American Journal of Computer Science and Technology
2023; 6(1): 20-24
http://www.sciencepublishinggroup.com/j/ajcst
doi: 10.11648/j.ajcst.20230601.13
ISSN: 2640-0111 (Print); ISSN: 2640-012X (Online)

Research on Principles and Problems of Computer Software Database Design

Zhang Daoping, Tuyatsetseg Badarch*

Department of Information Technology, School of Information Technology and Design, Mongolian National University, Ulaanbaatar, Mongolia

Email address:

ba.tuyatsetseg@mnun.edu.mn (Tuyatsetseg Badarch)

*Corresponding author

To cite this article:

Zhang Daoping, Tuyatsetseg Badarch. Research on Principles and Problems of Computer Software Database Design. *American Journal of Computer Science and Technology*. Special Issue: *Advances in Computer Science and Future Technology*. Vol. 6, No. 1, 2023, pp. 20-24. doi: 10.11648/j.ajcst.20230601.13

Received: January 12, 2023; **Accepted**: February 20, 2023; **Published**: March 4, 2023

Abstract: Database design is an unforgettable part of this evolution, it is the core of systems and is the foundation for creating a safe and reliable program. The design quality is directly correlated to the user's experience with the system. It has a huge effect on the smoothness, stability, and security of the system's operation. Problems also arise when the database is neglected. If the design lacks quality due to the developer's disregard for following design principles these problems arise, thus developers need to comprehensively analyze, test, and summarize their design. This paper aims at software design concepts and their importance from a researcher's point of view. We will discuss how someone could improve the transmission efficiency of data resources and the importance of updating the software database. Our study shows there is a difference between databases of different applications. For instance, the database of a supermarket chain must be fundamentally different from that of an information technology company. An insurance industry database is definitely different from a manufacturing enterprise database. From the micro perspective, although the same industry database is being used, there are differentiated designs and services. For example, a supermarket chain, international chain, and domestic supermarket chain databases are completely different. Another thing we want to highlight is about the caution that is advised in any industry, the systems design and development industry has to be extra careful. If a single aspect of a program database is organized wrong, the whole system could crash. It causes huge losses for the enterprises as the company is dependent on it. In addition, accuracy is crucial for maintaining the integrity of data. For example, if inaccurate data can lead to false conclusions and wasted time and resources, the users must enter data correctly and consistently to ensure accuracy. Data should also be verified regularly to identify any errors that may have been introduced. The study shows we lack vitality in the design phase of software systems. Regarding the literature, we evaluated that future work needed to be used as an architecture with improved hardware capability, and evaluate the architecture's adaptability in more scenarios. The obtained results from our study encourage and show the potential for applying a software architecture for data-intensive real-time applications in the future. The paper concludes even though the usage of real-time based database technologies increased in recent years, there is a lack of standardization and modularized architectures.

Keywords: Computer Software, Database Design, Principles of Design, Problems

1. Introduction

Database processing was initially used in major corporations and large organizations for large transaction processing systems. As a result of the database programming technology that promotes creation and development of computer database, currently, databases are being widely used today for Internet and intranet applications. In the 1940s and 1950s, data storage mainly depended on manual input, while computers were only used for scientific calculations, after which the data would be released. In the late 1950s and 1960s, computers were not only used for scientific calculation, but also for data management. Since the 1960s, it

transformed from the manual management stage, the file system stage, and finally to the database system stage [1]. There was an urgent need to solve the problems of data storage, data retrieval, and information maintenance. Thusly database technology developed rapidly.

A promising solution to manage the increasing amounts of data and complex computations is to use Real-Time DataBase Management Systems (RTDBMS) for structured data management [2]. The term "Database" was first used by the Systems Development Corporation in the 1960s to develop data for the U.S. naval base. Regarding the study, in recent years, the RDBMS is the dominant database type for enterprise computing. Especially, the SQL based RDBMS still make a huge percent of databases in deployment. In fact, this continued popularity of the SQL language has resulted in big data processing data bases including SQL-on-Hadoop and Apache Hive, just to identify a few [2].

The General Electric company designed and developed the IDS (IntegrateData Store) system, and successively developed the network database management system and hierarchical database management system, referred to as DBMS; In 1970, E. F. Codd, a researcher from IBM, put forward the relational model, ushering in the great development period of relational databases.

Regarding the historical view, commercial relational databases led by Oracle Database, Microsoft's SQLServer, and IBM's DB2 began to occupy the sales market of the enterprise-level database. Meanwhile, database technology was introduced in China. The first National Database Academic Conference held in November 1977 set up a communication platform for database researchers and database development and application enterprises, which marked a solid step in the development and application of Chinese database technology [3]. In addition to the intensive development for database system, in the mid-1970s, Britain, France, West Germany, Sweden, and Japan successively established and developed their own national databases. Then, in the late 1980s, online database retrieval began to rise in economically developed countries and the emergence of the CD-ROM databases. Since the 1990s, with the rapid development of computer technology and the Internet, the network database system accessed through the Internet began to appear. After entering the 21st century, the network database system has become the mainstream way of database retrieval.

