

**本 科 毕 业 论 文（设 计）**

题目（中文） **数据仓库分层架构模型的设计与搭建**

（英文）The Design and Construction of the Layered Architecture Model for Data Warehouses

学 院 信息与机电工程学院

年级专业 2020级计算机科学与技术

学生姓名 周文尧

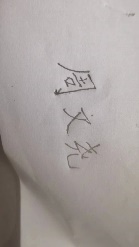
学 号 200141923

指导教师 马燕

**完 成 日 期 2024 年 4 月**

上海师范大学本科毕业论文（设计）  
诚信声明

本人郑重声明：所呈交的毕业论文（设计），题目《数据仓库分层架构模型的设计与搭建》是本人在指导教师的指导下，进行研究工作所取得的成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式注明。除此之外，本论文（设计）不包含任何其他个人或集体已经发表或撰写过的作品成果。本人完全意识到本声明应承担的法律责任。

作者签名： 

日期：2024年 3 月 20 日

上海师范大学本科毕业论文（设计）选题登记表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 学生姓名 | 周文尧 | 学号 | 200141923 | 专业 | 计算机科学与技术 |
| 学 院 | 信息与机电工程学院 | 指导教师姓名**/**职称 | | 马燕 | |
| 题 目 | 数据仓库分层架构模型的设计与搭建 | | | | |
| 本选题的意义、创新性：  在大数据时代下，面对爆炸式增长的的数据，我们需要设计并且建设高效的数据模型和体系来对这些数据进行有序和有结构地分类组织和存储，避免重复建设和数据不一致性，保证数据的规范性。  普通的操作型数据库主要面向事务性处理，而数据仓库中的所有数据一般按照主题进行划分。主题是对业务数据的一种抽象，是从较高层次上对信息系统中的数据进行归纳和整理。  面向主题的数据可以划分成两部分----根据原系统业务数据的特点进行主题的抽取和确定每个主题所包含的数据内容。例如客户主题、产品主题、财务主题等；而客户主题包括客户基本信息、客户信用信息、客户资源信息等内容。分析数据仓库主题的时候，一般方法是先确定几个基本的 主题，然后再将范围扩大，最后再逐步求精  面向操作型的数据库通常是异构的、并且相互独立，所以无法对信息进行概括和反映信息的本质。而数据仓库中的数据是经过数据的抽取、清洗、切换、加载得到的，所以为了保证数据不存在二义性，必须对数据进行编码统一和必要的汇总，以保证数据仓库内数据的一致性。数据仓库在经历数据集成阶段后，使数据仓库中的数据都遵守统一的编码规则，并且消除许多冗余数据。  数据仓库中的数据反映的都是一段历史时期的数据内容，它的主要操作是查询、分析而不进行一般意义上的更新（数据集成前的操作型数据库主要完成数据的增加、修改、删除、查询），一旦某个数据进入到数据仓库后，一般情况下数据会被长期保留，当超过规定的期限才会被删除。通常数据仓库需要做的工作就是加载、查询和分析，一般不进行任何修改操作，是为了企业高层 人员决策分析之用。  数据仓库不断从操作型数据库或其他数据源获取变化的数据，从而分析和预测需要的历史数据，所以一般数据仓库中数据表的键码（维度）都含有时间键，以表明数据的历史时期信息，然后不断增加新的数据内容。通过这些历史信息可以对企业的发展历程和趋势做出分析和预测。数据仓库的建设需要大量的业务数据作为积累，并将这些宝贵的历史信息经过加工、整理，最后提供给决策分析人员，这是数据仓库建设的根本目的。 | | | | | |
| 国内、外研究现状：  随着各种计算机技术，如数据模型、数据库技术和应用开发技术的不断进步，数据仓库技术也不断发展，并在实际应用中发挥了巨大的作用。IDC在1996年的一次对90年代前期进行的62个数据仓库项目的调查结果表明:进行数据仓库项目开发的公司在平均2.73年的时间内获得了平均为321%的投资回报率。使用数据仓库所产生的巨大效益同时又刺激了对数据仓库技术的需求，数据仓库市场正以迅猛势头向前发展:一方面，数据仓库市场需求量越来越大，每年约以400%的速度扩张;另一方面，数据仓库产品越来越成熟，生产数据仓库工具的厂家也越来越多。数据仓库技术及市场将向以下方向发展:   1、并行化和可扩展性  为提高数据仓库的性能和可扩展能力，数据仓库已趋向并行化。在硬件层次上，已越来越明显地采用多处理器并行结构;在数据库层次上，许多数据库厂商已推出并行产品，以适应数据仓库市场的需要。   2、集中化 数据仓库项目将越来越大，GartnerGroup预测:到2000年，约有70%的集中化信息管理将依赖于数据仓库市场。   3、数据仓库与Internet/Intranet的集成 随着Internet/Intranet技术的广泛应用和发展，数据仓库将Internet/Intranet进行很好的集成，即前台是Web服务器，后台是数据仓库系统。   4、数据挖掘工具的成熟和广泛使用 数据挖掘工具和人工智能代理将是以后5年推动决策支持演变过程的主要力量。   5、通用数据库 数据仓库将支持多媒体、支持结构化和非结构化数据，即向通用数据库发展，具有面向对象的能力。   6、数据仓库打包应用 数据仓库将集成一些工具和应用，打包推向用户。  国内外常见架构：较为常见的架构为lambda架构和kappa架构。  Lambda 架构的核心理念是“流批一体”，数据从底层的数据源开始，经过多种格式后进入大数据平台，被 Kafka、Flume 等数据组件收集后，再分成两条线进行计算：一条线是进入流式计算平台（例如 Storm、Flink或者Spark Streaming），去计算实时的一些指标，保证数据实时性；另一条线则进入批量数据处理离线计算平台（例如 Mapreduce、Hive，Spark SQL），去计算 T+1 的相关业务指标，这些指标需要隔日才能看见。  Lambda架构的缺陷：   * 同样的需求需要开发两套一样的代码； * 数据过多导致服务器存储大； * 集群资源使用增多； * 离线结果和实时结果不一致； * 批量计算T+1可能计算不完。   Kappa 架构的核心思想是通过改进流计算系统来解决数据全量处理的问题，使得实时计算和批处理过程使用同一套代码。而流计算天然的分布式特征，注定了他的扩展性更好，通过加大流计算的并发性，加大流式数据的“时间窗口”，来统一批处理与流式处理两种计算模式。  Kappa 架构的消息队列常选择 Kafka，因为它具有历史数据保存、重放的功能，并支持多消费者同时使用。而流处理集群一般选择 Flink ，因为 Flink 支持流批一体的处理方式，并且对 SQL 的支持率逐渐提高，因此可以减少流处理和批处理逻辑代码不一致的情况。  Kappa架构缺陷：   * Kafka 无法支持海量数据存储。 * Kafka 无法支持高效的 OLAP 查询， * 无法复用目前已经非常成熟的基于离线数仓的数据血缘、数据质量管理体系。 * Kafka 不支持 update / upsert ，目前 Kafka 仅支持 append 。 | | | | | |
| 研究内容及文章目录（至少二级标题）：  研究内容是基于现在最主流的lambda与kappa架构，设计一种混合架构方案，以满足特定条件下的公司数据仓库分层模型的底层设计与上层的搭建，将lambda架构需要维护两套逻辑，导致的维护复杂，以及kappa架构流式重新处理慢的弊端缩减，使得公司在使用该数据仓库，应对大数据的情况下，会有更好的性能体现  目录  [上海师范大学本科毕业论文（设计） 诚信声明 2](#_Toc161832929)  [上海师范大学本科毕业论文（设计）选题登记表 3](#_Toc161832930)  [摘 要 8](#_Toc161832931)  [目录 10](#_Toc161832932)  [1.绪论 11](#_Toc161832933)  [1.1 研究背景 12](#_Toc161832934)  [1.2 国内、外研究现状 12](#_Toc161832935)  [1.3. 国内外常见的数据仓库架构以及优缺点 13](#_Toc161832936)  [1.4 本文主要研究内容 14](#_Toc161832937)  [2相关知识介绍 15](#_Toc161832938)  [2.1 lambda架构 15](#_Toc161832939)  [2.2 kappa架构 16](#_Toc161832940)  [2.3 基于lambda架构与kappa架构的混合架构 17](#_Toc161832941)  [2.4 数据仓库的设计与建模 18](#_Toc161832942)  [2.5 SQLite数据库 19](#_Toc161832943)  [2.6 flask框架 20](#_Toc161832944)  [2.7 python 20](#_Toc161832945)  [2.8 Hadoop 20](#_Toc161832946)  [3.知了二手书交易平台 21](#_Toc161832947)  [3.1 项目背景 21](#_Toc161832948)  [3.2 项目目标 21](#_Toc161832949)  [3.2 项目功能需求 22](#_Toc161832950)  [3.3 用例图 22](#_Toc161832951)  [3.3 E-R图 23](#_Toc161832952)  [3.4数据库表结构 24](#_Toc161832953)  [4.系统功能设计与搭建 25](#_Toc161832954)  [4.1 系统总体功能设计 25](#_Toc161832955)  [4.2 数据仓库搭建 26](#_Toc161832956)  [4.3 前后端连接 29](#_Toc161832957)  [5.系统实现 30](#_Toc161832958)  [5.1 用户注册登录 31](#_Toc161832959)  [5.2 用户资料修改 31](#_Toc161832960)  [5.3 图书列表（首页） 32](#_Toc161832961)  [5.4 图书详情 32](#_Toc161832962)  [5.5 提交订单 32](#_Toc161832963)  [5.6 我的订单 33](#_Toc161832964)  [5.7 发布商品 33](#_Toc161832965)  [6.系统测试 34](#_Toc161832966)  [6.1测试目的以及方法 34](#_Toc161832967)  [6.2黑盒——错误推测法测试过程以及结果 35](#_Toc161832968)  [6.3白盒——提交规范测试 35](#_Toc161832969)  [7.总结与展望 36](#_Toc161832970)  [8.参考文献 37](#_Toc161832971) | | | | | |
| 研究方法、手段及研究工具：  调查法，探索性研究法，实验法。  通过查找相关文献资料，依据已经学习过的知识，进行探索性的分析与实践。  研究工具：python，SQLite等等 | | | | | |
| 主要参考文献：  [1] [校园二手书交易平台的可行性研究及构建策略](https://www.zhangqiaokeyan.com/academic-journal-cn_science-education-article-collects_thesis/0201222402548.html) [J] [. 熊鹏](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=%E7%86%8A%E9%B9%8F&option=202) [,李海杰](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=%E6%9D%8E%E6%B5%B7%E6%9D%B0&option=202) [. 科教文汇](https://www.zhangqiaokeyan.com/journal-cn-3331/). 2015,第30期  [2]赵京鹤,郭立,商秋平.企业生产管理数据仓库构建[J].中国自动识别技术,2023,(02):44-46.  [3] [Building Data Warehouses in the Era of Big Data: An Approach for Scalable and Flexible Big Data Warehouses](https://www.zhangqiaokeyan.com/academic-conference-foreign_meeting-208151_thesis/020514306690.html) [C] . [Carlos Costa,](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=Carlos%20Costa&option=202) [Maribel Yasmina Santos](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=Maribel%20Yasmina%20Santos&option=202)[International conference on advanced information systems engineering .](https://www.zhangqiaokeyan.com/conference-foreign-208151/) 2019  [4]宋传园.数据仓库的概念与技术分析[J].信息记录材料,2023,24(05):65-67.DOI:10.16009/j.cnki.cn13-1295/tq.2023.05.013.  [5]刘晓莉,李满,熊超等.基于Hadoop搭建高可用数据仓库的研究和实现[J].现代信息技术,2023,7(01):99-101.DOI:10.19850/j.cnki.2096-4706.2023.01.027. | | | | | |
| 时间进度：  2023年10月—2023年12月 完成选题和填写“选题登记表”  2023年12月—2024年02月 搜索资料，设计模型  2024年02月--2024年03月 搭建数据仓库模型，完成初稿  2024年04月 论文定稿  学生 （签名） 年 月 日 | | | | | |
| 指导教师意见：  指导教师 （签名） 年 月 日 | | | | | |
| 专业负责人意见：  专业负责人 （签名） 年 月 日 | | | | | |

注：本表与毕业论文（设计）一起存档，保存期为四年。

摘 要

随着大数据技术的快速发展，数据处理和分析在各行各业中扮演着愈发重要的角色。Lambda架构和Kappa架构作为两种主流的大数据处理架构，各自具有独特的优势。Lambda架构通过结合批处理和实时处理，实现了数据的稳定性和实时性的平衡；而Kappa架构则强调以流处理为核心，简化了数据处理流程，提高了处理效率。

在本文中，我们提出了一种基于Lambda与Kappa混合架构的设计方案，并将其应用于二手书交易平台的构建。该混合架构充分融合了Lambda架构的稳定性和Kappa架构的实时性，既能够处理大量的历史数据，又能够实时响应用户的行为变化。

具体有以下几方面工作：

1. Lambda架构与Kappa架构的分析
2. 混合架构的设计
3. 数据流处理
4. 数据仓库搭建
5. 二手书交易平台构建

综上所述，基于Lambda与Kappa混合架构的二手书交易平台设计与建模，实现了数据的实时采集、处理、分析和应用，为用户提供了更加便捷、智能的购书体验。该架构的设计与实践不仅为二手书交易平台提供了有力的技术支持，也为类似的大数据处理和实时分析场景提供了有益的参考和借鉴。通过深入研究和应用这两种架构的优势，我们能够更好地应对复杂多变的数据处理需求，推动数据驱动的业务发展。

关键词：大数据、数据仓库、Lambda、Kappa、二手书交易平台

**Abstract**

With the rapid development of big data technology, data processing and analysis have played an increasingly important role in various industries. Lambda architecture and Kappa architecture, as two mainstream big data processing architectures, each have unique advantages. Lambda architecture achieves a balance between data stability and real-time processing by combining batch processing with real-time processing, while Kappa architecture emphasizes stream processing as its core, simplifying the data processing flow and improving processing efficiency.

