Sndh Univ: Res: Jour:(Sc: Ser:) Vol: 31 (2) 31-38 (1999)

DBMS FOR COMING DECADE

Imdad Ail Ismaili* Nasrullah Memon** and A.H.S Bukhari*

Institute of Information & Technology, University of Sindh, Jamshoro.

Abstract

DBMS (Database Management System) grew up in an OLTP (Online Transaction Process) environment. During the last decade research and development tried to extend the usability of DBMS to new application areas: data warehouses, data mining, object-oriented applications, etc. This was either done by extending the existing SQL DBMS technology or by introducing OODBMS (Object Oriented Database Management System) technology.

In this paper, it is predicted that using OODBMSs, data mining purports to be the new database technology for the coming decade.

Key words: Data base Management system,

Introduction

A database management system, or **DBMS,** gives the user access to their data and helps them transform the data into information. Such database management ssystem includes **dBase, Paradox, IMS** and **Oracle**. These systems allow users to create, update, and extract information from their databases. Compared to manual filing system, the biggest advantages to a computerized database system are speed, accuracy, and accessibility.

Historical Background

The history of database research over the past 30 years is one of the exceptional productivity that has led to the database system becoming arguably the most important development in the field of software engineering. The database system is now the underlying framework of information system, and has fundamentally changed the way many organizations operate.

The database technology for information systems has undergone two generations of evolution, and its third generation is currently under development [1]. The transition from one generation to the next has always been necessitated by the ever increasing complexity of database applicants and the cost of implementing, maintaining, and extending these applications. The roots of **DBMS** lie in file-based system, such as **ISAM** and **VSAM.** The first generation was hierarchical, such as **IMS** and system 2000 and **CODASYL** (Conference on Data System Languages), such as **IDS, TOTAL, ADABAS, IDMS**, etc. These systems realized the sharing of an integrated database among many users within an application environment. The lack of data independence and the tedious navigational access to the database gave rise to the second-generation namely Relational Database Management System **(RDBMS).** It is characterized by notion of declarative query. Third generation will be characterized by richer data model and a richer set of database facilities necessary to meet the requirements of applications beyond the business data processing applications for which the first two generations have been developed.

Relational Database Management system

It is originally described by E.F. Codd in 1970 [1] and later implemented by **IBM**

**Mehran University of Jamshoro.

CamScanner

and others in products. It is completely different to earlier models and is based on the branches of Mathematics including Set Theory, Relational Algebra and Relational Calculus. It required data to have a simplified structure-atleast to First Normal Form **(1NF),** preferably at least **3NF.** It has more overheads than earlier models.

At the core of the relational model is a concept of a table (also called a relation) in which all data is stored. Each table is made up of records (rows also known as tuples) and fields (columns also known as attributes).

The eventual standard to describe **RDBMS** is published by **ANSI** as the **X3H2** specifications and its **ISO** counterpart. It is referred as **SQL** Plus version number.

In the relational model, operations that manipulate data do so on the basis of the data values themselves. Thus, if any one wishes to retrieve a row from a table, for example, he does so by comparing the value stored within a particular column for that row to some search criteria.

For example, he might say "Give me all rows from **"STUDENTS** table, which have "Memon in the **"FIRST_NAME"** column.

The database might return a list, which look essentially like this:

Memon	Nasrullah	SID-20000017
Memon	Akram	SID-20000018

He could then use data from a received row to query another table. For example, he might say **OK,** I want to know what grade "Memon Nasrullah" received in **'CMS-110** Database Systems'. So the 'Student **ID'** number from previous query as the keyword in the next query is used.

Thus every thing within the relational model can be stored in tables. A user can query information concerning table names, access rights, or some data and the results of these queries would then be presented to the user in the form of a table.

Relational **DBMSs** are currently the dominant **DBMS** with sales around US$5 Billions [2]. They have proved themselves for managing large volumes of well-structured data. The technology is mature and well understood with **SQL** skills in plentiful supply.

Relational **DBMSs** are commonly used for high volume online transaction processing **(OLTP)** systems by many large airlines and banks. Because the technology is also mature, they are able to provide 24x7x365 availability and security levels to **US DoD** standards.

