

# 《大数据开发实践》课程教学大纲

课程英文名	Practice of Big Data Development				
课程代码	B0501680	课程类别	专业课	课程性质	专业选修
学分	3		总学时数	48	
开课学院	计算机学院		开课基层教学组织	智能信息处理课程组	
面向专业	计算机科学与技术、软件工程、智能计算与数据科学(计算机科学与技术)、智能财务(软件工程)、信息与计算科学		开课学期	第6学期	

注：课程类别是指学科基础课/专业课/实践环节/通识公共课/公共基础课/其他；

课程性质是指通识必修/通识选修/学科必修/专业必修/专业选修/实践必修/实践选修。

## 一、课程目标

当下是大数据时代，“大数据”需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产。《大数据开发实践》课程培养掌握数据科学的基础知识及大数据相关技术，掌握大数据清洗和分析常用工具的使用，具有卓越的实践能力，能胜任数据清洗、数据存储、数据分析与挖掘、大数据系统开发与构建等工作的能力；在了解大数据开发实践相关技术及应用的同时，结合国家建设和民族复兴的新时代背景，增强学生家国情怀与文化自信，激发学生使命感和责任心。

通过理论教学和实践活动，达到以下课程目标：

**课程目标 1：**能够运用大数据处理主流技术和平台 Hadoop、MapReduce 并行处理和编程技术。

**课程目标 2：**能够应用现代技术和工具，通过文献检索及研究，分析大数据开发技术及思路，并能针对大数据问题提出初步的解决建议。

**课程目标 3：**针对特定的性能要求，初步具备分析并设计大数据系统架构及功能目标能力。

**课程目标 4：**具备基本的科学素养，及时了解大数据处理的国内外新技术和发展趋势，及时掌握国家相关方面的科技战略需求，树立强烈的爱国主义使命感与责任心。

## 二、课程目标与毕业要求对应关系

课程的课程目标对计算机科学与技术专业毕业要求指标点的支撑情况如表 1 所示：

表 1 课程目标与毕业要求对应关系

毕业要求	指标点	课程目标及支撑权重

毕业要求 1：工程知识：掌握数学、自然科学、工程基础、计算机专业领域的知识，并能应用于计算机领域复杂工程问题的解决方案中。	1-4 掌握计算机系统、物联网、人工智能、大数据、网络安全等某个专业领域的知识，并用于解决计算机领域的复杂工程问题。	目标 1：0.3 目标 2：0.3 目标 3：0.2 目标 4：0.2
毕业要求 3：设计/开发解决方案：能够设计计算机领域复杂工程问题的解决方案，设计与开发满足特定需求的软硬件系统、算法或部件，在设计中考虑社会、健康、安全、法律、文化以及环境等因素，并体现一定的创新意识。	3-1 具备计算思维和程序设计能力，能够针对计算机复杂系统设计与开发满足特定需求的模块或算法。	目标 2：0.5 目标 3：0.5
	3-2 掌握计算机硬件基础理论和设计方法，能够针对计算机复杂系统设计满足特定需求的功能部件或硬件系统。	目标 2：0.4 目标 3：0.3 目标 4：0.3
毕业要求 5：使用现代工具：能够针对计算机领域的复杂工程问题，开发、选择与使用恰当的技术、资源、现代工程工具和信息技术工具，包括对计算机领域的复杂工程问题进行预测与模拟，并能够理解其局限性。	5-2 针对计算机领域的复杂工程问题，能够开发、选用符合特定需求的技术、资源和现代工具，实现分析、计算或设计，并进行模拟和预测。	目标 2：0.3 目标 3：0.3 目标 4：0.4