In this paper, we propose a design scheme based on a hybrid architecture of Lambda and Kappa and apply it to the construction of a second-hand book trading platform. This hybrid architecture fully integrates the stability of the Lambda architecture with the real-time nature of the Kappa architecture, enabling it to process large amounts of historical data while also responding in real-time to changes in user behavior.

Specifically, the following aspects of work are included:

1. Analysis of Lambda architecture and Kappa architecture
2. Design of the hybrid architecture
3. Data stream processing
4. Construction of the data warehouse
5. Construction of the second-hand book trading platform

In summary, the design and modeling of the second-hand book trading platform based on the hybrid architecture of Lambda and Kappa enables real-time data collection, processing, analysis, and application, providing users with a more convenient and intelligent book-buying experience. The design and practice of this architecture not only provide powerful technical support for the second-hand book trading platform but also provide useful references and insights for similar big data processing and real-time analysis scenarios. Through in-depth research and application of the advantages of these two architectures, we can better cope with complex and diverse data processing needs and promote data-driven business development.

目录

[上海师范大学本科毕业论文（设计） 诚信声明 2](#_Toc161832929)

[上海师范大学本科毕业论文（设计）选题登记表 3](#_Toc161832930)

[摘 要 8](#_Toc161832931)

[目录 10](#_Toc161832932)

[1.绪论 11](#_Toc161832933)

[1.1 研究背景 12](#_Toc161832934)

[1.2 国内、外研究现状 12](#_Toc161832935)

[1.3. 国内外常见的数据仓库架构以及优缺点 13](#_Toc161832936)

[1.4 本文主要研究内容 14](#_Toc161832937)

[2相关知识介绍 15](#_Toc161832938)

[2.1 lambda架构 15](#_Toc161832939)

[2.2 kappa架构 16](#_Toc161832940)

[2.3 基于lambda架构与kappa架构的混合架构 17](#_Toc161832941)

[2.4 数据仓库的设计与建模 18](#_Toc161832942)

[2.5 SQLite数据库 19](#_Toc161832943)

[2.6 flask框架 20](#_Toc161832944)

[2.7 python 20](#_Toc161832945)

[2.8 Hadoop 20](#_Toc161832946)

[3.知了二手书交易平台 21](#_Toc161832947)

[3.1 项目背景 21](#_Toc161832948)

[3.2 项目目标 21](#_Toc161832949)

[3.2 项目功能需求 22](#_Toc161832950)

[3.3 用例图 22](#_Toc161832951)

[3.3 E-R图 23](#_Toc161832952)

[3.4数据库表结构 24](#_Toc161832953)

[4.系统功能设计与搭建 25](#_Toc161832954)

[4.1 系统总体功能设计 25](#_Toc161832955)

[4.2 数据仓库搭建 26](#_Toc161832956)

[4.3 前后端连接 29](#_Toc161832957)

[5.系统实现 30](#_Toc161832958)

[5.1 用户注册登录 31](#_Toc161832959)

[5.2 用户资料修改 31](#_Toc161832960)

[5.3 图书列表（首页） 32](#_Toc161832961)

[5.4 图书详情 32](#_Toc161832962)

[5.5 提交订单 32](#_Toc161832963)

[5.6 我的订单 33](#_Toc161832964)

[5.7 发布商品 33](#_Toc161832965)

[6.系统测试 34](#_Toc161832966)

[6.1测试目的以及方法 34](#_Toc161832967)

[6.2黑盒——错误推测法测试过程以及结果 35](#_Toc161832968)

[6.3白盒——提交规范测试 35](#_Toc161832969)

[7.总结与展望 36](#_Toc161832970)

[8.参考文献 37](#_Toc161832971)

1.绪论

1.1 研究背景

在大数据时代下

[[1]](#endnote-1)，面对爆炸式增长的的数据，我们需要设计并且建设高效的数据模型和体系来对这些数据进行有序和有结构地分类组织和存储，避免重复建设和数据不一致性，保证数据的规范性。

普通的操作型数据库主要面向事务性处理，而数据仓库中的所有数据一般按照主题进行划分。主题是对业务数据的一种抽象，是从较高层次上对信息系统中的数据进行归纳和整理。

面向主题的数据可以划分成两部分----根据原系统业务数据的特点进行主题的抽取和确定每个主题所包含的数据内容。例如客户主题、产品主题、财务主题等；而客户主题包括客户基本信息、客户信用信息、客户资源信息等内容。分析数据仓库主题的时候，一般方法是先确定几个基本的 主题，然后再将范围扩大，最后再逐步求精

面向操作型的数据库[[2]](#endnote-2)通常是异构的、并且相互独立，所以无法对信息进行概括和反映信息的本质。而数据仓库中的数据是经过数据的抽取、清洗、切换、加载得到的，所以为了保证数据不存在二义性，必须对数据进行编码统一和必要的汇总，以保证数据仓库内数据的一致性。数据仓库在经历数据集成阶段后，使数据仓库中的数据都遵守统一的编码规则，并且消除许多冗余数据。

数据仓库[[3]](#endnote-3)中的数据反映的都是一段历史时期的数据内容，它的主要操作是查询、分析而不进行一般意义上的更新（数据集成前的操作型数据库主要完成数据的增加、修改、删除、查询），一旦某个数据进入到数据仓库后，一般情况下数据会被长期保留，当超过规定的期限才会被删除。通常数据仓库需要做的工作就是加载、查询和分析，一般不进行任何修改操作，是为了企业高层 人员决策分析之用。

数据仓库不断从操作型数据库或其他数据源获取变化的数据，从而分析和预测需要的历史数据，所以一般数据仓库中数据表的键码（维度）都含有时间键，以表明数据的历史时期信息，然后不断增加新的数据内容。通过这些历史信息可以对企业的发展历程和趋势做出分析和预测。数据仓库的建设需要大量的业务数据作为积累，并将这些宝贵的历史信息经过加工、整理，最后提供给决策分析人员，这是数据仓库建设的根本目的。

1.2 国内、外研究现状

随着各种计算机技术，如数据模型、数据库技术和应用开发技术的不断进步，数据仓库技术也不断发展，并在实际应用中发挥了巨大的作用[[4]](#endnote-4)。IDC在1996年的一次对90年代前期进行的62个数据仓库项目的调查结果表明:进行数据仓库项目开发的公司在平均2.73年的时间内获得了平均为321%的投资回报率。使用数据仓库所产生的巨大效益同时又刺激了对数据仓库技术的需求，数据仓库市场正以迅猛势头向前发展:一方面，数据仓库市场需求量越来越大，每年约以400%的速度扩张;另一方面，数据仓库产品越来越成熟，生产数据仓库工具的厂家也越来越多。数据仓库技术及市场将向以下方向发展:   
1、并行化和可扩展性   
为提高数据仓库的性能和可扩展能力，数据仓库已趋向并行化。在硬件层次上，已越来越明显地采用多处理器并行结构;在数据库层次上，许多数据库厂商已推出并行产品，以适应数据仓库市场的需要。[[5]](#endnote-5)   
  
2、集中化  
数据仓库项目将越来越大，GartnerGroup预测:到2000年，约有70%的集中化信息管理将依赖于数据仓库市场。   
  
3、数据仓库与Internet/Intranet的集成  
随着Internet/Intranet技术的广泛应用和发展，数据仓库将Internet/Intranet进行很好的集成，即前台是Web服务器，后台是数据仓库系统。[[6]](#endnote-6)   
  
4、数据挖掘工具的成熟和广泛使用  
数据挖掘工具和人工智能代理将是以后5年推动决策支持演变过程的主要力量。   
  
5、通用数据库  
数据仓库将支持多媒体、支持结构化和非结构化数据，即向通用数据库发展，具有面向对象的能力。

6、数据仓库打包应用  
数据仓库将集成一些工具和应用，打包推向用户。

1.3. 国内外常见的数据仓库架构以及优缺点

国内外常见架构：较为常见的架构为lambda架构和kappa架构。

Lambda 架构的核心理念是“流批一体”[[7]](#endnote-7)，数据从底层的数据源开始，经过多种格式后进入大数据平台，被 Kafka、Flume 等数据组件收集后，再分成两条线进行计算：一条线是进入流式计算平台（例如 Storm、Flink或者Spark Streaming），去计算实时的一些指标，保证数据实时性；另一条线则进入批量数据处理离线计算平台（例如 Mapreduce、Hive，Spark SQL），去计算 T+1 的相关业务指标，这些指标需要隔日才能看见。

Lambda架构的缺陷：

* 同样的需求需要开发两套一样的代码；
* 数据过多导致服务器存储大；
* 集群资源使用增多；
* 离线结果和实时结果不一致；
* 批量计算T+1可能计算不完。

Kappa 架构的核心思想是通过改进流计算系统来解决数据全量处理的问题，使得实时计算和批处理过程使用同一套代码。而流计算天然的分布式特征，注定了他的扩展性更好，通过加大流计算的并发性，加大流式数据的“时间窗口”，来统一批处理与流式处理两种计算模式。

Kappa 架构的消息队列常选择 Kafka，因为它具有历史数据保存、重放的功能，并支持多消费者同时使用。而流处理集群一般选择 Flink ，因为 Flink 支持流批一体的处理方式，并且对 SQL 的支持率逐渐提高，因此可以减少流处理和批处理逻辑代码不一致的情况。

Kappa架构缺陷：

* Kafka 无法支持海量数据存储。
* Kafka 无法支持高效的 OLAP 查询，
* 无法复用目前已经非常成熟的基于离线数仓的数据血缘、数据质量管理体系。
* Kafka 不支持 update / upsert ，目前 Kafka 仅支持 append 。

1.4 本文主要研究内容

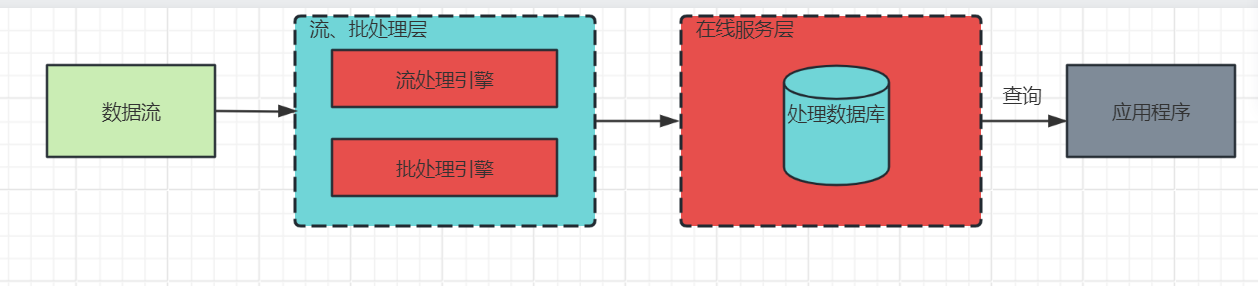
本文主要以国内外近年来数据仓库热门架构lambda框架与kappa架构为研究对象，分析其优势与劣势，扬长避短，设计一种有lambda架构与kappa架构同时存在的混合架构，使得底层数据可以根据需求和内容的不同，进入对应的数据流中，通过不同的架构进行流批处理，最终汇总在数据仓库中，供数据分析调用。而由于架构是属于底层的代码，不能很好的展示出其实际的应用，故而本文使用该混合架构设计研发一个二手书城交易平台——知了，底层数据库采用sqlite，使用flask框架连接前端与后端。

2相关知识介绍

2.1 lambda架构

Lambda架构是一种广泛应用于大数据处理领域的综合性解决方案，它融合了批处理和实时处理两种数据处理模式，旨在满足企业对数据处理的多样化需求。[[8]](#endnote-8)

Lambda架构如下图：



Lambda架构的核心思想是将数据处理流程划分为三个层次：批处理层、实时处理层以及服务层。批处理层专注于对历史数据的离线计算，利用分布式计算框架如Apache Spark，对海量数据进行深度分析和挖掘，生成全局一致的视图。实时处理层则聚焦于新产生的数据流，通过流式处理框架如Apache Flink，实时捕获、处理并响应数据变化，确保企业能够迅速响应市场动态和用户需求。

服务层作为Lambda架构的桥梁，负责将批处理层和实时处理层的结果进行合并和整合，为最终用户提供一致、准确的数据服务。通过数据整合、过滤和转换，服务层确保用户能够获取到满足业务需求的数据视图。

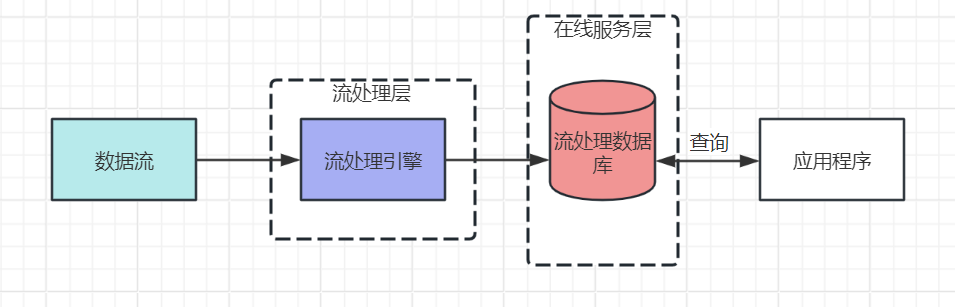
Lambda架构的优势在于其灵活性和可扩展性。它可以根据企业的实际需求，灵活调整批处理层和实时处理层的资源配置和处理能力，以适应不同规模的数据量和处理速度。此外，Lambda架构还具有良好的容错性和可维护性，能够确保数据处理的稳定性和可靠性。

