



Emerging Database Models and Related Technologies

Praseeda Manoj

Dept. of Computer Science, Muscat College
Oman

Abstract— *Traditional applications had a common platform that captured business transactions. The emerging class of applications requires new functionality that closes the loop between incoming transactions and the analytics that drive action on those transactions. Adopting a longer term and more abstract perspective, it is likely that new database management technologies and application areas will continue to emerge. As with any sufficiently fashionable technology, users should expect data management marketplace to yield solutions that are increasingly application/environment specific.*

Keywords— *Emerging Database, Multimedia, Temporal, Technology, BigData, GIS*

I. INTRODUCTION

The term database refers to the collection of related records, and the software should be referred to as the database management system or DBMS. Database management systems are usually categorized according to the data model that they support: relational, object-relational, network, and so on. The data model will tend to determine the query languages that are available to access the database. The world of database boasts many kinds of technologies, which cater to the need of many kinds of organizations. Since 1970 different models and methods have been developed to describe, analyse and design computer based files and databases. The existing relational DBMS technology has been successfully applied to many application domains. RDBMS technology has proved to be an effective solution for data management requirements in large and small organizations, and today this technology forms a key component of most information systems. However, Applications in domains such as Multimedia, Geographical Information Systems, digital libraries, mobile database etc. demand a completely different set of requirements in terms of the underlying database models. The conventional relational database model is no longer appropriate for these types of data. Furthermore the volume of data is significantly larger than in classical database systems. Finally, indexing, retrieving and analysing these data types require specialized functionality that are not available in conventional database systems. This paper will cover some requirements of these emerging databases such as multimedia database, spatial database, temporal database, biological/genome database, mobile database, big data, their underlying technologies, data models and languages. These trends have resulted into the development of new database technologies to handle new data types and applications.

II. MULTIMEDIA DATABASE

Multimedia computing has emerged as a major area of research and has started dominating all facets of lives of mankind. A multimedia database is a database that hosts one or more primary media file types such as video, audio, radar signals and documents or pictures in various encoding. These forms have in common that they are much larger than the earlier forms of data integers, character strings of fixed length and vastly varying size. These are fall into three main categories:

- Static media (time-independent, i.e. images and handwriting)
- Dynamic media (time-dependent, i.e. video and sound bytes)
- Dimensional media (i.e. 3D games or computer-aided drafting programs- CAD)

All primary media files are stored in binary strings of zeros and ones, and are encoded according to file type. The term "data" is typically referenced from the computer point of view, whereas the term "multimedia" is referenced from the user point of view. There are numerous different types of multimedia databases, including:

- The Authentication Multimedia Database is a 1:1 data comparison ratio.
- The Identification Multimedia Database is a data comparison of one-to-many

A newly-emerging type of multimedia database, is the Biometrics Multimedia Database, which specializes in automatic human verification based on the algorithms of their behavioural or physiological profile. This method of identification is superior to traditional multimedia database methods requiring the typical input of personal identification numbers and passwords. Due to the fact that the person being identified does not need to be physically present, where the identification check is taking place. This removes the need for the person being scanned to remember a PIN or password. Fingerprint identification technology is also based on this type of multimedia database. The historic relational databases (i.e. the Binary Large Objects - BLOBs- developed for SQL databases to store multimedia data) do not conveniently

support content-based searches for multimedia content. This is due to the relational database not being able to recognize the internal structure of a Binary Large Object and therefore internal multimedia data components cannot be retrieved.

III. TEMPORAL DATABASE

Time is an important aspect of real world phenomena. Events occur at specific points in time. Objects and relationships among objects exist over time. The ability to model this temporal dimension of real world is essential to many computer applications such as econometrics, inventory control, airline reservations, medical records, accounting, law, banking, land and geographical information systems. In contrast, existing database technology provides little support for managing such data.

