"SURVEY ON SPATIAL TEMPORAL DATA WAREHOUSING: - AN OVERVIEW"

[1]MISS. ROHINI A. DHOTE
Department of Computer Science & Information Technology, HVPM College of Engineering, Amravati, India
Rohinidhote6@gmail.com

[2]DR. S. P. DESHPANDE
Head of Department in PG Department of Computer Science & Technology, HVPMandals, Amravati, India
shrinivasdeshpande68@gmail.com

ABSTRACT: The field of Spatial Data warehouses has been emerging since the past decade due to the need to analyze large volumes of spatial data. The data once stored in a spatial data warehouse has to be queried using spatial online analytical processing (SOLAP) systems. The research in field of spatial data warehouses has been on conceptual models, materialization of spatial indexes, aggregation operations and SOLAP. In this paper we give an overview of the core concepts in a spatial data warehouse and recent advancements in the field. Spatial Data Warehouses (SDWs) allow analyzing historical data represented in a space supporting the Decision-making process. SDW applications require a multidimensional view of data that includes dimensions with hierarchies and facts with associated measures. In particular, hierarchies are important since traversing them users can analyze detailed and aggregated measures. To better represent users' requirements for SDWs, the conceptual model with spatial support should be used. Afterwards, the conceptual schema is translated to the logical and physical schemas. However, during the translation process the semantics can be lost. In this paper, we present the translation of spatial hierarchies from the conceptual to physical schemas represented in the MultiDimER model and Oracle 10g Spatial, respectively. Further, to ensure the semantic equivalence between the conceptual and the physical schemas, integrity constraints are exemplified mainly using triggers.

Keywords: Schema, SOLAP, Spatial, Temporal

1. INTRODUCTION

Spatio-temporal databases have received considerable attention during the past few years due to the accumulation of large amounts of multi-dimensional data evolving in time, and the emergence of novel applications such as traffic supervision, and mobile communication systems. Research has focused on modeling, indexing and query processing issues for problems involving historical information retrieval, motion and trajectory preservation [PJT00], future location estimation [SJLL00] etc. All these approaches assume that object locations are individually stored, and queries ask for objects that satisfy some spatio-temporal condition (e.g., mobile users inside a query window during a time interval, or the first car expected to arrive at a destination etc.). The motivation of this work is that many (if not most) current applications require summarized spatiotemporal data, rather than information about the locations of individual objects in time. As an example, traffic supervision systems need the number of cars in an area of interest, rather than their ids. Similarly mobile phone companies use the number of users serviced by individual cells in order to identify trends and prevent potential network congestion. Other spatio-temporal applications are by default based on arithmetic data rather than object locations. As an example consider a pollution monitoring system. The readings from several sensors are fed into a database which arranges them in regions of similar or identical values. These regions should then be indexed for the efficient processing of queries such as "find the areas near the center with the highest pollution levels yesterday".

The potentially huge amount of data involved in the above applications calls for pre-aggregation of results. In direct analogy with relational databases, efficient OLAP operations require materialization of summarized data. The motivation is even more urgent for spatiotemporal databases due to several reasons. First, in some cases, data about individual objects should not be stored due to legal issues. For instance, keeping the locations of mobile phone users through history may violate their privacy. Second, the actual data may not be important as in the traffic supervision system discussed. Third, although the actual data may be highly volatile and involve extreme space requirements, the summarized data are less voluminous and may remain rather constant for long intervals, thus requiring considerably less space for storage. In other words, although the number of moving cars (or mobile users) in some city area during the peak hours is high, the aggregated data may not change significantly since the number of cars (users) entering is similar to that exiting the area. This is especially true if only approximate information is kept, i.e., instead of the precise number we store values to denote ranges such as high, medium and low traffic.

2. SPATIAL DATA

2.1 What is Spatial Data?

Spatial data consist of spatial objects made up of points, lines, regions, rectangles, surfaces, volumes and even data of higher dimension which includes time. Example of spatial data include cities, rivers, roads, countries, states, crop coverage, mountain ranges, parts in a CAD system etc. Examples of spatial properties include the extent of a given river, or the boundary of a given country etc. often it is also desirable to attach non-spatial attribute information such as elevation heights, city names to the spatial data. Spatial database facilitate the storage of spatial and non-spatial information ideally without favoring one over the other.

