**基于Hadoop的大数据概论**

课程专注于Hadoop生态系统的基础构建与核心技术应用。课程从大数据的概念与特征出发，深入介绍Hadoop的架构原理，包括分布式文件系统（HDFS）对大规模数据存储的高效管理机制，以及MapReduce编程模型在分布式数据处理上的强大功能。学生将学习如何在Hadoop平台上进行数据的存储、读取、计算与分析操作，掌握YARN资源管理框架对集群资源的合理调配。同时，课程还会涉及Hive、HBase等相关组件的基本使用，让学生了解如何基于 SQL 类似的语言操作 Hive 数据仓库进行数据查询与分析，以及利用HBase构建面向列的分布式数据库存储海量结构化数据。通过丰富的实践案例与项目练习，培养学生具备运用Hadoop技术处理大数据的基本能力，为其深入探索大数据开发领域或从事相关大数据工程工作筑牢坚实的基础。