



北京交通大学

综合专题研究课程III

基于大数据的智能推荐系统设计与实现

张迪

dizhang@bjtu.edu.cn



北京交通大学

基于大数据的智能推荐系统设计与实现

课程项目简介

张迪

dizhang@bjtu.edu.cn

课程信息

- 软件工程专业 《综合专题研究课程III》
 - 实践训练为主，课程讲授和指导为辅
 - 面向学生：2019级软件工程专业本科生
 - 项目名称：基于大数据的智能推荐系统设计与实现
- 任课教师
 - 主讲教师：张迪
 - 联系方式：dizhang@bjtu.edu.cn
 - 助教：鲍鸿丞、刘书梦、张翼沁、逢思默

课程安排

- 时间安排
 - 7月4日-7月15日
 - 每天6.5个小时
 - 上午8:30-11:30, 下午13:30-17:00
- 项目汇报
 - 每日书面报告
 - 按小组汇报, 组长负责, 每人都要汇报, 每日24点之前发给助教
 - 中期汇报 (7月8日[周五]下午)
 - 进展汇报 (每组10分钟), 提交需求文档、中期汇报ppt (大数据处理)
 - 项目答辩 (7月15日[周五]全天)
 - 最终项目答辩 (每组20分钟), 提交需求文档、设计和开发文档、测试文档、使用说明书、项目代码和最终ppt

课程考核

- 考核形式

- 分组考核

- 分数组成

- 平时分（10分）
- 中期考核分（10分）
- 文档分（10分）
- 项目分（50分）
- 最终答辩分（20分）

- 组内得分

- 以小组分为基础划分三个段
 - 最高分：小组分加3分
 - 中间分：小组分
 - 最低分：小组分减3分

- 分数分配

- 组长给组员评价ABC
 - A一人，B两人，C两人
- 组员A得最高分，组员B得中间分，组员C得最低分

分组安排

- 5人一组
 - 1名组长，4名组员
 - 组长负责制，负责人员协调与任务分配
- 分组建议
 - 擅长不同任务的同学一起组队
 - 组长选择要慎重
 - 建议强弱联合
- 分组要求
 - 每个小组起一个组名

小提问

• 什么是大数据？

- 需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产

• 大数据的5V特征

- Volume（大量）
- Velocity（高速）
- Variety（多样）
- Value（低价值密度）
- Veracity（真实性）



小提问

- 什么是人工智能？
 - 它是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的技术科学
- 你认为的电商系统应该包含哪些基本功能？
 - 用户管理、商品管理、权限管理
 - 购物车、支付管理、数据分析和报表
- 什么是个性化推荐？
 - 它是建立在海量数据挖掘基础上的一种智能平台，向用户提供个性化的信息服务和决策支持

实训项目

- 基于大数据的智能推荐系统设计与实现
 - 大数据
 - 大数据平台：Hadoop, hive, sqoop
 - 数据采集、迁移、分析和清洗
 - 大数据系统与传统数据库之间的交互
 - 人工智能推荐系统
 - 聚类算法
 - 推荐算法
 - 图书电商Web平台
 - 个性化的图书销售平台

