

History of Database Applications

Moaz Mansour (47)
University of Rochester
500 Joseph C. Wilson Blvd., Rochester, NY 14627
+1(585)733-7670

moaz.mansour@rochester.edu

Wengang Yang (77)
University of Rochester
500 Joseph C. Wilson Blvd., Rochester, NY 14627
+1(585)629-2778

w.yang@rochester.edu

ABSTRACT

Databases and database management systems have become a very essential part of today's information systems. Most of our day-to-day activities certainly include a database interaction. It starts nowadays with the newly born children who get a record number in most of the governments in our modern world. It is involved in everything after that, going to school, graduating, getting a job, running or managing a business, buying travel tickets or even getting your groceries.

The purpose of this project is to take a look on the origins of databases from different perspectives (1) Its development over the years (2) different technologies adopted and (3) to take a look forward at what the future may hold for databases.

1. INTRODUCTION

The ancient or modern society, humans and organizations always have the everlasting demand to store and retrieve information and data. Long ago, simple and non-computerized database system was developed and used in hospitals, governments, and libraries, of which some basic concepts of the design are still applied in nowadays database application. Following is the revolution timeline of the databases and their different types.

Origins of the database go back to libraries, governmental, business and medical records before the computers were invented. And back to 1960's, two main data models were developed - network model CODASYL (Conference on Data System Language) and hierarchical model IMS (Information Management System).

The first generation of database was navigational. The relational database model was invented in 1970's by E.F. Codd. whose model later became the standard principal of database systems. Also during that time, IBM developed the prototype system called System R. During 1980's, it's the time when relational system began to commercialized and the concept of the object-oriented database was developed.

In 1990's, developments were more focused on the client tools such as Oracle Developer and VB. When came to the 21st century, more interactive applications appeared and there is a growing trend to

This work is to be used for class purposes at the University of Rochester as part of Databases Graduate Course. For any use beyond those covered by this note, obtain permission by emailing moaz.mansour@rochester.edu or w.yang@simon.rochester.edu

Copyright 2017 Moaz Mansour, Wengang Yang

provide more sophisticated programming logic within the database structure.

2. Database Development Timeline:

2.1 The Sixties

Back to 1960's, Cobol and with Cobol, and later Fortran, the first non-proprietary programming language has been developed. And this makes it possible to create enterprise computer systems. Based on this, two main data models were developed, which are network model by C.W.Bachman (Bachman 1965) and hierarchical model IMS (Information Management System).

In the late 1960's, IBM created the very first commercial system IMS for American Airlines to help them store reservation data. Both kinds of DBMSs (hierarchical and network) were accessible using Cobol, which makes it controllable to maintain and manage the database, but it's still complicated and time consuming.

2.2 The Seventies

In 1970, a totally different database model was created by Edgar F. Codd. He published his paper "A Relational Model of Data for Large Shared Data Banks". In the paper, Codd stated that all the data in a database could be represented as a tabular structure (tables with columns and rows). In the past, database applications are considered to search for data only by content instead of following links.

Peter Chen, proposed a new database model in 1976, which is the entity-relation model. And it serves as the foundation of many systems analysis and design methodologies.

2.3 The Eighties

In the early 1980s, research was centered on another type of database. It's mainly due to the demand to solve the requirement to deal with large scale of data and complicated object. To accomplish these tasks, the database had to be able to store classes and objects and the objects associations and methods, and the object-oriented DBMS (OODBMS) appeared. In the late 1980s several companies had invented OODBMSs (e.g. ObjectDesign, Versant).

Also, SQL has become the standard query language from then. And RDBMS was beginning to be widely used in market. Not long after the IMB PC's coming out, more and more DB companies were built as well as the products.

2.4 The Nineties

For 1990's, most of the progresses were made concentrating on client tools for application development, such as PowerBuilder (Sybase), Oracle Developer, and VB (Microsoft).

The client-server model for computing served as the norm for future business analysis and decision. around the mid 1990's after the internet emerged to the market, open source solutions came online with widespread use of GCC (GNU) Compiler Collection), CGI (Computer Generated Imagery), Apache, and MySQL. Online Transaction Processing (OLTP) and Online Analytic Processing (OLAP) became much more popular because of the wide application of POS.

2.5 Early 21st Century

When coming to 21st century, database applications continue to develop. Thanks to the development of PDAs (Personal Data Assistants), consolidation of vendors, transactions, more and more interactive database applications have been created. And apart from storing and retrieving data, the database application began to adopt more sophisticated programming language as well as more advanced features like triggers, cascading update, and delete. This makes database run more consistently within tables.

3. Database Types:

A database (DB) is basically grouped amount of data organized in a way that allows a computer program termed the Database Management System (DBMS) may easily manage it.

A database management system (DBMS) is a collection of software programs that gives a user the interaction ability to store, modify and extract data from a certain database. It enables the definition, creation, query, update and administration of databases.

There are three main categories of database management systems, and these are hierarchical, network and relational models. Each type differs in how the DBMS organizes data internally, and this determines the speed and efficiency of data retrieval.

3.1 Network Model

The network model is often used for a database management system if the relationships between the data records are defined in form of a graph. The records in this form are connected together via links and any given record may have several parent and/or dependent records. The network model permits cycles and allows many-to-many relationships to be expressed in a simple graph-like structure.

