

第二批国家级一流本科课程申报书

（社会实践课程）

课程名称：软件开发综合训练

专业类代码：0809

课程负责人：张宏鸣

联系电话：17795696615

申报学校：西北农林科技大学

填表日期：2021 年 5 月 5 日

推荐单位：信息工程学院

中华人民共和国教育部制
二〇二一年四月

填报说明

1.专业类代码指《普通高等学校本科专业目录（2020）》中的专业类代码（四位数字）。

2.以课程团队名义申报的，课程负责人为课程团队牵头人；以个人名义申报的，课程负责人为该课程主讲教师。团队主要成员一般为近5年内讲授该课程教师。

3.申报课程名称、所有团队主要成员须与教务系统中已完成的学期一致，并须截图上传教务系统中课程开设信息。

4.文本中的中外文名词第一次出现时，要写清全称和缩写，再次出现时可以使用缩写。

5.具有防伪标识的申报书及申报材料由推荐单位打印留存备查，国家级评审以网络提交的电子版为准。

6.涉密课程或不能公开个人信息的涉密人员不得参与申报。

一、课程基本信息

课程名称	软件开发综合训练	是否曾被推荐	○是 ●否
课程负责人	张宏鸣		
负责人所在单位	西北农林科技大学 信息工程学院		
课程编码+选课编码 (教务系统中的编码)	3095312		
课程分类	<input type="radio"/> 创新创业类 <input type="radio"/> 思想政治理论课类 <input checked="" type="radio"/> 专业类 <input type="radio"/> 其他(填写)		
课程性质	<input checked="" type="radio"/> 必修 <input type="radio"/> 选修		
开课年级	大学本科三年级		
面向专业	计算机科学与技术、软件工程、电子商务、信息系 统与管理、数据科学与大数据技术		
实践基地	名称及所在地：四川华迪信息技术有限公司、西安 达内优创信息科技有限公司，西北农林科技大学计 算机教学实验中心		
学 时	总学时：6 周（240 学时） 理论课学时：80 学时 实践学时：160 学时		
学 分	6 分		
最近两期开课时间	2020 年 8 月 3 日—2020 年 9 月 4 日（上传教务系统 截图）		
	2019 年 6 月 3 日—2019 年 7 月 12 日（上传教务系 统截图）		
最近两期学生总人数	634		

注：教务系统截图须至少包含课程编码、选课编码、开课时间、授课教师姓名等信息。

二、授课教师（教学团队）

课程团队主要成员（序号 1 为课程负责人，总人数限 5 人之内）								
序号	姓名	出生年月	单位	职务	职称	手机号码	电子邮箱	授课任务
1	张宏鸣	197908	西北农林科技大学信息 工程学院	院长	教授	13891990787	zhm@nwsuaf.edu.cn	课程总负责、总协调
2	聂炎明	197209	西北农林科技大学信息 工程学院	无	副教授	13572917517	yanmingnie@nwsuaf.edu.cn	大数据方向的辅导和答疑

3	张晓婷	197504	西北农林科技大学信息工程学院	无	讲师	13709129203	zxt@nwsuaf.edu.cn	人工智能方向教学指导和答疑
4	李宏利	197403	西北农林科技大学信息工程学院	无	讲师	13572869958	lhl@nwsuaf.edu.cn	Java EE 方向的辅导和答疑
5	霍迎秋	197810	西北农林科技大学信息工程学院	无	高级实验师	18092121869	fallying@nwsuaf.edu.cn	人工智能方向实践教学指导

课程负责人和团队其他主要成员教学情况（500 字以内）

（教学经历：近 5 年来在承担该门课程教学任务、开展教学研究、获得教学奖励方面的情况）

1. 课程负责人情况

课程负责人先后承担“面向对象程序设计”、“面向对象系统分析与设计”、“.Net 框架程序设计”、“软件开发综合训练”等课程的教学工作，年均教学工作量 210 学时。

2012 年开始承担该门课程的教学工作，2016 年成为课程负责人，主要负责该门课程的方案制定完善、企业对接、实践指导、参观学习、考核等工作。项目负责人担任陕西省计算机学会常务理事，陕西省计算机教育学会理事。近五年主持教育部首批新工科研究与实践项目 1 项（2018）、教育部协同育人项目 1 项（2017），参与陕西省教改重大攻关项目 1 项（2019，第 2）。发表相关教改论文 6 篇，其中核心教改 3 篇；主编或参编教材 4 部；获陕西省教学成果一等奖 1 项（2020，第 3）、校级教学成果二等奖（2017，第 1）、“唐立新教学名师奖”（2020）等奖项，先后两次被毕业生评为“我最喜爱的老师”十佳。

2. 团队其他成员情况

团队其他成员主要承担“JAVA EE”、“人工智能”、“大数据”等课程，发表教改论文 8 篇，其中核心 7 篇；主持或参与教学改革项目 14 项，其中省、部级以上教学改革项目 6 项；省级精品资源课程建设项目 1 项，校级优质资源建设课程 3 项，获校级教学成果一等奖、二等奖各 1 项，获校级先进个人 4 人次。

三、课程目标（300 字以内）

（结合本校办学定位、学生情况、专业人才培养要求，具体描述学生学习本课程后能够获取的知识、培养的能力和塑造的价值）

根据我校“厚基础，强能力，求创新”的人才培养要求，应对人才培养供给侧结构要素全方位融合的需求，对标新工科培养目标，采用校企合作方式，线上/线下结合模式，案例驱动的教学方法，将知识能力和爱岗敬业团队合作的价值观塑造等融为一体，达到以下培养目标：

