大数据编程

课程编号：（项目统一编码） 授课对象：本科生

学分： 4 任课教师：姚凯

课程类型：必修 开课学期：大二上

先修课程：计算机应用基础

任课教师联系方式：

办公电话：

电子邮箱：yaokai@cufe.edu.cn

答疑时间：邮件预约，请提前24小时。

一、课程概述

大数据是近年来最热门的研究话题之一，由此产生的技术也被广泛地应用在了多个领域。现在的社会是一个高速发展的社会，科技发达，信息流通，人们之间的交流越来越密切，生活也越来越方便，大数据就是这个高科技时代的产物。大数据分析常和云计算联系到一起，因为实时的大型数据集分析需要像MapReduce一样的框架来向数十、数百或甚至数千的电脑分配工作。因此，大数据编程需要适应大数据处理的特点，与传统编程方法有所区别。

该课程的主要内容包含大数据技术的发展现状和大数据技术在数据处理中主要包含的方法，让学生初步认识大数据编程方法的基础知识。同时，为了让同学们深入理解大数据技术，本课程将介绍营销建模、计算机技术及统计方法在真实商业中的应用场景，深入浅出地掌握支撑生活中大数据应用背后所用到的相关技术原理。了解大数据技术在大家生活中的应用场景外，本课程还着重提高大家的动手编程能力。课程中会选择适合学生的大数据编程技术进行上机实践，通过亲手编程实现平时广泛应用在大家身边的大数据分析技术。深入浅出地掌握大数据编程的实现方法和在真实商业中的应用场景，为以后在该领域的学习和研究打下坚实的基础。

二、课程目标

本处应当**至少涵盖两项项目的学习目标**。

在完成课程的基础上，学生应该能够：（√√为最适用的目标，√为一般适用的目标）

|  |  |  |  |
| --- | --- | --- | --- |
| 本课程适用目标 | 学习目标 | 具体描述 | |
| √ | 具有良好的沟通能力 | 能够在不同文化背景下的团队有效地开展工作 | |
| 具有良好的书面及口语表达能力 | |
| 具备团队协作能力 | |
| √√ | 具有解决问题的能力 | 掌握定量分析工具和方法，熟练收集和处理信息 | |
| 采用适当的分析框架，得出合理的结论 | |
| 具有反思和批判的能力 | |
| √ | 践行商业伦理 | 建立商业伦理准则 | |
| 进行商业伦理实践 | |
| √ | 全球化视野 | 了解全球经济环境 | |
| 掌握跨文化管理的知识与技能 | |
| √√ | 掌握专业知识 | 营销管理 | 具备良好的人文与社会素养 |
| 掌握基本的营销管理知识 |
| 具备良好的人文与社会素养 |
| 人力资源管理 | 具备良好的人文与社会素养 |
| 掌握基本的人力资源管理知识 |
| 工商管理 | 了解一定的人文数学与社会素养 |
| 掌握基本的企业管理与运营知识 |
| 物流管理 | 具备良好的数理基础与社会知识 |
| 掌握供应链管理理论与管理方法 |

除了上述目标以外，**学生还能够通过课程学习掌握可以包括且不限于以下内容**：（√√为最适合的知识点，√为一般适合的知识点）

|  |  |  |
| --- | --- | --- |
| No. | 课程预期学习知识 | 本课程适合的知识点 |
| 1. | 写作和口头上能有效交流 | √ |
| 2. | 能识别道德问题并用具有社会责任的方式去解决 |  |
| 3. | 能分析、构建问题 | √√ |
| 4. | 在商业、管理环境下能使用新技术 | √√ |
| 5 | 能有效同别人合作 | √√ |
| 6 | 能在不同环境下高效工作 |  |
| 7 | 反思思维，能明白自己在社会中的位置 |  |
| 8 | 能将商业及管理知识转化为实践 | √√ |
| 9 | 全球化背景下组织中的管理知识，包括：经济、政治、监管、法律、技术等 | √ |
| 10 | 社会责任，包括可持续发展，管理中的伦理问题，以及如何用具有社会责任的方式去解决伦理问题 |  |
| 11 | 金融理论、分析、报告、市场 | √ |
| 12 | 公司的系统和流程办法，包括规划和设计，生产/运营，供应链，营销和分销 |  |
| 13 | 组织中的团队行为和个人行为特征 |  |
| 14 | 影响商业实践的信息技术和统计/定量的方法，包括数据的创建、数据共享、数据分析、数据挖掘、数据报告和跨组织间的数据储存，以及伴随的相关道德问题 | √√ |

