**《Spark大数据技术》实验课程教学大纲**

**字体、字号请参考范例**

**注意：**

**首字母大写**

**植物拉丁学名斜体**

**一、课程简介**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **课程中文名** | **Spark大数据技术** | | | | | | | | |
| **课程英文名** | **Spark Big Data Technology** | | | | | | **双语授课** | | □是 □否 |
| **课程代码** | 10114088 | **课程学分** | | 2.5 | **总学时数** | | | 60 | |
| **课程类别** | **□**专业基础课程  **☑**专业核心课程  **□**专业选修课程  **□**其他 | **课程性质** | | ☑必修  □选修  □其他 | **课程形态** | | | □线上  ☑线下  □线上线下混合式  □社会实践  □虚拟仿真实验教学 | |
| **考核方式** | □闭卷 □开卷 □课程论文 ☑课程作品 ☑汇报展示 □报告  ☑课堂表现 □阶段性测试 ☑平时作业 □其他（可多选） | | | | | | | | |
| **开课学院** | 大数据与智能工程学院 | | **开课**  **系(教研室)** | | | 数据科学与大数据技术 | | | |
| **面向专业** | 数据科学与大数据技术 | | **开课学期** | | | 第5学期 | | | |
| **课程负责人** | 黄金龙 | | **审核人** | | | 程东东、孙建乐 | | | |
| **先修课程** | Hadoop大数据技术 | | | | | | | | |
| **后续课程** | 无 | | | | | | | | |
| **选用教材** | 1. 林子雨. Spark编程基础（Scala版）. 人民邮电出版社, 2022年. | | | | | | | | |
| **参考书目** | 1. [肖芳](https://book.jd.com/writer/%E8%82%96%E8%8A%B3_1.html)，[张良均](https://book.jd.com/writer/%E5%BC%A0%E8%89%AF%E5%9D%87_1.html). Spark大数据技术与应用. 人民邮电出版社, 2018年.  2. [刘彬斌](https://book.jd.com/writer/%E5%88%98%E5%BD%AC%E6%96%8C_1.html)，[李柏章](https://book.jd.com/writer/%E6%9D%8E%E6%9F%8F%E7%AB%A0_1.html)等. Hadoop+Spark大数据技术. [清华大学出版社](https://book.jd.com/publish/%E6%B8%85%E5%8D%8E%E5%A4%A7%E5%AD%A6%E5%87%BA%E7%89%88%E7%A4%BE_1.html), 2018年. | | | | | | | | |
| **课程资源** | 大数据spark精选全集视频教程. 网易云课堂. | | | | | | | | |
| **课程简介** | Spark大数据技术是数据科学与大数据技术专业必修的一门综合实践性课程，旨在让学生掌握Spark这一主流的大数据技术和工具，具备快速处理大数据的能力。课程内容包括Scala函数式编程的基本语法、Spark环境搭建、Spark Shell命令、Spark相关术语、RDD概念、RDD的转换操作和行动操作、Spark SQL概念等，同时还包括Spark大数据技术的实践分析等。通过该课程学习，学生能够掌握Spark框架的搭建、配置和使用，以及利用RDD和 Spark SQL处理不同类型大数据，并对其进行分析处理的能力。 | | | | | | | | |

**二、课程目标**

**表2-1 课程目标**

|  |  |
| --- | --- |
| **序号** | **具体课程目标** |
| **课程目标1** | 能够应用数学、自然科学、大数据处理框架Spark的理论与方法，以及计算机及软件工程等相关领域的专业知识、技能与工具，设计针对数据分析问题的解决方案，解决大数据分析过程中所面临的复杂工程问题，并能够在设计环节中体现创新意识。 |
| **课程目标2** | 能够应用大数据处理框架Spark的基本原理，在数据分析的构思与设计阶段，通过文献研究、实验试验、工程推理、数学建模等方法，识别、表达、分析复杂数据分析问题及其解决方法、识别和判断数据分析问题的关键环节和参数，已获得分析模型、工程知识库等有效结论。 |
| **课程目标3** | 能够针对实时数据分析处理过程中涉及的复杂工程问题，选择和使用恰当的分析处理工具，进行信息处理、表达、建模、设计、模拟、验证，并能够在实践中了解国内主流的数据实时处理平台在各个应用中的优势和局限性；树立使用国产软件、保护版权的意识，激发民族自豪感；坚持自主创新，为建设世界科技强国而奋斗。 |

