



# 湖南网络工程职业学院

---

## 大数据技术专业 专业技能考核题库

2021 级

2021 年 8 月

# 目录

|                                           |    |
|-------------------------------------------|----|
| 模块一 数据管理技术 .....                          | 1  |
| 项目 1: 建筑用地管理数据库设计 .....                   | 1  |
| 1.试题编号: J1-1-1: “建设用地行政管理模块”数据库设计 .....   | 1  |
| (1) 任务描述 .....                            | 1  |
| (2) 实施条件 .....                            | 4  |
| (3) 考核时量及要求 .....                         | 4  |
| (4) 评分细则 .....                            | 4  |
| 2.试题编号: J1-1-2: “建设用地地块管理模块”数据库设计 .....   | 5  |
| (1) 任务描述 .....                            | 5  |
| (2) 实施条件 .....                            | 8  |
| (3) 考核时量及要求 .....                         | 8  |
| (4) 评分细则 .....                            | 8  |
| 项目 2: 学生系统数据库设计 .....                     | 9  |
| 3.试题编号: J1-2-1: “学生成绩管理模块”数据库设计 .....     | 9  |
| (1) 任务描述 .....                            | 9  |
| (2) 实施条件 .....                            | 12 |
| (3) 考核时量及要求 .....                         | 12 |
| (4) 评分细则 .....                            | 12 |
| 4.试题编号: J1-2-2: “学生选课管理模块”数据库设计 .....     | 13 |
| (1) 任务描述 .....                            | 13 |
| (2) 实施条件 .....                            | 16 |
| (3) 考核时量及要求 .....                         | 17 |
| (4) 评分细则 .....                            | 17 |
| 项目 3: 学生食堂饭卡管理数据库设计 .....                 | 18 |
| 5.试题编号: J1-3-1: “学生食堂饭卡消费管理模块”数据库设计 ..... | 18 |
| (1) 任务描述 .....                            | 18 |

|                                            |    |
|--------------------------------------------|----|
| (2) 实施条件 .....                             | 21 |
| (3) 考核时量及要求 .....                          | 21 |
| (4) 评分细则 .....                             | 21 |
| 6. 试题编号: J1-3-2: “学生食堂饭卡充值管理模块”数据库设计 ..... | 22 |
| (1) 任务描述 .....                             | 22 |
| (2) 实施条件 .....                             | 26 |
| (3) 考核时量及要求 .....                          | 26 |
| (4) 评分细则 .....                             | 26 |
| 项目 4: 超市系统数据库设计 .....                      | 27 |
| 7. 试题编号: J1-4-1: “大型超市商品管理模块”数据库设计 .....   | 27 |
| (1) 任务描述 .....                             | 27 |
| (2) 实施条件 .....                             | 30 |
| (3) 考核时量及要求 .....                          | 30 |
| (4) 评分细则 .....                             | 30 |
| 8. 试题编号: J1-4-2: “大型超市库存管理模块”数据库设计 .....   | 31 |
| (1) 任务描述 .....                             | 31 |
| (2) 实施条件 .....                             | 34 |
| (3) 考核时量及要求 .....                          | 34 |
| (4) 评分细则 .....                             | 34 |
| 项目 5: 考试机构系统数据库设计 .....                    | 35 |
| 9. 试题编号: J1-5-1: “考试机构报名管理模块”数据库设计 .....   | 35 |
| (1) 任务描述 .....                             | 35 |
| (2) 实施条件 .....                             | 38 |
| (3) 考核时量及要求 .....                          | 38 |
| (4) 评分细则 .....                             | 38 |
| 10. 试题编号: J1-5-2: “教材管理模块”数据库设计 .....      | 39 |
| (1) 任务描述 .....                             | 39 |
| (2) 实施条件 .....                             | 42 |

|                                                 |    |
|-------------------------------------------------|----|
| (3) 考核时量及要求 .....                               | 42 |
| (4) 评分细则 .....                                  | 42 |
| 项目 6: 人力资源系统数据库设计 .....                         | 43 |
| 11. 试题编号: J1-6-1: “人力资源系统项目管理模块”数据库设计 .....     | 43 |
| (1) 任务描述 .....                                  | 43 |
| (2) 实施条件 .....                                  | 46 |
| (3) 考核时量及要求 .....                               | 47 |
| (4) 评分细则 .....                                  | 47 |
| 12. 试题编号: J1-6-2: “人力资源系统员工信息管理模块”数据库设计 .....   | 48 |
| (1) 任务描述 .....                                  | 48 |
| (2) 实施条件 .....                                  | 50 |
| (3) 考核时量及要求 .....                               | 51 |
| (4) 评分细则 .....                                  | 51 |
| 项目 7: 图书管理系统数据库设计 .....                         | 52 |
| 13. 试题编号: J1-6-3: “人力资源系统员工工资信息管理模块”数据库设计 ..... | 52 |
| (1) 任务描述 .....                                  | 52 |
| (2) 实施条件 .....                                  | 55 |
| (3) 考核时量及要求 .....                               | 55 |
| (4) 评分细则 .....                                  | 55 |
| 项目 7: 图书管理系统数据库设计 .....                         | 56 |
| 14. 试题编号: J1-7-1: “图书管理系统借阅管理模块”数据库设计 .....     | 56 |
| (1) 任务描述 .....                                  | 56 |
| (2) 实施条件 .....                                  | 59 |
| (3) 考核时量及要求 .....                               | 59 |
| (4) 评分细则 .....                                  | 59 |
| 15. 试题编号: J1-7-2: “图书订购系统管理模块”数据库设计 .....       | 60 |
| (1) 任务描述 .....                                  | 60 |
| (2) 实施条件 .....                                  | 63 |

|                                                   |    |
|---------------------------------------------------|----|
| (3) 考核时量及要求 .....                                 | 64 |
| (4) 评分细则 .....                                    | 64 |
| 模块二 大数据平台部署 .....                                 | 65 |
| 项目 1: Hadoop 伪分布式模式部署 .....                       | 65 |
| 1.试题编号: J2-1-1: 美团大数据平台 Hadoop 伪分布式模式部署 .....     | 65 |
| (1) 任务描述 .....                                    | 65 |
| (2) 实施条件 .....                                    | 68 |
| (3) 考核时量及要求 .....                                 | 68 |
| (4) 评分细则 .....                                    | 68 |
| 2.试题编号: J2-1-2: 去哪儿旅游网大数据平台 Hadoop 伪分布式模式部署 ..... | 71 |
| (1) 任务描述 .....                                    | 71 |
| (2) 实施条件 .....                                    | 73 |
| (3) 考核时量及要求 .....                                 | 73 |
| (4) 评分细则 .....                                    | 74 |
| 3.试题编号: J2-1-3: 赶集网大数据平台 Hadoop 伪分布式模式部署 .....    | 76 |
| (1) 任务描述 .....                                    | 76 |
| (2) 实施条件 .....                                    | 79 |
| (3) 考核时量及要求 .....                                 | 79 |
| (4) 评分细则 .....                                    | 79 |
| 4.试题编号: J2-1-4: 大麦网大数据平台 Hadoop 伪分布式模式部署 .....    | 82 |
| (1) 任务描述 .....                                    | 82 |
| (2) 实施条件 .....                                    | 85 |
| (3) 考核时量及要求 .....                                 | 85 |
| (4) 评分细则 .....                                    | 85 |
| 项目 2: Hadoop 完全分布式模式部署 .....                      | 88 |
| 5.试题编号: J2-2-1: 豆瓣网 Hadoop 完全分布式模式部署 .....        | 88 |
| (1) 任务描述 .....                                    | 88 |
| (2) 实施条件 .....                                    | 91 |

|                                                |     |
|------------------------------------------------|-----|
| (3) 考核时量及要求 .....                              | 91  |
| (4) 评分细则 .....                                 | 91  |
| 6.试题编号: J2-2-2: 起点网 Hadoop 完全分布式模式部署 .....     | 94  |
| (1) 任务描述 .....                                 | 94  |
| (2) 实施条件 .....                                 | 96  |
| (3) 考核时量及要求 .....                              | 97  |
| (4) 评分细则 .....                                 | 97  |
| 7.试题编号: J2-2-3: 夸克小说网 Hadoop 完全分布式模式部署 .....   | 99  |
| (1) 任务描述 .....                                 | 99  |
| (2) 实施条件 .....                                 | 101 |
| (3) 考核时量及要求 .....                              | 102 |
| (4) 评分细则 .....                                 | 102 |
| 8.试题编号: J2-2-4: 中关村在线平台 Hadoop 完全分布式模式部署 ..... | 104 |
| (1) 任务描述 .....                                 | 104 |
| (2) 实施条件 .....                                 | 107 |
| (3) 考核时量及要求 .....                              | 107 |
| (4) 评分细则 .....                                 | 107 |
| 9.试题编号: J2-2-5: 京东 Hadoop 完全分布式模式部署 .....      | 110 |
| (1) 任务描述 .....                                 | 110 |
| (2) 实施条件 .....                                 | 113 |
| (3) 考核时量及要求 .....                              | 113 |
| (4) 评分细则 .....                                 | 113 |
| 10.试题编号: J2-2-6: 西瓜视频 Hadoop 完全分布式模式部署 .....   | 115 |
| (1) 任务描述 .....                                 | 115 |
| (2) 实施条件 .....                                 | 118 |
| (3) 考核时量及要求 .....                              | 118 |
| (4) 评分细则 .....                                 | 118 |
| 模块三 Python 数据采集技术 .....                        | 121 |

|                                     |     |
|-------------------------------------|-----|
| 项目 1：数字内容服务行业数据采集与可视化 .....         | 121 |
| 1.试题编号：J3-1-1：时光网电影票房数据采集与可视化 ..... | 121 |
| (1) 任务描述 .....                      | 121 |
| (2) 实施条件 .....                      | 122 |
| (3) 考核时量及要求 .....                   | 123 |
| (4) 评分细则 .....                      | 123 |
| 2.试题编号：J3-1-2：豆瓣小说数据采集与可视化 .....    | 124 |
| (1) 任务描述 .....                      | 124 |
| (2) 实施条件 .....                      | 126 |
| (3) 考核时量及要求 .....                   | 126 |
| (4) 评分细则 .....                      | 126 |
| 3.试题编号：J3-1-3：猫眼电影数据采集与可视化 .....    | 128 |
| (1) 任务描述 .....                      | 128 |
| (2) 实施条件 .....                      | 130 |
| (3) 考核时量及要求 .....                   | 130 |
| (4) 评分细则 .....                      | 130 |
| 4.试题编号：J3-1-4：淘书网图书信息的采集与可视化 .....  | 131 |
| (1) 任务描述 .....                      | 131 |
| (2) 实施条件 .....                      | 133 |
| (3) 考核时量及要求 .....                   | 134 |
| (4) 评分细则 .....                      | 134 |
| 5.试题编号：J3-1-5：当当图书好评数据采集与可视化 .....  | 135 |
| (1) 任务描述 .....                      | 135 |
| (2) 实施条件 .....                      | 138 |
| (3) 考核时量及要求 .....                   | 138 |
| (4) 评分细则 .....                      | 138 |
| 项目 2：社会服务行业数据采集与可视化 .....           | 140 |
| 6.试题编号：J3-2-1：美食杰数据采集与可视化 .....     | 140 |

|                                             |     |
|---------------------------------------------|-----|
| (1) 任务描述 .....                              | 140 |
| (2) 实施条件 .....                              | 142 |
| (3) 考核时量及要求 .....                           | 142 |
| (4) 评分细则 .....                              | 142 |
| 7.试题编号: J3-2-2: 湖南人才网站数据采集与可视化 .....        | 144 |
| (1) 任务描述 .....                              | 144 |
| (2) 实施条件 .....                              | 146 |
| (3) 考核时量及要求 .....                           | 146 |
| (4) 评分细则 .....                              | 146 |
| 8.试题编号: J3-2-3: 链家长沙楼盘数据采集与可视化 .....        | 148 |
| (1) 任务描述 .....                              | 148 |
| (2) 实施条件 .....                              | 149 |
| (3) 考核时量及要求 .....                           | 150 |
| (4) 评分细则 .....                              | 150 |
| 9.试题编号: J3-2-4: 土流网数据采集与可视化 .....           | 151 |
| (1) 任务描述 .....                              | 151 |
| (2) 实施条件 .....                              | 153 |
| (3) 考核时量及要求 .....                           | 153 |
| (4) 评分细则 .....                              | 154 |
| 项目 3: 资讯服务行业数据采集与可视化 .....                  | 155 |
| 10.试题编号: J3-3-1: 中国大学 Top100 数据采集与可视化 ..... | 155 |
| (1) 任务描述 .....                              | 155 |
| (2) 实施条件 .....                              | 157 |
| (3) 考核时量及要求 .....                           | 157 |
| (4) 评分细则 .....                              | 157 |
| 11.试题编号: J3-3-2: 北京 2020 年天气数据采集与可视化 .....  | 159 |
| (1) 任务描述 .....                              | 159 |
| (2) 实施条件 .....                              | 161 |

|                                        |     |
|----------------------------------------|-----|
| (3) 考核时量及要求 .....                      | 161 |
| (4) 评分细则 .....                         | 161 |
| 12. 试题编号: J3-3-3: 糗事百科数据采集与可视化 .....   | 162 |
| (1) 任务描述 .....                         | 162 |
| (2) 实施条件 .....                         | 164 |
| (3) 考核时量及要求 .....                      | 164 |
| (4) 评分细则 .....                         | 165 |
| 13. 试题编号: J3-3-4: 胡润富豪榜数据采集与可视化 .....  | 166 |
| (1) 任务描述 .....                         | 166 |
| (2) 实施条件 .....                         | 168 |
| (3) 考核时量及要求 .....                      | 168 |
| (4) 评分细则 .....                         | 168 |
| 模块四 数据处理与可视化 .....                     | 170 |
| 项目 1: 数字内容服务行业数据处理与可视化 .....           | 170 |
| 1. 试题编号: J4-1-1: 时光网电影票房数据处理与可视化 ..... | 170 |
| (1) 任务描述 .....                         | 170 |
| (2) 实施条件 .....                         | 171 |
| (3) 考核时量及要求 .....                      | 171 |
| (4) 评分细则 .....                         | 171 |
| 2. 试题编号: J4-1-2: 豆瓣小说数据处理与可视化 .....    | 173 |
| (1) 任务描述 .....                         | 173 |
| (2) 实施条件 .....                         | 174 |
| (3) 考核时量及要求 .....                      | 174 |
| (4) 评分细则 .....                         | 174 |
| 3. 试题编号: J4-1-3: 猫眼电影数据处理与可视化 .....    | 176 |
| (1) 任务描述 .....                         | 176 |
| (2) 实施条件 .....                         | 177 |
| (3) 考核时量及要求 .....                      | 177 |

|                                      |     |
|--------------------------------------|-----|
| (4) 评分细则 .....                       | 177 |
| 4.试题编号: J4-1-4: 淘书网图书信息的采集与可视化 ..... | 179 |
| (1) 任务描述 .....                       | 179 |
| (2) 实施条件 .....                       | 180 |
| (3) 考核时量及要求 .....                    | 180 |
| (4) 评分细则 .....                       | 180 |
| 5.试题编号: J4-1-5: 小说排行榜数据处理与可视化 .....  | 182 |
| (1) 任务描述 .....                       | 182 |
| (2) 实施条件 .....                       | 183 |
| (3) 考核时量及要求 .....                    | 183 |
| (4) 评分细则 .....                       | 183 |
| 6.试题编号: J4-1-6: 当当图书好评数据处理与可视化 ..... | 185 |
| (1) 任务描述 .....                       | 185 |
| (2) 实施条件 .....                       | 186 |
| (3) 考核时量及要求 .....                    | 186 |
| (4) 评分细则 .....                       | 186 |
| 项目 2: 社会服务行业数据处理与可视化 .....           | 187 |
| 7.试题编号: J4-2-1: 链家长沙小区数据处理与可视化 ..... | 187 |
| (1) 任务描述 .....                       | 187 |
| (2) 实施条件 .....                       | 188 |
| (3) 考核时量及要求 .....                    | 189 |
| (4) 评分细则 .....                       | 189 |
| 8.试题编号: J4-2-2: 美食杰数据处理与可视化 .....    | 190 |
| (1) 任务描述 .....                       | 190 |
| (2) 实施条件 .....                       | 191 |
| (3) 考核时量及要求 .....                    | 191 |
| (4) 评分细则 .....                       | 192 |
| 9.试题编号: J4-2-3: 湖南人才网站数据处理与可视化 ..... | 193 |

|                                             |     |
|---------------------------------------------|-----|
| (1) 任务描述 .....                              | 193 |
| (2) 实施条件 .....                              | 195 |
| (3) 考核时量及要求 .....                           | 195 |
| (4) 评分细则 .....                              | 195 |
| 10. 试题编号: J4-2-4: 酒店数据处理与可视化 .....          | 197 |
| (1) 任务描述 .....                              | 197 |
| (2) 实施条件 .....                              | 199 |
| (3) 考核时量及要求 .....                           | 199 |
| (4) 评分细则 .....                              | 199 |
| 11. 试题编号: J4-2-5: 土流网数据处理与可视化 .....         | 201 |
| (1) 任务描述 .....                              | 201 |
| (2) 实施条件 .....                              | 203 |
| (3) 考核时量及要求 .....                           | 203 |
| (4) 评分细则 .....                              | 203 |
| 项目 3: 资讯服务行业数据处理与可视化 .....                  | 205 |
| 12. 试题编号: J4-3-1: 北京 2020 年天气数据处理与可视化 ..... | 205 |
| (1) 任务描述 .....                              | 205 |
| (2) 实施条件 .....                              | 206 |
| (3) 考核时量及要求 .....                           | 206 |
| (4) 评分细则 .....                              | 207 |
| 13. 试题编号: J4-3-2: 糗事百科数据处理与可视化 .....        | 208 |
| (1) 任务描述 .....                              | 208 |
| (2) 实施条件 .....                              | 209 |
| (3) 考核时量及要求 .....                           | 209 |
| (4) 评分细则 .....                              | 210 |
| 14. 试题编号: J4-3-3: 胡润富豪榜数据处理与可视化 .....       | 211 |
| (1) 任务描述 .....                              | 211 |
| (2) 实施条件 .....                              | 212 |

|                   |     |
|-------------------|-----|
| (3) 考核时量及要求 ..... | 212 |
| (4) 评分细则 .....    | 213 |

模块一 数据管理技术

项目 1：建筑用地管理数据库设计

1.试题编号：J1-1-1：“建设用地行政管理模块”数据库设计

(1) 任务描述

建设用地管理是国土资源管理的重要组成部分，是国土资源管理的核心，建设用地项目的审批将直接影响到城市的发展和规划。业务管理人员不仅需要及时准确的了解建设用地项目的基本情况，还要对土地的规划、权属等多种数据进行比对分析，以便正确划分土地用途,避免用地范围内可能存在的问题。运用 **MYSQL** 实现，监督管理、批后监管、统计分析等功能，提高业务管理人员的工作效率。

《行政管理》模块的 E-R 图如图 1-1 所示，物理数据模型如图 1-2 所示，数据表字段名定义见表 1-1。请按以下设计完成数据库创建、数据表创建和数据操作任务：

