Prehistoric data storage and organization methods, Similarities, and differences.

Data was always an integral part of human life. There were multiple ways employed to store information throughout history. With the current advanced methods of data management, we have access to all of information in the world history under centralized servers and can have instance access to such vast information with fewer clicks on the internet. With such improved data management systems, we have progressed on the aspect of keeping the data safe. Everyday limitless amount of sensitive data is transmitted via digital channels and keeping that data secure is a key factor. Thus, there is a significant demand for cybersecurity professionals across almost all industries to stop data breaches and the evil deeds of hackers and cybercriminals. We can currently experience sophisticated methods for keeping that data safe, like backing up files to vast server farms in the cloud.

Information storage has not always been simple, of course. Even before the development of the computer, data was still being produced, albeit much more slowly. Additionally, that data needed to be kept somewhere and in some way. Here is a summary of how information was saved in the 18th century before the advent of modern database management systems.

Six generation of data management

To get to the current state of modern data storage and management systems, data management underwent six significant periods of progression. Data was initially managed by **hand processing**. The use of **punch cards** and **electromechanical devices** was part of the subsequent generation. In the third stage, sequential files were processed in batches using computers and pre-stored programmes, and data was stored on **magnetic tape**. The fourth phase introduced the concept of a **database structure** and **online data navigation**. The fifth level introduced client-server processing that was dispersed, as well as automatic access to **relational databases**. Systems of the sixth generation, which can store more intricate data like as papers, pictures, speech, and video data, are now in the planning stages [5].

Let us go back to the basics. A simple pen and a sheet of paper was the sole way to store and transmit information for millennia. Everything had to be written out letter by letter prior to Johannes Gutenberg's creation of the printing machine in 1440. Such a system had its own fair share of advantages and disadvantages in relationship to modern database management system.

The following states the key factors needed to be analysed in terms of comparison between modern database management system against the prehistoric data storage and analysis methods.

Data Indexing:

The performance of a database management is determined in how fast it can retrieve data. Indexing is a technique followed by modern DBMS to improve performance, by reducing the number of times the data is accessed when query is performed. It helps users quickly locate and retrieve data and information stored in database tables. The search key is the database's first column, and it contains a duplicate or copy of the table's **candidate key** or **primary key** [12]. The primary key values are saved in sorted order so that the related data can be quickly accessible.

Similarly, in ancient records keeping, there was a vast amount of written material that necessitate the invention of some kind of retrieval system. The first use of the word "**index**" referred to the little slip attached to the scroll which contained the title and author of the work, but that was about it [7]. The next system used **table of content** as an alternate to what was called an "index". This can be correlated to the primary key of a table/object used in both relational and non-relational databases.

Data Retention Policy:

Modern database management systems come with **configuration** and **policies** to set when it comes to data retention. These define when to **remove** or **archive** data in the database. This helps in keeping the database clean and clear from outdated and obsolete information, as this helps in reducing cluttering of storage units (**servers).**

Historical data were stored in multiple forms such as rocks, paper, punch cards, or magnetic tapes. Some of the data in these last till the lifetime of the material. But some of these storage elements cannot be trusted with, in terms of longevity of data storage. For example, **Paper**, as everyone knows, has its own slew of problems as a data storage option. It is **expensive** to make, environmentally damaging and **harmful**, and **impossible to scale**. 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.

Data Validation and Authenticity:

Numerous safeguards are employed to protect database management systems from intentional cyberattacks and unauthorised use under the umbrella of database security. Data management systems themselves, as well as every application that uses them, are also protected from misuse, deterioration, and intrusion by **unauthorized** individuals. Modern DBMS includes provisions for restricting access to the data base by unauthorized users. One such method is via **granting Access control**, which is done by creating user accounts and controlled login process by the DBMS.

Whereas 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 are not at risk of being overlooked, and since they may be encrypted, the data would still be secure even if they were duplicated or stolen by an unauthorised person. Measures were taken to secure and authenticate paper record using methods such as **watermarks** or **invisible fibres** in paper.

User-friendly interface and experience:

It is an evident distinction that the prehistoric data storage and analysis involved computational machines and equipment’s that were not very user friendly .

Greek scientists created the **Antikythera Mechanism**, the earliest mechanical computer yet found. Its "CPU" is composed of thirty interlocking bronze gears, and it is believed to have been created to chronicle the cycle of Olympic Games and for astrological purposes. Its design indicates that it is most likely an advancement of an earlier computational device. The mechanism is remarkable for the degree of component miniaturisation and intricacy, which is comparable to that of astronomical clocks from the fourteenth century. Although mechanism specialist Michael Wright has indicated that the Greeks of this era could have implemented a system with many more gears, it has at least thirty gears [10]. Such complication means less user friendliness with respect to usage of these machineries, for data storage and computation.

