IS5102 – Database Management Systems

Practical 3

University of St Andrews

Group B — Topic 1: Before Databases

Main Parts	Lead Author
Hardware and software (2.)	220032472
Data structure (3.)	220028783
Data (4.)	220032632
Users and usage (5.)	220010697

18 November 2022

Word Count: 5899 words

Abstract

With the advancement of modern technologies, the current world is full of various kinds of data. The nowadays information systems generally use databases to store data, by which data can be managed in a high level of efficiency, including convenient organisation, querying and modification of data, while backup and recovery can also easily to be achieved; however, before the modern databases were invented, there was also demand for collecting, storing, tracking, etc. the information. This paper introduces how data was managed before digitisation as well as operated by the modern database management systems, showing the similarities and differences between the two different systems in terms of hardware, software, structure, data and users. Additionally, it also provides an outlook on the future direction of database development.

Keywords: *DBMS, Comparison, Pre-digital Data Management System*

Catalogue

1. Introduction

The information plays a significant role in humans' lives. Although databases were not clearly defined before digitisation, there were still many options to manage data. At the beginning, people used to paint on walls to record information in order to convey it (Miyagawa et al., 2018). As time passes by, with continuous development of economy and technologies, as well as mankind grew in size, there was more and more information needed to be handled. As a result of this requirement, database was finally devised in 1970.

A database is a collection of information that is stored in a certain and structured way (Oracle, n.d.). There are many advantages of using DBMS, including fast processing speed, reducing data redundancy, integrated data management, higher level of security, etc.; however, it can also be further improved in some aspects. This paper gives relatively brief introduction of the methods of data management before and after digitisation and focuses on comparing the similarities as well as differences in terms of hardware, software, structure, data content and users, which hope to provide a reference for the further development of the database in the future.

2. Hardware and software about data management

2.1 The evolution of hardware and software

Data was always an integral part of human life. There were multiple ways used to store information throughout history. With the current advanced methods of data management, we have access to all of the public information of the world under centralized servers and can have instant access to such vast information with fewer clicks on the internet.

Before the development of the computer, data was still being produced, although it was much more slowly than today. Additionally, the generated data needed to be kept somewhere in a specified way. The following is a summary of the evolution.

a) Data was initially managed by hand processing

Using a simple pen and a sheet of paper was the only way to store and transmit information before. Everything had to be written down letter by letter before Johannes Gutenberg created the printing machine in 1440. Such systems had their own advantages and disadvantages compared to modern database management.

Before the advent of the database system, the file-based system was relied upon to store and retrieve information; however, this was a more costly approach. One way of storing information was to keep them in form of permanent files. Each of the application programs was written as separate files, and added new contents according to the demands.

b) The use of punch cards and electromechanical devices

Using the punch card to store and read data was relatively fast and agile. The programs were read from a stack of cards with holes punched in them, which represented each character. Since a punch card could only hold a limited amount of data, storing program information required a stack of them.

People could represent a sequence of instructions that would translate to "on" or "off" orders, and looms would comprehend by determining whether a hole had been punched in the punch card. This straightforward approach established the fundamental blocks of the machine language that modern computers use to communicate.

c) Use magnetic tape as a management method

Punch cards were still in use until 1960, after which Magnetic storage came into existence. Magnetic tape encoders were considered as a replacement for punch cards and the idea grew wildly with the invention of computer systems. Magnetic tape was extremely prone to disintegration even if it was ideal for short-term use. This process could start after 10–20 years, depending on the environment.

d) The occurrence of DBMS and database structure

The first occurrence of a computerized form of storing data evolved in the year 1960, with the increase in production of cost-effective computers as an option for a private organization. IBM developed the Integrated Management System in the late 1960s, which was the origin of the database management system (GeeksforGeeks, 2022). In the same period, several theoretical data models were put forth (Quickbase, n.d.). The network data model and the hierarchical data model were the first two. The third theoretical model, which was the relational model, was proposed by Edgar F. Codd in the late 1960s (Foote, 2017). Towards the mid of 1970s, P. Chen proposed Entity-Relationship or ER model. This model made it possible for designers to focus on data application instead of logical table structure. Database management systems were built for computer by IBM (DB2) (Murphy, 2016) and gained popularity with the rapid rise of sales of the computer. Object-oriented database management system emerged towards mid of 1980, supporting object-oriented view of data. SQL became the standard of ANSI and OSI in 1986 and 1987 (Foote, 2017).

e) Client-server processing that was dispersed

The key event of the raise of World-wide-web in the 1990s functioned as a catalyst towards building client-server database systems. Database showed exponential growth in usage in this year. Internet-based transactions and processing began to become mature. Creation of MySQL, APACHE and other open-source systems were widely introduced via Internet.