There are some well-known shortcomings of the conventional **DBMS.**

It is too simple for modeling complex and nested entities. It has only a limited set of atomic data types. Application programs are implemented some algorithmic programming language and some database language embedded in it. Database languages are very different from programming languages, in both data model and data structure. The impedance-mismatch problem motivated the development of second-generation languages.

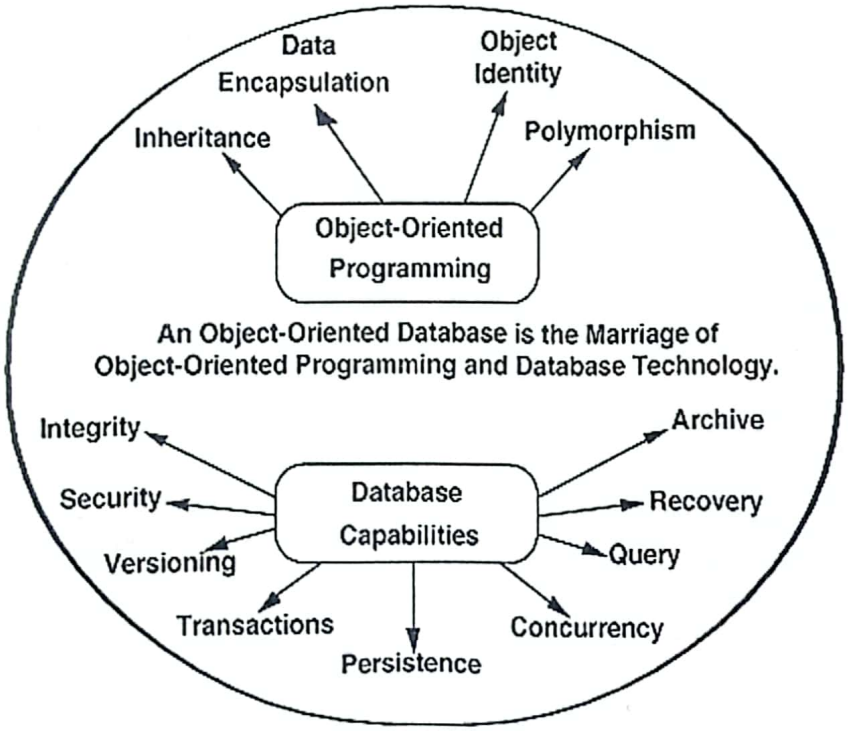
These shortcomings of the conventional **DBMS** has provided impetus for database professionals for most of the 1980s to pave the way for third-generation, known as Object Oriented Database Management System **(OODBMS).** There are currently at least two proposed approaches for transition from second generation to third generation technology: Extended Relational Database Management System

CamScanner

(also known as Object Relational database Management System **(ORDBMS)** and Object Oriented Database Technology. The fundamental difference between them is the basic data model and database language.

Object-Oriented Database Management System (OODBMS)

It is marriage of object-oriented programming and database systems [3]. As shown below:



They were invented around 1989. In simple terms, they are an extension of object-oriented software technology into database realm. The new object-oriented products typically emerged from startup companies as follows:

Relational vendors have responded to the new object-oriented database products by creating hybrid products incorporating relational and object-oriented functionality.

Object-Relational **DBMS** try to bridge the gap between the two previous technologies and attempt to provide the best of both. Most of the major Relational vendors have indicated that they moving towards Universal Servers, which provide support for complex objects as well as existing Relational structures.

It is not clear whether hybrids **(ORDMS)** or **OODBMS** will dominate.

However, Laker Consulting views [4], based on their years of experience with both object-oriented and relational databases, is that while major software vendors may like locking their customers into a "one size fits all" object-relational product, a sound approach is to use either "pure" object-oriented or relational technology on case-by-case basis. The hybrid approach often has more drawbacks than advantages.

