



湖南石油化工职业技术学院  
Hunan Petrochemical Vocational Technology College

# 课程标准

课程名称： Hadoop 大数据

课程代码： 31091020

适用专业： 大数据技术与应用

制订时间： 2020 年 2 月

湖南石油化工职业技术学院

## 目 录

1 课程概述.....	1
1.1 课程的性质.....	1
1.2 课程定位.....	1
1.3 课程设计思路.....	1
2. 课程基本目标.....	2
2.1 素质目标.....	2
2.2 知识目标.....	2
2.3 技能目标.....	2
2.4 职业资格证书融通要求.....	2
3、课程教学内容及学时安排.....	3
3.1 课程主要内容说明.....	3
3.2 课程组织安排说明.....	3
3.3 课程教学内容及要求.....	3
4 教学实施建议.....	8
4.1 教学组织建议.....	8
4.2 教学评价建议.....	8
4.3 教材选用.....	9
4.4 课程主讲教师和教学团队要求说明.....	9
4.5 课程思政要求.....	10
4.6 课程教学环境和条件要求.....	10
4.7 教学资源开发与利用.....	11
4.8 其它.....	11

# 《Hadoop 大数据》课程标准

课程名称：Hadoop 大数据

课程代码：31091020

总学时数：72 学时（理论课学时数：36； 实践课学时数：36）

适用专业：大数据技术与应用

## 1 课程概述

### 1.1 课程的性质

本课程是为大数据技术类相关专业学生开设的课程。随着时代的发展，大数据已经成为一个耳熟能详的词汇。与此同时，针对大数据处理的新技术也在不断的开发和运用中，逐渐成为数据处理挖掘行业广泛使用的主流技术之一。Hadoop 作为处理大数据的分布式存储和计算框架，得到了国内外大小型企业广泛的应用。Hadoop 是一个可以搭建在廉价服务器上的分布式集群系统架构，它具有可用性高、容错性高和可扩展性高等优点。由于它提供了一个开放式的平台，用户可以在完全不了解底层实现细节的情形下，开发适合自身应用的分布式程序。经过十多年的发展，目前 Hadoop 已经成长为一个全栈式的大数据技术生态圈，并在事实上成为应用最广泛最具有代表性的大数据技术。因此，学习 Hadoop 技术是从从事大数据行业工作所必不可少的一步。

### 1.2 课程定位

通过本课程的学习，使学生学会搭建 Hadoop 完全分布式集群，掌握 HDFS 的原理和基础操作，掌握 MapReduce 原理架构、MapReduce 程序的编写。为将来从事大数据挖掘研究工作以及后续课程的学习奠定基础。

### 1.3 课程设计思路

课程设计围绕大数据分析的基础和方法，考虑到实际企业生产环境中的应用场景，课程中采用实际案例，从案例出发、启动、实施、验证各环节剖析讲解，与真实环境同步，本课程通过大量的案例，引导学生思考问题、分析问题，学生在理解知识点的过程中遇到困难时，可以继续往后学习。可以通过对各学习情景中学习目标、主要内容、授课方式、师生要求等各项内容的描述，来规范课程所

要求的内容。真正做到在项目实施过程中各个环节的考察和评价，来完成对课程的评鉴与考核。

## 2. 课程基本目标

### 2.1 素质目标

- (1) 培养学生数据分析意识
- (2) 培养学生谦虚、好学的品质
- (3) 培养学生良好的自我表现、与人沟通的能力
- (4) 培养学生的团队精神和服务意识
- (5) 培养学生的自主学习能力

### 2.2 知识目标

- (1) 理解大数据概念及应用场景
- (2) 掌握开发语言如：Python
- (3) Hadoop 运行环境搭建
- (4) 掌握 Hadoop 三种模式
- (5) 搭建本地模式
- (6) 搭建伪分布式模式
- (7) 搭建完全分布式模式
- (8) 掌握大数据相关技术 HDFS、MapReduce
- (9) 掌握 Hadoop 集群及数据批处理
- (10) 掌握 Hive 数据库操作
- (11) 了解数据可视化应用

### 2.3 技能目标

- (1) 全局的眼光理解大数据，理解大数据的应用，并有总体的概念。
- (2) 理解企业中实际生产环境的大数据搭建，收集，分析，应用的过程。
- (3) 具备在实际应用场景中，能对数据处理流程，数据挖掘进行个性化处理。