课程的课程目标对软件工程专业毕业要求指标点的支撑情况如表 2 所示：

表 2 课程目标与毕业要求对应关系

毕业要求	指标点	课程目标及支撑权重
毕业要求 1：工程知识：掌握数学、自然科学、工程基础、软件工程专业领域的知识，并能应用于软件工程领域复杂工程问题的解决方案中。	1-4 掌握某个专业领域知识，并用于解决软件工程领域复杂工程问题。	目标 1：0.3 目标 2：0.3 目标 3：0.2 目标 4：0.2
毕业要求 3：设计/开发解决方案：能够设计软件工程领域复杂工程问题的解决方案，设计与开发满足特定需求的软件系统、模块或算法，在设计中考虑社会、健康、安全、法律、文化以及环境等因素，并体现一定的创新意识。	3-1 具备计算思维和程序设计能力，能够针对复杂软件系统设计与开发满足特定需求的模块或算法。	目标 1：0.3 目标 2：0.3 目标 3：0.4
毕业要求 5：使用现代工具：能够针对软件工程领域复杂工程问题，开发、选择与使用恰当的技术、资源、现代工程工具和信息技术工具，包括对软件工程领域复杂工程问题进行预测与模拟，并能够理解其局限性。	5-2 针对软件工程领域复杂工程问题，能够开发、选用符合特定需求的技术、资源和现代工具，实现分析、计算或设计，并进行模拟和预测。	目标 2：0.5 目标 3：0.5

课程的课程目标对智能财务(软件工程)毕业要求指标点的支撑情况如表 3 所示：

表 3 课程目标与毕业要求对应关系

毕业要求	指标点	课程目标及支撑权重

毕业要求 1: 工程知识: 掌握数学、自然科学、工程基础、财务领域基础知识和软件工程专业领域知识, 并能应用在智能财务软件领域复杂工程问题的解决方案中。	1-4 掌握某个专业领域知识, 并用于解决智能财务软件工程领域复杂工程问题。	目标 1: 0.3 目标 2: 0.3 目标 3: 0.2 目标 4: 0.2
毕业要求 3: 设计/开发解决方案: 能够设计智能财务软件领域复杂工程问题的解决方案, 设计与开发满足特定需求的软件系统、模块或算法, 在设计中考虑社会、健康、安全、法律、文化以及环境等因素, 并体现一定的创新意识。	3-1 具备计算思维和程序设计能力, 能够针对复杂智能财务软件系统设计与开发满足特定需求的模块或算法	目标 1: 0.3 目标 2: 0.3 目标 3: 0.4
毕业要求 5: 使用现代工具: 能够针对智能财务软件领域复杂工程问题, 开发、选择与使用恰当的技术、资源、现代工程工具和信息技术工具, 包括对复杂工程问题的预测与模拟, 并能够理解其局限性。	5-2 针对智能财务软件工程领域复杂工程问题, 能够开发、选用符合特定需求的技术、资源和现代工具, 实现分析、计算或设计, 并进行模拟和预测。	目标 2: 0.5 目标 3: 0.5

课程的课程目标对智能计算与数据科学(计算机科学与技术)专业毕业要求指标点的支撑情况如表 4 所示:

表 4 课程目标与毕业要求对应关系

毕业要求	指标点	课程目标及支撑权重
1. 工程知识: 能够掌握数学、自然科学、工程基础、计算机软硬件知识、人工智能、智能计算和数据科学的基础理论及专业知识, 并应用在人工智能、智能计算和大数据专业领域及其他相关交叉领域的复杂工程问题的解决方案中。	1-4 掌握人工智能、智能计算、大数据等某个专业领域的知识, 并用于解决计算机领域的复杂工程问题。	目标 1: 0.3 目标 2: 0.3 目标 3: 0.2 目标 4: 0.2
3. 设计/开发解决方案: 能够设计人工智能、智能计算和大数据专业领域及其他相关交叉领域复杂工程问题的解决方案, 能够设计与开发满足特定需求的计算机软硬件系统、模型或算法, 在设计中考虑社会、健康、安全、法律、文化以及环境等因素, 并体现一定的创新意识。	3-1 具备计算思维和程序设计能力, 能够针对计算机复杂系统设计与开发满足特定需求的模块或算法。 3-2 掌握计算机硬件基础理论和设计方法, 能够针对计算机复杂系统设计满足特定需求的功能部件或硬件系统。	目标 1: 0.3 目标 2: 0.3 目标 3: 0.4  目标 2: 0.4 目标 3: 0.3 目标 4: 0.3
5. 使用现代工具: 能够针对复杂工程问题, 开发、选择与使用恰当的技术、资源、现代工程工具和信息技术工具, 包括对复杂工程问题的预测与模拟, 并能够理解其局限性。	5-2 针对计算机领域的复杂工程问题, 能够开发、选用符合特定需求的技术、资源和现代工具, 实现分析、计算或设计, 并进行模拟和预测。	目标 2: 0.5 目标 3: 0.5