A temporal database is formed by compiling and storing temporal data. The difference between temporal data and non-temporal data is that a time period is appended to data expressing when it was valid or stored in the database. The data stored by conventional databases consider data to be valid at present time as in the time instance "now". When data in such a database is modified, removed or inserted, the state of the database is overwritten to form a new state. The state prior to any changes to the database is no longer available. Thus, by associate time with data, it is possible to store the different database states. In essence, temporal data is formed by time-stamping ordinary data (type of data we associate and store in conventional databases). In a relational data model, tuples are time-stamped and in an object-oriented data model, objects/attributes are time stamped. Each ordinary data has two time values attached to it, a start time and an end time to establish the time interval of the data. In a relational data model, relations are extended to have two additional attributes, one for start time and another for end time.

Different Forms of Temporal Databases

Time can be interpreted as valid time (when data occurred or is true in reality) or transaction time (when data was entered into the database).

- A historical database stores data with respect to valid time.
- A rollback database stores data with respect to transaction time.
- A bitemporal database stores data with respect to both valid and transaction time – they store the history of data with respect to valid time and transaction time.

A central goal of conventional relational database design is to produce a database schema consisting of a set of relational schemas. In normalization theory, normal forms constitute attempts at characterizing "good" relation schemas, and a wide variety of normal forms has been proposed, the most prominent being third normal form and Boyce-Codd normal form. An extensive theory has been developed to provide a solid formal footing for relational database design, and most database textbooks expose their readers to the core of this theory. In temporal databases, there is an even greater need for database design guidelines. However, the conventional normalization concepts are not applicable to temporal relational data models because these models employ relational structures different from conventional relations. New temporal normal forms and underlying concepts that may serve as guidelines during temporal database design are needed. Temporal data models generally define timeslice operators, which may be used to determine the snapshots contained in a temporal relation. Accepting a temporal relation as their argument and a time point as their parameter, these operators return the snapshot of the relation corresponding to the specified time point.

Adopting a longer term and more abstract perspective, it is likely that new database management technologies and application areas will continue to emerge that provide 'temporal' challenges. Due to the ubiquity of time and its importance to most database management applications, and because built-in temporal support generally offers many benefits and is challenging to provide, research in the temporal aspects of new database management technologies will continue to flourish for existing as well as new application areas.

IV. MOBILE DATABASE

Recent advances in portable and wireless technology led to mobile computing, a new dimension in data communication and processing. Portable computing devices coupled with wireless communications allow clients to access data from virtually anywhere and at any time. Now days you can even connect to your Intranet from an aeroplane. Mobile database are the database that allows the development and deployment of database applications for handheld devices, thus, enabling relational database based applications in the hands of mobile workers. The database technology allows employees using handheld to link to their corporate networks, download data, work offline, and then connect to the network again to synchronise with the corporate database. Mobile computing applications, residing fully or partially on mobile devices, typically use cellular networks to transmit information over wide areas, and wireless LANs over short distances. Some of the commercially available Common Mobile relational Database systems are IBM's DB2 Everywhere 1.0, Oracle Lite, Sybase's SQL etc.

These databases work on Palm top and hand held devices (Windows CE devices) providing a local data store for the relational data acquired from enterprise SQL databases. The main constraints for such databases are relating to the size of the program as the handheld devices have RAM oriented constraints. The commercially available mobile database systems allow wide variety of platforms and data sources. They also allows users with handheld to synchronise with Open Database Connectivity (ODBC) database content, and personal information management data and email from

Lotus Development's Notes or Microsoft's Exchange. These database technologies support either query-by-example (QBE) or SQL statements.

Mobile computing has proved useful in many applications. Many business travellers are using laptop computers to enable them to work and to access data while traveling. Delivery services may use/ are using mobile computers to assist in tracking of delivery of goods. Emergency response services may use/ are using mobile computers at the disasters sites, medical emergencies, etc. to access information and to provide data pertaining to the situation. Newer applications of mobile computers are also emerging.