### 2.4 职业资格证书融通要求

### 3、课程教学内容及学时安排

#### 3.1 课程主要内容说明

本课程重点是学习 Hadoop 核心技术，Hadoop 的核心是 HDFS 和 MapReduce，而两者只是理论基础，不是具体可使用的高级应用，Hadoop 生态有很多经典子项目，比如 HBase、Hive 等，这些都是基于 HDFS 和 MapReduce 发展出来的。本课程共 72 学时，其中 Hadoop 部分占 24 学时，HDFS 部分占 24 学时、MapReduce 占 20 学时和复习考试占 4 学时。

#### 3.2 课程组织安排说明

本课程主要使用集“教、学、做”于一体，采用案例演示法、项目教学法等教学方法，在电脑上理论结合实际，采用理实一体化教学模式完成课程组织和教学。

#### 3.3 课程教学内容及要求

序号	教学单元(或者模块)	素质内容及要求	知识内容及要求	技能内容及要求	参考学时
1	Hadoop 在云计算技术的作用和地位	培养学生谦虚、好学的品质 培养学生的自主学习的能力	<ul style="list-style-type: none"> <li>◆ 传统大规模系统存在的问题</li> <li>◆ Hadoop 概述</li> <li>◆ Hadoop 分布式文件系统</li> <li>◆ MapReduce 工作原理</li> <li>◆ Hadoop 集群剖析</li> <li>◆ Hadoop 生态系统对一种新的解决方案的需求</li> <li>◆ Hadoop 的行业应用案例分析</li> <li>◆ Hadoop 在云计算和大数据的位置和关系</li> </ul>	理解 Hadoop 分布式文件系统 理解 MapReduce 工作原理 理解 Hadoop 集群	6
2	Hadoop 生态系统介绍和演示	培养学生谦虚、好学的品质 培养学生的团队精神和服务意识 培养学生的自主学习的能力	<ul style="list-style-type: none"> <li>◆ Hadoop HDFS 和 MapReduce</li> <li>◆ Hadoop 数据库之 HBase</li> <li>◆ Hadoop 数据仓库之 Hive</li> <li>◆ Hadoop 数据处理脚本 Pig</li> <li>◆ Hadoop 数据接口 Sqoop 和 Flume, Scribe DataX</li> <li>◆ Hadoop 工作流引擎 Oozie</li> </ul>	理解 Hadoop HDFS 和 MapReduce 模式 理解 Hadoop 数据库之 HBase 理解 Hadoop 数据仓库之 Hive 理解 Hadoop 数据处理脚本 Pig	4

				理解 Hadoop 数据接口 Sqoop 和 Flume, Scribe DataX	
3	Hadoop 组件	培养学生数据分析意识 培养学生谦虚、好学的品质 培养学生良好的自我表现、与人沟通的能力 培养学生的团队精神和服务意识 培养学生的自主学习能力	<ul style="list-style-type: none"> <li>◆ Hadoop HDFS 基本结构</li> <li>◆ Hadoop HDFS 副本存放策略</li> <li>◆ Hadoop NameNode 详解</li> <li>◆ HadoopSecondaryNameNode 详解</li> <li>◆ Hadoop DataNode 详解</li> <li>◆ Hadoop JobTracker 详解</li> <li>◆ Hadoop TaskTracker 详解</li> </ul>	了解 Hadoop HDFS 基本结构、副本存放策略 了解 Hadoop NameNode、HadoopSecondaryNameNode、DataNode、JobTracker、TaskTracker	4
4	Hadoop 安装和部署	培养学生良好的自我表现、与人沟通的能力 培养学生的团队精神和服务意识 培养学生的自主学习能力	<ul style="list-style-type: none"> <li>◆ Hadoop 系统模块组件概述</li> <li>◆ Hadoop 试验集群的部署结构</li> <li>◆ Hadoop 安装依赖关系</li> <li>◆ Hadoop 生产环境的部署结构</li> <li>◆ Hadoop 集群部署</li> <li>◆ Hadoop 高可用配置方法</li> <li>◆ Hadoop 集群简单测试方法</li> <li>◆ Hadoop 集群异常 Debug 方法</li> </ul>	具备 Hadoop 的安装和部署以及集群简单测试方法的能力	6
5	Hadoop 集群规划	培养学生数据分析意识 培养学生谦虚、好学的品质 培养学生的自主学习能力	<ul style="list-style-type: none"> <li>◆ Hadoop 集群内存要求</li> <li>◆ Hadoop 集群磁盘分区</li> <li>◆ 集群和网络拓扑要求</li> <li>◆ 集群软件的端口配置</li> </ul>	理解 Hadoop 集群内存要求 理解 Hadoop 集群磁盘分区 理解集群和网络拓扑要求 理解集群软件的端口配置	4
6	MapReduce	培养学生数据分析	◆ Hadoop MapReduce 算法的	具备 Hadoop	4