2.2 Comparisons of pre-digital systems and modern DBMS based on hardware and software

a) Data Retention Policy

Historical data were stored in multiple forms such as rocks, papers, punch cards, or magnetic tapes. Some of the data might last till the end of lifetime of the material. But some of these storage elements could not, in terms of the longevity of data storage. For example, papers, as everyone knows, had its own slew of problems as a data storage option. It is expensive to make, harmful to the environment, and impossible to extend. Combined with its fragility to the elements like fire and water, it was clear there needed to be a better data storage system in the future.

The Modern database management system has attached configurations and policies when it comes to data retention. This helps to keep the database clean and clear from outdated and obsolete information by reducing the cluttering of storage units (servers).

b) Data Validation and Authenticity

Paper documents may be left on desks, printing presses, or in open spaces. Whether on purpose or not, this makes records and their contents accessible to unauthorised people. Electronic records have less risks of being overlooked, and since they may be encrypted, the data will still be secure even if they are copied or stolen by an unauthorised person.

Numerous safeguards are used to protect database management systems from intentional cyberattacks and unauthorised use under the umbrella of database security. Modern DBMS includes provisions for restricting access to the database by unauthorized users. One of the methods is via granting access control, which is done by creating user accounts and controlling the login process by DBMS.

c) User-friendly interface and experience

It is an evident distinction that pre-digital data storage and analysis were not very user-friendly.

The design of Antikythera Mechanism created by Greek scientists was remarkable for the degree of component miniaturisation and intricacy, which was comparable to that of astronomical clocks from the fourteenth century. Michael Wright pointed out that the Greeks of this era could have made systems with at least 30 gears (Wright, 2006) which could mean less user-friendliness on data storage and computation.

Modern database systems allow various interfaces, which simplify the ways of using the system to store and retrieve information. Some of the user-friendly interfaces provided by DBMS are menu-based, form-based, graphical user interface or etc. DBMS should have a contemporary, user-friendly layout and provide a wonderful experience for users in order to increase productivity and cut down processing time.

d) Automated data analytics

Automated data analytics or data visualisation are one of the major parts of management systems, as they help to better understand the stored information. By integrating the data management system with available analytics platforms, we can speed up data analytics and visualisation (Gaenger, 2020). This will prevent the need to switch between many platforms as data can be accessed from one central location. An analytics module can assist in understanding data, regardless of whether data management system is utilized at the source or destination. These tools also allow to predict the trends of data and make righteous decisions with modern technologies.

e) Data Recovery

It is quite hard for the pre-digital management system to recover data after deletion, but for the modern database system, there are multiple ways to realize recovery if the data is erased by mistake. For instance, IBM Database 2 (DB2) can provide the service of remaining all data alterations in the DB2 recovery log. When data is changed, both the previous and new versions of the log follow the update, therefore, if there is an unexpected situation, the log files can be retrieved for data recovery (Crus, 1984).

3. Data structure of data management systems

3.1 Data structures in pre-digital times

In ancient times, people already had the awareness to record data. As early as the Late Stone Age (50,000 to 12,000 years ago), animal bones for producing tally sticks, as exemplified by Ishango Bone, demonstrate that people used sticks or bones with notches marked on them for trade or supply activities (Heidari et al., 2022). This method of recording is based on the principle of split counting, taking advantage of the fact that the same stick or bone breaks a compatible interface and records the same data at both ends, thus ensuring that the transaction record is not altered by one side unilaterally. This structure is similar to two associated fields in a modern relational database.

Incan Quipu is a collection of coloured knotted cotton threads used as data storage equipment by the ancient Incas in 3000 B.C. It has one main rope that would have a variety of different coloured sub-ropes on it. Each colour of string had a different meaning, such as the number of men in the village, the number of pregnant women, the amount of food, etc. The ancient Incas tied knots in these different coloured sub-ropes to record data (Keister, 1964). This is similar to a table in a modern database, where the main rope can be a table that stores information about a village, the different coloured sub-ropes are fields in it, and the number of knots is a value of type int.

With the development of human civilisation, the scale of data that needs to be stored has grown, and with the advent of paper, paper-based data management systems have emerged as the longest-running form of record keeping until nowadays. (McCarthy, 2022) Take the example of the Great Library of Alexandria, whose first director, Zenodotus, divided the collection into three main categories and ordered them by author's name, which was the first recorded use of metadata. (Phillips, 2010) From here, the concept of a database management system began to emerge. The whole library as a database, and each category of books as a table, in which the books as fields and data are organized in a certain order by the librarian (database operator).

3.2 Modern data structures

After being first described by E.F. Codd in 1970, and after half a century of development, database-related concepts have matured considerably. Modern database management systems are divided into two main categories according to their storage structure: relational databases and non-relational databases.

