

《 大数据开发技术 》本科课程教学大纲

一、课程基本信息

课程名称	(中文) 大数据开发技术				
	(英文) Big Data Development Technology				
课程代码	2055036	课程学分		2	
课程学时	32	理论学时	12	实践学时	20
开课学院	信息技术学院	适用专业与年级		软件工程等相关专业 三、四年级	
课程类别与性质	专业 必修课	考核方式		考查	
选用教材	《Hadoop 大数据开发实战(慕课版)》. 千峰教育 编著. 人民邮电出版社. 2021			是否为马工程教材	否
先修课程	面向对象程序设计 数据库原理				
课程简介	<p>Hadoop 大数据开发技术是计算机、软件工程及相关专业学生的一门重要课程。本课程内容覆盖全面、讲解详细、注重实践，其中第 1 章首先让同学对大数据及 Hadoop 有一个总体的认识；第 2 章主要讲解如何搭建 Hadoop 分布式集群；第 3~第 5 章讲解 HDFS 分布式文件系统、MapReduce 分布式计算框架以及 Zookeeper 分布式协调服务；第 6 章讲解 Hadoop 2.0 的新特性；第 7~第 10 章主要讲解 Hadoop 生态圈中的相关辅助系统，包括 Hive、HBase 分布式存储系统、Flume、Sqoop；第 11 章讲解了综合项目——例如电商精准营销，该项目涵盖从前期设计到最终实施的整个过程的内容，对全书知识点进行串联和巩固，使同学们融会贯通，加深对 Hadoop 大数据开发技术的理解。</p>				
选课建议与学习要求	本课程是软件工程、计算机科学与技术等专业的专业必修(或选修)课，建议在第五学期开设。				
大纲编写人	朱丽娟		制/修订时间	2024 年 8 月	
专业负责人	朱丽娟		审定时间	2024 年 8 月	
学院负责人	陈桂斌		批准时间	2024 年 8 月	

二、课程目标与毕业要求

(一) 课程目标

类型	序号	内容
知识目标	1	工程知识：能够运用数学、大数据专业知识和工程基础知识对大数据应用复杂工程问题进行抽象表述和建模。
	2	方案设计：能够根据工程项目任务要求，选择技术路线，设计满足需求的大数据应用解决方案。
技能目标	3	使用现代工具：了解信息领域主要资料来源及获取方法，熟悉数学、程序设计等大数据应用工程领域现代工程工具。
	4	开发实现：能够综合运用大数据预处理、挖掘分析、可视化等技术手段，开发实现大数据应用系统案例。
	5	研究验证：能够选用或搭建开发环境进行软硬件实现与验证。
素养目标 (含课程思政目标)	6	具有思辨能力和批判精神：能够选择恰当的大数据应用复杂系统抽象模型，对模型的正确性进行推理和求解；能够分析系统的优缺点。
	7	具备自主学习的能力和终身学习的意识：包括对新技术的理解、提出问题、归纳问题和总结问题的能力。

(二) 课程支撑的毕业要求

L05 使用现代工具：能够针对软件系统中的复杂软件工程问题，开发、选择与使用恰当的技术、资源和工具，对软件系统中的复杂工程问题进行预测与模拟，并能够理解其局限性 ①了解软件开发过程中所涉及到的主要的软件开发平台、开发与管理工具的使用原理和方法，了解其差异和适用的领域，并理解其局限性。
L08 职业规范：具有人文社会科学素养、社会责任感，能够在工程实践中理解并遵守 IT 行业的职业道德和规范，履行责任 ③具有思辨能力和批判精神，能够表现出软件工程师对社会的安全和环境保护等的必须具备的社会责任，并在工程实践中自觉遵守和履行责任。

(三) 毕业要求与课程目标的关系

毕业要求	指标点	支撑度	课程目标	对指标点的贡献度
L05	①	M	1 工程知识：能够运用数学、大数据专业知识和工程基础知识对大数据应用复杂工程问题进行抽象表述和建模。	20%
			2 方案设计：能够根据工程项目任务要求，选择技	20%