	算法原理	析意识 培养学生谦虚、好学的品质 培养学生良好的自我表现、与人沟通的能力 培养学生的团队精神和服务意识 培养学生的自主学习能力	原理和优化思想 ◆ 灵活运用 MapReduce 实现算法	MapReduce 算法的实现和优化的能力	
7	编写 MapReduce 高级程序	培养学生数据分析意识 培养学生谦虚、好学的品质 培养学生良好的自我表现、与人沟通的能力 培养学生的团队精神和服务意识 培养学生的自主学习能力	<ul style="list-style-type: none"> <li>◆ 使用 Hadoop MapReduce Streaming 编程</li> <li>◆ MapReduce 流程</li> <li>◆ 剖析一个 MapReduce 程序</li> <li>◆ 基本 MapReduce API 概念</li> <li>◆ 驱动代码 Mapper、Reducer</li> <li>◆ Hadoop 流</li> <li>◆ API 使用 Eclipse 进行快速开发</li> <li>◆ 新 MapReduce API</li> <li>◆ MapReduce 的优化</li> <li>◆ MapReduce 的任务调度</li> <li>◆ MapReduce 编程实战</li> <li>◆ 如何利用其他 Hadoop 相关技术，包括 Apache Hive, Apache Pig, Sqoop 和 Oozie 等</li> <li>◆ 满足解决实际数据分析问题的高级 Hadoop API</li> </ul>	具备学习 Hadoop API 的能力 具备 MapReduce 算法的实现和优化的能力 具备 MapReduce 的任务调度的能力	8
8	集成 Hadoop 到现有 workflow 及 Hadoop API 深入探讨	培养学生数据分析意识 培养学生良好的自我表现、与人沟通的能力	<ul style="list-style-type: none"> <li>◆ 存储系统</li> <li>◆ 利用 Sqoop 从关系型数据库中导入数据到 Hadoop</li> <li>◆ 利用 Flume 导入实时数据到 Hadoop</li> <li>◆ ToolRunner 介绍、使用 MRUnit 进行测试</li> <li>◆ 使用 Configure 和 Close 方</li> </ul>	理解 Hadoop 存储系统原理 具备使用 Sqoop 导入数据的能力 具备使用 Flume 导入数据的能力	6

		培养学生的团队精神和服务意识 培养学生的自主学习能力	法来进行 Map/Reduce 设置和关闭		
9	使用 Hive 和 Pig 开发及技巧	培养学生数据分析意识 培养学生谦虚、好学的品质 培养学生的团队精神和服务意识 培养学生的自主学习能力	<ul style="list-style-type: none"> <li>◆ Hive 和 Pig 基础</li> <li>◆ Hive 的作用和原理说明</li> <li>◆ Hadoop 仓库和传统数据仓库的协作关系</li> <li>◆ Hadoop/Hive 仓库数据数据流</li> <li>◆ Hive 部署和安装</li> <li>◆ Hive Cli 的基本用法</li> <li>◆ HQL 基本语法</li> <li>◆ 运用 Pig 过滤用户数据</li> </ul>	理解 Hive 和 Pig 基础、作用和原理 理解 Hadoop 仓库和传统数据仓库的协作关系 具备 Hive 部署和安装的能力 具备 Hive 简单知识的基本使用的能力	4
10	Hbase 安装和使用	培养学生谦虚、好学的品质 培养学生的自主学习能力	<ul style="list-style-type: none"> <li>◆ Hbase 安装部署</li> <li>◆ Hbase 原理和结构</li> <li>◆ Hbase 运维和管理</li> </ul>	具备 HBase 安装部署的能力 理解 HBase 原理和运维管理	4
11	Hadoop2.0 集群探索	培养学生数据分析意识 培养学生谦虚、好学的品质 培养学生良好的自我表现、与人沟通的能力 培养学生的团队精神和服务意识 培养学生的自主学习能力	<ul style="list-style-type: none"> <li>◆ Hadoop2.0 HDFS 原理</li> <li>◆ Hadoop2.0 Yarn 原理</li> <li>◆ Hadoop2.0 生态系统</li> </ul>	理解 Hadoop2.0 HDFS 原理 理解 Hadoop2.0 Yarn 原理	4
12	Hadoop 企业级别案例解	培养学生数据分析意识	<ul style="list-style-type: none"> <li>◆ Hadoop 结构化数据案例</li> <li>◆ Hadoop 非结构化案例</li> <li>◆ Hbase 数据库案例</li> </ul>	具备 Hadoop 结构化和非结构化开发的能力	4