With the wide application of the embedded system, embedded database technology with easy to expand, small, easy to maintain, can be shared, has low redundancy, and independence of high advantages in our life. In order to solve the problem of low concurrency performance of embedded database SQLite, a concurrency control design based on Multi-Version Concurrency Control (MVCC) was proposed. [4]. The work brought the tremendous success into the operations such as add, delete, check and change and the index structure to modify the database work under MVCC. We can see the stages of database technology in Figure 1.

Database design has main stages that help to create, to implement, and to maintain a data management systems. The primary goal of designing a database is to produce physical and logical models of designs for the proposed database system. In general, the whole development process of database technology can be divided into three steps, the first two steps are manual management and file system, and the third stage refers to database system (Figure 1). In less than half a century, database technology has formed a solid theoretical foundation from its birth to the present. People put forward a variety of database models, such as the mesh model, hierarchical model, object-oriented model, relational model and semi-structured model, but also put forward a lot of new database technologies, such as XML data management, Web data integration, data flow management and data mining. Network technology and database technology have been widely used in various industries, for example, Compendium, the German ICT thermochemical database for medicine [6], chemical and chemical industries [7], Energy and metal Materials [8].

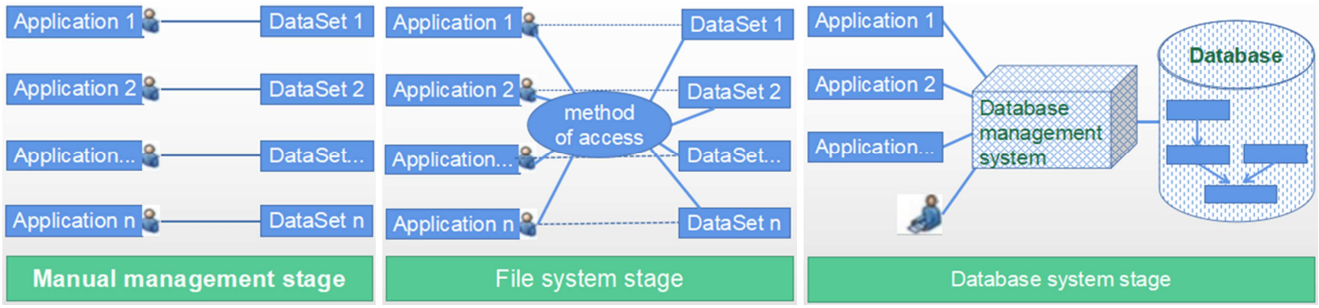

Figure 1. *The three stages of the database.*

2. Research of Database Management Systems

Data management services have become increasingly prominent now that we live in an era of big data. Data creates value, thus it has become important not only for all types of enterprises, not just the big ones. Even government institutions need to carry out accurate and effective management of their own data. However, with data management deepening, the form itself is facing a transformation.

Database technology mainly includes two important technologies: database and database management systems. As the foundation of the Internet, it is the guarantee for the

realization of various complex systems. The two aforementioned technologies are responsible for data collection and classification, organization and coding, storage and retrieval, and maintenance and dissemination. In plain terms, a database is a warehouse of data storage. The data is located in the computer storage equipment alike hard disk drives, servers, etc. Data inside has characteristics of permanent storage, and organization, and can be shared easily. The database management system is located between the data operator and the operating platform. It is mainly responsible for data management.

It has become a trend to manage one's own big data more economically and sustainably [5]. With the rapid development and prosperity of the Internet, social and economic, for example, into economics and Politics, Education and Culture, social and livelihood [9], and the Energetic Materials for military science, such as the NATO Center for Information Analysis of Munitions Security (EMC) database [10]. Network technology and database are in the core position to maintain the normal and stable operation of various systems. Nowadays, the Internet economy is rapidly penetrating into People's daily life, increasing people's dependence on computers and information networks. With the increasing maturity of machine learning, mass data processing, virtual reality, and artificial intelligence, computers, and databases will become an indispensable part of People's Daily production and life [11]. In the implementation of various Internet systems, the core is how to process data, namely how to collect and store data [12].

Once the data is stored and organized, the database management system takes control. It is structured to be high sharing, have low redundancy, have high independence, and have rapid expansion. Traditionally, databases are classified into relational databases and non-relational databases. Building a database on a relational model and managing the data organization allows the database to be relational. Examples include MySQL, Oracle database, DB2, etc. Those databases that have unstructured queries like NoSQL is an examples of non-relational databases.

