

《Hadoop 大数据技术》课程标准

一、管理信息

课程名称：Hadoop 大数据技术	批准人：
课程代码：	所属分院：广州城建职业学院
制定人：宋曼	制定时间：2019.9.1

二、基本信息

学 分：4	课程类型：限选课
学 时：64	授课对象：17 软件技术专业

三、课程设计

（一）课程目标设计

通过本课程的学习，让学生掌握大数据基本概念、大数据处理架构Hadoop、分布式文件系统HDFS、分布式数据库HBase、NoSQL数据库、云数据库、分布式并行编程模型MapReduce、大数据处理架构Spark和数据采集工具的基础编程和应用，为大数据领域深耕细作奠定基础、指明方向。

1. 知识目标

- (1) 了解大数据关键技术和大数据相关软件的用途。
- (2) 掌握 Linux 系统和相关软件的安装，掌握 Linux Shell 命令。
- (3) 掌握 Hadoop 和相关软件的安装，学会在伪分布模式下运行 Hadoop 代码。
- (4) 掌握 HDFS 操作常用 Shell 命令，掌握 HDFS 编程实践。
- (5) 掌握 HBase 的安装和配置过程，掌握 HBase 常用 Shell 命令和 HBase 编程实践。
- (6) 掌握典型 NoSQL 数据库 Redis 和 MongoDB 安装使用过程。
- (7) 能够使用 MapReduce 编写词频统计任务并进行程序的编译打包和运行。
- (8) 掌握 Hive 的安装和基本操作和 Hive 的应用实例。
- (9) 掌握 Spark 的安装和配置和 Spark Shell 编程实践，能够编写 Spark 独立程序。
- (10) 掌握典型可视化工具 Easel.ly、D3、Tableau、魔镜、Echarts 的安装和使用。
- (11) 掌握数据采集工具 Flume、Kafka、Sqoop 的安装和使用。

2. 能力目标

- (1) 能够建立对大数据知识体系的轮廓性认识，了解大数据发展历程、基本概念、主要影响、应用领域、关键技术、计算模式和产业发展，并了解云计算、物联网的概念及其与大数据之间的紧密关系；
- (2) 能够了解Hadoop的发展历史、重要特性和应用现状，Hadoop项目结构及其各个组件，

并熟练掌握Hadoop平台的安装和使用方法；

(3) 能够了解分布式文件系统的基本概念、结构和设计需求，掌握Hadoop分布式文件系统HDFS的重要概念、体系结构、存储原理和读写过程，并熟练掌握分布式文件系统HDFS的使用方法；

(4) 能够了解分布式数据库HBase的访问接口、数据模型、实现原理和运行机制，并熟练掌握HBase的使用方法；

(5) 能够了解NoSQL数据库与传统的关系数据库的差异、NoSQL数据库的四大类型以及NoSQL数据库的三大基石；基本掌握Redis、MongoDB等NoSQL数据库的使用方法；

