Time aspects in SAP Business Information Warehouse

Michael Hahne
cundus AG, Duisburg, Germany

With the Business Information Warehouse (BW) the SAP AG entered the business intelligence market.* Application components and services for decision support are summarized under the term mySAP Business Intelligence, its main component for persistent data storage is the Business Information Warehouse. Many SAP solutions are based upon the BW as core technology component following the paradigm of multidimensionality enabling numerous analytical applications. Data modeling and the capabilities of the underlying data model have a significant influence on acceptance and abilities of such systems. Possible solutions are limited by the data model. Main points in modeling multidimensional data structures are the possible hierarchical dimension structures and the treatment of structural changes in dimensions. The latter point is determined by reporting demands and discussed as time aspects in data warehouse literature. This article focuses on the possibilities of SAP Business Information Warehouse to handle dimension structures, especially changing dimensions.

1 MULTIDIMENSIONAL DATA STRUCTURES AND STAR SCHEMATA

Multidimensionality is the fundamental concept of a data warehouse. Stored data are based on multidimensional data structures with the cells of a n-dimensional cube representing the values. The edges of the cubes are called dimensions. They qualify the values and are the basis of deep analysis for business management. The elements in a dimension can be grouped. For example in a product dimension a grouping of elements to categories is possible. These classes of elements in an dimension are based upon 1:n-relationships. A set of such relationships form a navigational or consolidation path in a dimension and the base elements on lowest level form the level of deepest granularity and define therefore the most detailed level of analysis. The other elements are called derived or consolidated elements. The different parts of a dimension are summarized in figure 1.

Often there are parallel hierarchies in which the consolidation paths split into different ways that may join again at a higher level. If the condition of 1:n-relationships is dropped the resulting structures are based upon m:n-relationships. The assignment of calendar weeks to months is an example for this type of dimension structure. Furthermore the condition of balanced structures can be dropped resulting in unbalanced structures. In the case of unbalanced dimensions the grouping into classes of elements is a special case because the groups do not correspond with the levels that are defined by the length of a minimal path from the node to a root element. These groups have to follow rather the analytical needs than the counting of vertices.

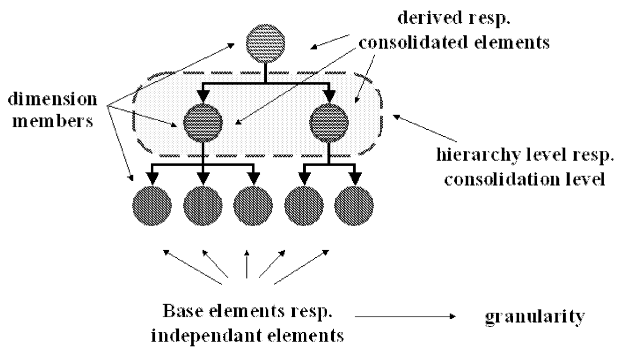

Figure 1: Components of structures in dimensions

There have been several approaches for storing multidimensional data structures in relational databases which are mainly discussed as *Star Schemata* (Hahne 1999). The facts are stored in a central fact table that builds the centre of the Star Schema surrounded by dimension tables (Nußdorfer 1998a).

*SAP, Business Information Warehouse, R/3, mySAP.com and BAPI are either registered trademarks or trademarks of SAP Aktiengesellschaft, Systems, Applications and Products in Data Processing, Neurottstrasse 16, 69190 Walldorf, Germany. SAP is not publisher of this paper and is not responsible for it under any aspect of press law.

The components of the primary key in the fact table reference the dimension tables. The dimension tables define analysis criteria and consolidation paths (Nußdorfer 1998b). In general these dimension tables are denormalized (Gluchowski et al. 1997). The requirement of consistent response time leads to variants considering especially this aspect. A central approach is to precalculate aggregates and store them in aggregate tables (Hahne 1999). The model of SAP BW is based essentially upon the Star Schema.

In the following section temporal structures in Data Warehousing are treated. Section three deals with the data model of the Business Information Warehouse called Enhanced Star Schema. The different variants for modeling hierarchical dimension structures especially in the BW are discussed in section four. The paper concludes with a summary in section five.

