

## 摘 要

随着因特网应用和计算机技术的飞速发展，数据库逐渐成为信息系统的核心部分并广泛应用于企业、金融机构、政府及国防等各个领域。其中，分布式关系型数据库以其低成本、高可靠性等特点成为当前数据库理论与应用领域的研究热点。

通过对分布式理论、关系型数据库理论及相关技术的学习和研究，本文基于Java语言实现了一个分布式的关系型数据库JSQL。JSQL主要包含五大模块，网络模块接受前端应用程序的连接请求，对客户端进行认证和授权，并对连接进行管理，可实现基于Mysql的通信协议，这样便于Mysql的用户方便的迁移到本系统上来；Sql的解析和执行模块接受客户端的SQL请求，然后解析和执行数据库存储的调用，返回执行结果；审计模块是存储和分析所有对数据库的更改情况，以可视化的方式向用户展示；数据库引擎模块利用了orientdb开源的数据库引擎，实现可靠的分布式存储；分布式模块利用hazlcast实现了数据库集群。基于以上功能模块，JSQL具有高可用性，可扩展性，负载均衡等特性，同时从数据库底层考虑了数据库安全审计需求，加入了数据审计图形化界面显示审计结果。

论文对系统的进行了功能和性能测试。功能测试结果表明，系统在功能上符合分布式数据库的基本要求，审计系统的功能也达到本论文的要求。论文通过对性能测试结果进行分析，认为系统的性能基本达到本论文的要求。但是本系统对复制SQL语句的支持还不是很完善，最后提出了改进的方案。

**关键词：**分布式数据库，mysql，安全审计，OLTP，NOSQL



## ABSTRACT

Database, operating system and compiler and called the three systems, can be said that the cornerstone of the entire computer software. Which is closer to the application layer database, is a lot of business support. This field after decades of development, There are new developments. From the beginning of the hierarchy database and relational database, to the recent hot Nosql database, and then to the recent Google Spanner and F1 as the representative of the NewSql database.

In the Internet age, the storage and access of massive data becomes the bottleneck of system design and use. For mass data processing, King, divided into two types: online transaction processing (OLTP) and online analytical processing (OLAP).

Relational database is based on the relational model of the database, which by means of aggregation algebra and other mathematical concepts and methods to deal with the database data. Which is the most popular mysql, mysql is an open source relational database, the advantage lies in the open source code, any business and individuals can according to their own needs to modify the source code mysql.

NoSQL database, called Not Only SQL, meaning that when the relational database is used when the relational database, not applicable There is no need to use relational database is not necessary, you can consider the use of more appropriate data storage.

Oracle, mysql and other traditional relational database is very mature and has been large-scale commercial, why use NoSQL database? mainly With the development of the Internet, the amount of data is growing, the performance requirements are getting higher and higher, the traditional database of congenital defects, namely stand-alone (single Library) performance bottlenecks, and difficult to expand. This is a stand-alone single library bottleneck, but difficult to expand, naturally unable to meet the growing mass of data storage And its performance requirements, so there will be a variety of different NoSQL products.

Although in the cloud computing era, the traditional database there are congenital defects, but NoSQL database can not be replaced, NoSQL can only For the traditional data supplement can not be replaced, so to avoid the shortcomings of traditional databases is the current era of large data must be resolved.

In order to solve these problems, such as mysql and other relational database, this article describes how to design and implement a compatible mysql protocol distributed database. He can automatically find the distributed database cluster nodes, Automatically allocate data to support massive data storage. Taking into account the increasingly important security of the database, from time to time the occurrence of database administrators or other attackers malicious changes in the database data, So the database developed by the database from the bottom of the database to join the audit function.

The main work of this paper is as follows:

Based on the realization of java language compatible mysql communication protocol database;

The realization of the database cluster, to achieve the database of high availability, scalability, load balancing and other characteristics;

From the bottom of the database to consider the database security, joined the data audit and other features.