3.2.1 Relational database

The relational database was first proposed by E.F. Codd in 1970 (Codd, 1970), the main feature of the relational database is the consistency of affairs. It emphasises the "ACID rules", namely Atomicity, Consistency, Isolation and Durability. Relational databases store data in a structured way, with data types defined before they are deposited and complex relationships between tables.

The logical structure of the relational model of a relational database is a two-dimensional data structure, which uses this simple logical structure to store the data of entities and describe the relationships between them (Haiyan et al., 2010). Relational databases generally store a structure of b+ tree (Kieseberg et al., 2011), which is a structure with a relatively stable time complexity for both insertion and modification.

3.2.2 Non-relational database

Unlike relational databases, non-relational databases do not use a relational model and are refined to support data types and query methods (EdPrice-MSFT, n.d.). Due to the dramatic increase in the volume of modern data, non-relational databases have an irreplaceable advantage over traditional relational databases when faced with large quantities of data. (Nishtha, 2012)

a) Key-value database model

The key-value data storage model is used by many enterprises to store data, such as Facebook's five Memcached use cases which receive millions of requests per day. (Atikoglu et al., 2012) It is a large Hashmap that stores data through the key and value association structure. Its most important feature is its high concurrency and performance. For example, SILT can reach 46k queries or 23k inserts per second (Lim et al., 2011).

b) The graphical database model

The graphical database model is a storage model that summarises the data and relationships of entities in diagrams, which consist of three parts: data structure, transformation language and holistic constraints (Angles & Gutierrez, 2008). It is mainly used in situations where the relationships are more important than the data itself. Its advantage is that all information about an entity can be explicitly and clearly represented in association with other nodes (Paredaens et al., 1995).

c) Network database model

The network database model allows tree-structured children to have multiple pairs of parents, solving the hierarchical structure of many-to-many relationships (MongoDB, n.d.) The network database model is simple to design and efficient to access (Raima, n.d.). This structure is well-suited for quick searches, data access or navigation. A representative database management system is Raima.

d) The document database model

A document database is a database that uses documents as a data storage structure, a database type that considers documents as part of the database (Clifton & Garcie-Molina, 1988). As the most similar non-relational database model to the relational database, the document database model is intuitive, flexible and has a variety of APIs that allow it to be adapted to a wide range of development options (MongoDB, n.d.). JSON and MongoDB are representatives of the document database model.

3.3 Comparisons between pre-digital structures and modern structures

3.3.1 Similarities

Both pre-digital and modern database storage structures have certain correspondences between data, including one-to-many, one-to-one, many-to-many, etc. They also have a simple reflection of the ACID principle, such as a certain Consistency and Durability of the bones of a record trade record. Secondly, the pre-digitised storage structure possesses a relatively simple storage structure and logical structure, such as the classification of books, sorting by author name, etc.

3.3.2 Differences.

a) Stronger management of data integration and cleansing

Pre-digital systems did not store logically linked information as tightly as modern databases, which led to inefficiencies in integration and cleansing. Pre-digital systems were generally paper-based, so data was managed manually. When integrating data, it was necessary to manually locate the paper-based information that needed to be used, bring it together and even transcribe it. Similarly, when cleaning data, it is necessary to manually check the data and eliminate errors manually. Modern databases, on the other hand, allow for uniform and rapid batch manipulation of data by methods internal or external to the database management system. As mentioned by Nin, manual operations are quickly integrated with the aid of algorithms to clean the data (Nin et al., 2007). Modern database management systems are considerably more efficient compared to pre-digitalisation.

b) Greater logic

Modern database management systems also have the advantage of greater logical expressiveness. Thanks to the development of computers, it has become possible to optimise storage methods in terms of structure. Before digitisation, it was difficult to represent tree hierarchies or complex entity relationships, and the description methods were more homogeneous, mostly sorted or ordered according to certain simple primary keys. In modern databases, on the other hand, we have different data structures to target our different situations. For example, the relational database model has foreign keys, the graph database model emphasises relationships more than the data itself, the file database model has a great deal of flexibility that was not available before digitisation, etc. Well-tested and validated data structures can be easily used by anyone and are more adapted to the processing of big data.

c) Lower redundancy and fewer errors improve data consistency

The adoption of a modern database structure has the added benefit of lower redundancy compared to pre-digitisation systems. Pre-digital systems had no logical structural constraints and the same information could be stored repeatedly. After a particular piece of information has been updated, there may also be information that has been forgotten to be updated leading to inconsistencies in the data. In modern databases, however, inconsistency errors can be substantially reduced through constraints and redundancy through database design paradigms, guided by database design principles, which are made possible by digitised storage and enhanced logical relationships.

d) Higher hit-rate searches

Searching with a modern database structure has a higher hit rate than with a pre-digital database structure. Before digitisation, people searched the contents of databases mainly manually and sequentially, requiring a traversal of a category of books. In modern databases, the storage structure may be a b+ tree (Kieseberg et al., 2011) and the logical structure may also refer to precise targeting as in the document example designed by Chris Clifton and others (Clifton & Garcie-Molina, 1988). This can make information retrieval much more efficient and reduce search times.

e) High reusability for analysis and exploitation

Modern databases tend to be more reusable than pre-digital systems due to their well-organised data structure. When analysing different aspects of information, one can quickly obtain results from a modern database with relatively good relationships and data, whereas pre-digital systems often take a lot of time to analyse due to the relative weakness of the relationships.