V. GEOGRAPHIC INFORMATION SYSTEMS

GIS is a technological field that incorporates geographical features with tabular data in order to map, analyse, and assess real-world problems. The key word to this technology is Geography – this means that some portion of the data is spatial. In other words, data that is in some way referenced to locations on the earth. Coupled with this data is usually tabular data known as attribute data. Attribute data can be generally defined as additional information about each of the spatial features. Geographic information systems(GIS) are used to collect, model, and analyse information describing physical properties of the geographical world. The scope of GIS broadly encompasses two types of data:

- spatial data, originating from maps, digital images, administrative and political boundaries, roads, transportation networks, physical data, such as rivers, soil characteristics, climatic regions, land elevations, and
- non spatial data, such as socio-economic data (like census counts), economic data, and sales or marketing information. GIS is a rapidly developing domain that offers highly innovative approaches to meet some challenging technical demands.

GIS Applications can be divided into three categories

- cartographic applications
- digital terrain modelling applications
- geographic objects applications

Figure 1 shows GIS categories and grouping of different GIS application areas. GIS data can be broadly represented in two formats, Vector data and Raster data. Vector data represents geometric objects such as points, lines and polygons. Raster data is characterized as an array of points, where each point represents the value of an attribute for a real-world location. Informally, raster images are n-dimensional array where each entry is a unit of the image and represents an attribute. Two-dimensional units are called pixels, while three-dimensional units are called voxels. Three-dimensional elevation data is stored in a raster-based digital elevation model (DEM) format. Another raster format called triangular irregular network (TIN) is a topological vector-based approach that models surfaces by connecting sample points as vector of triangles and has a point density that may vary with the roughness of the terrain. Rectangular grids (or elevation matrices) are two-dimensional array structures.