Lambda架构存在的缺陷：

* 同样的需求需要开发两套一样的代码；
* 数据过多导致服务器存储大；
* 集群资源使用增多；
* 离线结果和实时结果不一致；
* 批量计算T+1可能计算不完。

2.2 kappa架构

Kappa架构是一种简化且高效的流处理架构，它打破了传统数据处理中批处理和流处理的界限，将两者统一起来，以流处理的方式处理所有数据。[[9]](#endnote-9)



在Kappa架构中，不再单独设置批处理层来处理历史数据，而是将所有数据都视为流数据，并通过实时流处理框架（如Apache Flink或Kafka Streams）进行处理。这种设计使得Kappa架构更加简洁和高效，减少了系统的复杂性和维护成本。

Kappa架构的核心思想是将所有数据流化，包括历史数据和实时数据。通过流处理框架，Kappa架构可以实时捕获、处理和响应数据流，并提供低延迟的数据处理能力。这种架构使得企业能够实时获取数据洞察，迅速响应市场变化和用户需求。

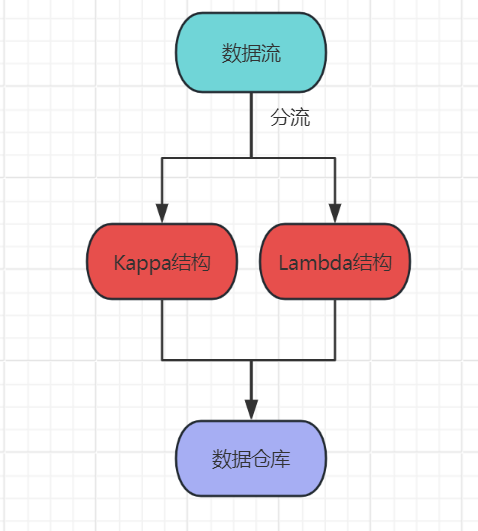
与传统Lambda架构相比，Kappa架构避免了批处理和实时处理之间的数据合并问题，简化了数据处理流程。它仅依赖于实时流处理框架，从而降低了系统的复杂性和出错的可能性。此外，Kappa架构还具备更好的扩展性和容错性，能够适应不同规模的数据量和处理需求。[[10]](#endnote-10)

Kappa架构缺陷：

* Kafka 无法支持海量数据存储。
* Kafka 无法支持高效的 OLAP 查询，
* 无法复用目前已经非常成熟的基于离线数仓的数据血缘、数据质量管理体系。

2.3 基于lambda架构与kappa架构的混合架构

Lambda架构与Kappa架构混合产生的架构是一种融合了批处理与流处理的综合性数据处理方案，旨在充分发挥两者的优势，同时满足数据处理的不同需求。[[11]](#endnote-11)



在这种混合架构中，Lambda架构和Kappa架构的核心理念得到了有机结合。Lambda架构的批处理层负责处理历史数据，利用大数据处理框架进行深度分析和挖掘，提供全局一致的视图。而Kappa架构的流处理思想则贯穿于整个数据处理流程，将所有数据视为流数据，通过实时流处理框架进行捕获、处理和响应。

这种混合架构既保留了Lambda架构对数据处理的全面性和准确性，又借鉴了Kappa架构的简洁性和高效性。它能够在保证数据一致性和准确性的同时，提供低延迟的数据处理能力，使我们能够实时响应市场变化和用户需求。

此外，混合架构还具备更好的灵活性和可扩展性。根据业务需求和数据量的变化，我们可以灵活调整批处理层和流处理层的资源配置和处理能力，以适应不同的处理场景。同时，混合架构也支持多种数据处理框架的集成，使得我们能够根据实际情况选择最适合的工具和技术。

在特定的环境中，我们应采用特定的方式来设计混合模式，对于本篇论文来说，设计混合模式的难点在于如何合理的分配数据流进入lambda架构与kappa架构，如何保持数据的一致性成为难点，这点在后文中会提出解决方案

2.4 数据仓库的设计与建模[[12]](#endnote-12)

数据分层每个企业根据自己的业务需求可以分成不同的层次，但是最基础的分层思想，理论上数据分为三个层，数据运营层、数据仓库层和数据服务层。基于这个基础分层之上添加新的层次，来满足不同的业务需求。[[13]](#endnote-13)

2.4.1数据运营层(ODS)

  Operate data store（操作数据-存储），是最接近数据源中数据的一层，数据源中的数据，经过抽取、洗净、传输，也就说传说中的ETL之后，装入ODS层。本层的数据，总体上大多是按照源头业务系统的分类方式而分类的。例如：MySQL里面的一张表可以通过sqoop之间抽取到ODS层 ODS层数据的来源方式：

业务库

   经常会使用sqoop来抽取，比如我们每天定时抽取一次。在实时方面， 可以考虑用canal监听mysql的binlog，实时接入即可。

埋点日志

线上系统会打入各种日志，这些日志一般以文件的形式保存，我们可以选择用flume定时抽取，也可以用用spark streaming或者Flink来实时接入，当然，kafka也会是一个关键的角色。

消息队列

  来自ActiveMQ、Kafka的数据等

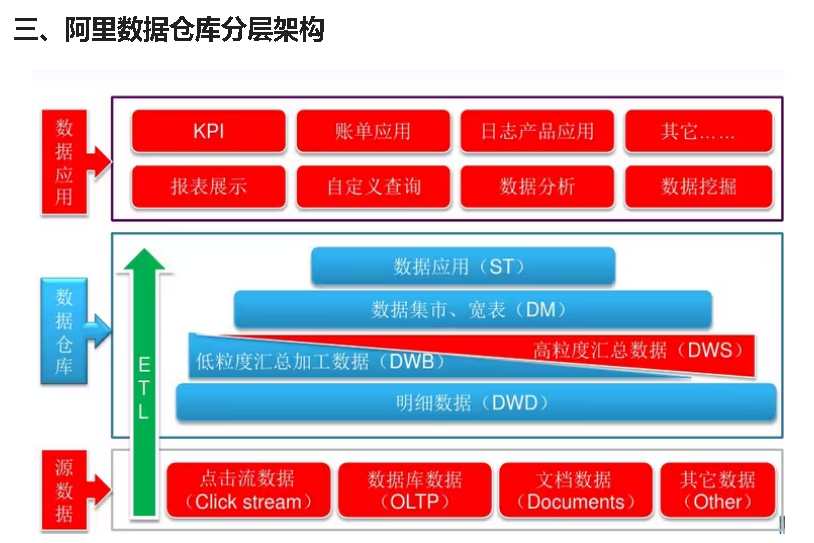
2.4.2数据仓库层（DW）

Data warehouse（数据仓库）。在这里，从ODS层中获得的数据按照主题建立各种数据模型。例如以研究人的旅游消费为主题的数据集中，便可以结合航空公司的登机出行信息，以及银联系统的刷卡记录，进行结合分析，产生数据集。在这里，我们需要了解四个概念：维（dimension）、事实（Fact）、指标（Index）和粒度（ Granularity）。

2.4.3数据服务层/应用层（ADS）：

Application Data Service(应用数据服务)。该层主要是提供数据产品和数据分析使用的数据，一般会存放在ES、MySQL等系统中供线上系统使用，也可能会存在Hive或者Druid中供数据分析和数据挖掘使用。例如：我们经常说的报表数据，或者说那种大宽表，一般就放在这里。

例如底下阿里数据仓库分层架构：

[[14]](#endnote-14)

2.5 SQLite数据库[[15]](#endnote-15)

SQLite数据库是一款轻量级的关系型数据库管理系统，以其小巧、快速、可靠和易用等特点而广受欢迎。

本文选用SQLite数据库，原因有四：其一，SQLite数据库小巧，可以以较小的开销搭建后台，方便数据的处理，以实现高并发的要求；其二，SQLite数据库可以支持事务的处理，由于我们是以Lambda架构与Kappa架构的混合模式进行开发，那么支持事务一致性，来保证数据不出错就成为了关键之所在。其三，SQLite数据库是无需搭建服务器的，直接以读写磁盘文件，可以实现在本机中直接运行，其四，SQLite数据库可移植性高，使得跨平台迁移数据变得简单。

2.6 flask框架

Flask的核心设计哲学是保持核心简单但易于扩展。它提供了构建Web应用所需的基本组件，如路由、模板和上下文管理，同时允许开发者根据需要添加其他功能或库。这种设计使得Flask既适合小型项目快速原型开发，也能够支持大型复杂应用的构建。[[16]](#endnote-16)

本文选取flask框架来连接前端和后端，主要目的是在于利用flask框架的高可扩展性，减少由于前后端连接而产生的问题。[[17]](#endnote-17)

2.7 python[[18]](#endnote-18)

Python 是一种解释型、面向对象、动态数据类型的高级程序设计语言。

本文选取python作为设计语言，原因有三：其一是它可以跨平台使用，兼顾windows系统，linux系统等；其二是它拥有强大的第三方库，如利用flask框架开发web程序等；其三是它拥有丰富的标准库，例如正则表达式，在处理数据上不容易出错。

2.8 Hadoop[[19]](#endnote-19)

Hadoop是一个由Apache基金会所开发的分布式系统基础架构，主要用于解决海量数据的存储和海量数据的分析计算问题。其设计初衷是为了利用普通机器组成的服务器群来分发以及处理数据，这些服务器群总计可达数千个节点，从而以可靠、经济、高效和可扩展的方式处理PB级的数据。[[20]](#endnote-20)

Hadoop实现了一个分布式文件系统（Distributed File System），其中最为核心的组件是HDFS（Hadoop Distributed File System）。HDFS具有高容错性，设计用来部署在低廉的硬件上，提供高吞吐量来访问应用程序的数据，特别适合于处理超大数据集。同时，Hadoop还包含了MapReduce框架，使得分布式计算变得简单和高效，可以在集群中并行执行任务，处理大规模数据集。

Hadoop的优势不仅在于其数据可靠性和可扩展性，还在于它可以通过简单的编程模型在不同的节点上进行大规模数据的分布式处理。用户可以在不了解分布式底层细节的情况下，开发分布式程序，从而充分利用集群的威力进行高速运算和存储。[[21]](#endnote-21)

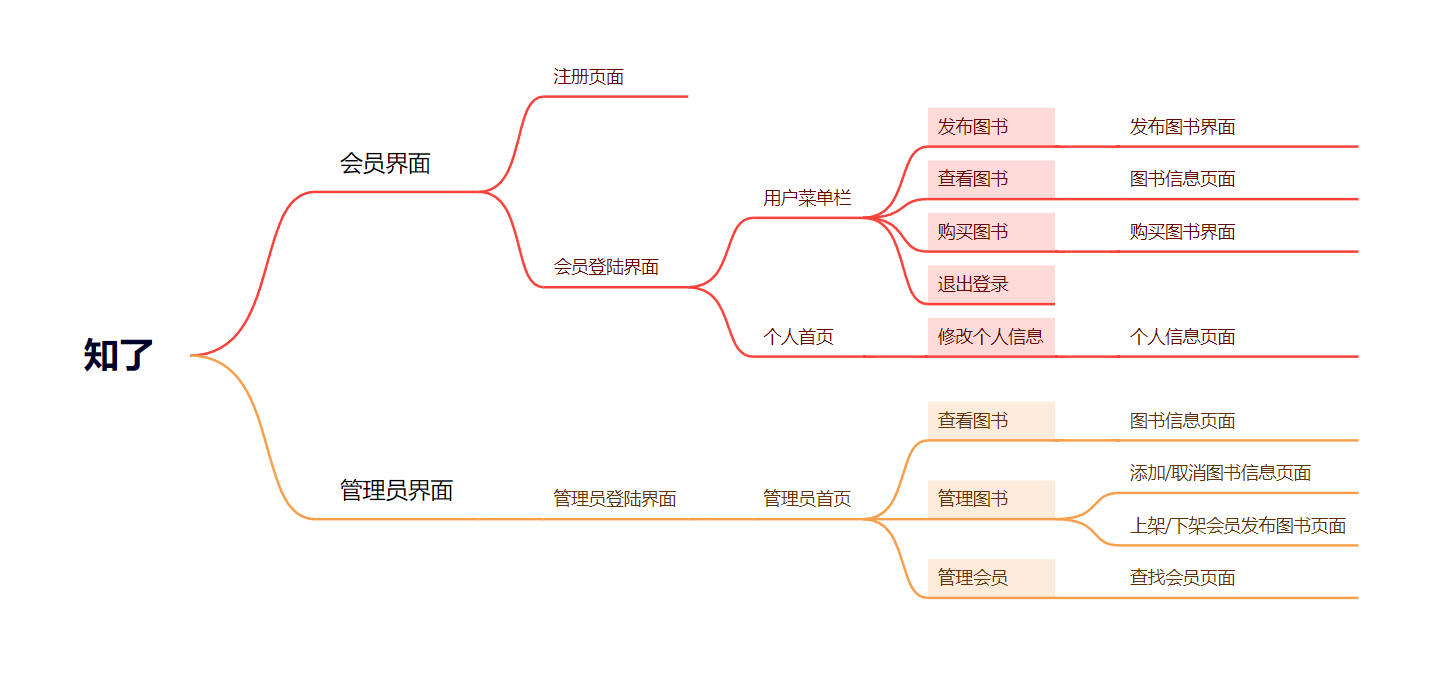
3.知了二手书交易平台

3.1 项目背景

现如今，环境问题备受关注，绿水青山就是金山银山的理念愈加强烈，创建资源节约型社会已经成为我国重要战略。[[22]](#endnote-22)然而，“废书”的问题已逐渐显现，尤其在校园中，每学期结束后，大量书籍被闲置或丢弃，或被学生几毛钱卖给收费站，资源浪费严重。另外，学生会因没有在学校订书打算自选教材、想学习本专业外的知识、想接触一下和本专业先关专业的学科、自己的课本丢失、打算考研过英语四六级计算机二三级等原因而四处找书,关系网好的人则托人借书，这样费了很大功夫也不能保证找到想要的书。而关系网不太好的学生借不到书.后往往会放弃找书，要么只能在书店里破费高昂的费用买新书。然而，这些买到的新书过一段时间便废弃不用，这样极大限度的浪费了资源。对此，建立二手书交易平台，合理利用二手书籍，具有重要意义。[[23]](#endnote-23)