知识目标：掌握软硬件系统设计与建模、大数据分析、人工智能应用及软件测试的最新方法；熟悉软件开发的全周期流程及具体的业务要求；了解专业新技术、新应用和新发展。

能力目标：具有根据软件项目的复杂应用场景和实施模式，选择最适合的开发方法、设计模式、最优算法设计，以及软硬件功能实现和部署的能力。

素养目标：具备独立思考、团队协作、探索精神、工匠精神以及服务三农的情怀。

四、课程建设及应用情况（2000 字以内）

（本课程的建设发展历程，课程与教学改革要解决的重点问题。课程内容与资源建设及应用情况，理论学习与社会实践内容的相关性，社会实践环节的动手训练内容，具体做法案例。课程学业考评方式，课程评价及改革成效等情况。

1. 课程建设发展历程

探索期：2011 年及以前。积极组织课程调研与方案研讨，探索学生外派和驻校的混合校企实践育人模式，方案由企业制定，课程主要由企业工程师承担，且只针对软件工程专业。在此阶段学院师资薄弱，课程资源匮乏，因此在双师型师资队伍打造、课程资源建设上逐渐加大投入。

建设期：2012 年-2017 年。将企业引入校园，分为 Java EE、C++和嵌入式

三个方向，覆盖软件工程、计算机科学与技术、信息管理与信息系统和电子商务四个专业，为期4周或6周。该阶段，师资得到有效补充，案例资源得到极大丰富，相关平台建设实现突破，教师和企业工程师共同参与课程方案打造、课堂讲授、考核等环节，实践效果大幅加强。获省教学成果二等奖（2013），获批省级人才培养模式创新实验区（2014），获省级精品资源共享课程（2014），出版相关教材（2015，2017）。

提升期：2018年-至今。优化实践方案，进一步对接社会需求，在涵盖原有Java EE、C++方向的基础上，增加大数据和人工智能两个新方向，引入更有实力的实践企业，教师全面参与实践课程，方案库、案例库、授课视频库、实践成果库等建设成效显著。主持获教育部首批新工科项目1项（2018），陕西省重点攻关项目1项（2020），省教学成果一等奖1项（2021）。

2. 课程与教学改革要解决的重点问题

（1）教学内容与行业的业务需求、技术要求和发展的不同步。用新理论、新知识、新技术更新教学内容，强化实践教学，这是本课程需要重点解决的问题之一；

（2）软件工程理论联系实践缺乏的问题。学生很少应用软件工程理论系统全面的进行软件设计，需要在一个真实业务环境中进行系统设计体验，实现所学知识的综合运用；

（3）团队协作精神不足的问题。团队中不同专业背景的学生缺乏有效沟通和协作。

3. 课程内容与资源建设

规划弹性化的实践方向，构建精品化的实践课程体系，针对不同方向定制差异化的课程内容。

（1）对于Java EE、C++以及嵌入式等软件开发类方向，全面实践软件开发过程、沉浸体验软件项目管理、实景面对实际项目需求、深入实践业界主流技术；

（2）对于大数据、人工智能等技术应用方向，聚焦大数据处理以及人工智能应用的全流程，应对真实的需要清洗、转换等预处理的领域数据，实践业界主流的方法库和技术栈。

（3）累计构建实践方案库（20个）、项目库（30个）和优秀成果库（50

个)等,为后续课程的开展提供资源基础,为课程的改革创新提供第一手材料。

(4) 理论授课注重对校内培养环节的拓展和深化,高度聚焦项目实践中关键理论知识的基本需求。实践环节的动手训练内容包括原型构建、技术实践、文档编制以及成果展示等。

4. 课程教学内容及组织实施情况

(1) 基于项目驱动和激励挑战的实践教学模式

将项目模块的设计与实现穿插到课程的所有内容章节中,以点连线、多线成面,使学生能够快速理解并掌握各章节重点知识。对项目逐步进行迭代、精化,最终完成一个项目的完整开发过程。同时运用有压力的实践学习方法,设置挑战内容以充分激发学生的潜能,提升学生的抗压能力。

(2) 采用结队编程、成员优势互补的团队协作开发方式

充分利用团队成员的经济管理/电子商务业务知识背景、软件工程专业技能以及学术研究经历,进行团队式项目开发研讨及实践,实现各专业业务领域与计算机学科的有机结合,增加学生的学习兴趣和参与度,培养沟通交流能力和团队合作精神。

(3) 运用线上线下混合式的高效教学方法

针对软件开发基础理论与核心技术、业务领域模型与规则等问题,采用翻转课堂。

课前:教师布置学习内容和实践任务,学生通过线上学习并带着问题进入线下课堂;

课中:学生通过小组汇报、提问、讨论等环节整理出问题清单,教师针对重点、难点和问题进行讲解和答疑,并引导学生进一步思考;

课后:学生线上测验、互动和同行点评,巩固延伸所学知识和技能。教师线上跟踪、监测和指导,持续性改进教学设计和实施方法。

5. 成绩评定方式

(1) 成绩评价标准如表 1 所示。

表 1 品质描述-标准表

编号	考核项目	品质描述	评估等级	备注
1	工作质量	工作质量严重不达标	不及格<70	
		工作质量存在瑕疵和纰漏	需改进(70-95)	
		能符合工作标准要求	达标(96-105)	
		工作质量超出预定标准	优秀(106-120)	
		工作质量远远超过预定标准,体现品质	卓越(121-150)	
2	工作效率	严重超时,无法按时完成任务	不及格<70	
		未能按时完成,有延迟	需改进(70-95)	
		能在规定的时限内完成任务	达标(96-105)	
		能够提前完成任务	优秀(106-120)	
		完成任务所需的时间远低于规定时间	卓越(121-150)	