### 三、课程目标与教学内容和方法的对应关系

《大数据开发实践》课程目标与教学内容、建议教学方法的对应关系如表 5 所示。教师可根据实际情况和自身特长采用合适的教学方法。

表 5 课程目标与教学内容、教学方法的对应关系

教学内容	教学方法	课程目标
1. 大数据处理技术简介	课堂讲授、课堂讨论	<b>2,4</b>
2. 数据采集与清洗-Python 爬虫工具	课堂讲授、课堂讨论、课堂实验	<b>1,2,3</b>
3. 数据采集与清洗-Flume、Kafka、Sqoop	课堂讲授、课堂讨论、课堂实验	<b>1,2</b>
4. 数据存储与管理—分布式文件系统 HDFS	课堂讲授、课堂讨论、课堂实验	<b>1,2</b>
5. 数据存储与管理-NoSQL 数据库（HBase、MongoDB）	课堂讲授、课堂讨论、课堂实验	<b>1,2,3,4</b>
6. 数据处理与分析- MapReduce 并行编程框架	课堂讲授、课堂讨论、课堂实验	<b>1,2</b>
7. 数据处理与分析- Spark 内存计算框架	课堂讲授、课堂讨论、课堂实验	<b>2,3</b>
8. 数据处理与分析- Hive 多维数据分析	课堂讲授、课堂讨论、课堂实验	<b>1,2</b>
9. 数据可视化展示-ECharts、D3	课堂讲授、课堂讨论、课堂实验	<b>1,2</b>
10. 资源管理与调度-YARN、Zookeeper	课堂讲授、课堂讨论、课堂实验	<b>2,3</b>

该课程详细教学内容和方法如下所：

#### 1. 大数据处理技术简介

##### (1) 教学内容：

- 大数据发展历程、主要特征；
- 大数据处理技术简介；
- 大数据处理流程简介；
- 大数据与云计算、人工智能、物联网。

(2) 教学重点：大数据处理技术有哪些，以及大数据的处理流程。

(3) 教学难点：大数据与云计算、人工智能、物联网这些新兴概念之间的关系。

(4) 教学要求：通过本小节学习，掌握大数据发展历程和主要特征，了解大数据处理技术及处理流程。

#### 2. 数据采集与清洗-Python 爬虫工具

**(1)教学内容:**

- 数据采集与清洗作用;
- 开放 API 接口、第三方库使用;
- Requests 库、BeautifulSoup 库使用;
- Scrapy 框架使用。

**(2)教学重点:** 学习第三方财经库 Tushare 库的使用，学习爬虫中 Requests 库、BeautifulSoup 库使用。

**(3)教学难点:** HTTP 协议消息传递的原理及利用 Requests 库、BeautifulSoup 库抓取网页内容的过程。

**(4)教学要求:** 通过本小节学习，了解数据采集与清洗的作用，并掌握网络数据获取的几种方式。除此之外，完成对应的上机实验，即 Tushare 库使用、python 爬虫实验。

### 3. 数据采集与清洗-Flume、Kafka、Sqoop

**(1) 教学内容:**

- 海量日志采集系统 Flume;
- 分布式发布订阅消息系统 Kafka;
- 数据迁移工具 Sqoop;

