IJITKM 0973-4414 Volume-10, Number-2 Jan-June 2017 pp. 12-17 **Impact Factor 2.8** available online at www.csjournalss.com

Spatial, Temporal and Cloud Data Warehouse and Mining

Dr. Prabhakar Laxmanrao Ramteke
Department of Information Technology, HVPM's College of Engineering & Technology, Amravati Maharashtra, India
pl_ramteke@rediffmail.com

Abstract: The big data in spatial, temporal and cloud data warehouses are an emerging area in the field of computer Science and Information Technology since past decade. To analyze these large volumes of data systematically online analytical processing of data is needed through queries. The innovations and advancement in spatial, temporal and cloud data warehouses has based on and design conceptual models, materialization of spatial indexes, aggregation operations. Data Mining is not specific to any particular kind of data and it operates on flat files, relational databases, data warehouses, transaction databases, multimedia databases and spatial databases. Cloud computing technology plays an important role to improve resource utilization of with the help of its unlimited storage capacity. It provides a shared infrastructure that can link a huge system pool together for various services and takes the large amount of data stored in many distributed computer.

A spatial data warehouse allows analyzing historical data represented in a space supporting the Decision-making process. Cloud computing allows the users to retrieve meaningful information from virtually integrated data warehouse that reduces the cost of infrastructure and storage. Spatial Data Warehouse applications require a multidimensional view of data that includes dimensions with hierarchies and facts with associated measures. The hierarchies are important since users can analyze detailed and aggregated measures. The conceptual model with spatial support should be used to represent users' requirements for Spatial Data warehouses. However, during the translation process the semantics can be lost.

The Data mining in Cloud Computing allows organizations/institutions to centralize the management of software and data storage, with assurance of efficient sharing of resources for their users. Data mining techniques are very much useful in cloud computing model.

Cloud computing users retrieve information from virtually integrated data warehouse that reduces the cost of infrastructure and storage. The Data mining in Cloud Computing allows centralizing the management of software and data storage, with assurance of efficient sharing of resources for their users.

Keywords: DBAAS, Spatiotemporal database, SOLAP, Spatial, Temporal and Cloud data warehouse

1. INTRODUCTION

A spatiotemporal database is a database that manages both space and time information. Common examples include, tracking of moving objects which typically can occupy only a single position at given time. A database of wireless communication networks, which may exist only for a short time span within geographic region. An index of species in a given geographic region, where over time additional species may be introduced or existing species migrate or die out and Historical tracking of plate tectonic activity [5].Due to huge growth in multi-dimensional data evolving in time, and the emergence of novel applications, spatio-temporal databases received closed attention to review. It is also focused on modeling, indexing and query processing issues for problems involving historical information retrieval, motion and trajectory preservation, future location estimation etc. [6].

All these approaches assume that object locations are individually stored, and queries ask for objects that satisfy some spatiotemporal condition e.g., mobile users inside a query window during a time interval or the car expected to arrive at a destination etc.. The 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. 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 center. The potentially large amount of data involved in the above applications. In direct analogy with relational databases, efficient online analytical processing (OLAP) operations require materialization of summarized data. 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. The actual data may not be important as in the traffic supervision system, 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 Concept of 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 Computer Aided Design system Examples of spatial properties include the extent of a given river or the boundary of given country. Often it is 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 reduction to a point in possible higher dimensional space. This is usually quite easy to do in conventional database management system since the system is just collection of records, where each record has many fields.

2.2 Spatial Database

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

- Geographic Space (Surface of the earth at large or small scales)

E.g. Geographic Information System

- The Universe E.g. Astronomy
- A Very Large Scale Integration Design

E.g. A model of the brain

2.3 Spatial Data Warehouse

The spatial data warehouse is an emerging area 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 online analytical processing. 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 AND CLOUD DATA

3.1 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, focuses on the visualization technique of temporal data. A wide range of visualization techniques has 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 evolution of an object characteristic over a period of time. The value of temporal data is called history. For the sake of simplicity, history as a collection of instant time-stamped or interval time-stamped data items, although there are many other ways of representing it.

3.2 Temporal Database

A Temporal database contains time varying data. Time is an important aspect of all real-world phenomenons and occurrence of events at specific points in time. Objects and its relationship among objects exist over time. The ability to model this temporal dimension of real world is essential to computer applications such as accounting, banking, econometrics, geographical information systems, inventory controls, law, medical records, multimedia process control, reservation system and scientific data analysis. Alternatively in temporal database, queries over previous states are easy to specify. Also modification to previous states and to future states is easier to express using a temporal DBMS.