Common way to deal with spatial data is to store it explicitly by parameterzing it and thereby obtaining a reduction to a point in a possible higher dimensional space. This is usually quite easy to do in a conventional database management system since the system is just a collection of records, where each record has many fields.

2.2 Spatial Database

Spatial database system is a database system. It offers spatial data types in its data model and query language. It supports spatial data types in its implementation. It manages data related to some space example

- Geographic space(surface of the earth at large or small scales)
  E.g. GIS, LIS
- The universe
  E.g. astronomy
- A VLSI design
  A model of the brain

2.3 Spatial Data Warehouse

The field of spatial data warehouse has been emerging since the past decade due to need to analyze large volume of spatial data. The data once stored in a spatial data warehouse has to be queried using spatial online analytical processing (SOLAP) systems. The research in the field of spatial data warehouse has been on conceptual models, materialization of spatial indexes, aggregation operation and SOLAP. In this paper we give an overview of the core concept in a spatial data warehouse and recent advancement in the field.

A data warehouse consists of facts and dimension modeled in star or snowflake schema. Data cube is a lattice of cuboids which represent hierarchies. The data cube may have cells which are pre computed for efficient query processing. Common OLAP include slicing, dicing, roll up, roll down. These concepts are extended to spatial data in a spatial data warehouse.

3. TEMPORAL DATA

3.1 What is Temporal Data?

Temporal data are abundantly present in many application domains such as banking, financial, clinical, geographical applications and so on. Temporal data have been extensively studied from data mining and database perspectives. Complementary to these studies, our work focuses on the visualization technique of temporal data. A wide range of visualization techniques have been designed to assist the users to visually analyze and manipulate temporal data. All the techniques have been designed independently. In such a context it is therefore difficult to systematically explore the set of possibilities as well as to thoroughly envision visualization technique of temporal data. A temporal data denotes the evolution of an object characteristic over a period of time. The value of temporal data is called a history. For the sake of simplicity, we define a history as a collection of instant time-stamped or interval time-stamped data items, although there are many other ways of representing a history.

3.2 Temporal Database

A Temporal database contains time varying data. Time is an important aspect of all real-world phenomenons. Events occur at specific points in time. Objects and the relationship among objects exist over time. The ability to model this temporal dimension of the real world is essential to much computer application such as accounting, banking, econometrics, geographical information systems, inventory controls, law, medical records, multimedia process control, reservation system and scientific data analysis.

Alternatively, in a temporal database, queries over previous states are easy to specify. Also modification to previous states and to future states is also easier to express using a temporal DBMS.

4. CONCEPTS OF DATA WAREHOUSE

The present section provides a global synthesis of the actual state of data warehousing and of the related concepts of multidimensional databases, data marts, online analytical processing and data mining. Specialized terms such as legacy systems, granularity, facts, dimensions, measures, snowflake schema, star schema, fact constellation, hypercube and N-tiered architectures are also defined. This synthesis is based on the theoretical concepts found in the pioneering literature of the mid-1990s, but it also reflects the most recent trends found in the literature as well as our own experiences.

4.1 Data Warehouse

An interesting paradox in the world of databases is that systems used for day-to-day operations store vast amounts of detailed information but yet are very inefficient for decision support and knowledge discovery. The systems used for day-to-day operations usually perform well for transaction processing where minimum redundancy and maximum integrity checking are key concepts, furthermore, this typically takes place within a context where the systems process large quantities of transactions involving small chunks of detailed data. On the other hand, decision makers need fast answers made of a few aggregated data summarizing large units of work. Something transactional systems do not achieve today with large databases. This difficulty of combining operational and decision-support databases within a single system gave rise to the dual-system approach typical of data warehouses. Although the underlying ideas are not new, the term 'data warehouse' originated ten years ago and rapidly became an explicit concept recognized by the community. It has been defined very similarly by pioneers such as Brackett (1996), Gill and Rao (I 996), Inmon et al (1996) and Poe (1995). In general, a data warehouse is an enterprise-oriented, integrated, non-volatile and read-only collection of data imported from heterogeneous sources and stored at several levels of detail to support decision-making. To facilitate the understanding of this definition, let us take each of these characteristics separately:

Subject-Oriented: Information is presented according to specific subjects or areas of interest, not simply as computer files. Data is manipulated to provide information about a particular subject. For example, the SRDB is not simply made accessible to end-users, but is provided structure and organized according to the specific needs

Integrated: A single source of information for and about understanding multiple areas of interest. The Data warehouse provides one-stop shopping and contains information about a variety of subject.

