures, the invention provides a new structure, the hyper-index. A hyper-index allows for data from one virtual table to be indexed by data from a different virtual table. A hyper-index, once built, is used just as any other index. The difference between a hyper-index and a conventional index is the location of the data being indexed, and in the maintenance of the hyper-index. An index is created to provide fast access to individual records through a specific virtual table. With a conventional index, the data used to create that index is data from the records retrieved through the virtual table for which the index is being created. With a hyper-index, the data is retrieved through a different virtual table whose have been linked by hyper-keys to the records of the virtual table where the index is being applied. In order for records in one virtual table to be indexed by data from another virtual table, there must be a direct path of hyper-keys from the virtual table to be indexed to the virtual table supplying the data for the index.

[0027] Restricting access to data within a virtual data warehouse using hyper-indexing is carried out in a very simple way. Each virtual table has attached to it an Access Mask -- a bit map of user groups that have access to the table. Each request for data has attached to it a Group Mask -- a bit map of user groups the requestor belongs to. For the requestor to be granted access, at least one bit that is set active in the Access Mask must also be set active in the Group Mask. This is true for each virtual table. Accordingly, if an individual had access to a virtual table containing a foreign hyper-key, but did not have access to the virtual table the foreign hyper-key pointed to, he would only retrieve results from the first virtual table. Any request to retrieve results through the foreign hyper-key would be denied.

[0028] A conventional index applied to data accessed through a virtual table inherits its access restrictions from that virtual table. A hyper-index inherits its access restrictions from the virtual table it is applied against, the virtual table providing the data for the conventional index it

is using as a basis, and each of the virtual tables providing the links between those two tables. For an individual to be able to search a hyper-index, he must have access privileges to each of those tables.

[0029] Incorporating data updates into the system is a two-pass process. In the first pass, records within the virtual tables are inserted, updated, or deleted; and conventional indices are adjusted accordingly. In the second pass, hyper-indices are adjusted to reflect the changes made in the first pass.

[0030] Referring to Fig. 1, an overview 100 of the totally virtualized system is shown. Specifically this figure shows an overview of a totally virtualized system, in which no data is stored outside the member databases. Steps 101 through 104 show the major steps in preparing the system for access. In step 101 virtual tables are created mapping actual tables within the member databases. At step 102 data mapped by the virtual tables is indexed (in any conventionally known manner) using data that is only extracted temporarily for indexing. At step 103 virtual tables are linked via hyper-keys which act to connect records from one table to records from another in order to form structures that facilitate searching and retrieval of the data within the member databases. At step 104 conventional indexes are attached to virtual tables that are lower down within the hierarchical structure and are used to create and maintain hyper-indices attached to the highest level of the hierarchy.

[0031] Steps 105 and 106 are the major steps involved in retrieving data. At step 105 the structure is searched using the hyper-indices thus avoiding the need for expensive table joins. At step 106 in response to a request, data is retrieved by following the hyper-keys to the virtual tables providing access to the databases containing the desired data, where queries by record ID are performed. The strict hierarchical structure is simply a representation of one of many possible structures that could be employed within the system. Also, certain data could also be

held outside the member databases, if desired, to facilitate fine tuning the search request before accessing the member databases.

[0032] Creating a Virtual Table. At the time the virtual table is created, each field is assigned a field id, and a special field, the identity field, is set aside for the identity value used to identify each record uniquely during update. Each table, and the associated primary hyper-key must be registered. (See Figure 2.) The table name will be used as the lookup name to search the global external name lookup structure. This same table name will also serve as the primary hyper-key name for the table.

[0033] As part of the registration process an external id is found, or generated and stored, within the global external name lookup structure under the lookup name. If the external id was not found but needed to be generated, the following actions are performed: (1) A hyper-key lookup structure, pointed to by the generated external id (step 203b), is created. That structure will be used to resolve hyper-key values for the primary hyper-key into the appropriate record ids. (2) The group access list to be associated with this table is added to the global group access list array structure Step 204. (3) The Modified Records List Structure is created, which will be used to identify modified records so that hyper-indices may be updated accordingly Step 205.

[0034] The external id, generated by the registration process and returned to the calling module, will be used as both, the table id and the primary hyper-key id step 206. As the table id, it will be used in any future access of the table and any structures, either local to the table or external lookup structures, containing information relevant to the table. Upon return from the lookup, the table id is stored along with the other table information. As the primary hyper-key id, it will be used to facilitate rapid location of the proper lookup structure for dereferencing the primary hyper-key for any record within this table, and returning a record id. The identity value for each record will be used as the primary hyper-key.

[0035] Referring now to Figure 2, this Figure shows the steps 200 involved in registering a table and the associated primary hyper-key. This process needs to be available to both the table creation and registering process and the foreign hyper-key creation and registering process to avoid requiring a specific order of table creation, and possibly a dead-lock. Steps 201 and 202 are carried out as follows. Step 201 involves finding a table name within the external name lookup structure or inserting it into the external name lookup structure and identifying the external type as a secondary hyper-key. Step 202 then retrieves the external key. If during step 201, the name is found within the lookup structure and this is a table create request, an error is returned at step 203a. If during step 201, the name is found within the lookup structure and this is not a table create request, the external id retrieved in step 202 is returned as both the table id and the primary hyper-key id. If during step 201, the name is not found within the lookup structure, steps 203b, 204 and 205 are carried out and the external id retrieved in step 202 is returned as both the table id and the primary hyper-key id.

[0036] Registering a Secondary Hyper-Key. Each secondary hyper-key must be registered. (See Figure 3.) The secondary hyper-key is registered using the table name, table id, and field name for the location where the secondary hyper-key resides. The table name and field name are mangled to produce a unique name. This unique name will be used as the lookup name to search the global external name lookup structure.

[0037] As part of the registration process an external id is found, or generated and stored, within the global external name lookup structure under the lookup name. If the external id was not found but needed to be generated, the following actions are performed: (1) A hyper-key lookup structure, pointed to by the generated external id, is created. That structure will be used to resolve hyper-key values for the secondary hyper-key into the appropriate hyper-ids. (2) A list array structure that may be accessed using the hyper-ids returned by the newly created A hyper-key lookup structure is created. That structure will be used to hold lists of record ids for records

that are supposed to be accessed by the secondary hyper-key associated with the hyper-id pointing to the list. (3) The table id is added to the global parent table array structure, which is used to identify the parent tables for all registered externals, at the location determined by the external id. Finally, the external id is returned so that it may be associated with the field within the table structure.

[0038] Specifically Figure 3 shows the steps involved in registering a secondary hyper-key 300. This process needs to be available to both the secondary hyper-key creation and registering process and the foreign hyper-key creation and registering process to avoid requiring a specific order of table creation, and possibly a dead-lock. Steps 301, 302 and 303 are carried out as shown. If during step 302, the name is found within the lookup structure and this is a key create request, an error is returned at step 304a. If during step 302, the name is found within the lookup structure and this is not a key create request, the external id retrieved in step 303 is returned as the secondary hyper-key id. If during step 302, the name is not found within the lookup structure, steps 304b, 305, and 306 are carried out and the external id retrieved in step 303 is returned as the secondary hyper-key id.

[0039] Registering a Foreign Hyper-Key Each foreign hyper-key field must also be registered in the same way as the hyper-key it is intended to access. If it is intended to access a primary hyper-key, it must register the name of the table it is intended to access. The registration proceeds just as with registering the table. (See Figure 2.) If it is intended to access a secondary hyper-key, it must register the name of the secondary hyper-key field along with the name of the table it is intended to access. The registration proceeds just as with registering the secondary hyper-key. (See figure 3.)

[0040] Registering a Hyper-Key Link. Finally, foreign hyper-keys within a virtual table, which are to be used as hyper-key links must be registered as such as shown by process steps 400. (See Figure 4.) The hyper-key link is registered using the table name, table id, field name,

and field id for the location where the foreign hyper-key resides, and the target hyper-id that was returned when the foreign key field was registered. The table name and index name are mangled to produce a unique name. This unique name will be used as the lookup name to search the global external name lookup structure.