3.3 Cloud Database

The Database management is much more complicated now as Big Data has arrived on the scene. In addition to traditional, structured data like business contacts and product intelligence, now have semi-structured and unstructured data coming at us fast and furious from all directions. The biggest source of this hard-to-analyze information is the mobile web. The flow of data just doesn't slow down as more and more people around the world access the Internet and use social media on mobile devices. And most organizations struggle to collect, organize, store and analyze all of it on their own. Enter the Cloud: a viable option for companies that don't have a lot of money for capital investments in equipment or the budget to maintain an IT department of the size needed to manage Big Data in house. Experts expect database-as-a-service (DBAAS), just like all the other 'as-a-service' options, to eventually become the standard solution for all but the most highly sensitive and mission-critical data.

A variety of cloud database management systems are available to store and analyze both relational (SQL) and non-relational (NoSQL) types of data. Here's a list of some of the leading service providers and their solutions:

Cloud database management options

- MicrosoftAzure/SQLDatabase – A full featured relational database-as-a-service, with Tables that offer NoSQL capabilities for storing large amounts of unstructured data and 'Blobs' (Binary Large Objects) for storing large amounts of unstructured text, video, audio and images.
- AmazonWebServices/Dynamo/RelationalDatabaseService – Amazon's offerings include NoSQL, MySQL, Oracle and MS SQL Server solutions. SimpleDB is Amazon's highly available and flexible non-relational data store that takes on the work of database administration.
- Xeround – A fully managed MySQL DBAAS that the vendor calls a 'drop-in solution' because it automates all configuration and ongoing Data Base operations.
- GoogleCloudSQL/GoogleAppEngineDatastore – Google's solutions for storing structured and unstructured data.
- ClearDB – This MySQL DBAAS boasts 100% uptime due to its multi-regional read/write mirroring.
- Database.com – A native cloud database service developed in house at Salesforce.com that became generally available in 2011. The vendor's website says it was built with the needs of social and mobile world at its core and not as an afterthought.

The unique features of cloud databases namely the ability to distribute data across wide geographical areas and among different servers in one physical data center which are based on cloud computing technology made possible by virtualization, something relational database management systems (RDBMS) were not designed for. To get around this limitation, leading DBAAS companies including Microsoft and Amazon offer their own RDBMS applications or software optimized for the cloud computing environment.

4. DATA WAREHOUSE

The present section provides a global synthesis of the actual state of data warehousing and 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 pioneering literature of the mid- 1990s, but it also reflects the most recent trends found in the literature

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 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 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. In general 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 Subject Related Data Base is not simply made accessible to end-users, but provided structure and organized according to the specific needs

Integrated- A single source of information for and about understanding multiple areas of interest is to be integrated. 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- It Contains the history of subject, as well as current information with respective to time as historical information is an important component of data warehouse.

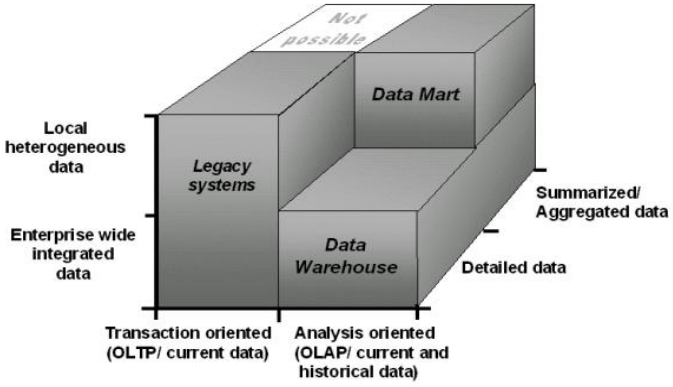
Accessible- The primary purpose of a data warehouse is to provide readily accessible information to end-users. Process-oriented- It is important to view data warehousing as a process for delivery of information. The maintenance of a data warehouse is ongoing and iterative in nature.

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 controversy however it is frequently defined as a specialized, subject-oriented, highly aggregated mini warehouse. It is a small data warehouse that satisfies the needs of reduced set of users [8]. 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 data warehouse. Figure 4.2.1 illustrates the distinction between legacy systems, data warehouses and data marts.



4.2.1 Comparison between Legacy system, Data mart and Data Warehouse

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 database and deliveries of partial decision-support solutions.It 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 is useless to say that the old chaotic situation prevailing between legacy systems will undoubtedly arise between data marts. This alternative presents several short term advantages.

4.3 Online Analytical Processing

OLAP is a very popular category of decision-support tools which typically are used as clients of data warehouse and of data marts. It provides functions for rapid, interactive and easy *ad hoc* exploration. Consequently, functions include previously defined drill-down, drill-up, and drill-across functions as well as other navigational functions such as filtering, slicing, dicing, pivoting, etc. Users may help to focus on locations which need special attention by methods. Multi-feature databases which incorporate multiple, sophisticated aggregates can be constructed to further facilitate data exploration and data mining. The objective is to extract structured information from unstructured or semi-structure data sources. The Data mining in Cloud Computing allows organizations/institutions to centralize the management of software and data storage, with assurance of efficient sharing of resources for their users.