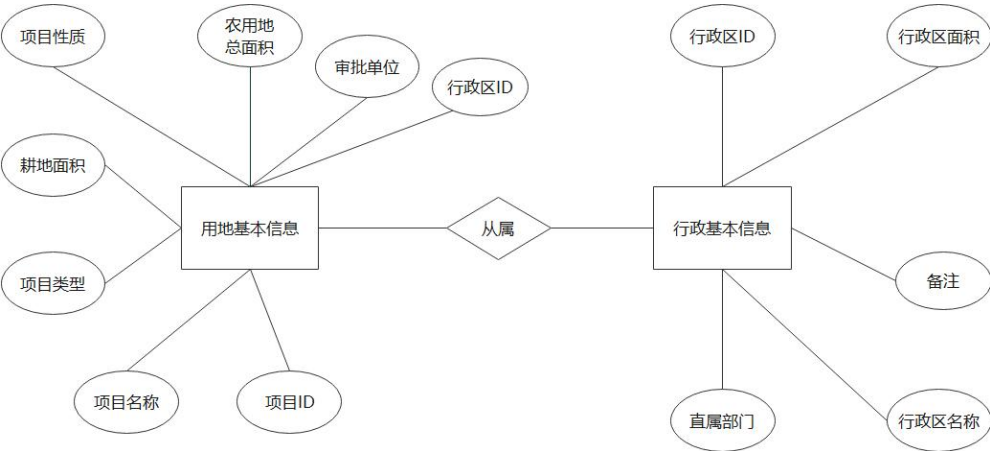


图 1-1 E-R 图

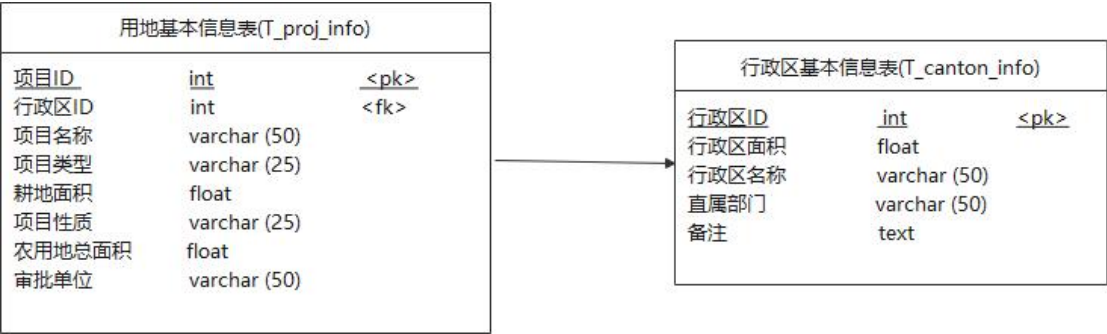


图 1-2 物理数据模型

表1-1字段名定义表

| 字段名         | 字段说明   | 字段名          | 字段说明   |
|-------------|--------|--------------|--------|
| proj_id     | 项目 ID  | farm_tot     | 农用地总面积 |
| canton_id   | 行政区 ID | approve_unit | 审批单位   |
| proj_name   | 项目名称   | canton_name  | 行政区名称  |
| proj_type   | 项目类型   | canton_tot   | 行政区面积  |
| proj_kind   | 项目性质   | branch       | 直属部门   |
| tilth_state | 耕地面积   | remark       | 备注     |

**任务一：创建数据库（10 分）**

创建数据库 ProjDB。新建 T1-1.sql 文件，在该文件中编写语句创建指定的数据库，使用语句显示已经创建的数据库，并使用语句使用 ProjDB 数据库。将上述操作截图，命名为 T1-1.jpg 文件，将 T1-1.sql、T1-1.jpg 保存到考生目录 T1 下。

**任务二：创建数据表及表间的关系及约束（30 分）**

创建 T2-1.sql 文件并完成以下功能：

- 根据图 1-1、1-2 和表 1-1，创建数据表 T\_canton\_info、T\_proj\_info。（15 分）
- 根据物理数据原型，创建数据表之间的关系。（15 分）

将创建表 T\_canton\_info、T\_proj\_info 的语句以及执行结果截图，命名分别为 T2-1.jpg、T2-2.jpg，将 T2-1.sql、T2-1.jpg、T2-2.jpg 三个文件保存到考生目录 T2 下。

**任务三：数据操作（50 分）**

考生目录 T3 下创建 T3-1.sql 文件并用 SQL 语句执行以下操作：

- 在每个表中插入以下测试数据；

向 T\_canton\_info 表插入数据：

| 行政区ID | 行政区面积    | 行政区名称  | 直属部门     | 备注   |
|-------|----------|--------|----------|------|
| 20001 | 23400.35 | 长沙市雨花区 | 长沙市国土资源局 | 省级直属 |
| 20002 | 15010.45 | 东屯渡街道  | 长沙市国土资源局 | 无    |
| 20003 | 40230.35 | 长沙市开福区 | 湖南省政府    | 省级直属 |
| 20004 | 55000.45 | 长沙市芙蓉区 | 湖南省政府    | 无    |
| 20005 | 805      | 长沙市行政区 | 长沙市国土资源局 | 无    |
| 20006 | 562      | 湖南省行政区 | 湖南省计划生育局 | 无    |

向 T\_proj\_info 表插入数据：

| 项目ID  | 行政区ID | 项目名称  | 项目类型 | 耕地面积  | 项目性质 | 农用地总面积 | 审批单位  |
|-------|-------|-------|------|-------|------|--------|-------|
| 10001 | 20001 | 西气东输  | 基础建设 | 710.5 | 国家   | 1002.2 | 政府    |
| 10002 | 20002 | 西部大开发 | 基础建设 | 615.4 | 国家   | 912.6  | 政府    |
| 10003 | 20003 | 广赛大厦  | 房地产  | 800.5 | 私用   | 1000   | 湖南省政府 |
| 10004 | 20004 | 职工医院  | 房地产  | 500.5 | 公用   | 700    | 湖南省政府 |

|       |       |      |      |       |    |      |       |
|-------|-------|------|------|-------|----|------|-------|
| 10005 | 20005 | 眼科中心 | 房地产  | 800.2 | 私用 | 1000 | 湖南省政府 |
| 10006 | 20006 | 公益学校 | 基础建设 | 5600  | 公用 | 700  | 湖南省政府 |

查询 T\_canton\_info、T\_proj\_info 两个表中的数据并截图，命名分别为 T3-1.jpg、T3-2.jpg，保存到考生目录 T3 下。（5 分）

- 查询出项目 ID 为“10002”的建设用地基本信息，包括：项目名称、项目类型、耕地面积、项目性质，并将结果截图，命名为 T3-3.jpg，保存到考生目录 T3 下。（5 分）
- 查询出行政直属部门为“长沙市国土资源局”的行政区名称、行政区面积，并将结果截图，命名为 T3-4.jpg，保存到考生目录 T3 下。（5 分）
- 查询出用地基本信息表中数据并按农用地总面积升序排序，并将结果截图，命名为 T3-5.jpg，保存到考生目录 T3 下。（5 分）
- 查询出用地基本信息表中数据并按耕地面积降序排序，并将结果截图，命名为 T3-6.jpg，保存到考生目录 T3 下。（5 分）
- 删除用地基本信息表中耕地面积大于“720”的数据，删除数据后，使用查询语句显示删除后的结果并截图，命名为 T3-7.jpg，保存到考生目录 T3 下。（5 分）
- 请把直属部门由“长沙市国土资源局”修改为“永州市国土资源局”，修改数据后，使用查询语句显示修改后的结果并截图，命名为 T3-8.jpg，保存到考生目录 T3 下。（5 分）
- 在行政区域信息表中插入数据，行政区 ID 为“83100”，行政名称为“星沙”，行政面积为“142”，直属部门为“长沙市国土资源局”，数据插入后，使用查询语句显示插入后的结果并截图，命名为 T3-9.jpg，保存到考生目录 T3 下。（5 分）
- 创建名为 ProjInfo\_view1 的视图，视图的数据为项目编号是“10001”的用地基本信息，使用语句显示视图 ProjInfo\_view1 的数据并截图，命名为 T3-10.jpg，保存到考生目录 T3 下。（5 分）
- 删除上题中所创建的 ProjInfo\_view1 视图，使用 show tables 语句查看所有数据库表，验证视图 ProjInfo\_view1 已经删除并将结果截图，命名为 T3-11.jpg，保存到考生目录 T3 下。（5 分）

良好的习惯和职业素养（10 分）

- 答题过程中保持安静，有问题举手示意，有礼貌。（3分）
- 数据库名命名规范，数据表名命名规范，字段名命名规范。（3分）
- 答题完毕离场时整理好工位。（4分）

## （2）实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。

计算机安装了 Windows 操作系统，MySQL 数据库软件以及 Office 办公软件。

## （3）考核时量及要求

- 考核时长 3 小时；
- 创建答题文件夹创建以“学号\_题号”命名的文件夹，存放所有答题文件；
- 创建答题文件，包括：SQL 脚本文件，如 J3-1.sql

## （4）评分细则

表1-2 考核评价细则表

| 评价项            | 分值   | 评分细则                                                   |
|----------------|------|--------------------------------------------------------|
| 数据库创建          | 10 分 | 创建考生目录 T1，文件 T1-1.sql、T1-1.jpg，得 1.5 分                 |
|                |      | 使用语句创建了数据库，得 3 分                                       |
|                |      | 使用语句显示已经创建的数据库的名字，得 3 分                                |
|                |      | 使用新创建的数据库，得 2.5 分                                      |
| 创建数据表及表间的关系及约束 | 30 分 | 创建考生目录 T2，文件 T2-1.sql、T2-1.jpg、T2-2.jpg，得 2 分          |
|                |      | 表 T_canton_info 创建的字段名、字段类型、主键正确，并正确显示表结构及主键约束，得 12 分  |
|                |      | 表 T_proj_info 创建的字段名、字段类型、主键正确，并正确显示表结构及主键和外键约束，得 16 分 |
| 数据操作           | 50 分 | 表 T_canton_info 和 T_proj_info 正确插入数据，得 5 分             |
|                |      | 查询出项目 ID 为“10002”的建设用地基本信息，得 5 分                       |
|                |      | 查询出行政直属部门为“长沙市国土资源局”的行政区名称、行政区面积，得 5 分                 |
|                |      | 查询出用地基本信息表中数据并按农用地总面积升序排序，得 5 分                        |
|                |      | 查询出用地基本信息表中数据并按耕地面积降序排序，得 5 分                          |
|                |      | 删除用地基本信息表中耕地面积大于“720”，得 5 分                            |
|                |      | 直属部门由“长沙市国土资源局”修改为“永州市国土资源局”，得 5 分                     |
|                |      | 正确插入数据（行政区 ID 为“83100”、行政名称为“星                         |

|            |                         |     |                                       |
|------------|-------------------------|-----|---------------------------------------|
|            |                         |     | 沙”、行政面积为“142”、直属部门为“长沙市国土资源局”), 得 5 分 |
|            |                         |     | 创建的 ProjInfo_view1 视图正确, 得 5 分        |
|            |                         |     | 删除视图成功, 得 5 分                         |
| 良好的习惯和职业素养 | 代码书写格式规范                | 4 分 | 脚本缩进、换行规范, 得 4 分                      |
|            | 答题过程中保持安静, 有问题举手示意, 有礼貌 | 3 分 | 遵守考场纪律、有礼貌, 得 3 分                     |
|            | 答题完毕离场时整理好工位            | 3 分 | 鼠标、键盘、凳子等摆放整齐, 得 3 分                  |

## 2.试题编号: J1-1-2: “建设用地地块管理模块”数据库设计

### (1) 任务描述

近年来,在国家大力推进城镇化和工业化进程的大背景下,到处是在建的高楼和厂房,城市用地存量严重不足,甚至出现超额使用用地指标的现象。土地资源在我国日趋紧张已成了不争的事实。长沙市政府在用地紧张的危机意识,想了解土地的具体使用情况。《地块管理》模块的 E-R 图如图 2-1 所示,物理数据模型如图 2-2 所示,数据表字段名定义见表 2-1。请按以下设计完成数据库创建、数据表创建和数据操作任务:

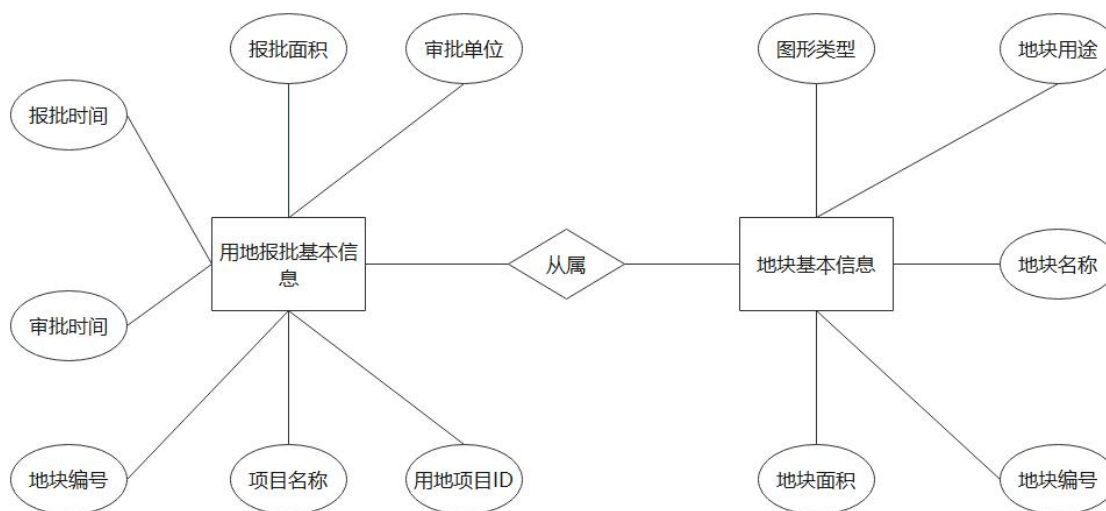


图 2-1 E-R 图



图2-2物理数据模型

表2-1字段名定义表

| 字段名        | 字段说明    | 字段名        | 字段说明 |
|------------|---------|------------|------|
| proj_id    | 用地项目 ID | appr_date  | 审批时间 |
| proj_name  | 项目名称    | plot_id    | 地块编号 |
| appr_area  | 报批面积    | plot_name  | 地块名称 |
| appr_date  | 报批时间    | total_area | 地块面积 |
| appr_unit  | 审批单位    | purpose    | 项目用途 |
| shape_type | 图形类型    |            |      |

### 任务一：创建数据库（10 分）

创建数据库 ContractionDB。新建 T1-1.sql 文件，在该文件中编写语句创建指定的数据库，使用语句显示已经创建的数据库，并使用语句使用 ContractionDB 数据库。将上述操作截图，命名为 T1-1.jpg 文件，将 T1-1.sql、T1-1.jpg 保存到考生目录 T1 下。

### 任务二：创建数据表（30 分）

创建 T2-1.sql 文件并完成以下功能：

- 根据图 2-1、2-2 和表 2-1，创建数据表 T\_appr\_area、T\_plot\_info。（15 分）
- 根据物理数据原型，创建数据表之间的关系。（15 分）

将创建表 T\_appr\_area、T\_plot\_info 的语句以及执行结果截图，命名分别为 T2-1.jpg、T2-2.jpg，将 T2-1.sql、T2-1.jpg、T2-2.jpg 三个文件保存到考生目录 T2 下。

### 任务三：数据操作（50 分）

考生目录 T3 下创建 T3-1.sql 文件并用 SQL 语句执行以下操作：

- 在每个表中插入 2 条测试数据：

向 T\_appr\_area 表插入数据：

| 用地项目 ID | 项目名称     | 审批时间       | 报批时间       | 报批面积   | 审批单位  |
|---------|----------|------------|------------|--------|-------|
| 101     | 王府花园小区项目 | 2018-06-18 | 2018-06-30 | 1086.2 | 长沙工商局 |

|     |           |            |            |        |         |
|-----|-----------|------------|------------|--------|---------|
| 102 | 天虹超市项目    | 2018-06-14 | 2018-07-14 | 1325.0 | 湖南国土资源局 |
| 103 | 枫树山小学食堂项目 | 2017-06-18 | 2017-06-30 | 86.2   | 长沙教育局   |
| 104 | 海德公园      | 2016-07-14 | 2016-07-14 | 325.0  | 长沙市规划局  |

向 T\_plot\_info 表插入数据:

| 地块编号  | 用地项目 ID | 地块面积   | 地块名称    | 项目用途 | 图形类型 |
|-------|---------|--------|---------|------|------|
| 10001 | 101     | 900    | 长沙市王府花园 | 商业用地 | 菱形   |
| 10002 | 102     | 925.7  | 长沙市天虹超市 | 商业用地 | 正方形  |
| 10003 | 103     | 1725.4 | 枫树山小学   | 商业用地 | 菱形   |
| 10004 | 104     | 925.7  | 海德公园    | 商业用地 | 正方形  |

查询 T\_appr\_area、T\_plot\_info 两个表中的数据并截图，命名分别为 T3-1.jpg、T3-2.jpg，保存到考生目录 T3 下。（5 分）

- 查询出报批面积大于“1000”的建设用地的报批基本信息，并将结果截图，命名为 T3-3.jpg，保存到考生目录 T3 下。（5 分）
- 查询出报批时间为 2018 年 6 月份的所有建设用地的报批基本信息，并将结果截图，命名为 T3-4.jpg，保存到考生目录 T3 下。（5 分）
- （4）查询出地块名称为“长沙市王府花园”的用地的报批的项目名称和审批时间，并将结果截图，命名为 T3-5.jpg，保存到考生目录 T3 下。（5 分）
- 查询出所有的地块基本信息，要求按地块面积升序排序，并将结果截图，命名为 T3-6.jpg，保存到考生目录 T3 下。（5 分）
- 删除地块面积大于“1100”的地块基本信息，删除数据后，使用查询语句显示删除后的结果并截图，命名为 T3-7.jpg，保存到考生目录 T3 下。（5 分）
- 把地块名称为“长沙市王府花园”修改为“永州市王府花园”，修改数据后，使用查询语句显示修改后的结果并截图，命名为 T3-8.jpg，保存到考生目录 T3 下。（10 分）
- 创建名为 Plot\_id\_view1 的视图，视图的数据为地块编号是“10002”的地块基本信息，使用 select 语句显示视图 Plot\_id\_view1 的数据并截图，命名为 T3-9.jpg，保存到考生目录 T3 下。（5 分）
- 删除上题中所创建的 Plot\_id\_view1 视图，使用 show tables 语句查看所有数据库表，验证视图 Plot\_id\_view1 已经删除并将结果截图，命名为 T3-10.jpg，保存到考生目录 T3 下。（5 分）

### 良好的习惯和职业素养（10分）

- 答题过程中保持安静，有问题举手示意，有礼貌。（3分）
- 数据库名命名规范，数据表名命名规范，字段名命名规范。（3分）
- 答题完毕离场时整理好工位。（4分）

### （2）实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。

计算机安装了 Windows 操作系统，MySQL 数据库软件以及 Office 办公软件。

### （3）考核时量及要求

- 考核时长 3 小时；
- 创建答题文件夹创建以“学号\_题号”命名的文件夹，存放所有答题文件；
- 创建答题文件，包括：SQL 脚本文件，如 J3-1.sql

### （4）评分细则

表2-2 考核评价细则表

| 评价项            | 分值   | 评分细则                                                   |
|----------------|------|--------------------------------------------------------|
| 数据库创建          | 10 分 | 创建考生目录 T1，文件 T1-1.sql、T1-1.jpg，得 1.5 分                 |
|                |      | 使用语句创建了数据库，得 3 分                                       |
|                |      | 使用语句显示已经创建的数据库的名字，得 3 分                                |
|                |      | 使用新创建的数据库，得 2.5 分                                      |
| 创建数据表及表间的关系及约束 | 30 分 | 创建考生目录 T2，文件 T2-1.sql、T2-1.jpg、T2-2.jpg，得 2 分          |
|                |      | 表 T_appr_area 创建的字段名、字段类型、主键正确，并正确显示表结构及主键约束，得 12 分    |
|                |      | 表 T_proj_info 创建的字段名、字段类型、主键正确，并正确显示表结构及主键和外键约束，得 16 分 |
| 数据操作           | 50 分 | 表 T_appr_area 和 T_proj_info 正确插入数据，得 5 分               |
|                |      | 查询出报批面积大于“1000”的建设用地的报批基本信息，得 5 分                      |
|                |      | 出报批时间为 2018 年 6 月份的所有建设用地的报批基本信息，得 5 分                 |
|                |      | 查询出地块名称为“长沙市王府花园”的用地的报批的项目名称和审批时间，得 5 分                |
|                |      | 查询出所有的地块基本信息，要求按地块面积升序排序，得 5 分                         |
|                |      | 删除地块面积大于“1100”的地块基本信息，得 5 分                            |
|                |      | 把地块名称为“长沙市王府花园”修改为“永州市王府花园”，得 5 分                      |

|            |                       |     |                                                                                                          |
|------------|-----------------------|-----|----------------------------------------------------------------------------------------------------------|
|            |                       |     | 在 T_plot_info 表中插入数据(地块编号为“10005”，用地项目编号为“104”，地块名称为“新型工业区”，地块面积为“1500”，地块用途为“工业用地”，图形类型为“长方形”的信息)，得 5 分 |
|            |                       |     | 创建的 Plot_id_view1 视图正确，得 5 分                                                                             |
|            |                       |     | 删除视图成功，得 5 分                                                                                             |
| 良好的习惯和职业素养 | 代码书写格式规范              | 4 分 | 脚本缩进、换行规范，得 4 分                                                                                          |
|            | 答题过程中保持安静，有问题举手示意，有礼貌 | 3 分 | 遵守考场纪律、有礼貌，得 3 分                                                                                         |
|            | 答题完毕离场时整理好工位          | 3 分 | 鼠标、键盘、凳子等摆放整齐，得 3 分                                                                                      |