[0041] As part of the registration process an external id is found or generated and stored within the global external name lookup structure under the lookup name. If the external id was not found but needed to be generated, the following actions are performed: (401) The target hyper-id is added to the global link target array structure (step 404b), at the location determined by the external id. (2) The table id is added to the global parent table array structure step 405, which is used to identify the table exporting the link information, at the location determined by the external id. (3) The field id is added to the global link field array structure step 406, which is used to identify the field exporting the link information, at the location determined by the external id. Finally, the external id is returned Step 407 so that it may be associated with the index as the index id.

[0042] Registering a Conventional Index. Any conventional index that is to be used as part of a hyper-index must be registered. A conventional index is one that is created on a virtual table using data within that table, which may have been extracted from the data source being mapped by that virtual table. The data may be maintained permanently within that virtual table, or only exist for the period of time during which it is being accessed.

[0043] Referring now to Figure 5, the conventional index is registered as shown by process steps 500 using the table name, table id for the table that the index is attached to, and the index name, index type, and index descriptor. The descriptor contains all the information necessary for allowing portions of the index, based upon the record id, to be extracted and placed into another index – a hyper-index. Both, the descriptor information, and the process of moving indexing structures and data between indexes are unique to the individual indexing algorithms as is

conventionally known in the art. The table name and index name are mangled at step 510 to produce a unique name. This unique name will be used as the lookup name to search the global external name lookup structure.

[0044] As part of the registration process an external id is found at step 502, or generated and stored, within the global external name lookup structure under the lookup name. If the external id was not found but needed to be generated as shown at step 503, the following actions are performed: (1) The descriptor information for the conventional index is added to the global index descriptor array structure at step 504b, at the location determined by the hyper-id, so that one or more hyper-indexes may be created using information from the conventional index. (2) The type of the index: keyword, range, timeline, grid, etc., is added to the global index type array structure, at the location determined by the hyper-id at step 505. (3) At step 506, the table id is added to the global parent table array structure, which is used to identify the parent tables for all registered externals, at the location determined by the external id. Finally, the external id is returned at step 507 so that it may be associated with the index as the index id.

[0045] Creating a Hyper-Index. Referring to Figure 6, at the time a hyper-index is created, all components to be used in the hyper-index must already exist. The virtual table upon which the hyper-index is to be applied must have a foreign hyper-key that points either directly or indirectly to the virtual table attached to the conventional index. That is the foreign hyper-key must be a primary or secondary hyper-key from the virtual table attached to the conventional index, or it must point to the first hyper-key link in a chain of one or more hyper-key links (exported foreign hyper-keys) that terminate by pointing to the virtual table attached to the conventional index.

[0046] The first step of process 600 in creating the hyper-index is to vary the path described in the previous paragraph. Once that has been done, a descriptor is created at step 601 that identifies the conventional index involved, as well as each step in the chain of hyper-key links

from the initial foreign hyper-key to the final primary or secondary hyper-key exported by the virtual table attached to the conventional index.

[0047] The second step is to attach to the descriptor a list of Access Masks (step 603), a copy of the one for the virtual table it is attached to, a copy of the one for the virtual table providing the data for the index, and copies of the Access Masks for each of the virtual tables containing a foreign hyper-key used as a hyper-key link in the chain from the hyper-index to the conventional index.

[0048] Next, the index descriptor is added to the global Index Descriptor Array structure, and the table id of the parent virtual table, over which the hyper-index has been created, is added to the global Parent Table Array structure – both using the index id as the access point.

[0049] The final step in creating the hyper-index is to identify all the dependencies of the hyper-index. A Record Dependency Tree is created for identifying the dependency of individual sections of the hyper-index, which are identified by the record ids for records within the parent virtual table, to records existing in other virtual tables. The entries are entered and accessed by the table id and record id of the foreign record upon which the section is dependent. A Hyper-Index Dependency Map is created for each hyper-index. Within that map, a bit is assigned for each virtual table in the system. The map has all bits set to inactive except for the bit corresponding to the virtual table attached to the hyper-index, the bit corresponding to the virtual table attached to the conventional index, and the bits corresponding to each virtual table containing a foreign hyper-key used as a hyper-key link in the chain from the hyper-index to the conventional index.

[0050] Referring now to Fig. 7A, this figure shows the steps involved in retrieving an external id for a component that has been registered. Steps 701, 702, and 703 are carried out whereby step 701 creates a unique name by combining the table name with the field or index name. If during step 702, the name is found within the lookup structure, the external id retrieved

in step 703 is returned. If during step 702, the name is not found within the lookup structure, an error code is returned at step 704b.

[0051] In Figure 7-B, the steps involved in creating a hyper-key link path 750 as part of a hyper-index descriptor are illustrated. For each hyper-key link in the source hyper-key link path, the link id is retrieved at step 751 using the table and field names where the hyper-key exists, and added to the resulting hyper-key link path at 752. The path is then terminated with a zero at step 753 and returned to the calling routine at 754.

[0052] Modifying Virtual Tables. A virtual table is modified in direct response to changes in the actual table within the member database. The changes can be to data the virtual table is caching or to data being indexed by a conventional index attached to the virtual table, or to local or foreign hyper-key values.

[0053] Any time a record is accessed in the update process, the primary hyper-key, or identity value, is automatically exported and dereferenced to retrieve the record id. If a record is inserted, the primary hyper-key is added to the lookup structure. If a record is deleted, the primary hyper-key is removed from the lookup structure. However, a deleted hyper-key may not be reused until the next time the virtual table is being updated.

[0054] If a new or updated record contains one or more fields with a foreign hyper-key, each foreign hyper-key is dereferenced using the hyper-key lookup structure for the specific hyper-key, and the resulting hyper-id stored in its place within the record. Next, if the foreign hyper-key is to be exported as a hyper-key link, the hyper-id from dereferencing the foreign hyper-key is stored in the lookup structure for the hyper-key link, accessed by the record id.

[0055] Finally, if one or more fields contain a secondary hyper-key for export, each secondary hyper-key is exported in the following way. (1) The hyper-key string is located within the lookup structure for the hyper-key, which returns the id of a list of record ids identifying

records associated with the secondary hyper-key. (2) The record id for the record being modified is added to that list.

[0056] Each time a virtual table is modified, the appropriate bit is set in the Updated Virtual Table Map structure. Additionally, the record id of each record modified or deleted is added to the Modified Records List for that virtual table.

[0057] Updating Conventional Indexes. Each time a record is updated within a virtual table, any conventional indexes attached to that virtual table must be updated as well.

[0058] Each time a conventional index is modified due to changes in a record, the information from the index relating to that record must be exported. The actual information that is exported will depend on the type of index method being used. The important point is that whatever information is necessary for inserting the record into the actual index be exported. That information is added to the Exported Index Updates List for that index -- a list of index update information that can be accessed directly by record id, and will be used in updating hyper-indexes based upon the conventional index.

[0059] Figure 8 shows the steps involved in registering a hyper-index 800. Steps 801, 802, and 803 are carried out as follows: at step 801 a unique name is created by combining the table name with the field name for the index. If during step 802, the name is found within the lookup structure, the index is being doubly defined and an error is returned at step 804a. Otherwise, steps 804b, and 805 through 808 are carried out as described in the Figure and the external id retrieved in step 803 is returned as the index id.

[0060] Updating Hyper-Indexes. As shown in Fig. 9, once all virtual tables and conventional indexes have been modified, it is time to update the hyper-indexes through process 900. Whether or not an individual hyper-index needs to be updated is determined by comparing the Hyper-Index Dependency Map for the hyper-index with the Updated Virtual Table Map at step 902. If

it occurs at least once that a bit set in the Hyper-Index Dependency Map is also set in the Updated Virtual Table Map, the hyper-index must be processed at 902a.