(2) 成绩组成如表 2 所示。

表 2 西北农林科技大学信息工程学院-实践绩效考核

姓名				岗位名称				班级			
小组名称				直管领导				填表日期			
序号	考核项目	实训目标	考核标准		指标权重	完成情况说明		结果得分	最终得分		
1	平时成绩	作业完成度	以作业表信息为准		15%				0.0		
		课堂表现	以授课老师打分为准		15%				0.0		
2	过程考核成绩	代码质量	以编码规范检查结果为准		30%				0.0		
3	答辩及代码成绩	答辩表现	以答辩结果为准		20%				0.0		
		项目完成度	以授课老师检查结果为准		20%				0.0		
合计					100%				0.0		
班主任复核：											
					签名：						
讲师复核：											
					签名：						
学生签名：						分管领导签名：					
日期：						日期：					

6. 课程评价及改革成效

(1) 校企联合育人的实践教学新模式,认可度高

累计 10 届学生参与实践教学课程的学习,受益学生达 3 千多人;校内兄弟学院来院交流、取经 20 余次,先后被校内新闻报道 15 余次;得到了广大师生的认可。校企协同育人 在学生就业、深造中已经凸显优势,是学生口中最有价值、最有收获、最值得期盼的教学环节。

(2) 学生实践能力的提高,分析问题、解决问题能力大幅提升

通过理论联系实践,实景案例化教学、加深了学生对软件开发理论知识的理解,提升了实践动手能力;新理论、新知识、具体的市场需求的引入,让学生深

切体会到企业开发流程、模式，为学生毕业工作，继续读研奠定了技术基础。

(3) 培育了一支优秀的实践教学队伍

2011 年以来，先后有 15 位校内指导教师参与课程建设。团队成员主持省级以上教改项目 8 项，出版教材 5 部，获省级以上教学成果奖 2 项，指导科创 32 项；获唐立新教学名师奖、学校首批卓越教师奖、校级“优秀教师”和学校“学生最喜爱的教师”；讲课比赛奖 5 人次，入选课程思政案例 3 项。

五、课程特色与创新（500 字以内）

（概述本课程的特色及教学改革创新点）

课程特色：

(1) 校企协作实践育人新模式

紧跟市场对软件人才的业务、技术和能力的要求，把企业引进校园，把学生带入企业，校企共同完善课程内容、开发课程资源、协同打造实践教学方案。学生多角色、团队分工与合作完成项目全流程开发，充分体验企业软件开发的过程及规范、组织与管理。实景案例培养学生复杂工程开发能力

(2) 软件开发模式驱动的多元化考核

以实际商业项目案例为原型，以团队为基本单位，开展多元化考核。针对不同角色，将终结性考核与形成性评价相结合，基于过程激励机制，引导学生重视学习过程，开发学生潜力及探究能力。

(3) 德育教育与智育教育有机结合

结合软件开发发展历史特点，挖掘吃苦耐劳、艰苦奋斗、工匠精神等思政元素融入课程教学，根植学生“经国本，解民生，尚科学”的高尚情怀。

创新点：

(1) 基于实景案例，实践软件开发全流程，融入软件开发新理念、新技术，真正实现高校的人才培养与职场有效对接。

(2) 基于产教融合校企合作，教师深度参与，工程师与教师高度配合，培养了一批高素质、高水平、双师型教师队伍。

(3) 坚持立德树人，以学为中心，引导学生，互帮互助，掌握团队协作方法，培养团队合作精神，熏陶企业文化，学生得到全面培养。

六、课程建设计划（500 字以内）

（今后五年课程的**持续建设**计划、需要进一步解决的问题，改革方向和改进措施等）

结合我校“经国本、解民生、尚科学”的办学理念，本课程以“学”为中心，聚焦知识、能力、素养三个层面的人才培养，使课程育人目标向更深更广延伸：

1. 持续改进教学内容，赋予课程新内涵

坚持立德树人，以学生发展为中心，开发学生内在潜力和学习动力的初心不变，基于信息技术发展与市场需求的变化，不断重塑课程知识体系、能力体系和评价体系，优化教学模式。

2. 持续改进教学方法，提高学生服务三农的能力

持续加大资源共享平台建设和资源整合力度，并融入先导课程的教学；出版优秀案例教材，推广教学成果，对接乡村振兴、现代农业信息化建设的新需求，丰富教学选题渠道和形式；不断探索学生自主实践过程管理制度和方法，在实践中培养学生团队合作能力、独立解决问题能力、压力承受能力。

3. 创新合作模式、深化合作内容

不断创新合作模式，深化合作内容，深入推进产、学、研、用融合，比如与企业共建研发中心，联合开展科技攻关、项目合作。

七、附件材料清单

1. 课程负责人和团队成员的 10 分钟“说课”视频

〔含课程概述、教学设计思路、教学环境（课堂内外）、教学方法、创新特色、教学效果评价与比较等。技术要求：分辨率 720P 及以上，MP4 格式，图像清晰稳定，声音清楚。视频中标注出镜人姓名、单位，课程负责人出镜时间不得少于 3 分钟。“说课”使用的语言及字幕为国家通用语言及文字。〕

2. 教学设计样例说明

（提供一节代表性课程的完整教学设计和教学实施流程说明，尽可能细致地反映出教师的思考和教学设计，在文档中应提供不少于 5 张教学活动的图片。要求教学设计样例应具有较强的可读性，表述清晰流畅。课程负责人签字。）

3. 最近一学期的教学日历

（申报学校教务处盖章。）

4. 最近一学期的社会实践考核作业及学生代表性的实践报告

（申报学校教务处盖章。）

5. 最近两学期的学生成绩分布统计

（申报学校教务处盖章。）

6. 最近一学期的课程教案

（课程负责人签字。）

7. 最近一学期学生评教结果统计

（申报学校教务处盖章。）

8. 最近一次学校对课堂教学评价

（申报学校教务处盖章。）

9. 教学（课堂或实践）实录视频

提供完整的一节课堂实录视频（标注课程内容、课程对象、上课时间以及上课地点，至少 40 分钟。技术要求：分辨率 720P 及以上，MP4 格式，图像清晰稳定，声音清楚。教师必须出境，视频中需标注教师姓名、单位；要有学生的镜头，并须告知学生可能出现在视频中，此视频会公开。少数民族语言视频须配国家通用语言字幕。）