Modern database systems allow for various interfaces that eases the ways of using the system to store and retrieve information. Some of the user-friendly interfaces provided by the database management system (DBMS) are - **Menu-based, Form-based, Graphical user interface** and much more. A good user experience for database systems aids in quick and simple information discovery. DBMS should have a **contemporary**, **user-friendly layout** and provide a **tailored** experience. Users **productivity** will increase as a result, and data search time will be cut down.

Automated data analytics:

Automated data **analytics** or **data visualisation** is a part of most data management system, as they help in better understanding of the stored information. By integrating data management system with available analytics platform (Power Bi, tableau, looker, etc.) we can speed up data analytics and visualisation [11]. This will prevent the need to switch between many platforms because data can be accessed from one central location. An analytics module will assist in making sense of the data, regardless of whether data management system is utilized at source or destination. These tools also allow to predict the trends in data and allow for making righteous decisions making and with modern data analytics technology, you can continuously collect and analyse new data to update the understanding as conditions change.

One of the earliest examples of prehistoric data storage is the **Ishango Bone**, which was found in 1960 Democratic Republic of Congo. To keep track of trade or supply operations, Palaeolithic tribespeople carved grooves into bones or sticks. To perform simple arithmetic calculation, they would compare sticks and notches, allowing them to forecast things like how long their food supplies would last [8].
In case of File-based management system the data format and storage details are provided by the file system. Since the users create and manage these data on their own, they are free to follow any format of creation. This inconsistency made it difficult in finding trends among the data.

Around the year 1880, The US Census Bureau had a problem it will take 8 years to process all the 1880 census data, and it is anticipated that the 1890 census data would take over 10 years, meaning it won't even be ready until the 1900 census, when it will be out of date. Such longer period of analysis will render impractical when it comes to data analysis.

The **Hollerith Tabulating Machine** was created in 1881 by Herman Hollerith, a young engineer hired by the bureau. He accomplishes his goal of being recognised in history as the founder of modern automated computation by turning ten years' worth of work into three months using punch cards. Herman Hollerith's tabulator consisted of electrically operated components that captured and processed census data by "reading" holes on paper punch cards. Because of his outstanding outcomes, Herman Hollerith was given the job of processing and tabulating the 1890 census data. The Census Bureau would employ modified versions of his method until the 1950s, when computers took over. He founds a business that will eventually be known as IBM [9].

Different Means of storing and analysing data before modern database management systems

Paper

Since it would be centuries before magnetic tape would revolutionize data storage, a simple pen and a piece of paper was the only option to store and relay information. Before the invention of the printing press by Johannes Gutenberg in 1440, everything had to be written out letter by letter. Books were expensive to reproduce since only monks and religious figures had the time (and patience) to rewrite books by hand.

Even though paper is still commonly employed as a data storage medium today, a digital scan of the physical paper is frequently used in addition to it as a backup in case.

Libraries – [Needs to be proofread as it is directly taken from online source]

After the printing press became more widespread, newspaper and book production soared. Daily newspapers became popular around the start of the 18th century—with an estimated 337,000 different book titles printed throughout the 1700s.

Since the Age of Enlightenment brought a new respect and interest to knowledge and education for all classes, cultural institutions like museums and libraries started being built. The British Museum, the first national public museum, was created in 1753 with a collection of over 40,000 books and manuscripts. The United States Library of Congress was created in 1800 and currently has over 170 million books. Public libraries such as these helped bridge the educational gap between the classes during this period while also ensuring books were safe and available for future generations.

Punch Cards

The punch card way of storing and reading data was fast and agile. The programs were read from a stack of cards with holes punched in them, marking each character in a program. Since a punch card can only hold a limited amount of data, a stack of them was required to store a program information.

Punch cards were used in the textile industry, to store data and control mechanized looms. People could represent a sequence of instructions that would translate to "on" or "off" orders that the looms would comprehend by determining whether a hole had been punched in the punch card. This straightforward approach established the fundamental building blocks of the binary language that modern computers use to communicate.

Less information could fit on a punch card back then than they can now. It could hold about eighty distinct characters. Punch cards are still used in voting machines and high school and college standardised tests even though they quickly lost favour as magnetic storage technology advanced.

Magnetic tapes

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

Data Modeling Timeline

1960

Hierarchical Data Model.

The Network Data Model

The first occurrence of computerized form of storing data was evolved in the year 1960, with the increase in production of cost-effective computers as an option for private organization. Before 1960s limited data storage options were available. In the 1960s, a number of theoretical data models were put forth [1]. The network data model and the hierarchical data model are the first two. The third theoretical model, the relational model, was proposed by Edgar F. Codd in the late 1960s. [2]

1970

Relational Database systems

A paper on relational database model was published by Edgar F. Codd, which laid the standard principle for database systems. In this model the database schema and logical organization is disconnected from the physical world.

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.