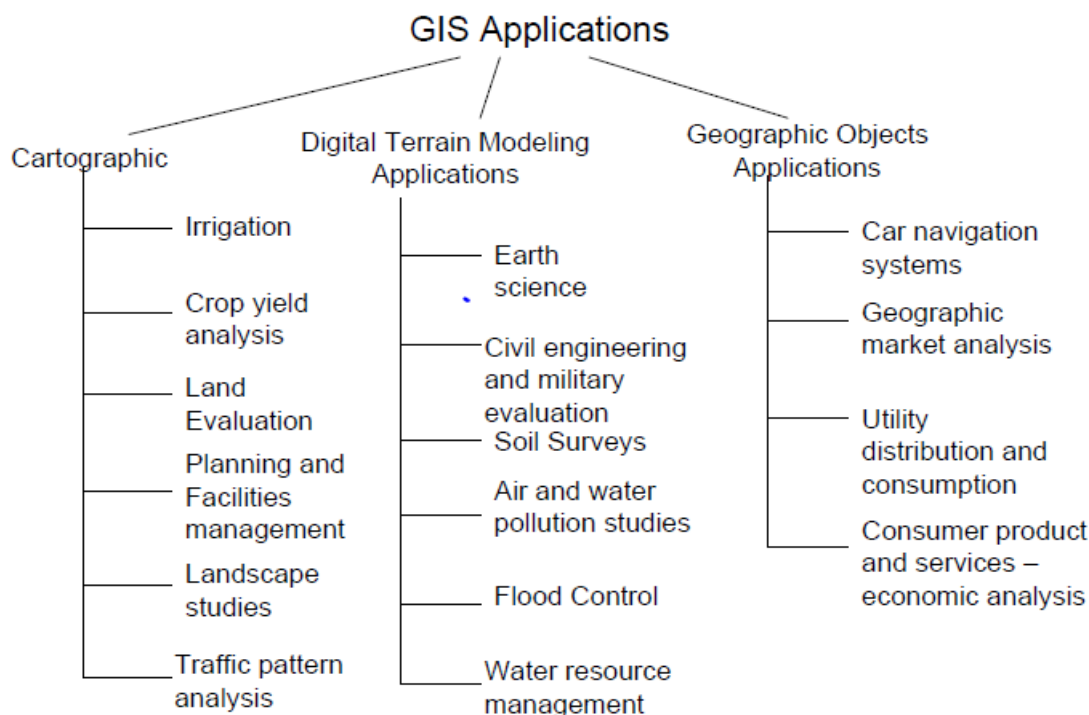


Figure 1: GIS applications category and grouping

VI. GENOME DATA

The biological sciences encompass an enormous variety of information. Environmental science gives us a view of how species live and interact in a world filled with natural phenomena. Biology and ecology study particular species. Anatomy focuses on the overall structure of an organism, documenting the physical aspects of individual bodies. Traditional medicine and physiology break the organism into systems and tissues and strive to collect information on the workings of these systems and the organism as a whole. Histology and cell biology delve into the tissue and cellular levels and provide knowledge about the inner structure and function of the cell. This wealth of information that has been generated, classified, and stored for centuries has only recently become a major application of database technology. Genetics has emerged as an ideal field for the application of information technology. In a broad sense, it can be taught of as the construction of models based on information about genes – which can be defined as units of heredity – and population and the seeking out of relationships in that information. The study of genetics can be divided into three branches:

- Mendelian genetics. This is the study of the transmission of traits between generations.
- Molecular genetics. This is the study of the chemical structure and function of genes at the molecular level.
- Population genetics. This is the study of how genetic information varies across populations of organisms.

Biological data exhibits many special characteristics that make management of biological information a particularly challenging problem. The characteristics related to biological information, and focusing on a multidisciplinary field called bioinformatics that has emerged. Bioinformatics addresses information management of genetic information with special emphasis on DNA sequence analysis. Applications of bioinformatics span design of targets for drugs, study of mutations and related diseases, anthropological investigations on migration patterns of tribes and therapeutic treatments. The term genome is defined as the total genetic information that can be obtained about an entity. The human genome, for example, generally refers to the complete set of genes required to create a human being –estimated to be more than 30,000 genes spread over 23 pairs of chromosomes, with an estimated 3 to 4 billion nucleotides. The goal of the Human Genome Project(HGP) has been to obtain the complete sequence – the ordering of the bases – of those nucleotides.

VII. DIGITAL LIBRARY

Digital libraries are an important and active research area. Conceptually, a digital library is an analog of a traditional library—a large collection of information sources in various media—coupled with the advantages of traditional technologies. However, digital libraries differ from their traditional counter-parts in significant ways: storage is digital, remote access is quick and easy, and materials are copied from a master version. Furthermore, keeping extra copies on hand is easy and is not hampered by budget and storage restrictions, which are major problems in traditional libraries. Thus, digital technologies overcome many of the physical and economic limitations of traditional libraries.

The Digital Library Initiative (DLI), jointly focused by SNF, DARPA, and NASA, has been a major accelerator of the development of digital libraries. This initiative provided significant funding to six major projects at six universities in its first phase covering a broad spectrum of enabling technologies. The initiative's web page define its focus as “dramatically advance the means to collect, store, and organize information in digital forms, and make it available for searching, retrieval, and processing via communication networks—all in user-friendly ways. The magnitude of these data collections as well as their diversity and multiple formats provides challenges on a new scale. The future progression of the development of digital libraries is likely to move from the present technology of retrieval via the internet, though net searches of indexed information in repositories, to a time of information correlation and analysis by intelligent networks. Techniques for collecting information, storing it, and organizing it to support informational requirements learned in decades of design and implementation of database will provide the baseline for development of approaches appropriate for digital libraries.

VIII. BIG DATA

Now a days advancement of technology generate large, diverse, longitudinal, complex, and/or distributed data sets mainly from instruments, sensors, Internet transactions, email, video, click streams, and/or all other digital sources. Individuals with smartphones and on social network sites and multimedia will continue to fuel exponential growth of data. The large pools of data that can be captured, communicated, aggregated, stored, and analysed is part of every sector and function of the global economy. This amount of data has been exploding. Companies capture trillions of bytes of information about their customers, suppliers, and operations, and millions of networked sensors are being embedded in the physical world in devices such as mobile phones and automobiles, sensing, creating, and communicating data. Multimedia and individuals with smartphones and on social network sites will continue to fuel exponential growth. Big data—large pools of data that can be captured, communicated, aggregated, stored, and analysed—is now part of every sector and function of the global economy. Like other essential factors of production such as hard assets and human capital, it is increasingly the case that much of modern economic activity, innovation, and growth simply couldn't take place without data. Big data represents a sea change in the technology we draw upon for making decisions. Organizations will integrate and analyse data from diverse sources, complementing enterprise databases with data from social media, video, smart mobile devices, and other sources. The evolution of information architectures to include big data will likely provide the foundation for a new generation of enterprise infrastructure. To exploit these diverse sources