10. 课程团队成员和课程内容政治审查意见

（申报课程高校党委负责对本校课程团队成员以及申报课程的内容进行政审，出具政审意见并加盖党委印章；团队成员涉及多校时，各校党委分别对本校人员出具意见；非高校成员由其所在单位党组织出具意见。团队成员政审意见内容包括政治表现、是否存在违法违纪记录、师德师风、学术不端、五年内是否出现过重大教学事故等问题；课程内容审查包括价值取向是否正确，对于我国政治制度以及党的理论、路线、方针、政策等理解和表述是否准确无误，对于国家主权、领土表述及标注是否准确，等等。）

11. 课程内容学术性评价意见

〔由学校学术性组织（校教指委或学术委员会等），或相关部门组织的相应学科专业领域专家（不少于 3 名）组成的学术审查小组，经一定程序评价后出具。须由学术性组织盖章或学术审查小组全部专家签字。无统一格式要求。〕

12. 社会实践所在地相关机构的评价

（实践基地相关机构出具并盖章。）

13. 其他材料，不超过 2 份（选择性提供）

以上材料均可能在网上公开，请严格审查，确保不违反有关法律及保密规定。

2.教学设计样例说明



西北农林科技大学
Northwest A&F University

Spark RDD 编程的教学设计

所属学科： 工 学

专 业： 计算机/软件工程/电子商务/信息管理与信息系统

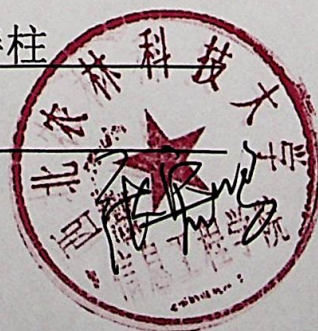
课 程： 软件开发综合训练

适用对象： 计算机、软工、电商、信管专业本科三年级学生

主讲教师： 张宏鸣、聂炎明、张晓婷

李宏利、霍迎秋、夏善柱

教学单位： 信息工程学院



课 程 说 明			
课程	软件开发综合训练	题目	Spark RDD 编程
章次	第五章 Spark	时长	约 100 分钟
适用对象	计算机、软工、电商、信管专业本科三年级学生		
使用教材	《Spark 编程：Python 版》，主编：林子雨等，人民邮电出版社（第 1 版）		
教 学 内 容 分 析			
教学内容	本节课介绍了 Spark RDD，以及基于 Spark RDD 的 API 的编程。		
教学背景	第四章介绍了 Hadoop HDFS 和 Hadoop MapReduce，但由于 Hadoop MapReduce 自身的缺陷而在实际大数据批处理运算中采用 Spark。Spark RDD 作为 Spark 的核心概念,对于理解 Spark 的架构、运行过程以及 Spark 编程，在 Spark 的知识体系中具有极其重要的地位。		
教学目标	1. 理解 Spark RDD 的概念、特性以及在 Spark 执行过程中的角色； 2. 掌握几种常用的 Spark RDD 算子（转换算子和行动算子），熟悉基于 RDD API 的 Spark 编程； 3. 了解包括分区、缓存等的 Spark RDD 的几个高级特性或操作，以及 Spark 广播变量的概念和使用。		
教学重点	1. Spark RDD 的概念； 2. Spark RDD API 编程。		
教学难点	1. Spark RDD 之间的宽依赖和窄依赖； 2. Spark RDD 的惰性计算机制。		
教 学 思 路 和 方 法			
教学思路	提出问题	以基于 Hadoop MapReduce 实现的 WordCount 为例，讲述 MapReduce 用于批处理时的缺陷。特别是针对 KMeans 等迭代式处理的场景下 Hadoop MapReduce 所存在的问题。	
	探索方法	在上述问题的基础上，引导学生进入到 Spark RDD 的基本概念的学习中。引出基于 Spark 的 WordCount 程序的简要的说明。	

	解决问题	首先基于 Hadoop MapReduce 的 WordCount 程序，然后引导学生通过逐步分析处理过程中各个步骤，意识到需要反复存取文件并因此而造成的效率底下的问题，从而解决本节课一开始所提出的核心问题。
	工程应用	通过 Spark 自带的 KMeans 程序，说明基于 Spark RDD 的 KMeans 程序，在需要多轮迭代的程序中是如何避免反复存取文件的，进而极大地提升了程序的执行效率。这种理论与工程实际相结合的方式，能够让学生进一步认识到 Spark 在大数据批处理计算中目前所处的核心地位。
教学方法	<p>本次微课主要采用以问题驱动为主的“<u>启发式</u>”教学法，在介绍新知识点的过程中，基于学生已知的现象和原理，设计适当的问题，从而逐步引导学生积极思考，这种方式有助于学生对新概念——Spark RDD 的理解和掌握。</p> <p>采用 PPT 的教学形式，利用 Spark RDD、Spark 架构以及运行机理等多张图片解释 Spark 以及 Spark RDD 的含义，形象直观便于学生理解。此外，运用<u>边讲解边上机实践</u>的方式丰富教学手段，帮助学生加深对于 Spark RDD 相关理论知识的理解，对于 Spark RDD API 编程获得第一手的直观体验和排错经历等，从而激发了学生对于 Spark 的学习兴趣。</p>	
教 学 过 程 设 计		
<p>本节课的教学内容分为以下 7 个部分：</p> <p>1) ★Spark RDD 原理（15 分钟）；2) Spark 任务阶段（Task Stage）的划分（10 分钟）；3) Spark 运行过程（15 分钟）；4) ★Spark RDD 的创建（10 分钟）；5) ★Spark RDD 算子（30 分钟）；6) Spark 的广播变量（10 分钟）；7) Spark 和 YARN 中各个进程的对比（3 分钟）。【注：预留 7 分钟做本节课的总结】</p> <p>本节课的教学过程分为以下 6 个部分：</p> <p>1. 提出问题</p> <p>以学生熟悉的基于 Hadoop MapReduce 的 WordCount 程序为例，通过该程序在运行的而过程中需要多次存取文件而造成效率低下的问题，引出 Spark 以及 Spark RDD 的概念，同时抛出迭代式批处理问题如聚类等算法如何 MapReduce 化，具体给出 Spark 自带的 KMeans 程序，从而激发学生的学习兴趣。通过定性分析可计算出，在基于 Hadoop MapReduce 的 WordCount 程序中存取文件的次数以及数据量，由此提出</p>		