1980

Structured Query Language

This is the year in which the Structured query language that we use today, became the standard language for database. The earlier network and hierarchical data model saw a decline, leading to raise of relational database. Database management systems were built for computer by IBM (DB2) [3] and gained popularity with the rapid rise of computer sales. Object oriented database management system emerged towards mid of 1980, supporting object-oriented view of data.
SQL became ANSI and OSI standards in 1986 and 1987[4]

1990

Client-server database

Object oriented database

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

Data Storage Model before 1970

File-based System: Before the advent of the database system, file-based system was relied upon to store and retrieve information. This However, was a less elegant and costly approach. One way of storing information was to keep them in form of permanent files. Each of the application programs were written as separate files as per the requirements by the users and new ones were added on demand.

This system led to couple of disadvantages on comparing to the current database management system. Few notable differences are:

- **Data Inconsistency**: The data format and storage details are provided by the file system. Since the users create and manage these data on their own, they are free to follow any format of creation. This leads to data inconsistency. Whereas, modern database management system (DBMS) provides an abstract representation of data, leading to consistent data format across all the system. This helps unity in creating and reading information irrespective of the system being used.

- **Data Recovery**: The data in a file-based system does not allow for efficient data storage mechanism, leading to failures to recover data in case of system crash, whereas the current DBMS allows data recovery at its core as an essential feature.

- **Data Redundancy**: Substantial number of files were created by different programmers within an organization. With larger the number of files, and no means of validating the data duplication, can lead to data redundancy. In this event, there is a possibility of memory wastage to take place resulting in higher storage costs.

Hierarchical Data model: The data in this model was set up in a way that resembled a tree, with children represented by the branches and parents by the top of the tree. Because it makes data overview easier to understand, this kind of modelling was able to manage vast amounts of data. It enhanced data sharing throughout the network and communicated their relationship. The navigation was challenging because of the style of model's intricacy, structural limitations, and lack of data independence. Only a hierarchical data structure can be used to organize data without sacrificing its integrity [6]. **The IBM Informational Management System (IMS)** [3]**,** a database management system, released in the year 1966 was based out of Hierarchical data model. IMS operates on a hierarchical model, which means that it views data as trees rather than tables that may be joined together using JOIN operations. Each type of record you store can have other record types as children; these child record types indicate additional information that, given a record of the parent type, you might be interested in.

Network Data Model – It contained a collection of data items called ‘Records’, and each record related to an entity in the area it is being modelled to. Each data had a name and a value property, which can be stored in the database as undivided object. This allowed the network database management system to store, retrieve, delete, or modify within the database.

Codasyl's Data Base Management Methodology - The **CODASYL** method was an excessively complex technique that demanded extensive training. It was dependent on a "manual" navigation technique employing a network of linked data sets [4]. One of three methods could be used to search for records:

- Using the primary key (also known as the CALC key)
- Moving relationships from one record to another

- Scanning all records in sequential order

The CODASYL technique eventually fell out of favor as more user-friendly, less complicated systems entered the market.

Relational Model: The rigidity and the duplication issue in the aforementioned models led E F Codd to propose relational model. The relational model is an abstract model based on relational algebra and is top-down approach in that the data storage scheme.

REFERENCE

[1] A Timeline of Database History & Database Management, https://www.quickbase.com/articles/timeline-of-database-history, accessed 09-11-2022

[2] A Brief History of Data Modeling, https://www.dataversity.net/brief-history-data-modeling/ , accessed 09-11-2022

[3] Remembrance of IBM DB2 Events of the 1980's which Led to the Creation of IDUG, 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 10-11-2022

[4] A Brief History of Database Management, https://www.dataversity.net/brief-history-database-management/ , accessed 10-11-2022

[5] J. Gray, "Evolution of data management", Oct. 1996.
Available: https://ieeexplore.ieee.org/document/539719

[6] UKEssays. November 2018. The Hierarchical Model. [online]. Available from: https://www.ukessays.com/essays/information-technology/hierarchical-data-model.php?vref=1 [Accessed 15 November 2022].

[7] History of Information Retrieval, https://www.asindexing.org/about-indexing/history-of-information-retrieval/, accessed 16-11-2022

[8] Tally stick, https://en.wikipedia.org/wiki/Tally_stick , accessed on 13-11-2022

[9] The Hollerith Machine
https://www.census.gov/history/www/innovations/technology/the_hollerith_tabulator.html

[10] Antikythera mechanism - https://en.wikipedia.org/wiki/Antikythera_mechanism#Operation

[11] Data Visualization Tools for SQL, https://chartio.com/learn/business-intelligence/the-11-best-data-visualization-tools-for-sql/, accessed 16-11-2022

[12] Indexing in database management, https://byjus.com/gate/indexing-in-dbms-notes/ , accessed on 15-11-2022

Additonal References

A brief history of software and computing, https://www.quickbase.com/resources/articles/a-brief-history-of-software-and-computing/ , accessed 09-11-2022

A History of Data Collection, Storage, and Analysis, https://gutcheckit.com/blog/a-history-of-data/ , accessed 09-11-2022

Data Architectures' Evolution and Protection, https://ieeexplore.ieee.org/document/9872597 , accessed 09-11-2022