**Keywords:** mysql, java, mysql, java, mysql

# 目 录

<b>第一章 绪论</b> .....	1
1.1 研究工作的背景与意义 .....	1
1.2 国内外研究历史与现状 .....	1
1.3 论文的主要工作 .....	1
1.4 本论文的结构安排 .....	2
1.5 本章要求 .....	2
<b>第二章 理论基础和相关的技术</b> .....	3
2.1 分布式数据库概念和技术 .....	3
2.1.1 数据库的分类和架构 .....	3
2.1.2 分布式数据库的介绍 .....	3
2.1.3 分布式数据库相关技术 .....	3
2.1.3.1 负载均衡 .....	3
2.1.3.2 数据分片技术 .....	3
2.1.3.3 数据库高可用技术 .....	3
2.1.4 Mysql数据库体系结构 .....	5
2.2 Mysql通信协议的研究 .....	5
2.2.1 交互过程 .....	5
2.2.2 协议基本类型 .....	6
2.2.3 mysql报文结构 .....	6
2.2.4 报文类型 .....	7
2.3 本章小结 .....	9
<b>第三章 需求分析</b> .....	10
3.1 系统概述 .....	10
3.2 需求分析 .....	10
3.3 通信协议分析 .....	10
3.4 编程模型分析 .....	10
3.5 本章小结 .....	10
<b>第四章 系统设计</b> .....	11
4.1 总体架构设计 .....	11
4.2 系统模块划分 .....	11

4.2.1 数据库模块设计 .....	11
4.2.2 集群架构设计 .....	11
4.2.3 数据审计模块设计 .....	13
4.3 各模块详细设计 .....	13
4.3.1 数据库模块设计 .....	13
4.3.2 集群架构的设计 .....	14
4.3.3 数据审计架构的设计 .....	14
4.4 本章小结 .....	14
<b>第五章 系统实现 .....</b>	<b>15</b>
5.1 数据库模块实现 .....	15
5.1.1 网络模块 .....	15
5.1.2 sql解析模块 .....	15
5.1.3 存储引擎模块 .....	15
5.2 集群架构的实现 .....	15
5.2.1 通信模块 .....	15
5.2.2 数据复制 .....	15
5.2.3 负载均衡 .....	16
5.3 数据审计架构的实现 .....	16
5.3.1 审计数据库 .....	16
5.3.2 审计管理器 .....	16
5.3.3 审计可视化模块的实现 .....	16
5.4 本章小结 .....	16
<b>第六章 系统测试 .....</b>	<b>17</b>
6.1 测试环境 .....	17
6.2 功能和性能测试 .....	17
6.2.1 基本功能测试 .....	17
6.2.2 数据库集群测试 .....	17
6.3 数据库审计和可视化测试 .....	17
6.4 本章小结 .....	17
<b>致 谢 .....</b>	<b>18</b>
<b>参考文献 .....</b>	<b>19</b>

## 第一章 绪论

在互联网时代，海量数据的存储与访问成为系统设计与使用的瓶颈问题，对于海量数据处理，按照使用场景，主要分为两种类型：联机事务处理（OLTP）和联机分析处理（OLAP）。

联机事务处理（OLTP）也称为面向交易的处理系统，其基本特征是原始数据可以立即传送到计算中心进行处理，并在很短的时间内给出处理结果。

联机分析处理（OLAP）是指通过多维的方式对数据进行分析、查询和报表，可以同数据挖掘工具、统计分析工具配合使用，增强决策分析功能。

### 1.1 研究工作的背景与意义

web2.0的到来，存储海量数据的需求越来越大。

在互联网时代，海量数据的存储与访问成为系统设计与使用的瓶颈问题，对于海量数据处理，按照使用场景，主要分为两种类型：联机事务处理（OLTP）和联机分析处理（OLAP）。