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 called the 'corporate architecture' or the 'generic architecture'. It is represented in Fig.5.1 & Fig. 5.2.

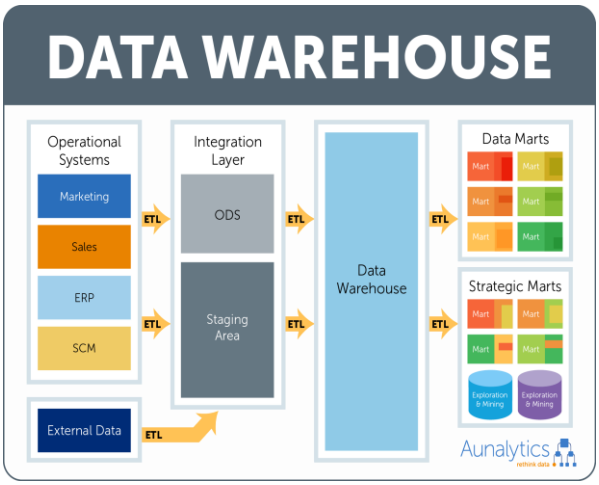


Fig. 5.1 Data Warehouse Components

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.

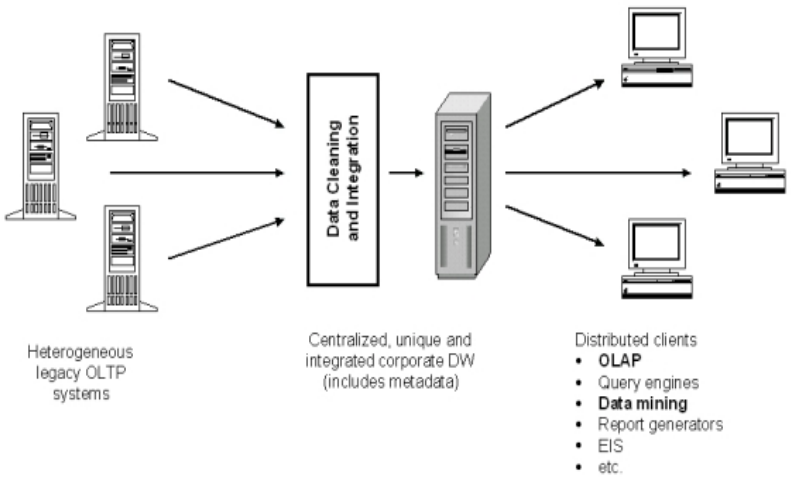


Fig. 5.2 Generic Architecture of Data Warehouse

5.1 Patio-Temporal Data Warehouse

Several aim at extending Data ware house and OLAP with spatial/temporal features

- Proposed solutions vary considerably in the kind of data and queries that can be represented and expressed
- Conceptual framework for spatio-temporal Data Warehouses using an extensible type system
- Taxonomy of several classes of queries increase expressive power extending tuple relational with aggregate functions
- This provides the underlying basis for implementing spatio-temporal Data Warehouses

6. FUTURE SCOPE

Data mining techniques with Cloud computing can be used in institutions/organizations to extract meaningful knowledge from large data. Data mining techniques can be applicable to any kind of information such as scientific data, medical data, satellite sensing, text report, business transactions, educational systems and government offices.

7. CONCLUSION

Spatial, Temporal and cloud database are an exciting and rapidly advancing field. Above are few areas for research. These huge collections of Spatial-temporal data often hide possibly interesting information and valuable knowledge. It is obvious that manual analysis of these data is impossible and data warehousing might provide useful tools and technology. Spatial temporal or cloud 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 databases using analytics.

References

[1] Agarwal, S., Agrawal, P. M., Gupta, "On the computation of multidimensional aggregates"', In Proceedings of International Conference
on Very Large Data Bases. Bombay, 1996

[2] Taylor & Francis, "Fundamentals of spatial data warehousing for geographic knowledge discovery", (pg. 53 – 73), 2001

[3] Bédard, Y. Rivest, S., & Prolix, M.J, "Spatial online analytical processing (SOLAP): Concepts, architectures and solutions from a
Geometrics engineering perspective", In Data Warehouses and OLAP: Concepts, architectures and solutions (p. 298 – 319),
IGI Global, 2007

[4] S.Bimonte, A. Tchounikine, & M. Miquel, "Towards a Spatial Multidimensional Model in DOLAP", (pg. 39-46), 2005

[5]https://en.wikipedia.org/wiki/Spatiotemporal_database

[6] Yufei Tao, Dimitris Papadias Hong Kong University of Science and Technology, Hong Kong, China, "Historical Spatio-Temporal
Aggregation",2016

[7] Jiawei Han and Micheline Kamber, "Data Mining: Concepts and Techniques", 2nd Edition, 2006

[8] Ralph Kimball, "The data warehouse tool kit", John Wiley & Sons, 1996.