**北 京 邮 电 大 学**

**本科毕业设计（论文）开题报告**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 学院 | 计算机学院 | 专业 | | 计算机科学与技术 | | 班级 | 2012211308 |
| 学生姓名 | 胡官涛 | 学号 | | 2012211320 | | 班内序号 | 18 |
| 指导教师姓名 | 王红熳 | 所在单位 | | 教育部信息网络工程研究中心 | | 职称 | 教授 |
| 设计（论文）题目 | （中文） HBASE与关系数据库同步工具的设计与实现 | | | | | | |
| （英文） The Design and Implementation of synchronization tool between HBASE and relational database | | | | | | |
| 毕业设计（论文）开题报告内容：  1：**课题背景和意义**  因特网的发展已经彻底改变了人们的生活，每时每刻都有大量数据被传送到互联网上。伴随着数据规模越来越大，越来越多的数据存取功能开始水平扩展，这样就可以删除和增加存储服务，从而适应多变的业务需求。而传统的关系数据库虽然在SQL查询方面表现良好，但是他们更侧重在一台服务器上，海量数据的存储成为瓶颈，单台服务器无法加载海量数据。而HBASE不同于传统的关系数据库，它是基于列存储的非关系数据库，在海量数据的存储和查询方面有很大优势。但是HBASE不支持SQL查询，而且对于繁复的业务流程难以满足需求。而传统的关系数据库恰好能和HBASE互补，如果能使用工具将二者随意转换，那么刚好在弥补彼此的缺点之后，在相应的需要的场景使用不同的数据模式，便能在大数据时代得心应手。  本次论文研究HBASE与关系数据库同步工具的设计与实现，旨在提供一个方便快捷的工具，从而使企业在使用的时候可以避免传统关系数据库和HBASE的不足点，充分利用二者的优点，适应大数据时代背景和多变的业务需求，从而更好的提供互联网服务。  2：**研究的基本内容和拟解决的主要问题**  实现从HBASE导出数据到关系数据库、从关系数据库导出数据到HBASE的功能，导出时支持按照时间、按表、按列导出、对导出过程进行监控等。  **2.1 主要内容**：  熟悉传统数据库技术，熟悉HBASE相关技术，  调研HBASE与关系型数据库开源同步工具，  搞清在关系数据库和HBASE转换时，各关键字之间的依赖关系如何消除并存储进HBASE，在还原的时候能重现  运用软件工程思想完成HBASE与关系数据库同步工具设计  完成HBASE与关系数据库同步工具设计的单元测试和集成联调。  **2.2 拟解决的主要问题**：关系数据库和HBASE之间互相转换时候，数据间各种依赖关系的解析和消除并再现。  当从HBASE转换到关系数据库的时候，只需要将每个列族单独存储为一个表即可；  当从关系数据库转换到HBASE的时候，需要重新设计HBASE的表模式，步骤为：a、按照传统关系数据库的设计  方法对所有表进行初步设计，b、对表模式进行BCNF分解（减少表间数据的冗余和依赖关系），并用几种转换方法进  行处理，得到转换后的HBASE表。  3：**研究方法及措施**  **3.1 表模式转换方法**：基本转换、内嵌转换、分割转换、内联转换。  **3.2 基本解决思路和过程：**  学习已有开源工具，熟悉其代码和思想；  在学习的基础上，利用软件工程思想，设计工具的基础架构并设计功能等；  利用数据库相关知识，解决在数据库转化过程中依赖关系的处理（BCNF）范式的分解；  利用上述表模式转换方法，对表进行转换；  采用JAVA语言进行开发，实现所有功能，并实现联调和单元测试。  4：**研究工作的步骤与进度**  第1周：基础知识准备，回顾关系数据库数据库相关知识、了解NoSQL数据库HBASE相关技术；  第2周：准备开题汇报，编写任务书，开题报告书；  第3-4周：确定技术方案，调研HABASE与关系数据库开源同步工具，了解其基本运作机制，分析实现方式，掌握关键技术；基于调研，确定HBASE与关系数据库同步工具技术方案；  第5-7周：编写设计文档，编写HBASE与关系数据库同步工具概要设计文档，编写HBASE与关系数据库同步工具详细设计文档；  第7-9周：根据设计文档进行开发，搭建开发环境，并开始开发工作，进行HBASE与关系数据库同步工具开发，完成HBASE与关系数据库同步工具开发工作；  第10周：测试优化，对HBASE与关系数据库同步工具进行功能、性能测试；  第11-12周：总结并编写论文，整理项目文档，开始毕业论文编写，撰写毕业设计论文，准备答辩。  5**：主要参考文献等项目**  [1]：周海汉 Apache Software Foundation 《HBASE官方文档中文版》<http://abloz.com/hbase/book.html>  [2]：明日科技 《JAVA从入门到精通（第三版）》 清华大学出版社，  [3]：[美] Lars George，译者:代志远/刘佳/蒋杰 《Hbase权威指南》 东南大学出版社  [4]：Baron Schwartz译者: 王小东/李军/康建勋 《高性能mysql》 电子工业出版社  [5]：肖丁 修佳鹏 《软件工程模型与方法》 北京邮电大学出版社  [6]：杨寒冰 同济大学毕业设计论文 2013 《关系型数据库与HBASE间的数据迁移介绍》  [7]：Abraham Silberschatz等 《数据库系统概念》 机械工业出版社  [8]：作者不详 百度文库 《HBase基本shell命令介绍》  [9]：HBASE视频教程（极客学院） <http://www.jikexueyuan.com/course/hbase/>  [10]：java课程视频（极客学院） <http://search.jikexueyuan.com/course/?q=java>  [11]：intelij idea 官方网站 <http://www.jetbrains.com/idea/> | | | | | | | |
| 指导教师签字 |  | | 日期 | | 年 月 日 | | |

注：可根据开题报告的长度加页。