**《云计算与大数据技术》教学大纲**

**一、课程基本信息**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **课程类别** | 专业拓展课程 | | **课程性质** | 理论 | **课程属性** | 选修 | |
| **课程名称** | 云计算与大数据技术 | | | **课程英文名称** | cloud computing and big data technology | | |
| **课程编码** | J35X058D | | | **适用专业** | 软件工程 | | |
| **考核方式** | 考查 | | | **先修课程** | 面向对象程序设计 | | |
| **总学时** | 32 | **学分** | | **2** | **理论学时** | | 20 |
| **实验学时/实训学时/ 实践学时/上机学时** | | | | **上机学时:**12 | | | |
| **开课单位** | | | | 计算机与信息学院 | | | |

**二、课程简介**

《云计算与大数据技术》是软件工程专业中一门综合性很强的基础课程，主要内容包括云计算概论、云计算基础（云计算关键技术、云交付模式、云部署模式、云计算的优势与挑战以及典型云应用）、虚拟化相关知识、云计算应用、大数据的基本概念、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、基于内存的大数据处理架构Spark、大数据在各个领域的应用。

本课程的目的与任务是使学生通过本课程的学习，从基本概念入手，由浅入深学习各种相关知识，学会云计算的相关关键技术和云部署模式、大数据相关技术，介绍Hadoop MapReduce和Spark等大数据相关技术，掌握云计算和大数据的相关思想。本课程要求学生掌握云计算和大数据的基础知识和理论，让学生掌握集群的构建和如何使用大数据分析技术解决特定业务领域的问题，培养学生大数据分析的能力，为将来从事大数据挖掘研究工作，为更深入地学习和今后的实践打下良好的基础。

**三、课程教学目标**

|  |  |  |  |
| --- | --- | --- | --- |
| **课程教学目标** | | **支撑人才培养规格指标点** | **支撑人才培养规格** |
| **知**  **识**  **目**  **标** | **目标1：**培养学生分析问题和解决问题的能力，使学生掌握云计算和大数据存储的基本技巧和方法，熟悉云计算和大数据分析的基本技术，并能熟练运用一些常用的计算框架、数据库，解决一些较综合的问题。 | 4-1:具有运用工程基础知识的能力。 | 4.工程基础知识 |
| **能**  **力**  **目**  **标** | **目标2：**  在理论学习和实践中掌握云计算和大数据存储、运算、分析的基本能力，逐步掌握云计算和大数据分析的思想和方法。 | 5-2:具备应用型软件开发实践能力，能根据软件需要，设计简单的解决方案。 | 5.工程实践能力。 |
| **素**  **质**  **目**  **标** | **目标3：**  通过本课程的学习，培养作为一个软件  工程技术人员必须具备的坚持不懈的学习精神，严谨治学的科学态度和积极向上的价值观，为未来的学习、工作和生活奠定良好的基础。 | 2-1：具有综合运用各种手段查阅文献、获取信息的能力。  2-3：具有较好的创新创业能力。  能针对个人或职业发展的需求，采用合适的方法自主学习，以适应社会和行业发展。 | 2.综合素质能力 |

