Study on Simplified Database Management System for Auxiliary Computer Science Education for College Student

Taizhi Lv[1, a], Xuejun You[1, b] and Yong Chen[2, c]

[1] School of Information Technology, Jiangsu Maritime Institute, Jiangsu Nanjing 211170, China.

[2] Nanjing Longyuan Microelectronic Company Limited, Jiangsu Nanjing 211106, China.

[a]lvtaizhi@163.com, [b]68151247@qq.com, [c]68151247@qq.com

Abstract

With the rapid development of society, we are in an era of information explosion. Now more and more data needs to be stored and processed, and the requirements for data process are getting higher and higher. The demand for talent of data process software development also has increased rapidly. In order to make students of computer science major master the technologies of data process and software development, the simplified database management system for auxiliary computer science education is designed and implemented. It helps students to master the techniques of massive data processing, database paradigm, network database, relational model, indexing mechanism, network programming, data compression and other technologies. Although deficiency of commercial value, some technologies of this system is used in commercial software, and this system can improve students' professional level and mastery of current main technologies. This system have both theoretical and practical significance.

Keywords

Simplified database management system, Java, Regular expression, computer science education, Software development.

1. Introduction

Computer Science is the study of algorithms, data and computing systems. A computing system can accept, store, transform and present data in ways which contribute to knowledge and the well-begin of society [1]. As a computer science and technology major, college students will focus on designing and implementing systems. Applied higher education is not only the requirement of the development of society, economy and science and technology, but also the requirement of popularization of higher education.

Now, there are many problems in computer science and technology education for application-oriented college. Weak ability of engineering practice is one of most important problems. It is difficult to make students master engineering practice skill only by theory learning. It is helpful to improve the practice ability of students by introducing actual application to teaching work. Many colleges use as Linux as an example to teach operating system principles course, and MySQL as an example to teach database principles course, and Java language as an example to teach programming course. Most of these systems are complex and not open, and it is difficult to strengthen students learning by analyzing them. In order to improve the quality of teaching, many famous universities develop simplified system to assist computer science teaching. Karel created by Richard E. Pattis is an educational programming language for beginners and firstly used in programming methodology course teaching in Stanford University [2]. Then many universities and colleges use it in programing teaching. It has proved that Karel can significantly decrease the learning difficult of Java beginner and improve the interest of them. Andrew S. developed a simple operating system Minix which is similar to Unix operating system and is open source [3]. It is used in the teaching of operating system courses. There are many knowledge points and complex concepts in the compiler principle course. These are too theoretical that very difficult to understand for students. To overcome the defects, Ding

implements auxiliary teaching software about process visualization class c compiler [4]. The better teaching results are achieved by combining theory with practice.

The methods above all focus on improvement of single course of the computer science major by the study on teaching auxiliary software. The actual software development involves many courses of the major. It is useful to help students understand their major comprehensively by the research and development of an example system. For improving teaching of computer science major, a simplified database management system is developed. The system can be applied teaching of some courses, such as computer programming, database technology, data structure and algorithm, software engineering, design pattern, computer network, visual programming, design pattern. This system can also let students have a systematic understanding of software development.

The rest of this paper is organized as follows. Section 2, at first, introduces the design of this system. There related tools are introduced in Section 3. Finally, Section 4 concludes this paper.

2.   System Design

This database management system consists of a database service, a server configuration tool, a client tool, and a JDBC interface for the Java programmer. The main functions of the database management system include database schema definition, data manipulation, data query and big data file import. The database schema definition function includes creating databases, creating data tables, creating indexes, creating views, and so on. The data operations function includes adding records, modifying records, and deleting records. The data query function provides the query operation by select query language. This function realizes the projection operation by the choose columns, realizes the selection operation by the where clause, and realizes the sorting operation by the order clause. Big data file import supports more than 10G text files imported into a database.

