

上海建桥学院课程教学进度计划表

一、基本信息

课程代码	2050559	课程名称	大数据挖掘、分析与应用
课程学分	4	总学时	64
授课教师	蔡文英	教师邮箱	Cai.wenying@huatec.com
上课班级	计科中兴 B15-2	上课教室	四教 303; 信息 418
答疑时间	周一下午三点到五点半 周三早上 9 点到十一点, 下午三点到五点半		
主要教材	大数据技术基础: 基于 Hadoop 与 Spark		
参考资料	百度; 慕课网		

二、课程教学进度

周次	教学内容	教学方式	作业
1	大数据的发展历程以及大数据对国内外各行各业的影响; 大数据的概念及其特征; 大数据的来源, 理解大数据在技术、安全等方面面临的挑战和研究大数据的意义; 大数据的存储与计算模式的相关概念, 了解其中的关键技术及基本思想; 大数据的典型应用场景, 学会用创新性思维来看待大数据	讲授	课后复习, 下节课提问
2	1、Hadoop 的发展过程和优势; Hadoop 的生态系统以及其中的基本概念; Hadoop 的版本发行情况; 掌握 Hadoop 的基本结构	讲授	课后复习, 下节课提问
3	Hadoop 的分布式文件系统	讲授; 实操	课后复习, 下节课提问
4	Hadoop 分布式计算框架	讲授	课后复习, 下节课提问
5	MapReduce API 的种类; MapReduce Java API 的常用类	讲授	课后复习, 下节课提问
6	MapReduce 的数据类型; MapReduce Java API 的编程思路	讲授	课后复习, 下节课提问

注: 课程教学进度计划表电子版公布在本学院课程网站上, 并发送到教务处存档。

7	序列化；MapReduce 的输入输出类型与格式	讲授；实操	课后复习，下节课提问
8	MapReduce 的 map 任务和 reduce 任务； MapReduce 任务的配置和执行	讲授	课后复习，下节课提问
9	掌握 Spark 环境搭建；Spark 的发展历程和应用现状，熟悉 Spark 的特点	讲授	课后复习，下节课提问
10	Spark 的数据模型 RDD 的处理过程；常用的 RDD 算子的使用方法；	讲授	课后复习，下节课提问
11	Spark 的任务调度机制；Spark 的技术体系架构	讲授	课后复习，下节课提问
12	Spark Streaming 的优势；Spark Streaming 的工作机制	讲授	课后复习，下节课提问
13	DStream 流的 Window 操作机制；DStream 的使用方法	讲授	课后复习，下节课提问
14	Spark Streaming 的性能优化；Spark SQL 的工作原理	讲授	课后复习，下节课提问
15	Spark SQL 中的 DataSet 与 DataFrame 类及其常用方法；掌握常见数据源的获取与查询	讲授	课后复习，下节课提问
16	复习周	讲授	课后复习，下节课提问
17	考试周	无	

三、评价方式以及在总评成绩中的比例

总评构成(1+X)	评价方式	占比
1	期末闭卷考	40%
X1	个人项目报告	40%
X2	课堂小测验	20%

任课教师：蔡文英 系主任审核： 日期：