[0061] Determining which record entries must be updated within the hyper-index is more complicated. This requires the use of the Record Dependency Tree for the hyper-index, and the Modified Records List for each virtual table for which bits were set in both the Hyper-Index Dependency Map and the Updated Virtual Table Map tested in the preceding paragraph. The record ids for the record entries that must be updated within the hyper-index are stored a structure known as the Update Request Tree. A new Update Request tree is created each time a hyper-index is to be updated and deleted after the update is completed. The Record Dependency Tree is a lookup structure that stores a list of local record ids for sections of the hyper-index that are dependent on a specific record within some other virtual table. The list is stored and accessed by a combination of remote virtual table id and remote record id.

[0062] The process of filling the Update Request Tree is shown through the following steps: (1) All record ids from the Modified Records List for the virtual table attached to the hyper-index are loaded into the Update Request Tree at step 1002a. (2) Each record id in the Modified Records List for each of the modified virtual tables upon which the hyper-index is dependent is used to locate a possible list of record ids within the Record Dependency Tree at step 1002b. (3) If a match is found, all record ids in the list found are added to the Update Request Tree and the matching entry is removed from the Record Dependency Tree at step 1003.

[0063] Once the Update Request Tree has been properly filled, the hyper-index data for each local record id found in the Update Request Tree is modified in the following way: (1) Any data in the hyper-index for that local record id is removed. (2) The hyper-key path is followed from virtual table to virtual table until the final virtual table in the path is encountered and the remote record ids for all the relevant sections of the conventional index associated with it and the hyper-index have been identified. If only primary hyper-keys were included in the path, only data for

one remote record id would need to be imported. However, if one or more hyper-keys were secondary, data from multiple remote record ids may be required. (3) Remote record id by remote record id, the exported sections of the conventional index are extracted from the Exported Index Updates List for that index and inserted into the section of the hyper-index associated with the local record id.

[0064] In order to maintain this system, the Record Dependency Tree must be kept up to date. To that end, each time a remote record id is encountered while following the hyper-key path, including the record ids within the final virtual table, and all remote record ids associated with hyper-key links in the hyper-key path; the local record id for which the data is being imported is added to a list of local record ids pointed to by the remote virtual table id and the remote record id within the Record Dependency Tree.

[0065] Once all hyper-indexes have been processed, the Updated Virtual Table Map is reset to all zeros and the Modified Records List for each virtual table is cleared.

[0066] Searching Indexes. Searching any index, whether a conventional index or a hyper-index is carried out the same way. However, access to the index differs between the two index types. For a conventional index, the users Group Mask is compared against the Access Mask attached to the index. If it occurs at least once that a bit set in the Group Mask is also set in the Access Mask, then access to the index is granted. For a hyper-index, the user Group Mask is tested against each of the Access Masks attached to the hyper-index. For each Access Mask tested, If it occurs at least once that a bit set in the Group Mask is also set in the Access Mask then that Access Mask passes the test. If every Access Mask passes the test, access is granted. If even one Access Mask does not pass the test access is denied. If access is granted the search is carried out.

[0067] Retrieving Data. Retrieving data, either from cached fields within virtual tables or fields within tables mapped by virtual tables is the same as with a conventional virtual data,

warehouse after access has been granted to the virtual table, unless hyper-key retrieval is involved. If no hyper-key retrieval is involved, the users Group Mask is compared against the Access Mask attached to the virtual table. If it occurs at least once that a bit set in the Group Mask is also set in the Access Mask, then access to the virtual table and the cached data it holds and the actual table it maps is granted.

[0068] In a hyper-key retrieval, the user identifies a sequence of hyper-keys to follow from virtual table to virtual table until, in the end, he identifies the field or fields to extract from the cached data and/or the mapped database table. As with building a hyper index, if all the hyper-keys are primary, only data from one record will be returned. However, if one or more of the hyper-keys is secondary, data from multiple records could be returned. To gain access to the data the Group Mask must be tested against the Access Mask for each virtual table being accessed while following the hyper-keys. For each Access Mask tested, If it occurs at least once that a bit set in the Group Mask is also set in the Access Mask then that Access Mask passes the test. . If every Access Mask passes the test, access is granted. If even one Access Mask does not pass the test access is denied. If access is granted the data retrieval is carried out.

CLAIMS

What is claimed as new and desired to be protected by Letters Patent of the United States is:

1. A virtual data warehouse, comprising:

a plurality of virtual tables each comprising at least one virtual record with accompanying virtual fields corresponding to at least one record and fields located in at least one remote database table or file system;

a plurality of local hyper-keys identifying unique virtual fields within the plurality of virtual tables; and,

a plurality of foreign hyper-keys that link virtual records to the plurality of local hyper-keys.

2. The virtual data warehouse as described in claim 1, wherein each of the plurality of foreign hyper-keys identifies both a specific local hyper-key of the plurality of local hyper-keys and a specific virtual table of the plurality of virtual tables, the specific local hyper-key located on the specific virtual table.

3. The virtual data warehouse as described in claim 2, wherein the plurality of local hyper-keys includes both primary hyper-keys that identify a single unique virtual field, and secondary hyper-keys that identify multiple virtual fields that are collectively unique.

4. The virtual data warehouse as described in claim 3, wherein each of the plurality of virtual tables comprises only one primary hyper-key.

5. The virtual data warehouse as described in claim 2, further comprising:

a plurality of virtual linking tables, each comprising at least one local hyper-key and one foreign hyper-key, but containing no virtual records, each of the plurality of virtual linking tables configured as a link between the plurality of foreign hyper-keys on ones of the plurality of virtual tables and the plurality of local hyper-keys on either other ones of the plurality of virtual tables or other ones of the plurality of virtual linking tables.

6. The virtual data warehouse as described in claim 2, further comprising:

a plurality of hyper-indices, each comprising one primary hyper-key and one foreign hyper-key so as to directly link the at least one virtual record from a virtual table of the plurality of virtual tables to one of the plurality of local hyper-keys on another virtual table of the plurality of virtual tables.

7. The virtual data warehouse as described in claim 2, wherein the plurality of foreign hyper-keys may be replaced by hyper-IDs which directly reference the at least one record and fields linked via local hyper-keys and virtual records and virtual fields.

8. The virtual data warehouse as described in claim 2, wherein each of the plurality of virtual tables further comprise an access mask configured to indicate a plurality of user groups that have access to the plurality of virtual tables.

9. The virtual data warehouse as described in claim 8, wherein requests for data are accompanied by a group mask indicating from which ones of the plurality of user groups the requests originated,

10. The virtual data warehouse as described in claim 9, wherein access to a virtual table is only granted if a match occurs between an associated access mask and a requesting group mask.

11. The virtual data warehouse as described in claim 6, wherein each of the plurality of hyper-indices contains information concerning access rights to each of the plurality of virtual tables linked by the plurality of hyper-indices.

12. A virtual data warehouse, comprising:

storage means to store a plurality of virtual records with accompanying virtual fields, the plurality of virtual records corresponding to a plurality of records and accompanying fields located in at least one remote database table or file system;

identifying means to identifying unique virtual fields stored by the storage means; and,

~~linking means to link the plurality of virtual records to the unique virtual fields identified~~
by the identifying means.

13. The virtual data warehouse as described in claim 12, wherein the linking means is configured to link a specific virtual record of the plurality of virtual records to a unique virtual field identified by the identifying means and stored by an identified storage means.

14. The virtual data warehouse as described in claim 13, wherein the identifying means is configured to identify both a single unique virtual field and multiple virtual fields that are collectively unique.

15. The virtual data warehouse as described in claim 14, wherein each of the each of the specific storage means comprises only one unique virtual field identified by the identifying means.

16. The virtual data warehouse as described in claim 13, further comprising:

indexing means to index ones of the plurality of virtual records to the unique virtual fields identified by the identifying means.

17. The virtual data warehouse as described in claim 13, further comprising:

security means to restrict user access to the plurality of virtual records.

18. The virtual data warehouse as described in claim 17, wherein the security means include a comparison of user group data submitted with an access request to access data stored by the storage means.