联机事务处理（OLTP）也称为面向交易的处理系统，其基本特征是原始数据可以立即传送到计算中心进行处理，并在很短的时间内给出处理结果。

联机分析处理（OLAP）是指通过多维的方式对数据进行分析、查询和报表，可以同数据挖掘工具、统计分析工具配合使用，增强决策分析功能。

### 1.2 国内外研究历史与现状

google spanner和t1等系统。在互联网时代，海量数据的存储与访问成为系统设计与使用的瓶颈问题，对于海量数据处理，按照使用场景，主要分为两种类型：联机事务处理（OLTP）和联机分析处理（OLAP）。

联机事务处理（OLTP）也称为面向交易的处理系统，其基本特征是原始数据可以立即传送到计算中心进行处理，并在很短的时间内给出处理结果。

联机分析处理（OLAP）是指通过多维的方式对数据进行分析、查询和报表，可以同数据挖掘工具、统计分析工具配合使用，增强决策分析功能。

### 1.3 论文的主要工作

结合nosql的关系型数据库的特点，同时从数据库自身考虑数据库安全。 ...

## 1.4 本论文的结构安排

本论文后续各章结构如下：

第二章介绍了

第三章介绍了

第四章描述了审计子系统是如何实现，同时将涉及的技术方法作了介绍和分析。

第五章对审计管理器模块进行了分析和设计。首先提出该模块的概述，并结合这些目标对系统总体结构和重要的功能进行了详细的分析和设计，从而保证设计目标的实现。

第六章主要通过介绍审计子系统的具体应用，说明了审计子系统的实用价值和地位。

结论部分详细总结了本论文的研究工作，并针对存在的主要不足指出了改进的重点和今后的方向。

## 1.5 本章要求

...



## 第二章 理论基础和相关的技术

本章主要写了又那些数据库，每种数据库的特点。

...

非关系型数据库包括下面3个

### 2.1 分布式数据库概念和技术

#### 2.1.1 数据库的分类和架构

分布式数据库技术的概述

...

#### 2.1.2 分布式数据库的介绍

分布式数据库技术的概述

...

#### 2.1.3 分布式数据库相关技术

##### 2.1.3.1 负载均衡

##### 2.1.3.2 数据分片技术

##### 2.1.3.3 数据库高可用技术

...

### 负载均衡

#### 负载均衡的概念

分布式数据库技术的概述

...

#### 负载均衡算法

论云

...

## 一致性hash算法

分布式数据库技术的概述

...

## 数据分片技术

...

## 数据分片概述

分布式数据库技术的概述

...

## 数据库分片策略

分布式数据库技术的概述

...

## 数据库高可用技术

分布式数据库技术的概述

...

## 集群可靠性技术

分布式数据库技术的概述

...

## 数据复制

分布式数据库技术的概述 ...

## 数据一致性

分布式数据库技术的概述

## 数据库的分类和架构

### 非关系型数据库

#### 键值存储数据库

主要是内存数据库

...

## 列存储数据库

比如hbase数据库

...

## 面向文档数据库

主要的是mangdb数据库

...

## 图形数据库

主要有orientdb

...

## 关系型数据库

关系数据库是主要的数据库

...

## 关系数据库重要理论

主要有acid 事务是数据库的主要特点

...

## 关系数据库架构

sql解析

...

### 2.1.4 Mysql数据库体系结构

从上到下

## 2.2 Mysql通信协议的研究

...

### 2.2.1 交互过程

...

### 握手认证阶段

本章主要说了java的相关技术，因为jsql是基于java开发的数据库。java是一个跨平台的语言。

...

### 命令执行阶段

本章主要说了java的相关技术，因为jsql是基于java开发的数据库。java是一个跨平台的语言。

...

## 2.2.2 协议基本类型

### 整型值

本章主要说了java的相关技术，因为jsql是基于java开发的数据库。java是一个跨平台的语言。

...

### null结尾的字符串

本章主要说了java的相关技术，因为jsql是基于java开发的数据库。java是一个跨平台的语言。

...

### 长度编码的字符串

本章主要说了java的相关技术，因为jsql是基于java开发的数据库。java是一个跨平台的语言。

...

### 二进制数据(长度编码)

本章主要说了java的相关技术，因为jsql是基于java开发的数据库。java是一个跨平台的语言。

...