Non-Volatile: stable information that doesn't change each time an operational process is executed. Information is consistent regardless of when the warehouse is accessed.

Time-variant: Containing the history of the subject, as well as current information. Historical information is an important component of a data warehouse.

Accessible: The primary purpose of a data warehouse is to provide readily accessible information to end-users.

Process-oriented: it is important to view data warehousing as a process for delivery of information. The maintenance of a data warehouse is ongoing and iterative in nature.

Read-only: The warehouses can import the needed data but they cannot alter the state of the source databases, making sure that the original data always rest within the source. Such a requirement is necessary for technical concerns (e.g. to avoid update loops and inconsistencies) but mandatory to minimize organizational concerns (where is the original data? who owns it? who can change it? do we still need the legacy system? etc.) Thus, by definition, data warehouses are not allowed to write back into the legacy systems. However, although a data warehouse is not an Online Transaction Processing (OLTP) system (a system oriented towards the entering, storing, updating, integrity checking, securing and simple querying of data), it can be built to enter directly new information which is of high value for strategic decision-making but which does not exist in legacy systems.

To support decision-making

It is the sum of all the previous characteristics which make data warehouses the best source of information to support decision-making. Data warehouses provide the needed data stored in a structure which is built specifically to perform with global, homogeneous, multi-levels and multi-epochs queries from decision-makers. This allows for the use of new decision-support tools and new types of data queries, exploration and analyses which were too time consuming in the past.

4.2 Data Marts

The exact definition of data mart is still a controversy (Date 2000); however it is frequently defined as a specialized, subject-oriented, highly aggregated mini warehouse. It is more restricted in scope than the warehouse and can be seen as a departmental or partial special purpose warehouse usually dealing with coarser granularity. Several data marts can be created in an enterprise. Most of the time, it is built from a subset of the data warehouse. Figure 2.2.1 illustrates the distinction between legacy systems, data warehouses and data marts. However, in face of the major technical and organizational challenge of building an enterprise-wide warehouse, one may be temple to build subject-specific data marts without building the data warehouse first. This may, accelerate the feeding of the database and the delivery of partial decision-support solutions. This may solve temporary problems with small investments and minimum political struggle. But, there is a risk of seeing several data marts emerging throughout the organization and still having trouble in getting the global picture. It also is useless to say that the old chaotic situation prevailing between legacy systems will undoubtedly arise between data marts. In spite of these problems and an unavoidable chaos, this alternative presents several short term advantages. Thus, it is frequently adopted and may sometimes be the only possible alternative.

4.3 Online analytical processing

OLAP is a very popular category of decision-support tools which are typically used as clients of the data warehouse (and of data marts). OLAP provides functions for the rapid, interactive and easy *ad hoc* exploration and analysis of data with a multidimensional user interface. Consequently, OLAP functions include the previously defined drill-down. drill-up, and drill-across functions as well as other navigational functions such as filtering, slicing, dicing, pivoting, etc. (see OLAP Council 1995: Thomsen 1997). Users may also be helped to focus on exceptions or Locations which need special attention by methods which mark the interesting cells and paths. This kind of discovery-driven exploration of data has been studied by Sarawagi *et al.* (1998). Also, multi-feature databases which incorporate multiple, sophisticated aggregates can be constructed, as shown by Ross *et al.* (1998), to further facilitate data exploration and data mining.