3.2 项目目标

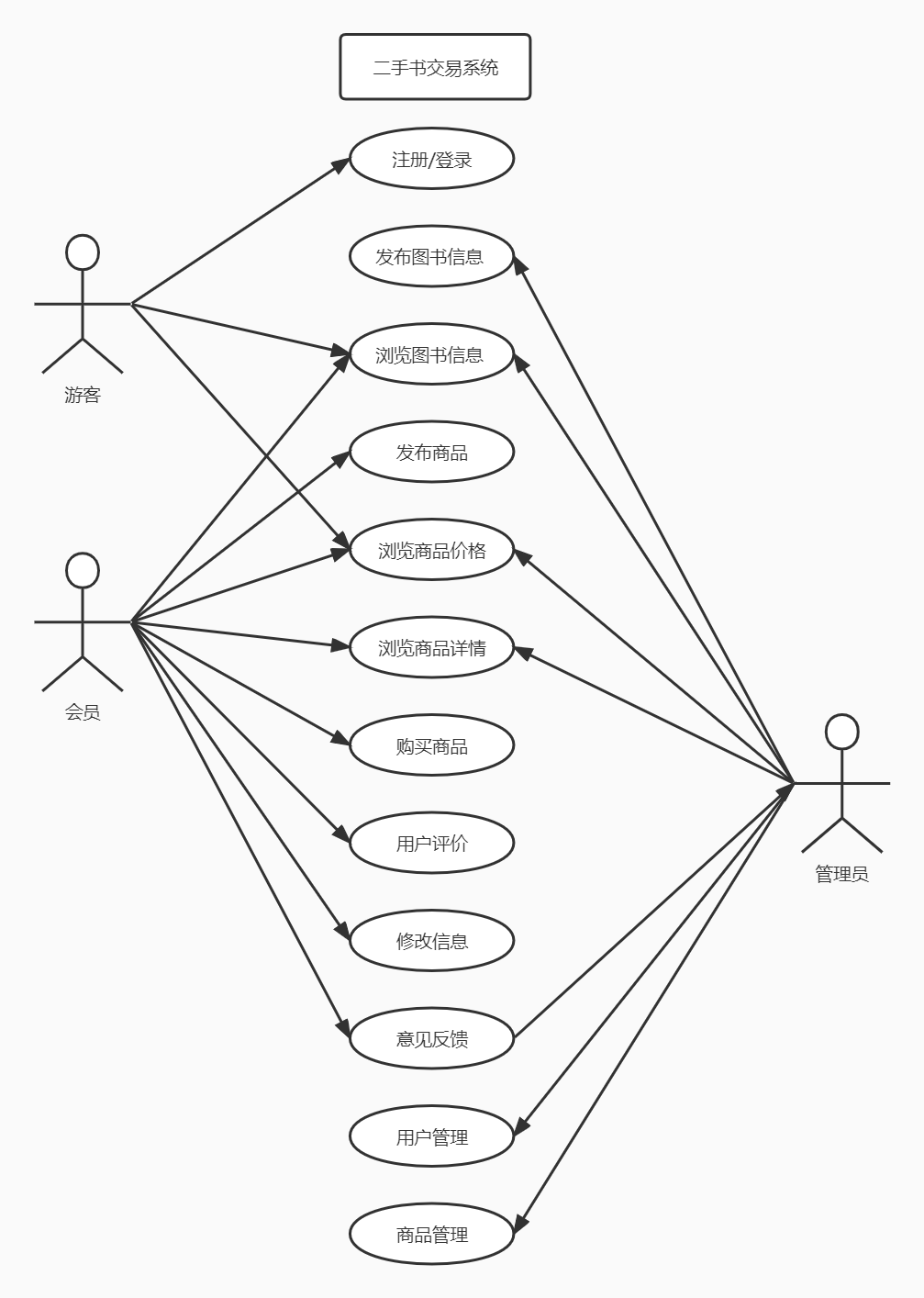
建立二手书交易系统，创建对应的数据库，以创造一个二手书在线交易平台。本系统旨在减少交易所需资源，增强交易目的性，帮助人们及时、便捷地进行二手物品交易。交易双方可在系统中进行便捷沟通。本系统还应保证书籍、需求和交易的合法性。[[24]](#endnote-24)



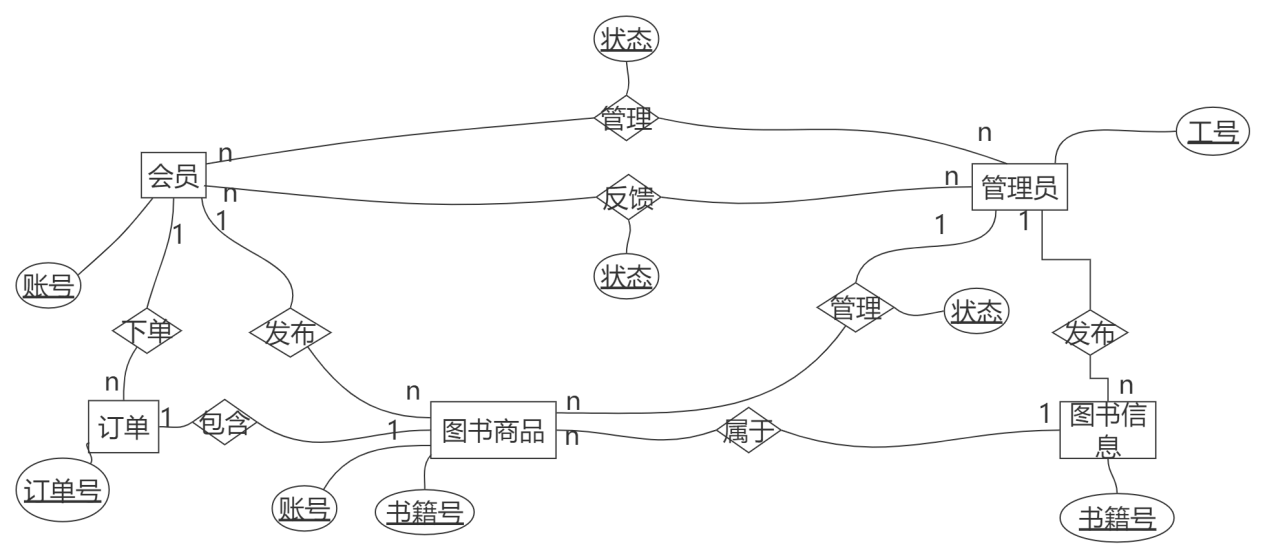
3.2 项目功能需求

|  |  |
| --- | --- |
| 游客： | 1、注册成为会员 |
| 2、可查看图书信息 |
| 3、仅可查看二手书价格 |
| 会员： | 1、基本会员资料设置 |
| 2、可查看图书信息 |
| 3、查看二手书的介绍、价格 |
| 4、选择购买/出售/求购 |
| 5、订单状态查询 |
| 6、对买/卖家进行评价 |
| 7、意见反馈 |
| 管理员： | 1、导入图书信息 |
| 2、对会员信息的管理，设置会员访问权限 |
| 3、对成交情况进行统计 |
| 4、网站维护 |

3.3 用例图

****

3.3 E-R图



3.4数据库表结构

**3.4.1会员表**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 字段名称 | Member\_Id | Member\_Name | Member\_Sex | Member\_Age | Member\_Register | Member\_Status |
| 字段含义 | 会员账号 | 会员昵称 | 会员性别 | 会员年龄 | 会员注册日期 | 会员状态 |
| 类型 | Int | varchar(20) | Char(2) | Int | datatime | Int |
| 约束 | 主码 |  |  |  |  |  |
| 样本 | 2001111 | 左日月木辛 | 男 | 20 | 2024-1-16 | 1 |

**3.4.2管理员表**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 字段名称 | Admin\_Id | Admin\_Name | Admin \_Sex | Admin \_Age | Admin \_Register | Admin \_Status |
| 字段含义 | 管理员账号 | 管理员姓名 | 管理员性别 | 管理员年龄 | 管理员创建日期 | 管理员状态 |
| 类型 | Int | varchar(20) | Char(2) | Int | datatime | Int |
| 约束 | 主码 |  |  |  |  |  |
| 样本 | 3001111 | 我不叫zwy | 男 | 20 | 2023-12-16 | 1 |

**3.4.3图书信息表**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 字段名称 | Book\_Id | Book\_Name | Book\_Publish | Book\_picture | Writer\_Name | Book\_Status |
| 字段含义 | 书籍编号 | 书名 | 出版社 | 书图片 | 作者 | 书籍状态 |
| 类型 | Int | varchar(40) | Varchar(40) | Image | Varchar(20) | Int |
| 约束 | 主码 |  |  |  |  |  |
| 样本 | 4001111 | 周老师教你10天学会C语言 | 努力学习图书出版社 | 1.jpg | 周文尧 | 1 |

**3.4.4订单表**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 字段名称 | Order\_Id | Book\_Id | ship\_Adress | Order\_date | Member\_Id | Order\_Status |
| 字段含义 | 订单编号 | 书籍号 | 送货地址 | 下单日期 | 收件人编号 | 订单状态 |
| 类型 | Int | Int | Varchar(40) | Datetime | Int | Int |
| 约束 | 主码 | 外码 |  |  | 外码 |  |
| 样本 | 5001111 | 4001111 | 上师大 | 2022-8-8 | 200111 | 1 |

**3.4.5图书商品表**

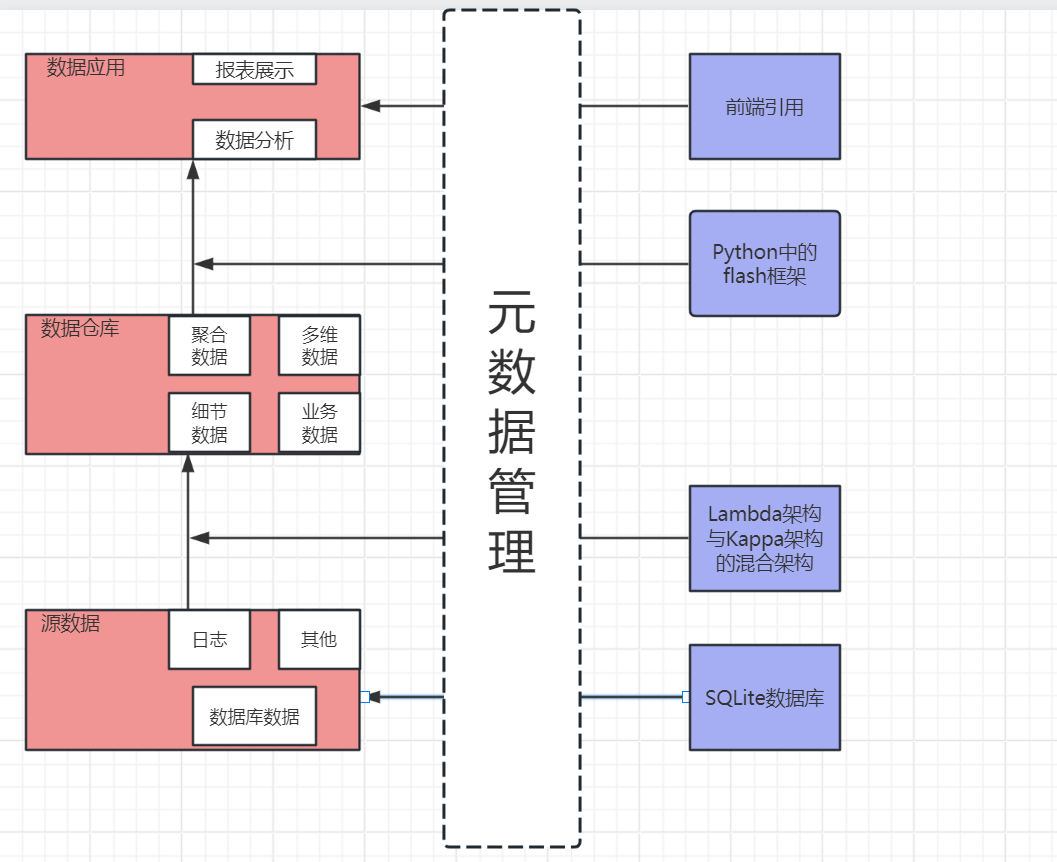
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 字段名称 | Goods\_Id | Book\_Id | Member\_Id | Book\_value | Goods\_picture | Goods\_Status |
| 字段含义 | 商品编号 | 书籍编号 | 售卖人编号 | 单价 | 商品图片 | 商品状态 |
| 类型 | Int | Int | Int | float | Image | Int |
| 约束 | 主码 | 外码 |  |  |  |  |
| 样本 | 600111 | 4001111 | 200111 | 23.98 | 1.jpg | 1 |

4.系统功能设计与搭建

4.1 系统总体功能设计

底层源数据我们将其分为日志数据与数据库数据，对于日志数据我们采取Kappa架构，对于数据库数据以及其他数据，我们都采用lambda数据。这样做是因为，日志数据的实时变更性，使得我们可以很好的使用kafka队列进行维护，而数据库数据因为有高并发的需求，我们采取lambda架构则可以更好的做到这一点。当然，在未来的开发中，可以跟随实际需求，灵活的选择对数据进行分类，然后采取kappa架构亦或者是Lambda架构。