2 TIME ASPECTS IN A DATA WAREHOUSE

Time aspects are very important in multidimensional models (Chamoni and Stock 1999). This is also expressed by the especial role of the time dimension. Facts are assigned to the time dimension inherently but the tracking of changes in dimensions is not supported in general. This aspect is essential for analytical possibilities and the reporting demands define which kind of support of time aspects in dimensions is needed.

2.1 Changing dimensions and types of reporting demands

In practical analytical demands the data often must be analyzed according to the actual dimension structure and according to some kind of historical structures. Another reporting demand requires consideration of the structure for each fact tuple stored in the cube valid at loading time. This is known as the *historical truth*. Another reporting requirement is the consideration of only those structure components that are valid in all reported periods. This concept is described as *not to compare apples and pears*. There are four principle possibilities to implement these reporting demands:

1. Update the historical data according to the new valid structures.
2. Store the data with historical structure separately in addition to the complete data with the actual structure.
3. Build parallel hierarchies with old and new dimension structures.
4. Temporal databases.

The simplest way of treating changing dimensions is updating historical data to match actual dimension structures which implies that all old structures are lost and therefore cannot be reported but on the other hand the resulting models are small and not too complex. The concept of additionally stored data with historical and actual structure enables both dimension structures to be reported. But this concept also results in big models. With parallel hierarchies it is possible to consolidate data with any structure but the resulting dimension structures are confusing. Only the usage of databases with time stamps enables comprehensive business reporting but is connected with a slightly poorer performance.

2.2 Classification of time and database systems

According to the interpretation of time as an isomorphic image of the set of positive integers called *temporal universe* the time axis is divided into equidistant intervals called *chronon* which for example represents a second or minute. A fixed point of the time axis is called an *instant* and the time between two instances is called *time interval* (Clifford et al. 1977, p. 180). In the context of database systems at least two different concepts of time have to be considered: the *valid time* and the *transaction time*. The former considers the instances for which an object version is valid and the latter represents the time of update in the database.

Database systems can be classified according to the extent of implemented concepts of time aspects. The terms of valid time and transaction time lead to the following classification matrix (Ahn and Snodgrass 1986, p. 97):

	no valid time	*valid time*
no transaction time	snapshot database	historical database
transaction time	rollback database	bitemporal database

Table 1: classification of database systems

Database systems that consider at least transaction time or valid time are called temporal databases.

2.3 Valid time in multidimensional data models

In general both valid time and transaction time have to be considered in the context of multidimensional datamodels, because all facts refer to time implicitly given by the time dimension only the classification paths are considered. The discussion of transaction time is omitted in the following and only the concept of valid time is dicussed more closly.

Typically there are dimensions which are very static and others that are not. For example a product dimension is a dynamic dimension that changes very often. The consideration of updates in classification hierarchies results in the possibilities of implementation mentioned earlier. In the case of a relational data model the concept of *slowly changing dimensions* has been discussed in literature (Kimball 1996, p. 100ff.). According to Kimball the simplest case is overwriting old values implying the loss of possibilities for history tracking. In the second case there are single history tuples in the dimension table listing the dimension versions separately. The third case implements additional fields in the table with actual field values in addition to the original values.

More generally tracking history of dimension structures defined as a graph leads to a concept called *validity matrix* (Stock 2001, p. 121ff.). According to Stock it is sufficient to consider the edges in the dimension graph for putting time stamps. Hence the edges as pair of nodes are completed with additional valid time information represented by a time interval which can be defined by the starting and ending point. From this information of validity of edges the validity of nodes can be derived. The data of validity can be visualized in a symmetric matrix defined by the set of nodes as description for rows as well as for columns. Every cell represents a possible edge in the graph and contains the validity information for this especial edge. Another approach is to time stamp the whole classification hierarchy in the case of an update but this leads to a new complete hierarchy version in the case of just a small structural change.

3 ENHANCED STAR SCHEMA OF SAP

Data Storage in BW is based on relational database technology. The multidimensional data is also stored in some slightly modified kind of Star Schema with some. It improves some points that are weak supported in the classical Star Schema approach. Main aspects are the treatment of m:n-relationships in dimensions, support of unbalanced structures and aspects of structural changes in dimensions. A further point is the possibility to save data at different levels in cubes, because in general only input on the lowest level is considered. For example the case of actual data on a daily level and budget data on a monthly level.