本节课要解决的两个问题。

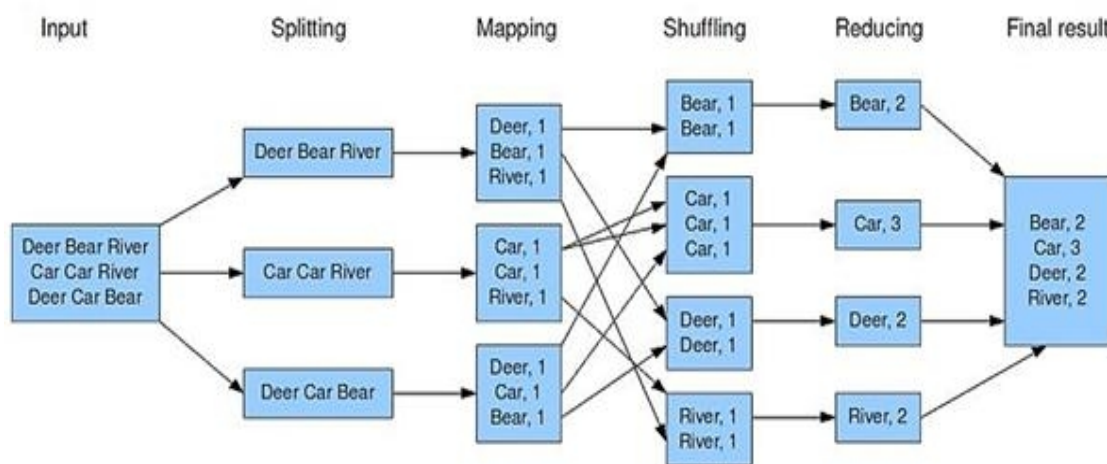
(1) ★如何尽量避免批处理计算中的中间结果频繁存取文件？

(2) Spark 是如何解决上述问题？亦即 Spark RDD 的内在机理。

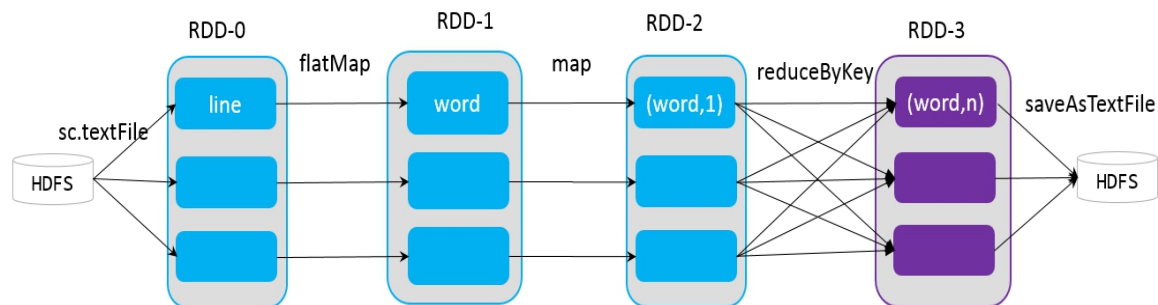
* 问题(1)是研究本节课所学内容要解决的核心问题，通过提问的方式从而让学生明确学习目标，明确 Spark RDD 设计的内在主旨。问题(2)的答案是对问题(1)的有效解答。

2. 探索方法

通过展示基于 Hadoop MapReduce 的 WordCount 程序的处理流程示意图，讨论可能涉及到中间结果存取文件的环节，并分析其对于处理性能的影响。



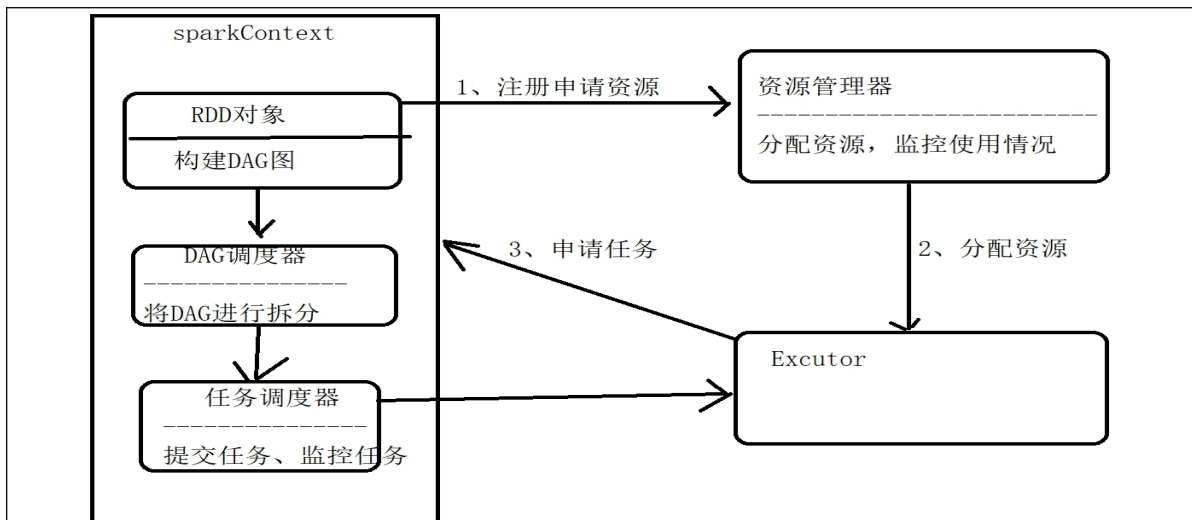
给出基于 Spark RDD 的 WordCount 程序的处理流程示意图，以示对比。