再使用混合架构后，我们还需要将数据汇总到数据仓库中，而后由flask框架让后端数据库与前端应用进行对接，使得我们在前端的操作，可以反馈到后端。



4.2 数据仓库搭建

4.2.1 数据仓库数据分层

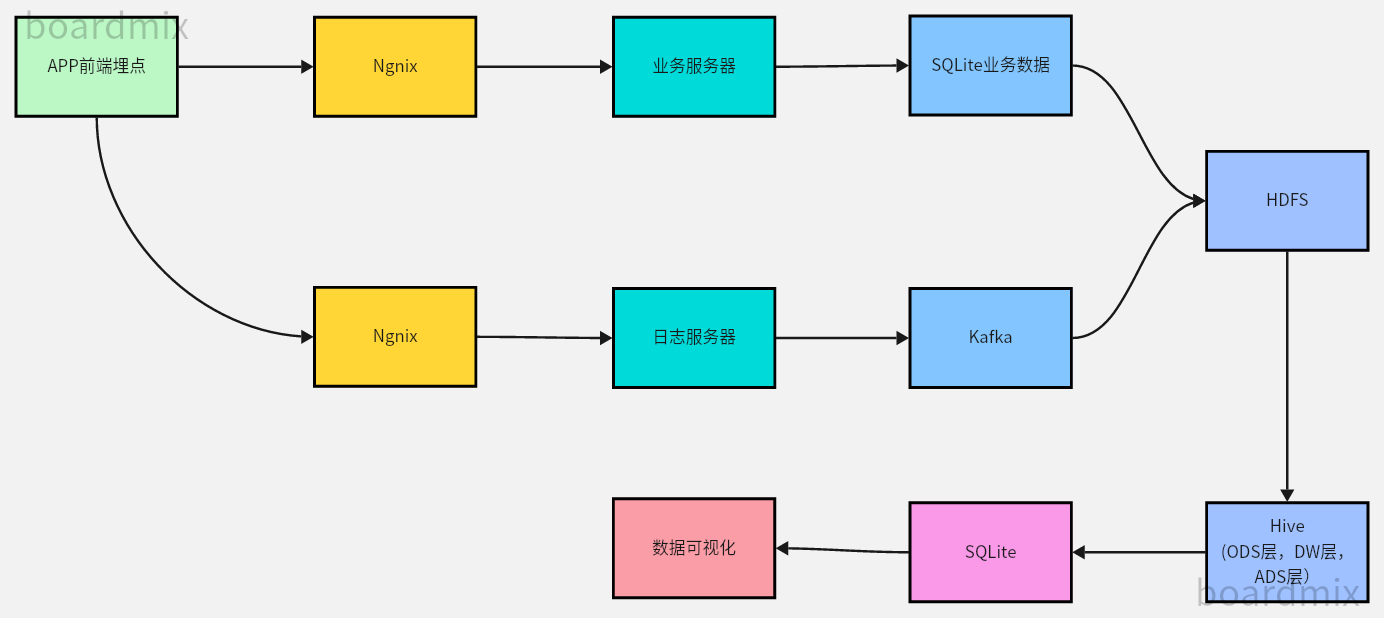
数据仓库分为三层，分别为ODS层（原始数据层），DW层（明细数据层），ADS层（数据应用层）

在ODS层中，我们存放着原始数据。直接加载原始日志，数据，数据保持原貌不做处理。

在DW层中，我们存放轻度汇总后的表单。对于ODS中的数据进行清洗（去除空值，脏数据等），然后汇总成宽表，例如：商品详情表，用户资料表

在ADS层中，我们为各种统计数据提供报表。设计粒度，存放高度汇总后的表单，例如：当日用户增加表。

4.2.2 环境配置



我们采用基于Hadoop搭建数据仓库。其中最为核心的组件是HDFS（Hadoop Distributed File System）。HDFS具有高容错性，设计用来部署在低廉的硬件上，提供高吞吐量来访问应用程序的数据，特别适合于处理超大数据集。[[25]](#endnote-25)

我们还需要用到基于Hadoop的工具hive。Hive的主要功能是将结构化的数据文件映射为Hive数据库表，并提供简单的SQL查询功能。这使得熟悉SQL的用户能够方便地查询存储在Hadoop中的大规模数据。[[26]](#endnote-26)

我们将数据分为日志数据与业务数据，将日志数据使用kappa架构，而业务数据使用lambda架构。[[27]](#endnote-27)

具体过程如下：

* 配置以linux为内核的虚拟机集群[[28]](#endnote-28)，命名为hadoop101，hadoop102，hadoop103，hadoop101中存放数据库SQLite，hadoop102中为HDFS，hadoop103中为YARN
* Hive的安装以及配置，主要需要配置好hive-env.sh文件内容，以及路径
* Hadoop集群配置，主要是在hadoop102中启动HDFS[[29]](#endnote-29)，在hadoop103中启动YARN。
* SQLite数据库配置，主要需要配置好主机，使得在任何主机上，只需要使用root用户+密码即可使用
* Hive元数据拷贝到SQLite，主要需要拷贝驱动以及配置Metastore[[30]](#endnote-30)

4.2.3 ODS层数据搭建

原始数据层的搭建，对于初始数据无需做改变

具体过程如下：

* 创建数据库
* 创建启动日志表ods\_start\_log
* 创建事件日志表ods\_event\_log
* 在hadoop102中加载数据脚本，设置脚本启用时间

4.2.4 DW层数据搭建

对ODS层数据进行清洗（去除空值，脏数据，超过极限范围的数据，行式存储改为列存储，改压缩格式）并且轻度汇总成宽表。

具体过程如下:

* 创建启动表（启动表即为宽表）
* 从ODS层中导入启动表数据
* 创建基础日志表，需要自定义UDF函数和UDTF函数
* 创建商品详情表，商品列表页表，用户资料表等，根据前面所规划内容建表即可。
* 根据表单，导入相应数据。

4.2.5 ADS层数据搭建

需根据实际需求搭建高度汇总表，这里只搭建了新增用户表

具体过程如下：

* 创建新增用户表，并且导入数据

附代码样例(ADS层数据搭建)

建表语句：

hive (gmall)>

drop table if exists ads\_new\_mem \_count;

create external table ads\_new\_mem\_count

(

`create\_date` string comment '创建时间' ,

`new\_mem\_count` BIGINT comment '新增用户数量'

) COMMENT '每日新增用户信息数量'

row format delimited fields terminated by '\t'

location '/warehouse/gmall/ads/ads\_new\_mid\_count/';

导入数据语句：

hive (gmall)>

insert into table ads\_new\_mem\_count

select

create\_date,

count(\*)

from dws\_new\_mem\_day

where create\_date='2024-03-12'

group by create\_date;

4.3 前后端连接

使用python中的flask框架，将SQLite中已经汇总好的数据仓库中的数据提取出，并且前端已经开好的端口，将数据传递到前端中。[[31]](#endnote-31)

具体过程：

* 编写前端文件zhiliao\_api
* 编写后端文件zhiliao\_client
* 编写启动脚本start\_bat
* 使用flask框架连接前后端内容，写成文件main.py

附代码：

import logging

from flask import Flask, render\_template, request

# from flask\_cors import CORS

from api import api\_blueprint

app = Flask(\_\_name\_\_, static\_folder="./root/static", template\_folder="./root")

app.register\_blueprint(api\_blueprint)

# CORS(app, supports\_credentials=True)

@app.route('/')

def root():

return render\_template("index.html")

if \_\_name\_\_ == '\_\_main\_\_':

app.logger.setLevel(logging.DEBUG)

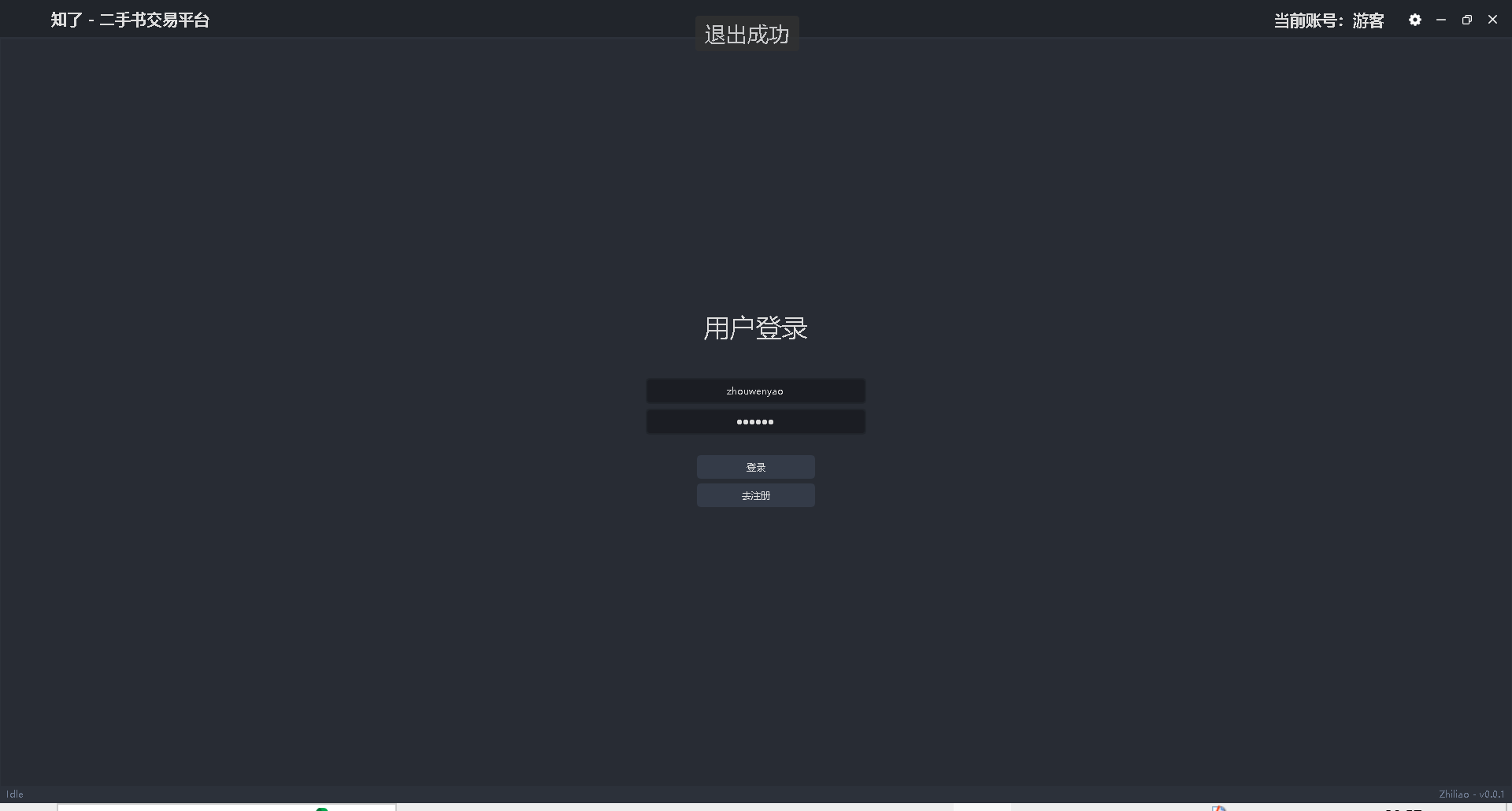
FORMAT = "%(message)s"

app.run(debug=False, host='0.0.0.0', port=23818)

5.系统实现

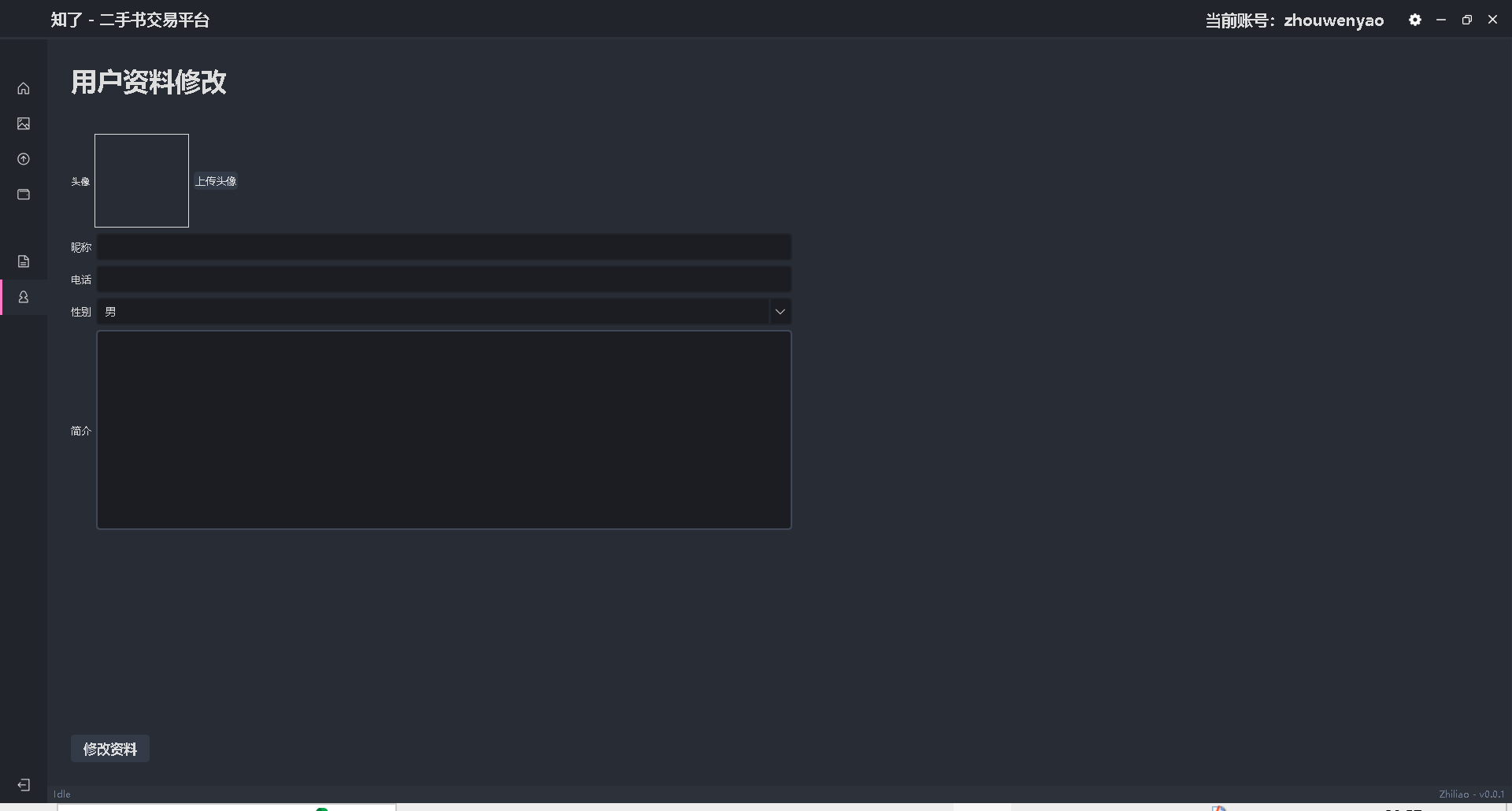
5.1 用户注册登录

在登录界面上，可以看到两个主要的输入框，分别是用户名和密码框。用户需要在此处输入正确的个人账号信息和密码，以验证身份并获取访问权限。为了增强安全性，登录界面还采用了加密技术，保护用户的密码信息不被泄露。如果未存在账号，则会提醒是否去注册账号。未输入用户名以及刚进入平台，为默认游客账号。