Typically, MongoDB and HBase are the representatives of non-relational databases.

As the internet evolves, the number of users increases gradually. Complex requirements need to be met, which forces data management systems to evolve too. Today, embedded databases are becoming more popular. As long as there are embedded devices, embedded databases will exist. This evolution is thanks to the era of mobile computing, intelligent mobile terminals, big data, AI, 4G, and 5G systems. Looking at the future, embedded databases are developing towards decentralization and miniaturization. At present, the development of computer application greatly promotes the database technology and system [13].

3. Design of Database Principles

The development of computer hardware platforms is still practicing Moore's Law, and the application of databases is expanding rapidly to depth and breadth. This has the possibility of becoming a problem now that user numbers are rising rapidly. There are four core principles of good database design, namely, accuracy, accessibility, minimizing redundancy, and meeting expectations. For these characteristics, we are describing the main principles for them.

3.1. Understand User Requirements in Depth

The most important design principle of a database is based on the customer's needs, that is, what the customer wants to do with the database. To keep the customer requirements, database administrators should consult with users to understand their needs and expectations. Then, the database can be designed to make it easy for users to find and retrieve the data they need. Database administrators can improve accessibility by creating clear and concise indexes and documenting the database structure [14-16].

This needs to be from the macro and micro aspects. The macro is the industry category that the database is serving. Here are some principle database rules that the service category involves. For example, service class, production class, manufacturing class, etc.

3.2. Database Design Needs to Be Made with Caution

Although caution is advised in any industry, the systems design and development industry have to be extra careful. If a single aspect of a program is written wrong, the whole system could crash. It causes huge losses for the enterprises as the company is dependent on it. In addition, accuracy is crucial for maintaining the integrity of data. For example, if inaccurate data can lead to false conclusions and wasted time and resources, the users must enter data correctly and consistently to ensure accuracy. Data should also be verified regularly to identify any errors that may have been introduced.

To avoid a catastrophe like the one mentioned above, it is extremely important that the design is detailed and normative. This allows for easy information extraction thus the system becomes more precise.

Another important principle is to minimize redundancy that is critical for reducing storage costs and eliminating errors. Duplicate copies of data waste space and can introduce inaccuracies if not correctly synced.

Data base management system often helps to ensure that only the necessary data is retrieved, further speeding up the search process.

3.3. Database Expansion Methods

Database power and flexibility is more important. In table 1, we present the expansion methods and their characteristics to determine the methods, respectively. From the table, it is a clear to understand we need to choose the composite index based method. This method usually.

Table 1. *Expand database by seven methods.*

Method	Example	Unreasonable use	Fair use
Creating an Index		Full table scan	Scan by index
Composite Index	select * from users where area='Ulan Bator' and age=22;	(area, age, salary), (area, age), (area)	(area, age, salary)
Index will not contain NULL columns	NULL values for the composite index.	Any column containing NULL values in a composite index is invalid for the composite index.	
Using Short Indexes	Slow, Waste of disk space and I/O operations	Slow, Waste of disk space and I/O operations	Fast, Saves disk space and I/O operations [14]
Statement Operations	If it is necessary like "aaa%"	Discouraged	If it is necessary like "aaa%"
Perform operations on columns	select * from users where YEAR (adddate)<2022	The operation will be performed on each row, which will result in a full table scan for index failure.	select * from users where adddate<'2022-01-01';
The NOT IN and <> operations are not used	NOT IN id<>5	Full table scan	NOT EXISTS id>5 or id<5

4. Common Problems in the Design of Software Database

This section describes the common problems as well as suggests some technologies to solve the problems. Table 1 shows how the methods of the architecture were characterized [17-18]. According the table, the chosen technologies or methods seek to meet the requirements described in the table. Note that each dimension presents the suggested methods, which is implemented by using technologies for building real-time data pipelines and streaming applications.

4.1. Too Much Emphasis on Some Functions Leads to Poor Scalability

If too much emphasis is in the functionality, it may increase the database modification difficulty. This means that the design must be made detail-oriented. If the design phase takes too long or too short, it will cause an inconvenient product. Therefore, the design process must be made with the goal of using said design in mind. The most important aspect to anchor yourself is the ease of use and functionality of the system.

4.2. Database Design Is Not Perfect

Many developers have trouble with connecting multiple databases. This causes problems like not being able to display every piece of information we seek, and the inability to comprehensively update databases. This directly affects the user's access to pertinent data.

4.3. Design of Real-Time Database

Real-time databases are responsible for real-time data and resource management, data transfer, and other work. Real-time databases reside in memory and process data [19].