4. Data of database management systems

According to different needs of management, database is mainly used to organise, store and manage data according to various data structures.

4.1 History of data

People use data in almost every aspect in their daily lives, and the origin of data is as ancient as the characters.

When it dates back to the ancient time, at around 19,000BC, people believed that the Ishango bone could be acted as a tally stick which was used to implement simple data collection, calculation and storage (Shatby, 2022, section 2). They draw scratches to represent data, and people can presume that they just did the manipulations by making use of different forms or length of lines. In this way, data, or perhaps just some simple numbers, were collected and stored directly on the bone by recording by hands, and could be checked by identifying distinct symbols.

Time moved forward to 1640s, when the word “data” first appeared in English, John Graunt, who is believed to be the earliest person to apply data analysis into real situations, did some researches on the death records of London, and even attempted to predict further information. After the book written by Graunt was published, the city began to report the information data about death weekly, and wished to do the preparations for sudden circumstances (Shatby, 2022, section 2). At that time, people mainly kept records of data in chronological order. When there was new data which needed to be kept, they first recorded the information and then made a paper record in chronological order, for example in the form of a report, which was either kept by the management department or used to be released to the public.

In 1880s, with the gradual expansion of the data range, there was more data emerged than they could cope with. At that time, Herman Hollerith developed a kind of machine which could make use of punch cards in order to read data electronically (Shatby, 2022, section 2). Since then, the administrations of data could record various data in a more efficient way by using this kind of machines, however, keeping track of the information still needed to be done manually.

When it came to 1900s, the crux of data management shifted onto how to collect and store data. At that time, the scientists came up with various original ideas and innovations, including magnetic data storage, cloud data storage and also relational database management system (Shatby, 2022, section 2), which is mainly used to manage data in the modern society.

With the development of Internet, data became more and more extensively accessible in recent years. The history of data is a quite long story, and is likely to evolve towards diversity. In recent years, when people think of the pre-digital data management system and the modern database management system, we can find that there are some similarities as well as differences of the data inside the two kinds of systems, and they can be divides into some sections as follows.

4.2 Elements of data management

a) Forms of data

The two systems have similar forms of data storage, the former one draw tables or lists manually and the latter one mainly used digital tables to perform. Although the forms resemble each other, the pre-digital system can have more formats while the database system use relational model which contains columns and tuples to present data.

b) Types of data

At the beginning of the development of data, people mostly utilised integers, because it seemed complex for them to note down the decimals or other kinds of data. On the contrary, in the modern society, people use integers in some special attributes, such as age, the count of individuals, etc. Besides that, there are more various types, including float, character (CHAR), text and so on, which can meet different usages of data.

c) Orders of data

For the pre-digital system, data must be inserted and stored in a specific ordered, such as ordered by ID; otherwise, when the users want to find a piece of specified record, the administrations do not have to search every page in order to find it. As for the database system, orders of the tuples are not important, because the users could select information by using SQL or other data query languages rapidly.

d) Data Integrity

Both of the two versions of systems have to guarantee the integrity, although maybe in different ways. For pre-digital systems, the management are supposed to make sure to update associated data and clean useless data promptly; additionally, they also have to always keep an eye to ensure the data that is planned to insert is not meaningless. As for the database system, the administrations can manage data integrity by identifying primary keys as well as foreign keys, making use of cascading actions, default values, etc.

e) Rules of updating and deleting data

The two systems are different from the rules of deleting tables. In particular, the database management system must delete the tuples that are not referred by other attributes from other tables first, or there will be an error. In addition, updating data that has foreign key constraints also has to be careful, otherwise the manipulation is likely to be interrupted by causing faults. From this aspect, there are no reminders for pre-digital system, then the administrations can only be aware of the abnormal situations when the next step cannot continue if they make faults before. Therefore, the database system offers a much better way for data integration and cleansing.