19. The virtual data warehouse as described in claim 16, wherein the indexing means is configured to contain access rights for each of the plurality of virtual records.

20. A computer program product, comprising:

a computer useable medium upon which computer program logic is recorded, the computer program logic configured to create a virtual data warehouse comprising:

a plurality of virtual tables each comprising at least one virtual record with accompanying virtual fields corresponding to at least one record and fields located in at least one remote database table or file system;

a plurality of local hyper-keys identifying unique virtual fields within the plurality of virtual tables; and,

a plurality of foreign hyper-keys that link virtual records to the plurality of local hyper-keys.

21. The computer program product as described in claim 20, wherein the computer logic is configured to ensure that each of the plurality of foreign hyper-keys identifies both a specific local hyper-key of the plurality of local hyper-keys and a specific virtual table of the plurality of virtual tables, the specific local hyper-key located on the specific virtual table.

22. The computer program product as described in claim 21, wherein the computer logic is configured to ensure that the plurality of local hyper-keys includes both primary hyper-keys that identify a single unique virtual field, and secondary hyper-keys that identify multiple virtual fields that are collectively unique.

23. The computer program product as described in claim 22, wherein the computer logic is configured to ensure that each of the plurality of virtual tables comprises only one primary hyper-key.

24. The computer program product as described in claim 21, the computer program logic configured to create a virtual data warehouse further comprising:

a plurality of virtual linking tables, each comprising at least one local hyper-key and one foreign hyper-key, but containing no virtual records, each of the plurality of virtual linking tables configured as a link between the plurality of foreign hyper-keys on ones of the plurality of virtual tables and the plurality of local hyper-keys on either other ones of the plurality of virtual tables or other ones of the plurality of virtual linking tables.

25. The computer program product as described in claim 21, the computer program logic configured to create a virtual data warehouse further comprising:

a plurality of hyper-indices, each comprising one primary hyper-key and one foreign hyper-key so as to directly link the at least one virtual record from a virtual table of the plurality of virtual tables to one of the plurality of local hyper-keys on another virtual table of the plurality of virtual tables.

26. The computer program product as described in claim 21, wherein the computer logic is configured to ensure that the plurality of foreign hyper-keys may be replaced by hyper-IDs which directly reference the at least one record and fields linked via local hyper-keys and virtual records and virtual fields.

27. The computer program product as described in claim 21, wherein the computer logic is configured to ensure that each of the plurality of virtual tables further comprise an access mask configured to indicate a plurality of user groups that have access to the plurality of virtual tables.

28. The computer program product as described in claim 27, wherein the computer logic is configured to ensure that requests for data are accompanied by a group mask indicating from which ones of the plurality of user groups the requests originated,

29. The computer program product as described in claim 28, wherein the computer logic is configured to ensure that access to a virtual table is only granted if a match occurs between an associated access mask and a requesting group mask.

30. The computer program product as described in claim 25, wherein the computer logic is configured to ensure that each of the plurality of hyper-indices contains information concerning access rights to each of the plurality of virtual tables linked by the plurality of hyper-indices.

100
↓

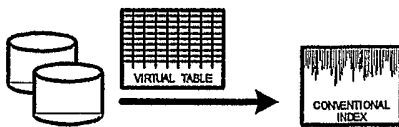
Figure 1

OVERVIEW OF A TOTALLY VIRTUALIZED SYSTEM

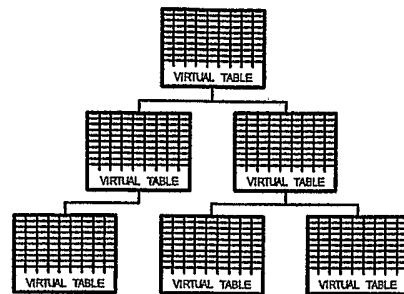
101 Virtual tables are created mapping actual tables within the member databases



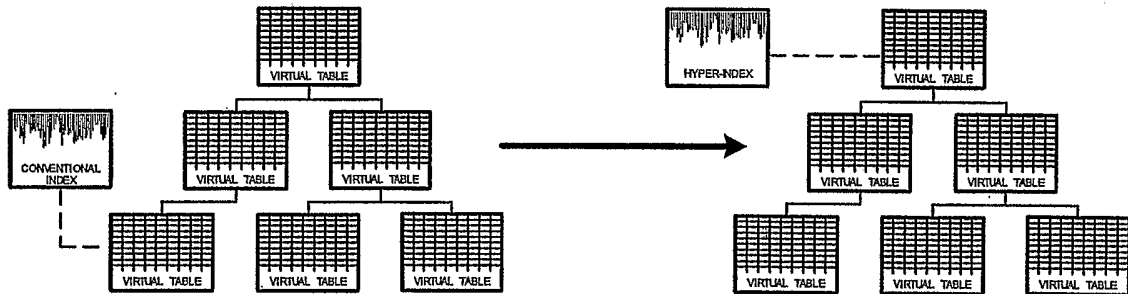
102 Data mapped by the virtual tables is indexed, in a conventional manner, using data that is only extracted temporarily for indexing.



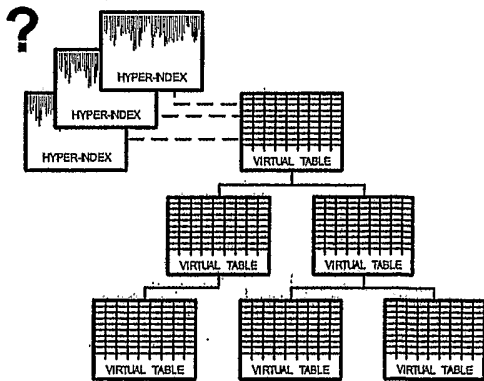
103 Virtual tables are linked via hyper-keys, which connect records from one table to records from another, to form structures that facilitate searching and retrieval of the data within the member databases.



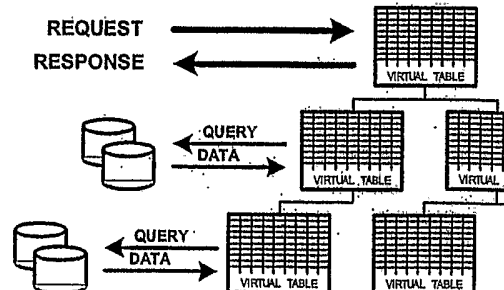
104 Conventional indexes attached to virtual tables lower down within the hierarchical structure are used to create and maintain hyper-indexes attached to the highest level of the hierarchy.



105 The structure is searched using the hyper-indexes, thus avoiding the need for expensive table joins.



106 In response to a request, data is retrieved by following the hyper-keys to the virtual tables providing access to the databases containing the desired data, where queries by record id are performed.



+

Figure 2

REGISTERING A TABLE AND THE ASSOCIATED PRIMARY HYPER-KEY

① Find the table name within the external name lookup structure or insert it into the external name lookup structure and identify the external type as a secondary hyper-key.

② Retrieve external id .