(6) 能够了解云数据库的概念、基本原理和代表性产品的使用方法；

(7) 能够熟练掌握分布式编程框架MapReduce的基本原理和编程方法；

(8) 能够了解大数据处理架构Spark的基本原理和编程方法；

3. 素质目标

(1) 养成善于思考、深入研究的良好自主学习的习惯和创新精神；

(2) 培养结构化程序设计思想和良好的编码规范；

(3) 培养细致缜密的工作态度、团结协作的良好品质、沟通交流和书面表达能力；

(4) 养成爱岗敬业、遵守职业道德规范、诚实、守信的高尚品质。

(二) 课程内容设计

本课程以“构建知识体系、阐明基本原理、开展初级实践、了解相关应用”为原则，围绕主流的大数据技术开展实践，主要内容以基础编程和案例教程为主。具体内容设计如下：

1. 课程内容总体设计

序号	单元名称	主要内容	学时安排
1	大数据技术概述	大数据关键技术、大数据相关软件	2
2	Linux 系统的安装和使用	Linux 系统安装、Linux 系统和相关软件、Shell 命令、综合实训 1：熟悉常用 Linux 操作	6
3	Hadoop 的安装和使用	Hadoop 简介、Hadoop 安装前期准备、Hadoop 安装、伪分布模式下 Hadoop 运行过程	4
4	HDFS 操作方法和基础编程	HDFS 操作常用 Shell 命令、HDFS 编程实践、综合实训 2：熟悉常用 HDFS 操作	6
5	HBase 的安装和基础编程	HBase 的安装和配置、HBase 常用 Shell 命令、HBase 编程实践、综合实训 3：熟悉常用 HBase 操作	6
6	典型 NOSQL 数据库的安装和使用	Redis 安装和使用、MongoDB 安装和使用、综合实训 4：NoSQL 和关系数据库的操作比较	6
7	MapReduce 基础编程	词频统计任务的编写、编译打包程序、综合实训 5：MapReduce 初级编程实践	6
8	数据仓库 Hive 的安装和使用	Hive 的安装和基本操作、Hive 应用实例	4
9	Spark 的安装和基础编程	Spark 的安装和配置、Spark Shell 编程、编写 Spark 独立程序	4

10	典型的可视化工具的使用方法	Easel.ly 信息图制作、D3 可视化库的使用、可视化工具 Tableau 的使用、使用“魔镜”制作图表、使用 Echarts 制作图表	6
11	数据采集工具的安装和使用	Flume 工具、Kafka 工具、Sqoop 工具、编程实例	4
12	大数据课程综合实训案例	需求分析、环境搭建、本地数据集上传到 Hive、Hive 数据分析、Hive、MySQL、HBase 数据互导、利用 R 进行数据可视化分析	10
总计			64

2. 课程内容详细设计

序号	单元名称	单元内容	知识目标	学时安排
1	大数据技术概述	单元 01: 大数据技术概述	了解大数据的关键技术; 掌握大数据相关软件;	2
		小计		2
2	Linux 系统的安装和使用	单元 02: Linux 系统安装	掌握Linux系统安装; 学会使用Linux系统和相关软件; 掌握基础的Linux Shell命令; 掌握综合实训1: 熟悉常用Linux操作	2
		单元 03: Linux 系统 Shell 命令		2
		单元 04: 综合实训 1: 熟悉常用 Linux 操作		2
		小计		6
3	Hadoop 的安装和使用	单元 05: Hadoop 安装	了解Hadoop历史和用途; 能够完成Hadoop和其他配套软件的安装; 能够完成Hadoop伪分布模式配置, 并编写命令, 测试成功;	2
		单元 06: 伪分布模式下 Hadoop 的运行		2
		小计		4
4	HDFS 操作方法和基础编程	单元 07: HDFS 操作常用 Shell 命令	掌握HDFS操作常用Shell命令; 掌握HDFS编程实践; 掌握综合实训2: 熟悉常用HDFS操作;	2
		单元 08: HDFS 编程实践		2
		单元 09: 综合实训 2: 熟悉常用 HDFS 操作		2
		小计		6
5	HBase 的安装和基础编程	单元 10: HBase 的安装和配置	能够完成HBase的安装和配置; 掌握HBase常用Shell命令; 掌握HBase编程实践; 掌握综合实训3: 熟悉常用HBase操作	2
		单元 11: HBase 编程实践		2
		单元 12: 综合实训		2

		3: 熟悉常用 HBase 操作		
		小计		6
6	典型 NOSQL 数据库的安装和使用	单元 13: Redis 安装和使用	掌握Redis安装和使用; 掌握MongoDB安装和使用;	2
		单元 14: MongoDB 安装和使用	掌握综合实训4: NoSQL和关系数据库的操作比较	2
		单元 15: 综合实训 4: NoSQL 和关系数据库的操作比较		2
		小计		6
7	MapReduce 基础编程	单元 16 : MapReduce 编程 (1)	能够使用MapReduce编写词频统计任务的编写; 能够在Eclipse下运行java程序, 并进行程序编译、打包、在MapReduce下运行; 掌握综合实训5: MapReduce初级编程实践	2
		单元 17 : MapReduce 编程 (2)		2
		单元 18: 综合实训 5: MapReduce 初级编程实践		2
		小计		6
8	数据仓库 Hive 的安装和使用	单元 19: Hive 的安装和基本操作	掌握Hive的安装和基本操作; 掌握Hive应用实例;	2
		单元 20: Hive 应用实例		2
		小计		4
9	Spark 的安装和基础编程	单元 21: Spark 的安装和配置	掌握Spark的安装和配置; 掌握Spark Shell编程; 能够编写Spark独立程序;	2
		单元 22: 编写 Spark 独立程序		2
		小计		4
10	典型的可视化工具的使用方法	单元 23: Easel.ly 信息图和 D3 可视化库	掌握Easel.ly信息图制作; 掌握D3可视化库的使用; 掌握可视化工具Tableau的使用; 学会使用“魔镜”制作图表; 学会使用Echarts制作图表;	2
		单元 24: 可视化工具 Tableau 的使用		2
		单元 25: 实验“魔镜”和 Echarts 制作图表		2
		小计		6
11	数据采集工具的安装和使用	单元 26: Flume 工具和 Kafka 工具	掌握Flume工具的使用; 掌握Kafka工具的使用; 掌握Sqoop工具的使用; 能完成编程实例;	2
		单元 27: Sqoop 工具和编程实践		2