三、教学进度安排

下表中，教学活动可以包括且并不止于（**根据课程实际情况，不必全选，也可以列出以下没有的教学活动**）：

* 小组作业（包含论文、案例分析等）
* 小组展示
* 课堂案例分析
* 课堂视频播放
* 课程实验
* 在线答疑和交流

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 周次 | 教学内容 | 教学活动（请在上面的教学活动中选出合适的教学活动，并非每一周都需要这些活动，但是每学期使用两次以上的教学活动） | 教学活动简单描述  （主题、参与者、大致要求） | 教学活动所满足的学习目标 | 阅读材料 |
| 第一周 | 大数据概述 |  | 介绍大数据发展过程 |  |  |
| 第二周 | 大数据应用技术 | 案例分析 | 介绍广泛应用在大家身边的大数据应用技术 |  | 1.阅读材料  2.案例分析视频 |
| 第三周 | 数据获取方法 |  | 数据获取是大数据营销中至关重要的环节 |  | 1. 阅读材料 |
| 第四周 | 编写数据爬虫 | 上机演示 | 学生自行编写程序获取数据 |  | 1. 阅读材料 |
| 第五周 | 互联网广告 |  | 介绍 |  | 1. 阅读材料  2. 参考教材 |
| 第六周 | 精准广告投放算法 |  | 介绍精准广告投放中用到的核心技术 |  | 1. 阅读材料  2. 参考教材 |
| 第七周 | 个性化推荐应用 |  | 讲述互联网营销中应用最为广泛的应用 |  | 1.参考教材 |
| 第八周 | 个性化推荐应用 |  | 讲述互联网营销中应用最为广泛的应用 |  | 1.参考教材 |
| 第九周 | 小组 |  |  |  |  |
| 第十周 | 并行计算和云计算 |  | 介绍如果并行处理海量数据 |  | 1. 参考教材 |
| 第十一周 | 并行编程 | 上机演示 | 动手实现并行编程并处理海量数据 |  | 1.阅读资料 |
| 第十二周 | 线程同步 |  | 介绍线程之间如何共享资源并避免死锁 |  | 1、阅读资料 |
| 第十三周 | 分布式计算 |  | 如何利用多台计算机处理海量数据 |  | 1. 阅读资料  2. 参考教材 |
| 第十四周 | 海量数据处理实验 | 上机实验 | 给学生海量数据，上机实验，快速将数据处理完 |  | 1. 阅读资料 |
| 第十五周 | Hadoop |  | 讲解hadoop的基本原 |  | 1. 阅读资料  2. 参考教材 |
| 第十六周 | Hadoop平台下的大数据分析工具 |  | 介绍基于hadoop平台下的主要分析工具 |  | 1. 阅读资料  2. 参考教材 |
| 第十七周 | 成果展示 |  | 小组形式汇报团队学习成果 |  |  |
| 第十八周 | 期末考试 |  |  |  |  |

六、课程使用的IT技术手段

七、教材

Anand Rajaraman，Jeffrey David Ullman著 王斌译， 大数据：互联网大规模数据挖掘与分布式处理，人民邮电出版社

黄宜华，深入理解大数据：大数据处理与编程实践, 机械工业出版社

八、参考书目和阅读材料

迈尔-舍恩伯格，库克耶 著，盛杨燕，周涛 译，大数据时代，浙江人民出版社

麦德奇,布朗，大数据营销:定位客户，机械工业出版社

课堂分发相关阅读材料

九、教学辅助材料，如CD、录像等

十、教学要求（根据课程需要自己提要求）

学生必须出席每一堂课，如果临时出现身体或其他特殊情况不能上课，应提前与老师联系并填写正式请假条，请假条在课前提交。旷课将影响本课程成绩。

学生必须课堂小组报告，不得无故缺席。如果临时因身体等特殊原因不能参加小组报告和考试，应事前请假。

不迟到、不早退。上课时将手机等电子设备关闭或置于静音状态。

按时、按规定格式提交作业。

十一、考核方法（根据课程需要自行设定考核方法）

小组作业（40%）

期末考试（40%）

课堂参与和讨论（10%）

出勤（10%）

十二、学术诚信与学术不端（这部分所有课程都一样，教师不填也可）

本课程所安排的商业计划书报告，或其他任何其他形式的口头或书面作业，必须是原创的，同时是专门为本课程所做的。任何抄袭或伪造行为都将导致该名同学在本课程将得到0分的成绩，这意味着该名学生将需重修该课程或改选其他课程以完成学校对学分的要求。

学生可以引用他人的学术论述和成果，如果引用他人成果，必须注明来源和作者，以尊重他人知识产权。