**III Data**

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

3.1 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. It was his effort made it possible to process data much more conveniently as well as quickly (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.

3.2 Elements of management

·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 use 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.

·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.

·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.

·Data Integrity

Both of the two versions of systems have to guarantee the integrity, although maybe in different ways. For pre-digital sysyem, 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.

·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 offer a much better way for data integration and cleansing.

3.3 Performance of management

·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 used of out daily lives.

·Speed

The advance of science and technology really facilitates faster implementation of management. At the beginning, people could only hire more employees to enhance the speed of processing, even in this way, the speed of executing was still a great problem. There are multiple fresh techniques in current years which can greatly accelerate the speed of operations. For example, the database systems can make use of B- tree to speed up the selecting operations, and complicated parallel algorithms can also boost the querying speed of data (Pavlo et al., 2009, p. 177), etc.

·Efficiency

As the experiment of benchmark shows, scanning data sequentially costs a lot of time, and hashing as well as ISAM is likely to cause the decrease of 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 quering methods according to unique characters of specific databases (Ahn and Snodgrass, 1986, p. 103), which can enhance efficiency significantly.

·Data Security

The security and privacy of data has always been a serious problem to data management. For the past times, people were not adware of the importance of protecting data, so the abuse of data might be serious, however, with the limitations of technology, people also could not access to the data management system very easily. These days, with the quick spread of computers and the growing expertise of users, the data of organisations and individuals are more susceptible to be misused by computer specialists. The complexity of assaults on computer systems will rise as the benefits of development (Denning and Denning, 1979, p. 228). Under these kind of circumstances, the data security also has a long way to go. In fact, there is no perfect solutions to this, however, a positive mechanism can reduce the risk of disclosure of data.

·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 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, p. 179).

·Data Usability

The government, business or institution that collect 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 in an easier way. The usability of data is still on the way and needs to be improved.

·Data 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.

**References:**

[1] Ahn, I. and Snodgrass, R., 1986. Performance evaluation of a temporal database management system. *ACM SIGMOD Record,* 15(2), pp.96-107.

[2] Batory, D., 1987. Principles of database management system extensibility. *Database eeri*.

[3] Bertino, E. and Martino, L., 1991. Object-oriented database management systems: concepts and issues. Computer, 24(4), pp.33-47.

[4] 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.

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

Kent, W., 1979. Limitations of record-based information models. *ACM Transactions on Database Systems (TODS)*, 4(1), pp.107-131.

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

[7] InfoVis: Wiki, 2009. *Information for "Data Scale".* [Online]. [Accessed 13 November 2022]. Available from: https://infovis-wiki.net/wiki/Data\_Scale

[8] Pavlo, A., Paulson, E., Rasin, A., Abadi, D.J., DeWitt, D.J., Madden, S. and Stonebraker, M., 2009, June. A comparison of approaches to large-scale data analysis. *In Proceedings of the 2009 ACM SIGMOD International Conference on Management of data* (pp. 165-178).

[9] Shatby, S.E., 2022. *The History of Data: From Ancient Times to Modern Day.* [Online]. [Accessed 12 November 2022]. Available from: https://365datascience.com/trending/history-of-data/