	析	<p>培养学生谦虚、好学的品质</p> <p>培养学生良好的自我表现、与人沟通的能力</p> <p>培养学生的团队精神和服 务意识</p> <p>培养学生的自主学习能力</p>	◆ Hadoop 视频分析案例	具备 Hadoop 食品分析的能力	
13	RedHadoop 企业版本	<p>培养学生数据分析意识</p> <p>培养学生谦虚、好学的品质</p> <p>培养学生良好的自我表现、与人沟通的能力</p> <p>培养学生的团队精神和服 务意识</p> <p>培养学生的自主学习能力</p>	<p>◆ 运用 RedHadoop 快速构建服务集群</p> <p>◆ 运用 RedHadoop DW 构建数据仓库</p>	<p>具备运用 RedHadoop 快速构建服务集群的能力</p> <p>具备运用 RedHadoop DW 构建数据仓库的能力</p>	4
14	Spark 原理和入门	<p>培养学生数据分析意识</p> <p>培养学生谦虚、好学的品质</p> <p>培养学生良好的自我表现、与人沟通的能力</p> <p>培养学生的团队精神和服 务意识</p>	<p>◆ Spark 原理；Spark 的架构图；Spark 运行模式介绍</p> <p>◆ <code>--local</code>；<code>--standalone</code>；<code>--mesos</code>；<code>--yarn</code>；Spark 的 RDD</p> <p>◆ 什么是 RDD；RDD 的种类；<code>--Transformation</code>；<code>--Action</code></p> <p>◆ Spark 的存储级别；Cache 介绍；Spark 的容错原理</p> <p>◆ Lineage 容错；Checkpoint 容错；RDD 的创建</p> <p>◆ 案例—统计单词的个数</p>	具备 Spark 相关开发能力	6

	培养学生的自主 学习能力		
复习、考试			4
合计学时			72

## 4 教学实施建议

### 4.1 教学组织建议

- 建议选择 **Hadoop2.0** 版本教学；
- 教师要在开展教学前应有一个已准备好的项目为教学依据，围绕着开展教学；
- 项目教学要尽量采用分组教学法；
- 要注重学生自主编程的培养，避免养成抄写代码的习惯；
- 适当组织些活动，化解抽象枯燥的教学；
- 使用通俗易懂的例证，方便学生简单直观理解。

### 4.2 教学评价建议

#### 1、课程内容评价要点

序号	单元（模块）	考核标准	权重比例%
1	课堂学习	包括出勤、课堂表现及课堂积极回答问题等	15
2	课后作业	是否按时、按质、按量完成教师布置的课后练习	15
3	课堂实训	能否实操出课堂练习	20
4	期末考试	由教师评定的笔试成绩	50

#### 2、课程评价方法和内容

评价类型	评价方法	评价内容
职业素养	过程性评价	到课考勤，学习及工作态度、安全意识、质量观念、

(10%)	(10%)	合作精神、敬业精神等纳入职业素养考核，在具体考核指标中体现。
理论知识 (50%)	过程性评价 (20%)	主要是课堂提问、平时作业、单元测验、期中测验等。
	终结性评价 (30%)	主要是期末考试，评价综合专业理论知识掌握和运用能力，由计算机随机命题或人工命题组成标准试卷，尽量与国家临床医学检验技士职称资格考试接轨。
职业技能 (40%)	过程性评价 (20%)	实训报告、实际操作过程评价。
	终结性评价 (20%)	建议考核核心技能项目 参照技能考核标准与要求，编制核心技能项目的评分标准，评分标准应涵盖操作规范性、结果准确性、人文关怀、沟通交流、操作安全等。