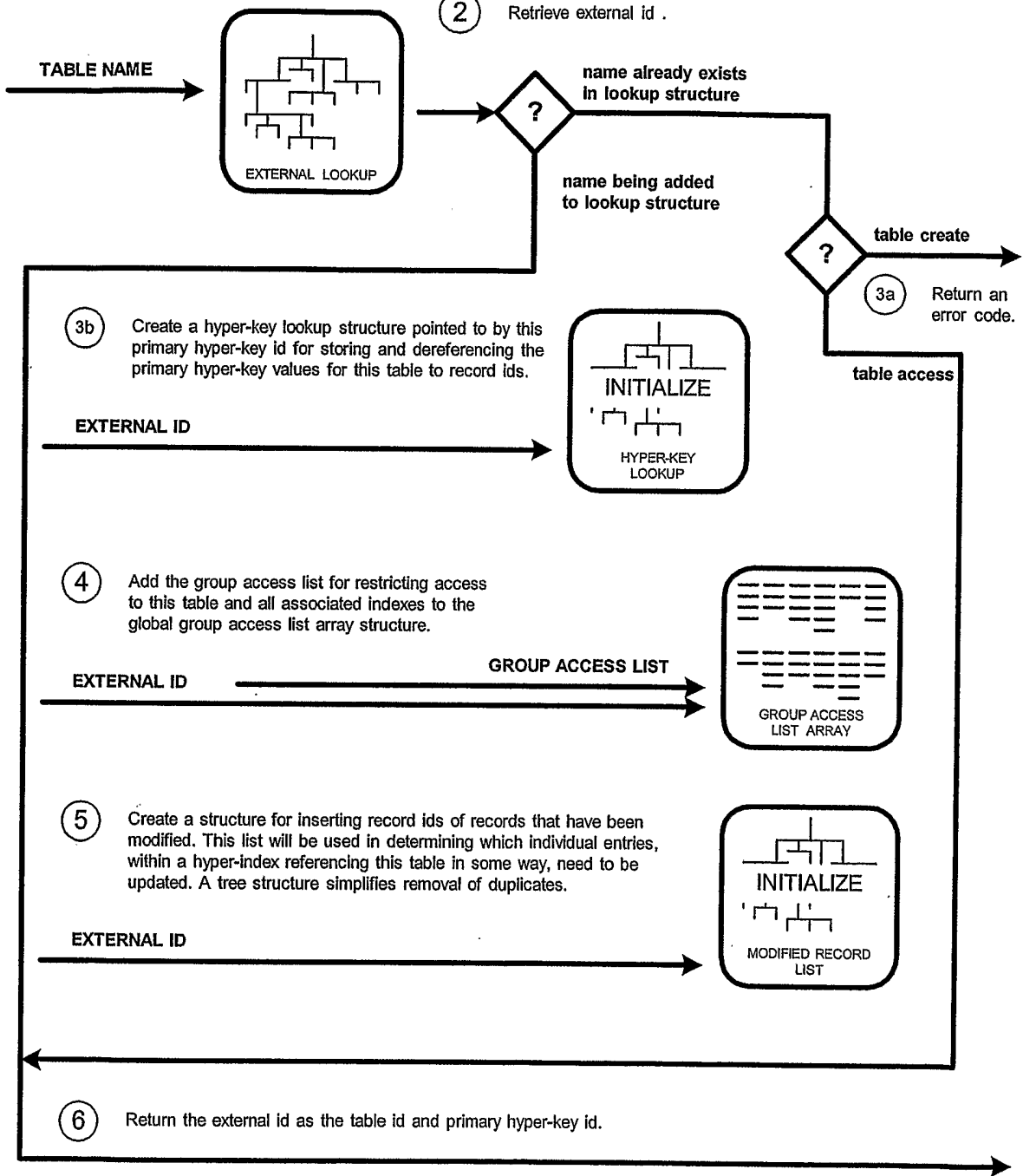


Figure 3 REGISTERING A SECONDARY HYPER-KEY

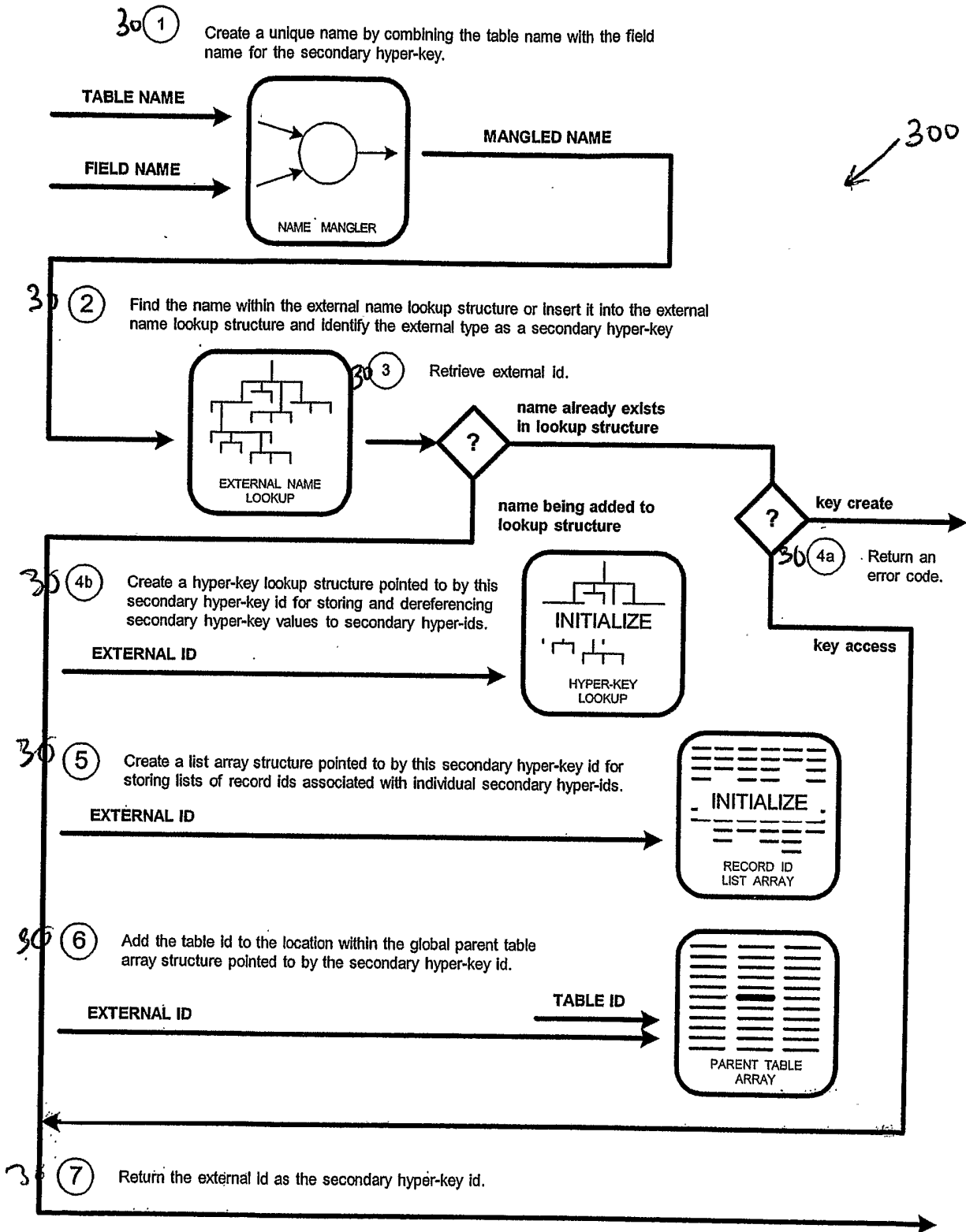
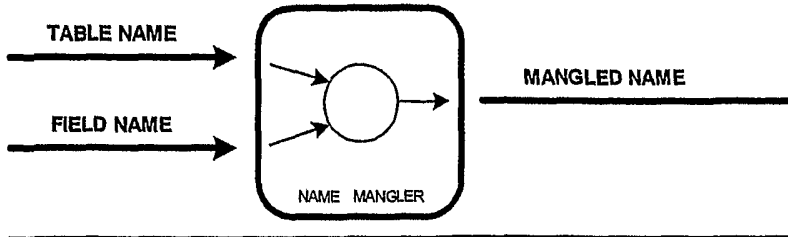


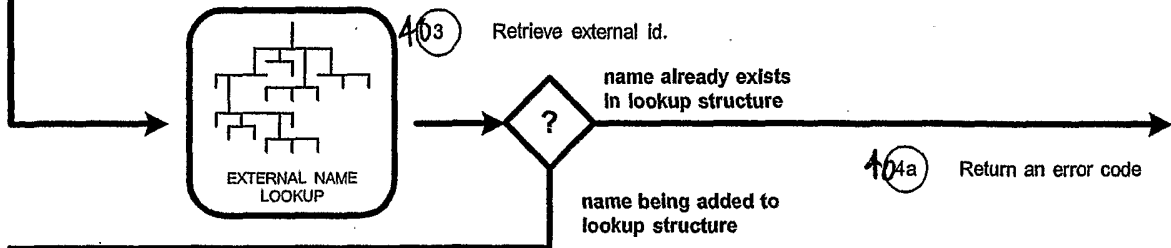
Figure 4 REGISTERING A HYPER-KEY LINK

401 Create a unique name by combining the table name with the field name for the foreign hyper-key.



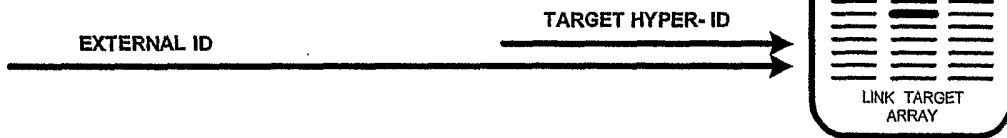
406

402 Find the name within the external name lookup structure or insert it into the external name lookup structure and identify the external type as a hyper-key link.