重点讨论是 Spark 在哪些处理环节避免了中间结果的文件存取开销并因而能够大幅度提升处理效率的。通过层层设问以期引发学生积极参与思考，从而达到自然地过渡到对 Spark RDD 的学习。

3. 内容展开

1) Spark RDD 原理 (15 分钟): 由 Hadoop MapReduce 的缺点，引出 Spark RDD。



- 创建 sparkContext 对象;
- 向资源管理器注册;
- 为 Executor' 申请资源;
- 资源管理器分配资源;
- 在 work 中启动 Executor' 进程, Executor' 通过心跳反馈信息;
- sparkContext 构建 DAG 有向无环图, 通过 DAG 调取解析, 把任务集 TaskSet 交给任务调度器 TaskScheduler;
- 任务调度器 TaskScheduler 分发任务给 Executor;
- Executor 执行任务反馈结果信息给任务调度器 TaskScheduler。

4) Spark RDD 创建方式 (10 分钟)。讲解 5 分钟, 上机实践 5 分钟。

- 方式一。使用 parallelize 方法基于非 RDD 数据结构创建 RDD;
- 方式二。读取外部文件: 1) 读取本地文件创建 RDD; 2) 读取 HDFS 文件创建 RDD。

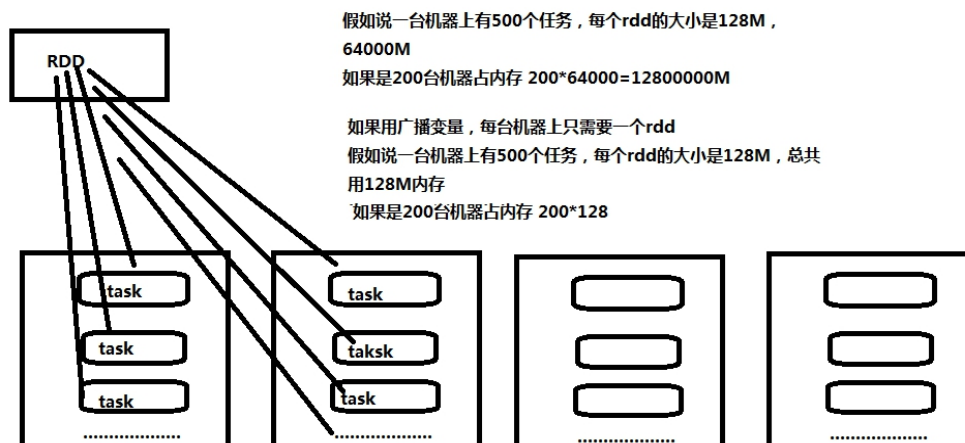
RDD 的分区数设定与重分区。

5) Spark RDD 算子 (30 分钟)。讲解 15 分钟, 上机实践 15 分钟。

- 转换算子:
 - reduceByKey();
 - groupByKey();
 - keys() 和 values() 与 reduceByKey 及 groupByKey 的对比;
 - mapValues 是对 key-value 的中所有 value 进行 map 操作;
 - sortByKey: 按照 key 值进行排序;
 - sortBy();
 - join();
 - filter() 过滤;
 - distinct() 去重。
- 行动算子:
 - collect();
 - take();
 - foreach();
 - reduce()。