## 项目 2：学生系统数据库设计

### 3.试题编号：J1-2-1：“学生成绩管理模块”数据库设计

#### (1) 任务描述

随着计算机的发展，学生的成绩管理摆脱了传统的人工统计的方式。通过计算机来进行成绩的管理，效率高、检索速度快、查找方便、可靠性高、保密性好。《成绩管理》模块的 E-R 图如图 3-1 所示，物理数据模型如图 3-2 所示，数据表字段名定义见表 3-1。请按要求完成数据库创建、数据表创建和数据操作任务：

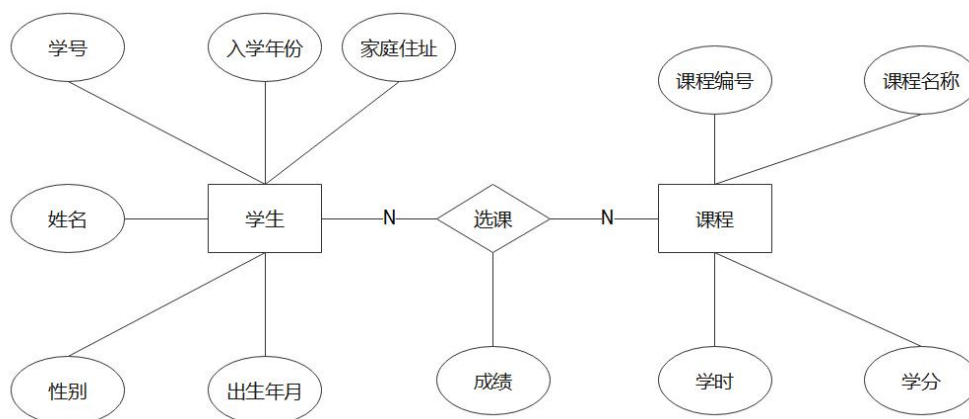


图 3-1E-R图

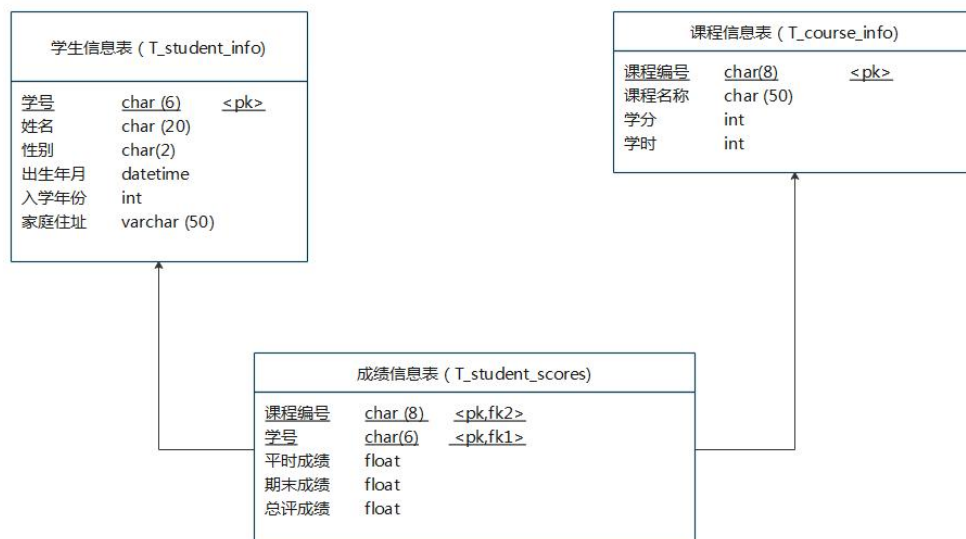


图 3-2 物理数据原型

表 3-1 字段名定义表

| 字段名          | 字段说明 | 字段名            | 字段说明 |
|--------------|------|----------------|------|
| Student_no   | 学号   | Credit         | 学分   |
| Student_name | 姓名   | Classhour      | 学时   |
| Sex          | 性别   | Course_name    | 课程名  |
| Birth        | 出生年月 | Course_no      | 课程编号 |
| Enter_date   | 入学年份 | Ordinary_score | 平时成绩 |
| Address      | 家庭住址 | End_scores     | 期末成绩 |
| Course_no    | 课程编号 | Total_scores   | 总评成绩 |

### 任务一：创建数据库（10 分）

创建数据库 ScoresDB。新建 T1-1.sql 文件，在该文件中编写语句创建指定的数据库，使用语句显示已经创建的数据库，并使用语句使用 ScoresDB 数据库。将上述操作截图，命名为 T1-1.jpg 文件，将 T1-1.sql、T1-1.jpg 保存到考生目录 T1 下。

### 任务二：创建数据表及表间的关系及约束（30 分）

创建 T2-1.sql 文件并完成以下功能：

- 根据图 3-1、3-2 和表 3-1，创建数据表 T\_student\_info、T\_course\_info、T\_student\_scores。

（15 分）

- 根据物理数据原型，创建数据表之间的关系。（15 分）

将创建表 T\_student\_info、T\_course\_info、T\_student\_scores 的语句以及执行结果截图，命名分别为 T2-1.jpg、T2-2.jpg、T2-3.jpg，将 T2-1.sql、T2-1.jpg、T2-2.jpg、T2-3.jpg 四个文件

保存到考生目录 T2 下。

### 任务三：数据操作（50 分）

考生目录 T3 下创建 T3-1.sql 文件并用 SQL 语句执行以下操作：

➤ 向表中插入以下数据：

向 T\_student\_info 表插入数据：

| 学号     | 姓名 | 性别 | 出生年月       | 入学年份 | 家庭住址   |
|--------|----|----|------------|------|--------|
| S00001 | 张三 | 男  | 1994-10-07 | 2012 | 长沙市天心区 |
| S00002 | 李四 | 女  | 1992-11-11 | 2012 | 长沙市雨花区 |
| S00003 | 张五 | 男  | 1993-10-07 | 2013 | 长沙市天心区 |
| S00004 | 李六 | 女  | 1994-11-11 | 2014 | 长沙市芙蓉区 |

向 T\_course\_info 表插入数据：

| 课程编号   | 课程名称          | 学分 | 学时 |
|--------|---------------|----|----|
| c10001 | 数据库原理         | 4  | 64 |
| c10002 | hadoop 大数据原理  | 4  | 72 |
| c10003 | Python 程序设计基础 | 4  | 64 |
| c10004 | 信息技术          | 2  | 32 |

向 T\_student\_scores 表插入数据：

| 课程编号   | 学号     | 平时成绩 | 期末成绩 | 总评成绩 |
|--------|--------|------|------|------|
| c10001 | s00001 | 40   | 80   | 75   |
| c10003 | s00001 | 40   | 72   | 75   |
| c10004 | s00002 | 60   | 60   | 60   |
| c10002 | s00003 | 60   | 90   | 60   |

查询 T\_student\_info、T\_course\_info、T\_student\_scores 三个表中的数据并截图，命名分别为 T3-1.jpg、T3-2.jpg、T3-3.jpg，保存到考生目录 T3 下。（10 分）

- 查询出出生年份在 1994 年-1995 年的所有学生信息，并将结果截图，命名为 T3-4.jpg，保存到考生目录 T3 下。（5 分）
- 统计出学生“张三”所选课程的总学分，并将结果截图，命名为 T3-5.jpg，保存到考生目录 T3 下。（5 分）
- 统计出课程号为“c10001”学生期末成绩的平均分，并将结果截图，命名为 T3-6.jpg，保存到考生目录 T3 下。（5 分）
- 创建视图 ProjInfo\_view1，查询“张三”的出生年月、入学年份、家庭住址，使用语句显示视图 ProjInfo\_view1 中的数据并截图，命名为 T3-7.jpg，保存到考生目录

T3 下。(5 分)

- 删除上题中所创建的 ProjInfo\_view1 视图，使用 show tables 语句查看所有数据库表，验证视图 ProjInfo\_view1 已经删除并将结果截图，命名为 T3-8.jpg，保存到考生目录 T3 下。(5 分)
- 统计每门课程的选课人数，结果列显示课程号和人数，并将结果截图，命名为 T3-9.jpg，保存到考生目录 T3 下。(10 分)
- 删除学号为“s00001”的学生的所有成绩，删除数据后，使用查询语句显示删除后的结果并截图，命名为 T3-10.jpg，保存到考生目录 T3 下。(5 分)

#### 良好的习惯和职业素养 (10 分)

- 答题过程中保持安静，有问题举手示意，有礼貌。(3 分)
- 数据库名命名规范，数据表名命名规范，字段名命名规范。(3 分)
- 答题完毕离场时整理好工位。(4 分)

#### (2) 实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。

计算机安装了 Windows 操作系统，MySQL 数据库软件以及 Office 办公软件。

#### (3) 考核时量及要求

- 考核时长 3 小时；
- 创建答题文件夹创建以“学号\_题号”命名的文件夹，存放所有答题文件；
- 创建答题文件，包括：SQL 脚本文件，如 J3-1.sql

#### (4) 评分细则

表3-2 考核评价细则表

| 评价项            | 分值   | 评分细则                                                  |
|----------------|------|-------------------------------------------------------|
| 数据库创建          | 10 分 | 创建考生目录 T1，文件 T1-1.sql、T1-1.jpg，得 1.5 分                |
|                |      | 使用语句创建了数据库，得 3 分                                      |
|                |      | 使用语句显示已经创建的数据库的名字，得 3 分                               |
|                |      | 使用新创建的数据库，得 2.5 分                                     |
| 创建数据表及表间的关系及约束 | 30 分 | 创建考生目录 T2，文件 T2-1.sql、T2-1.jpg、T2-2.jpg，得 2 分         |
|                |      | 表 T_course_info 创建的字段名、字段类型、主键正确，并正确显示表结构及主键约束，得 8 分  |
|                |      | 表 T_student_info 创建的字段名、字段类型、主键正确，并正确显示表结构及主键约束，得 8 分 |

|            |                       |     |                                                                                      |
|------------|-----------------------|-----|--------------------------------------------------------------------------------------|
|            |                       |     | 表 T_student_scores 创建的字段名、字段类型、主键正确，并正确显示表结构及主键和外键约束，得 12 分                          |
| 数据操作       | 50 分                  |     | 表 T_course_info 和 T_student_info 和 T_student_scores 正确插入数据，得 10 分                    |
|            |                       |     | 查询出生年份在 1994 年-1995 年的所有学生信息，得 5 分                                                   |
|            |                       |     | 统计出学生“张三”所选课程的总学分，得 5 分                                                              |
|            |                       |     | 统计出课程号为“c10001”学生期末成绩的平均分，得 5 分                                                      |
|            |                       |     | 创建视图 ProjInfo_view1，查询“张三”的出生年月、入学年份、家庭住址，使用语句显示视图 ProjInfo_view1 中的数据，得 5 分         |
|            |                       |     | 删除上题中所创建的 ProjInfo_view1 视图，使用 show tables 语句查看所有数据库表，验证视图 ProjInfo_view1 已经删除，得 5 分 |
|            |                       |     | 统计每门课程的选课人数，结果列显示课程号和人数，得 10 分                                                       |
|            |                       |     | 删除学号为“s00001”的学生的所有成绩，删除数据后，使用查询语句显示删除后的结果，得 5 分                                     |
| 良好的习惯和职业素养 | 代码书写格式规范              | 4 分 | 脚本缩进、换行规范，得 4 分                                                                      |
|            | 答题过程中保持安静，有问题举手示意，有礼貌 | 3 分 | 遵守考场纪律、有礼貌，得 3 分                                                                     |
|            | 答题完毕离场时整理好工位          | 3 分 | 鼠标、键盘、凳子等摆放整齐，得 3 分                                                                  |

#### 4.试题编号：J1-2-2：“学生选课管理模块”数据库设计

##### (1) 任务描述

学生选课系统作为一种现代化的教学技术，越来越受到人们的重视，是一个学校不可或缺的部分，学生选课系统就是为了管理好选课信息而设计的。学生选课系统的将使选课管理工作规范化、系统化、程序化，避免选课管理的随意性，提高信息处理的速度和准确性，能够准确、及时、有效的查询和修改学生选课情况。《选课管理》模块的 E-R 图如图 4-1 所示，物理数据模型如图 4-2 所示，数据表字段名定义见表 4-1。请按以下设计完成数据库创建、数据表创建和数据操作任务：

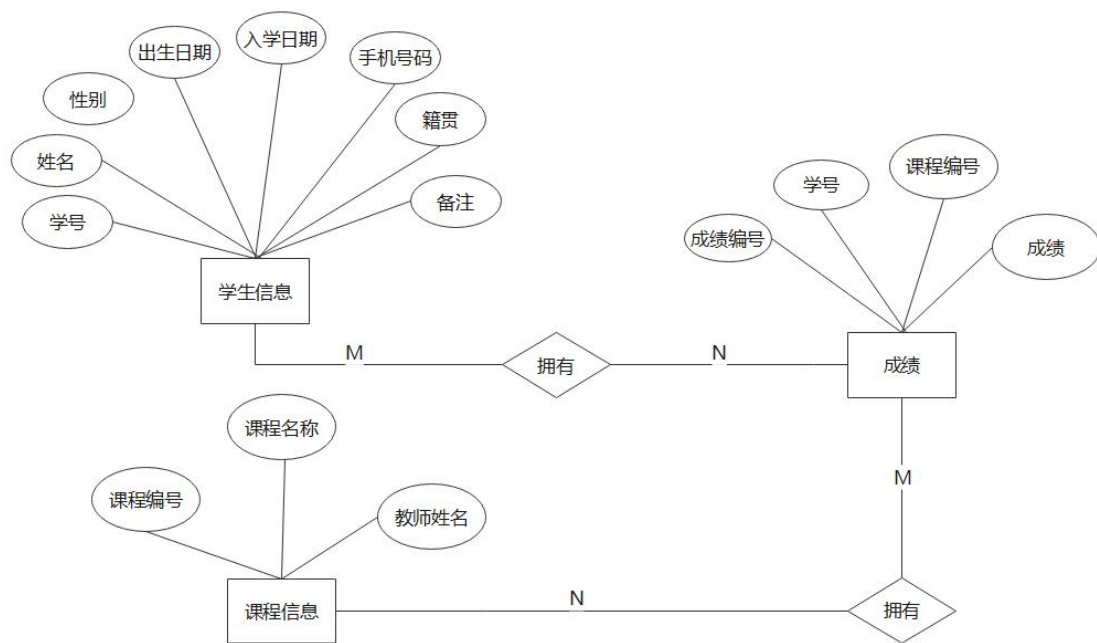


图 4-1 E-R 图

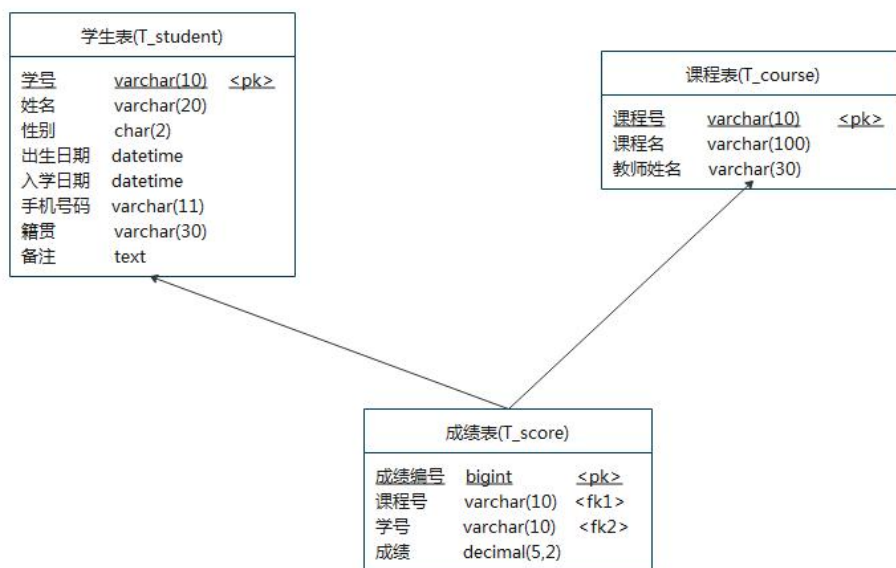


图 4-2 物理数据原型

表 4-1 字段名定义表

| 字段名        | 字段说明 | 字段名          | 字段说明 |
|------------|------|--------------|------|
| Stud_id    | 学号   | Reserve      | 备注   |
| Stud_name  | 姓名   | Course_id    | 课程编号 |
| Stud_sex   | 性别   | Course_name  | 课程名称 |
| Birth_date | 出生日期 | Teacher_name | 教师姓名 |
| Entry_Date | 入学日期 | Score_id     | 成绩编号 |
| Mobile     | 手机号码 | Score        | 成绩   |

|             |    |  |  |
|-------------|----|--|--|
| Birth_place | 籍贯 |  |  |
|-------------|----|--|--|

### 任务一：创建数据库（10 分）

创建数据库 CourseDB。新建 T1-1.sql 文件，在该文件中编写语句创建指定的数据库，使用语句显示已经创建的数据库，并使用语句使用 CourseDB 数据库。将上述操作截图，命名为 T1-1.jpg 文件，将 T1-1.sql、T1-1.jpg 保存到考生目录 T1 下。

### 任务二：创建数据表及数据表间的关系及约束（30 分）

创建 T2-1.sql 文件并完成以下功能：

- 根据图 4-1、4-2 和表 4-1，创建数据表 T\_student、T\_course、T\_score。（15 分）
- 根据物理数据原型，创建数据表之间的关系。（15 分）

将创建表 T\_student、T\_course、T\_score 的语句以及执行结果截图，命名分别为 T2-1.jpg、T2-2.jpg、T2-3.jpg，将 T2-1.sql、T2-1.jpg、T2-2.jpg、T2-3.jpg 四个文件保存到考生目录 T2 下。

### 任务三：数据操作（50 分）

考生目录 T3 下创建 T3-1.sql 文件并用 SQL 语句执行以下操作：

- 在每个表中插入以下测试数据：

向 T\_student 表中插入数据：

| 学号     | 姓名 | 性别 | 出生日期       | 入学日期       | 手机号码        | 籍贯 | 备注 |
|--------|----|----|------------|------------|-------------|----|----|
| s00001 | 张三 | 男  | 1991-01-01 | 2018-09-01 | 15100000000 | 北京 | 无  |
| s00002 | 李四 | 女  | 1992-02-02 | 2018-09-01 | 17811111111 | 上海 | 无  |
| s00003 | 赵柳 | 男  | 1990-11-01 | 2018-09-01 | 19999999999 | 南京 | 休学 |
| s00004 | 宋其 | 女  | 1992-02-16 | 2018-09-01 | 13457777777 | 长沙 | 无  |

向 T\_course 表中插入以下数据：

| 课程号    | 课程名       | 教师姓名  |
|--------|-----------|-------|
| c00001 | 数据结构      | 石正喜   |
| c00002 | python 基础 | 黑马程序员 |
| c00003 | 大数据平台搭建   | 夏丽琴   |
| c00004 | java 高级编程 | 尹一奇   |
| c00005 | 数据库原理     | 周军    |
| c00006 | 大学英语      | jack  |