5.2 用户资料修改

在用户资料修改界面上，可以修改头像、昵称、电话、性别等信息。



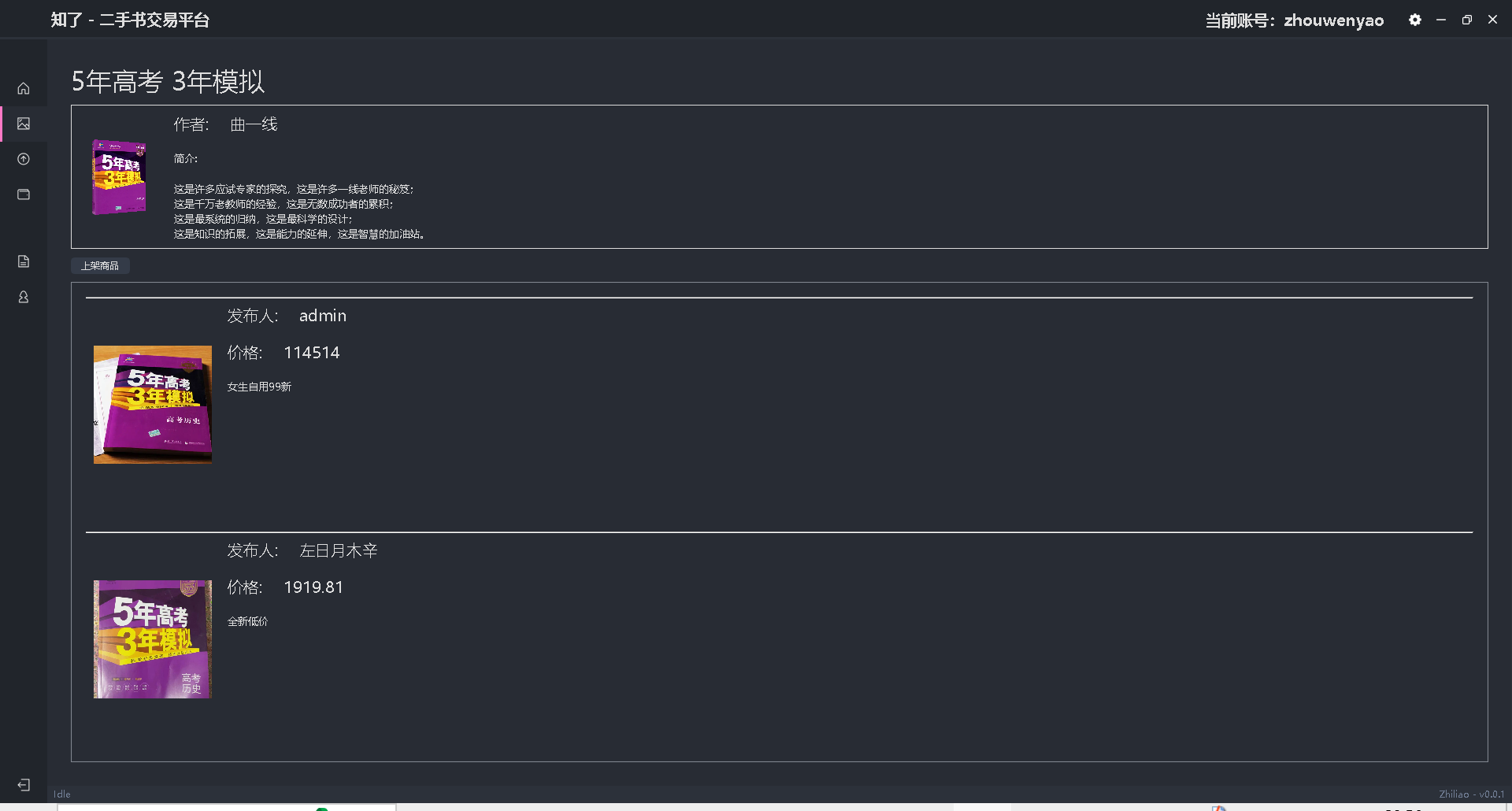
5.3 图书列表（首页）

在图书列表（首页）中，可以查看到图书信息，作者，简介等信息。



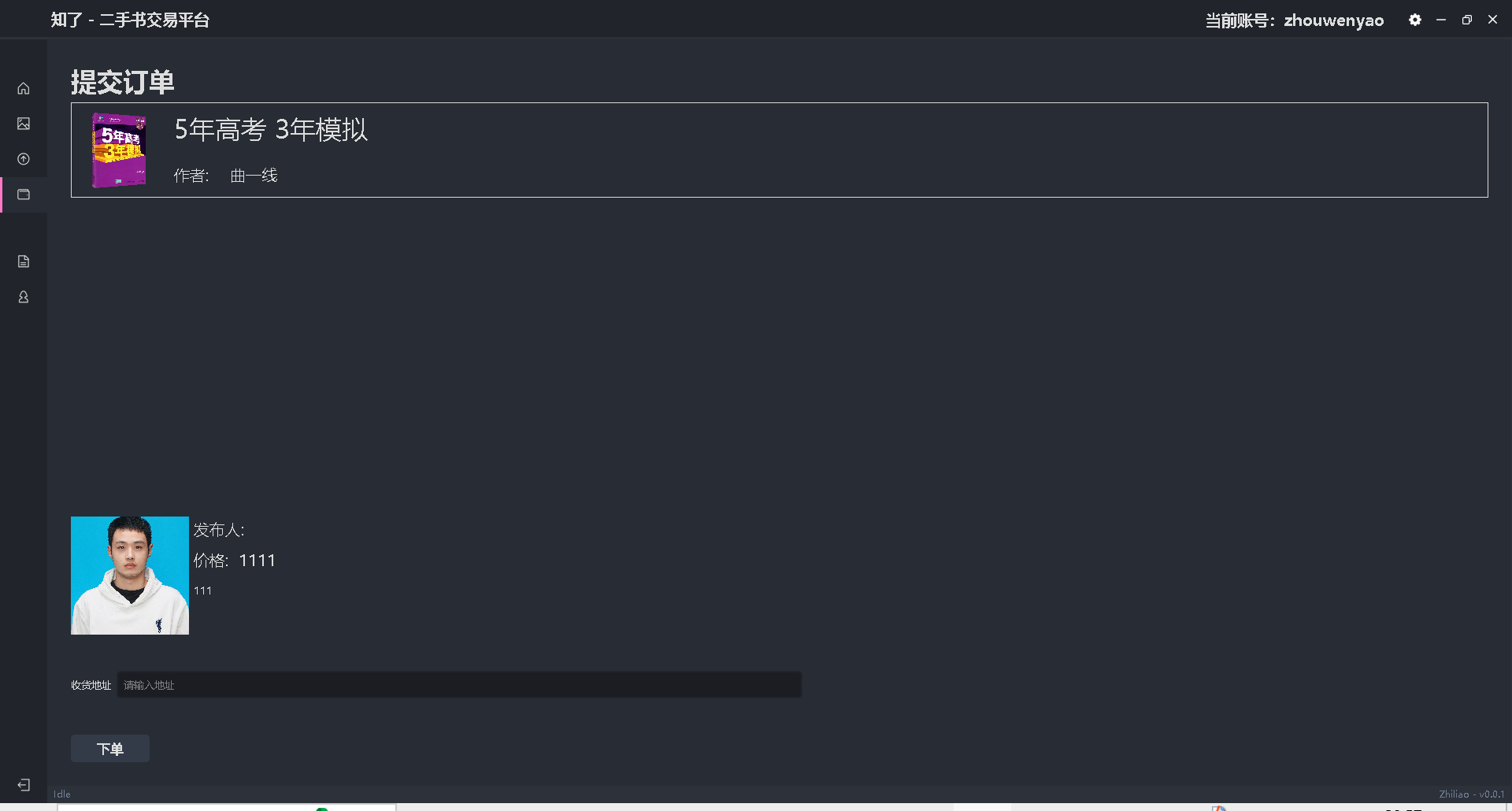
5.4 图书详情

在图书列表界面中点击图书可以进入图书详情界面，在此界面中可以看到这本书更为详细的介绍，以及有多少人在卖这本书，以及价格，便于选购。



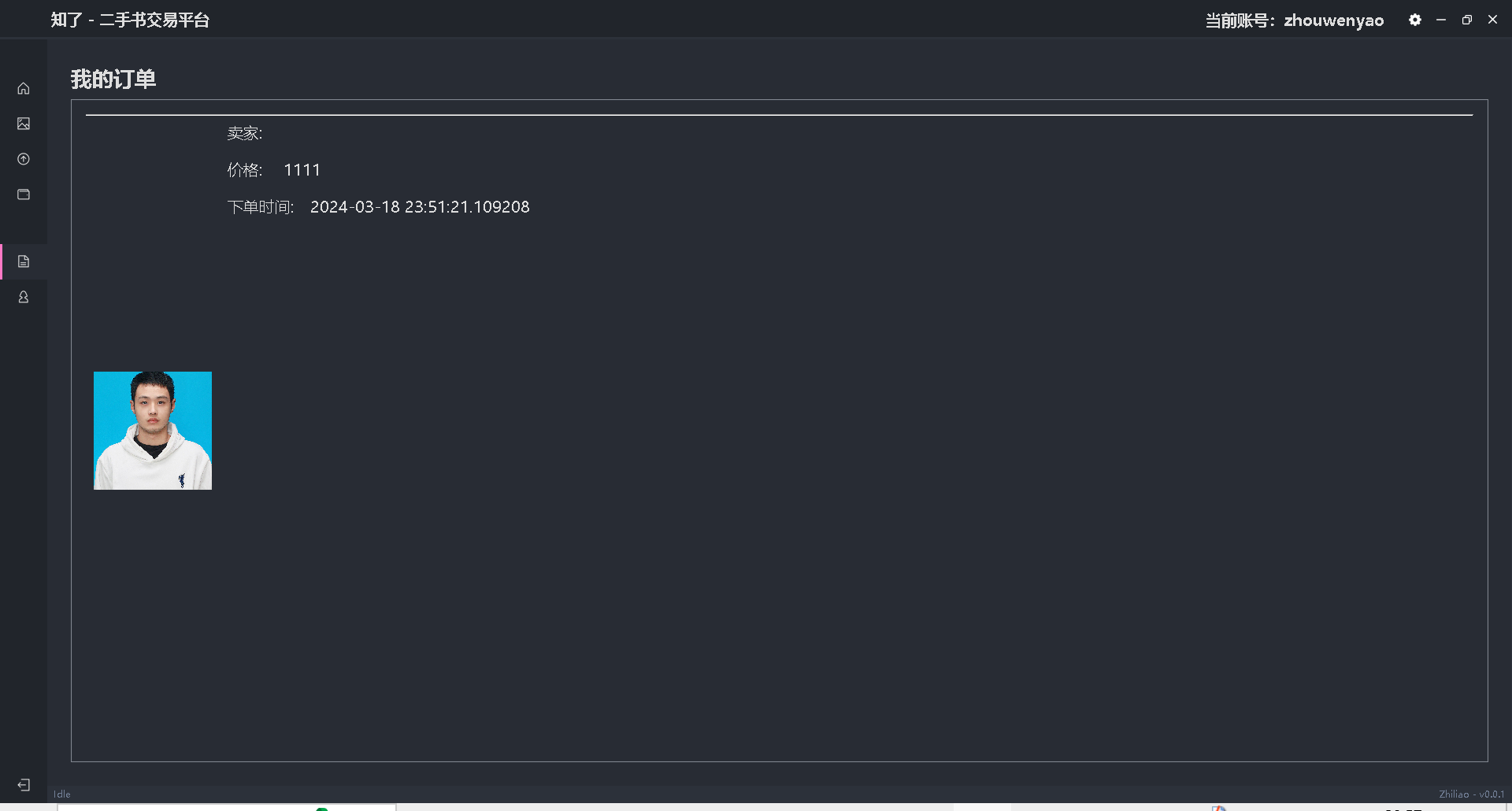
5.5 提交订单

在图书列表中选择一本想买的二手书，可进入订单界面，在提交订单界面，需要填入收货地址，方可下单成功。（暂未实现付款功能）



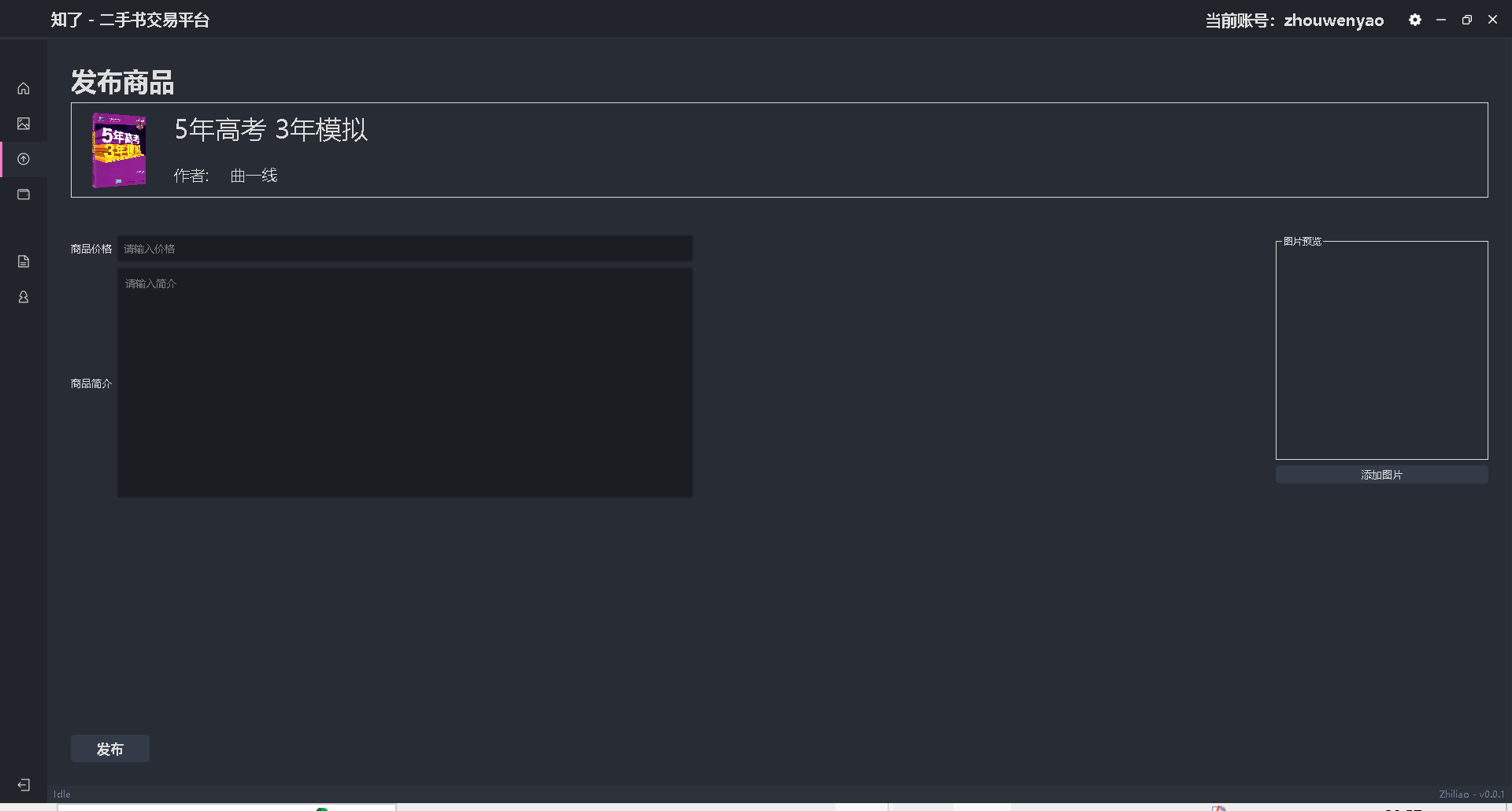
5.6 我的订单

下单完成后，在我的订单界面中，可以查看到本次订单的内容。



5.7 发布商品

在发布商品界面，可以输入价格、作者、简介等，最后必须拍照上传图片后方可发布。



6.系统测试

6.1测试目的以及方法

系统测试的主要目的在于全面验证软件系统的功能和性能，以确保它们能够满足预先设定的要求，并且在实际的运行环境中表现出色。在这个过程中，我们的核心关注点不仅仅是软件能否运行，更重要的是软件如何运行，以及它在各种条件和情境下的响应和稳定性。

首先，从功能角度来看，系统测试能够验证系统是否按照设计和规格说明书实现了所有的功能点。这包括但不仅限于用户界面的一致性、业务流程的准确性、数据处理和存储的正确性等。

其次，从性能角度出发，系统测试将测量系统的响应时间、吞吐量、错误率等关键指标，以判断系统是否能在高负载或极端情况下保持稳定的运行状态。这有助于预防系统在实际应用中可能出现的性能瓶颈或崩溃风险。

在本文中，我们更偏重于功能角度，去测试我们数据集的一致性、数据处理的准确性等。

我们首先采用黑盒测试的方式进行测试[[32]](#endnote-32)，使用错误推测法进行测试系统可能存在的漏洞。

最后，从维护角度触发，系统测试，需要测试代码提交规范，编码规范等。方便日后管理者对其维护和更新。

所以，我们采用白盒测试的方法进行测试，去测试系统中可能存在的代码不规范问题

6.2黑盒——错误推测法测试过程以及结果

|  |  |  |  |
| --- | --- | --- | --- |
| **测试用例** | **步骤** | **预期结果** | **实际结果** |
| 用户输入商品信息的部分信息以查询相关的商品信息。 | 在商品信息界面输入的商品信息与数据库中的商品信息表记录不匹配 | 搜索失败 | 搜索失败 |
| 用户输入商品信息的相关信息以及上传图片以完成商品信信息的添加 | 在发布商品界面输入的商品信息与数据库中的商品信息表记录发生冲突 | 添加失败 | 添加失败 |
| 用户输入订单信息的相关信息以下单 | 在订单界面中未填入收货地址，而点击下单 | 下单失败 | 下单失败 |
| 用户修改或者填入自身资料以及上传照片以完成用户资料的上传 | 在用户资料界面输入的用户信息与数据库中的用户信息表记录发生冲突 | 修改（上传）失败 | 修改（上传）失败 |
| 同时有多个用户下单填写完信息并且下单同一个商品 | 使用虚拟机，模拟多个用户同时下单同一件商品的场景 | 第一个用户下单成功，其余失败 | 第一个用户下单成功，其余失败 |

6.3白盒——提交规范测试

6.3.1代码规范

|  |  |  |  |
| --- | --- | --- | --- |
| **编号** | **测试内容** | **预期结果** | **实际结果** |
| 1 | 检查单行注释方式是否符合规范 | 单行注释以#开头 | 单行注释以#开头 |
| 2 | 检查多行注释方式是否符合规范 | 多行以块注释方式进行注释 | 多行以块注释方式进行注释 |
| 3 | 检查文档注释方式是否符合规范 | 文档注释以【‘’’文档当注释’’’】的方式进行注释 | 文档注释以【‘’’文档当注释’’’】的方式进行注释 |
| 4 | 检查项目代码上的每个类是否有详细注释 | 有 | 有 |
| 5 | 检查项目代码上的每个方法是否有详细注释 | 有 | 有 |
| 6 | 检查项目代码上的每个变量是否有详细注释 | 有 | 有 |
| 7 | 当进行修改代码后，检查注释是否有进行了相应的修改 | 有 | 有 |
| 8 | 检查注释内容是否通俗易懂 | 是 | 是 |
| 9 | 检查循环或逻辑分支代码结构上是否有详细注释 | 有 | 有 |
| 10 | 检查代码是否有多余无用的注释 | 无 | 无 |

6.3.2 命名以及编码规范

|  |  |  |  |
| --- | --- | --- | --- |
| 编号 | 测试内容 | 预期结果 | 实际结果 |
| 1 | 检查代码的类、方法、变量名命名是否能够体现变量的数据类型和具体意义 | 是 | 是 |
| 2 | 检查代码的包名、模块名、局部变量名、函数名的命名是否符合规范（全小写+下划线式驼峰） | 符合 | 符合 |
| 3 | 检查代码的全局变量的命名是否符合规范（全大写+下划线式驼峰） | 符合 | 符合 |
| 4 | 检查代码的类的命名是否符合规范（首字母大写式驼峰） | 符合 | 符合 |
| 5 | 检查函数参数的命名是否符合规范（小写+下划线） | 符合 | 符合 |
| 6 | 检查代码的常量的命名是否符合规范（全大写+下划线式驼峰） | 符合 | 符合 |

7.总结与展望

在信息化时代，数据仓库设计与建模作为数据处理与分析的核心环节，已经日益凸显其重要性。尤其是在面对大规模、复杂化的数据场景时，高效、灵活的数据仓库架构更是成为了关键所在。[[33]](#endnote-33)在近年来的实践中，Lambda架构与Kappa架构的混合应用逐渐受到青睐，其在处理批处理与实时数据流方面的优势得到了广泛的认可。

Lambda架构以其批处理与实时处理并行的特点，为数据仓库提供了强大的数据处理能力。批处理层能够处理历史数据，进行深度挖掘和分析，保证了数据的全局一致性；而实时处理层则能够迅速响应新数据，满足业务对实时性的需求。然而，Lambda架构也存在一定的复杂性，需要维护两套独立的处理流程。

而Kappa架构则提出了流处理的思想，将所有数据视为流数据，简化了数据处理流程。它通过实时流处理框架进行数据的捕获、处理和响应，既降低了系统的复杂性，又提高了数据处理的效率。但是，Kappa架构在处理历史数据方面可能存在一些局限性。[[34]](#endnote-34)

因此，将Lambda架构与Kappa架构进行混合，可以充分发挥两者的优势。混合架构既能够处理历史数据，保证数据的全局一致性，又能够实时响应新数据，满足业务的实时性需求。同时，通过合理的架构设计和技术选择，可以降低系统的复杂性，提高数据处理的效率和稳定性。

具体到二手书交易平台的应用场景中，混合架构的数据仓库设计与建模能够支持平台对大量交易数据、用户行为数据等进行深度分析和挖掘。通过对这些数据的处理和分析，平台可以更好地了解用户需求、市场趋势以及业务运营状况，为决策提供有力支持。同时，混合架构的实时处理能力也能够帮助平台及时响应市场变化，优化用户体验，提升竞争力。

展望未来，随着技术的不断发展和业务需求的不断变化，数据仓库设计与建模将面临更多的挑战和机遇。[[35]](#endnote-35)一方面，需要继续探索更加高效、灵活的数据处理架构和技术；另一方面，也需要结合具体业务场景，进行更加深入的数据分析和挖掘，为业务发展提供更有价值的支持。同时，还需要关注数据安全、隐私保护等方面的问题，确保数据处理的合法性和合规性。

综上所述，数据仓库设计与建模是数据处理与分析的重要环节，Lambda架构与Kappa架构的混合应用为其提供了更加高效、灵活的处理方式。在二手书交易平台等具体业务场景中，混合架构的应用将为平台提供有力支持，助力业务发展。展望未来，我们需要不断探索和创新，以更好地满足业务需求，推动数据处理的发展。

8.参考文献

1. 基于大数据的数据仓库研究现状 [J] .中国技术新技术新产品 . [罗巍,](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=%E7%BD%97%E5%B7%8D&option=202)[刘功总;](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=%E5%88%98%E5%8A%9F%E6%80%BB&option=202)2020.17期 [↑](#endnote-ref-1)