3.1 Multidimensional data provider in the BW

The basic object for storing multidimensional data in the Business Information Warehouse is an *Info-Cube*. One fundamental principle in BW is the sharing of structural informations in dimensions across all Info-Cubes. Master data therefore is valid for all multidimensional structures in a system (SAP 2000b). The fact tables in the enhanced Star Schema as used in BW are also in the center of the model. Facts are called key figures. The fact table is surrounded by the dimension tables with appending master data tables.

A slight confusion lies in the usage of the terms *dimension* and *dimension table* in SAP BW and in the general terminology of multidimensional modeling. In the Enhanced Star Schema of BW the master data are joined to the facts with dimension tables. In these tables only the reference is stored, the real dimension member values are stored in the separated master data tables.

The basic elementary object to build multidimensional models in BW is an *Info-Object*. It is uniquely identified by its technical name and is comparable to a field in a table defined by name and data type. Info-Objects are defined globally in the BW-system and represent the objects to be analyzed, i.e. the key figures and also the criteria for analysis given as characteristics. All data objects in BW are based on Info-Objects as elementary constituent.

3.2 Concept of master data in BW

Very often the objects building dimension members are given as numerical key for example like a material number. An explanation is given by a textual description of the elements as the material text according to the material number. In Business Information Warehouse an essential feature is the consideration of textual descriptions in different languages. Hence they are stored in separate tables with the texts together with the language code. The ISO country code for example matches to different country names in different languages. The Enhanced Star Schema considers attributes of dimension elements as master data attributes of characteristics. They are stored in separate tables. For example a customer object has attributes like the address and contact person. In the BW attributes of caracteristics are divided into those which can be queried without the base characteristic and those with only descriptive function. Former attributes are called *navigational attributes* and the latter are *display attributes*.

An example of a hierarchy defined on a characteristic is the grouping of countries into regions and continents. These explicitly defined hierarchies are stored in the Business Information Warehouse also in the separated master data tables. The usage of the term *hierarchy* in the BW is unclear. The standard meaning like the structure in a dimension corresponds to 1:n-relationships between dimension members and hence in BW there are the three different forms of hierarchical structures defined by external hierarchies, characteristics in the cube and navigational attributes of characteristics (SAP 2000b). The master data tables are not directly joined with the fact table of the info cube but linked by further SID-tables that translate artificial dimension keys (surrogate identifiers) to keys for characteristics. This enables m:-n-relationships between characteristics in a dimension. This concept is shown in figure 2.

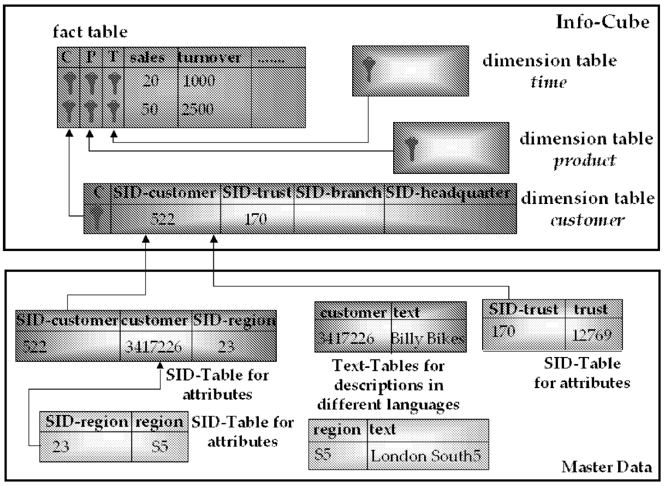

Figure 2: Enhanced Star Schema and concept of Master Data in SAP BW

The core meta data information of an Info-Object is its data type. The classical numerical data types are supplemented by types of quantities and amounts that are given with a unit of measure. For amounts of money the corresponding currency is the unit. Units are stored in BW automatically and conversions between different units like currency conversion are a possibility the system enables. Conversion rules are not only static, also different types can be used like currency conversion with average exchange rates, fixed courses and others.