向 T\_score 表中插入以下数据：

| 课程编号 | 课程号    | 学号     | 成绩 |
|------|--------|--------|----|
| 101  | c00001 | s00001 | 40 |
| 102  | c00001 | s00002 | 60 |
| 103  | c00001 | s00003 | 78 |
| 104  | c00001 | s00004 | 98 |
| 105  | c00002 | s00001 | 80 |

|     |        |        |    |
|-----|--------|--------|----|
| 106 | c00002 | s00002 | 65 |
| 107 | c00003 | s00003 | 77 |
| 108 | c00004 | s00004 | 69 |
| 109 | c00004 | s00002 | 65 |
| 110 | c00005 | s00003 | 66 |
| 111 | c00005 | s00004 | 88 |

查询 T\_student、T\_course、T\_score 三个表中的数据并截图,命名分别为 T3-1.jpg、T3-2.jpg、T3-3.jpg, 保存到考生目录 T3 下。(5 分)

- 查询出所有选修了课程号为“c00005”的学生学号、姓名和籍贯, 并将结果截图, 命名为 T3-4.jpg, 保存到考生目录 T3 下。(5 分)
- 查询出学生学号为“s00001”所有选修课程的最高分, 并将结果截图, 命名为 T3-5.jpg, 保存到考生目录 T3 下。(10 分)
- 查询出选修了课程编号为“c00005”这门课的学生的学号和成绩, 并按成绩降序拍, 并将结果截图, 命名为 T3-6.jpg, 保存到考生目录 T3 下。(10 分)
- 查询有哪些课程没有被任何同学选修, 要求显示未被选修的课程名字, 并将结果截图, 命名为 T3-7.jpg, 保存到考生目录 T3 下。(5 分)
- 创建视图 Student\_info\_view1, 查询“数据库原理”教师的名字, 使用语句显示视图 Student\_info\_view1 中的数据并截图, 命名为 T3-8.jpg, 保存到考生目录 T3 下。(10 分)
- 删除上题中所创建的 Student\_info\_view1 视图, 使用 show tables 语句查看所有数据库表, 验证视图 ProjInfo\_view1 已经删除并将结果截图, 命名为 T3-9.jpg, 保存到考生目录 T3 下。(5 分)

#### 良好的习惯和职业素养 (10 分)

- 答题过程中保持安静, 有问题举手示意, 有礼貌。(3 分)
- 数据库名命名规范, 数据表名命名规范, 字段名命名规范。(3 分)
- 答题完毕离场时整理好工位。(4 分)

## (2) 实施条件

安排在教学实训机房, 计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。计算机安装了 Windows 操作系统, MySQL 数据库软件以及 Office 办公软件。

### (3) 考核时量及要求

- 考核时长 3 小时；
- 创建答题文件夹创建以“学号\_题号”命名的文件夹，存放所有答题文件；
- 创建答题文件，包括：SQL 脚本文件，如 J3-1.sql

### (4) 评分细则

表4-2 考核评价细则表

| 评价项            |                       | 分值   | 评分细则                                                                                     |
|----------------|-----------------------|------|------------------------------------------------------------------------------------------|
| 数据库创建          |                       | 10 分 | 创建考生目录 T1，文件 T1-1.sql、T1-1.jpg，得 1.5 分                                                   |
|                |                       |      | 使用语句创建了数据库，得 3 分                                                                         |
|                |                       |      | 使用语句显示已经创建的数据库的名字，得 3 分                                                                  |
|                |                       |      | 使用新创建的数据库，得 2.5 分                                                                        |
| 创建数据表及表间的关系及约束 |                       | 30 分 | 创建考生目录 T2，文件 T2-1.sql、T2-1.jpg、T2-2.jpg，得 2 分                                            |
|                |                       |      | 表 T_student 创建的字段名、字段类型、主键正确，并正确显示表结构及主键约束，得 8 分                                         |
|                |                       |      | 表 T_course 创建的字段名、字段类型、主键正确，并正确显示表结构及主键约束，得 8 分                                          |
|                |                       |      | 表 T_scores 创建的字段名、字段类型、主键正确，并正确显示表结构及主键和外键约束，得 12 分                                      |
| 数据操作           |                       | 50 分 | 表 T_student 和 T_course 和 T_scores 正确插入数据，得 5 分                                           |
|                |                       |      | 查询出所有选修了课程号为“c00005”的学生学号、姓名和籍贯，得 5 分                                                    |
|                |                       |      | 查询出学生学号为“s00001”所有选修课程的最高分，得 10 分                                                        |
|                |                       |      | 查询出选修了课程编号为“c00005”这门课的学生的学号和成绩，得 10 分                                                   |
|                |                       |      | 查询有哪些课程没有被任何同学选修，要求显示未被选修的课程名字，得 5 分                                                     |
|                |                       |      | 创建视图 Student_info_view1，查询“数据库原理”教师的名字，使用语句显示视图 Student_info_view1 中的数据，得 10 分           |
|                |                       |      | 删除上题中所创建的 Student_info_view1 视图，使用 show tables 语句查看所有数据库表，验证视图 ProjInfo_view1 已经删除，得 5 分 |
| 良好的习惯和职业素养     | 代码书写格式规范              | 4 分  | 脚本缩进、换行规范，得 4 分                                                                          |
|                | 答题过程中保持安静，有问题举手示意，有礼貌 | 3 分  | 遵守考场纪律、有礼貌，得 3 分                                                                         |

|  |                  |     |                     |
|--|------------------|-----|---------------------|
|  | 答题完毕离场时<br>整理好工位 | 3 分 | 鼠标、键盘、凳子等摆放整齐，得 3 分 |
|--|------------------|-----|---------------------|

### 项目 3：学生食堂饭卡管理数据库设计

#### 5.试题编号：J1-3-1：“学生食堂饭卡消费管理模块”数据库设计

##### (1) 任务描述

饭卡管理系统是一套针对大学校园食堂交费和一般消费等方面的信息管理系统,它包括学生在校内消费的各方面内容:刷卡消费、查询、存款和持卡者信息管理等,方便对饭卡信息进行各项操作。《饭卡管理》模块的 E-R 图如图 5-1 所示,物理数据模型如图 5-2 所示,数据表字段名定义见表 5-1。请按以下设计完成数据库创建、数据表创建和数据操作任务:

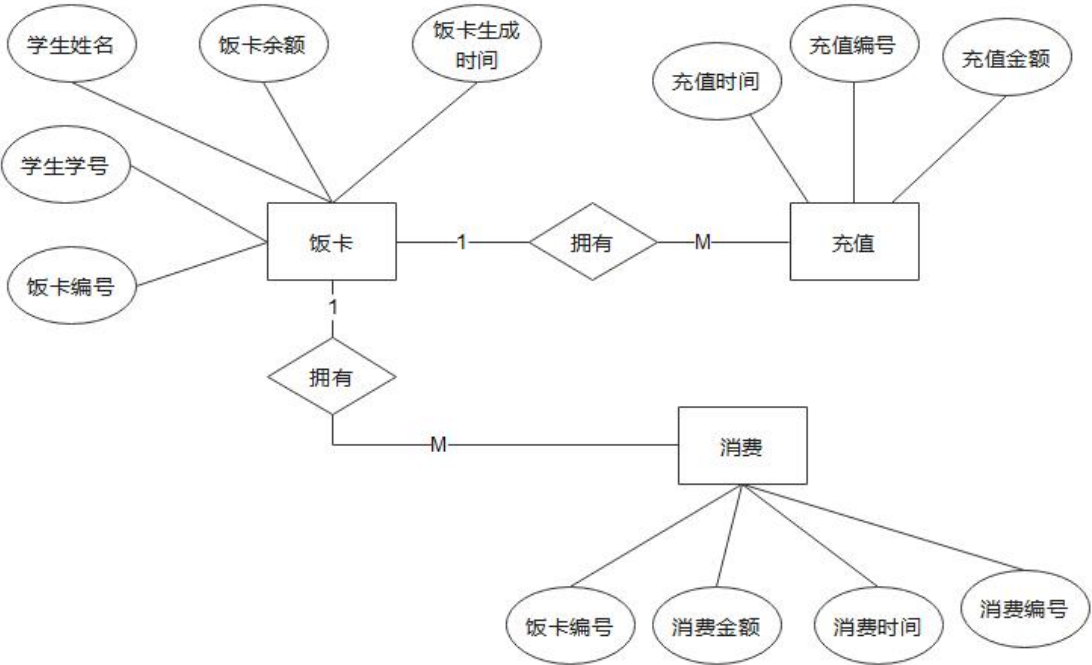


图 5-1 E-R

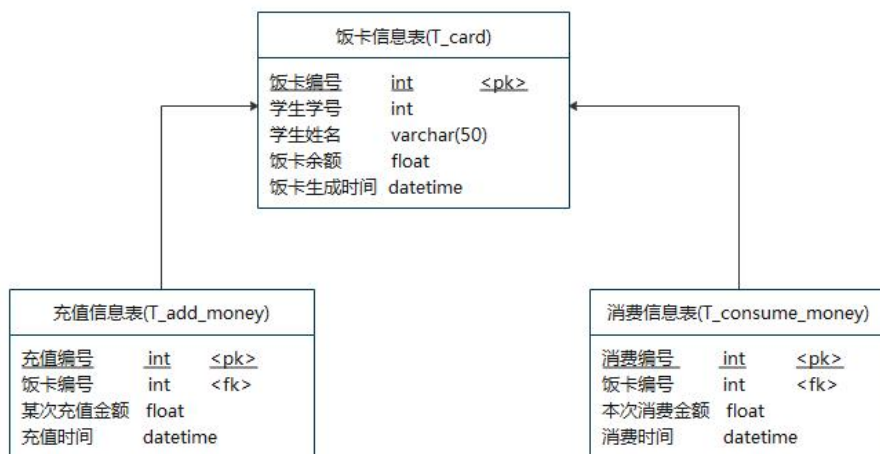


图5-2 物理数据模型图

表5-1 字段名定义表

| 字段名           | 字段说明         | 字段名           | 字段说明   |
|---------------|--------------|---------------|--------|
| card_id       | 饭卡编号         | add_money     | 某次充值金额 |
| student_id    | 学生学号         | recharge_date | 充值时间   |
| student_name  | 学生姓名         | consume_id    | 消费编号   |
| curr_money    | 饭卡余额(初始值为 0) | consume_money | 本次消费金额 |
| register_date | 饭卡生成时间       | consume_date  | 消费时间   |
| add_id        | 充值编号         |               |        |

### 任务一：创建数据库（10 分）

创建数据库 CardDB。新建 T1-1.sql 文件，在该文件中编写语句创建指定的数据库，使用语句显示已经创建的数据库，并使用语句使用 CardDB 数据库。将上述操作截图，命名为 T1-1.jpg 文件，将 T1-1.sql、T1-1.jpg 保存到考生目录 T1 下。

### 任务二：创建数据表间的关系及约束（30 分）

创建 T2-1.sql 文件并完成以下功能：

- 根据图 5-1、5-2 和表 5-1，创建数据表 T\_card、T\_add\_money、T\_consume\_money。

（15 分）

- 根据物理数据原型，创建数据表之间的关系。（15 分）

将创建表 T\_card、T\_add\_money、T\_consume\_money 的语句以及执行结果截图，命名分别为 T2-1.jpg、T2-2.jpg、T2-3.jpg，将 T2-1.sql、T2-1.jpg、T2-2.jpg、T2-3.jpg 四个文件保存到考生目录 T2 下。

### 任务三：数据操作（50 分）

考生目录 T3 下创建 T3-1.sql 文件并用 SQL 语句执行以下操作：

- 在每个表中插入以下测试数据:

向 T\_card 表中插入数据:

| 饭卡编号 | 学生学号      | 学生姓名 | 饭卡余额  | 饭卡生成时间     |
|------|-----------|------|-------|------------|
| 1    | 201865111 | 张三   | 100.1 | 2018-06-02 |
| 2    | 201865112 | 李四   | 156   | 2018-06-04 |
| 3    | 201865113 | 王五   | 148.5 | 2018-06-07 |
| 4    | 201865114 | 赵六   | 26.8  | 2018-06-09 |

向 T\_add\_money 表中插入数据:

| 充值编号 | 饭卡编号 | 充值金额 | 充值时间       |
|------|------|------|------------|
| 1001 | 1    | 50   | 2018-10-02 |
| 1002 | 2    | 100  | 2018-11-04 |
| 1003 | 3    | 150  | 2018-12-07 |
| 1004 | 1    | 50   | 2019-01-09 |
| 1005 | 4    | 40   | 2019-06-02 |

向 T\_consume\_money 表中插入数据:

| 消费编号 | 饭卡编号 | 消费金额 | 消费时间       |
|------|------|------|------------|
| 2001 | 1    | 5.5  | 2018-07-02 |
| 2002 | 2    | 15   | 2018-10-04 |
| 2003 | 3    | 13   | 2018-12-07 |
| 2004 | 1    | 7.5  | 2019-03-09 |
| 2005 | 1    | 8    | 2019-05-02 |

查询 T\_card、T\_add\_money、T\_consume\_money 三个表中的数据并截图，命名分别为 T3-1.jpg、T3-2.jpg、T3-3.jpg，保存到考生目录 T3 下。(5 分)

- 查询饭卡生成时间为 2018 年 6 月的饭卡信息，包括：卡号、姓名、余额，并将结果截图，命名为 T3-4.jpg，保存到考生目录 T3 下。(5 分)
- 查询所有饭卡的余额之和，并将结果截图，命名为 T3-5.jpg，保存到考生目录 T3 下。(5 分)
- 查询消费金额大于 10 的消费编号和消费时间，并将结果截图，命名为 T3-6.jpg，保存到考生目录 T3 下。(5 分)
- 查询 2019 年学生单次消费金额最高的消费编号，并将结果截图，命名为 T3-7.jpg，保存到考生目录 T3 下。(5 分)
- 查询 T\_add\_money 表中的数据，将数据按照充值金额降序排取第一条数据，并将结果截图，命名为 T3-8.jpg，保存到考生目录 T3 下。(10 分)
- 创建视图 consume\_view，查询 2018 年学生的月平均消费金额 (2018 年全年的消

费总额除以 12),使用语句显示视图 consume\_view 中的数据并截图,命名为 T3-9.jpg,保存到考生目录 T3 下。(10 分)

- 删除上题中创建的视图,使用 show tables 语句查看所有数据库表,验证视图 consume\_view 已经删除并将结果截图,命名为 T3-10.jpg,保存到考生目录 T3 下。(5 分)

### 良好的习惯和职业素养 (10 分)

- 答题过程中保持安静,有问题举手示意,有礼貌。(3 分)
- 数据库名命名规范,数据表名命名规范,字段名命名规范。(3 分)
- 答题完毕离场时整理好工位。(4 分)

## (2) 实施条件

安排在教学实训机房,计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。

计算机安装了 Windows 操作系统,MySQL 数据库软件以及 Office 办公软件。

## (3) 考核时量及要求

- 考核时长 3 小时;
- 创建答题文件夹创建以“学号\_题号”命名的文件夹,存放所有答题文件;
- 创建答题文件,包括:SQL 脚本文件,如 J3-1.sql

## (4) 评分细则

表5-2 考核评价细则表

| 评价项            | 分值   | 评分细则                                                         |
|----------------|------|--------------------------------------------------------------|
| 数据库创建          | 10 分 | 创建考生目录 T1, 文件 T1-1.sql、T1-1.jpg, 得 1.5 分                     |
|                |      | 使用语句创建了数据库, 得 3 分                                            |
|                |      | 使用语句显示已经创建的数据库的名字, 得 3 分                                     |
|                |      | 使用新创建的数据库, 得 2.5 分                                           |
| 创建数据表及表间的关系及约束 | 30 分 | 创建考生目录 T2, 文件 T2-1.sql、T2-1.jpg、T2-2.jpg, 得 2 分              |
|                |      | 表 T_card 创建的字段名、字段类型、主键正确, 并正确显示表结构及主键约束, 得 8 分              |
|                |      | 表 T_add_money 创建的字段名、字段类型、主键正确、外键正确, 并正确显示表结构及主键约束, 得 10 分   |
|                |      | 表 T_consume_money 创建的字段名、字段类型、主键正确, 并正确显示表结构及主键和外键约束, 得 10 分 |
| 数据操作           | 50 分 | 表 T_card 和 T_add_money 和 T_consume_money 正确插入数据, 得 5 分       |

|            |                       |     |                                                                                             |
|------------|-----------------------|-----|---------------------------------------------------------------------------------------------|
|            |                       |     | 查询饭卡生成时间为 2018 年 6 月的饭卡信息，包括：卡号、姓名、余额，得 5 分                                                 |
|            |                       |     | 查询所有饭卡的余额之和，得 5 分                                                                           |
|            |                       |     | 查询消费金额大于 10 的消费编号和消费时间，得 5 分                                                                |
|            |                       |     | 查询 2019 年学生单次消费金额最高的消费编号，得 5 分                                                              |
|            |                       |     | 查询 T_add_money 表中的数据，将数据按照充值金额降序排取第一条数据，得 10 分                                              |
|            |                       |     | 创建视图 consume_view，查询 2018 年学生的月平均消费金额（2018 年全年的消费总额除以 12），使用语句显示视图 consume_view 中的数据，得 10 分 |
|            |                       |     | 删除上题中创建的视图，使用 show tables 语句查看所有数据库表，验证视图 consume_view 已经删除，得 5 分                           |
| 良好的习惯和职业素养 | 代码书写格式规范              | 4 分 | 脚本缩进、换行规范，得 4 分                                                                             |
|            | 答题过程中保持安静，有问题举手示意，有礼貌 | 3 分 | 遵守考场纪律、有礼貌，得 3 分                                                                            |
|            | 答题完毕离场时整理好工位          | 3 分 | 鼠标、键盘、凳子等摆放整齐，得 3 分                                                                         |

## 6.试题编号：J1-3-2：“学生食堂饭卡充值管理模块”数据库设计

### （1）任务描述

饭卡管理系统是一套针对大学校园食堂交费和一般消费等方面的信息管理系统，它包括学生在校内消费的各方面内容：刷卡消费、查询、存款和持卡者信息管理等，方便对饭卡信息进行各项操作。《饭卡管理》模块的 E-R 图如图 6-1 所示，物理数据模型如图 6-2 所示，数据表字段名定义见表 6-1。请按以下设计完成数据库创建、数据表创建和数据操作任务：