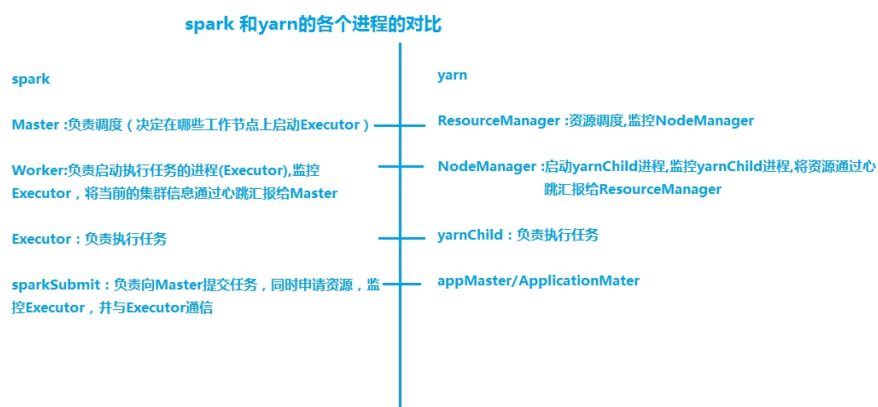
RDD 的缓存等高级特性操作等。

6) 广播变量 (10 分钟)。讲解 5 分钟，上机实践 5 分钟。



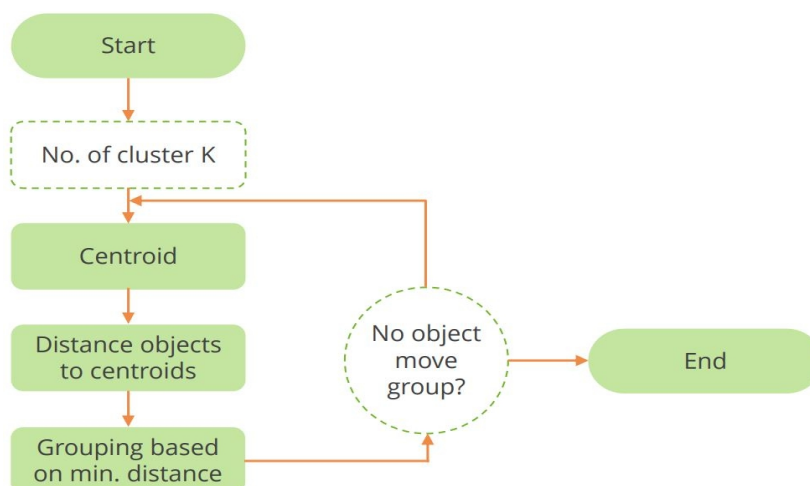
广播变量与累加器及其操作 API。

7) Spark 和 YARN 中各个进程的对比 (3 分钟)。



8) 讨论与答疑 (7 分钟)。

4. 解决问题及举例 (时间可以纳入讨论与答疑环节中)。



以聚类算法 KMeans 为例，如果是采用基于 Hadoop MapReduce 的 KMeans 算法，在一个迭代中以及上下游迭代之间将产生大量的中间结果的文件读写操作，从而会极大地影响计算效率。而采用 Spark RDD 机制，将有效地避免上述的大量文件存取操作，从而也提升了运算效率。同时，还列举大家所熟知的数据关系表的自然连接操作的 MapReduce 化程序，探讨基于 Hadoop MapReduce 与基于 Spark 的区别，从而提高学生的学习兴趣。

并通过学生自身的上机实践，进而可以在一定程度上激发学生对本章节知识的积极性。

5. 内容小结

(1) 重点

- ① Spark RDD 的概念。Spark RDD 是高度受限的、分布式数据集结构；
- ② Spark RDD 算子及其 API 编程。两种算子的区别。Spark RDD API 的语法等。

(2) 难点

- ① Spark RDD 之间的宽依赖和窄依赖。且这是 Spark 将处理任务（Task）划分为执行阶段（Stage）的依据；
- ② Spark RDD 的惰性计算机制。特别是在 Spark RDD 编程中的排错处理中可能由于不理解该惰性机制而造成的困惑。

6. 课后思考

(1) 基于 Spark RDD API 实现诸如关系的自然连接操作等

目的：深入理解程序的（MapReduce 或）Spark 化。

(2) Spark RDD 及其 API 的代码内核分析

目的：深入理解 Spark RDD 的机理。

注：途径可以是分析如关系的自然连接操作等的相应算法，尝试对其进行（MapReduce 或）Spark 化改造。并上机查看 Spark RDD 相关源码。

教 学 总 结

学生在学习 Spark RDD 的初期，不易深刻理解其背后的设计主旨，可能导致将 Spark 仅仅当作是另一种大数据批处理框架。本节课在以问题为驱动的“启发式”教学的基础上，主要从以下两方面改善教学方式，以期获得较好的教学效果：

1. 以 PPT 讲解形式，以学生熟悉的基于 Hadoop MapReduce 的 WordCount 程序导入要 Spark RDD 的基本。
2. 通过相关的视频和图片资料，尽可能地做到理论联系工程实际。

3.最近一学期的教学日历

西北农林科技大学教学日历

20 19 — 20 20 学年第 3 学期

课程名称: 软件开发综合训练-Java EE 方向 授课专业: 计算机、软件、信管、电商 17 级

总学时: 240 本学期学时: 240 讲课学时: 80 实验学时: 152 其它环节: 8 学时

教研室(系)主任签字: 刘钟

教学院长(主任)签字: 李

天	学时	其中			计划教学内容	方式
		讲课	实验	其它环节		
1	8	8			Java、Java EE、数据库、软件工程、软件测试、UML 等训前测评	
2	8	4	4		ORM 原理、常见 ORM 介绍 (Hibernate、MyBatis)、MyBatis 系统对象、MyBatis 作用域和系统对象、MyBatis XML 配置文件、MyBatis 对象工厂	
3	8	4	4		MyBatis SQL 映射语句 (SELECT、INSERT、UPDATE、DELETE)、MyBatis SQL 参数、MyBatis resultMap、MyBatis 动态 SQL	
4	8	4	4		MyBatis JavaAP、Maven 构建原理、SpringMVC 框架简介、SpringMVC 控制器类、SpringMVC 和 web.xml 映射等	
5	8	4	4		SpringMVC 请求数据的绑定、SpringMVC 数据转换、格式化、校验、SpringMVC 数据模型控制、SpringMVC 视图及解析器	
6	8	4	4		IOC 与 DI 的基本概念、Spring 整体框架结构、Spring IoC 容器、Spring AOP、Spring 事务 Spring Bean 及工厂、Spring 与 MyBatis 集成、SSM 框架集成	
7	8	4	4		IntelliJ IDEA 主要功能, 工具安装, 安装目录与配、编译, Project 和 Module、项目创建和配置、版本控制配置与使用、实时代码模板等	
8	8	4	4		SpringBoot 简介、开发环境搭建, maven 工程创建、测试类编写与测试、SpringMVC 框架的实现、整合 Mybatis, 整合 Redis, 整合 Junit 等	
9	8	4	4		jQuery 语法、选择器、对象, 使用 jQuery 进行 DOM 操作、jQuery 中的事件处、Ajax 工作原理、XMLHttpRequest 对象、Ajax 发送请求和处理响应	
10	8	4	4		Ajax 与 DOM 编程、使用 Ajax 技术实现表单验证、使用 Ajax 技术实现动态加载列表框	
11	8	4	4		RUP 与 CMMI3、软件项目计划	
12	8	4	4		配置管理、软件需求开发 (UML 与用例模型)	
13	8	4	4		软件分析设计 (分析、设计模型)、软件分析设计 (数据建模)	
14	8	4	4		软件测试技术、先启阶段 (Inception)、精化阶段 (Elaboration)	
15	8	4	4		构建阶段 (Construction)、产品化阶段 (Transition)、过程管理	
16	8	4	4		最终确认分组选题情况, 老师介绍行业开发规范, 学员完成项目需求分析	
17	8	4	4		根据需求分析完成概要设计、详细设计、画原型	
18	8	4	4		完成系统技术架构设计	
19-28	80		80		各小组完成系统开发、测试、优化、撰写功能测试报告和性能测试报告	
29	8		4	4	PPT 项目讲述/项目演示/项目技术答疑/项目总结/答辩材料收集	
30	8	4		4	完善项目内容, 材料归档, 对实训过程进行量化总结	