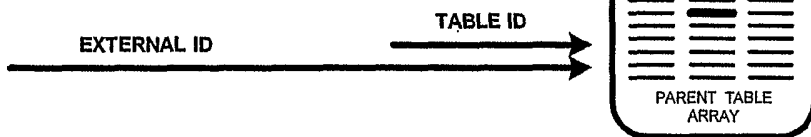


403 Retrieve external id.

404b Add the hyper-key id for the target of the foreign hyper-key to the location within the global link target array structure pointed to by the hyper-key link id just retrieved.



405 Add the table id to the location within the global parent table array structure pointed to by the hyper-key link id.



406 Add the field id for the field within the current table where the foreign hyper-key is located to the location within the global link target array structure pointed to by the hyper-key link id just retrieved.



407 Return the external id as the hyper-key link id.



Figure 5 REGISTERING A CONVENTIONAL INDEX

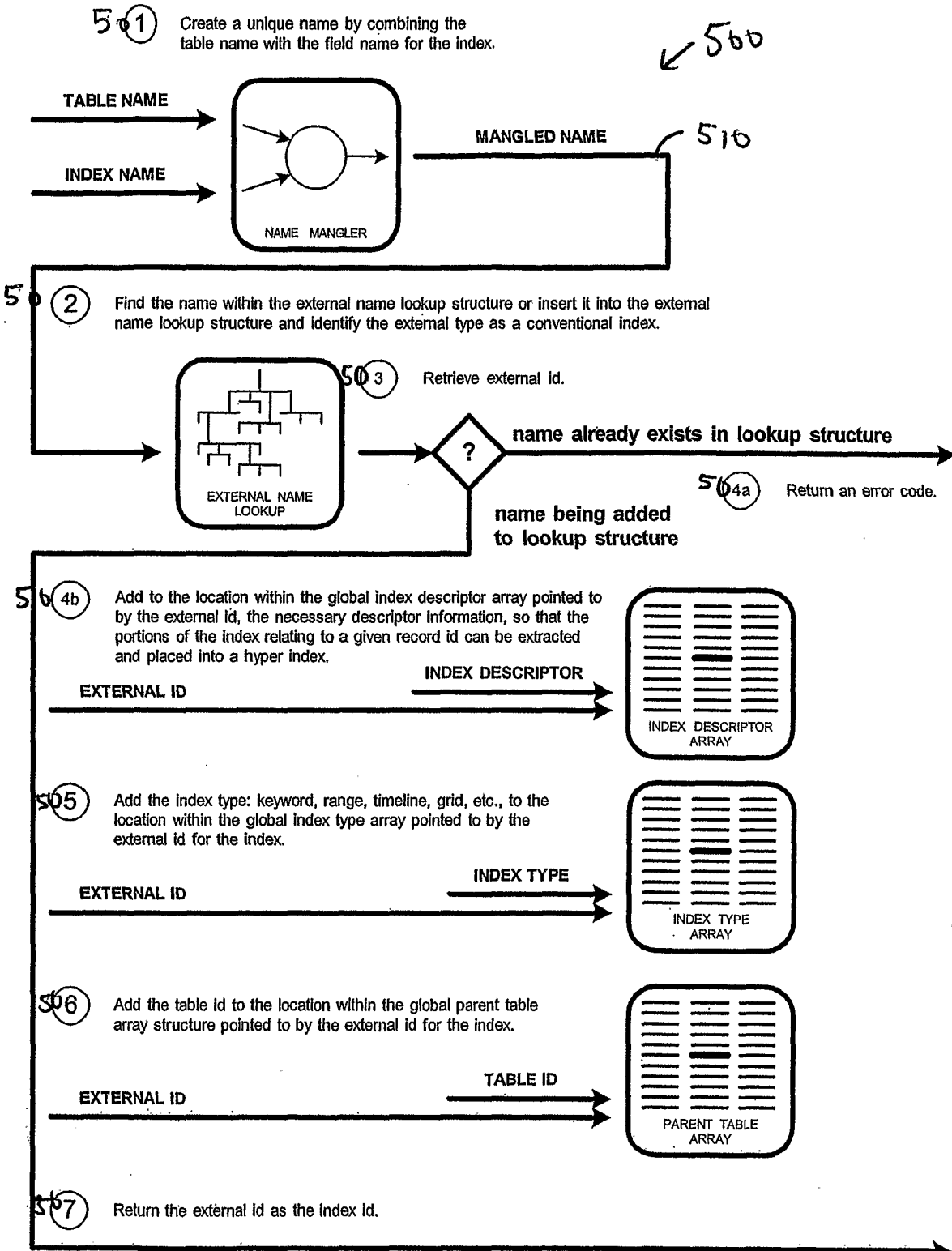
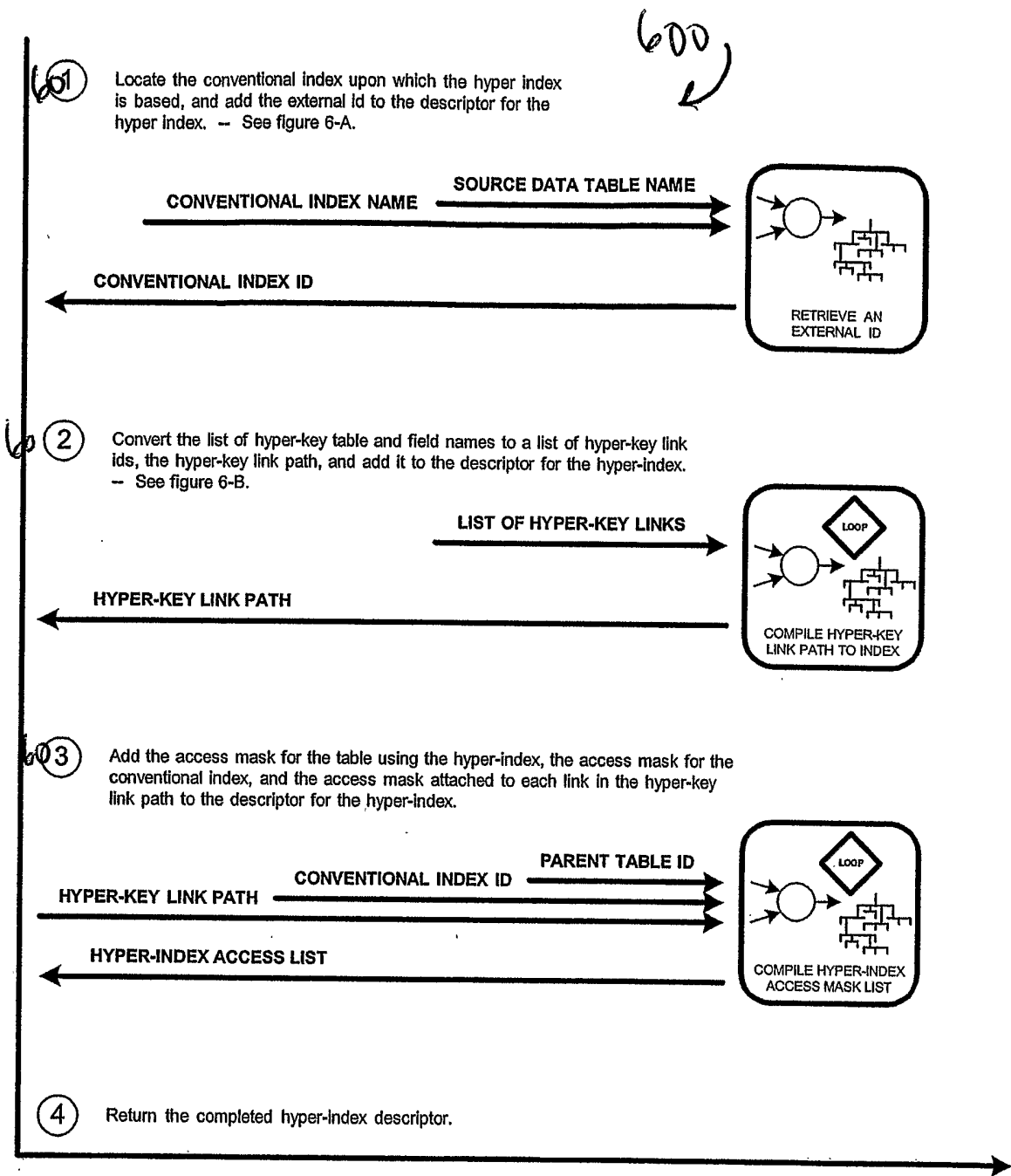