**(2)教学重点:** 学习完上一节学习的网络数据采集后，继续学习其他数据采集工具，包括 日志数据采集 Flume、数据分发中间件 Kafka、数据导入工具 Sqoop。

**(3)教学难点:** Flume 的基本架构及工作原理、Kafka 的工作原理及应用场景、Sqoop 的工作原理及导入导出流程。

**(4)教学要求:** 通过本小节学习，了解 Flume、Kafka、Sqoop 三种不同的数据采集工具的原理和使用。除此之外，完成对应的上机实验，即海量日志采集系统 Flume、分布式发布订阅消息系统 Kafka、数据迁移工具 Sqoop。

**思政融合点 1:** 引导学生查阅资料，查找学习的数据采集与清洗软件国产化替代产品。通过对国内外数据采集与清洗技术的差距，激发学生爱国主义使命感、责任心和学习先进技术的热情。

### 4. 数据存储与管理—分布式文件系统 HDFS

**(1)教学内容:**

- HDFS 的基本特征与构架;
- HDFS 可靠性设计;
- HDFS 文件存储组织与读写;
- HDFS 文件系统操作命令;

- HDFS 基本编程接口与示例。

(2) **教学重点:** HDFS 的设计目标和基本概念; HDFS 的体系结构; HDFS 的运行机制; HDFS 文件系统的命令操作。

(3) **教学难点:** HDFS 写数据和读数据的过程; HDFS 编程接口和示例。

(4) **教学要求:** 通过本小节学习, 掌握 HDFS 的特征和设计目标, 了解 HDFS 中数据流读写的运行机制。除此之外, 完成对应的上机实验, 即分布式文件系统 HDFS。

## 5. 数据存储与管理—NoSQL 数据库 (HBase、MongoDB)

(1) **教学内容:**

- NoSQL 数据库起源;
- NoSQL 数据库与关系型数据库的区别;
- NoSQL 数据库类别、原理;
- 列式数据库 HBase;
- 文档数据库 MongoDB。

(2) **教学重点:** HDFS 的设计目标和基本概念; HDFS 的体系结构; HDFS 的运行机制; HDFS 文件系统的命令操作。

(3) **教学难点:** HDFS 写数据和读数据的过程; HDFS 编程接口和示例。

(4) **教学要求:** 通过本小节学习, 掌握 HDFS 的特征和设计目标, 了解 HDFS 中数据流读写的运行机制。除此之外, 完成对应的上机实验, 即分布式文件系统 HDFS。

## 6. 数据处理与分析- MapReduce 并行编程框架

(1) **教学内容:**

- MapReduce 基本编程模型和框架;
- Hadoop MapReduce 基本框架与工作过程;
- Hadoop MapReduce 主要组件与编程接口。

(2) **教学重点:** Hadoop MapReduce 基本框架与工作过程; Hadoop MapReduce 主要组件与编程接口。

(3) **教学难点:** MapReduce 的设计思想、系统架构; MapReduce 的程序设计方法, 能够编写 Mapper 类和 Reducer 类; MapReduce 在集群上的运作; Hadoop MapReduce 主要组件与编程接口。

(4) **教学要求:** 通过本小节学习, 能够编写 MapReduce 的 Mapper 类和 Reducer 类。能够在集群下运作 MapReduce 项目。能够将 HDFS 文件系统中的文件作为输入开发 MapReduce 项目。除此之外, 完成对应的上机实验, 即 MapReduce 数据统计分析。

## 7. 数据处理与分析- Spark 内存计算框架

(1)教学内容:

- Spark 简介包括出现的原因;
- Spark 生态圈;
- Spark 工作流程。

(2)教学重点: Spark 简介包括与工作流程包括基本概念、架构设计、RDD 算子及运行流程。

(3)教学难点: Spark 工作流程包括与 MapReduce 的对比、SparkSQL、Spark Streaming、GraphX、MLlib 等使用。

(4)教学要求: 通过本小节学习, 掌握 Spark 的基本概念、生态圈及使用场景。特别地, 了解 Spark 的工作流程, 对 RDD 的概念进行深入理解。除此之外, 完成对应的上机实验, 即 Spark 案例分析。