西北农林科技大学教学日历

20 19 — 20 20 学年第 3 学期

课程名称: 软件开发综合训练-大数据方向 授课专业: 计算机、软件、信管、电商 17 级

总学时: 240 本学期学时: 240 讲课学时: 80 实验学时: 149 其它环节: 11 学时

教研室(系)主任签字: 刘钟

教学院长(主任)签字: 李

教务处

信息工程学院

天	学时	其中			计划 教 学 内 容	方式
		讲课	实验	其它环节		
1	8	5	3		实训课程安排、考评要求、开发环境检查; CMMI3 规范介绍; Hadoop 与 Hbase、Spark、Hive 基础与环境搭建	
2	8	3	5		大数据开发: 案例演示及技术说明、 Python 编程与网络数据采集	
3	8	3	4	1	Python 编程与网络数据采集、 MapReduce 编程模型	
4	8	4	4		MapReduce 编程模型、 Hive 的原理和常用操作	
5	8	3	4	1	SparkRDD 编程、例会	
6	8	4	4		SparkSQL 编程	
7	8	3	5		Spark 机器学习工作流、 Spark-回归模型的应用、 Spark-分类模型	
8	8	6	2		Spark-聚类模型的应用、 Spark-关联规则模型的应用	
9	8	1	7		大数据开发: 案例开发过程演示	
10	8	6	1	1	大数据开发: 案例开发过程演示、例会	
11	8	5	2	1	项目需求讲解、项目组队、确定项目、项目业务分析	
12	8	8			项目业务分析与算法建模	
13	8	2	6		数据建模与构建数据中心	
14	8	6	2		数据采集、存储与清洗	
15	8	7	0	1	数据采集、存储与清洗; 例会	
16	8	8	0		数据采集、存储与清洗	
17	8	0	7	1	数据采集、存储与清洗、评审	
18	8	3	5		数据探索、数据转换与统计分析	
19-28	80		80		回归模型、分类模型、聚类模型、关联规则应用与优化	
29	8	0	8		数据可视化	
30	8	3	0	5	数据可视化、项目展示与评审	

西北农林科技大学教学日历

20 19 ——— 20 10 学年第 3 学期

课程名称: 软件开发综合训练-人工智能方向 授课专业: 计算机、软件、信管、电商 17

总学时: 240 本学期学时: 240 讲课学时: 80 实验学时: 15 其它环节: 8 学时

教研室(系)主任签字: 刘斌

教学院长(主任)签字: 李斌

天	学时	其中			计划教学内容	方式
		讲课	实验	其它环节		
1	8	8			人工智能综述和项目开发环境介绍	
2	8	4	4		机器视觉综述和 Linux 文件管理、用户管理、软件安装维护、其他常用操作命令	
3	8	4	4		人工智能算法简介和人工智能算法演练	
4	8	4	4		机器视觉算法简介和 Linux 网络管理、容器技术(软件容器封装与发布)	
5	8	4	4		C++基本语法、C++面向对象编程、C++高级编程和C++资源库讲解、C++与机器视觉	
6	8	4	4		python 基本语法、python 数据结构、python 常用开发包、python 矩阵运算和 python 连接数据库等	
7	8	4	4		图像处理基本知识讲解、图像处理基础算法: 图像增强、频域处理、灰度处理、二值化处理、Python-OpenCV 图像处理等具体技术讲解演练	
8	8	4	4		人脸检测与识别算法讲解, 人脸检测与识别算法设计、实现	
9	8	4	4		人体姿态识别算法讲解, 人体姿态识别算法设计、实现	
10	8	4	4		人脸表情识别算法讲解, 人脸表情识别算法设计、实现	
11	8	4	4		人脸表情识别算法讲解, 人脸表情识别算法设计、实现	
12	8	4	4		算法优化策略讲解, 算法模型评价方法讲解。学员调优实践。	
13	8	4	4		前端和后端开发技术讲解。包括: web 端开发、服务端开发、桌面程序开发、数据库开发等, 同步进行演练	
14	8	4	4		数据库设计、AI 推理模型归一化、视频数据采集模块、视频数据分析模块、视频分析数据存储、视频数据可视化模块打包集成	
15	8	4	4		最终确认分组选题情况, 老师介绍行业开发规范, 学员完成项目需求分析	
16	8	4	4		老师指导学员根据需求分析完成概要设计、详细设计、画原型	
17	8	4	4		老师指导学员完成系统技术架构设计	
18-27	80		80		各小组完成系统开发、测试、优化、撰写功能测试报告和性能测试报告	
28	8	4		4	小组完成系统部署和系统运行环境的搭建	
29	8	4		4	PPT 项目讲述/项目演示/项目技术答疑/项目总结/答辩材料收集	
30	8		8		完善项目内容, 材料归档, 对实训过程进行量化总结	