《基于Hadoop平台的大数据迁移与查询方法研究及应用》

1. 高效的数据关系系统对于数据应用管理非常重要；
2. 数据规模持续增长，数据类型不断变化、非结构化数据成为存储和处理对象等因素-------->导致传统关系型数据库主导地位得到动摇，不能满足有效的存储、分析和访问数据
3. Hadoop：高性能集群计算和存储能力，在分布式计算和大量数据处理方面脱颖而出，使用Hadoop平台对大数据进行处理，并实现高效查询，需对Hadoop与DBMS进行迁移，使用Hadoop来优化数据库核心工作效率及数据库性能，完成不断更新的查询需求。
4. 研究Hadoop平台相关架构及数据交换原理基础上，对基于Hadoop平台的数据迁移提出了MR技术的实现方案，MR可实现更优的并发性，体现优化数据转化的能力。
5. 深入分析MR的工作机制及Hadoop三种常用的作业调度器。在此基础上，对MR的作业调度器进行优化设计，将公平调度器与优先权(高响应比优先)调度算法相结合，提出基于优先权的公平调度器的改进算法，同时以TaskTracker的负载均衡监听器辅助进行调度工作。
6. 分析HBase和Hive各自的机制，对Hive-HBase相结合的查询方法进行探讨，设计了一个基于Hive-HBase的数据查询方案。
7. 搭建实验环境，对于Hadoop平台的数据迁移及数据查询进行实验与分析，对比不同调度算法的迁移性能