2.1 Service Execution Flow

The database service can be installed as a start-up service which can de started with the operating system. It also can be started by the user. After the service program starts, it first loads the configuration file, starts monitoring according to the configuration parameters, and waits for the user to connect. When a client connects to the server, the server creates a new thread to complete the interaction with the client, and the service continues to wait for a new connection from client.
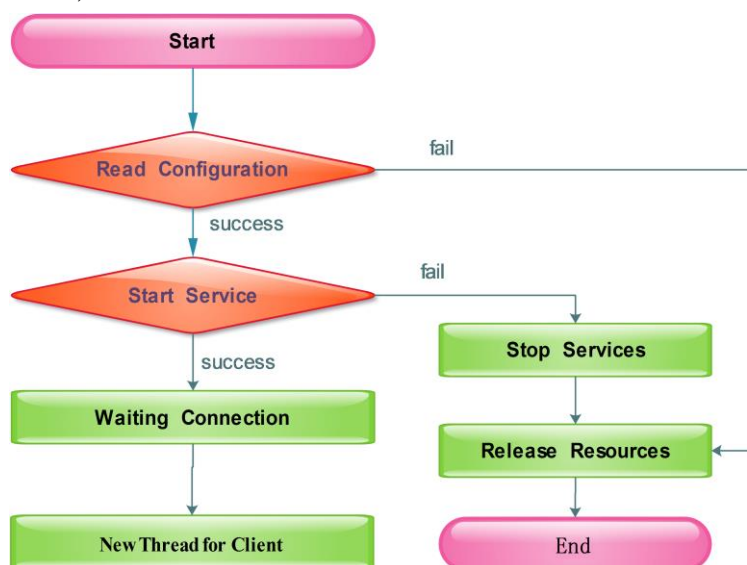

Figure 1. The service execution flowchart

The new thread first verifies whether the client is authorized or not. If the user is not authorized, the connection is disconnected and the thread is terminated. If the client is an authorized user, the thread is continued to wait for the interactive command from the client. There are three client commands

supported, namely DDL (Data Define Language), DML (Data Manipulation Language), and DQL (Data Query Language). The thread execution process is shown in Figure 2.
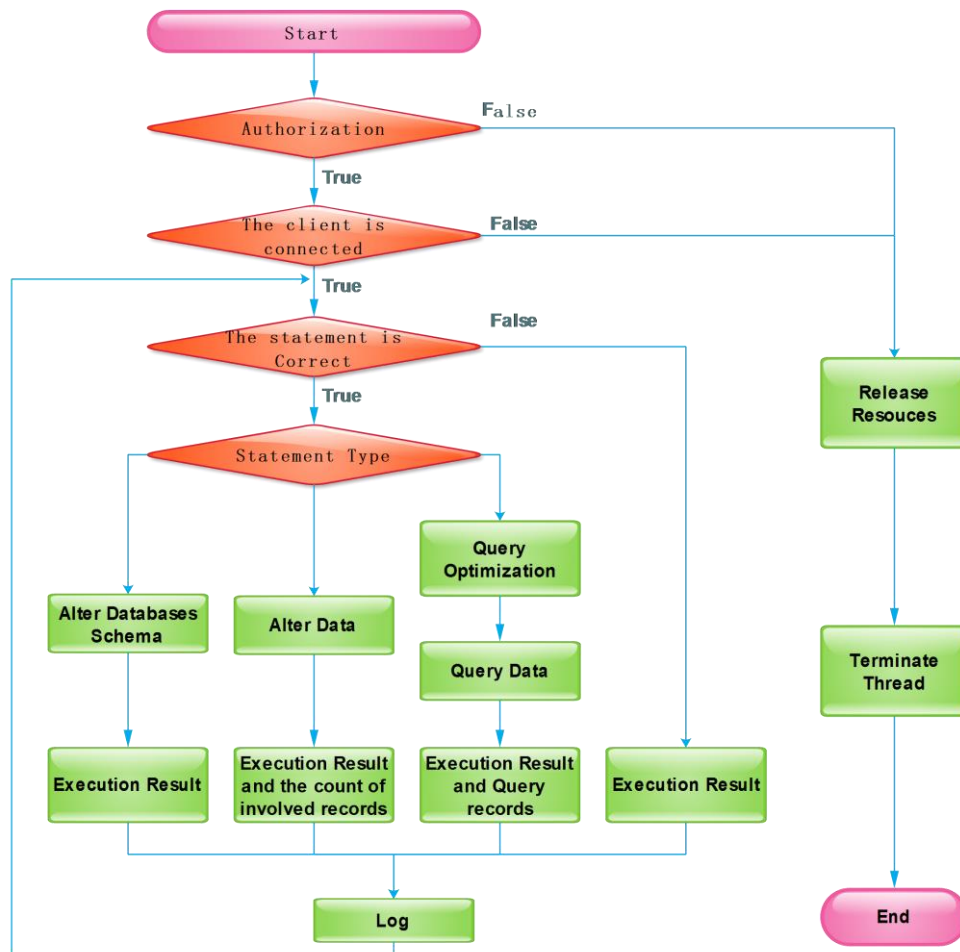

