



长沙职业技术学院
CHANGSHA VOCATIONAL & TECHNICAL COLLEGE

《Spark 快速数据运算》课程标准

教 学 单 位 : 长沙职业技术学院

执 笔 人 : 吴细花

教 研 室 主 任 审 核 : 何政山

分 院 (教 学 部) 负 责 人 审 核 : 桑子华

教 务 处 审 核 : 王聪

教 学 副 校 长 审 定 : 罗慧玲

制 订 日 期 : (2019.7.30)

修 订 日 期 :

长沙职业技术学院教务处 编制

二〇一九年〇七月

【课程信息】

表 1 课程信息表

课程名称	Spark 快速数据运算	开课院部	经济贸易与信息技术学院	
课程代码		考核性质		考试 <input checked="" type="checkbox"/> 考查 <input type="checkbox"/>
总学时	72	课程类型	公共基础课	是 <input type="checkbox"/>
			专业基础课	是 <input type="checkbox"/>
			专业核心课	是 <input checked="" type="checkbox"/>
			专业方向课	是 <input type="checkbox"/>
			专业拓展课	是 <input type="checkbox"/>
			其它	
适用专业	大数据技术与应用专业			

表 2 课程标准开发团队名单

序号	姓名	工作单位	职称/职务
1	凌敏	长沙职业技术学院	讲师
2	吴细花	长沙职业技术学院	副教授
3	殷正坤	长沙职业技术学院	副教授
4	何政山	长沙职业技术学院	副教授

目 录

一、课程概述.....	- 1 -
二、课程定位.....	- 1 -
三、教学目标.....	- 1 -
四、课程设计.....	- 2 -
五、 课程内容与要求.....	- 3 -
六、教学建议.....	- 4 -
七、教学评价与考核.....	- 7 -
八、附录.....	错误！未定义书签。

《Spark 快速数据运算》课程标准

一、课程概述

（一）课程性质

《Spark 快速数据运算》是大数据技术与应用专业必修、核心技能课程。

Spark 拥有 Hadoop MapReduce 所具有的优点，能更好地适用于数据挖掘与机器学习中需要迭代的算法，从而大大减少了读写 HDFS 的次数，Spark 在企业中的运用越来越广泛，学习 Spark 分布式计算框架已然是进入大数据行业所必不可少的一步。

（二）与其它课程的联系

本课程为后续课程《大数据应用开发》和《大数据运维技术》及学院的大数据相关课程的学习提供了专业知识和技能基础，也为《顶岗实习》和《毕业设计》等打下良好的基础，更为职业生涯发展奠定基础。

二、课程定位

本课程是大数据技术与应用专业的专业核心课程，主要目标是培养大数据技术与应用专业学生的能用 Spark 进行实时数据分析和能构建不同业务场景的解决方案和决策的职业能力、具有大数据意识的职业素养和在 Spark 框架基础上进行二次开发的创新能力。

通过本课程的学习，使学生具备数据分析和处理的职业能力和职业素养，能胜任大数据分析技术员、大数据运维技术员等岗位工作。本课程以“餐饮平台菜品智能推荐”项目为导向，通过与长沙中电软件园建立密切合作关系，将企业的教育资源融入到教学体系中，确保学生学习到最先进和实用的数据分析、处理、开发的新技术。学完本课程后，学生可以参加大数据分析与应用职业技能等级证书认证考试，为将来走向工作岗位奠定坚实的基础。

三、教学目标

（一）素质目标

1. 具备良好的动手习惯、严谨的行事风格、踏实的工作作风
2. 具备创新、创意的思维和团队协作精神

3. 形成爱岗、敬业、细致、求精”的职业道德

4. 形成高度社会责任感

(二) 知识目标

1. 了解 Spark 平台的特性和优势；

2. 掌握 Spark 基本原理与架构、集群安装配置、Scala 与 Spark 编程、Spark 代表组件的功能；

3. 掌握利用 Spark 进行数据分析的基本方法；

4. 掌握利用 Spark 进行数据挖掘/机器学习的基本方法；

5. 掌握 Spark 分布式计算框架，掌握 Spark Streaming；实时计算框架、Spark GraphX：图计算框架。

6. 掌握常用推荐算法、数据变换方法、异常数据处理方法，能完成完整项目案例。

(三) 能力目标

1. 规范熟练地使用 Scala 语言进行应用和编程。

2. 能规范熟练地使用 Spark SQL、Spark Streaming、Spark GraphX 等使用方法，进行数据的分析、处理、方案决策。

3. 能熟练编写 Spark 内存计算程序；

4. 能利用大数据相关的工具对数据进行清洗、预处理和可视化分析；

5. 具备初步从事 Spark 项目的能力。

四、课程设计

本课程贴合企业真实需求，以“餐饮平台菜品智能推荐”项目为载体，沿着实际项目的开发流程，课程围绕两大工作领域展开教学：第一大领域，Spark 编程，对接大数据挖掘岗位能力要求，能够编写较为复杂的数据提取语句，准确地横跨众多业务线的底层数据表提取数据。第二大领域是 Spark 对数据分析与处理。对接大数据分析能力和数据仓库搭建，能够根据业务需求，进行数据仓库方案设计，选择合适的大数据产品进行数据仓库搭建，使用数据仓库进行数据的分析处理，使用 SQL 进行数据清洗，复杂的数据分析处理，够，使独立使用大数据平台实现数据库与大数据平台之间的数据迁移。在教与练中培养学生的实践能力。基本每个章节都包含了实训与课后习题，通过练习和操作，帮助读者巩固