Figure 6 BUILDING A HYPER-INDEX DESCRIPTOR



600

Figure 7-A

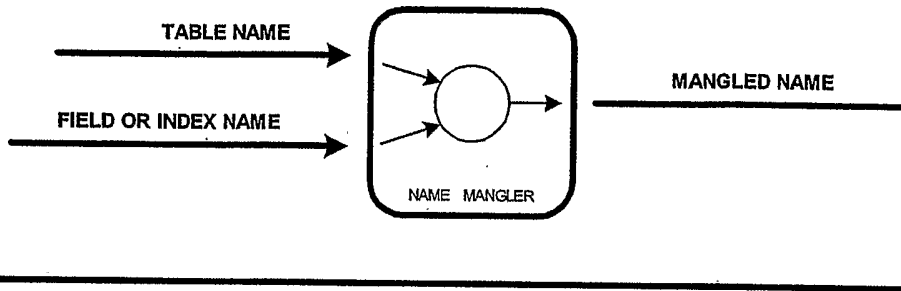
RETRIEVING AN EXTERNAL ID

This function is used to retrieve:

- (1) the index id for a conventional index;
- (2) the index id for a hyper- index;
- (3) the table id for a virtual table;
- (4) the hyper-key id for a primary hyper-key;
- (5) the hyper-key id for a secondary hyper-key;

700

701 Create a unique name by combining the table name with the field or index name. Use a null string for a primary index or a virtual table.



702 Find the name within the external name lookup structure or insert it into the external name lookup structure and identify the type of the external.