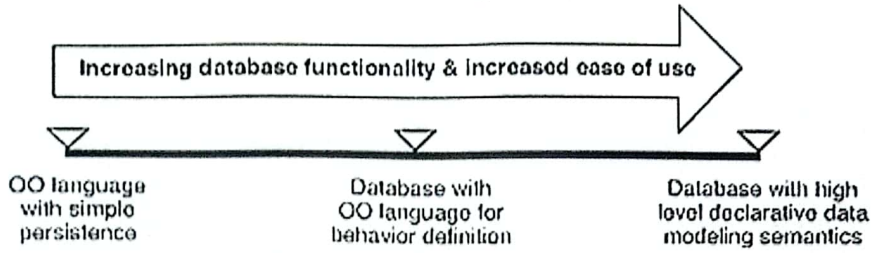
CamScanner

OODBMS started initially in the engineering and design domains, but have recently become favored for financial and telecommunications applications as well. Whilst the market is still small (a few percent of Relational) [2], the technology have proved itself and although may not supersede Relational, continues to find new application areas, such as World Wide Web (**WWW**).

These systems require plenty of training and guidance for programmers to be skilled in using the sophisticated tools.

The sign of maturation of **OODBMS** is the establishment of industry groups to standardize on different aspects of technology. A significant is noticed for the development of such standards. There is no official standard of **OODBMS** [5]. The de-facto standard is, the Object Database Standard: **ODMG**-V2.0.

Today, **OODBMS** vendors are adding more database features to their products to provide the functionality one would expect from a mature **DBMS**. This evolution moves us to the mid-point of the evolutionary scale shown in the following figure:



By using this technology and to meet the multimedia demands of tomorrow, multimedia database system are under development. The Object-oriented method will help to meet the goal of becoming database for 'advance' applications. The Object-Oriented **DBMS** helps multimedia information systems to meet their requirements and increase their capability. One of the characterizing features of these systems is their integration of large amounts of complex structured data. Since multiple streams need to be coordinated both spatially and temporally for data presentation. This implies the need for a good **DBMS** that could support data integration and provide an integrated repository to store large amount of multimedia data. The repository problem can be solved by Data Warehouse research that is conducted by many industries. The use of data warehouses will satisfy the need of large and integrated database system.

During last 12 to 18 months there has been major shift toward the Internet as the platform for remote access. In future, database vendors must ensure the integration of databases and their components and universal access to them from anywhere in the world.

Internet technologies provide unparalleled advantage because the Internet is a global network that provides universal access.

Therefore the new frontier is the use of Internet technologies to facilitate database access and to provide services to customers, partners, employees, and the general public.

Data Warehouses

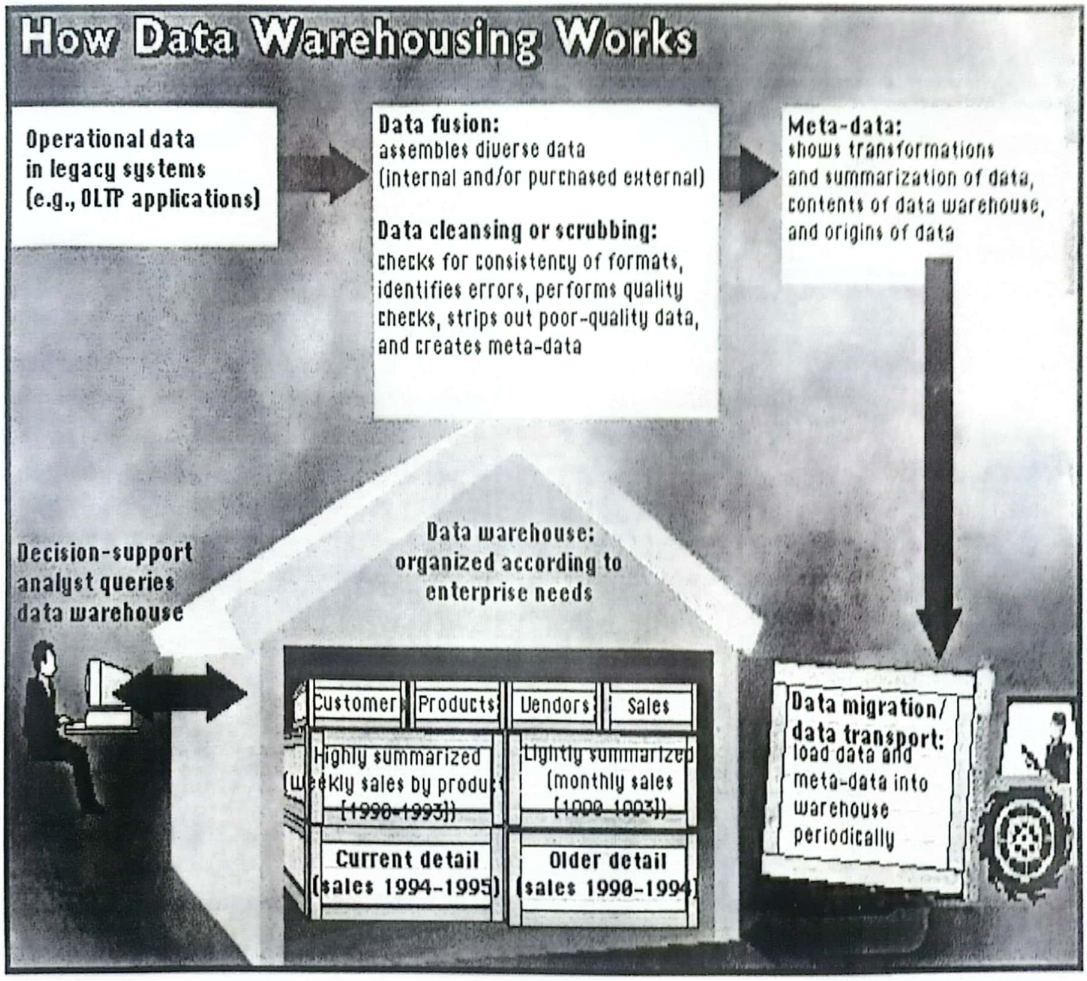
The data warehouse model [6] is a new and immature database concept, which is itself, a new information technology architecture whose focus is to alter the