所学的内容。通过创造了更加符合企业大数据应用真实场景，帮助广大学生提升对大数据专业的理解能力、操作能力和执行能力。

五、 课程内容与要求

表 1 课程内容与进程安排一览表

序号	项目（模块）	任务（单元）	教学内容			目标要求	学时安排	
			知识点	技能点	思想素质		讲授	实践
1	Spark 概述		1. Spark 基本概念 2. Spark 发展和应用 3. Spark 环境配置方法 4. Spark 架构及原理	SPart 环境配 置、架构	爱岗 敬业	1. 了解 Spark 概念 2. 掌握如何搭建 Spark 分布式环境 3. 掌握 Spark RDD 的概念 4. 了解 Spark 的运行流程和原理	2	2
2	Scala 基础		1. Scala 简介 2. Scala 特性 3. Scala 的环境设置及安装方法 4. Scala 类的定义和运行方法 5. Scala 数据类型、函数、表达式的应用 6. Scala if 判断和 for 循环 7. Scala 集合应用	Scala 的环 境设 置及 数据、 函数 的应 用	爱动 手 爱动 脑	1. 掌握 Scala 环境的安装 2. 掌握 Scala 语言的应用和编程	2	4
3	Spark 编程		1. 从内存中已有数据创建 RDD 2. 从外部存储创建 RDD 3. RDD 转化操作和行动操作方法 4. RDD 键值对操作 5. 文件读取与存储	Spark 中数 据 RDD 的 创 建、 存 储、 转 化 和 行 动。	细 致 求 精	1. 掌握 RDD 的转化操作和行动操作方法 2. 掌握键值对 RDD 的操作 3. 掌握文本文件的读取和存储	4	4
4	Spark 编程进阶		1. 下载与安装 IntelliJ IDEA 2. Scala 插件安装与使用 3. 配置 Spark 运行环境的方法 4. 运行 Spark 程序的方法 5. 持久化方法	配 置 Spark 运 行 环 境 及 方 法	细 致 求 精	1. 掌握如何在 IDEA 中配置 Spark 编程环境 2. 掌握 Spark 程序的编写	4	4

			6. 数据分区方法					
5	Spark SQL: 结构化数据文件处理		1. Spark SQL 简介 2. Spark SQL CLI 配置 3. Spark SQL 与 Shell 交互 4. DataFrame 基础操作方法	Spark SQL 数据处理	严谨 踏实	掌握 Spark SQL 使用方法	4	4
6	Spark Streaming: 实时计算框架		1. Spark Streaming 运行原理 2. Spark Streaming 使用方法 3. DStream 编程模型基本方法	Spark Streaming: 实时计算方法	严谨 踏实	掌握 Spark Streaming 使用方法	4	4
7	Spark GraphX: 图计算框架		1. 图的基本概念 2. 图计算的应用 3. GraphX 的基础概念 4. 图的创建与存储 5. 数据查询与数据转换 6. 结构转换与关联聚合	Spark GraphX: 图计算框架	团队协作 创新 创意思维	掌握 Spark GraphX 使用方法	4	6
8	Spark MLlib: 功能强大的算法库		1. 机器学习 2. MLlib 介绍和应用	MLlib 介绍和应用	团队协作 创新 创意思维	掌握 Spark MLlib 使用方法	4	6
9	项目案例: 餐饮平台菜品智能推荐		1. 常用推荐算法 2. 异常数据处理方法 3. 数据变换方法	常用推荐算法	团队协作 创新 创意思维	1. 学习 ALS 算法 2. 掌握 MLlib 算法包的使用	4	6
汇总							32	40

六、教学建议

(一) 授课教师基本要求

本课程的教学团队成员必须是本科以上学历，中级以上职称，有一定网络、数据库、大数据等课程授课经验，另外还应具备如下素质：

1、语言表达能力：流利的普通话，丰富的课堂语言，有感染力，能够用简洁的语言表达准确的信息；

2、课堂调控能力：具有灵活的课堂调控能力，能够机智、妥善地处理课堂教学中的突发事件；

3、熟知本专业人才培养方案，本课程《课程标准》的基本内容和对本专业学生的基本素质的培养要求；

4、团队成员具有较强团队合作意识，能主动承担并解决教育教学中的问题；

5、团队成员应具有较强的科研能力，能够独立承担各类教育、教学课题研究，帮助本专业尽快成长。

（二）实践教学条件

1. 校内实训室

该实训室硬件主要包括计算机服务器、存储设备、网络设备、机架、UPS 电源。该实训室软件主要包括云计算平台、桌面云平台、云计算实训平台、大数据实训平台。满足于大数据技术基础、数据采集与预处理、大数据存储与处理、大数据分析挖掘技术、数据可视化、Spark 快速数据运算、大数据运维技、Python 爬虫等专业核心课程的教学与实训。