3.3 Concept of valid time in SAP BW

For all master data in the Business Information Warehouse it is possible to define them with time stamps and therefore attributes, texts and hierarchies can be given with validity information.

To any Info-Object as the basic object in BW texts in possible different languages can be stored in seperated text tables. In BW it is possible to define these texts as time dependant. In this case the corresponding text table has additional fields *valid from* and *valid to* to track the validity of texts.

Attributes of an Info-Object regardless if they are navigational or display attributes may be defined as time dependant as well. Hence every Info-Object may have time dependant as well as not time dependant attributes. The concept of implementing history tracking is also based upon two additional fields *valid from* and *valid to*. These time dependant and also not time dependant attributes are joined to info cubes by different SID-tables which are summarized in figure 3.

Time dependancy of external hierarchies is special because there are different possibilities in BW. The most simple case is building different versions of a hierarchy. This is a good approach if a hierarchy structure at the end of a year has to be saved in addition to the actual structure and hence only very few different versions are given corresponding to fixed points of time. Over and above that it is possible to define the whole hierarchy as well as the hierarchy structure as time dependant. These are the two concepts of time stamping the complete hierarchy versus time stamping only the edges. The selected type of considering time dependance of external hierarchies is a property of the Info-Object and hence valid for all hierarchies of the underlying Info-Object.

4 VARIANTS FOR MODELING HIERARCHICAL STRUCTURES IN DIMENSIONS

Besides the different possible types of structures particularly the changes in dimensions are important. This aspect must be considered in the phase of modeling.

4.1 Hierarchical structures between characteristics in a dimension

The first possibility to model dimensional structures in BW is given by mapping each named level of consolidation of the dimension as defined in the conceptual model to a characteristic in one dimension (remember the different meaning of the term dimension). Therefore the assigning of data to consolidation elements is given by the fact data and fixed at load time. Hence this relationship between the different levels of consolidation are not stored in master data valid for all cubes but stored in the Info-Cube. This principle is shown in figure 4.

Unbalanced structures are impossible with this kind of modeling dimension structures because every level maps to one characteristic. The Business Information Warehouse does not know about the hierarchical structure of characteristics in a dimension as customer, region and country for example. Therefore

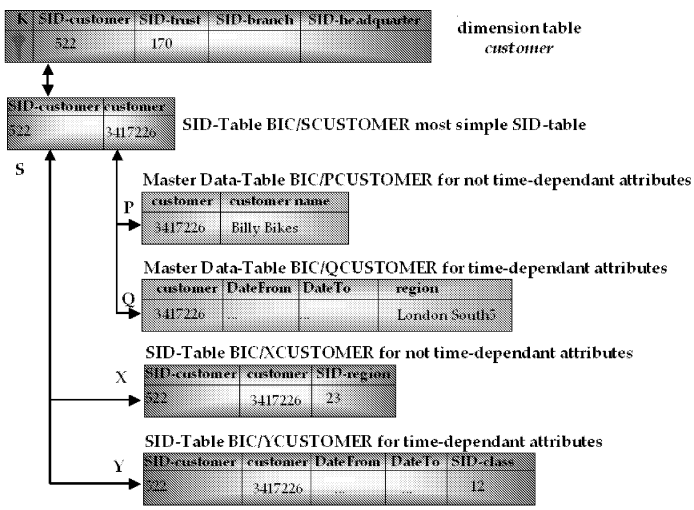

Figure 3: Variants of SID-Tables

there is no predefined path of navigation. Hierarchical relationships between characteristics cannot be restructered without general reorganisation of data and hence reporting is based on the historical truth.

4.2 Hierarchical structures modeled with navigational attributes

In the case of a hierarchy represented by a set of characteristics a reorganisation is needed to recognize different structures in one dimension. This is a reason for modeling dimension structures in the master data. Then every level of consolidation corresponds with an attribute of the base characeristic. This is shown in figure 5.

In the most extreme case where the dimension consists only of one characteristic all hierarchical information is stored in the master data. In those cases with only one characteristic in it, the dimension can be defined as a *line item dimension* in which the dimension table is omitted and the fact table is directly joined to the master data table.