## 8. 数据处理与分析- Hive 多维数据分析

(1)教学内容:

- 了解 Hive 的作用与结构组成;
- Hive 的数据模型;
- Hive 的安装;
- Hive 的查询语言—HiveQL;
- Hive JDBC 编程接口与功能增强。

(2)教学重点: Hive 的数据模型; Hive 的查询语言-HiveQL; Hive JDBC 编程接口与功能增强。

(3)教学难点: HiveQL 查询语言的编程接口。

(4)教学要求: 通过本小节学习, 能够使用 Hive 的数据模型, 通过 HiveQL 查询语言的编程接口实现数据查询分析。除此之外, 完成对应的上机实验, 即大数据仓库 Hive。

**思政融合点 2:** 引导学生了解大数据处理技术在的各种框架及其应用场景, 帮助学生以辩证的、战略的思维理解科技第一生产力对于国家建设与民族复兴的重要意义。

## 9. 数据可视化展示-ECharts、D3

(1)教学内容:

- 高维数据可视化
- 网络和层次化数据可视化
- 时空数据可视化
- 文本数据可视化
- 数据可视化工具: ECharts、D3。

(2) 教学重点：数据可视化的概念，每类数据常用的可视化形式。

(3) 教学难点：每类数据可视化的基本原理及意义、ECharts、D3 的使用。

(4) 教学要求：通过本小节学习，了解数据可视化的概念和作用，并掌握常见的四类数据的可视化形式。除此之外，完成对应的上机实验，即 Echarts 数据可视化。

**思政融合点 3：**数据可视化目前已成为计算机领域的热门研究课题，介绍数据可视化在显示生活中的应用，激发学生对于参与到前沿技术研究的自信感和热情。

## 10. 资源管理与调度-YARN、Zookeeper

(1) 教学内容：

- YARN 的由来；
- YARN 的体系架构；
- Zookeeper 概述；
- Zookeeper 数据模型；
- Zookeeper 服务架构。

(2) 教学重点：YARN 的由来及其功能，Zookeeper 数据模型及原理。

(3) 教学难点：YARN 的体系架构及 Zookeeper 的数据模型。

(4) 教学要求：通过本小节学习，了解数 YARN 的由来、作用及体系架构，同时掌握 Zookeeper 的数据模型及服务架构。

**思政融合点 4：**引导学生完成课前课后自主学习任务，培养责任心和终身学习能力；设计解决方案并完成编程项目，精益求精改进方案及程序性能，培养工匠精神。

**思政融合点 5：**总结课堂内容，让学生了解大数据相关技术在各个领域的实际应用，强调动手能力对于解决实际问题的重大意义，鼓励学生通过现代化手段解决各类实际应用问题。

## 四、 实践环节及要求

### 1. 实验项目和基本要求

表 6 实验项目和基本要求

序号	实验项目	学时	基本要求	实验性质	实验类别
1	Tushare 库使用	3	熟悉 Tushare 财经库的使用，包括基本方法等	必做	综合
2	python 爬虫实验	3	掌握 python 爬虫，包括 Requests 和 BeautifulSoup 库的使用	必做	综合
3	Flume、Kafka、Sqoop 使用	3	熟悉海量日志采集系统 Flume、分布式发布订阅消息系统 Kafka、数据迁移工具 Sqoop 的使用	必做	综合

4	分布式文件系统 HDFS	3	熟悉 HDFS 编程实践，特别是 HDFS 读写文件	必做	综合
5	MapReduce 数据统计分析	3	熟悉 MapReduce 编程实践，也别是 map 函数和 reduce 函数的撰写	必做	综合
6	Spark 案例分析	3	熟悉 Spark 案例分析，包括 Scala 语言使用	必做	综合
7	大数据仓库 Hive	3	熟悉 Hive 进行大数据分析的流程	必做	综合
8	Echarts 数据可视化	3	掌握 pycharts 库，对一些常见的可视化算法进行数据	必做	综合