**表2-2 课程目标与毕业要求对应关系**

| **毕业要求** | **指标点** | **课程目标** |
| --- | --- | --- |
| **毕业要求1：**能够将数学、自然科学、工程基础、数据科学与大数据技术专业知识用于解决大数据应用领域的复杂工程问题。【H】 | 1.1能够掌握数学、自然科学、工程基础和专业知识，用于大数据应用领域相关工程问题的表述。 | **课程目标1** |
| **毕业要求3：**能够针对大数据应用领域的复杂工程问题，综合应用数据科学与大数据技术基本原理和方法，设计、开发满足特定应用需求的系统方案，并能够在设计环节中体现创新意识，考虑社会、健康、安全、法律、文化以及环境等多维度协同发展因素。【M】 | 3.3能够综合考虑社会、健康、安全、法律、文化以及环境因素，并能够进行系统测试、维护或优化升级。 | **课程目标3** |
| **毕业要求4：**能够基于数据科学与大数据技术的相应原理，采用科学方法对大数据应用领域中的复杂工程问题进行研究，制定技术路线、设计实验方案并开展实验，通过实验分析得到合理有效的结论。【H】 | 4.1能够综合应用数据科学与大数据技术基本原理，针对大数据应用领域复杂工程问题，通过文献研究、调研和分析，设计合适的解决方案。 | **课程目标2** |

**三、课程教学内容与方法**

**表3课程目标、教学内容和方法对应关系**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **序号** | **项目名称** | **项目来源** | **教学目标（观测点、重难点）** | **学时数** | **项目类型** | **要求** | **每组人数** | **教学方法** | **课程目标** |
| 1 | 项目1：Scala函数式编程 | 实验教材 | 掌握Scala基本语法 | 12 | 验证性 | 必做 | 1 | 课堂讲授  实验指导 | 课程目标1 |
| 完成Scala安装搭建实验环境 |
| 掌握Scala常用的集合操作（重点） |
| 掌握Scala函数式编程（难点） |
| 2 | 项目2：初识Spark及其环境搭建 | 实验教材 | 熟悉Spark构架与原理，及其三种集群模式 | 16 | 验证性 | 必做 | 1 | 课堂讲授  实验指导 | 课程目标2 |
| 掌握Spark环境搭建及其相关配置 | 课程目标3 |
| 掌握Spark交互式shell命令（重点） |
| 能够通过IDEA进行Spark作业的编程环境搭建（难点） |
| 了解Spark在国内的应用现状和案例 |
| 3 | 项目3：Spark核心概念（RDD） | 实验教材 | 熟悉RDD的概念及其创建方式（重点） | 20 | 综合性 | 必做 | 1 | 课堂讲授  实验指导 | 课程目标1 |
| 掌握RDD常见的转换操作（难点） |
| 掌握RDD常见的行动操作（难点） |
| 掌握RDD键值对操作算子 | 课程目标2 |
| 掌握RDD存储操作算子 |
| 完成Word Count综合案例 | 课程目标3 |
| 完成二次排序综合案例 |
| 4 | 项目4：Spark SQL概念 | 实验教材 | 理解Spark SQL概念和特点 | **12** | 综合性 | 必做 | 1 | 课堂讲授  实验指导 | 课程目标1 |
| 掌握DataFrame和DataSet的用法 |
| 掌握DataFrame的三种构建方法 |
| 掌握DataFrame常见的SQL相关操作 |
| 能够利用IDEA工程创建SparkSession对象 | 课程目标3 |
| 能够对DataFrame进行持久化 | 课程目标2 |
| 利用SparkSQL完成足球运动员的数据分析 |

## 四、课程考核

**（一）考核内容与考核方式**

**表4-1 课程目标、考核内容与考核方式对应关系**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **课程目标** | **考核内容** | **所属**  **学习项目** | **考核占比** | **考核方式** |
| 课程目标 1 | 1.Scala基本语法、环境搭建、集合操作、函数式编程 | 项目1 | 46% | 课堂交互  实验报告  课后练习  期末考核 |
| 2.RDD的概念及其创建方式、转换操作、行动操作 | 项目3 |
| 3.Spark SQL概念、DataFrame和DataSet的用法，及其常见的SQL操作 | 项目4 |
| 课程目标 2 | 1.Spark构架与原理、及其三种集群模式 | 项目2 | 35% | 课堂交互  实验报告  课后练习  期末考核 |
| 2.RDD键值对操作算子和存储操作算子 | 项目3 |
| 3.DataFrame的持久化 | 项目4 |
| 4.利用SparkSQL实现数据的交互式分析 |
| 课程目标 3 | 1.Spark交互式shell命令和IDEA的Spark项目创建，及Spark国内外研究现状 | 项目2 | 19% | 课堂交互  实验报告  课后练习  期末考核 |
| 2.利用RDD完成Word Count和二次排序综合案例 | 项目3 |
| 3.SparkSession的IDEA工程创建 | 项目4 |
| …… |  |  | 100% |  |