Figure 2. The thread execution flowchart

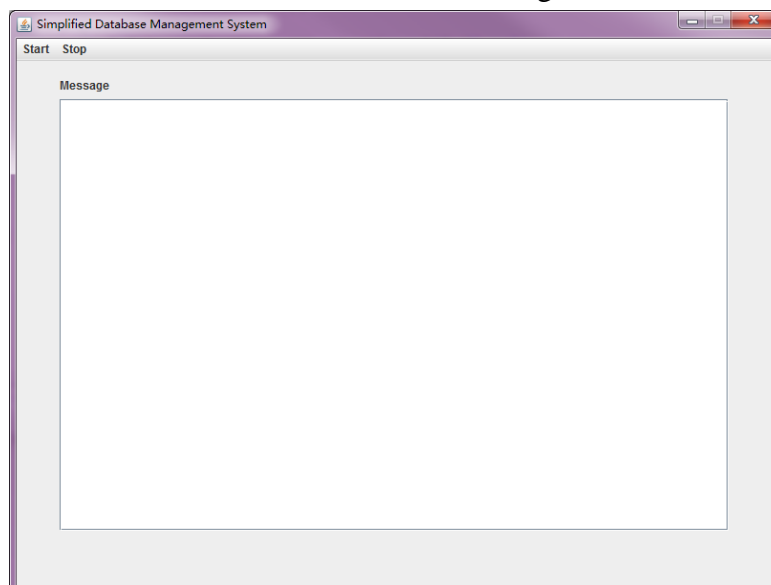The graphical interface of the database service is shown in Fig. 3.

Figure 3. The start-up graphical interface

2.2 Data Structure And Index

Records are divided by the separator of 'FF FE FD FC', and the line number of the record is followed by the separator. The line number facilitates the quick search of the record. Fields of each record are

divided by 'FF FA F9'. The index file structure is preceded by the index field value, followed by the corresponding row number. The index can be only a fixed length such as Integer, char (32), double, etc. The index cannot be the type of varchar.

Since the data is stored on the hard disk and I/O operations are very time consuming. For massive data, the index is also very large and cannot be loaded into memory at one time. In order to reduce I/O operation, The B+ tree is used to this system for achieving the index [5]. It can improve the efficiency of the search by the locality principle and the characteristics of disk read-ahead.

2.3 Grammar Check Based on Regular Expression

Regular expression is a tool used to describe a string match. It can be used to retrieve text, verify whether a string symbolizes a pattern, and replace the text content of a pattern [6-7]. This system uses it to complete the verification of SQL statements and the extraction of content. The following is the validation mode of the select query statement:

"^select(\\s+district){0,1}\\s+([\\w,*]+)+\\s+from\\s+(\\w+)(\\s+where\\s+[\\w<=>!\u4e00-\u9fa5']+){0,1}(\\s+order\\s+by\\s+(?:[\\w,\\s]+?)){0,1}"+"(\\s+limit\\s+[0-9]+(?:\\s*,\\s*[0-9]+){0,1}){0,1}"+"$";

2.4 Big Data Import

In order to speed up the transmission of massive data between the client and the server, and improve the data uploading efficiency, an improved block compression algorithm is adopted. The client uses the RandomAccessFile class to read in a large amount of data files, and then passes the block compression to the server. The server decompresses according to the tag, and stores the data in the server. Compression uses a parallel-based Gzip compression algorithm, and its performance has been greatly improved [8]. The data block uploading and storage system runs in a block transfer allocator that connects multiple clients and one storage server through a network. The server is a receiving module, configured to receive a hash list of file information and files uploaded by the multiple clients at the same time, and store them in a hash database, where each file corresponds to a hash list. The client is an uploading module, the data block is uploaded to the server.