4.3 Performance of management about data

a) Scale

The scale of the data can be indicated by the domain of attributes' values (InfoVis, 2009, para. 1). For the previous systems, the data scale is very restricted because of the limitations of technology, so most of the time before database formally appeared, administrations mainly used discrete set, which was numeric range. Then with the development of techniques, the modern database systems can put nominal, ordinal, discrete, continuous and even binary data (InfoVis, 2009, para. 1) into use of our daily lives.

b) Extensibility

In fact, both of the two data management systems do not provide the service of data extensibility at the beginning, especially it is hard to require the manual work of data manipulation to offer data extension; however, database systems also lack provisions for the inclusion of extensions in either their query languages or structures (Batory, 1987, p. 25). Luckily, the Object-oriented database has a meaningful service of extension (Bertino, 1991, p. 37), which can be applied to data extensibility explicitly.

c) Speed

The advance of science and technology really facilitates faster implementation of management. At the beginning, there was much redundancy of data, administrators could only hire more employees to enhance the speed of processing. There are multiple fresh techniques in current years to reduce redundant data, which can greatly accelerate the speed of operations.

d) Efficiency

According to the elements mentioned in 4.2, the administrators have distinguished different types of data in DBMS, which can store and transform different kinds of data without type conversion. Besides, organising data in a reasonable order, form, etc. can also increase the processing efficiency.

e) Complexity

Although the data is organised in different ways, the complexity of data types and forms are similar in the two kinds of data management systems; however, there are less data redundancy as better integrity in DBMS, therefore the performance about complexity of DBMS is likely to be better.

5. Users and usage of data management systems

5.1 Users and usage before digitisation

Databases had a simple structure and a single user, with only end users before digitization. They were relatively expensive with a small number of users, and only able to be created by large national institutions. For example, databases were used in ancient Chinese academies to record information about books and their storage locations, pharmacies used databases to record prescriptions and herbs for various diseases, post offices used databases to record information about each movement of goods, pawnbrokers and banks used databases to record information about each transaction, etc.

Non-digital databases are mainly used for storing information, retrieving information, and disseminating and exchanging information. Ancient government departments, school libraries, hospitals, and commercial companies all have their own complex database systems, many of which are still in use today. These databases are typically smaller in size, with simpler relationships and practical applications.

5.2 Users and usage of modern DBMS

5.2.1 Users

Modern DBMS users include database administrators (DBA), application programmers and end users. For the most part, the database administrator is the only one who can interact directly with the database, while the programmer uses the API and the end user uses the front and back-end interfaces (Techopedia, 2022).

a) DBA

The DBA, as an IT person, is required to implement and configure the database and is also responsible for the effectiveness of its access and use. At the same time, they need to specify solutions for their maintenance, use, security and management (TechTarget, n.d.).

The rapid development of new technologies such as big data, cloud computing and virtualisation has placed a higher demand on the skills of DBAs. The skills are mainly technical skills and non-technical skill. Non-technical skills include excellent written and verbal communication skills, focus and close attention to detail, problem-solving, and the ability to interact with different teams, end users and executives in the organisation. Professional and technical skills include knowledge of computer systems, database management systems, etc. (Ikoli, 2014).

As new technologies emerge, the technical requirements for DBAs are increasing and the roles of DBA, system administrator and developer are blurring, so DBAs need to have a knowledge base in information science, computer science and practical experience using Hadoop systems, also they need to have strong organisational skills, logical and analytical skills, the ability to learn throughout life and a sense of confidentiality (Simmonds, 2013).

Database administrators with legitimate access are also the biggest potential threat to the database. While existing access control mechanisms and intrusion detection systems allow for some level of insider access, as well as a range of models to prevent information misuse by insiders in database systems through database applications, a sense of confidentiality among database managers is important to ensure data security (Chagarlamudi et al., 2009).

b) Application programmers.

As databases evolved, so did how database programmers could use them, and with the advent of relational databases in 1970, programmers could replace complex algorithms that would otherwise need to be written by specifying a predicate. The ability to use database management systems with greater ease brought about greater programmer productivity (Silberschatz et al., 1991).

c) End users of DBMS.

The end users of a modern database are usually people who can commit individual transactions that query and update the database under a transaction management mechanism. Examples include bank staff, corporate ticketing staff, shipping couriers, etc. (Geeksforgeeks, n.d.).

Modern users are much more efficient than in the past. As the experiment of benchmark shows, scanning data sequentially costs a lot of time and hashing as well as ISAM is likely to cause a decrease in performance (Ahn and Snodgrass, 1986, p. 103). Those methods are commonly used in pre-digital systems, which would result in a severe reduction in implementation efficiency. Things are quite different in the database systems, the administrators of the database can develop particular access as well as querying methods according to the unique characters of specific databases (Ahn and Snodgrass, 1986, p. 103), which can enhance efficiency significantly.