			术路线，设计满足需求的大数据应用解决方案。	
			3 使用现代工具：了解信息领域主要资料来源及获取方法，熟悉数学、程序设计等大数据应用工程领域现代工程工具。	20%
			4 开发实现：能够综合运用大数据预处理、挖掘分析、可视化等技术手段，开发实现大数据应用系统案例。	20%
			5 研究验证：能够选用或搭建开发环境进行软硬件实现与验证。	20%
L08	③	H	6 具有思辨能力和批判精神：能够选择恰当的大数据应用复杂系统模型，对模型的正确性进行推理和求解；能够分析所开发系统的优缺点。	70%
			7 具备自主学习的能力和终身学习的意识：包括对新技术的理解、提出问题、归纳问题和总结问题的能力。	30%

三、课程内容与教学设计

(一) 各教学单元预期学习成果与教学内容

<p>第 1 单元 初识 Hadoop</p> <p>主要内容：掌握大数据的类型和特征，了解 Hadoop 框架下的解决方案，掌握 Hadoop 核心构成、Hadoop 生态系统的主要组件、Hadoop 发行版本的差异及如何选择；了解 Hadoop 典型应用场景</p> <p>重点：大数据的类型和特征和 Hadoop 框架下的解决方案，Hadoop 核心构成、Hadoop 生态系统的主要组件，Hadoop 典型应用场景</p> <p>难点：Hadoop 框架下的解决方案、Hadoop 核心构成和 Hadoop 生态系统的主要组件</p>
<p>第 2 单元 搭建 Hadoop 集群</p> <p>主要内容：了解 Hadoop 环境搭建前所需做的准备工作，虚拟机软件 VMware 的安装与使用，Ubuntu 操作系统及 Linux 常用命令，掌握 Hadoop 单机和分布式的环境搭建</p> <p>重点：Hadoop 分布式的环境搭建</p> <p>难点：Hadoop 完全分布式安装过程中的集群 SSH 互访，配置文件的编写，有能力解决 Hadoop 启动过程中的集群问题</p>
<p>第 3 单元 HDFS</p> <p>主要内容：HDFS 存储架构和数据读写流程，HDFS 的 Shell 命令，Java 程序操作 HDFS，Hadoop 序列化和小文件处理，等</p> <p>重点：掌握 HDFS 的组成与架构及其工作流程；从 Web 访问、HDFS Shell、HDFS API 三个方面掌握 HDFS 操作</p> <p>难点：HDFS 存储架构和数据读写流程</p>

第 4 单元 MapReduce

主要内容: MapReduce 编程组件、作业解析和工作原理; Shuffle 阶段; 优化—数据倾斜;
典型案例: 排序、倒排索引、连接、平均分以及百分比, 过滤敏感词汇

重点: MapReduce 运行流程, MapReduce 程序设计方法

难点: 掌握 MapReduce 编程模型, 掌握 MapReduce 编程实例

第 5 单元 Yarn 和 Zookeeper

主要内容: YARN 资源管理框架, Hadoop 的 HA 模式; Zookeeper 的安装和常用命令, 客户端编程, 典型应用场景

重点: 理解 YARN 架构的理由、理解 Hadoop 的 HA 模式; 掌握 Zookeeper 的安装方法, 理解 Zookeeper 的基本原理, 理解 Zookeeper 的应用场景

难点: 理解 Yarn 和 Zookeeper 的应用场景, 并掌握 Yarn 和 Zookeeper 的使用

第 6 单元 Hive 和 HBase

主要内容: 认识数据仓库, Hive 安装、Hive 数据库操作、Hive 表及其查询, Hive 函数、性能优化以及 Hive 案例分析; 认识 HBase, HBase 安装、Shell 常用操作, HBase 编程, HBase 过滤器和比较器, HBase 和 Hive 结合, HBase 性能优化

重点: 掌握 Hive 的安装、Hive 的数据库和表的操作方法、Hive 函数的使用, Hive 的查询数据; 掌握 HBase 的安装与部署, JAVA API 操作 HBase, HBase 的存储流程

难点: Hive 架构及其原理、Hive 的性能调优, 掌握 HBase 架构及其原理、理解 HBase 与 Hive 之间的关系

第 7 单元 Flume 和 Sqoop

主要内容: Flume 的基本组件、安装, Flume 数据流模型, Flume 可靠性保证, Flume 拦截器; Sqoop 的基本概念、安装和命令, Sqoop 数据导入和数据导出, Sqoop job