General Electric's Integrated Data Store (IDS) and the Integrated Database Management System (IDMS) are two significant examples depicting the adoption of network model in databases.

3.2 Hierarchical Model

The hierarchical model is often used for a database management system if the relationships between data records are defined in form of a tree-like structure. It implies simple relationships yet inflexible. The records in this form are connected via links and any given record may have only one parent, though, it can have several child nodes.

The hierarchical model is a restricted version of the network model where at any given point the whole tree needs to be traversed in

order to retrieve data. The IBM Information Management System (IMS) is one of the most widely used hierarchical databases.

3.3 Relational Model

A Relational Database Management System (RDBMS) is a system where a relation is defined as a set of tuples represented by a table. Each column in the table contains data of the same type. Business rules are implemented in this kind of database by employing constraints in form of Boolean expressions to provide restrictions on the kinds of data to be stored.

Relations' tuples are identified by unique fields called keys. The relational database can be accessed rapidly by an index as it allows direct look-up rather than checking all tuples.

The structured query language (SQL) is the computer language that deals with relational database. It is used to perform common operations like data storage and retrieval on the database.

The relational database deals with undefined or missing information by a three-valued logic where test results would be True, False or Null. The data in the relational model are represented as a mathematical n-ary relation. Each row in a table represents one n-tuple of a relation and cardinality is defined as the number of tuples in a relation. A tuple of a relation is considered a set, thus there is no defined ordering in a relation.

The relational model does not itself define the way in which the database handles concurrent data change requests. These changes are handled by a transaction model. A transaction is a transformation of state which has the properties of atomicity, consistency, isolation and durability, ACID.

3.3.1 Structured Query Language (SQL)

The most popular and widely used database language SQL which was initially called SEQUEL. It was designed by IBM to retrieve and manipulate data in their database. It included common operations like insert, delete, update, query, schema creation and modification.

The most common operation in SQL is the query command, also known as SFW. It is performed by the SELECT statement to retrieve data FROM one or more tables WHERE target tuples satisfy certain condition.

The Data Manipulation Language (DML) is a subset of SQL used to add, update and delete data. The Data Definition Language (DDL) manages table and index structure.

3.3.2 Oracle Database

An Oracle database is a collection of data treated as a unit, and it is used to store and retrieve related information. It allows concurrent access to the data, thus it is a multi-user environment friendly. It is a secure database as it has the ability to prevent unauthorized access to the database and it also provides smooth recovery of information in the event of an outage.

The oracle database consists of one or more physical data files which contain all the data, along with a control file which specifies the database physical structure. In addition, it includes a logical storage that controls disc space usage.

The schema in an Oracle database is a group of objects which defines the logical structures to the database's data. They include structures such as tables, views and indexes.

Similar to relational model, tables are the basic unit of data storage in an Oracle database, and each table has several rows and columns. The index function is not different from a relational database where it quickens access to table data, thus, enhances the data retrieval performance.

Data dictionaries is a common part of any Oracle database, it is the place where information about the logical and physical structure of the database can be found. It is created and updated automatically to ensure data consistency and accuracy.

The Oracle database server and application tools are managed by a Database Administrator (DBA) who allocates system storage and requirements. The access to the database is always monitored and controlled as well as its performance.

4. Next Generation Databases

With the emerging of big data in the market nowadays, a need for more sophisticated database system arouse. That leads to new technologies called NOSQL databases.

The successors of IBM in the databases field now like Google started developing new systems to handle massive amounts of data. Google bigtable was introduced early 2000s to cope with data variety, velocity and volume increase, aside, the new SQL analytics which was also introduced at the same time.

NOSQL databases came into play to fill in the gap of handling unstructured data and that when we started hearing of CouchDB, MongoDB, Cassandra and Redis. Aside, Hadoop which is another merging player too in this space, has gain traction for analyzing petabytes of data especially in the business intelligence area.

Today, businesses require real-time analytics on operational data. The SQL scale-up proves too costly but scale-out removes resource constraint. Scale-out provides real-time analytics with high volume transactions. Google and Clustrix are pioneers in this space.

Database experts are forecasting that the future belongs to the scale-out SQL which will replace the single node SQL. In addition, Data warehouse type analytics will become available in real-time database.

5. CONCLUSION

The future of databases would be of huge varieties and possibilities. The volume of database could be bigger and bigger and can tackle with huge amount of data set in a more innovative way. In the market there already exists mobile database. Technology keeps changing and improving, will solve people's demand in a more convenient way, all you need to do is click. In business area, Distributed transaction processing is already becoming the standard for it. The future of database will continue to revolutionize and beyond our imagination.

6. ACKNOWLEDGMENTS

Our thanks to ACM SIGCHI for allowing us to modify VLDB conferences templates they had developed.

7. REFERENCES

- [1] Berg, K. L., Seymour, T., & Goel, R. History of databases. International Journal of Management & Information Systems (Online), 17(1), 29. (2013) Retrieved from <https://search.proquest.com/docview/1418458139?accountid=13567>
- [2] Gerard O'Regan. History of Databases. Introduction to the History of Computing. Springer International Publishing Switzerland. (2016). Retrieved from <http://download.springer.com/static/pdf/64>
- [3] Guy Harrison. Next Generation Databases: NoSQL, NewSQL, and Big Data. Apress Media, LLC. (2015)