Also, modern users have strict privacy politics. The government, business or institution that collects user information are responsible and accountable for the proper use of the data and its security. Before the database was created, this could only be done by the power of long-term supervision manually, which was easy to cause problems. To date, one of the most famous W3C's Platform for Privacy Preference (P3P) enables websites to apply their privacy policies in a special manner which can be recognised by machines (Byun, 2006, p. 9), so that the users may evaluate the published privacy policies against their privacy preferences more easily. The usability of data is still on the way and needs to be improved.

5.2.2 The usage of DBMS

Digital databases have driven the development of information services and other applications in an increasingly wide range of applications and fields. Digital databases underpin the industrialised economies of the early 21st century, communication systems, logistics and transport, financial management, knowledge-based systems, accessibility of scientific literature and civil and defense applications, laying the foundations for fundamental disciplines in various fields such as computational science and biology. From its beginnings, in industrial research laboratories, it expanded to universities, commerce, government laboratories and research etc. (Silberschatz et al., 1991).

Databases have evolved from initially concentrating on data management in business applications such as automated banking, record keeping and reservation systems to being used in all computing environments to organise, create and maintain important collections of information. At the same time scientific databases, design databases and universal access to information are taking databases in a variety of new and important directions. The next generation of database applications and heterogeneous distributed databases will be the key areas that will have a significant impact in the future (ACM, 1991).

Databases provide products and technologies of great social and economic significance, improving the delivery of health care, advanced design and manufacturing systems, enhanced tools for scientists, increased productivity per capita through increased individual capabilities to access information and new military applications. Databases are added as a backbone to computer-aided design systems, for example, civil engineers envisage a facility engineering design system to manage all information on related projects. The US National Institutes of Health (NIH) and the Department of Energy (DOE) build databases to implement search techniques for matching data on individual medical problems with differences in genetic makeup. In the commercial sector, sellers can run ad hoc queries on historical databases to find buying patterns and make inventory decisions, and NASA scientists could use the vast amount of information collected from space to build databases for storing and finding relevant images (ACM, 1991).

Modern databases can be used in a much wider range of applications and fields than pre-digital databases and can bring benefits to a much wider range of users or industries. At the same time, thanks to technological developments, databases can bring benefits to a much wider range of fields than before digitisation.

6. Future development

Although database management systems are now far more advanced than they used to be before digitisation, the rapid growth of all industries in the world today has led to an explosion in the volume, diversity and speed of various data, which has promoted the advancement of new technologies for processing as well as analysing large-scale data (Ikoli, 2014). This exponential growth in data has led to an increase in the frequency and granularity of data collection while increasing the challenges for the data management industry (Davenport et al., 2012).

The technology of database has various kinds of applications; therefore, more and more attention has been drawn to its technological research. Although the technologies still have some limitations; however, according to the benefits and convenient that database has brought to our lives, it is expected to get further development.

7. Conclusion

The technology of database management is one of the most effective ways to manage data and handle large-scale information in modern society, which has brought the advancement of the economy as well as civilisation; however, in ancient times, mankind also had to process information, including figures, characters, dates, etc. Without the advanced development of the Internet and database service, although humans could still accomplish control of data, it was much less convenient not to do that in a systematic approach.

In this paper, the elements of data management have been divided into four main sections, including hardware as well as software, structure, data, and users. Based on this structure and current research, this paper focuses on the comparisons between the pre-digital data management system and the modern DBMS, also with envision of future developments, which has meaningful significance to the true understanding of the advantages of DBMS. Unluckily, the materials of pre-digital data management are limited; therefore, the description of that might seem to be less clear and convincing.

The technology of DBMS plays a really important role in the management of large-scale information in modern society, which is believed to get further development as well as applications in the future.

References

[1] ACM. (1991) *Database systems: achievements and opportunities*. Available at: https://dl.acm.org/doi/10.1145/125223.125272 (Accessed: November 14, 2022).

[2] Ahn, I. and Snodgrass, R. (1986) "Performance evaluation of a Temporal Database Management System," *ACM SIGMOD Record*, 15(2), pp. 96–107. Available at: https://doi.org/10.1145/16856.16864.

[3] Angles, R. and Gutierrez, C. (2008) "Survey of graph database models," *ACM Computing Surveys*, 40(1), pp. 1–39. Available at: https://doi.org/10.1145/1322432.1322433.

[4] Atikoglu, B. et al. (2012) "Workload Analysis of a large-scale key-value store," Proceedings of the *12th ACM SIGMETRICS/PERFORMANCE joint international conference on Measurement and Modeling of Computer Systems - SIGMETRICS '12* [Preprint]. Available at: https://doi.org/10.1145/2254756.2254766.