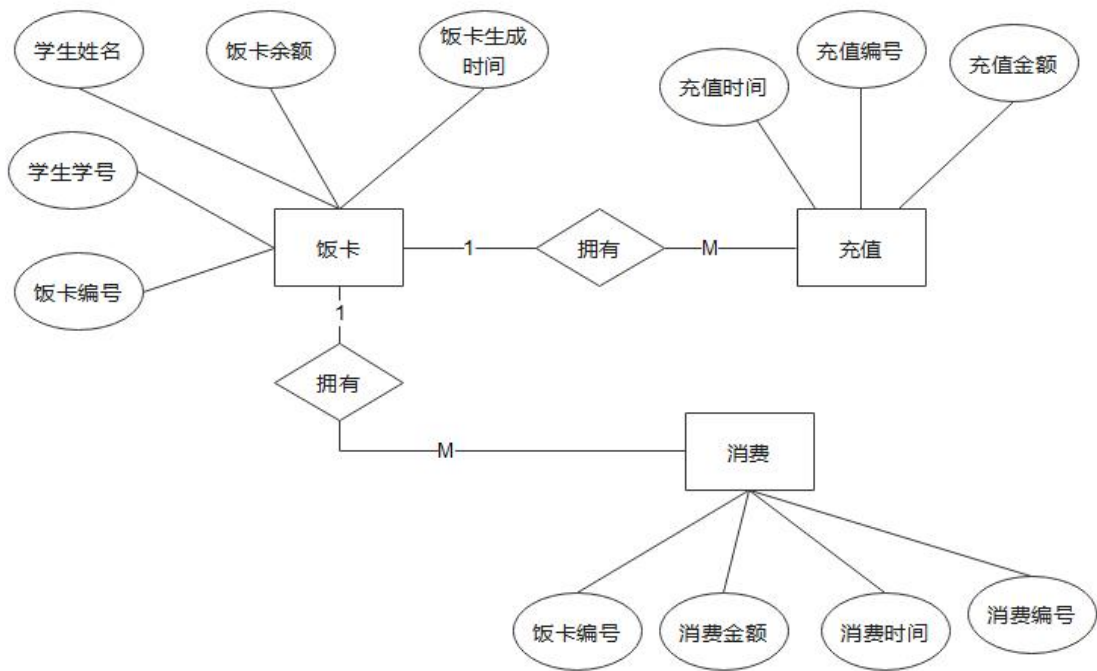


图 6-1 E-R 图

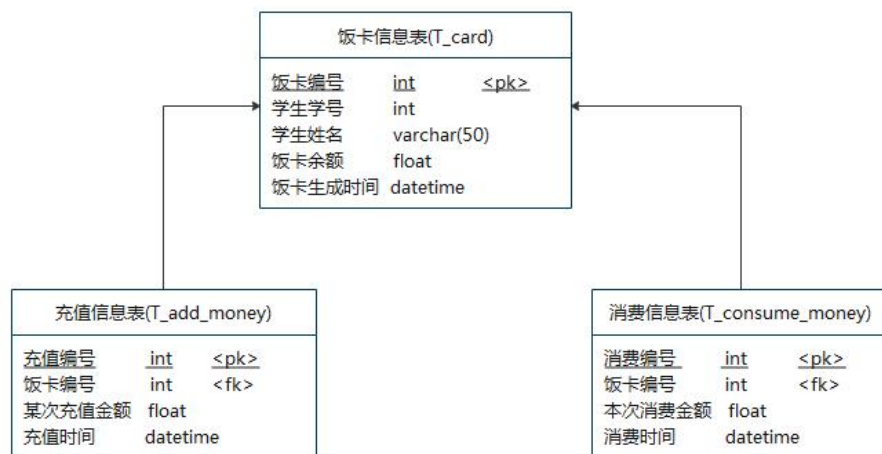


图 6-2 物理数据模型图

表 6-1 字段名定义表

| 字段名           | 字段说明              | 字段名           | 字段说明            |
|---------------|-------------------|---------------|-----------------|
| card_id       | 饭卡编号              | add_money     | 某次充值金额          |
| student_id    | 学生学号              | recharge_date | 充值时间；默认值为当前录入时间 |
| student_name  | 学生姓名              | consume_id    | 消费编号            |
| curr_money    | 饭卡余额(初始值为 0)      | consume_money | 本次消费金额          |
| register_date | 饭卡生成时间；默认值为当前录入时间 | consume_date  | 消费时间；默认值为当前录入时间 |
| add_id        | 充值编号              |               |                 |

### 任务一：创建数据库（10 分）

创建数据库 CardDB。新建 T1-1.sql 文件，在该文件中编写语句创建指定的数据库，使用语句显示已经创建的数据库，并使用语句使用 CardDB 数据库。将上述操作截图，命名为 T1-1.jpg 文件，将 T1-1.sql、T1-1.jpg 保存到考生目录 T1 下。

### 任务二：创建数据表及表间的关系及约束（40 分）

创建 T2-1.sql 文件并完成以下功能：

- 根据图 6-1、6-2 和表 6-1，创建数据表 T\_card、T\_add\_money、T\_consume\_money，其中将饭卡表的饭卡编号 (card\_id)列设置为标识列，自动从 1 开始增长。（25 分）
- 根据物理数据原型，创建数据关系表。（15 分）

将创建表 T\_card、T\_add\_money、T\_consume\_money 的语句以及执行结果截图，命名分别为 T2-1.jpg、T2-2.jpg、T2-3.jpg，将 T2-1.sql、T2-1.jpg、T2-2.jpg、T2-3.jpg 四个文件保存到考生目录 T2 下。

### 任务三：数据操作（40 分）

考生目录 T3 下创建 T3-1.sql 文件并用 SQL 语句执行以下操作：

- 在每个表中插入以下测试数据：

向 T\_card 表插入数据：

| 饭卡编号    | 学生学号 | 学生姓名 | 饭卡余额   | 饭卡生成时间     |
|---------|------|------|--------|------------|
| default | 101  | 张滨   | 100.0  | 2019-09-01 |
| default | 102  | 周杰   | 1000.0 | 2019-09-01 |
| default | 103  | 张三   | 500.0  | 2019-09-01 |
| default | 104  | 张艳   | 112.0  | 2019-09-01 |
| default | 105  | 代香   | 109.0  | 2019-09-01 |
| default | 106  | 刘强   | 512.1  | 2019-09-01 |

向 T\_add\_money 表插入数据

| 充值编号 | 饭卡编号 | 充值金额  | 充值时间                 |
|------|------|-------|----------------------|
| 1001 | 1    | 150.0 | 2019-09-18 08:31:00  |
| 1002 | 2    | 500.0 | 2019-09-18 12:12:00  |
| 1003 | 3    | 22.0  | 2019-07-18 08:31:00  |
| 1004 | 4    | 500.0 | 2019-09-18 12:12:00' |
| 1005 | 5    | 150.0 | 2019-10-18 10:15:00  |

向 T\_consume\_money 表插入数据

| 消费编号  | 饭卡编号 | 消费金额 | 消费时间                |
|-------|------|------|---------------------|
| 10001 | 1    | 5.0  | 2019-09-19 14:12:00 |
| 10002 | 2    | 15.0 | 2019-09-20 18:12:00 |
| 10003 | 3    | 7.0  | 2019-09-21 17:12:00 |
| 10004 | 4    | 50.0 | 2019-05-19 13:12:00 |
| 10005 | 5    | 45.0 | 2019-11-20 18:21:00 |
| 10006 | 3    | 15.0 | 2019-09-25 09:12:00 |
| 10007 | 3    | 45.0 | 2019-09-25 17:56:00 |
| 10008 | 3    | 23.0 | 2019-09-29 07:32:00 |
| 10009 | 6    | 67.0 | 2019-10-21 19:12:00 |

查询 T\_card、T\_add\_money、T\_consume\_money 三个表中的数据并截图，命名分别为 T3-1.jpg、T3-2.jpg、T3-3.jpg，保存到考生目录 T3 下。（5 分）

- 查询饭卡编号为“2”的同学所有的消费记录，并将结果截图，命名为 T3-4.jpg，保存到考生目录 T3 下。（5 分）
- 查询所有饭卡的余额之和，并将结果截图，命名为 T3-5.jpg，保存到考生目录 T3 下。（5 分）
- 查询学生“张三”2019 年消费总金额，并将结果截图，命名为 T3-6.jpg，保存到考生目录 T3 下。（5 分）
- 查询 T\_consume\_money 表中单次消费最多的金额，并将结果截图，命名为 T3-7.jpg，保存到考生目录 T3 下。（10 分）
- 查询每个月份充值金额的总数，查询结果的字段命名为：充值金额的总数和月份，并将结果按充值金额的总数降序排。结果截图，命名为 T3-8.jpg，保存到考生目录 T3 下。（10 分）
- 创建视图 ProjInfo\_view1 用于查询学生单次消费的平均值，使用语句显示视图 ProjInfo\_view1 中的数据并截图，命名为 T3-9.jpg，保存到考生目录 T3 下。（5 分）
- 删除上题中创建的视图，使用 show tables 语句查看所有数据库表，验证视图 ProjInfo\_view1 已经删除并将结果截图，命名为 T3-10.jpg，保存到考生目录 T3 下。

(5 分)

### 良好的习惯和职业素养 (10 分)

- 答题过程中保持安静，有问题举手示意，有礼貌。(3 分)
- 数据库名命名规范，数据表名命名规范，字段名命名规范。(3 分)
- 答题完毕离场时整理好工位。(4 分)

### (2) 实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。

计算机安装了 Windows 操作系统，MySQL 数据库软件以及 Office 办公软件。

### (3) 考核时量及要求

- 考核时长 3 小时；
- 创建答题文件夹创建以“学号\_题号”命名的文件夹，存放所有答题文件；
- 创建答题文件，包括：SQL 脚本文件，如 J3-1.sql

### (4) 评分细则

表6-2考核评价细则表

| 评价项            | 分值   | 评分细则                                                       |
|----------------|------|------------------------------------------------------------|
| 数据库创建          | 10 分 | 创建考生目录 T1，文件 T1-1.sql、T1-1.jpg，得 1.5 分                     |
|                |      | 使用语句创建了数据库，得 3 分                                           |
|                |      | 使用语句显示已经创建的数据库的名字，得 3 分                                    |
|                |      | 使用新创建的数据库，得 2.5 分                                          |
| 创建数据表及表间的关系及约束 | 30 分 | 创建考生目录 T2，文件 T2-1.sql、T2-1.jpg、T2-2.jpg，得 2 分              |
|                |      | 表 T_card 创建的字段名、字段类型、主键正确，并正确显示表结构及主键约束，得 8 分              |
|                |      | 表 T_add_money 创建的字段名、字段类型、主键正确、外键正确，并正确显示表结构及主键约束，得 10 分   |
|                |      | 表 T_consume_money 创建的字段名、字段类型、主键正确，并正确显示表结构及主键和外键约束，得 10 分 |
| 数据操作           | 50 分 | 表 T_card 和 T_add_money 和 T_consume_money 正确插入数据，得 5 分      |
|                |      | 查询饭卡编号为“2”的同学所有的消费记录，得 5 分                                 |
|                |      | 查询所有饭卡的余额之和，得 5 分                                          |
|                |      | 查询学生“张三”2019 年消费总金额，得 5 分                                  |
|                |      | 查询 T_consume_money 表中单次消费最多的金额，得 10 分                      |
|                |      | 查询每个月份充值金额的总数，查询结果的字段命名为：                                  |

|            |                       |     |                                                                       |
|------------|-----------------------|-----|-----------------------------------------------------------------------|
|            |                       |     | 充值金额的总数和月份，并将结果按充值金额的总数降序排，得 10 分                                     |
|            |                       |     | 创建视图 ProjInfo_view1 用于查询学生单次消费的平均值，使用语句显示视图 ProjInfo_view1 中的数据，得 5 分 |
|            |                       |     | 删除上题中创建的视图，使用 show tables 语句查看所有数据库表，验证视图 ProjInfo_view1 已经删除，得 5 分   |
| 良好的习惯和职业素养 | 代码书写格式规范              | 4 分 | 脚本缩进、换行规范，得 4 分                                                       |
|            | 答题过程中保持安静，有问题举手示意，有礼貌 | 3 分 | 遵守考场纪律、有礼貌，得 3 分                                                      |
|            | 答题完毕离场时整理好工位          | 3 分 | 鼠标、键盘、凳子等摆放整齐，得 3 分                                                   |

## 项目 4：超市系统数据库设计

### 7.试题编号：J1-4-1：“大型超市商品管理模块”数据库设计

#### (1) 任务描述

最初的超市资料管理，都是靠人力来完成的。但近几年我国超市经营规模日趋扩大，销售额和门店数量大幅度增加，而且许多超市正在突破以食品为主的传统格局，向品种多样化发展。小型超市在业务上需要处理大量的库存信息，还要时刻更新产品的销售信息，不断添加商品信息，并对商品各种信息进行统计分析。《商品管理》模块的 E-R 图如图 7-1 所示，物理数据模型如图 7-2 所示，数据表字段名定义见表 7-1。请按以下设计完成数据库创建、数据表创建和数据操作任务：

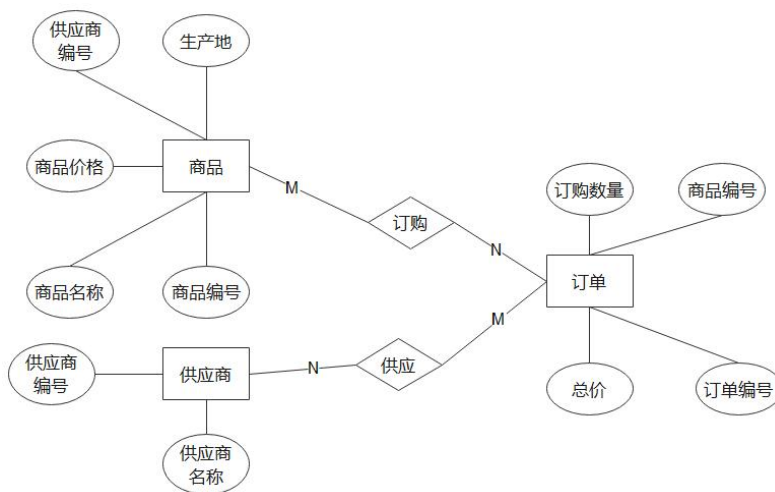


图 7-1 E-R 图

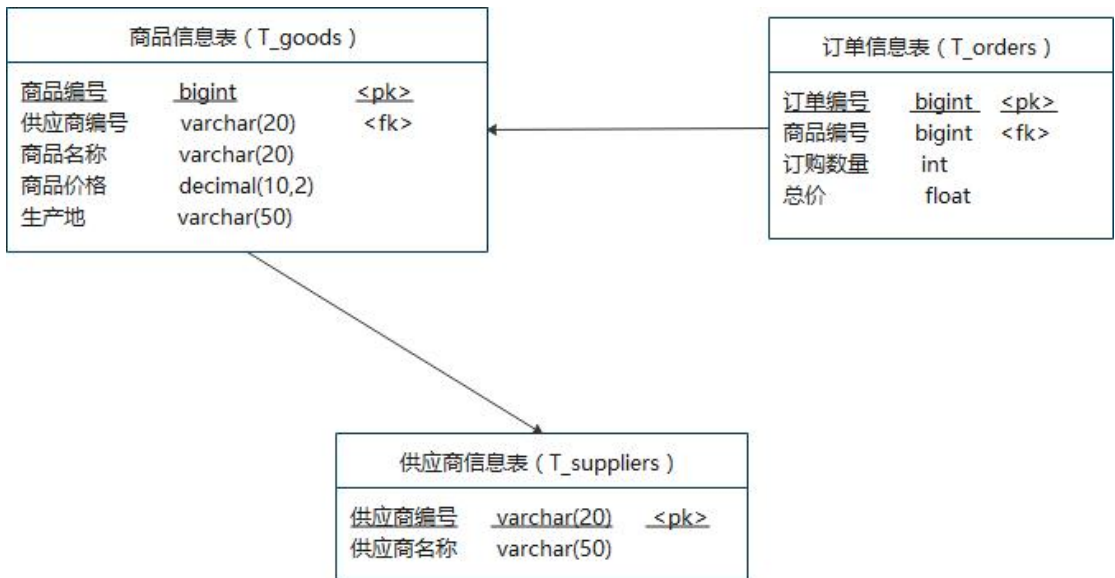


图 7-2 物理数据模型

表 7-1 字段名定义表

| 字段名          | 字段说明 | 字段名         | 字段说明  |
|--------------|------|-------------|-------|
| goods_id     | 商品编号 | quantity    | 订购数量  |
| goods_name   | 商品名称 | total_price | 总价    |
| goods_price  | 商品价格 | supp_code   | 供应商编号 |
| goods_adress | 生产地  | supp_name   | 供应商名称 |
| order_id     | 订单编号 |             |       |

任务一：创建数据库（10 分）

创建数据库 StoresDB。新建 T1-1.sql 文件，在该文件中编写语句创建指定的数据库，使用语句显示已经创建的数据库，并使用语句使用 StoresDB 数据库。将上述操作截图，命名为 T1-1.jpg 文件，将 T1-1.sql、T1-1.jpg 保存到考生目录 T1 下。

任务二：创建数据表及表间的关系及约束（30 分）

创建 T2-1.sql 文件并完成以下功能：

- 根据图 7-1、7-2 和表 7-1，创建数据表 T\_suppliers、T\_goods、T\_orders。（15 分）
- 根据物理数据原型，创建数据表之间的关系。（15 分）

将创建表 T\_suppliers、T\_goods、T\_orders 的语句以及执行结果截图，命名分别为 T2-1.jpg、T2-2.jpg、T2-3.jpg，将 T2-1.sql、T2-1.jpg、T2-2.jpg、T2-3.jpg 四个文件保存到考生目录 T2 下。

任务三：数据操作（50 分）

考生目录 T3 下创建 T3-1.sql 文件并用 SQL 语句执行以下操作：

- 在 3 个表中分别插入表种样本数据，查询 T\_suppliers、T\_goods、T\_orders 三个表中的数据并截图，命名分别为 T3-1.jpg、T3-2.jpg、T3-3.jpg，保存到考生目录 T3 下。（5 分）

供应商表(T\_suppliers样本数据)

| 供应商编号  | 供应商名称 |
|--------|-------|
| 430102 | 联想科技  |
| 540199 | 宏基科技  |
| 440708 | 惠普科技  |

商品信息表(T\_goods样本数据)

| 商品编号 | 供应商编号  | 商品名称  | 商品价格 | 生产地 |
|------|--------|-------|------|-----|
| 1000 | 430102 | 联想笔记本 | 6600 | 广东  |
| 1001 | 540199 | 宏基笔记本 | 7200 | 香港  |
| 1002 | 440708 | 惠普笔记本 | 8600 | 广东  |

订单信息表 (T\_orders样本数据)

| 订单编号     | 商品编号 | 订购数量 | 总价    |
|----------|------|------|-------|
| 11070232 | 1000 | 2    | 13200 |
| 11060343 | 1001 | 3    | 21600 |
| 11050322 | 1002 | 1    | 8600  |
| 11040322 | 1001 | 2    | 14400 |
| 11020322 | 1002 | 3    | 25800 |

- 将数据表中商品名为“宏基笔记本”的价格上调 5%，使用查询语句显示修改后的数据，并将结果截图，命名为 T3-4.jpg，保存到考生目录 T3 下。（5 分）
- 查询出商品编号为“1001”的总的订购数量，并将结果截图，命名为 T3-5.jpg，保存到考生目录 T3 下。（5 分）
- 查询供应商为“联想科技”的所有订单的商品名称、商品价格，并将结果截图，命名为 T3-6.jpg，保存到考生目录 T3 下。（10 分）
- 查询商品信息表(T\_goods)中的数据，并按价格降序排，将结果截图，命名为 T3-7.jpg，保存到考生目录 T3 下。（5 分）
- 查询商品信息表 (T\_goods) 中商品的平均价格，并将结果截图，命名为 T3-8.jpg，保存到考生目录 T3 下。（10 分）
- 订单信息表中增加一个字段“备注”（字段名：Reserve、字段类型：text）。修改完成后显示表结构，并将结果截图，命名为 T3-9.jpg，保存到考生目录 T3 下。（5 分）
- 查询所有供应商名称包含“科技”的供应商编号、供应商名称，并将结果截图，命

名为 T3-10.jpg，保存到考生目录 T3 下。（5 分）

### 良好的习惯和职业素养（10 分）

- 答题过程中保持安静，有问题举手示意，有礼貌。（3 分）
- 数据库名命名规范，数据表名命名规范，字段名命名规范。（3 分）
- 答题完毕离场时整理好工位。（4 分）

### （2）实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。

计算机安装了 Windows 操作系统，MySQL 数据库软件以及 Office 办公软件。

### （3）考核时量及要求

- 考核时长 3 小时；
- 创建答题文件夹创建以“学号\_题号”命名的文件夹，存放所有答题文件；
- 创建答题文件，包括：SQL 脚本文件，如 J3-1.sql