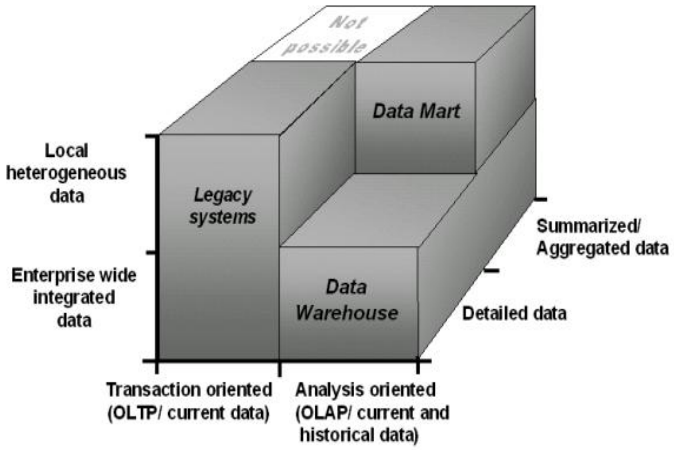

Figure 1: Comparison between Legacy system, Data mart and Data Warehouse

5. ARCHITECTURE OF DATA WAREHOUSE

Data warehouses can be implemented with different architectures depending on technological and organizational needs and constraints. The most typical one is also the simplest: it is called the 'corporate architecture' (Weldon 1997) or the 'generic architecture' (Poe 1995). It is represented in Figure 4.1. In such an architecture, the warehouse imports and integrates the desired data directly from the heterogeneous source systems, stores the resulting homogeneous enterprise-wide aggregated/summarized data in its own server, and lets the clients access these data with their own knowledge discovery software package (e.g. OLAP, data mining, query builder, report generator, executive information system). This two-tiered client-server architecture is the most centralized architecture.

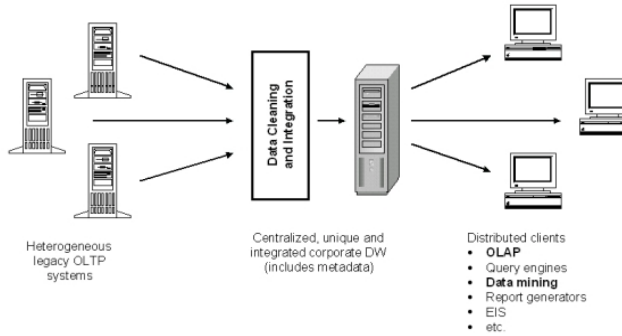

Figure 2: Generic Architecture of Data Warehouse

6. SPATIO-TEMPORAL DATA WAREHOUSE

- Several proposals aim at extending DW and OLAP with spatial/temporal features
- No commonly agreed definition of what is a spatio-temporal DW and what functionality it should support
- Proposed solutions vary considerably in the kind of data that can be represented and the kind of queries that can be expressed
- [Vaisman & Zim´anyi 2009] defined
    - Conceptual framework for spatio-temporal DWs using an extensible type system
    - Taxonomy of several classes of queries of increasing expressive power extending the tuple relational calculus with aggregated functions [Klug 1982]
- This provides the underlying basis for implementing spatio-temporal DWs

7. CONCLUSION

Spatial and Temporal database are an exciting and rapidly advancing field and we have outlined above few areas we consider worthwhile for doctoral candidates just starting their research Career. This paper aims at bringing together researchers and practitioners of spatial, temporal and spatio-temporal data warehousing.

These huge collections of Spatial-temporal data often hide possibly interesting information and valuable knowledge. It is obvious that a manual analysis of these data is impossible and data warehousing might provide useful tools and technology I this setting. Spatial temporal data warehousing is an emerging research area that is dedicated to the development of novel algorithm and computational technique for the successful analysis of large spatial-temporal database

8. REFERENCES

[1] Agarwal, S., Agrawal, R., Deshpande, P. M., Gupta, A., Naughton, J. F., Ramakrishnan, R. and Sarawagi, S. (1996) 'On the computation of multidimensional aggregates'. In Proceedings of the 1996 International Conference on Very Large Data Bases. Bombay.

[2] Bédard, Y., Merret, T., & Han, J. (2001). Fundamentals of spatial data warehousing for geographic knowledge discovery. In Geographic data mining and knowledge discovery (p. 53 – 73). Taylor & Francis.

[3] Bédard, Y., Rivest, S., & Proulx, M.-J. (2007). spatial online analytical processing (SOLAP): Concepts, architectures, and solutions from a geomatics engineering perspective. In Data warehouses and OLAP: Concepts, architectures and solutions (p. 298 – 319). IGI Global.

[4] Bimonte, S., Tchounikine, A., & Miquel, M. (2005).Towards a spatial multidimensional model. In DOLAP (p. 39-46).

9. AUTHOR PROFILE

	Miss. Rohini A. Dhote pursuing Master of Engineering in Computer Science & Information Technology from HVPM College of Engineering, Amravati, Maharashtra, India
	Dr. S. P. Deshpande Head of Department in PG Department of Computer Science & Technology HVPMandal Amravati, Maharashtra