[5] Batory, D., (1987) "Principles of database management system extensibility," *Database eeri*. Available at: https://d1wqtxts1xzle7.cloudfront.net/83348786/87JUN-CD-with-cover-page-v2.pdf?Expires=1668785409&Signature=M9LPSodyD71BKAmvkk51EhOznam07pQhBuazLqmRyjYLREVpf93F3QtsQvK0YpZLPkHuCZDgwDB0rmA~snIbW0Oj7j~vzI-q0uYMHxeGB0ctC5Y3Su0BSbikC-dbkSWTZ4YyR~Eg5CIUphVSNugNuMhIXovgattbsIXeQeScvRoCrddBowGVn-IoU9GgiKfsJ7~OP9VVTw5uQm4jOH0hyThfRQsEHIKW0OnGx8fuEZ9hBcEWCChMO8-S8WU3AamIUa~~vZ1sBluB2KJ6kDoF7Em4nMLayd7n0fmrxJN-PxpulGv0qO4pDHG0OWG~Aqro6l~wz-5oz7y6RSL8TB1~6g__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA#page=42

[6] Bertino, E. and Martino, L., (1991) "Object-oriented database management systems: Concepts and issues," *Computer*, 24(4), pp. 33–47. Available at: https://doi.org/10.1109/2.76261.

[7] Byun, J.W. and Bertino, E., (2006) "Micro-views, or on how to protect privacy while enhancing data usability: concepts and challenges," *ACM SIGMOD Record*, 35(1), pp.9-13.

[8] Chagarlamudi, M., Panda, B. and Hu, Y. (2009). "Insider threat in database systems: Preventing malicious users' activities in databases," In *2009 Sixth International Conference on Information Technology: New Generations*, pp. 1616-1620. IEEE. Available at: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5070873

[9] Clifton, C. and Garcie-Molina, H. (1988) "The design of a document database," *Proceedings of the ACM conference on Document processing systems - DOCPROCS '88* [Preprint]. Available at: https://doi.org/10.1145/62506.62528.

[10] Codd, E.F. (1970) “A relational model of data for large shared data banks,” *Communications of the ACM*, 13(6), pp. 377–387. Available at: https://doi.org/10.1145/362384.362685.

[11] Crus, R.A., (1984) “Data recovery in IBM Database 2,” *IBM Systems Journal*, 23(2), pp.178-188.

[12] Davenport, T.H., Barth, P. and Bean, R. (2012). *How 'big data' is different*. Available at: https://www.hbs.edu/ris/Publication%20Files/SMR-How-Big-Data-Is-Different_782ad61f-8e5f-4b1e-b79f-83f33c903455.pdf

[13] Denning, D.E. and Denning, P.J., (1979) “Data security,” *ACM computing surveys (CSUR)*, 11(3), pp.227-249.

[14] EdPrice-MSFT (no date) Non-relational data and NoSQL - Azure Architecture Center, *Azure Architecture Center | Microsoft Learn*. Available at: https://learn.microsoft.com/en-us/azure/architecture/data-guide/big-data/non-relational-data (Accessed: November 13, 2022).

[15] Foote, K.D. (2017) *A brief history of data modeling, DATAVERSITY*. Available at: https://www.dataversity.net/brief-history-data-modeling/ (Accessed: November 17, 2022).

[16] Gaenger, S. (2020) *The 11 best data visualization tools for SQL, Chartio*. Available at: https://chartio.com/learn/business-intelligence/the-11-best-data-visualization-tools-for-sql/ (Accessed: November 17, 2022).

[17] Geeksforgeeks. (no date) *Categories of End Users in DBMS*. Available at: https://www.geeksforgeeks.org/categories-of-end-users-in-dbms/ (Accessed: November 13, 2022).

[18] GeeksforGeeks (2022) *History of DBMS*. Available at: https://www.geeksforgeeks.org/history-of-dbms/ (Accessed: November 17, 2022).

[19] Haiyan, Y. et al. (2010) “Performance evaluation of post-relational database in hospital information systems,” *2010 Second International Workshop on Education Technology and Computer Science* [Preprint]. Available at: https://doi.org/10.1109/etcs.2010.134.

[20] Heidari, A., Jafari Navimipour, N. and Unal, M. (2022) “The history of computing in Iran (persia)—since the achaemenid empire,” *Technologies*, 10(4), p. 94. Available at: https://doi.org/10.3390/technologies10040094.

[21] Ikoli, G.I. (2014) *The Impact of Big Data on Database Administrators from a Skills Perspective*. Available at: https://www.scss.tcd.ie/publications/theses/diss/2014/TCD-SCSS-DISSERTATION-2014-060.pdf

[22] InfoVis: Wiki, (2009) “Information for "Data Scale"”. Available at: https://infovis-wiki.net/wiki/Data_Scale (Accessed: November 13, 2022).