## 2.2.3 mysql报文结构

...

## 报文头

本章主要说了java的相关技术，因为jsql是基于java开发的数据库。java是一个跨平台的语言。

...

## 报文数据

本章主要说了java的相关技术，因为jsql是基于java开发的数据库。java是一个跨平台的语言。

...

## 2.2.4 报文类型

...

## 登陆认证交互报文

本章主要说了java的相关技术，因为jsql是基于java开发的数据库。java是一个跨平台的语言。

...

## 客户端命令请求报文

本章主要说了java的相关技术，因为jsql是基于java开发的数据库。java是一个跨平台的语言。

...

## 服务器响应报文

本章主要说了java的相关技术，因为jsql是基于java开发的数据库。java是一个跨平台的语言。

...

## JAVA网络和多线程技术

## TCP/IP相关协议

...

## TCP/ip协议概述

本章主要说了java的相关技术，因为jsql是基于java开发的数据库。java是一个跨平台的语言。

...

## TCP通信的连接和终止

本章主要说了java的相关技术，因为jsql是基于java开发的数据库。java是一个跨平台的语言。

...

## java流IO技术

本章主要说了java的相关技术，因为jsql是基于java开发的数据库。java是一个跨平台的语言。

...

## Socket编程技术

本章主要说了java的相关技术，因为jsql是基于java开发的数据库。java是一个跨平台的语言。

...

## NIO技术

本章主要说了java的相关技术，因为jsql是基于java开发的数据库。java是一个跨平台的语言。

...

## 线程池技术

本章主要说了java的相关技术，因为jsql是基于java开发的数据库。java是一个跨平台的语言。 ...

## Reactor模式

本章主要说了java的相关技术，因为jsql是基于java开发的数据库。java是一个跨平台的语言。

...

## Netty简介

本章主要说了java的相关技术，因为jsql是基于java开发的数据库。java是一个跨平台的语言。

## 2.3 本章小结

存储部分的数据库主要分为两大类：关系型数据库与 NoSQL 数据库。

...

## 第三章 需求分析

本章主要写了又那些数据库，每种数据库的特点。

### 3.1 系统概述

### 3.2 需求分析

### 3.3 通信协议分析

### 3.4 编程模型分析

..

### 3.5 本章小结

存储部分的数据库主要分为两大类：关系型数据库与 NoSQL 数据库。

...



## 第四章 系统设计

### 4.1 总体架构设计

jsql是一个可以替代mysql的分布式java数据库

...