## 2. 实验报告基本要求

实验报告至少包含以下几个部分：(1)实验目的；(2)实验环境；(3)实验过程(含实验方案、步骤、程序等)；(4)实验结果及结果分析；(5)实验总结。

## 五、与其它课程的联系

**先修课程：**程序设计基础、数据结构、大数据基础。

**后续课程：**大数据实用案例与分析

## 六、学时分配

总学时 48 学时，其中讲课 24 学时，课内上机 24 学时。如表 7 所示。

表 7《大数据开发实践》学时分配表

教学内容	讲课时数	实验时数	实践学时	课内上机时数	课外上机时数	自学时数	习题课时	讨论时数
1. 大数据处理技术简介	2							
2. 数据采集与清洗-Python 爬虫工具	2			3		2		
3. 数据采集与清洗-Flume、Kafka、Sqoop	2			3				
4. 数据存储与管理—分布式文件系统 HDFS	3			3		2		
5. 数据存储与管理-NoSQL 数据库（HBase、MongoDB）	3			3				
6. 数据处理与分析- MapReduce 并行编程框架	3			3		2		
7. 数据处理与分析- Spark 内存计算框架	3			3				
8. 数据处理与分析- Hive 多维数据分析	2			3		2		
9. 数据可视化展示-ECharts、D3	2			3				

10.资源管理与调度-YARN、Zookeeper	2						
合 计	24			24	8		
总 计	48						

## 七、课程目标达成途径及学生成绩评定方法

### 1.课程目标达成途径

课程目标的达成途径如表 8 所示。

表 8 课程目标与达成途径

课程目标	达成途径
<b>课程目标 1:</b> 能够运用大数据处理主流技术和平台 Hadoop、MapReduce 并行处理和编程技术。	以引导式、启发式和总结式教学方法为主，通过重点/难点内容讲解、课后作业、进行随堂提问、课堂程序演示等模式，帮助学生能够运用大数据分析技术分析实际应用案例。
<b>课程目标 2:</b> 能够应用现代技术和工具，通过文献检索及研究，分析大数据开发技术及思路，并能针对大数据问题提出初步的解决建议。	以启发式、分析式和研讨式教学方法为主，结合典型案例，通过上机实践、作业、随堂提问、课堂讨论等模式，帮助学生具备运用大数据建模和分析的思维方式，为实际应用问题建立问题模型，具备数据分析的能力。
<b>课程目标 3:</b> 针对特定的性能要求，初步具备分析并设计大数据系统架构及功能目标能力。	以引导式、启发式和总结式教学方法为主，通过重点/难点内容讲解、上机实践、课后作业、进行随堂提问、课堂讨论等模式，帮助学生掌握大数据的新技术与发展趋势，具备分析并设计大数据系统架构及功能目标能力。
<b>课程目标 4:</b> 具备基本的科学素养，及时了解计算机体系结构的国内外新技术和发展趋势，及时掌握国家相关方面的科技战略需求，树立强烈的爱国主义使命感与责任心。	通过课堂讲授、课后自学、文献查阅、课堂讨论、分析对比、总结报告等各种方式，让学生对大数据技术及其应用有了解、掌握及实际运用解决问题的能力，建立终生学习的意识；同时，进一步了解目前国内大数据先进技术与取得的成就，从而建立强烈的民族自豪感与爱国主义使命感。

### 2.学生成绩评定方法

该课程为考查课程，考查方式为期末提交学生作品。该课程采用形成性评价与终结性评价相结合的评价方法，学期总评成绩由二部分构成：平时成绩，占比 50%，包括课程思政实践、课堂成绩、课后作业、实验操作及实验报告等；期末考试成绩，占比 50%。各部分的具体评价环节、关联课程目标、评价依据及方法和在总成绩中的占比，如表 9 所示。