**表4-2 课程目标与考核方式矩阵关系**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 课程  目标 | 考核方式 | | | | | | | | 考核占比 | |
| 平时考核（40%） | | | | | | 期末考核  （60%） | |
| 课堂交互（10%） | | 实验报告  （20%） | | 课后练习  （10%） | |
| 课程目标1 | 60% | 60% | | 40% | | 40% | | 46%=10%\*60%+20%\*60%+10%\*40%+60%\*40% | |
| 课程目标2 | 30% | 20% | | 40% | | 40% | | 35%=10%\*30%+20%\*20%+10%\*40%+60%\*40% | |
| 课程目标3 | 10% | 20% | | 20% | | 20% | | 19%=10%\*10%+20%\*20%+10%\*20%+60%\*20% | |

**（二）成绩评定**

**1.平时成绩评定**

（1）课堂交互：通过学生考勤情况及其在课堂上发言、回答提问情况，评价学生的课程参与能力。（包括理想信念、家国情怀、社会责任、学习态度、学习兴趣、团队合作与终身学习意识等）

（2）实验报告：评价学生收集课程知识相关资料、撰写文本的能力；研究并提出问题，进而解决问题和合作研究的能力。

（3）课后练习：围绕课程的学习目标进行作业的设计。如让学生简述大数据的认识，考核学生对于大数据相关概念的理解情况，帮助学生将定义转化为自己的理解。

平时成绩（100%）=课堂交互（25%）+实验报告（50%）+课后练习（25%）

**2.期末成绩评定**

期末考试主要考察学生对数据分析知识体系、研究范畴和研究方向，为大数据分析与应用提供了强有力的工具。了解大数据项目开发中大数据计算框架发挥的重要作用。本课程以理论指导实践，以实践加深理论的理解，循序渐进通过Spark的讲解与实操，对大数据分析与应用产生更直观的认识。掌握专业相关复杂工程问题与社会可持续发展的客观评价能力和反思能力，方式为实验设计。

**3.总成绩评定**

总成绩由平时考核成绩和期末考核成绩构成。

总成绩（100%）=平时成绩（40%）+期末成绩（60%）

1. **评分标准**

**表5评分标准（非试卷考核项目）**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **考核项目** | **评分标准** | | | | |
| **优秀**  **(100>x≥90)** | **良好**  **(90> x≥80)** | **中等**  **(80> x≥70)** | **及格**  **(70> x≥60)** | **不及格**  **(x <60)** |
| 课程设计 | （1）课程设计符合课程性质，选题范围适中，具有较高的研究价值和意义，表现出很强的问题意识。（2）课程设计思路清晰且具有创新性。（3）理解课程设计原理，且代码能够正常运行（4）课程报告结构合理，组织严密，连贯一致。语言表达准确，叙述清楚，符合专业术语规范。（5）课程报告符合学术规范。 | （1）课程设计选题恰当合理，具有较高的研究价值和意义，表现出较强的问题意识。（2）课程设计思路较为清晰且具有一定创新性。（3）基本理解课程设计原理，且代码能够正常运行（4）课程报告结构较为合理，语言表达较为准确，叙述清楚，所使用的专业术语较为规范。（5）课程报告基本符合学术规范，无明显错误。 | （1）课程设计选题较为合理，具有一定的研究价值和意义，表现出一定的问题意识。（2）课程设计思路基本清晰且具有一定创新性（3）在一定程度上理解课程设计原理，且代码能够正常运行（4）课程报告结构较为合理，语言表达较为准确，叙述清楚，所使用的专业术语较为规范。  （5）课程报告基本符合学术规范，无明显错误。 | （1）课程设计选题具有一定的研究价值和意义，但凝练不够，问题意识欠佳。（2）课程设计思路基本清晰，但创新性不足（3）在一定程度上理解课程设计原理，且代码能够正常运行（4）课程报告结构基本合理，语言表达基本清楚，所使用的教育专业术语基本规范。（5）课程报告基本符合学术规范，有部分错误。 | （1）课程设计选题不符合课程性质，或主题不明确（2）课程设计思路不清晰（3）不理解课程设计原理，代码不能正常运行（4）课程报告结构不合理，语言不通顺，所使用的教育专业术语不规范。（5）课程报告明显不符合学术规范，或存在抄袭现象。 |

## 五、其他说明

本课程大纲依据2023版数据科学与大数据技术专业人才培养方案，由大数据与智能工程学院（部）数据科学与大数据技术教学系（教研室）讨论制定，大数据与智能工程学院（部）教学工作委员会审定，教务处审核批准，自2023级开始执行。

**字体、字号请参考范例**

**注意：**

**首字母大写**

**植物拉丁学名斜体**

**字体、字号请参考范例**

**注意：**

**首字母大写**

**植物拉丁学名斜体**

**字体、字号请参考范例**

**注意：**

**首字母大写**

**植物拉丁学名斜体**