### 4.2 系统模块划分

下面是分布式环境下的部署图：

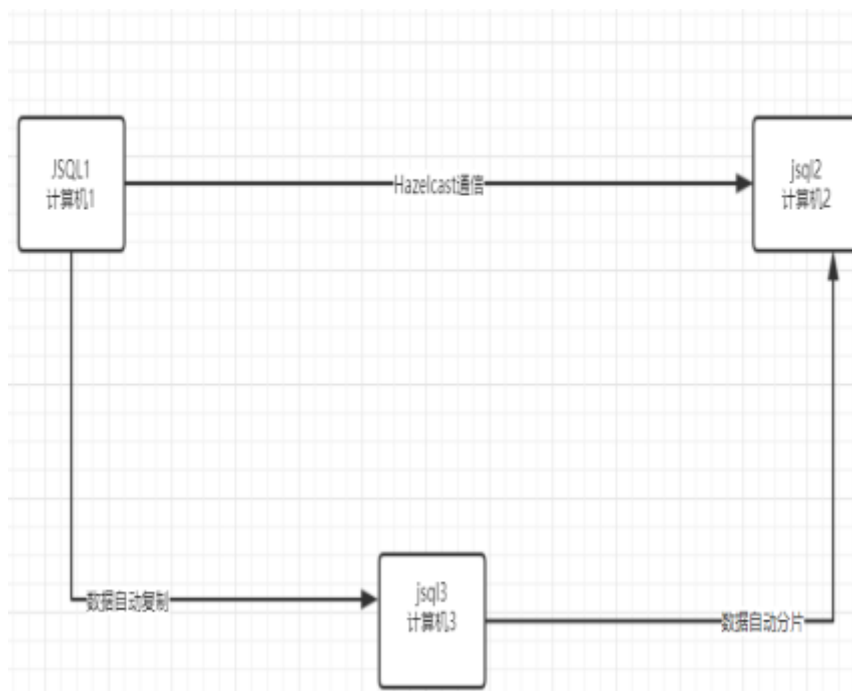


图 4-1 分布式环境下架构

图6-2是jsql的架构图

#### 4.2.1 数据库模块设计

数据库主要是sqk到粗存

...

#### 4.2.2 集群架构设计

主要用hazelcast

...

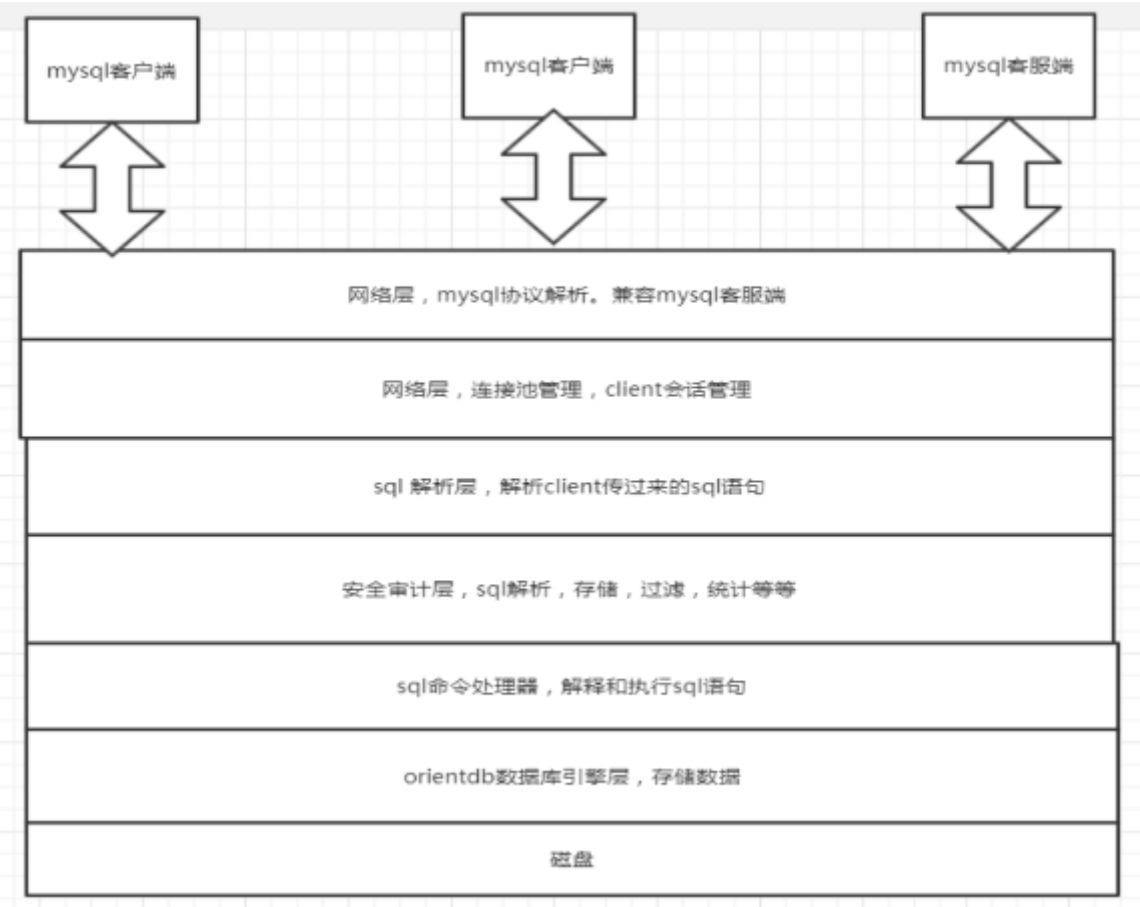


图 4-2 JSQL架构

### 4.2.3 数据审计模块设计

数据库审计的

...

elasticsearch简介

...

grafana简介

...