2. [大数据时代传统关系数据库与NoSQL数据库的对比与分析](https://www.zhangqiaokeyan.com/academic-journal-cn_china-computer-communication_thesis/0201291141479.html) [J] [. 段晨辉](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=%E6%AE%B5%E6%99%A8%E8%BE%89&option=202) [,张小女](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=%E5%BC%A0%E5%B0%8F%E5%A5%B3&option=202) [. 信息与电脑](https://www.zhangqiaokeyan.com/journal-cn-8946/). 2021,第015期 [↑](#endnote-ref-2)
3. 大数据时代数据仓库技术的发展及应用 [J] . 中国传媒科技 . [薛振文,](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=%E8%96%9B%E6%8C%AF%E6%96%87&option=202)[李彤,](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=%E6%9D%8E%E5%BD%A4&option=202)[牛帆](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=%E7%89%9B%E5%B8%86&option=202) .2023 .10期 [↑](#endnote-ref-3)
4. [Building Data Warehouses in the Era of Big Data: An Approach for Scalable and Flexible Big Data Warehouses](https://www.zhangqiaokeyan.com/academic-conference-foreign_meeting-208151_thesis/020514306690.html) [C] . [Carlos Costa,](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=Carlos%20Costa&option=202) [Maribel Yasmina Santos](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=Maribel%20Yasmina%20Santos&option=202)[International conference on advanced information systems engineering .](https://www.zhangqiaokeyan.com/conference-foreign-208151/) 2019 [↑](#endnote-ref-4)
5. 赵京鹤,郭立,商秋平.企业生产管理数据仓库构建[J].中国自动识别技术,2023,(02):44-46. [↑](#endnote-ref-5)
6. [Estudi i solució Datawarehouse basada en arquitectures big data](https://www.zhangqiaokeyan.com/open-access_resources_thesis/0100029193432.html) [O] . [Dalmau Carmona Joan](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=Dalmau%20Carmona%20Joan&option=202)2016 [↑](#endnote-ref-6)
7. [基于Lambda架构的移动互联大数据平台架构的设计与应用](https://www.zhangqiaokeyan.com/academic-journal-cn_sci-tech-development-enterprise_thesis/0201244961304.html) [J] [. 苏树鹏](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=%E8%8B%8F%E6%A0%91%E9%B9%8F&option=202) [. 企业科技与发展](https://www.zhangqiaokeyan.com/journal-cn-1921/). 2016,第006期 [↑](#endnote-ref-7)
8. [Strategies for Big Data Analytics through Lambda Architectures in Volatile Environments](https://www.zhangqiaokeyan.com/academic-journal-foreign_detail_thesis/0204118543106.html) [J] . [Veith Alexandre da Silva,](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=Veith%20Alexandre%20da%20Silva&option=202) [Anjos Julio C.S. dos,](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=Anjos%20Julio%20C.S.%20dos&option=202) [de Freitas Edison Pignaton,](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=de%20Freitas%20Edison%20Pignaton&option=202) [IFAC PapersOnLine .](https://www.zhangqiaokeyan.com/journal-foreign-20181/) 2016,第30期 [↑](#endnote-ref-8)
9. [A Personal Analytics Platform for the Internet of Things - Implementing Kappa Architecture with Microservice-based Stream Processing](https://www.zhangqiaokeyan.com/academic-conference-foreign_meeting-247620_thesis/0705014150010.html) [C] . [Theo Zschornig,](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=Theo%20Zschornig&option=202) [Robert Wehlitz,](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=Robert%20Wehlitz&option=202) [Bogdan Franczyk](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=Bogdan%20Franczyk&option=202)[International Conference on Enterprise Information Systems .](https://www.zhangqiaokeyan.com/conference-foreign-247620/) 2017 [↑](#endnote-ref-9)
10. # [肖睿](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=%E8%82%96%E7%9D%BF&option=202),[许红涛](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=%E8%AE%B8%E7%BA%A2%E6%B6%9B&option=202),[吴保杰](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=%E5%90%B4%E4%BF%9D%E6%9D%B0&option=202),[蔚赵春](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=%E8%94%9A%E8%B5%B5%E6%98%A5&option=202),[王洪涛](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=%E7%8E%8B%E6%B4%AA%E6%B6%9B&option=202)等 基于Kappa架构的实时日志分析平台研究与实践[J]. 中国金融电脑.2021.第八期