4.4. Lack of Data Log Information

The information log should be established in the design process. This allows for database analysis and establishing design quality later on. This part of the database ought not to be skipped. The information log will affect the accurate control of the entire database. Information log is necessary in the continuous refining of the system, which is the only way to create a perfect system.

Finally, we would like to emphasize the Doric architecture was implemented using Apache Kafka technologies for building real-time data pipelines and streaming applications. The obtained results were encouraging and show the potential for applying Doric as a structure to foster the development of modern information systems in organizations and to support serve as a guideline for new corporate architectures [20]. Regarding the research, the prototype was evaluated in five scenarios containing different volumes of data and the architecture is composed by the following six steps to process data and six architectural components. These steps are used when consumers desire to access the result of the analysis [20]. The solution can help to enhance a flexible and modular architecture for data-intensive and real-time systems.

5. Conclusion

The paper concludes even though the usage of real time based database technologies increased in recent years, there is a lack of standardization and modularized architectures. In practice, there are many problems in the design phase of software database development. The study shows we lack the vitality of the design phase of software systems.

Database design is much harder than software design. It is a comprehensive integration of the database into the system that causes problems. Designers ought to have a good knowledge of database design and constantly be improving their skills. This will help achieve an excellent database with great functionality. Another tip for database designers is to follow certain principles and look at the system from the user's point of view. For the software to have a good overall performance, it all starts with the database design. If the foundation is poured correctly, the software can run stable and safely. Lastly, future work will need to consider the architecture with improved hardware capability and evaluate the architecture's adaptability in more scenarios.

The obtained results from our study encourage and show the potential for applying the software architecture for data-intensive real-time applications in the future.

References

[1] Mingwu F, Analysis of database programming technology based on computer software engineering in the new era Hubei Agricultural Mechanization 21 157-8, 2019.

[2] L. Almeida, F. Santos, Coordinating distributed autonomous agents with a real-time database: The cambada project. In Proceedings of the 19th International Symposium of Computer and Information Sciences, 2004.

[3] Dan Jia, Research on Method of Database Platform Design and Machine Learning Performance Prediction of Lubricating Materials. General Institute of Mechanical Research, 2021.

[4] Jing Ziqi, Zhaonian, Design and implementation of multi-version concurrency control on embedded database SQLite. Journal of Computer Applications, Journal of Plant Genetic Resources, 2022.

[5] Ni Jiacai, Key Techniques of Multi-Tenant Database Design, Tsinghua University, 2015.

[6] Warr WNick Laus C et al. Exploration of ultra large compound collections for drug discovery, Journal of Chemical Information and Modeling, 62, 9: 2021-2034, 2022.

[7] PubChem project. PubChem compound. https://www.ncbi.nlm.nih.gov/pccompound accessed July 1, 2022.

[8] Jain A, Hauter R, The materials project: A materials genome approach to accelerating materials innovation, APL Materials, 2013.

[9] Wang Yu-fu, Chen Li, Function and Use of the World Major Grape Germplasm Resources Database [J]. 2022.

[10] Energetic materials compendium (EMC) v5.0. https://www.msi-ac.nato.int/news/energetic-materials-compendium-emc-v50 accessed July 1, 2022.

[11] Jingjing Chen. Research and Visualization of SQLite Database. Nanjing University of Posts and Telecommunications. 2020.

[12] Zhiyi Wang. Design of Data Upload System Based on Database Cluster. China Electronics Technology Group Corporation Electronic Science Research Institute. 2021.

[13] Bei Han. Research and Application of SQLite Database. Nanjing University of Posts and Telecommunications for the Degree of Master of Engineering. 2019.

[14] ZHANG Hao, REN Jing. Design and research of smart tourism information service platform. journal of Computer Applications. 2021.

[15] Ju ruli, Ying Chunma. Design of Information Management Database Based on Computer Technology. Information and computers. 2022.

[16] Tao Qiu, Bin Wang, Shaowei Shu, Zhibo Zhao, Zi wen song, Yan hui zhong. Intelligent index tuning method for relational database. Journal of software. 2020.

[17] Yanqing Yang, Xian Chongguo. Problems and design principles in computer database design. Wireless Internet Technology Phase II. 2021.

[18] Jiyang HU. Research on Necessity and Design Method of Computer Software Database Development. Computer Knowledge and Technology, 2020 (29): 36-37.

[19] Yangdong He. Design and Implementation of Embedded real-time Database for Intelligent Agricultural Machinery. Nanjing University. 2019.

[20] Cadaviz, Miguel & Farias, Kleinner & Gonçales, Lucian & Bischoff, Vinicius. (2018). DORIC: An Architecture for Data-intensive Real-time Applications. 1-7. 10.1145/3229345.3229416.