表 9 课程考核与成绩评定方法

成绩构成	考核项目	考核关联的课程目标	考核依据与方法	占总评成绩的比重
平时成绩	课程思政实践	4	基于大数据应用主题，通过课外文献查阅、课堂展示、课堂小组讨论、阅读报告等	5%

			多种形式，考查学生对我国大数据技术及其应用的了解情况以及核心价值观状况。	
	课堂成绩	1,2,3	考察考勤、课堂讨论、课堂互动、课堂练习等。	15%
	课后作业	1,2,3,4	考察作业是否及时完成、作业完成质量。	10%
	实验操作及实验报告	1,2,3,4	布置大数据技术及在各个领域的开发应用等方面的新技术相关主题的若干个实验，提交实验报告。	20%
期末成绩	期末考查	1,2,3,4	考查学生提交的期末作品的成绩。	50%
总评成绩		1,2,3,4	=平时成绩*50%+期末成绩*50%	100%

表 10. 考核内容详细评分标准

考核内容	评分标准			
	90-100	75-90	60-75	<60
课程思政实践	报告条理清晰，文字流畅，字数 $\geq$ 4000，参考文献数量 $\geq$ 8 且相关性强；内容完整且材料丰富，体现强烈的使命感、责任心与民族自豪感	报告条理清楚，字数 $\geq$ 3000，参考文献数量 $\geq$ 5 且相关性较好；内容完整，材料不够丰富，能体现学生的使命感、责任心与民族自豪感	报告有一定条理，字数 $\geq$ 1000，参考文献数量 $\geq$ 2 且基本相关；内容基本完整但材料较少，能体现学生的使命感与民族自豪感	报告字数 $<$ 1000，参考文献数量 $<$ 2；内容少，或有抄袭现象，体现不出学生的使命感与民族自豪感
课堂成绩	雨课堂测试、课堂练习、回答问题等据实评价；或参与回答次数在教学班前 15%	雨课堂测试、课堂练习、回答问题等据实评价；或参与回答次数在教学班前 50%	雨课堂测试、课堂练习、回答问题等据实评价；或参与回答次数在教学班前 85%	雨课堂测试、课堂练习、回答问题等据实评价；或参与回答次数在教学班后 15%
课后作业	非标作业：方案等设计合理，分析准确，能满足问题全部要求	非标讨论题：方案较合理，分析较正确，能基本满足问题全部要求	非标讨论题：方案基本合理，能满足问题大部分要求	非标讨论题：方案不够合理，只能满足问题少量要求
标准题目：按照作业题目评分标准据实评价				
实验操作及实验报告	程序运行流畅，功能完善，性能好；代码独创性好；算法与数据结构设计或代码实现等有 2 处及以上创新点	程序运行正常，功能基本实现，但不够完善、优化等，代码自创率应高于 50%；算法与数据结构设计或代码实现等至少有一处创新点	程序运行基本正常，有少量 bug；功能实现至少达到 80%，代码自创率应高于 30%；基本没有创新性	程序运行 bug 多，功能实现低于要求的 80%
期末考查	报告条理清晰，文字流畅内容完整，字数 $\geq$ 3500，数据结构及算	报告条理清楚，内容较完整，字数 $\geq$ 2500，数据结构及算法设计	报告内容基本完整，字数 $\geq$ 1500，数据结构及算法设计基本合	报告内容不完整，字数少于 1500，数据结构及算法设计不能

	法设计合理且效率高，有 2 个以上创新点或改进，参考文献 ≥5 篇，查重率≤20%，汇报 PPT 图表清晰，设计美观，答辩过程脱稿讲解，分析条理清晰，问题回答准确，小组协作好	合理且性能较好，有 1 个以上创新点或改进，参考文献≥3 篇，查重率≤35%，汇报 PPT 设计较美观，答辩过程脱稿讲解，条理较清楚，问题回答基本正确，分工较合理	理，参考文献≥2 篇，查重率≤50%，汇报 PPT 美观性及内容一般，答辩过程大部分内容脱稿讲解，大部分问题能基本正确回答，分工基本合理	满足题目基本要求，查重率>60%，有抄袭现象，汇报 PPT 界面文字多，内容不完整，答辩过程基本念 PPT，只能回答少量问题，分工不够合理
--	---	---	--	---