### （4）评分细则

表 7-2 考核评价细则表

| 评价项            | 分值   | 评分细则                                                 |
|----------------|------|------------------------------------------------------|
| 数据库创建          | 10 分 | 创建考生目录 T1，文件 T1-1.sql、T1-1.jpg，得 1.5 分               |
|                |      | 使用语句创建了数据库，得 3 分                                     |
|                |      | 使用语句显示已经创建的数据库的名字，得 3 分                              |
|                |      | 使用新创建的数据库，得 2.5 分                                    |
| 创建数据表及表间的关系及约束 | 30 分 | 创建考生目录 T2，文件 T2-1.sql、T2-1.jpg、T2-2.jpg，得 2 分        |
|                |      | 表 T_suppliers 创建的字段名、字段类型、主键正确，并正确显示表结构及主键约束，得 8 分   |
|                |      | 表 T_goods 创建的字段名、字段类型、主键正确、外键正确，并正确显示表结构及主键约束，得 10 分 |
|                |      | 表 T_orders 创建的字段名、字段类型、主键正确，并正确显示表结构及主键和外键约束，得 10 分  |
| 数据操作           | 50 分 | 表 T_suppliers 和 T_goods 和 T_orders 正确插入数据，得 5 分      |
|                |      | 将数据表中商品名为“宏基笔记本”的价格上调 5%，使用查询语句显示修改后的数据，得 5 分        |
|                |      | 查询出商品编号为“1001”的总的订购数量，得 5 分                          |
|                |      | 查询供应商为“联想科技”的所有订单的商品名称、商品价格，得 10 分                   |
|                |      | 查询商品信息表（T_goods）中的数据，并按价格降                           |

|            |                       |     |                                                          |
|------------|-----------------------|-----|----------------------------------------------------------|
|            |                       |     | 序排，得 5 分                                                 |
|            |                       |     | 查询商品信息表（T_goods）中商品的平均价格，得 10 分                          |
|            |                       |     | 订单信息表中增加一个字段“备注”（字段名：Reserve、字段类型：text）。修改完成后显示表结构，得 5 分 |
|            |                       |     | 查询所有供应商名称包含“科技”的供应商编号、供应商名称，得 5 分                        |
| 良好的习惯和职业素养 | 代码书写格式规范              | 4 分 | 脚本缩进、换行规范，得 4 分                                          |
|            | 答题过程中保持安静，有问题举手示意，有礼貌 | 3 分 | 遵守考场纪律、有礼貌，得 3 分                                         |
|            | 答题完毕离场时整理好工位          | 3 分 | 鼠标、键盘、凳子等摆放整齐，得 3 分                                      |

## 8.试题编号：J1-4-2：“大型超市库存管理模块”数据库设计

### （1）任务描述

最初的超市资料管理，都是靠人力来完成的。但近几年我国超市经营规模日趋扩大，销售额和门店数量大幅度增加，而且许多超市正在突破以食品为主的传统格局，向品种多样化发展。小型超市在业务上需要处理大量的库存信息，还要时刻更新产品的销售信息，不断添加商品信息，并对商品各种信息进行统计分析。《库存管理》模块的 E-R 图如图 8-1 所示，物理数据模型如图 8-2 所示，数据表字段名定义见表 8-1。请按以下设计完成数据库创建、数据表创建和数据操作任务：

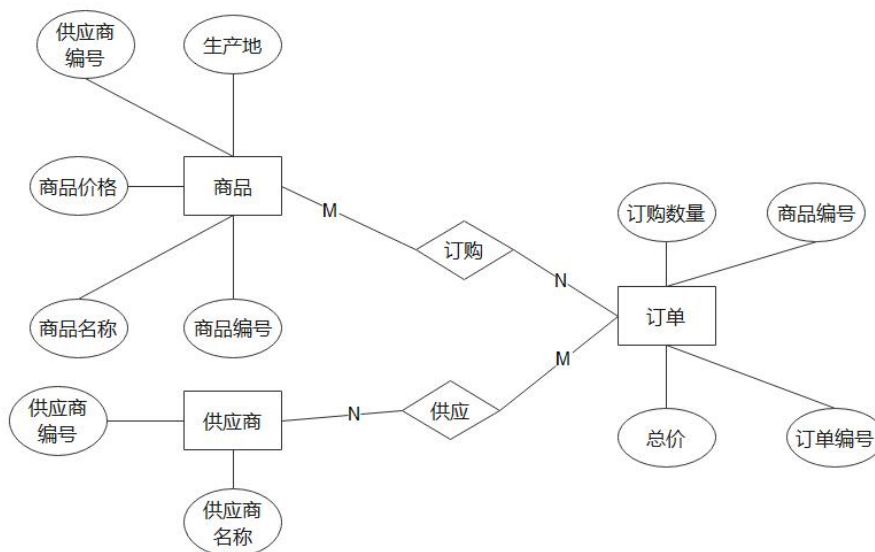


图8-1 E-R图

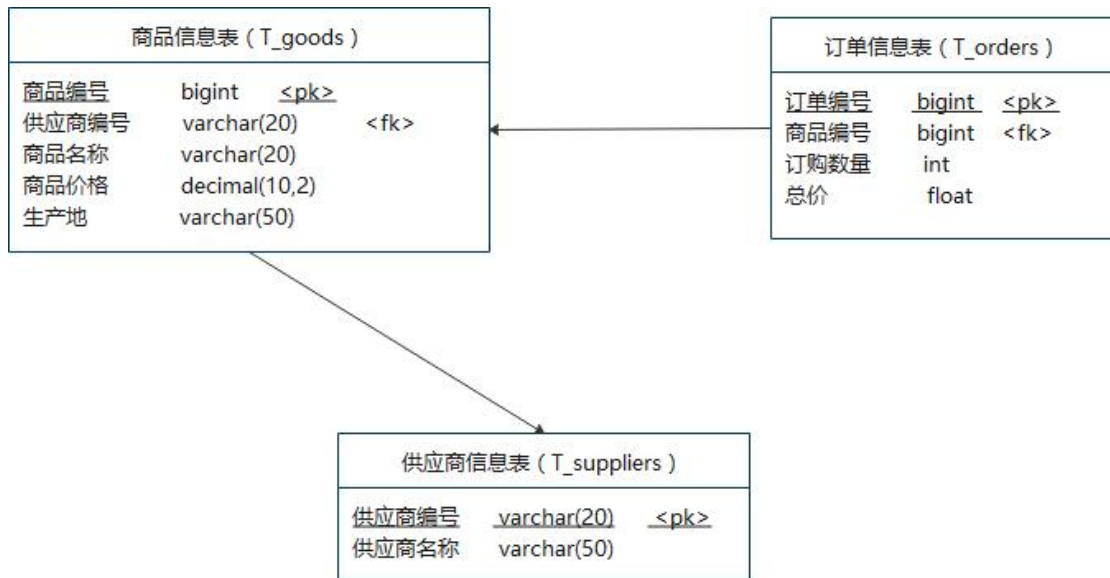


图8-2物理数据模型

表8-1字段名定义表

| 字段名           | 字段说明 | 字段名         | 字段说明  |
|---------------|------|-------------|-------|
| Goods_id      | 商品编号 | Quantity    | 订购数量  |
| Goods_name    | 商品名称 | Total_price | 总价    |
| Goods_price   | 商品价格 | Supp_code   | 供应商编号 |
| Goods_address | 生产地  | Supp_name   | 供应商名称 |
| Order_id      | 订单编号 |             |       |

### 任务一：创建数据库（10 分）

创建数据库 StoresDB。新建 T1-1.sql 文件，在该文件中编写语句创建指定的数据库，使用语句显示已经创建的数据库，并使用语句使用 StoresDB 数据库。将上述操作截图，命名为 T1-1.jpg 文件，将 T1-1.sql、T1-1.jpg 保存到考生目录 T1 下。

### 任务二：创建数据表及表间的关系及约束（30 分）

创建 T2-1.sql 文件并完成以下功能：

- 根据图 8-1、8-2 和表 8-1，创建数据表 T\_goods、T\_orders、T\_suppliers。（15 分）
- 根据物理数据原型，创建数据表之间的关系。（15 分）

将创建表 T\_goods、T\_orders、T\_suppliers 的语句以及执行结果截图，命名分别为 T2-1.jpg、T2-2.jpg、T2-3.jpg，将 T2-1.sql、T2-1.jpg、T2-2.jpg、T2-3.jpg 四个文件保存到考生目录 T2 下。

### 任务三：数据操作（40 分）

考生目录 T3 下创建 T3-1.sql 文件并用 SQL 语句执行以下操作：

- 在每个表中插入以下测试数据；

向 T\_suppliers 表插入数据：

| 供应商编号   | 供应商名称 |
|---------|-------|
| s430101 | 联想科技  |
| s430102 | 盛基科技  |
| s430103 | 惠普科技  |

向 T\_goods 表插入数据：

| 商品编号  | 供应商编号   | 商品名称  | 商品价格 | 生产地 |
|-------|---------|-------|------|-----|
| 11000 | s430101 | 联想笔记本 | 6600 | 广东  |
| 11001 | s430102 | 盛基笔记本 | 7200 | 香港  |
| 11002 | s430103 | 惠普笔记本 | 8600 | 广东  |
| 11003 | s430103 | 惠普打印机 | 1200 | 广东  |

向 T\_orders 表插入数据：

| 订单编号      | 商品编号  | 订购数量 | 总价     |
|-----------|-------|------|--------|
| 211070232 | 11000 | 20   | 132000 |
| 211060343 | 11001 | 101  | 727200 |
| 211050322 | 11002 | 18   | 154800 |
| 211040322 | 11001 | 20   | 144000 |
| 211020322 | 11002 | 39   | 335400 |

查询 T\_goods、T\_orders、T\_suppliers 三个表中的数据并截图，命名分别为 T3-1.jpg、T3-2.jpg、T3-3.jpg，保存到考生目录 T3 下。（5 分）

- 将商品名为“惠普笔记本”的价格下调 5%，数据修改后进行查询并将结果截图，命名为 T3-4.jpg，保存到考生目录 T3 下。（5 分）
- 查询所有商品名称包含“笔记本”的商品编号、商品名称、商品价格，并将结果截图，命名为 T3-5.jpg，保存到考生目录 T3 下。（5 分）
- 查询订购数量大于 100 的商品名称，并将结果截图，命名为 T3-6.jpg，保存到考生目录 T3 下。（10 分）
- 查询出订购了供应商编号为“s430101”的生产地，并将结果截图，命名为 T3-7.jpg，保存到考生目录 T3 下。（10 分）
- 创建视图 ProjInfo\_view1，查询所有商品的订单信息包括：商品名称、订购数量，使用语句显示视图 ProjInfo\_view1 中的数据并截图，命名为 T3-8.jpg，保存到考生目录 T3 下。（10 分）
- 删除上题中所创建的 ProjInfo\_view1 视图，使用 show tables 语句查看所有数据库

表，验证视图 ProjInfo\_view1 已经删除并将结果截图，命名为 T3-9.jpg，保存到考生目录 T3 下。（5 分）

### 良好的习惯和职业素养（10 分）

- 答题过程中保持安静，有问题举手示意，有礼貌。（3 分）
- 数据库名命名规范，数据表名命名规范，字段名命名规范。（3 分）
- 答题完毕离场时整理好工位。（4 分）

### （2）实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。

计算机安装了 Windows 操作系统，MySQL 数据库软件以及 Office 办公软件。

### （3）考核时量及要求

- 考核时长 3 小时；
- 创建答题文件夹创建以“学号\_题号”命名的文件夹，存放所有答题文件；
- 创建答题文件，包括：SQL 脚本文件，如 J3-1.sql

### （4）评分细则

表8-2 考核评价细则表

| 评价项            | 分值   | 评分细则                                                 |
|----------------|------|------------------------------------------------------|
| 数据库创建          | 10 分 | 创建考生目录 T1，文件 T1-1.sql、T1-1.jpg，得 1.5 分               |
|                |      | 使用语句创建了数据库，得 3 分                                     |
|                |      | 使用语句显示已经创建的数据库的名字，得 3 分                              |
|                |      | 使用新创建的数据库，得 2.5 分                                    |
| 创建数据表及表间的关系及约束 | 30 分 | 创建考生目录 T2，文件 T2-1.sql、T2-1.jpg、T2-2.jpg，得 2 分        |
|                |      | 表 T_suppliers 创建的字段名、字段类型、主键正确，并正确显示表结构及主键约束，得 8 分   |
|                |      | 表 T_goods 创建的字段名、字段类型、主键正确、外键正确，并正确显示表结构及主键约束，得 10 分 |
|                |      | 表 T_orders 创建的字段名、字段类型、主键正确，并正确显示表结构及主键和外键约束，得 10 分  |
| 数据操作           | 50 分 | 表 T_suppliers 和 T_goods 和 T_orders 正确插入数据，得 5 分      |
|                |      | 将商品名为“惠普笔记本”的价格下调 5%，数据修改后进行查询，得 5 分                 |
|                |      | 查询所有商品名称包含“笔记本”的商品编号、商品名称、商品价格，得 5 分                 |
|                |      | 查询订购数量大于 100 的商品名称，得 10 分                            |

|            |                       |     |                                                                                      |
|------------|-----------------------|-----|--------------------------------------------------------------------------------------|
|            |                       |     | 查询出订购了供应商编号为“s430101”的生产地，得 10 分                                                     |
|            |                       |     | 创建视图 ProjInfo_view1，查询所有商品的订单信息包括：商品名称、订购数量，使用语句显示视图 ProjInfo_view1 中的数据，得 10 分      |
|            |                       |     | 删除上题中所创建的 ProjInfo_view1 视图，使用 show tables 语句查看所有数据库表，验证视图 ProjInfo_view1 已经删除，得 5 分 |
| 良好的习惯和职业素养 | 代码书写格式规范              | 4 分 | 脚本缩进、换行规范，得 4 分                                                                      |
|            | 答题过程中保持安静，有问题举手示意，有礼貌 | 3 分 | 遵守考场纪律、有礼貌，得 3 分                                                                     |
|            | 答题完毕离场时整理好工位          | 3 分 | 鼠标、键盘、凳子等摆放整齐，得 3 分                                                                  |

## 项目 5：考试机构系统数据库设计

### 9.试题编号：J1-5-1：“考试机构报名管理模块”数据库设计

#### (1) 任务描述

《报名管理》模块的 E-R 图如图 9-1 所示，物理数据模型如图 9-2 所示，数据表字段名定义见表 9-1。请按以下设计完成数据库创建、数据表创建和数据操作任务：

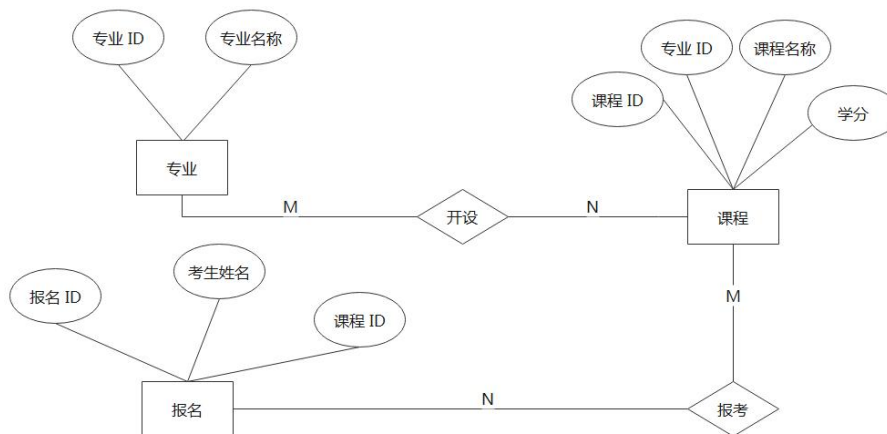


图 9-1 E-R 图

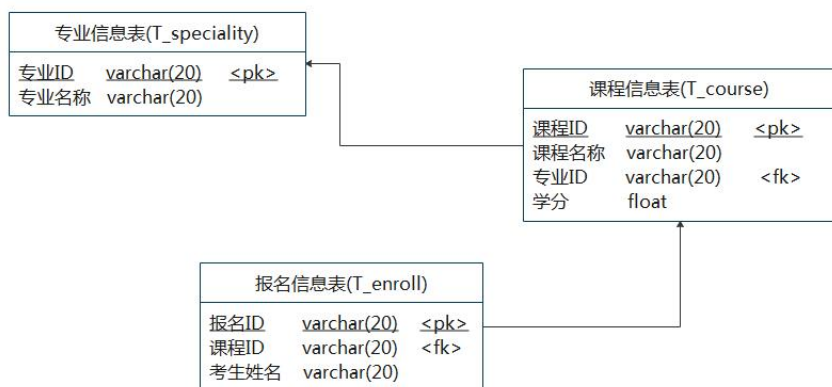


图 9-2 物理数据原型

表 9-1 字段名定义表

| 字段名             | 字段说明  | 字段名         | 字段说明  |
|-----------------|-------|-------------|-------|
| speciality_id   | 专业 ID | course_mark | 课程学分  |
| speciality_name | 专业名称  | enroll_id   | 报名 ID |
| course_id       | 课程 ID | enroll_name | 考生姓名  |
| course_name     | 课程名称  |             |       |

### 任务一：创建数据库（10 分）

创建数据库 CourseExamDB。新建 T1-1.sql 文件，在该文件中编写语句创建指定的数据库，使用语句显示已经创建的数据库，并使用语句使用 CourseExamDB 数据库。将上述操作截图，命名为 T1-1.jpg 文件，将 T1-1.sql、T1-1.jpg 保存到考生目录 T1 下。

### 任务二：创建数据表及表间的关系及约束（30 分）

创建 T2-1.sql 文件并完成以下功能：

- 根据图 9-1、9-2 和表 9-1，创建数据表 T\_speciality、T\_course、T\_enroll。（15 分）
- 根据物理数据原型，创建数据关系。（15 分）

将创建表 T\_speciality、T\_course、T\_enroll 的语句以及执行结果截图，命名分别为 T2-1.jpg、T2-2.jpg、T2-3.jpg，将 T2-1.sql、T2-1.jpg、T2-2.jpg、T2-3.jpg 四个文件保存到考生目录 T2 下。

### 任务三：数据操作（50 分）

考生目录 T3 下创建 T3-1.sql 文件并用 SQL 语句执行以下操作：

- 在每个表中插入以下测试数据；

T\_speciality 表测试数据

| speciality_id | name     |
|---------------|----------|
| s001          | 大数据技术与应用 |
| s002          | 数字媒体     |
| s003          | 艺术设计     |

T\_course表测试数据

| course_id | name    | speciality_id | mark |
|-----------|---------|---------------|------|
| c001      | 网络操作系统  | s001          | 4    |
| c002      | 数据库应用技术 | s001          | 4    |
| c003      | 数据处理技术  | s001          | 4    |
| c005      | 界面板式与色彩 | s003          | 4    |
| c006      | 图形图像处理  | s003          | 4    |

T\_enroll表测试数据

| enroll_id | course_id | name |
|-----------|-----------|------|
| e001      | c001      | 刘长   |
| e002      | c002      | 王梅   |
| e003      | c003      | 李磊   |
| e004      | c001      | 刘文武  |
| e005      | c002      | 陈诚   |
| e006      | c003      | 赵雅芝  |
| e007      | c005      | 彭志向  |
| e008      | c006      | 杨扬   |

查询 T\_speciality、T\_course、T\_enroll 三个表中的数据并截图，命名分别为 T3-1.jpg、

T3-2.jpg、T3-3.jpg，保存到考生目录 T3 下。（5 分）

- 查询 T\_course 表中学分为 4 的课程的名字，并将结果截图，命名分别为 T3-4.jpg，保存到考生目录 T3 下。（5 分）
- 在表 T\_enroll 中增加一个字段“备注”（字段名：remark、字段类型：varchar(20)”)放在现有字段的最后,增加字段后显示表结构并截图，命名分别为 T3-5.jpg，保存到考生目录 T3 下。（5 分）
- 查询“大数据技术与应用”专业开设的课程名称，命名分别为 T3-6.jpg，保存到考生目录 T3 下。（5 分）
- 查询报考“大数据技术与应用”专业的学生姓名，命名分别为 T3-7.jpg，保存到考生目录 T3 下。（10 分）
- 查询出报考课程号为“c003”的考生的名字，命名分别为 T3-8.jpg，保存到考生目录 T3 下。（5 分）
- 创建视图 ProjInfo\_view1，查询报名 ID 为“e001”的考生姓名，使用语句显示视图 ProjInfo\_view1 中的数据并截图，命名为 T3-9.jpg，保存到考生目录 T3 下。（10 分）
- 删除上题中创建的视图 ProjInfo\_view1，使用 show tables 语句查看所有数据库表，验证视图 ProjInfo\_view1 已经删除并将结果截图，命名为 T3-10.jpg，保存到考生目