**四、课程主要教学内容、学时安排及教学策略**

**（一）理论教学**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **教学模块** | **学时** | **主要教学内容与策略** | **学习任务安排** | **支撑课程目标** |
| 大数据概述和大数据处理架构Hadoop | 2 | **指导内容：**介绍大数据的基本概念和应用领域，并阐述大数据、云计算和物联网的相互关系。介绍大数据处理架构Hadoop。Hadoop生态系统、Hadoop的安装与使用。  **重点：**大数据的基本概念，Hadoop生态系统。  **难点：**Hadoop的安装与使用。  **思政元素：**培养学生科学探索精神。  **教学方法与策略：**线下教学。对于思想、原理在课堂上予以讲授，对于实践过程部分安排上机实践。课堂主要运用讲授法和案例法开展教学，辅以启发式提问拓宽学生学习思路。 | 课前：预习并了解Hadoop。  课堂：学习本教学模块。  课后：Hadoop安装。 | 目标1  目标2  目标3 |
| 分布式文件系统HDFS | 2 | **指导内容：**分布式文件系统HDFS的基本原理和使用方法。讲述分布式文件系统、HDFS的存储原理、HDFS的数据读写过程、HDFS编程实践。  **重点：**HDFS的存储原理、HDFS的数据读写过程。  **难点：**HDFS编程实践。  **思政元素：**培养学生科学探索精神。  **教学方法与策略：**线下教学。对于思想、原理在课堂上予以讲授，对于实践过程部分安排上机实践。课堂主要运用讲授法和案例法开展教学，辅以启发式提问拓宽学生学习思路。 | 课前：预习并了解HDFS。  课堂：学习本教学模块。  课后：HDFS编程实践。 | 目标1  目标2  目标3 |
| 分布式数据库HBase | 2 | **指导内容：**分布式数据库HBase的基本原理和使用方法。讲述HBase数据模型、HBase的实现原理、HBase运行机制、HBase编程实践。  **重点：H**Base数据模型  **难点：**HBase编程实践  **思政元素：**培养学生科学探索精神。  **教学方法与策略：**线下教学。对于思想、原理在课堂上予以讲授，对于实践过程部分安排上机实践。课堂主要运用讲授法和案例法开展教学，辅以启发式提问拓宽学生学习思路。 | 课前：预习并了解HBase。  课堂：学习本教学模块。  课后：HBase编程实践。 | 目标1  目标2  目标3 |
| NoSQL数据库  和云数据库 | 2 | **指导内容：**NoSQL数据库的概念和基本原理。讲述NoSQL的四大类型、NoSQL的三大基石。云数据库。讲述云数据库产品、云数据库系统架构。  **重点：**NoSQL的四大类型。云数据库系统架构。  **难点：**NoSQL的三大基石。  **思政元素：**培养学生科学探索精神。  **教学方法与策略：**线下教学。对于思想、原理在课堂上予以讲授，对于实践过程部分安排上机实践。课堂主要运用讲授法和案例法开展教学，辅以启发式提问拓宽学生学习思路。 | 课前：预习并了解NoSQL数据库。预习并了解云数据库。  课堂：学习本教学模块。  课后：NoSQL和关系数据库的操作比较。熟悉云数据库产品 | 目标1  目标2  目标3 |
| MapReduce | 2 | **指导内容：**分布式并行编程模型MapReduce原理和使用方法。讲述MapReduce的工作流程、实例分析：WordCount、MapReduce的具体应用、MapReduce编程实践。  **重点：**MapReduce的工作流程。  **难点：**实例分析：WordCount。  **思政元素：**培养学生科学探索精神。  **教学方法与策略：**线下教学。对于思想、原理在课堂上予以讲授，对于求解过程部分安排上机实践。课堂主要运用讲授法和案例法开展教学，辅以启发式提问拓宽学生学习思路。 | 课前：预习并了解MapReduce。  课堂：学习本教学模块。  课后：MapReduce编程实践。 | 目标1  目标2  目标3 |