## 八、 教学资源

表 11 课程的基本教学资源

资源类型	资源
教材	黄宜华, 苗凯翔编:《深入理解大数据:大数据处理与编程实践》, 机械工业出版社, 2014。
参考书籍或文献	1. 林子雨:《大数据基础编程、实验和案例教程》 2. 陆价恒等:《Hadoop 实战(第 2 版)》, 机械工业出版社, 2014。 3. 彭作文等:《大数据分行业大解析》, 中国铁道出版社, 2016。 4. 赵敏哲:《64 位 Linux 操作系统与应用实例》, 机械工业出版社 2000。 5. Daniel P.Bovet & Marco Cesati 著, 陈莉君等译:《深入理解 Linux 内核》, 中国电力出版社, 2007。
教学文档	无

## 九、 课程目标、毕业要求指标点达成度定量评价

在课程结束后，需要对每一个课程目标(含思政课程目标)进行达成度的定量评价，用以实现课程的持续改进。

课程目标达成度的定量评价算法：

- 1、 使用教学活动(如课程思政实践、课后作业、课堂表现、文献阅读报告、期末作品等等)成绩或期末考试部分题目得分率作为评价项目，来对某个课程目标进行达成度的定量评价；
- 2、 为保证考核的全面性和可靠性，要求对每一个课程目标的评价项目选择超过两种；
- 3、 根据施教情况，评价项目可以由教师自行扩展，权重比例可以由教师自行设计；
- 4、 对某一个课程目标有支撑的各评价项目权重之和为 1；
- 5、 使用所有学生(含不及格)的平均成绩计算。

表 12. 课程目标达成度定量评价方法

课程目标	课程目标达成度评价方式
课程目标 1：能够运用大数据处理主流技术和平台 Hadoop、MapReduce 并行处理和编程技术。	课程思政实践: 0.2 课堂成绩: 0.4 期末考查: 0.4
课程目标 2：能够应用现代技术和工具，通过文献检索及研究，分析大数据开发技术及思路，并能针对大数据问题提出初步的解决建议。	课堂成绩: 0.4 实验操作及实验报告: 0.4 期末考查: 0.2
课程目标 3：针对特定的性能要求，初步具备分析并设计大数据系统架构及功能目标能力。	课堂成绩: 0.2 实验操作及实验报告: 0.4

课程目标	课程目标达成度评价方式
	期末考查: 0.4
<b>课程目标 4:</b> 具备基本的科学素养, 及时了解计算机体系结构的国内外新技术和发展趋势, 及时掌握国家相关方面的科技战略需求, 树立强烈的爱国主义使命感与责任心。	课程思政实践: 0.2 课堂成绩: 0.2 实验操作及实验报告: 0.3 期末考查: 0.3

## 十、说明

本课程大纲主要用于规范计算机科学与技术、软件工程、智能财务(软件工程)、信息与计算科学、智能计算与数据科学(计算机科学与技术)专业的《大数据开发实践》课程的教学目标、教学内容、教学方法、教学要求以及考核评价方法等, 承担该课程的教师必须遵照本大纲安排授课计划、实施教学过程, 完成学生各个阶段与各方面的学习成果考核与评价; 在学期末, 需对课程目标和课程支撑的毕业要求指标点进行达成度评价。

本课程大纲于 2021 级开始执行, 生效之日原先版本均不再使用。

## 十一、编制与审核

表 13 大纲编制与审核信息

工作内容	责任部门或机构	负责人	完成时间
编制	智能信息处理组	俞东进、陈洁、孙笑笑	2022.03
审核	智能信息处理组	彭勇	2022.03
审定	计算机学院教学工作委员会	黄孝喜	2022.05