录 T3 下。(5 分)

### 良好的习惯和职业素养 (10 分)

- 答题过程中保持安静，有问题举手示意，有礼貌。(3 分)
- 数据库名命名规范，数据表名命名规范，字段名命名规范。(3 分)
- 答题完毕离场时整理好工位。(4 分)

### (2) 实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。

计算机安装了 Windows 操作系统，MySQL 数据库软件以及 Office 办公软件。

### (3) 考核时量及要求

- 考核时长 3 小时；
- 创建答题文件夹创建以“学号\_题号”命名的文件夹，存放所有答题文件；
- 创建答题文件，包括：SQL 脚本文件，如 J3-1.sql

### (4) 评分细则

表9-2 考核评价细则表

| 评价项            | 分值   | 评分细则                                                                           |
|----------------|------|--------------------------------------------------------------------------------|
| 数据库创建          | 10 分 | 创建考生目录 T1，文件 T1-1.sql、T1-1.jpg，得 1.5 分                                         |
|                |      | 使用语句创建了数据库，得 3 分                                                               |
|                |      | 使用语句显示已经创建的数据库的名字，得 3 分                                                        |
|                |      | 使用新创建的数据库，得 2.5 分                                                              |
| 创建数据表及表间的关系及约束 | 30 分 | 创建考生目录 T2，文件 T2-1.sql、T2-1.jpg、T2-2.jpg，得 2 分                                  |
|                |      | 表 T_speciality 创建的字段名、字段类型、主键正确，并正确显示表结构及主键约束，得 8 分                            |
|                |      | 表 T_course 创建的字段名、字段类型、主键正确、外键正确，并正确显示表结构及主键约束，得 10 分                          |
|                |      | 表 T_enroll 创建的字段名、字段类型、主键正确，并正确显示表结构及主键和外键约束，得 10 分                            |
| 数据操作           | 50 分 | 表 T_speciality 和 T_course 和 T_enroll 正确插入数据，得 5 分                              |
|                |      | 查询 T_course 表中学分为 4 的课程的名字，得 5 分                                               |
|                |      | 在表 T_enroll 中增加一个字段“备注”(字段名：remark、字段类型：varchar(20))放在现有字段的最后，增加字段后显示表结构，得 5 分 |
|                |      | 查询“大数据技术与应用”专业开设的课程名称，得 5 分                                                    |

|            |                       |     |                                                                                    |
|------------|-----------------------|-----|------------------------------------------------------------------------------------|
|            |                       |     | 查询报考“大数据技术与应用”专业的学生姓名，得 10 分                                                       |
|            |                       |     | 查询出报考课程号为“c003”的考生的名字，得 5 分                                                        |
|            |                       |     | 创建视图 ProjInfo_view1，查询报名 ID 为“e001”的考生姓名，使用语句显示视图 ProjInfo_view1 中的数据，得 10 分       |
|            |                       |     | 删除上题中创建的视图 ProjInfo_view1，使用 show tables 语句查看所有数据库表，验证视图 ProjInfo_view1 已经删除，得 5 分 |
| 良好的习惯和职业素养 | 代码书写格式规范              | 4 分 | 脚本缩进、换行规范，得 4 分                                                                    |
|            | 答题过程中保持安静，有问题举手示意，有礼貌 | 3 分 | 遵守考场纪律、有礼貌，得 3 分                                                                   |
|            | 答题完毕离场时整理好工位          | 3 分 | 鼠标、键盘、凳子等摆放整齐，得 3 分                                                                |

## 10.试题编号：J1-5-2：“教材管理模块”数据库设计

### (1) 任务描述

《教材管理》模块的 E-R 图如图 10-1 所示，物理数据模型如图 10-2 所示，数据表字段名定义见表 10-1。请按以下设计完成数据库创建、数据表创建和数据操作任务：

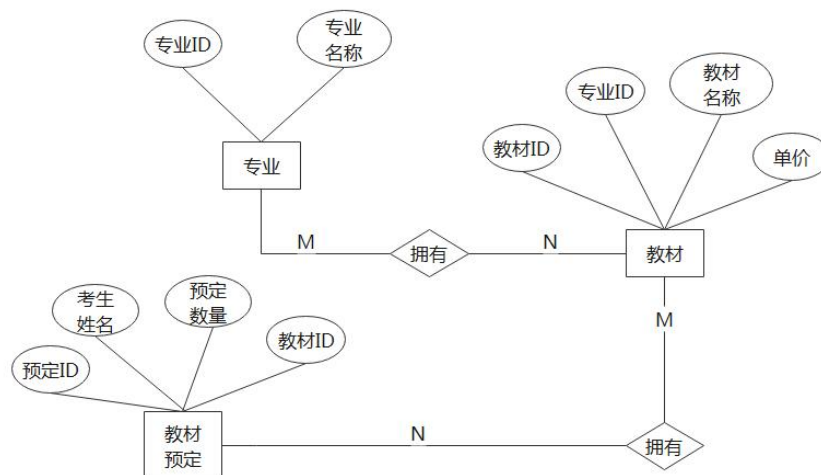


图 10-1 E-R 图

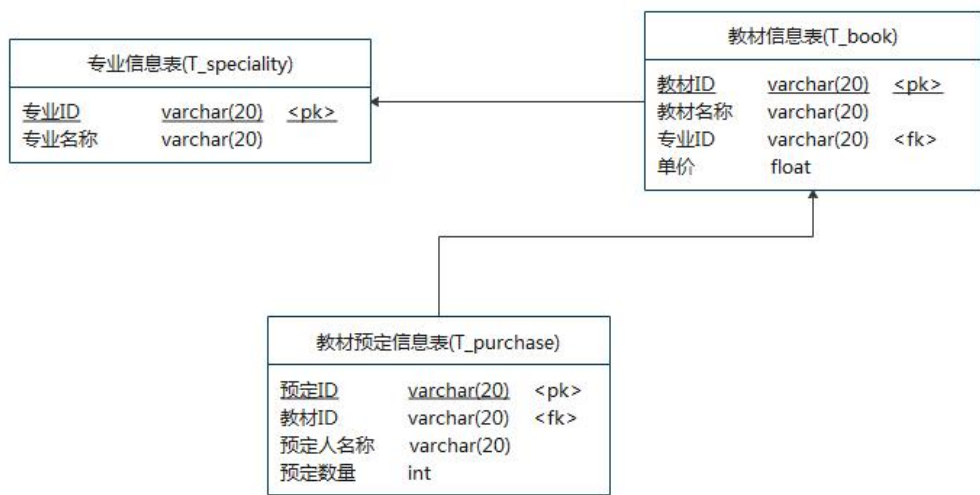


图 10-2 物理数据原型

表 10-1 字段名定义表

| 字段名             | 字段说明  | 字段名               | 字段说明  |
|-----------------|-------|-------------------|-------|
| speciality_id   | 专业 ID | book_price        | 单价    |
| speciality_name | 专业名称  | purchase_id       | 预订 ID |
| book_id         | 教材 ID | purchase_name     | 预订人名称 |
| book_name       | 教材名称  | purchase_quantity | 预订数量  |

### 任务一：创建数据库（10 分）

创建数据库 BooksDB。新建 T1-1.sql 文件，在该文件中编写语句创建指定的数据库，使用语句显示已经创建的数据库，并使用语句使用 BooksDB 数据库。将上述操作截图，命名为 T1-1.jpg 文件，将 T1-1.sql、T1-1.jpg 保存到考生目录 T1 下。

### 任务二：创建数据表及表间的关系及约束（30 分）

创建 T2-1.sql 文件并完成以下功能：

- 根据图 10-1、10-2 和表 10-1，创建数据表 T\_speciality、T\_book、T\_purchase。（15 分）
- 根据物理数据原型，创建数据关系。（15 分）

将创建表 T\_speciality、T\_book、T\_purchase 的语句以及执行结果截图，命名分别为 T2-1.jpg、T2-2.jpg、T2-3.jpg，将 T2-1.sql、T2-1.jpg、T2-2.jpg、T2-3.jpg 四个文件保存到考生目录 T2 下。

### 任务三：数据操作（30 分）

考生目录 T3 下创建 T3-1.sql 文件并用 SQL 语句执行以下操作：

- 在每个表中插入以下测试数据；  
T\_speciality 表测试数据

| 专业 ID | 专业名称     |
|-------|----------|
| s001  | 大数据技术与应用 |
| s002  | 数字媒体     |
| s003  | 艺术设计     |
| s004  | 计算机软件    |
| s005  | 移动应用开发   |
| s006  | 计算机技术    |

T\_book表测试数据

| 教材 ID | 教材名称         | 专业 ID | 单价   |
|-------|--------------|-------|------|
| b001  | Linux 网络操作系统 | s001  | 52.8 |
| b002  | mysql 数据库    | s001  | 45.5 |
| b003  | Hadoop 平台搭建  | s001  | 39.5 |
| b004  | 大数据概论        | s001  | 50.8 |
| b005  | 数据库系统概论      | s001  | 45.5 |
| b006  | 大数据开发基础      | s001  | 39.5 |
| b007  | 数字媒体技术基础     | s002  | 50.8 |
| b008  | 数字三维雕塑       | s002  | 45.5 |
| b009  | 三维可视化制作      | s002  | 39.5 |
| b010  | 二维动画制作       | s002  | 50.8 |
| b011  | 计算机应用基础      | s003  | 41.5 |
| b012  | 微信小程序开发      | s004  | 30.5 |

T\_purchase表测试数据

| 预订 ID | 教材 ID | 预订人名称 | 预订数量 |
|-------|-------|-------|------|
| p001  | b001  | 李四    | 2    |
| p002  | b002  | 李四    | 1    |
| p003  | b003  | 李四    | 2    |
| p004  | b001  | 王五    | 2    |
| p005  | b002  | 王五    | 1    |
| p006  | b003  | 王五    | 2    |

查询 T\_speciality、T\_book、T\_purchase 三个表中的数据并截图，命名分别为 T3-1.jpg、T3-2.jpg、T3-3.jpg，保存到考生目录 T3 下。（5 分）

- 查询教材名称为“大数据开发基础”的价格，并将结果截图，命名为 T3-4.jpg，保存到考生目录 T3 下。（5 分）
- 查询出属于“大数据技术与应用”专业的教材名字，并将结果截图，命名为 T3-5.jpg，保存到考生目录 T3 下。（5 分）
- 查询出教材信息表中价格最高的图书的名字，并将结果截图，命名为 T3-6.jpg，保存到考生目录 T3 下。（10 分）

- 查询出预订教材“mysql 数据库”的预订人名称，并将结果截图，命名为 T3-7.jpg，保存到考生目录 T3 下。（5 分）
- 创建视图 ProjInfo\_view1 查询预定人姓名为李四的预订教材 ID、预订数量，使用语句显示视图 ProjInfo\_view1 中的数据并截图，命名为 T3-8.jpg，保存到考生目录 T3 下。（5 分）
- 删除上题中创建的视图，使用 show tables 语句查看所有数据库表，验证视图 ProjInfo\_view1 已经删除并将结果截图，命名为 T3-9.jpg，保存到考生目录 T3 下。（5 分）
- 删除 T\_purchase 表中教材 ID 为“b001”的教材预订信息，查询删除后的表的数据并将结果截图，命名为 T3-10.jpg，保存到考生目录 T3 下。（10 分）

#### 良好的习惯和职业素养（10 分）

- 答题过程中保持安静，有问题举手示意，有礼貌。（3 分）
- 数据库名命名规范，数据表名命名规范，字段名命名规范。（3 分）
- 答题完毕离场时整理好工位。（4 分）

### （2）实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。计算机安装了 Windows 操作系统，MySQL 数据库软件以及 Office 办公软件。

### （3）考核时量及要求

- 考核时长 3 小时；
- 创建答题文件夹创建以“学号\_题号”命名的文件夹，存放所有答题文件；
- 创建答题文件，包括：SQL 脚本文件，如 J3-1.sql

### （4）评分细则

表10-2 考核评价细则表

| 评价项        | 分值   | 评分细则                                          |
|------------|------|-----------------------------------------------|
| 数据库创建      | 10 分 | 创建考生目录 T1，文件 T1-1.sql、T1-1.jpg，得 1.5 分        |
|            |      | 使用语句创建了数据库，得 3 分                              |
|            |      | 使用语句显示已经创建的数据库的名字，得 3 分                       |
|            |      | 使用新创建的数据库，得 2.5 分                             |
| 创建数据表及表间的关 | 30 分 | 创建考生目录 T2，文件 T2-1.sql、T2-1.jpg、T2-2.jpg，得 2 分 |
|            |      | 表 T_speciality 创建的字段名、字段类型、主键正确，并             |

|            |                       |      |                                                                                |
|------------|-----------------------|------|--------------------------------------------------------------------------------|
| 系及约束       |                       |      | 正确显示表结构及主键约束，得 8 分                                                             |
|            |                       |      | 表 T_book 创建的字段名、字段类型、主键正确、外键正确，并正确显示表结构及主键约束，得 10 分                            |
|            |                       |      | 表 T_purchase 创建的字段名、字段类型、主键正确，并正确显示表结构及主键和外键约束，得 10 分                          |
| 数据操作       |                       | 50 分 | 表 T_speciality 和 T_book 和 T_purchase 正确插入数据，得 5 分                              |
|            |                       |      | 查询教材名称为“大数据开发基础”的价格，得 5 分                                                      |
|            |                       |      | 查询出属于“大数据技术与应用”专业的教材名字，得 5 分                                                   |
|            |                       |      | 查询出教材信息表中价格最高的图书的名字，得 10 分                                                     |
|            |                       |      | 查询出预订教材“mysql 数据库”的预订人名称，得 5 分                                                 |
|            |                       |      | 创建视图 ProjInfo_view1 查询预定人姓名为李四的预订教材 ID、预订数量，使用语句显示视图 ProjInfo_view1 中的数据，得 5 分 |
|            |                       |      | 删除上题中创建的视图，使用 show tables 语句查看所有数据库表，验证视图 ProjInfo_view1 已经删除，得 5 分            |
|            |                       |      | 删除 T_purchase 表中教材 ID 为“b001”的教材预订信息，查询删除后的表的数据，得 10 分                         |
| 良好的习惯和职业素养 | 代码书写格式规范              | 4 分  | 脚本缩进、换行规范，得 4 分                                                                |
|            | 答题过程中保持安静，有问题举手示意，有礼貌 | 3 分  | 遵守考场纪律、有礼貌，得 3 分                                                               |
|            | 答题完毕离场时整理好工位          | 3 分  | 鼠标、键盘、凳子等摆放整齐，得 3 分                                                            |

## 项目 6：人力资源系统数据库设计

### 11. 试题编号：J1-6-1：“人力资源系统项目管理模块”数据库设计

#### (1) 任务描述

《项目管理》模块的 E-R 图如图 11-1 所示，物理数据模型如图 11-2 所示，数据表字段名定义见表 11-1。请按以下设计完成数据库创建、数据表创建和数据操作任务：

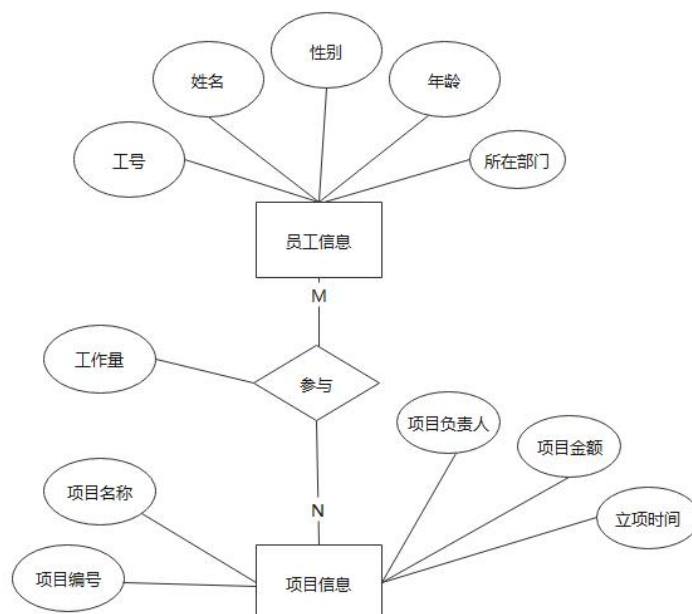


表11-1 字段名定义表

### 任务一：创建数据库（10 分）

T1-1.jpg 文件，将 T1-1.sql、T1-1.jpg 保存到考生目录 T1 下。

### 任务二：创建数据表及表间的关系及约束（30 分）

创建 T2-1.sql 文件并完成以下功能：

- 根据图 11-1、11-2、和表 11-1，创建数据表 T\_staff\_info、T\_project\_info、T\_staff\_workload。（15 分）
- 根据物理数据原型，创建数据关系表。（15 分）

将创建表 T\_staff\_info、T\_project\_info、T\_staff\_workload 的语句以及执行结果截图，命名分别为 T2-1.jpg、T2-2.jpg、T2-3.jpg，将 T2-1.sql、T2-1.jpg、T2-2.jpg、T2-3.jpg 四个文件保存到考生目录 T2 下。

### 任务三：数据操作（50 分）

考生目录 T3 下创建 T3-1.sql 文件并用 SQL 语句执行以下操作：

- 在每个表中插入以下测试数据；

向 T\_staff\_info 表插入数据：

| 工号         | 姓名  | 性别 | 年龄 | 所在部门 |
|------------|-----|----|----|------|
| 2016010002 | 李四  | 男  | 35 | 销售部  |
| 2016010003 | 刘琳  | 女  | 26 | 销售部  |
| 2016010001 | 谢云  | 男  | 30 | 研发部  |
| 2016010004 | 王占非 | 女  | 29 | 研发部  |

向 T\_project\_info 表插入数据：

| 项目编号     | 项目名称       | 项目负责人 | 项目金额  | 立项时间       |
|----------|------------|-------|-------|------------|
| 20200002 | 天河中心小学门户网站 | 张某    | 26000 | 2020-10-03 |
| 20200003 | web 网站开发   | 刘某    | 6000  | 2020-05-23 |
| 20200001 | 超市库存管理系统   | 萧峰    | 56120 | 2021-05-03 |
| 20200004 | 学校网站开发     | 王琳    | 8900  | 2019-05-20 |
| 20200005 | 招聘网站开发     | 谢志强   | 28000 | 2019-12-30 |

向 T\_staff\_workload 表插入数据:

| 项目编号     | 工号         | 工作内容 | 工作量 |
|----------|------------|------|-----|
| 20200002 | 2016010001 | 前端开发 | 5   |
| 20200003 | 2016010004 | 软件测试 | 3   |
| 20200001 | 2016010002 | 系统开发 | 9   |
| 20200004 | 2016010003 | 网站推广 | 12  |

查询 T\_staff\_info、T\_project\_info、T\_staff\_workload 三个表中的数据并截图，命名分别为 T3-1.jpg、T3-2.jpg、T3-3.jpg，保存到考生目录 T3 下。(5 分)

- 查询出项目金额小于 10000 的所有项目名称，并将结果截图，命名为 T3-4.jpg，保存到考生目录 T3 下。(5 分)
- 查询出员工“刘琳”参与的所有项目名称，并将结果截图，命名为 T3-5.jpg，保存到考生目录 T3 下。(10 分)
- 查询所有女员工的姓名，并将结果截图，命名为 T3-6.jpg，保存到考生目录 T3 下。(5 分)
- 查询 2020 年立项的项目的名称，并将结果截图，命名为 T3-7.jpg，保存到考生目录 T3 下。(5 分)
- 查询出参与项目编号为“20200002”项目的所有员工姓名，并将结果截图，命名为 T3-8.jpg，保存到考生目录 T3 下。(10 分)
- 创建视图 ProjInfo\_view1，查询项目的项目名称，立项时间，使用语句显示视图 ProjInfo\_view1 的结构并截图，命名为 T3-9.jpg，保存到考生目录 T3 下。(10 分)

良好的习惯和职业素养 (10 分)