The storage of dimension structures as attributes in master data implies that they are valid for all Info-Cubes and that changes are very simple to implement because only master data table updates are needed. In this type of modeled hierarchies in BW there are also no predefined paths of consolidation and hence levels can be left out in drilling down or rolling up. In general this possibility of modeling enables reporting only with the actual hierarchical structure.

If different structures are needed to be recognized in order to fulfill the reporting requirements then in BW there is the possibility to implement the attributes as time-dependant. This means all attribute values are stored together with the validity information in two additional fields *valid from* and *valid to*. On time-dependant attributes aggregates for realizing performance improvements cannot be defined.

4.3 External hierarchies in master data

The most flexible way for modeling dimension structures in BW is the way of defining *external hierarchies* that are based on a parent-child-representation (SAP 2000a) and hence enable unbalanced structures. They are also stored in master data and therefore are valid for each Info-Cube. For every Info-Object many different external hierarchies in different versions can be defined additionally. External hierarchies are based upon nodes with predecessor and successor if need be. Elements without successor are called leaf elements and the root elements are those without a predecessor. Nodes in an external hierarchy can be textual nodes or nodes based upon characteristics.

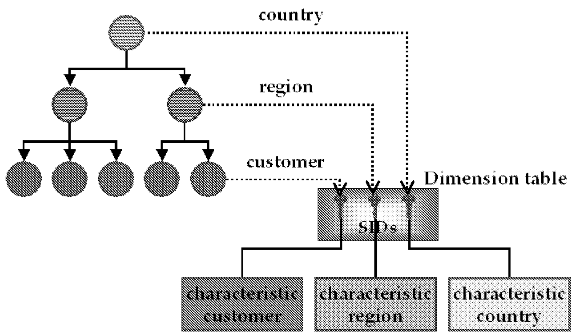

Figure 4: Dimension structures given by characteristics of fact data

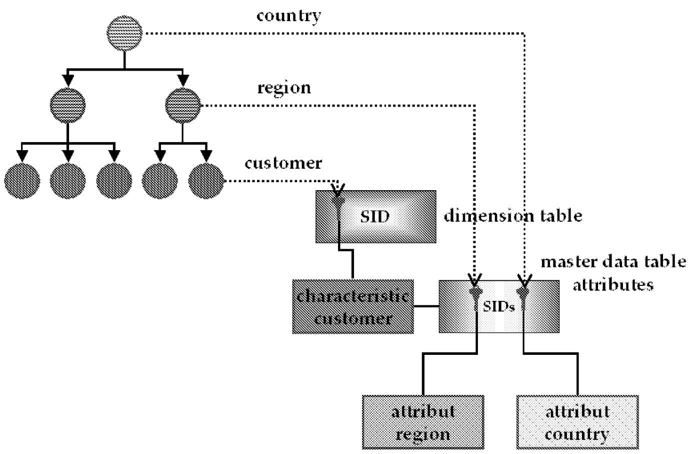

Figure 5: Hierarchical structures based on Navigational Attributes

External hierarchies are stored in master data spread over different tables. In the *inclusion table* the parent-child tuples are stored. Nodes for characteristics are stored with the SID and textual nodes get an artificial identifier of negative numbers. The different tables and their connection is summarized in figure 6. One can assign multiple parent elements to a node dropping the restriction to 1:n-relationships between consolidation levels. The consistent calculation is ensured by the BW automatically.