表 2 大数据分析与应用实训室

实训室名称	大数据分析与应用实训室	面积要求	100m ²
序号	核心设备	数量要求	备注
1	SHARPXG-MB55XA 投影机	1 台	
2	计算机服务器	1 台	
3	存储设备	1 台	
4	交换机	4 台	
5	教师机电脑	1 台	
6	学生机电脑	56 台	
7	教学管理软件	1 台	
8	大数据实训平台	1 套	

2. 校外实习基地

配备多个校外实习基地，组织学生到实习基地见习或实习，请企业业务骨干承担学生在企业的实践教学工作，帮助学生对企业实际业务流程进行了解或动手。毕业实习以学生在实习单位顶岗操作为主，直接担当所在单位的某一角色进

行业务处理，直接检验其操作技能与动手能力。

表 3 大数据专业校外实习基地

序号	校外实习基地名称	合作企业名称	用途	合作深度要求
1	华为数据挖掘实习基地	华为技术有限公司	认识实习、跟岗实习、顶岗实习	紧密合作
2	中电软件园实习基地	长沙中电软件园	认识实习、跟岗实习、顶岗实习	深度合作
3	自兴人工智能实习基地	自兴人工智能	认识实习、跟岗实习	紧密合作
4	强智科技实习基地	湖南强智科技发展有限公司	认识实习、跟岗实习、顶岗实习	深度合作
5	拓维信息实习基地	长沙拓维信息技术有限公司	认识实习、跟岗实习	深度合作
7	北京易观智库网络科技有限公司长沙研发基地实习基地	北京易观智库网络科技有限公司长沙研发基地	认识实习、跟岗实习、顶岗实习	一般合作
8	湖南艾瓦特网络科技有限公司实习基地	湖南艾瓦特网络科技有限公司	认识实习、跟岗实习、顶岗实习	一般合作

3. 教学方法与策略

(1) 教学方法

项目驱动法：本课程精选 18 个工作项目，并将这些项目分为 8 个类型，形成 8 个教学单元，每个教学单元由 2-3 个教学项目，1 个实践项目组成，其中每个项目由任务清单-必备知识-任务实施-课后拓展几个环节组成，采用“项目导向-任务驱动-精讲多练-理论实践一体化”的教学方法。

混合式教学法：通过超星学习通课前发布任务清单，学生进行自学，课中教师进行导学，课后进行督学，课中，以学生为主体，教师为主导，增加课堂活力，提高学生学习的自觉性和主动性；激发学生的课堂活力和创新思维能力；培养学生团队合作能力、获取知识的能力。

问题探究法：课堂上引导学生思考项目需要分为几部分完成。用到哪些知识点？通过学生讨论，解决在预习中存在的疑惑问题，然后老师讲解知识的重难点和项目的操作步骤，根据讨论的方法，要求学生们协作完成项目的制作。

(2) 教学策略

利用超星网络教学平台，课前发布学习任务清单，让学生了解每次课的学习

任务，通过完成预习检测，疑问讨论，了解学生预习情况，调整教学内容，课中进行以实训操作为主，教师进行答疑、知识串讲和随堂指导，课后提交作业、完成项目拓展、项目实践，课后测试，项目评价等。

4. 参考教材选用

表 4 《Spark 快速数据运算》课程参考教材选用表

序号	教材名称	教材类型	出版社	主编	出版日期
1	Spark 大数据分析 与实战	高职高专	清华大学出版社	黑马程序员	2019-09
2	Spark 大数据分析 与实战	高职高专	清华大学出版社	黑马程序员	2019-02
3	Spark 快速数据 运算与应用	规划教材	人民邮电出版社	肖芳, 张良均	2018-07

5. 课程资源的利用与开发

充分参与调研，确保教学项目来源于实际应用，并围绕项目建设教学大纲、建设、配备与大数据技术与应用专业有关的音视频素材、教学课件、数字化教学案例库、虚拟仿真软件、数字教材等专业教学资源库，种类丰富、形式多样、使用便捷、动态更新、满足教学，现已建有智慧商务专业群资源库，职教云《大数据技术与应用》专业教学资源库，长沙职业技术学院智慧网络教学平台。

七、教学评价与考核

本课程为考试课程，突出学生解决实际问题的能力，加强过程性考核，期末考试采用百分制的闭卷考试模式。学生的考试成绩由平时成绩（60%）和期末考试（40%）组成，其中平时成绩包括课堂表现、作业、项目完成情况、出勤，在线学习情况组成，具体如下表 5 所示：

表 5 课程考核评价表

考核项目	考核点	建议考核方式	权重 (%)
教学单元	知识点	预习检测	5
	技能点	参与主题讨论	5
	思想素质	参与主题讨论	5

	课堂表现	课堂互动	5
	作业	师评	10
	项目完成情况	师评、互评	10
日常行为表现	考勤	线上平台统计	5
	在线学习情况	线上平台统计	15
期末考试		机试	40