重点: 掌握 Flume 的安装和使用; 掌握 Sqoop 的安装, 掌握 Sqoop 的 import、export、job 等命令的用法

难点: 掌握 Flume 架构及其原理, 掌握 Flume 常用的配置与应用场景; Sqoop 的框架及其应用场景

第 8 单元 综合项目

主要内容: 项目概述, 项目详细介绍, 项目模块分析; 数据采集, 数据清洗; 使用数据仓库进行数据分析, 可视化

重点: 项目背景及需求, 项目模块分析和架构设计, 数据来源、数据清洗流程, 数据仓库操作流程

难点: 项目需求分析和架构设计, 数据采集与清洗, 应用测试方法

(二) 教学单元对课程目标的支撑关系

教学单元	课程目标						
	1	2	3	4	5	6	7
第 1 单元 初识 Hadoop	√	√	√			√	√
第 2 单元 搭建 Hadoop 集群			√		√		
第 3 单元 HDFS	√	√	√	√	√		

第 4 单元 MapReduce	√	√	√	√	√		
第 5 单元 Yarn 和 Zookeeper	√	√	√	√			
第 6 单元 Hive 和 HBase	√	√	√	√	√		
第 7 单元 Flume 和 Sqoop	√	√	√	√	√		
第 8 单元 综合项目		√	√	√	√	√	√

(三) 课程教学方法与学时分配

教学单元	教与学方式	考核方式	学时分配		
			理论	实践	小计
第 1 单元 初识 Hadoop	课堂讲授为主；通过大数据的历史发展，引入 Hadoop 生态系统的介绍，加入 Hadoop 典型应用场景教学案例，由学生讨论，教师总结引导	课后作业	2	0	2
第 2 单元 搭建 Hadoop 集群	理论与上机实验讲授为主，通过环境搭建，引导学生理解 Hadoop 集群的安装原理	课后作业 实验报告	2	2	4
第 3 单元 HDFS	理论与上机实验讲授为主，辅之以 HDFS 的命令与编程演示	课后作业 实验报告	2	2	4
第 4 单元 MapReduce	理论与上机实验讲授为主，辅之以 MapReduce 编程演示	课后作业 实验报告	3	3	6
第 5 单元 Yarn 和 Zookeeper	理论与上机实验讲授为主，辅之以 Zookeeper 的演示	课后作业 实验报告	1	2	3
第 6 单元 Hive 和 HBase	理论与上机实验讲授为主，辅之以 Hbase 和 Hive 编程演示	课后作业 实验报告	1	2	3
第 7 单元 Flume 和 Sqoop	理论与上机实验讲授为主，辅之以 Flume, Sqoop 的编程演示	课后作业 实验报告	1	3	4
第 8 单元 综合项目	上机实践为主，辅之以项目介绍；学生团队合作	大作业	0	6	6
合计			12	20	32

(四) 课内实验项目与基本要求

序号	实验项目名称	目标要求与主要内容	实验时数	实验类型
1	Hadoop 集群搭建	虚拟机和 Linux 系统安装、克隆，网络配置，SSH 服务配置，Hadoop 分布式集群	4	验证型
2	案例编程	HDFS、MapReduce、Hive、Hbase、Flume、	10	设计型

		Sqoop 等编程案例		
3	综合案例	综合性的大数据应用程序的设计和编程	6	综合型

实验类型：①演示型 ②验证型 ③设计型 ④综合型

四、课程思政教学设计

《大数据开发技术》课程思政教学设计将大数据开发的专业知识与思想政治教育相结合，既培养学生的专业技能，又加强其思想道德素养。主要教学设计如下：

- (1) 引导学生认识大数据在经济发展、社会治理等场景中的重要作用；
- (2) 介绍大数据应用程序的开发流程和案例，培养学生的科技道德和社会责任感，使其在使用大数据技术时能够遵循法律法规和伦理规范；
- (3) 鼓励学生分组讨论，分享学习心得和见解；通过组队完成项目实战，培养团队合作和沟通能力；
- (4) 定期对教学效果进行总结和反思，及时调整教学方案和方法。

五、课程考核

总评构成	占比	考核方式	课程目标							合计
			1	2	3	4	5	6	7	
X1	10	课堂表现(出勤、预习、回答问题等)	10	10	10	10	10	30	20	100
X2	20	课后作业	40	10	10	10	10	10	10	100
X3	40	实验报告	10	10	30	10	10	20	10	100
X4	30	大作业	10	10	20	20	20	10	10	100

六、其他需要说明的问题

无