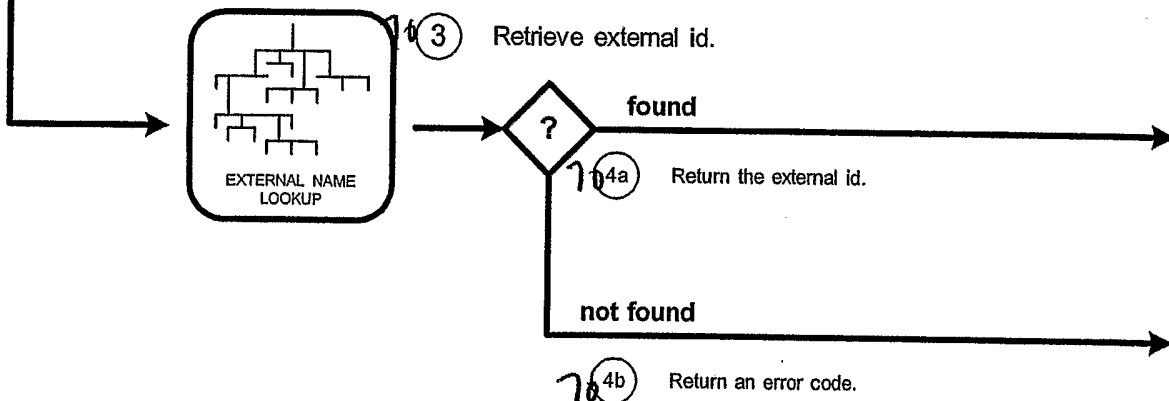


Figure 7-B

CREATE HYPER-KEY LINK PATH

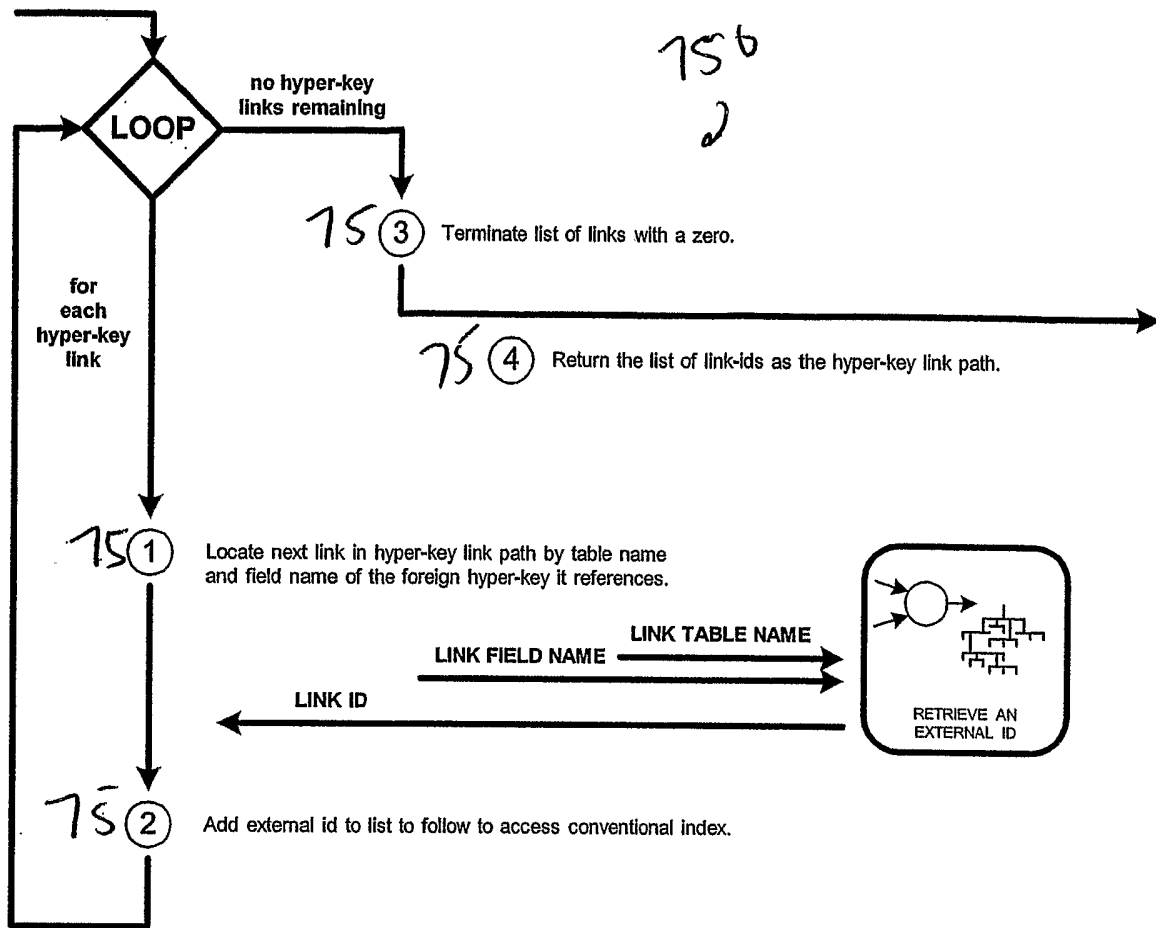


Figure 8

REGISTERING A HYPER-INDEX

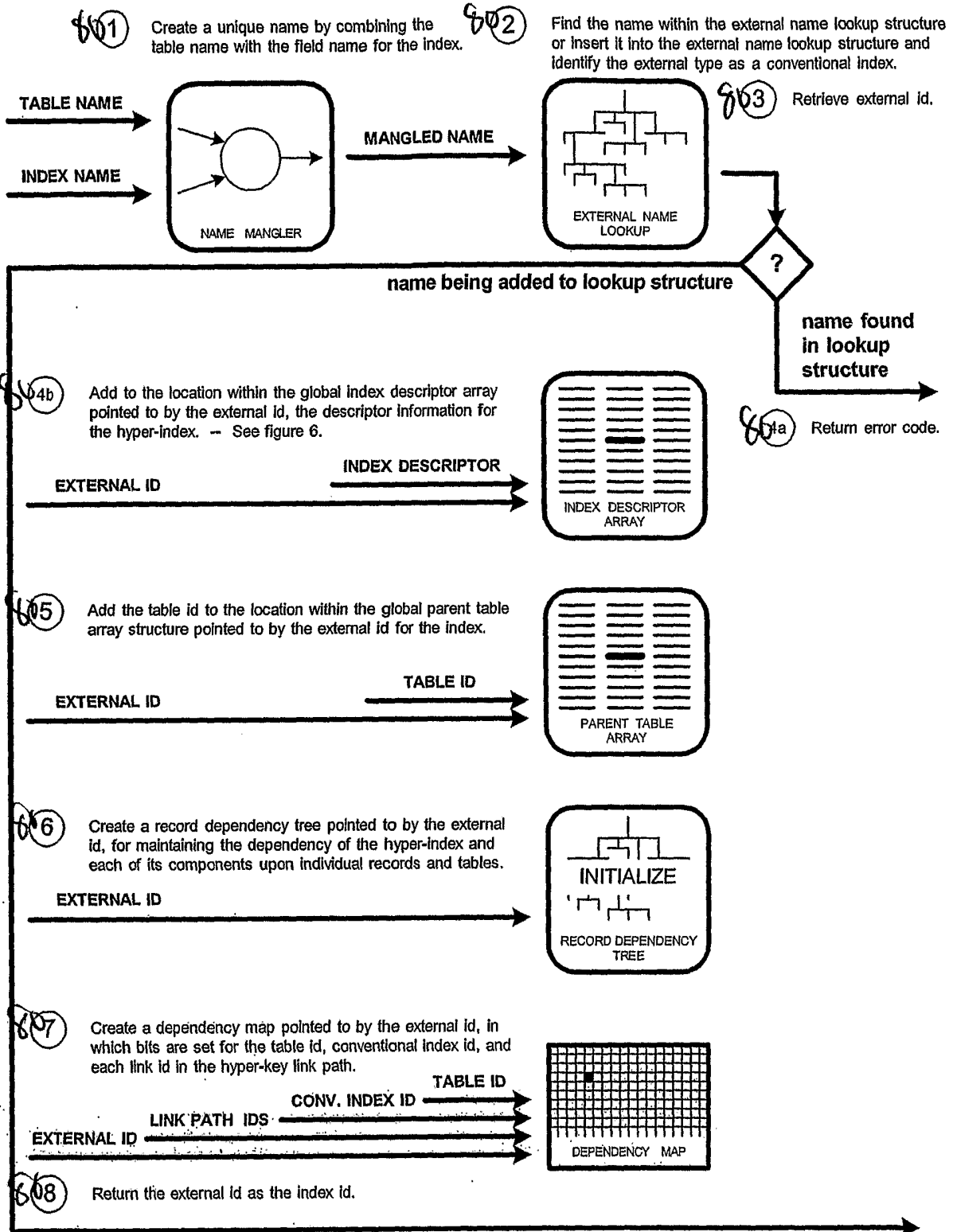


Figure 9

UPDATING A HYPER-INDEX

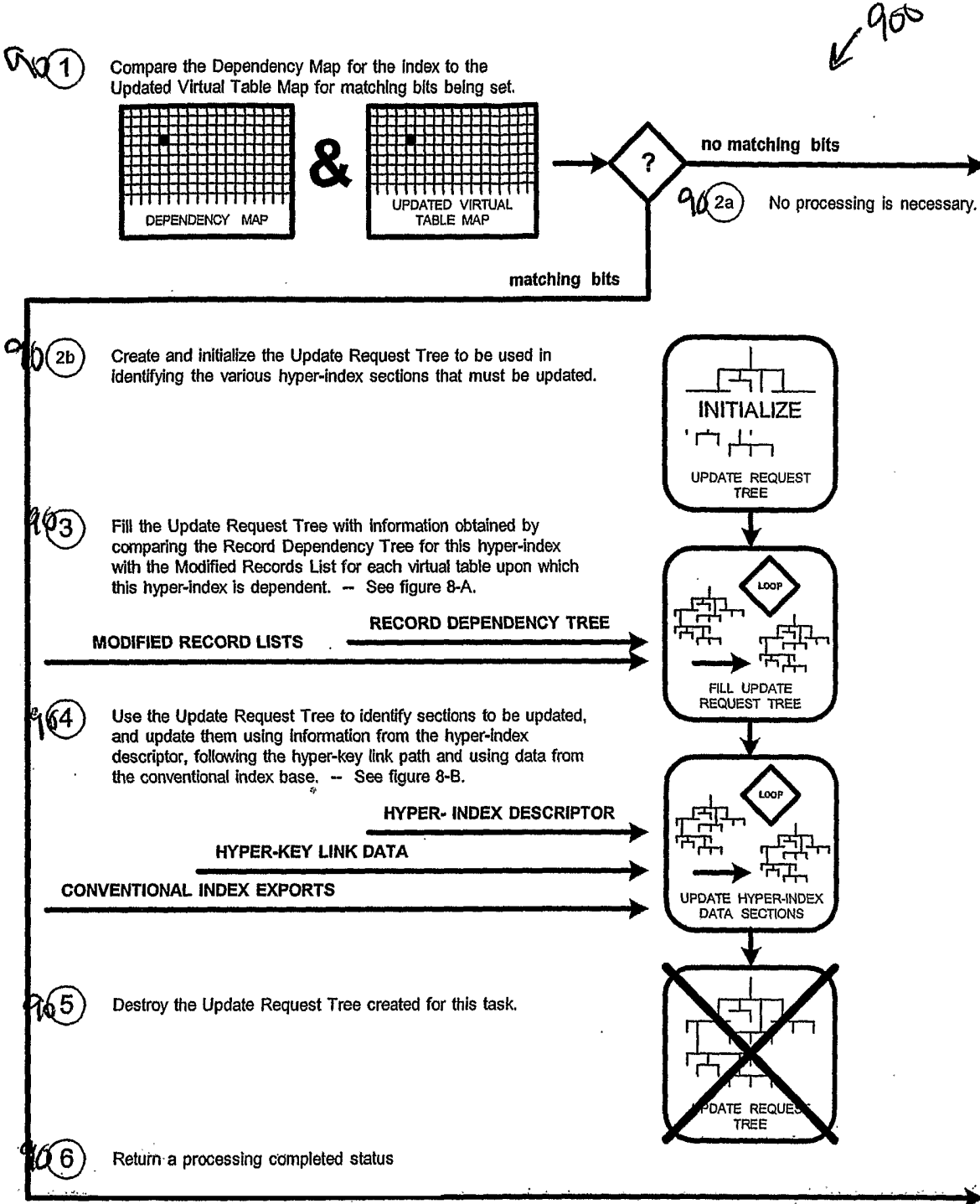


Figure 10-A

FILL UPDATE REQUEST TREE