3.   Tools and Interface Design

3.1 Configuration Tool

The configuration module for this system is to configure the key parameters with the server. There are eight parameters which should be configured.

(1) Port number: It is a sock port address by which the client can connect the service.

(2) Database path: The database path is a position in where data store.

(3) Maximum number of connections: the maximum number of connections that the client connects to the database, if the client

(4) Max Connection: It is a max connection number. If the connection number is exceeded this number, the service will reject the new connection.

(5) Whether the null value is sorted before the not null value: It decides the sequence of data query.

(6) Date format: It is used for recognizing date in import big data text file.

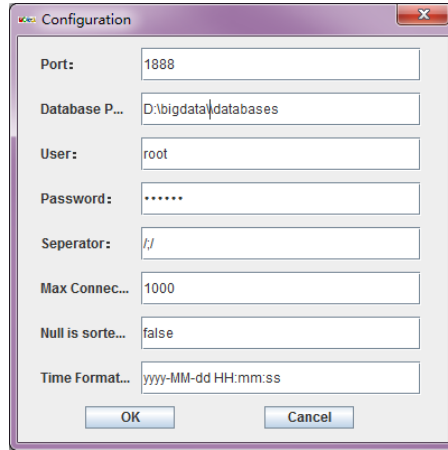(7) User name and password: They are used for verifying the authorization of client connection.

Figure 4. Configuration tool

3.2 Client Tool

After the client tool is started, users should input the IP address, port number, user name and password. The graphical interface is shown in Fig. 5.
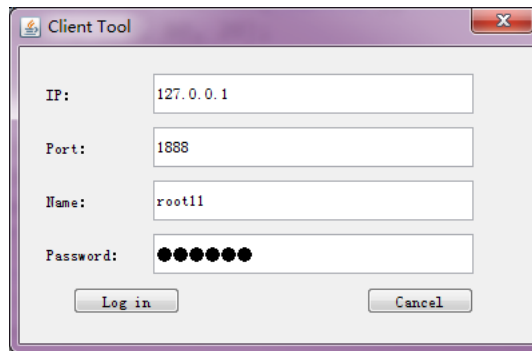

Figure 5. Login

These four input item should be correct. If one of them is error, this tool will prompt "Server connection failed" and user should re-enter the IP address, port number, username and password. If users have successfully connected to the server, users can enter the data operation command. The main interface is shown in Fig. 6.
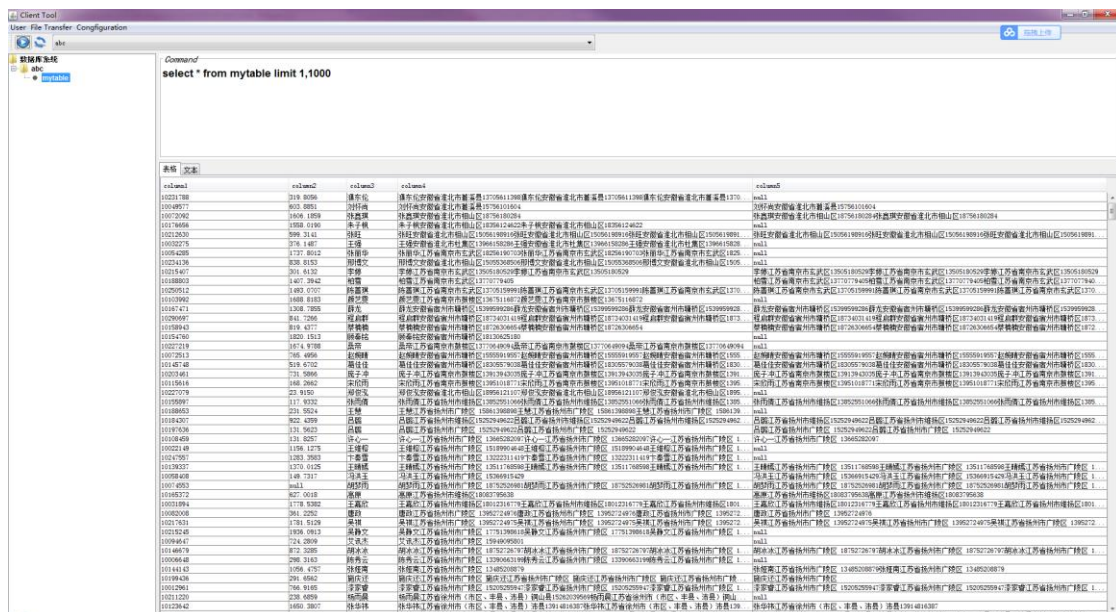

Figure 6. Main graphical interface of the client tool

The client has the following functions: login, connection, switch database, query statement, database and table display, display data table structure, data upload, create database and table, insert data and so on.

