

《 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	大数据开发技术				
	Big Data Development Technology				
课程代码	2055036	课程学分		2	
课程学时	32	理论学时	16	实践学时	16
开课学院	信息技术学院	适用专业与年级		大三软件工程	
课程类别与性质	专业选修课	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 主编. 人民邮电出版社			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>《大数据开发技术》是计算机和软件工程专业学生的一门重要专业拓展选修课。本课程主要学习当前广泛使用的大数据 Hadoop 平台及其主要组件的作用及使用。通过学习 Hadoop 平台框架, 学会手动搭建 Hadoop 环境, 掌握 Hadoop 平台上存储及计算的原理、结构、工作流程, 掌握基础的 MapReduce 编程, 掌握 Hadoop 生态圈常用组件的作用、结构、配置和工作流程, 并具备大数据的动手及问题分析能力, 使用掌握的知识应用到实际的项目实践中。</p> <p>课程由理论及实践两部分组成, 课程理论部分的内容以介绍 Hadoop 平台主要组件的作用、结构、工作流程为主, 对 Hadoop 平台组件的作用及其工作原理有比较深入的了解; 课程同时为各组件设计有若干实验, 使学生在理论学习的同时, 提高实践动手能力, 为在 Hadoop 的大数据平台上进行大数据项目开发做准备。最后, 以综合项目“电商精准营销”对课程知识点进行串联和巩固, 该项目涵盖从前期设计到最终实施的整个过程的内容, 使同学们融会贯通, 从而加深对 Hadoop 技术的理解。</p>				
选课建议与学习要求	本课程是软件工程专业、计算机科学与技术专业的选修课, 建议在第五学期开设。希望在完成本课程学习后, 同学们具备从事大数据开发相关工作岗位的基础知识技能与基本职业素养。				
大纲编写人	卢凯		制/修订时间	2022 年 8 月	
专业负责人	朱丽娟		审定时间	2022 年 8 月	
学院负责人	邓桂娥		批准时间	2022 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、自然科学、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。
	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。
	7	可持续发展：能够理解大数据应用工程实践中节能与绿色计算技术的手段和方法对环境和社会可持续发展的影响。

(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、自然科学、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。	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 可持续发展：能够理解大数据应用工程实践中节能与绿色计算技术的手段和方法对环境和社会可持续发展的影响。	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，Xshell 和 xftp 的安装，虚拟机软件 VMware 的安装与使用，掌握 Hadoop 伪分布式和完全分布式的环境搭建。</p> <p>重点：Hadoop 伪分布式和完全分布式的环境搭建。</p> <p>难点：Hadoop 伪分布式和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：掌握 HDFS 的组成与架构：NameNode、DataNode、SecondaryNameNode 等，掌握工作流程。掌握机架感应、HDFS 文件读取写入流程、掌握 HDFS 数据容错相关处理机制。从 Web 访问、HDFS Shell、HDFS API 三个方面掌握 HDFS 操作。</p> <p>重点：掌握 HDFS 的组成与架构：NameNode、DataNode、SecondaryNameNode 等，掌握工作流程。从 Web 访问、HDFS Shell、HDFS API 三个方面掌握 HDFS 操作。</p> <p>难点：通过 HDFS Shell 和 HDFS API 访问 HDFS 的操作。</p>

第 4 单元 MapReduce 和 Yarn

主要内容：掌握 MapReduce 设计思想，掌握 MapReduce 编程模型，掌握 MapReduce 编程实例之 wordcount、统计各个部门员工薪水总和、序列化。了解 MapReduce 的数据输入、输出格式，掌握分区和合并的概念。掌握 MapReduce 作业运行机制，掌握 Shuffle 与排序，了解 YARN 调度组件，了解 FIFO 调度器，了解 Capacity 调度器和 Fair 调度器。

重点：掌握 MapReduce 编程模型，掌握 MapReduce 编程实例之 wordcount、统计各个部门员工薪水总和、序列化。了解 MapReduce 的数据输入、输出格式，掌握分区和合并的概念。掌握 MapReduce 作业运行机制，掌握 Shuffle 与排序。

