黄河交通学院本科毕业设计（论文）任务书

学院：智能工程学院

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 学生姓名 | 刘丹杰 | | 专业班级 | 大数据20-1班 | 学号 | 2080910T01521 |
| 指导教师 | 炎士涛 | | 职称 | 副教授 | 学位 | 硕士 |
| 题目名称 | 基于Hadoop的新能源汽车销售数据分析系统的设计与实现 | | | | | |
| 起止时间 | 2024年1月2日 至 2024年6月15日 | | | | | |
| 选题性质 | □理论研究 应用研究 □技术开发 □产品设计 □其他 | | | | | |
| 是否在实践中完成 | | □是 否 | | | | |
| 设计（研究）目标：  设计并实现一个可扩展、高性能的数据分析系统，能够处理大规模的新能源汽车销售数据，并提供多维度的数据分析功能，包括销售趋势分析、地域分布分析、车型偏好分析等。实现数据可视化功能，以图表、报表等形式直观展示分析结果，并构建友好的用户界面，方便用户进行数据查询、筛选和导出。  设计（研究）内容及具体要求：  1. 数据采集模块  数据源接入：支持多种数据源的接入，如销售平台、社交媒体、政府统计数据等。  数据预处理：对采集的数据进行清洗、去重、格式化等预处理操作，确保数据质量。  2.数据存储模块  Hadoop分布式文件系统（HDFS）：用于存储大规模的销售数据，提供高可靠性和高吞吐量。  数据仓库：构建基于Hadoop的数据仓库，对数据进行组织、索引和优化查询。  3. 数据处理与分析模块  MapReduce编程模型：用于处理大规模数据集，进行复杂的数据转换和聚合操作。Hive/Pig：提供SQL-like语言或脚本语言，用于数据的批处理和分析。  4. 实时数据处理模块  Spark Streaming：用于处理实时销售数据流，提供实时数据分析和处理能力。  Kafka：作为消息队列，支持高吞吐量的实时数据摄入。  5. 数据挖掘与机器学习模块  机器学习算法库：集成机器学习算法，用于预测销售趋势、客户偏好分析等。  数据挖掘工具：提供关联规则、聚类分析、分类等数据挖掘技术。  6. 数据可视化模块  图表生成工具：将分析结果通过图表、图形等形式直观展示。  仪表板：为用户提供实时数据监控和历史数据分析的仪表板。  7. 用户交互与报告模块  用户界面：提供友好的用户界面，支持数据查询、报告生成和下载。  报告自动生成：根据用户需求自动生成销售报告和分析报告。  8. 安全与权限管理模块  用户认证与授权：确保只有授权用户才能访问敏感数据。  数据加密：对存储和传输的数据进行加密，保障数据安全。  9. 系统监控与维护模块  日志管理：记录系统操作日志，便于问题追踪和性能分析。  性能监控：监控系统性能，确保数据处理的高效和稳定。  10. 扩展与集成模块  API接口：提供API接口，支持与其他系统或应用的集成。  模块化设计：系统采用模块化设计，便于未来功能的扩展和升级。  11.必须在规定时间内按质按量地完成论文，观点正确，结构合理，条理清晰，论据有理有据，具备一定的分析能力和概括能力。 | | | | | | |
| 进度安排：  1. 2024年1月2日--2024年3月15日， 完成选题以及开题工作。  2. 2024年3月16日--2024年4月26日，完成初稿，中期检查。  3. 2024年4月27日--2024年5月25日，完成第二稿。  4. 2024年5月26日--2024年5月31日，完成论文查重与修改。  5. 2024年6月1日--2024年6月15日，整理资料，完成定稿，完成答辩。  指导教师签字： 年 月 日 | | | | | | |
| 主要参考文献：  [1]周德,杨成慧,罗佃斌.基于Hadoop的分布式日志分析系统设计与实现[J].现代信息科技,2023,7(23):57-60.DOI:10.19850/j.cnki.2096-4706.2023.23.012.  [2]任宏,李春林,李晓峰.基于Hadoop技术的物联网大数据同步存储系统设计[J].网络安全和信息化,2023(12):85-87.  [3]谢盛嘉.基于Hadoop平台的学情分析系统设计[J].电子技术,2023,52(11):408-409.  [4]王子昱.基于Hadoop的大数据云计算处理的实现[J].无线互联科技,2023,20(19):89-91+104.  [5]李威,邱永峰.基于Hadoop的电商大数据可视化设计与实现[J].现代信息科技,2023,7(17):46-49.DOI:10.19850/j.cnki.2096-4706.2023.17.009.  [6]邹文景,唐良运,甘莹等.基于Hadoop技术的物联网大数据同步存储系统设计[J].电子设计工程,2023,31(18):114-117+122.DOI:10.14022/j.issn1674-6236.2023.18.024.  [7]Liuqi Z ,Xing W ,Zhenlin H , et al.Power Big Data Analysis Platform Design Based on Hadoop[J].Journal of Physics: Conference Series,2023,2476(1):  [8]Ning X .Individual Online Learning Behavior Analysis Based on Hadoop[J].Computational Intelligence and Neuroscience,2022,20221265340-1265340.  [9]陶淘,彭颖,张晨亮.基于Hadoop技术的气象数据实时传输监控系统设计[J].计算机测量与控制,2024,32(01):114-120.DOI:10.16526/j.cnki.11-4762/tp.2024.01.017.  [10]赵建立,汤卓凡,姚孟阳.基于Hadoop的配电网需求数据存储控制技术优化[J].粘接,2024,51(02):182-185.  [11]那蓉萃.基于Hadoop的工业物联网大数据处理及应用[J].信息记录材料,2023,24(12):221-223+226.DOI:10.16009/j.cnki.cn13-1295/tq.2023.12.061.  [12]石文昭.基于Hadoop的自动化设备监管系统设计[J].信息记录材料,2023,24(11):178-180.DOI:10.16009/j.cnki.cn13-1295/tq.2023.11.025. | | | | | | |
| 系（教研室）意见：  主任签字： 年 月 日 | | | | | | |
| 学院意见：  负责人签字： 年 月 日 | | | | | | |

备注：1.任务书由指导教师填写并下发给学生；若是学生自选设计（论文）题目，任务书可在导师指导下由学生填写；

2.本表一式四份，在毕业设计（论文）开始前提交，学院、系（教研室）、导师、学生各一份。