### 4.3 教材选用

#### 1、教材选用建议

为了让学生掌握职业岗位工作所需的技术知识，顺利实施职业技能训练，授课承担部门应选用近几年出版的全国优秀的高职规划教材，并且采用项目驱动式的编写思路为宜。

#### 2、教材编写建议

为了使教材适合高职教育以及现代技术发展快、创新多的特点，突出强调理论教学与实践操作紧密结合的一体化教学模式，自编教材应以“项目导向，任务驱动”为主线。

### 4.4 课程主讲教师和教学团队要求说明

本课程要求任课教师首先牢固树立中国特色社会主义理想信念，践行社会主义核心价值观，自觉增强立德树人、教书育人的荣誉感和责任感，学为人师，行为世范。最好由具有双师型素质的高学历的教程承担。要求教师具有扎实的专业知识和丰富的相关行业实际工作经验，具有一定职业教学教学能力，能够开展课程教学改革和科学研究。

#### 4.5 课程思政要求

全面推进课程思政建设，发挥好专业课程的育人作用。专业课程教学过程以专业知识和技能为载体，加强思想政治教育，充分发挥课堂主渠道功能，努力发掘课程中立德树人的要素，与思想政治理论课同向同行，形成协同效应。本专业课程思政具体要求如下。

##### 1、课程教学与爱国主义教育相结合

通过选择优秀典型的行业企业案例、视频题材等重要思政教育内容，激发爱国热情，培养家国情怀。在专业教师引导之下，通过我国 IT 行业 and 大数据技术应用发展成就和实力的展示，开展爱国主义教育、中国梦教育，增强学生的国家认同感与民族自豪感。

##### 2、课程教学与团队合作精神相结合

专业核心课程实训教学过程中，以实训任务为载体，以工作小组为单元，引导学生将企业本职工作经历融入学习过程，调动学习积极性，重点强调项目成员团队合作的原动力和凝聚力，树立了正确的集体观，培养团队合作精神。

##### 3、课程教学与职业素养培养相结合

通过实践教学环节和企业经历，结合企业生产实际和行业人才素养需求，引入企业对优秀员工必备素质和基本规范的要求，引导学生自觉实践相关行业的职业精神和职业规范，增强职业责任感，培养学生良好的职业品德、职业纪律及职业责任心，教育学生爱岗敬业、讲究诚信、精益求精，在潜移默化中提高了学生未来岗位的适应能力。

##### 4、课程教学与高职学生学情相结合

高职院校学生普遍基础薄弱、学习主动性不强，在这样的学情下，课程教学中教师应实时自我反思和自我总结，不断完善教学手段，增强学生的学习兴趣，提升学生的信心，提高学生的专业能力。

##### 5、课程教学与实际项目案例相结合

教学中，引入实际企业或公司案例，通过理论课程教学结合实际项目案例的教学模式，引导学生提升自我意识、养成良好的职业精神和职业规范，在实际项目案例中不断总结自己、提升自我、提升团队作战意识和团队协作能力。

#### 4.6 课程教学环境和条件要求

主要能够满足正常的课程教学、实习实训所需的专业教室、实训室。

#### 1、专业教室基本条件

配备交互智能教育平板、黑（白板）、多媒体计算机、投影设备、音响设备，互联网接入或 WIFI 环境，并具有网络安全防护措施。安装应急照明装置并保持良好状态，符合紧急疏散要求、标志明显、保持逃生通道畅通无阻。

序号	教学场地	设施配置	功能
1	投影室	投影仪、相关软件等	公共课程教学
2	多媒体机房	电脑、投影仪、相关软件等	专业课理实一体化教学

### 4.7 教学资源开发与利用

#### 1、常规教学文件

常规教学文件应包括：授课计划、教案、讲稿、教学课件等资料。

#### 2、教学资源

应建立适合教师教学的《教学案例库》和适合学生自主学习的《导学手册》和《习题集》。

#### 3、网络资源

有条件的情况下，可以建立本课程相关学习网站，放置教师的授课视频和学习手册、在线测试等教学资源，方便学生在线学习、下载资源、学习讨论。

### 4.8 其它