

普通专升本免试入学专业综合知识答辩测试大纲

（数据科学与大数据技术专业）

一、测试目标

本专业大类坚持立德树人根本任务，适应区域经济社会发展需求，立足甘肃、面向西北，培养德、智、体、美、劳全面发展的高素质应用型人才，具有良好政治素养和职业道德，知识结构合理，富有创新精神和实践能力，具备数据预处理、数据存储、数据分析等方面的知识。能够适应未来职业和社会发展，能够数据采集、数据预处理、数据存储、数据分析、数据可视化以及大数据工程运维等工作的专门人才。

二、命题的指导思想和原则

（一）命题内容紧密围绕 Python 程序设计、Linux 系统基础、大数据导论、人工智能基础等基础课程考核，重点考核学生的理解、掌握和综合运用的能力。

（二）命题以科学性、明确性、全面性、整体性原则为指导，注重考查学生的基础知识、数据可视化、大数据分析及应用能力。问答题目数量共计 15 个，每问难度一致，问答题目表达简练、明了、准确。

三、测试方式

专业综合知识答辩实行在公布的测试考核知识点范围内随机抽题，每人抽取一题，现场作答。

四、测试时间

每生测试时长为 15 分钟。

五、试卷结构

测试总分为 100 分，每个考核知识点的题目应能够体现语言表达（10 分）、应试态度（10 分）、知识内容（30 分）、专业水平及综合能力（50 分）四部分内容。

六、参考书目

1. 《大数据导论》，刘鹏主编，清华大学出版社，2018年
2. 《Linux 系统应用基础教程》，张小进主编，机械工业出版社，2020年
3. 《Python 程序设计基础》，董付国主编，清华大学出版社，2018年
4. 《人工智能导论》，黄河燕等主编，高等教育出版社，2024年

七、测试的基本要求

1. 本专业测试大纲适用于电子专业大类下数据科学与大数据技术专业的普通专升本免试入学专业综合知识问答测试。
2. 本门测试主要考核学生大数据导论、程序设计与分析的专业综合能力。
3. 考生需按照公布的测试考核知识点内容及数量进行答辩准备。
4. 考生在测试当日进行现场抽签抽取答辩题号，并按照所抽取的题号进行现场作答。
5. 考生的应试过程中不得出现涵盖有标识考生姓名、考号、毕业学校等明显的标示性内容，如果发现按作弊处理。

八、测试考核知识点范围

考核知识点及要求：

本次面试主要考查数据科学与大数据技术专业相关的基础理论、核心技能与实践能力，特别注重在数据工程、人工智能应用开发及大数据系统操作等方面的综合素养。具体考核知识点及要求如下：

考核知识点 1：Linux 与 Shell 编程

编写一个 Shell 脚本程序“compare.sh”，要求随机生成一个小于 100 的整数，用户从终端输入猜测数字，程序将持续判断并提示“大”或“小”，直至猜中为止。

考核知识点 2：Linux 系统用户与文件管理

运用 Shell 命令完成以下操作：

- 1) 创建名为‘team’的用户组；

2) 创建用户 ‘member’，将其加入 ‘team’ 组；

3) 在 ‘member’ 主目录下创建 ‘project’ 目录，设置读写执行权限并进入；

4) 在 ‘project’ 中创建 ‘doc1.txt’ 与 ‘doc2.txt’，分别设置读写执行与只读权限，并确保所有者与组为 ‘member’ 与 ‘team’；

5) 在 ‘project’ 下创建子目录 ‘sub_project’，设置权限并进入，同样确保其所有者与组正确。

考核知识点 3: Linux 系统运行级别管理

简述 Linux 系统的运行级别及其含义，并演示使用 systemd 命令进行运行级别的切换。

考核知识点 4: Python 基础编程

编写 Python 函数，从终端输入两个正整数，计算并返回其最大公因数。

考核知识点 5: Python 文本处理与数据分析基础

编写 Python 程序，接收一个文件路径列表，统计多个文本文件中各单词的出现频率，要求使用异常处理机制打开文件，并利用正则表达式或字符串方法进行词频统计，返回词频字典。

考核知识点 6: Python 数据管理实践

实现一个简易学生信息管理系统，基于字典结构存储学号、姓名和成绩，支持添加、查询、展示全部信息、计算平均成绩等功能，并通过菜单循环实现交互操作。

考核知识点 7: Python 数据结构操作

针对给定数字列表，完成以下操作：计算元素和、找出最大值与最小值、转换为元组、生成元素与对应平方值的字典，并输出结果。

考核知识点 8: Python 数据分析工具 Pandas 应用

概述 Pandas 库在数据处理与分析中的主要功能，并举例说明其在实际数据科学项目中的典型应用。

考核知识点 9: 请叙述大数据、云计算和物联网三者之间的关系。

考核知识点 10: 大数据与人工智能概述

阐述大数据、人工智能与机器学习三者之间的内在联系与区别。

考核知识点 11: 人工智能与深度学习概念

简述人工智能、机器学习、深度学习的基本定义及其层次关系。

考核知识点 12: 主流大数据技术架构

列举常见大数据处理框架,并分析其各自特点、适用场景及优缺点。

考核知识点 13: 深度学习框架比较

说明当前主流的深度学习框架,并简述其应用领域及优势。

考核知识点 14: Hadoop 系统基础

解释 Hadoop 的基本概念及其在大数据生态中的核心作用。

考核知识点 15: 人工智能模型训练流程理解

请简述一个典型的监督式机器学习项目,如分类任务的主要流程,并说明在数据准备、特征工程、模型训练、评估与调优等阶段常用的工具或技术,可结合 Scikit-learn、TensorFlow/PyTorch 等框架举例。