| Spark | 2 | **指导内容：**Spark原理与基础编程。讲述Spark生态系统、Spark的部署和应用方式、Spark编程实践**。**  **重点：**Spark的部署和应用方式。  **难点：**Spark编程实践**。**  **思政元素：**培养学生科学探索精神。  **教学方法与策略：**线下教学。对于思想、原理在课堂上予以讲授，对于求解过程部分安排上机实践。课堂主要运用讲授法和案例法开展教学，辅以启发式提问拓宽学生学习思路。 | 课前：预习并了解Spark。  课堂：学习本教学模块。  课后：Spark编程实践**。** | 目标1  目标2  目标3 |
| 流计算 | 2 | **指导内容：**流计算。讲述流计算的处理流程、流计算的应用、开源流计算框架Storm、Spark Streaming。  **重点：**流计算的处理流程。  **难点：**流计算的应用。  **思政元素：**培养学生科学探索精神。  **教学方法与策略：**线下教学。对于思想、原理在课堂上予以讲授，对于求解过程部分安排上机实践。课堂主要运用讲授法和案例法开展教学，辅以启发式提问拓宽学生学习思路。 | 课前：预习并了解流计算。  课堂：学习本教学模块。  课后：流计算的应用。 | 目标1  目标2  目标3 |
| 云计算基础 | 2 | **指导内容：**分布式计算和云计算的关键技术。云计算的关键技术包括虚拟化技术、分布式海量数据存储技术、云平台技术、并行编程技术和数据管理技术；掌握四种云交付模型，分别是软件即服务、平台即服务、基础设施即服务和容器即服务，并掌握前三种基本模型之间的区别；掌握三种云部署模式，分别是公有云、私有云和混合云；了解云计算的优势和典型的云应用。  **难点：**云计算的关键技术。  **思政元素：**培养学生科学探索精神。  **教学方法与策略：**线下教学。对于思想、原理在课堂上予以讲授，对于求解过程部分安排上机实践。课堂主要运用讲授法和案例法开展教学，辅以启发式提问拓宽学生学习思路。 | 课前：预习并了解分布式计算和云计算的关键技术。  课堂：学习本教学模块。  课后：分布式计算和云计算的关键技术应用。 | 目标1  目标2  目标3 |
| 云计算机制与应用 | 2 | **指导内容：**云计算设施机制、云管理机制、云监控机制和特殊云机制；掌握每一种机制的概念、特点和作用。介绍常见的云计算应用，包括谷歌的云计算平台和应用、亚马逊的弹性计算云、IBM的蓝云计算平台、清华大学透明计算平台、阿里云和Microsoft Azure。  **难点：**云计算设施机制、云管理机制。  **思政元素：**培养学生科学探索精神。  **教学方法与策略：**线下教学。对于思想、原理在课堂上予以讲授，对于求解过程部分安排上机实践。课堂主要运用讲授法和案例法开展教学，辅以启发式提问拓宽学生学习思路。 | 课前：预习并了解云计算机制与应用。  课堂：学习本教学模块。  课后：云计算应用。 | 目标1  目标2  目标3 |
| 虚拟化 | 2 | **指导内容：**了解虚拟化技术；了解虚拟化技术的分类；掌握系统虚拟化、虚拟化与云计算；了解相关开源技术以及虚拟化未来的发展趋势；了解虚拟化的发展历史以及虚拟化带来的好处。  **难点：**虚拟化与云计算。  **思政元素：**培养学生科学探索精神。  **教学方法与策略：**线下教学。对于思想、原理在课堂上予以讲授，对于求解过程部分安排上机实践。课堂主要运用讲授法和案例法开展教学，辅以启发式提问拓宽学生学习思路。 | 课前：预习并了解虚拟化与云计算。  课堂：学习本教学模块。  课后：虚拟化与云计算。 | 目标1  目标2  目标3 |