3.3 JDBC Driver

JDBC (Java Database Connectivity) provides an API for executing SQL statements in the Java language. JDBC only provides basic interfaces, and the specific implementation requires various database system vendors to implement them. In this way, this system should complete operations of the database through a unified interface. Fig. 7 shows how the JDBC interface is called. The simplified DBMS (Database management system) represents this system proposed in this paper.
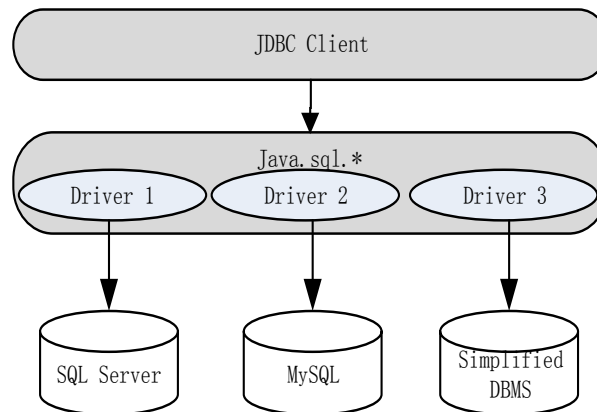
Figure 7. JDBC Interface

4. Conclusion

This paper introduces a simplified database management system for auxiliary computer science education for college student. It adopts technologies such as Java, C/S architecture, network, thread. Although this system has a big gap compared with the commercial DBMS software, the implementation of this system can help students have a deeper understanding of the basic principles of database technology, such as data models, relational models, indexing mechanisms, SQL language, network databases, data files, and so on. Applying the simplified database management system to the teaching work not only helps the students to understand the basic principles of the database, but also the application of the system involves multi-threading, network programming, Java language, regular expression and other technologies.

Acknowledgements

References

[1] Gurevich Y. Logic and the Challenge of Computer Science[J]. Trends in Theoretical Computer Science, 2017:1-57.
[2] Coufal P, Hornik T, Hubalovsky S, et al. The Development of KarelNXT Robot as a Simulation of xKarel Programming Language[C]// International Conference on Applied Physics, System Science and Computers. Springer, Cham, 2017:105-111.
[3] Tanenbaum A S. Lessons learned from 30 years of MINIX[J]. Communications of the Acm, 2016, 59(3):70-78.
[4] Ding Z, Zhou Z, Wei Z. An Auxiliary Teaching Software About Process Visualization Class C Compiler[J]. Computer and Modernization, 2017(6): 34-39.

[5] Zhang D, Baclawski K P, Tsotras V J. B+-Tree[J]. Encyclopedia of Database Systems, 2009, 288(22):15537-15546.

[6] Li Y, Krishnamurthy R, Raghavan S, et al. Regular Expression Learning for Information Extraction.[C]// Conference on Empirical Methods in Natural Language Processing, EMNLP 2008, Proceedings of the Conference, 25-27 October 2008, Honolulu, Hawaii, Usa, A Meeting of Sigdat, A Special Interest Group of the ACL. DBLP, 2008:21-30.

[7] Garofalakis M N, Rastogi R, Shim K. SPIRIT: Sequential Pattern Mining with Regular Expression Constraints[C]// Proc. International Conference on Very Large Data Bases, September 7-10, Edinburgh, Scotland, Uk. 1999:223--234.

[8] Song G, Jiang M Q, Zhang Y Q, et al. Parallel compression with gzip on shared-memory systems[J]. Computer Engineering & Design, 2009, 30(4):781-784.

[9] Gould C, Su Z, Devanbu P. JDBC Checker:A Static Analysis Tool for SQL/JDBC Applications[C]// International Conference on Software Engineering, 2004. ICSE 2004. Proceedings. IEEE, 2004:697-698.