

《大数据概论》(课程代码: 13421)

自学考试实践考核大纲

华南理工大学自学考试

I 课程性质与目标

一、课程性质和特点

本课程是为培养学生的数据素养，掌握典型的大数据分析过程，并运用掌握的理论知识解决现实的大数据分析问题而设置的一门专业实践课程。

二、课程目标

通过本课程的学习，使考生了解大数据应用在其他领域中的应用，熟悉大数据处理的基本流程，深入了解典型的大数据分析过程，具备运用大数据思维、模型和工具解决实际问题的能力，为考生之后的学习和工作奠定坚实的基础。

课程目标如下：

1. 了解大数据在各领域的应用现状，包括互联网领域、生物医学领域、物流领域等。
2. 理解典型的大数据分析流程。
3. 了解数据存储与管理概念及其常用技术，掌握文件系统、关系数据库等方法。
4. 了解数据处理与分析概念及其常用技术，掌握分类、聚类等方法，能熟练运用表格处理工具或关系数据库进行数据的分类或聚类等。
5. 了解数据可视化技术的发展历程、功能及其常用技术，能熟练运用信息图表工具等工具进行数据可视化。

三、与相关课程的联系与区别

学习本课程前，考生应具备一定的数据素养，对大数据概念有基本的认识、熟悉大数据各个环节的相关技术。考生应掌握了一定的数据处理能力，包括数据采集、数据清洗、数据转换、数据存储与管理、数据可视化等基础课程知识。

本课程为学习大数据相关应用系统分析和开发系列课程打下基础，并能够帮助考生较好地完成毕业设计。

四、课程的重点和难点

本课程的重点内容包括：数据采集与预处理、数据存储与管理、数据处理与分析等相关内容，以及熟练运用各相关大数据处理方法解决实际问题的能力。

本课程的难点内容包括：关于数据采集与预处理、数据存储与管理、数据处理与分析的理解，数据处理等方法与实际问题实践的相结合，对数据分析过程的熟练掌握。

五、考核目标

本课程的实践考核目标要求考生能运用大数据思维、模型和工具，熟练掌握典型的大数据分析过程来合理解决实际问题；能够灵活运用数据存储系统（如 MySQL、文件系统等）、可视化工具（如 Excel）和数据处理算法（如分类、聚类算法）等解决实际的大数据分析问题。

在学习本课程之前，考生应完成大数据概论课程的理论学习。

II 课程内容与考核要求

第1章：数据采集与预处理和数据存储

（包含教材第5章5.1-5.3节，第6章6.1节）

一、课程内容

- 1.1 数据采集
- 1.2 数据清洗
- 1.3 数据转换
- 1.4 数据存储

二、学习目的与要求

通过对本章的学习，达到如下要求：

1. 数据采集、数据清洗和数据转换。

2. 掌握数据清洗的主要内容并能选用合适的工具进行数据清洗，如缺失值处理、异常值处理、数据类型转换、重复值处理等。

3. 掌握数据转换策略的主要内容并能选用合适的工具进行数据转换，如平滑处理、聚集处理、数据泛化处理、规范化处理和属性构造处理等。

4. 掌握文件系统和关系数据库进行数据存储的方法。

三、考核知识点与考核要求

1. 数据采集

领会：传统数据采集与大数据采集的区别；数据采集方法。

应用：网络爬虫。

2. 数据清洗

领会：数据清洗的内容；数据清洗的注意事项。

综合应用：运用工具（如 Excel、MySQL）进行数据清洗，如缺失值处理、异常值处理、数据类型转换、重复值处理等。

3. 数据转换

领会：数据转换策略。

综合应用：运用工具（如 Excel、MySQL）进行数据转换，如平滑处理、聚集处理、数据泛化处理、规范化处理和属性构造处理等。

4. 数据存储

领会：文件系统和关系数据库。

应用：运用工具（如 Excel、MySQL）进行数据存储。

第2章：数据处理分析与数据可视化

（包含教材第7章7.1-7.2节，第8章8.1-8.3节）

一、课程内容

2.1 数据处理与分析

2.2 数据可视化

二、学习目的与要求

通过对本章的学习，达到如下要求：

1. 领会数据处理与分析的概念、部分数据挖掘算法（例如分类）；能选用合适工具进行数据处理与分析。

2. 领会数据可视化的概念及功能；能选用合适工具生成可视化图表。

三、考核知识点与考核要求

1. 数据处理与分析

领会：数据处理与数据挖掘、数据分析。

综合应用：运用工具（如 Excel、MySQL）进行数据的分类等处理。

2. 数据可视化

领会：数据可视化的发展历程、关键技术及功能。

应用：运用工具（如 Excel）生成数据的可视化图表。

III 关于实践考核实施的说明

一、关于能力层次的说明

在大纲的考核要求中，提出了“领会”、“简单应用”、“综合应用”等三个能力层次，它们的含义是：

领会：掌握课程中的基本知识点，熟悉基本的软件操作。

简单应用：掌握课程中较为深入的知识点，熟练完成较为复杂的软件操作。

综合应用：掌握课程中多个复杂知识点，并能将其结合起来，完成完整的解决方案。

二、自学教材

本实践课程使用教材为：《大数据导论》，林子雨编著，人民邮电出版社，2020年版。

三、关于命题考试的要求

1. 本课程的命题考试，应根据本大纲所规定的考试内容和考试目标来确定范围和考核要求。考试命题适当突出重点，体现本课程综合应用的内容。

2. 本课程考试场所为计算机机房，采用现场考核方式，考生按照考核题目的要求在规定时间内完成。

3. 本课程考核使用的工具：图表工具Excel 2016版，关系数据库工具MySQL 5.7。

4. 考生参加本课程考核可以携带教材。

5. 考核时间为120分钟。

6. 实践考核采用五级评分制评定考核成绩：满分100分，0~59分为不合格，60~69分为合格，70~79分为中等，80~89分为良好，90~100分为优秀。

附录

题型举例

大数据处理综合题

1. 题目：零售销售数据分析

2. 背景：有一家零售公司的分析师，需要分析过去一年的销售数据，以了解哪些产品类别最受欢迎，以及不同季节的销售趋势。

3. 数据集：假设已有一个CSV文件（考试现场提供），包含以下字段：

- 日期（年-月-日）
- 产品ID
- 产品类别
- 销售数量
- 销售金额

4. 处理过程：请按照以下大数据分析过程的步骤，综合利用Excel和MySQL完成。

1) 数据导入：将CSV文件导入MySQL中创建的数据库中。

使用MySQL（脚本和截图）：

2) 数据清洗：包括检查数据是否有缺失值或异常值，并对日期字段进行排序。

使用MySQL（脚本和截图）：

3) 数据分析：分析不同产品类别的销售数量和金额。

使用MySQL（脚本和截图）：

4) 数据可视化：使用EXCEL创建销售趋势图，例如折线图或柱状图；将不同季节的销售数据用不同颜色表示。