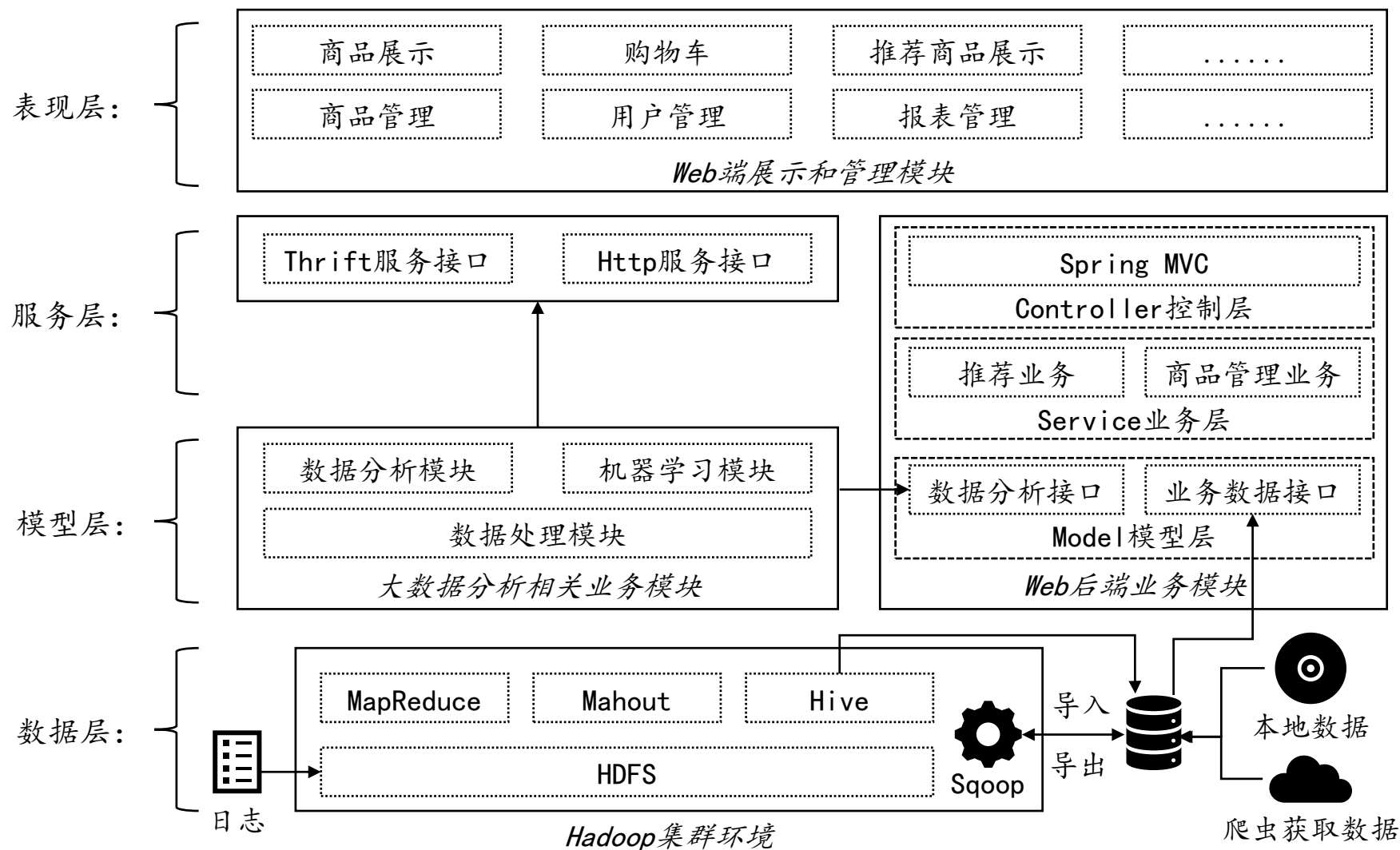
项目特色

- 1. 涵盖大数据项目开发的整个流程
 - 从数据采集、迁移到ETL，再到数据开发和数据分析
- 2. 跨越和整个多个技术方向
- 3. 基于传统Web电商网站改造升级，帮助传统企业进行大数据转型
- 4. 项目的扩展性高
- 5. 整合人工智能的相关算法，具备一定的新颖性和实用性

项目展示



项目模块划分



项目任务划分

• 1. 系统设计与环境搭建

子模块	功能	技术点
Hadoop集群	搭建高可用的Hadoop分布式集群环境	Hadoop、Zookeeper
安装相关大数据组件	安装项目所依赖的相关软件和大数据组件	MySQL、Hive、Sqoop、Python
安装配置开发工具	配置开发环境，安装配置开发工具	IntelliJ IDEA、Anaconda

项目任务划分

• 2. 数据处理模块

子模块	功能	技术点
数据采集	使用爬取图书信息并保存到数据库	Python、Scrapy
数据导入导出	将Excel数据导入HDFS 使用Sqoop将HDFS数据导入MySQL 使用Sqoop将MySQL数据导入HDFS	Sqoop、mysql、HDFS
业务数据的数据清洗	编写业务数据清洗模块	ETL
点击流日志的数据清洗	编写点击流日志清洗模块	ETL

项目任务划分

• 3. 数据分析与机器学习模块

子模块	功能	技术点
数据分析工具	编写Hive for Hadoop数据分析的工具类 编写通用的工具类HiveUtil、SqoopUtil	Hive、Hadoop
购书转化率分析	编写购书转化率模块	Hive
购书用户聚类	基于Kmeans的聚类分析	Mathout、numpy、pandas、scikit-learn
个性化推荐系统	基于协同过滤的推荐系统	Mathout、numpy、pandas、scikit-learn

项目任务划分

• 4. Web端功能模块

子模块	功能	技术点
系统架构设计	搭建基于SSM的web开发框架	Spring、Struts、MyBatis
后端开发	图书管理 用户管理 购物车 支付管理 推荐管理 权限管理 数据分析与报表 图书投放预测 销售额预测	MVC

项目任务划分

• 4. Web端功能模块

子模块	功能	技术点
前端开发	后台管理页面 首页展示 图书列表 图书详情 购物车 支付页面 登录和注册页面 个性推荐商品展示 图书推荐词云展示 个性化聚类展示 图书热点图及Top20展示	JSP、JavaScript、JQuery、echart、vue



北京交通大学

Thank you!

Q & A