难点：掌握 MapReduce 编程模型，掌握 MapReduce 编程实例之 wordcount、统计各个部门员工薪水总和、序列化。

第 5 单元 Zookeeper

主要内容：掌握 Zookeeper 的安装方法，理解 Zookeeper 的基本原理，理解 Zookeeper 的应用场景。

重点：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：掌握 Zookeeper 的安装方法，理解 Zookeeper 的应用场景。

第 6 单元 HBase 和 Hive

主要内容：掌握 HBase 的基本概念，掌握 HBase 的体系结构、数据模型、HBase 的安装与部署，掌握采用 HBase Shell 操作 HBase、掌握 JAVA API 操作 HBase，掌握 Hive 数据类型与存储格式，掌握 Hive 数据模型，掌握查询数据等。

重点：掌握 HBase 的基本概念，HBase 的安装与部署，掌握采用 HBase Shell 操作 HBase、掌握 JAVA API 操作 HBase，掌握 Hive 数据类型与存储格式和查询数据。

难点：HBase 的安装与部署，JAVA API 操作 HBase，Hive 的查询数据。

第 7 单元 Flume 和 Sqoop

主要内容：了解 Flume 的基本概念，掌握 Flume 安装，掌握 Flume 常用的配置与应用场景；了解 Sqoop 的基本概念，掌握 Sqoop 的安装，掌握 Sqoop 的应用场景。

重点：掌握 Flume 安装，掌握 Flume 常用的配置与应用场景；掌握 Sqoop 的安装，掌握 Sqoop 的应用场景。

难点：Flume 安装与应用场景；Sqoop 的安装，Sqoop 的应用场景。

第 8 单元 综合项目—电商精准营销

主要内容：项目详细介绍，包括概述、架构设计等，了解数据来源，掌握数据清洗流程，掌握数据仓库操作流程，掌握应用测试方法。

重点：项目模块分析，数据采集与清洗，使用数据仓库进行数据分析并进行可视化

难点：数据采集与清洗，数据分析和可视化

(二) 教学单元对课程目标的支撑关系

课程目标	1	2	3	4	5	6	7
	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 和 Yarn	√	√	√	√	√		
第 5 单元 Zookeeper	√	√	√	√			
第 6 单元 HBase 和 Hive	√	√	√	√	√		
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 和 Yarn	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	2	2	4
第 5 单元 Zookeeper	理论与上机实验讲授为主，辅之以 Zookeeper 的演示。	课后作业	2	0	2
第 6 单元 HBase 和 Hive	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	3	3	6
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	3	3	6
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。	项目报告	0	4	4
合计			16	16	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	实验一 Hadoop 环	Hadoop 环境搭建准备工作；Hadoop 伪分布	2	④

	境搭建	式和完全分布式的环境搭建。		
2	实验二 通过 API 和 Shell 访问 HDFS	Hadoop 集群的伪分布式和完全分布式的环境搭建关键技术点。	2	②
3	实验三 MapReduce 基础编程与 WordCount 程序	MapReduce 编程模型, MapReduce 编程实例之 wordcount、统计各个部门员工薪水总和、序列化。	2	②
4	实验四 HBase 基础编程	HBase 的安装与配置、Shell 访问与 Java API 访问	2	④
5	实验五 Hive 常用操作	基于 Local 模式的 Hive 环境搭建和常用操作	1	①
6	实验六 Flume 的安装与基本使用	Flume 的安装和使用, 掌握 Source、Sink、Channel 的使用方法, 掌握拦截器的用法	2	②
7	实验七 Sqoop 的安装与基本使用	Sqoop 的安装和使用, 掌握 import、export、job 命令的用法	1	①
8	实验八 电商精准营销	项目模块分析, 数据采集与清洗, 使用数据仓库进行数据分析, 可视化	4	④

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	30	项目报告		10	10	10	20	30	20	100

X2	20	平时出勤 课堂表现	10	10	10	10		30	30	100
X3	20	课后作业	20	20	10	10		20	20	100
X4	30	实验报告		20	20	20	20	10	10	100

六、其他需要说明的问题

无