		小计		4
12	大数据课程综合实训案例	单元 28：需求分析和环境搭建	能够进行综合实训案例的需求分析和环境搭建； 完成本地数据集上传到Hive； 完成Hive数据分析； 完成Hive、MySQL、HBase数据互导；能够利用R进行数据可视化分析；	2
		单元 29：上传本地数据集		2
		单元 30：Hive 数据分析		2
		单元 31：数据互导		2
		单元 32：数据可视化分析		2
		小计		10
总计				64

(三) 能力训练设计

大单元名称	单元	能力目标	能力目标描述	知识目标	考核	学时
大数据技术概述	单元 01：大数据技术概述	能够列出大数据关键技术和分类	能够列出大数据关键技术和分类	了解大数据的关键技术；掌握大数据相关软件；	程序展示	2
Linux系统的安装和使用	单元 02：Linux 系统安装	能使用Linux系统	能够安装Linux系统和相关软件	掌握Linux系统安装；学会使用Linux系统和相关软件；	程序展示	2
	单元 03：Linux 系统Shell命令		能够编写基础Linux Shell命令	掌握基础的Linux Shell命令；	程序展示	2
	单元 04：综合实训 1：熟悉常用Linux操作		能够熟练操作Linux Shell命令；	掌握综合实训 1：熟悉常用Linux操作	程序展示	2
Hadoop的安装和使用	单元 05：Hadoop 安装	能安装Hadoop软件并进行编程	能够完成Hadoop和其他配套软件的安装；	了解Hadoop历史和用途；掌握Hadoop和其他配套软件的安装；	程序展示	2
	单元 06：伪分布模式下Hadoop的运行		能够完成Hadoop伪分布模式配置，并编写命令，测试成功；	掌握 Hadoop 伪分布模式配置和命令编写；	程序展示	2
HDFS操作方法和基础编程	单元 07：HDFS 操作常用Shell命令	能编写HDFS的shell命令	能够编写HDFS的shell命令	掌握HDFS操作常用Shell命令；	程序展示	2
	单元 08：		能够进行HDFS编程实践	掌握HDFS编程实践；	程序	2

	HDFS 编程实践				展示	
	单元 09：综合实训 2：熟悉常用 HDFS 操作		能够进行基本的HDFS操作	掌握综合实训 2：熟悉常用 HDFS 操作；	程序展示	2
HBase 的安装和基础编程	单元 10：HBase 的安装和配置	能完成 HBase 的安装和编程	能够完成HBase的安装和配置	掌握HBase的安装和配置；掌握HBase常用Shell命令；	程序展示	2
	单元 11：HBase 编程实践		能够编写HBase常用Shell命令	掌握HBase编程实践；	程序展示	2
	单元 12：综合实训 3：熟悉常用 HBase 操作		能够进行HBase编程实践	掌握综合实训 3：熟悉常用 HBase 操作	程序展示	2
典型 NOSQL 数据库的安装和使用	单元 13：Redis 安装和使用	能进行典型数据编程	能够进行Redis数据库的安装和编程实践	掌握Redis安装和使用；	程序展示	2
	单元 14：MongoDB 安装和使用		能够进行MongoDB数据库的安装和编程实践	掌握MongoDB安装和使用；	程序展示	2
	单元 15：综合实训 4：NoSQL 和关系数据库的操作比较		能够熟练进行Redis和MongoDB数据库编程实践	掌握综合实训 4：NoSQL 和关系数据库的操作比较；	程序展示	2
MapReduce基础编程	单元 16：MapReduce 编程（1）	能进行 MapReduce编程	能够使用MapReduce编写词频统计任务的编写；	掌握MapReduce的处理单元Map单元和Reduce单元的编程；	程序展示	2
	单元 17：MapReduce 编程（2）		能够在Eclipse下运行java程序，并进行程序编译、打包、在MapReduce下运行；	掌握MapReduce程序的译、打包和运行过程	程序展示	2
	单元 18：综合实训 5：MapReduce 初级编程实践		能够熟练掌握 MapReduce初级编程实践	掌握综合实训 5：MapReduce 初级编程实践；	程序展示	2
数据仓库Hive的安装和使用	单元 19：Hive 的安装和基本操作	能完成 Hive的安装和编程	能够安装Hive并进行基本操作	掌握Hive的安装和基本操作；	程序展示	2
	单元 20：		能够应用Hive进行编程	掌握 Hive 应用实例；	程序	2