of data for decision-making, an organization must develop an effective strategy for acquiring, organizing, and analysing big data, using it to generate new insights about the business and make better decisions.

The previously nebulous definition of “big data” is growing more concrete as it becomes the focus of more applications. As seen in Figure 2 (below), volume, velocity and variety make up three key characteristics of big data:

- Volume. Rather than just capturing business transactions and moving samples and aggregates to another database for analysis, applications now capture all possible data for analysis.
- Velocity. Traditional transaction-processing applications might have captured transactions in real time from end users, but newer applications are increasingly capturing data streaming in from other systems or even sensors. Traditional applications also move their data to an enterprise data warehouse through a deliberate and careful process that generally focuses on historical analysis.
- Variety. The variety of data is much richer now, because data no longer comes solely from business transactions. It often comes from machines, sensors and unrefined sources, making it much more complex to manage.

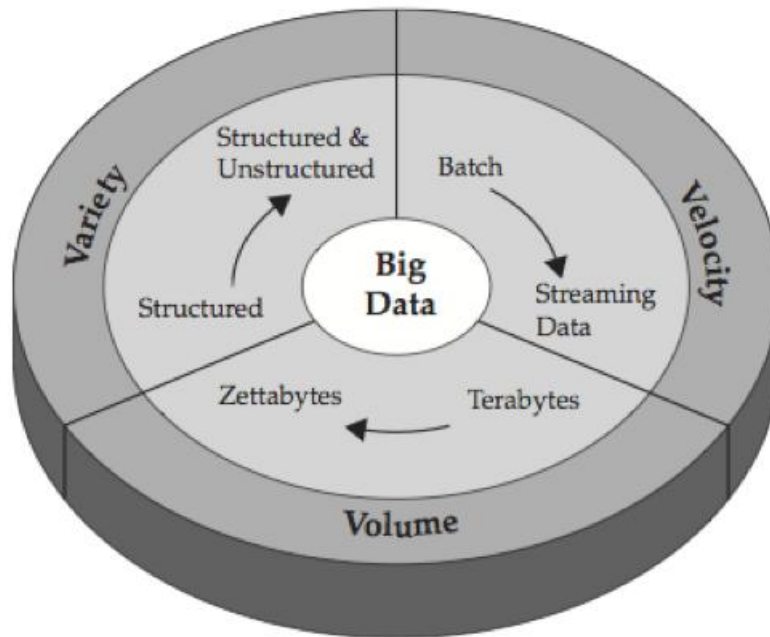


Figure 2: The three characteristics of big data: volume, velocity and variety

IX. NOSQL DATABASES

The term NoSQL has been around for just a few years and was invented to provide a descriptor for a variety of database technologies that emerged to cater for what is known as "Web-scale" or "Internet-scale" demands. In computing, NoSQL (commonly interpreted as "not only SQL") is a broad class of database management systems identified by non-adherence to the widely used relational database management system model. NoSQL databases are not built primarily on tables, and generally do not use SQL for data manipulation. NoSQL database systems are often highly optimized for retrieval and appending operations and often offer little functionality beyond record storage (e.g. key-value stores). The reduced run-time flexibility compared to full SQL systems is compensated by marked gains in scalability and performance for certain data models. In short, NoSQL database management systems are useful when working with a huge quantity of data when the data's nature does not require a relational model. The data can be structured, but NoSQL is used when what really matters is the ability to store and retrieve great quantities of data, not the relationships between the elements. Usage examples might be to store millions of key-value pairs in one or a few associative arrays or to store millions of data records.