CamScanner

model of industrial production. Data warehouses are being developed for the purpose of providing a framework in which to perform Data Mining, which is the discovery of useful patterns in massive databases with mass collections of data stored in tertiary storage for the purposes of driving decision support systems.

business questions that traditionally were too timed consuming to resolve.

Although data mining is still in its infancy, companies in wide range of industries including retail, finance, health care, manufacturing transportation, and aerospace are already using data mining tools and techniques to take advantage of

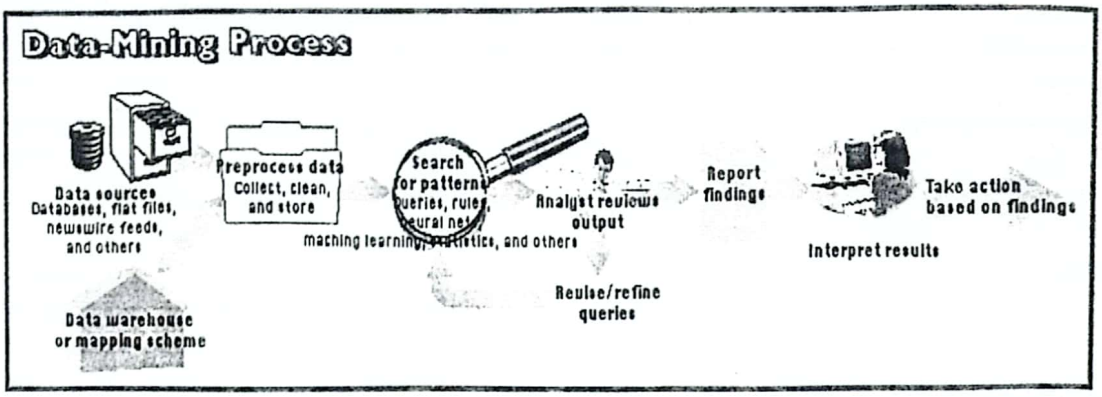


Data Mining

Data Mining or Data Discovery [6], is the computer-assisted process of digging through and analyzing enormous sets of data and then extracting the meaning of the data. Data Mining tools predicts behaviors and future trends, allowing businesses to make proactive, knowledge driven decisions. Data mining tools can answer historical data. By using pattern recognition technologies and statistical mathematical techniques to sift through warehouse information, data mining helps analysts recognize significant facts, relationships, trends, patterns, exceptions and anomalies that might otherwise go unnoticed.