审计的架构

...

...

## 4.3 各模块详细设计

本文主要是

...

### 4.3.1 数据库模块设计

主要分为3个部分

...

网络模块的设计

数据库主要是sqk到粗存

...

sql解析模块的设计

主要用hazelcast

...

存储引擎模块的设计

数据库审计的

...

### 4.3.2 集群架构的设计

主要用hazelcast

数据库审计的

...

### 4.3.3 数据审计架构的设计

...

#### 审计数据库设计

数据库主要是sqk到粗存

...

#### 审计管理器设计

主要用hazelcast

...

#### 审计可视化模块的实现

数据库审计的

## 4.4 本章小结

分布式数据库

...

## 第五章 系统实现

本文主要是

...

### 5.1 数据库模块实现

主要分为3个部分

...

#### 5.1.1 网络模块

数据库主要是sqk到粗存

...

#### 5.1.2 sql解析模块

主要用hazelcast

...

#### 5.1.3 存储引擎模块

数据库审计的

...

### 5.2 集群架构的实现

主要用hazelcast

...

#### 5.2.1 通信模块

数据库主要是sqk到粗存

...

#### 5.2.2 数据复制

主要用hazelcast

...

### 5.2.3 负载均衡

数据库审计的

...

## 5.3 数据审计架构的实现

...

### 5.3.1 审计数据库

数据库主要是sqk到粗存

...

### 5.3.2 审计管理器

主要用hazelcast

...

### 5.3.3 审计可视化模块的实现

数据库审计的

...

## 5.4 本章小结

分布式数据库

...

## 第六章 系统测试

...

### 6.1 测试环境

...

### 6.2 功能和性能测试

...

#### 6.2.1 基本功能测试

.....

#### 6.2.2 数据库集群测试

.....

### 6.3 数据库审计和可视化测试

...

### 6.4 本章小结

...

## 致 谢

在攻读硕士学位期间，首先衷心感谢我的导师曹晟老师...



## 参考文献

- [1] 王浩刚, 聂在平. 基于mysql的分布式数据库的研究和实现[J]. 电子学报, 1999, 27(12):68–71
- [2] X. F. Liu, B. Z. Wang, W. Shao. A marching-on-in-order scheme for exact attenuation constant extraction of lossy transmission lines[C]. China-Japan Joint Microwave Conference Proceedings, Chengdu, 2006, 527–529
- [3] 竺可桢. 物理学[M]. 北京: 科学出版社, 1973, 56–60
- [4] 陈念永. 基于mysql的分布式数据库的研究和实现[D]. 成都: 电子科技大学, 2001, 50–60
- [5] 顾春. 基于mysql的分布式数据库的研究和实现[N]. 人民日报, 2012年3月31日
- [6] 冯西桥. 基于mysql的分布式数据库的研究和实现[R]. 北京: 清华大学核能技术设计研究院, 1997年6月25日
- [7] 肖珍新. 基于mysql的分布式数据库的研究和实现[P]. 中国, 实用新型专利, ZL201120085830.0, 2012年4月25日
- [8] 中华人民共和国国家技术监督局. GB3100-3102. 中华人民共和国国家标准—量与单位[S]. 北京: 中国标准出版社, 1994年11月1日
- [9] M. Clerc. Discrete particle swarm optimization: a fuzzy combinatorial box[EB/OL]. [http://clere.maurice.free.fr/ps0/Fuzzy\\_Discrere\\_PS0/Fuzzy\\_DPS0.htm](http://clere.maurice.free.fr/ps0/Fuzzy_Discrere_PS0/Fuzzy_DPS0.htm), July 16, 2010