**（二）实践教学**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **实践类型** | **项目名称** | **学时** | **主要教学内容** | **项目**  **类型** | **项目**  **要求** | **支撑课程目标** |
| 上机 | 实验1:熟悉常用的Linux操作和Hadoop操作 | 2 | 指导内容：熟悉常用的Linux操作和Hadoop操作。  重点：Hadoop操作。  难点：Hadoop操作。  思政元素：要求学生必须坚持实事求实、严谨的科学态度。培养学生科学探索精神。 | 验证 | 学习实践一些常用的Linux命令、熟悉常用的Hadoop操作 | 目标1  目标2  目标3 |
| 上机 | 实验2:熟悉常用的HDFS操作 | 2 | 指导内容：HDFS在Hadoop体系结构中的角色、熟练使用HDFS操作常用的Shell命令、熟悉HDFS操作常用的Java API。  重点：HDFS操作常用的Shell命令。  难点：HDFS操作常用的Java API。  思政元素：要求学生必须坚持实事求实、严谨的科学态度。培养学生科学探索精神。 | 验证 | 操作系统：Linux（建议Ubuntu16.04及以上版本）、Hadoop版本：3.1.3及以上版本、JDK版本：1.8及以上版本。 | 目标1  目标2  目标3 |
| 上机 | 实验3:熟悉常用的HBase操作 | 2 | 指导内容：HBase操作常用的Shell命令、HBase操作常用的Java API。  重点：HBase操作常用的Shell命令。  难点：HBase操作常用的Java API。  思政元素：要求学生必须坚持实事求实、严谨的科学态度。培养学生科学探索精神。 | 验证 | HBase版本：2.2.2或以上。 | 目标1  目标2  目标3 |
| 上机 | 实验4：NoSQL和关系数据库的操作比较 | 2 | 指导内容：（1）四种数据库(MySQL、HBase、Redis和MongoDB)的概念以及不同点；  （2）熟练使用四种数据库操作常用的Shell命令；  （3）熟悉四种数据库操作常用的Java API。  重点：四种数据库操作常用的Shell命令。  难点：四种数据库操作常用的Java API。  思政元素：要求学生必须坚持实事求实、严谨的科学态度。培养学生科学探索精神。 | 验证 | （1）MySQL版本：5.6或以上（2）HBase版本：2.2.2  或以上（3）Redis版本：5.0.5  或以上（4）MongoDB版本：4.2.0或以上。 | 目标1  目标2  目标3 |
| 上机 | 实验5：MapReduce初级编程实践 | 2 | 指导内容：（1）基本的MapReduce编程方法（2）用MapReduce解决一些常见的数据处理问题，包括数据去重、数据排序和数据挖掘等。  重点：基本的MapReduce编程方法。  难点：用MapReduce解决一些常见的数据处理问题。  思政元素：要求学生必须坚持实事求实、严谨的科学态度。培养学生科学探索精神。 | 验证 | Hadoop版本：3.1.3或以上版本。 | 目标1  目标2  目标3 |
| 上机 | 实验6:阿里云 | 2 | 指导内容：创建阿里云服务器  重点：服务的申请。  难点：云服务器的使用。  思政元素：要求学生必须坚持实事求实、严谨的科学态度。培养学生科学探索精神。 | 验证 | 创建阿里云服务器和云服务器的使用。 | 目标1  目标2  目标3 |
|  | 备注： 项目类型填写验证、综合、设计、训练等。 | | | | | |