The fledgling NoSQL marketplace is going through a rapid transition – from the predominantly community-driven platform development to a more mature application-driven market. Scaling up web infrastructure on NoSQL basis have proven successful for Facebook, Digg and Twitter. Successful attempts have been made to develop NOSQL applications in the biotechnology, defence and image/signal processing. Interest in using key-value pair (KVP) technology has re-emerged to the point where the traditional RDMS vendors evaluate strategy of developing in-house NoSQL solutions and integrating them in current product offers. It will not take long before we'll see acquisitions driven by emerging NoSQL technology. The future deals will likely be made to better compete both in platform offering and in vertical market segments.

X. CONCLUSIONS

Applications in domains such as Multimedia, Geographical Information Systems, digital libraries, big data demand a completely different set of requirements in terms of the underlying database models which conventional relational database can no longer handle. The conventional relational database model is no longer appropriate for these types of data. Furthermore the volume of data is typically significantly larger than in classical database systems. Finally, indexing, retrieving and analyzing these data types require specialized functionality, which is not available in conventional database systems. Hence, a new direction, such as described above, in DBMS is necessary.

REFERENCES

- [1] G. Slivinskas, C. S. Jensen, and R. T. Snodgrass. Query Plans for Conventional and Temporal Queries Involving Duplicates and Ordering. In Proceedings of the 16th International Conference on Data Engineering, San Diego, California, February/March 2000
- [2] Silberwschatz., A., and Korth, H., [2000]. "Database System Concepts", 4th ed., McGraw – hill, 2000.
- [3] <http://answers.yahoo.com/question/index?qid=20070705045634AAcuhyF>
- [4] <http://www.gislounge.com/what-is-gis/>
- [5] <http://www.tech-faq.com/multimedia-database.html>
- [6] J. Clifford and A. Tuzhilin (eds.). Recent Advances in Temporal Databases: Proceedings of the International Workshop on Temporal Databases. Workshops in Computing Series. Springer- Verlag 1995.
- [7] NoSQL Market Forecast 2013-2018 <http://www.marketresearchmedia.com/?p=568>
- [8] A guide to big data workload management challenges <http://www.datastax.com/wp-content/uploads/2012/08/DataStax-AnalystReport-GigaOM.pdf>
- [9] M. Levene and G. Loizou. A Guided Tour of Relational Databases and Beyond. Springer- Verlag 1999.
- [10] <http://my.safaribooksonline.com/book/databases/9788177585674/emerging-database-technologies/ch21lev1sec2#X2ludGVybmFsX0h0bWxWaWV3P3htbGlkPTk3ODgxNzc1ODU2NzQIMkZjaDE3JnF1ZXJ5PQ>
- [11] <http://en.wikipedia.org/wiki/GIS>
- [12] Rob, P., and Coronel, C. [2000] Database Systems, Design, Implementation, and Management, 4th ed., Course Technology, 2000.
- [13] <http://blogs.gartner.com/doug-laney/deja-ppvvue-others-claiming-gartners-volume-velocity-variety-construct-for-big-data/>
- [14] "NoSQL 2009". Blog.sym-link.com. 12 May 2009. Retrieved 29 March 2010.
- [15] Jaymin Patel, Temporal Database System
- [16] Thomas Connolly, Carolyn Begg, Database Systems – A practical approach to Design, Implementation, and Management. Addison – Wesley 2010.
- [17] http://books.google.com.om/books?id=AyY1a6-k3PIC&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false
- [18] <http://www.tech-faq.com/multimedia-database.html>
- [19] Jiawei Han and Micheline Kamber, Data Mining Concepts and Techniques 2006
- [20] <http://en.wikipedia.org/wiki/NoSQL>
- [21] Tweed, Rob; George James (2010). "A Universal NoSQL Engine, Using a Tried and Tested Technology"
- [22] Christian S. Jensen, Introduction to Temporal Database Research
- [23] McKinsey Global Institute, Big data: The next frontier for innovation, competition, and productivity