For business, data mining is used to discover patterns and relationships in the data in order to help make better business

CamScanner



decisions. Data mining can help spot sales trends, develop smarter marketing campaigns, and accurately predict customer loyalty. Specific uses of data mining include market segmentation, customer churn, fraud detection, direct marketing, interactive marketing, market basket analysis and trend analysis.

Data mining can generate new business opportunities by Automated prediction of trends and behaviors or Automated discovery of previously unknown patterns.

By using modeling technique, data mining is able to tell important things that any one didn't know or going to happen next. As a simple example of building a model, consider the director of marketing for Telecommunication Company. He would like to focus his marketing and sales efforts on segments of population most likely to become big users of long distance services. He knows a lot of his customers, but it is impossible to discern the common characteristics of his best customers because there are so many variables. From his existing database of customers, which contains information such as age, sex, credit history, income zip code, occupation, etc., he can use data mining tools, such as neural networks, to identify the characteristics of those customers who makes lots of distance calls. For instance, he might learn that his best customers are unmarried females between the age of 24 and 32 who makes in excess of $60,000 per year. This, then, is his model for high value customers, and he would budget his marketing efforts accordingly.

The analytical techniques used in data mining are often well-known mathematical algorithms and techniques. What are new are the applications of those techniques to general business problems made possible by the increased availability of data and inexpensive storage and processing power. Also, the use of graphical interfaces has led to tools becoming available that business experts can easily use.

Some of the tools used for data mining are known as Artificial Neural Networks and Genetic Algorithms [6].

The future of data mining is described as under:

In short-term, the results of data mining will be in profitable.

In medium-term, data mining may be as common and easy to use as email. By using these tools, it would be easy to root out a phone number of long lost classmate.

The long-term prospectus is truly exciting. Computers may reveal new treatments for diseases or new insights into the nature of universe.

CamScanner

Conclusion

Relational **DBMSs** are currently dominant and mature database technology and have proved themselves for managing large volumes of well-structured data. Where as Object **DBMSs** started initially in the engineering and design domains, but have recently become favored for financial and telecommunication applications as well. Whilst the market is still small (a few percent of Relational), the technology has proved itself and although may not supersede Relational, continues to find new application areas, such as World Wide Web.

Using **OODBMSs,** a data warehouse environment can offer an organization a competitive advantage over their business rivals, if implemented in the right way. It should be kept simple, with clearly defined aims and goals from the outset and include long-term planning for information delivered in the future.

Data Mining (also known as Knowledge Data Discovery-**KDD)** is a rapidly expanding field with promise for great applicability. Knowledge Discovery purports to be the new database technology for the coming decade.

It is anticipated that commercial database systems of the future will include **KDD** capabilities in the form of intelligent database interfaces.

It is advisable for an organization that is using **RDBMS** for a number of years to choose **OODBMS** having data mining applications after 10 years time.

References

[1] Thomas M.Conolly, Carolyn E. Begg "Database Systems", Addisson Wessley.

[2] Choosing a strategy for Object Data Management-Relational, Object, or Hybrid

http://www.soi.city.ac.uk/~akmal/ot97.html

[3] **Object Oriented Database Management Systems Revisited**

http://www.dacs.dtic.mil/tech/oodm2/introduction.html

[4] **Database Innovations**

http://www.laker.com.au/database.htm

[5] **IDC Bulletin # 14821E**

http://www.cai.com/products/jasmine/analyst/ide/14821E.htm

[6] **Data Mining**

http://www.eco.utexas.edu/~norman/BUS.FOR/course.mat/Alex/

[7] **Knowledge Discovery in Databases: Tools and Techniques**

http://www.acm.org/crossroads/xrds5-2/kdd.html

[8] **Data Warehousing and Data Mining**

http://www.web.city.ac.uk/~dz542/cw.html

[9] **AIT-Lecture 11-Databases**

http://www.csm.uwe.ac.uk/~tdrewry/aitlec11.htm

[10] **Atzeni, Ceri, Paraboschi and Torlone "Database Systems: Concepts, Languages, and Architectures, McGraw-Hill 1999.**

CamScanner