    [↑](#endnote-ref-10)
11. # 大数据技术4：Lambda和Kappa架构区别. 何哥. 大数据治理. 2024.

    [↑](#endnote-ref-11)
12. 宋传园.数据仓库的概念与技术分析[J].信息记录材料,2023,24(05):65-67.DOI:10.16009/j.cnki.cn13-1295/tq.2023.05.013. [↑](#endnote-ref-12)
13. . [数据仓库维度建模设计原则及应用](https://www.zhangqiaokeyan.com/academic-journal-cn_journal-yichun-university_thesis/0201234550480.html) [J] [. 廖飒](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=%E5%BB%96%E9%A3%92&option=202) [,黄光明](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=%E9%BB%84%E5%85%89%E6%98%8E&option=202) [. 宜春学院学报](https://www.zhangqiaokeyan.com/journal-cn-3848/). 2007,第004期 [↑](#endnote-ref-13)
14. 作者：大数据老哥 https://www.bilibili.com/read/cv9116595/?from=search&spm\_id\_from=333.337.0.0 [↑](#endnote-ref-14)
15. [陈敬静](https://www.zhangqiaokeyan.com/search.html?doctypes=1_2_3_4_5_6_7_1-0_9_4-0&sertext=%E9%99%88%E6%95%AC%E9%9D%99&option=202).[SQLite数据库研究与可视化](javascript:goDetail('/academic-degree-domestic_mphd_thesis/020316042470.html','020316042470',15)).南京邮电大学 .2020 [↑](#endnote-ref-15)
16. [基于Spark+Flask的大数据可视化系统设计与实现](https://www.zhangqiaokeyan.com/academic-journal-cn_science-information_thesis/02012100105784.html) [J] [. 王源](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=%E7%8E%8B%E6%BA%90&option=202) [,陈智勇](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=%E9%99%88%E6%99%BA%E5%8B%87&option=202) [. 科学与信息化](https://www.zhangqiaokeyan.com/journal-cn-55169/). 2022,第22期 [↑](#endnote-ref-16)
17. [Implementation of database using python flask framework](https://www.zhangqiaokeyan.com/open-access_resources_thesis/01000131368082.html) [O] . [Nidhi Chauhan,](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=Nidhi%20Chauhan&option=202) [Mandeep Singh,](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=Mandeep%20Singh&option=202) [Ayushi Verma,](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=Ayushi%20Verma&option=202) 2019 [↑](#endnote-ref-17)
18. [Python语言在“程序设计基础”课程中的应用探讨](javascript:goDetail('/academic-journal-cn_way-success_thesis/02012100407668.html','02012100407668',3)).[史文](https://www.zhangqiaokeyan.com/search.html?doctypes=1_2_3_4_5_6_7_1-0_9_4-0&sertext=%E5%8F%B2%E6%96%87&option=202),[孙曼曼](https://www.zhangqiaokeyan.com/search.html?doctypes=1_2_3_4_5_6_7_1-0_9_4-0&sertext=%E5%AD%99%E6%9B%BC%E6%9B%BC&option=202),[李春辉](https://www.zhangqiaokeyan.com/search.html?doctypes=1_2_3_4_5_6_7_1-0_9_4-0&sertext=%E6%9D%8E%E6%98%A5%E8%BE%89&option=202).[《成才之路》](https://www.zhangqiaokeyan.com/journal-cn-916/).2023.第3期 [↑](#endnote-ref-18)
19. 刘晓莉,李满,熊超等.基于Hadoop搭建高可用数据仓库的研究和实现[J].现代信息技术,2023,7(01):99-101.DOI:10.19850/j.cnki.2096-4706.2023.01.02 [↑](#endnote-ref-19)
20. [Research in Big Data Warehousing using Hadoop](https://www.zhangqiaokeyan.com/open-access_resources_thesis/01000132628404.html) [O] . [Abderrazak Sebaa,](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=Abderrazak%20Sebaa&option=202) [Fatima Chikh,](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=Fatima%20Chikh&option=202) [Amina Nouicer,](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=Amina%20Nouicer&option=202) 2017 [↑](#endnote-ref-20)
21. [基于Hadoop的大数据仓库建设及应用](https://www.zhangqiaokeyan.com/academic-journal-cn_china-digital-cable-tv_thesis/0201290099523.html) [J] [. 邓戴伟](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=%E9%82%93%E6%88%B4%E4%BC%9F&option=202) [. 中国有线电视](https://www.zhangqiaokeyan.com/journal-cn-8815/). 2021,第006期 [↑](#endnote-ref-21)
22. [媒体关注、绿色技术创新与企业环境绩效](https://www.zhangqiaokeyan.com/academic-journal-cn_pioneering-with-science-technology-monthly_thesis/0201299940452.html) [J] [. 罗宇洁](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=%E7%BD%97%E5%AE%87%E6%B4%81&option=202) [,刘佳丽](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=%E5%88%98%E4%BD%B3%E4%B8%BD&option=202) [. 科技创业月刊](https://www.zhangqiaokeyan.com/journal-cn-5177/). 2022,第8期 [↑](#endnote-ref-22)
23. [基于微信小程序的校园二手书籍交易平台设计与实现](https://www.zhangqiaokeyan.com/academic-journal-cn_digital-community-smart-home_thesis/0201291139699.html) [J] [. 李沛熹](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=%E6%9D%8E%E6%B2%9B%E7%86%B9&option=202) [,朱晓君](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=%E6%9C%B1%E6%99%93%E5%90%9B&option=202) [,姜建](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=%E5%A7%9C%E5%BB%BA&option=202) [. 电脑知识与技术](https://www.zhangqiaokeyan.com/journal-cn-8877/). 2021,第023期 [↑](#endnote-ref-23)
24. [校园二手书交易平台的可行性研究及构建策略](https://www.zhangqiaokeyan.com/academic-journal-cn_science-education-article-collects_thesis/0201222402548.html) [J] [. 熊鹏](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=%E7%86%8A%E9%B9%8F&option=202) [,李海杰](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=%E6%9D%8E%E6%B5%B7%E6%9D%B0&option=202) [. 科教文汇](https://www.zhangqiaokeyan.com/journal-cn-3331/). 2015,第30期 [↑](#endnote-ref-24)
25. [Research in Big Data Warehousing using Hadoop](https://www.zhangqiaokeyan.com/academic-journal-foreign_detail_thesis/0204117854780.html) [J] . [Abderrazak Sebaa,](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=Abderrazak%20Sebaa&option=202) [Fatima Chikh,](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=Fatima%20Chikh&option=202) [Amina Nouicer,](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=Amina%20Nouicer&option=202) [Journal of Information Systems Engineering and Management .](https://www.zhangqiaokeyan.com/journal-foreign-34727/) 2017,第2期 [↑](#endnote-ref-25)
26. [基于Hive数据仓库的物流大数据平台的研究与设计](https://www.zhangqiaokeyan.com/academic-journal-cn_electronic-design-engineering_thesis/0201235186757.html) [J] [. 张锐](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=%E5%BC%A0%E9%94%90&option=202) [. 电子设计工程](https://www.zhangqiaokeyan.com/journal-cn-8722/). 2017,第009期 [↑](#endnote-ref-26)
27. [基于hive的日志仓库构建研究](https://www.zhangqiaokeyan.com/academic-journal-cn_computer-era_thesis/0201241328645.html) [J] [. 蒋焕亮](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=%E8%92%8B%E7%84%95%E4%BA%AE&option=202) [. 计算机时代](https://www.zhangqiaokeyan.com/journal-cn-8903/). 2016,第011期 [↑](#endnote-ref-27)
28. [虚拟化项目中Linux虚拟机置备技巧](https://www.zhangqiaokeyan.com/academic-journal-cn_security-informatization_thesis/02012100591250.html) [J] [. 赵明立](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=%E8%B5%B5%E6%98%8E%E7%AB%8B&option=202) [. 网络安全和信息化](https://www.zhangqiaokeyan.com/journal-cn-24216/). 2023,第1期 [↑](#endnote-ref-28)
29. [Dual-JT: Toward the high availability of JobTracker in Hadoop](https://www.zhangqiaokeyan.com/academic-conference-foreign_2012-ieee-4th-international-conference-cloud-computing-technology_thesis/020512208805.html) [C] . [Wan Jian,](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=Wan%20Jian&option=202) [Liu Minggang,](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=Liu%20Minggang&option=202) [Hu Xixiang,](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=Hu%20Xixiang&option=202) [2012 IEEE 4th International Conference on Cloud Computing Technology and Science. .](https://www.zhangqiaokeyan.com/conference-foreign-33669/) 2012 [↑](#endnote-ref-29)
30. [Evaluating partitioning and bucketing strategies for Hive-based Big Data Warehousing systems](https://www.zhangqiaokeyan.com/academic-journal-foreign_detail_thesis/0204117673303.html) [J] . [Eduarda Costa,](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=Eduarda%20Costa&option=202) [Carlos Costa,](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=Carlos%20Costa&option=202) [Maribel Yasmina Santos](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=Maribel%20Yasmina%20Santos&option=202)[Journal of Big Data .](https://www.zhangqiaokeyan.com/journal-foreign-33698/) 2019,第1期 [↑](#endnote-ref-30)
31. [基于Python Flask的运维信息管理系统设计与实现](https://www.zhangqiaokeyan.com/academic-journal-cn_computer-programming-skills-maintenance_thesis/0201289467586.html) [J] [. 冷四军](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=%E5%86%B7%E5%9B%9B%E5%86%9B&option=202) [. 电脑编程技巧与维护](https://www.zhangqiaokeyan.com/journal-cn-8872/). 2021,第005期 [↑](#endnote-ref-31)
32. [基于黑盒测试与白盒测试的比较探究](https://www.zhangqiaokeyan.com/academic-journal-cn_electronics-world_thesis/0201290232895.html) [J] [. 妥泽花](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=%E5%A6%A5%E6%B3%BD%E8%8A%B1&option=202) [. 电子世界](https://www.zhangqiaokeyan.com/journal-cn-8721/). 2021,第011期 [↑](#endnote-ref-32)
33. [Open Cloud deep-learning architecture for big data analytics and real-time applications.](https://www.zhangqiaokeyan.com/academic-degree-foreign_mphd_thesis/02061474699.html) [D] . [Das, Arun.](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=Das,%20Arun.&option=202)2016 [↑](#endnote-ref-33)
34. [A Personal Analytics Platform for the Internet of Things - Implementing Kappa Architecture with Microservice-based Stream Processing](https://www.zhangqiaokeyan.com/academic-conference-foreign_meeting-247620_thesis/0705014150010.html) [C] . [Theo Zschornig,](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=Theo%20Zschornig&option=202) [Robert Wehlitz,](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=Robert%20Wehlitz&option=202) [Bogdan Franczyk](https://www.zhangqiaokeyan.com/search.html?doctypes=4_5_6_1-0_4-0_1_2_3_7_9&sertext=Bogdan%20Franczyk&option=202)[International Conference on Enterprise Information Systems .](https://www.zhangqiaokeyan.com/conference-foreign-247620/) 2017 [↑](#endnote-ref-34)
35. [数据仓库的概念与技术分析](javascript:goDetail('/academic-journal-cn_information-recording-materials_thesis/02012101435002.html','02012101435002',10))[宋传园](https://www.zhangqiaokeyan.com/search.html?doctypes=1_2_3_4_5_6_7_1-0_9_4-0&sertext=%E5%AE%8B%E4%BC%A0%E5%9B%AD&option=202).[《信息记录材料》](https://www.zhangqiaokeyan.com/journal-cn-9244/) 2023年第5期 [↑](#endnote-ref-35)