	Hive 应用实例		实践		展示	
Spark的安装和基础编程	单元 21：Spark 的安装和配置	能完成 Spark 的安装和编程	能够安装和配置Spark	掌握Spark的安装和配置； 掌握Spark Shell编程；	程序展示	2
	单元 22：编写 Spark 独立程序		能够使用Shell命令进行Spark编程	掌握 Spark 独立程序的编写；	程序展示	2
典型的可视化工具的使用方法	单元 23：Easel.ly 信息图和 D3 可视化库	能使用典型可视化工具进行数据分析	能够使用Easel.ly进行信息图的制作； 能够使用D3进行可视化操作；	掌握Easel.ly信息图制作； 掌握D3可视化库的使用；	程序展示	2
	单元 24：可视化工具 Tableau 的使用		能使用Tableau工具进行可视化分析	掌握可视化工具Tableau的使用；	程序展示	2
	单元 25：实验“魔镜”和 Echarts 制作图表		能使用“魔镜”和 Echarts制作图表	学会使用“魔镜”制作图表； 学会使用 Echarts 制作图表；	程序展示	2
数据采集工具的安装和使用	单元 26：Flume 工具和 Kafka 工具	能使用数据采集工具	能使用Flume和Kafka工具进行数据采集	掌握Flume工具的使用； 掌握Kafka工具的使用；	程序展示	2
	单元 27：Sqoop 工具和编程实践		能使用Sqoop工具进行数据采集	掌握Sqoop工具的使用； 能完成编程实例；	程序展示	2
大数据课程综合实训案例	单元 28：需求分析和环境搭建	能独立完成一个大数 据实践任务	能够进行综合实训案例的需求分析和环境搭建；	掌握需求分析的方法	程序展示	2
	单元 29：上传本地数据集		能够完成本地数据集上传到Hive；	掌握Hive的使用方法	程序展示	2
	单元 30：Hive 数据分析		能够完成Hive数据分析；	掌握Hive数据分析方法	程序展示	2
	单元 31：数据互导		能够进行 Hive、MySQL、HBase数据互导；	掌握 3 种数据库的互导编程	程序展示	2
	单元 32：数据可视化分析		能够利用R进行数据可视化分析；	掌握 R 数据可视化分析方法	程序展示	2

(4) 进度设计

周次	教学内容	课时分配				教学活动安排	作业安排
		讲授	实验	实训	实习		
1	单元 1: 大数据技术概述	1		1		课堂讲授及操练	适情况安排
1	单元 2: Linux 系统安装	1		1		课堂讲授及操练	适情况安排
1	单元 3: Linux 系统 Shell 命令	1		1		课堂讲授及操练	适情况安排
2	单元 4: 综合实训 1: 熟悉常用 Linux 操作	1		1		课堂讲授及操练	适情况安排
2	单元 5: Hadoop 安装	1		1		课堂讲授及操练	适情况安排
2	单元 6: 伪分布模式下 Hadoop 的运行	1		1		课堂讲授及操练	适情况安排
3	单元 7: HDFS 操作常用 Shell 命令	1		1		课堂讲授及操练	适情况安排
3	单元 8: HDFS 编程实践	1		1		课堂讲授及操练	适情况安排
3	单元 9: 综合实训 2: 熟悉常用 HDFS 操作	1		1		课堂讲授及操练	适情况安排
4	单元 10: HBase 的安装和配置	1		1		课堂讲授及操练	适情况安排
4	单元 11: HBase 编程实践	1		1		课堂讲授及操练	适情况安排
4	单元 12: 综合实训 3: 熟悉常用 HBase 操作	1		1		课堂讲授及操练	适情况安排
5	单元 13: Redis 安装和使用	1		1		课堂讲授及操练	适情况安排
5	单元 14: MongoDB 安装和使用	1		1		课堂讲授及操练	适情况安排
5	单元 15: 综合实训 4: NoSQL 和关系数据库的操作比较	1		1		课堂讲授及操练	适情况安排
6	单元 16: MapReduce 编程 (1)	1		1		课堂讲授及操练	适情况安排
6	单元 17: MapReduce 编程 (2)	1		1		课堂讲授及操练	适情况安排
6	单元 18: 综合实训 5: MapReduce 初级编程实践	1		1		课堂讲授及操练	适情况安排
7	单元 19: Hive 的安装和基本操作	1		1		课堂讲授及操练	适情况安排
7	单元 20: Hive 应用实例	1		1		课堂讲授及操练	适情况安排
7	单元 21: Spark 的安装和配	1		1		课堂讲授及操练	适情况安排