**五、学生学习成效评估方式及标准**

考核与评价是对课程教学目标中的知识目标、能力目标和素质目标等进行综合评价。在本课程中，学生的最终成绩是由平时成绩（作业、考勤）、期末考试2个部分组成。

1.平时成绩（占总成绩的40%）：

|  |  |
| --- | --- |
| **等级** | **评 分 标 准** |
| **1.作业；2.考勤** |
| 优秀  （90～100分） | 1.作业书写工整、书面整洁；90％以上的习题解答正确或实验习题结果准确无误。  2.出勤率100%。 |
| 良好  （80～89分） | 1.作业书写工整、书面整洁；80％以上的习题解答正确或实验习题结果准确无误。  2.出勤率90%。 |
| 中等  （70～79分） | 1.作业书写工整、书面整洁；70％以上的习题解答正确或实验习题结果准确无误。  2.出勤率80%。 |
| 及格  （60～69分） | 1.作业书写工整、书面整洁；60％以上的习题解答正确或实验习题结果准确无误。  2.出勤率70%。 |
| 不及格  （60以下） | 1.作业书写不工整、书面不整洁；60％以下的习题解答正确或实验习题结果准确无误。  2.出勤率60%以下。 |

2.期末考试（占总成绩的60%）：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **考核**  **模块** | **考核内容** | **主要**  **题型** | **支撑目标** | **分值** |
| 大数据概述和大数据处理架构Hadoop | 介绍大数据的基本概念和应用领域，并阐述大数据、云计算和物联网的相互关系。介绍大数据处理架构Hadoop。Hadoop生态系统、Hadoop的安装与使用。 | 填空题、判断题、选择题 | 目标1目标2 | 15 |
| 分布式文件系统HDFS | 分布式文件系统HDFS的基本原理和使用方法。分布式文件系统、HDFS的存储原理、HDFS的数据读写过程、HDFS编程实践。 | 填空题、判断题、选择题、编程题、分析题 | 目标1  目标2 | 10 |
| 分布式数据库HBase | 分布式数据库HBase的基本原理和使用方法。HBase数据模型、HBase的实现原理、HBase运行机制、HBase编程实践。 | 填空题、判断题、选择题、编程题、分析题 | 目标1  目标2 | 10 |
| NoSQL数据库  和云数据库 | NoSQL数据库的概念和基本原理。讲述NoSQL的四大类型、NoSQL的三大基石、编程实践。云数据库产品、云数据库系统架构 | 填空题、判断题、选择题、编程题、分析题 | 目标1  目标2 | 10 |
| MapReduce | 分布式并行编程模型MapReduce原理和使用方法。MapReduce的工作流程、实例分析：WordCount、MapReduce的具体应用、MapReduce编程实践。 | 填空题、判断题、选择题、编程题、分析题 | 目标1  目标2 | 10 |
| Spark | Spark原理与基础编程。讲述Spark生态系统、Spark的部署和应用方式、Spark编程实践**。** | 填空题、判断题、选择题、编程题、分析题 | 目标1  目标2 | 5 |
| 流计算 | 流计算的处理流程、流计算的应用、开源流计算框架Storm、Spark Streaming。 | 填空题、判断题、选择题 | 目标1  目标2 | 5 |
| 云计算基础 | 分布式计算和云计算的关键技术。虚拟化技术、分布式海量数据存储技术、云平台技术、并行编程技术和数据管理技术。四种云交付模型，分别是软件即服务、平台即服务、基础设施即服务和容器即服务。三种云部署模式，分别是公有云、私有云和混合云。 | 填空题、判断题、选择题 | 目标1  目标2 | 20 |
| 云计算机制与应用 | 云计算设施机制、云管理机制、云监控机制和特殊云机制；云计算应用，包括谷歌的云计算平台和应用、亚马逊的弹性计算云、IBM的蓝云计算平台、清华大学透明计算平台、阿里云和Microsoft Azure。 | 填空题、判断题、选择题 | 目标1  目标2 | 10 |
| 虚拟化 | 虚拟化技术；虚拟化技术的分类；系统虚拟化、虚拟化与云计算；相关开源技术以及虚拟化未来的发展趋势；虚拟化的发展历史以及虚拟化带来的好处。 | 填空题、判断题、选择题 | 目标1  目标2 | 5 |

|  |  |  |
| --- | --- | --- |
| **序号** | **教学安排事项** | **要 求** |
| 1 | 授课教师 | 职称：助教及以上 学历（位）：本科以上  其他：具有硕士以上学位的高级工程师 |
| 2 | 课程时间 | 周次：1-16  节次：1-8节均可，2节联排。 |
| 3 | 授课地点 | ☑教室 ☑实验室 □室外场地  □其他 |
| 4 | 学生辅导 | 线上方式及时间安排：微信或企业微信等，每周一次，每次45分钟。  线下地点及时间安排：固定教室，每4周一次，每次45分钟。 |

1. **教学安排及要求**

**七、选用教材**

# [1] 大数据技术原理与应用（第3版），**林子雨，**人民邮电出版社，2021年1月。

# [2] 大数据基础编程、实验和案例教程（第2版），**林子雨，**清华大学出版社，2020年10月。

# [3] 云计算与大数据技术（第2版），吕云翔，清华大学出版社，2023年7月

**八、参考资料**

[1]安俊秀，王鹏，靳宇倡主编.《Hadoop大数据处理技术基础与实践》.北京：人民邮电出版社，2015年。

[2]张伟洋 主编.《Hadoop大数据技术开发实战》.北京：清华大学出版社，2019年。

[3]黑马程序员主编.《Hadoop大数据技术原理与应用》. 北京：清华大学出版社，2019年。

大纲执笔人： 雷明敏

讨论参与人: 王浩亮，刘天贻

学院院长：王浩亮

学院（部）审核人：刘天贻