[23] Keister, O. R. (1964). *The Incan Quipu. The Accounting Review*, 39(2), 414–416. Available at: http://www.jstor.org/stable/243533

[24] Kieseberg, P. et al. (2011) “Using the structure of b+-trees for enhancing logging mechanisms of databases,” Proceedings of the *13th International Conference on Information Integration and Web-based Applications and Services - iiWAS '11* [Preprint]. Available at: https://doi.org/10.1145/2095536.2095588.

[25] Lim, H. et al. (2011) “Silt,” Proceedings of the *Twenty-Third ACM Symposium on Operating Systems Principles - SOSP '11* [Preprint]. Available at: https://doi.org/10.1145/2043556.2043558.

[26] McCarthy, G. (2022) *The past, present and the future of Data Management: GSM barcoding, Barcoding*. Available at: https://www.barcoding.co.uk/the-past-present-and-the-future-of-data-management/ (Accessed: November 13, 2022).

[27] Miyagawa, S., Lesure, C. and Nóbrega, V.A. (2018) “Cross-modality information transfer: A hypothesis about the relationship among prehistoric cave paintings, symbolic thinking, and the emergence of language,” *Frontiers in Psychology*, 9. Available at: https://doi.org/10.3389/fpsyg.2018.00115.

[28] MongoDB (no date) *Document database - nosql, MongoDB*. Available at: https://www.mongodb.com/document-databases (Accessed: November 13, 2022).

[29] MongoDB (no date) *Understanding the network database model, MariaDB Knowledge Base*. Available at: https://mariadb.com/kb/en/understanding-the-network-database-model/ (Accessed: November 13, 2022).

[30] Murphy, N. (2016) *Idug Content Committee, Remembrance of IBM DB2 Events of the 1980's which Led to the Creation of IDUG*. Available at: https://www.idug.org/blogs/nate-murphy1/2020/07/27/remembrance-of-ibm-db2-events-of-the-1980s-which-led-to-the-creation-of-idug (Accessed: November 17, 2022).

[31] Nin, J. et al. (2007) “On the use of semantic blocking techniques for data cleansing and integration,” *11th International Database Engineering and Applications Symposium* (IDEAS 2007) [Preprint]. Available at: https://doi.org/10.1109/ideas.2007.4318104.

[32] Nishtha, J. et al. (2012) “A Survey and Comparison of Relational and Non-Relational Database”, *International Journal of Engineering Research & Technology (IJERT)*, 1(6). A

vailable at: https://www.academia.edu/44482078/IJERT_A_Survey_and_Comparison_of_Relational_and_Non_Relational_Database.

[33] Oracle (no date) *What is a database?*. Available at: https://www.oracle.com/database/what-is-database/ (Accessed: November 17, 2022).

[34] Paredaens, J., Peelman, P. and Tanca, L. (1995) “G-log: A graph-based query language,” *IEEE Transactions on Knowledge and Data Engineering*, 7(3), pp. 436–453. Available at: https://doi.org/10.1109/69.390249.

[35] Phillips, H. (2010) *Great Library of Alexandria. Library Philosophy and Practice, Aug*. Available at: https://nbn-resolving.org/urn:nbn:de:0168-ssoar-190807

[36] Quickbase (no date*) A timeline of database history*. Available at: https://www.quickbase.com/articles/timeline-of-database-history (Accessed: November 17, 2022).

[37] Raima (no date) *Network database, relational DB, and graph DB compared, Raima*. Available at: https://raima.com/network-database-relational-db-and-graph-db-compared/ (Accessed: November 13, 2022).

[38] Shatby, S.E., (2022) *The History of Data: From Ancient Times to Modern Day*. Available at: https://365datascience.com/trending/history-of-data/ (Accessed: November 10, 2022).

[39] Silberschatz, A., Stonebraker, M. and Ullman, J. eds. (1991). “Database systems: Achievements and opportunities,” *Communications of the ACM*, 34(10), pp.110-120.

[40] Simmonds, T. (2013) *Teaching Database Administration in the World of Big Data and Small Budgets*. Available at: https://gala.gre.ac.uk/id/eprint/19913/7/19913%20SIMMONDS_Teaching_Database_Administration_2013.pdf

[41] Techopedia. (2022) *Database Management System* (DBMS). Available at: https://www.techopedia.com/definition/24361/database-management-systems-dbms (Accessed: November 17, 2022).

[42] TechTarget. (no date) *database administrator* (DBA). Available at: https://www.techtarget.com/searchdatamanagement/definition/database-administrator (Accessed: November 14, 2022).

[43] Wright, Michael T. (2006). "The Antikythera Mechanism and the early history of the moon phase display". *Antiquarian Horology*. 29 (3): 319–29. Available at: http://fsoso.online.fr/antikythera/DOCS/The%20Antikythera%20Mechanism%20and%20the%20Early%20History%20of%20the%20Moon%20.pdf