	置					
8	单元 22: 编写 Spark 独立程序	1		1		课堂讲授及操练 适情况安排
8	单元 23: Easel.ly 信息图和 D3 可视化库	1		1		课堂讲授及操练 适情况安排
8	单元 24: 可视化工具 Tableau 的使用	1		1		课堂讲授及操练 适情况安排
9	单元 25: 实验“魔镜”和 Echarts 制作图表	1		1		课堂讲授及操练 适情况安排
9	单元 26: Flume 工具和 Kafka 工具	1		1		课堂讲授及操练 适情况安排
9	单元 27: Sqoop 工具和编程实践	1		1		课堂讲授及操练 适情况安排
10	单元 28: 需求分析和环境搭建	1		1		课堂讲授及操练 适情况安排
10	单元 29: 上传本地数据集	1		1		课堂讲授及操练 适情况安排
10	单元 30: Hive 数据分析	1		1		课堂讲授及操练 适情况安排
11	单元 31: 数据互导	1		1		课堂讲授及操练 适情况安排
11	单元 32: 数据可视化分析	1		1		课堂讲授及操练 适情况安排
合计		32		32		

四、考核方案设计

本课程考核采用形成性考核方式，总评成绩由形成性考核的各项成绩组成，全面考核学生的动手能力、基础理论和平时学习状况。分值比例如下：

综合素质评价（平时表现+通用能力考核） 20%

过程考核（课堂实践考核+课后实践考核+项目开发实践） 50%

终结考核(期终考核) 30%

（一）综合素质评价标准

考核内容	评价标准	考核分值
出勤情况	迟到、早退 1 次扣 0.5 分，旷课 1 次扣 2 分，扣完 5 分为止，旷课 3 次及以上不得参加终结性考核。	5%
学习态度	不认真听课、不按时完成课堂或课外实践 1 次扣 0.5 分，扣完 5 分为止。	5%
学习表现	主动提出问题、回答问题 1 次加 0.5 分；课堂或课外实践完成质量高（正确率 80%以上）1 次加 0.5 分，加满 5 分为止。	5%
通用能力	提出问题、回答问题时语言表达准确、逻辑性强 1 次加 0.5 分，加	5%

满 5 分为止。

(二) 过程考核标准

考核内容	评价标准	考核分值
课堂实践	课堂实践完成质量高（正确率 80%以上）1 次加 0.5 分，加满 15 分为止。	15%
课外实践	课外实践完成质量高（正确率 80%以上）1 次加 0.5 分，加满 5 分为止。	5%
项目开发 实践	需求分析 0 至 3 分、整体设计 0 至 10 分、界面设计 0 至 2 分、程序编码 0 至 10 分、程序测试 0 至 5 分。	30%

(三) 终结考核标准

采用笔试闭卷方式，题型包括：选择题、知识填空题、程序填空题、程序设计题。

分值比例：选择题占 15%，知识填空题占 15%，程序填空题占 30%，程序设计题占 40%。

五、参考资料

1. 林子雨. 大数据技术原理与应用——概念、存储、处理、分析与应用（第2版）. 人民邮电出版社. 2017年.
2. 陆嘉恒. Hadoop实战. 机械工业出版社. 2011年.
3. 曾大聃, 周傲英(译). Hadoop权威指南中文版. 清华大学出版社. 2010年.
4. 迪米达克 (Nick Dimiduk), 卡拉纳 (Amandeep Khurana), 谢磊. HBase实战中文版. 人民邮电出版社; 第1版 (2013年9月1日)