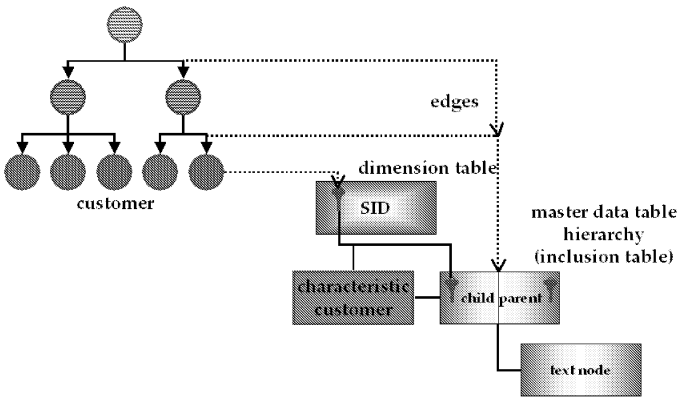

Figure 6: Explicit modeled structures in External Hierarchies

With the technique of different versions and multiple hierarchies modeled as external hierarchies different reporting demands can be implemented. To match the validity with a special time they can be defined as time-dependent with the two possibilities of recognizing the whole hierarchy or each parent-child tuple individually. This enables reporting according to all possible dimension structures given in the complete period of time.

5 SUMMARY

On-Line Analytical Processing (OLAP) as a fundamental principle of building analytical systems to support users in their business administration needs is based upon the multidimensional conceptual model with cubes and dimensions as main components enabling multiple variants of navigation and projections to different consolidation levels.

The aspect of modeling hierarchical dimension structures is an essential point in designing analytical information systems and furthermore in BW many different types of hierarchical structures are supported. Hierarchies can be implemented in three different ways in the Business Information Warehouse. Besides the possibility of hierarchies based upon characteristics in dimensions it is possible to implement hierarchical structures in the master data either with navigational attributes or with external hierarchies. The decision which choice fits best has to be done in the phase of modeling and the different possibilities can be used simultaniously to recognize various reporting demands. There are many aspects that have to be considered in this decision process. One very important aspect is the performance. Generally speaking hierarchies in facts are faster to query than those in master data. Poorest performance is connected with time-dependent external hierarchies. Hierarchies based upon characteristics is the only way to model structures valid only for one cube because master data are valid for all cubes in a system.

A further criterion is the flexibility in reorganizing dimension structures. This can can easily be done with master data only. The flexibility of various dimension types that are possible with external hierarchies are often an argument for this way of modeling hierarchical structures.

The treatment of all reporting demands is possible with the technique of time-dependent structures in BW. But there is a point that has to be considered: For querying such structures a key query date that sets the point in time to be recognized must be defined and in each query only one key date is valid for all time-dependent objects of the query. The more flexibility of modeling time-dependent structures results in more complexity and poorer performance.

REFERENCES

Ahn, I. and R. Snodgrass (1986). Performance evaluation of a temporal database management system. *SIGMOD record 15*(2), 96–107.

Chamoni, P. and S. Stock (1999). Temporal structures in data warehousing. In *Proceedings of the First International Conference on Data Warehousing and Knowledge Discovery (DaWaK'99)*, Florenz.

Clifford, J., C. Dyreson, T. Isakowitz, C. Jensen, and R. Snodgrass (1977). On the semantics of 'now' in databases. *ACM transactions on database systems 22*(2), 171–214.

Gluchowski, P., R. Gabriel, and P. Chamoni (1997). *Management-Support-Systeme : computergestützte Informationssysteme für Führungskräfte und Entscheidungsträger*. Berlin et al: Springer-Verlag.

Hahne, M. (1999). Logische Datenmodellierung für das Data Warehouse - Bestandteile und Varianten des Star Schemas. In P. Chamoni and P. Gluchowski (Eds.), *Analytische Informationssysteme: Data Warehouse, On-Line Analytical Processing, Data Mining* (2 ed.)., pp. 145–170. Berlin, Heidelberg et al: Springer.

Kimball, R. (1996). *The Data Warehouse Toolkit*. New York et al: John Wiley & Sons.

Nußdorfer, R. (1998a). E/R-Modell im Jungbrunnen. *Datenbank Fokus* (10), 16–19.

Nußdorfer, R. (1998b). Star Schema, Das E/R-Modell steht auf dem Kopf: Teil 1: Modellieren von Fakten-Tabellen. *Datenbank Fokus* (10), 22–28.

SAP (2000a). Hierarchies in SAP BW. ASAP for BW Accelerator.

SAP (2000b). Multidimensional Modeling with BW. ASAP for BW Accelerator.

Stock, S. (2001). *Modellierung zeitbezogener Daten im Data Warehouse*. Wiesbaden: Deutscher Universitäts-Verlag.