- 答题过程中保持安静，有问题举手示意，有礼貌。(3 分)
- 数据库名命名规范，数据表名命名规范，字段名命名规范。(3 分)
- 答题完毕离场时整理好工位。(4 分)

## (2) 实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。

计算机安装了 Windows 操作系统，MySQL 数据库软件以及 Office 办公软件。

### (3) 考核时量及要求

- 考核时长 3 小时；
- 创建答题文件夹创建以“学号\_题号”命名的文件夹，存放所有答题文件；
- 创建答题文件，包括：SQL 脚本文件，如 J3-1.sql

### (4) 评分细则

表 11-2 考核评价细则表

| 评价项            |                       | 分值   | 评分细则                                                                  |
|----------------|-----------------------|------|-----------------------------------------------------------------------|
| 数据库创建          |                       | 10 分 | 创建考生目录 T1，文件 T1-1.sql、T1-1.jpg，得 1.5 分                                |
|                |                       |      | 使用语句创建了数据库，得 3 分                                                      |
|                |                       |      | 使用语句显示已经创建的数据库的名字，得 3 分                                               |
|                |                       |      | 使用新创建的数据库，得 2.5 分                                                     |
| 创建数据表及表间的关系及约束 |                       | 30 分 | 创建考生目录 T2，文件 T2-1.sql、T2-1.jpg、T2-2.jpg，得 2 分                         |
|                |                       |      | 表 T_project_info 创建的字段名、字段类型、主键正确，并正确显示表结构及主键约束，得 8 分                 |
|                |                       |      | 表 T_staff_info 创建的字段名、字段类型、主键正确，并正确显示表结构及主键约束，得 8 分                   |
|                |                       |      | 表 T_staff_workload 创建的字段名、字段类型、主键正确，并正确显示表结构及主键和外键约束，得 12 分           |
| 数据操作           |                       | 50 分 | 表 T_project_info 和 T_staff_info 和 T_staff_workload 正确插入数据，得 5 分       |
|                |                       |      | 查询出项目金额小于 10000 的所有项目名称，得 5 分                                         |
|                |                       |      | 查询出员工“刘琳”参与的所有项目名称，得 10 分                                             |
|                |                       |      | 查询所有女员工的姓名，得 5 分                                                      |
|                |                       |      | 查询 2020 年立项的项目的名称，得 5 分                                               |
|                |                       |      | 查询出参与项目编号为“20200002”项目的所有员工姓名，得 10 分                                  |
|                |                       |      | 创建视图 ProjInfo_view1，查询项目的项目名称，立项时间，使用语句显示视图 ProjInfo_view1 的结构，得 10 分 |
| 良好的习惯和职业素养     | 代码书写格式规范              | 4 分  | 脚本缩进、换行规范，得 4 分                                                       |
|                | 答题过程中保持安静，有问题举手示意，有礼貌 | 3 分  | 遵守考场纪律、有礼貌，得 3 分                                                      |
|                | 答题完毕离场时整理好工位          | 3 分  | 鼠标、键盘、凳子等摆放整齐，得 3 分                                                   |

12.试题编号：J1-6-2：“人力资源系统员工信息管理模块”数据库设计

(1) 任务描述

《员工信息管理模块》模块的 E-R 图如图 12-1 所示，物理数据模型如图 12-2 所示，数据表字段名定义见表 12-1。请按以下设计完成数据库创建、数据表创建和数据操作任务：

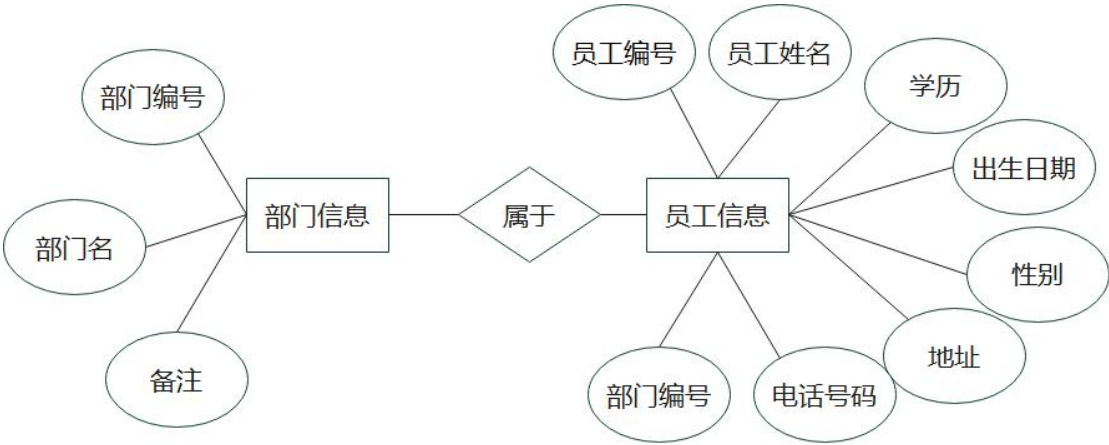


图 12-1 E-R 图

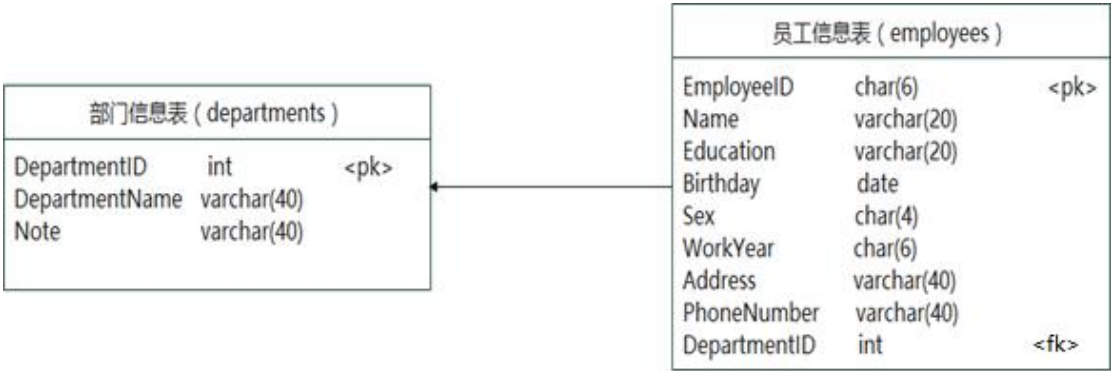


图 12-2 物理数据原型

表 12-1 字段名定义表

| 字段名            | 字段说明 | 字段名          | 字段说明 |
|----------------|------|--------------|------|
| DepartmentID   | 部门编号 | Birthday     | 出生日期 |
| DepartmentName | 部门名  | Sex          | 性别   |
| Note           | 备注   | WorkYear     | 工作年限 |
| EmployeeID     | 员工编号 | Address      | 地址   |
| Name           | 员工姓名 | PhoneNumber  | 电话号码 |
| Education      | 学历   | DepartmentID | 部门编号 |

### 任务一：创建数据库（10 分）

创建数据库 EmployDB。新建 T1-1.sql 文件，在该文件中编写语句创建指定的数据库，使用语句显示已经创建的数据库，并使用语句使用 EmployDB 数据库。将上述操作截图，命名为 T1-1.jpg 文件，将 T1-1.sql、T1-1.jpg 保存到考生目录 T1 下。

### 任务二：创建数据表及表间的关系及约束（30 分）

创建 T2-1.sql 文件并完成以下功能：

- 根据图 12-1、12-2 和表 12-1，创建数据表 departments、employees。（15 分）
- 根据物理数据原型，创建数据关系表。（15 分）

将创建表 departments、employees 的语句以及执行结果截图，命名分别为 T2-1.jpg、T2-2.jpg，将 T2-1.sql、T2-1.jpg、T2-2.jpg 三个文件保存到考生目录 T2 下。

### 任务三：数据操作（50 分）

考生目录 T3 下创建 T3-1.sql 文件并用 SQL 语句执行以下操作：

- 在每个表中插入以下测试数据；

向 departments 表插入数据：

| 部门编号 | 部门名   | 备注   |
|------|-------|------|
| 1    | 财务部   | null |
| 2    | 人力资源部 | null |
| 3    | 经理办公室 | null |
| 4    | 研发部   | null |
| 5    | 市场部   | null |

向 employees 表插入数据：

| 员工编号   | 员工姓名 | 学历 | 出生日期        | 性别 | 工作年限 | 地址           | 电话号码     | 部门编号 |
|--------|------|----|-------------|----|------|--------------|----------|------|
| 000001 | 王林   | 大专 | 1966-01-23  | 1  | 8    | 中山路 32-1-508 | 83355668 | 2    |
| 010008 | 伍容华  | 本科 | 1976-03-28  | 1  | 3    | 北京东路 100-2   | 83321321 | 1    |
| 020010 | 王向容  | 硕士 | 1982-12-09  | 1  | 2    | 四牌楼 10-0-108 | 83792361 | 1    |
| 020018 | 李丽   | 大专 | 1960-07-30  | 0  | 6    | 中山东路 102-2   | 83413301 | 1    |
| 102201 | 刘明   | 本科 | 1972-10-18  | 1  | 3    | 虎距路 100-2    | 83606608 | 5    |
| 102208 | 朱俊   | 硕士 | 1965-09-28  | 1  | 2    | 牌楼巷 5-3-106  | 84708817 | 5    |
| 108991 | 钟敏   | 硕士 | 1979-08-10  | 0  | 4    | 中山路 10-3-105 | 83346722 | 3    |
| 111006 | 张石兵  | 本科 | 1974-10-01  | 1  | 1    | 解放路 34-1-203 | 84563418 | 5    |
| 210678 | 林涛   | 大专 | 1977-04-02' | 1  | 2    | 中山北路 24-35   | 83467336 | 3    |
| 302566 | 李玉珉  | 本科 | 1968-09-20  | 1  | 3    | 热和路 209-3    | 58765991 | 4    |
| 308759 | 叶凡   | 本科 | 1978-11-18  | 1  | 2    | 北京西路 3-7-52  | 83308901 | 4    |

|        |     |    |            |   |   |               |          |   |
|--------|-----|----|------------|---|---|---------------|----------|---|
| 504209 | 陈林琳 | 大专 | 1969-09-03 | 0 | 5 | 汉中路 120-4-12' | 84468158 | 4 |
|--------|-----|----|------------|---|---|---------------|----------|---|

查询 departments、employees 两个表中的数据并截图，命名分别为 T3-1.jpg、T3-2.jpg，保存到考生目录 T3 下。（5 分）

- 公司新成立了一个销售部，部门代码为 6，备注中注明“筹建”，请将上述信息插入到表 departments 中并将结果截图，命名为 T3-3.jpg，保存到考生目录 T3 下。（5 分）
- 王林辞职离开公司请将 employees 表中王林的相关信息删除，删除数据后对表中的数据进行查询，并将结果截图，命名为 T3-4.jpg，保存到考生目录 T3 下。（5 分）
- 查询每个雇员的地址和电话，显示的列标题要求显示“地址”、“电话”，并将结果截图，命名为 T3-5.jpg，保存到考生目录 T3 下。（5 分）
- 计算每个部门雇员的人数，并将结果截图，命名为 T3-6.jpg，保存到考生目录 T3 下。（5 分）
- 查询财务部、研发部、市场部的员工的姓名、地址和电话信息，并将结果截图，命名为 T3-7.jpg，保存到考生目录 T3 下。（5 分）
- 查找员工编号中倒数第二个数字为 0 的员工的姓名、地址和学历，并将结果截图，命名为 T3-8.jpg，保存到考生目录 T3 下。（5 分）
- 创建视图 project\_view，视图的数据为工作年限最长的员工的工号，使用语句显示视图 project\_view 的结果并截图，命名为 T3-9.jpg，保存到考生目录 T3 下。（10 分）
- 删除上题中创建的视图，使用 show tables 语句查看所有数据库表，验证视图 project\_view 已经删除并将结果截图，命名为 T3-10.jpg，保存到考生目录 T3 下。（5 分）

#### 良好的习惯和职业素养（10 分）

- 答题过程中保持安静，有问题举手示意，有礼貌。（3 分）
- 数据库名命名规范，数据表名命名规范，字段名命名规范。（3 分）
- 答题完毕离场时整理好工位。（4 分）

## （2）实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。计算机安装了 Windows 操作系统，MySQL 数据库软件以及 Office 办公软件。

### (3) 考核时量及要求

- 考核时长 3 小时；
- 创建答题文件夹创建以“学号\_题号”命名的文件夹，存放所有答题文件；
- 创建答题文件，包括：SQL 脚本文件，如 J3-1.sql

### (4) 评分细则

表 12-2 考核评价细则表

| 评价项            |                 | 分值   | 评分细则                                                                  |
|----------------|-----------------|------|-----------------------------------------------------------------------|
| 数据库创建          |                 | 10 分 | 创建考生目录 T1，文件 T1-1.sql、T1-1.jpg，得 1.5 分                                |
|                |                 |      | 使用语句创建了数据库，得 3 分                                                      |
|                |                 |      | 使用语句显示已经创建的数据库的名字，得 3 分                                               |
|                |                 |      | 使用新创建的数据库，得 2.5 分                                                     |
| 创建数据表及表间的关系及约束 |                 | 30 分 | 创建考生目录 T2，文件 T2-1.sql、T2-1.jpg、T2-2.jpg，得 2 分                         |
|                |                 |      | 表 departments 创建的字段名、字段类型、主键正确，并正确显示表结构及主键约束，得 12 分                   |
|                |                 |      | 表 employees 创建的字段名、字段类型、主键正确，并正确显示表结构及主键和外键约束，得 16 分                  |
| 数据操作           |                 | 50 分 | 表 departments 和 employees 正确插入数据，得 5 分                                |
|                |                 |      | 公司新成立了一个销售部，部门代码为 6，备注中注明“筹建”，请将上述信息插入到表 departments 中，得 5 分          |
|                |                 |      | 王林辞职离开公司请将 employees 表中王林的相关信息删除，删除数据后对表中的数据进行查询，得 5 分                |
|                |                 |      | 查询每个雇员的地址和电话，显示的列标题要求显示“地址”、“电话”，得 5 分                                |
|                |                 |      | 计算每个部门雇员的人数，得 5 分                                                     |
|                |                 |      | 查询财务部、研发部、市场部的员工的姓名、地址和电话信息，得 5 分                                     |
|                |                 |      | 查找员工编号中倒数第二个数字为 0 的员工的姓名、地址和学历，得 5 分                                  |
|                |                 |      | 创建视图 project_view，视图的数据为工作年限最长的员工的工号，使用语句显示视图 project_view 的结果，得 10 分 |
|                |                 |      | 删除上题中创建的视图，使用 show tables 语句查看所有数据库表，验证视图 project_view 已经删除，得 5 分     |
| 良好的习惯和职业       | 代码书写格式规范        | 4 分  | 脚本缩进、换行规范，得 4 分                                                       |
|                | 答题过程中保持安静，有问题举手 | 3 分  | 遵守考场纪律、有礼貌，得 3 分                                                      |

|    |              |     |                     |
|----|--------------|-----|---------------------|
| 素养 | 示意，有礼貌       |     |                     |
|    | 答题完毕离场时整理好工位 | 3 分 | 鼠标、键盘、凳子等摆放整齐，得 3 分 |

## 项目 7：图书管理系统数据库设计

### 13.试题编号：J1-6-3：“人力资源系统员工工资信息管理模块”数据库设计

#### （1）任务描述

《员工工资信息管理模块》模块的 E-R 图如图 13-1 所示，物理数据模型如图 13-2 所示，数据表字段名定义见表 13-1。请按以下设计完成数据库创建、数据表创建和数据操作任务：

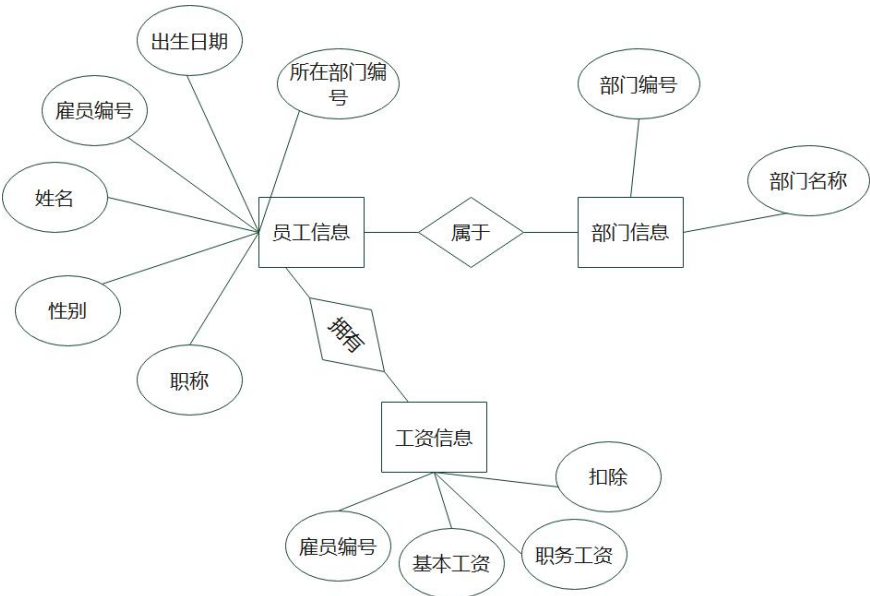


图 13-1 E-R 图

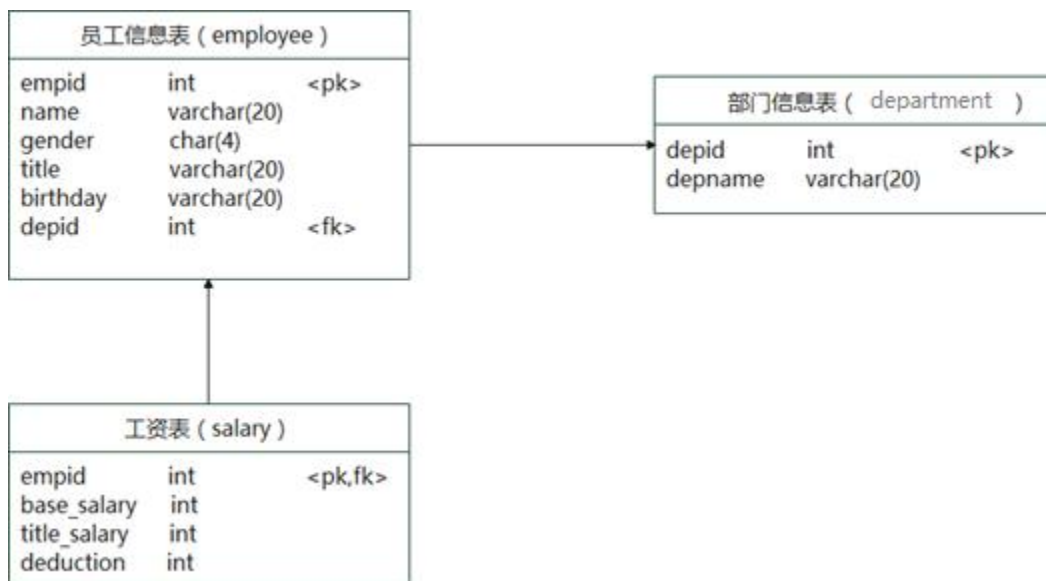


图13-2 物理数据原型

表13-1 字段名定义表

| 字段名      | 字段说明 | 字段名          | 字段说明 |
|----------|------|--------------|------|
| empid    | 雇员编号 | depid        | 部门编号 |
| name     | 姓名   | depname      | 部门名称 |
| gender   | 性别   | base_salary  | 基本工资 |
| birthday | 出生日期 | title_salary | 职务工资 |
| title    | 职称   | deduction    | 扣除   |

### 任务一：创建数据库（10 分）

创建数据库 SalaryDB。新建 T1-1.sql 文件，在该文件中编写语句创建指定的数据库，使用语句显示已经创建的数据库，并使用语句使用 SalaryDB 数据库。将上述操作截图，命名为 T1-1.jpg 文件，将 T1-1.sql、T1-1.jpg 保存到考生目录 T1 下。

### 任务二：创建数据表及表间的关系及约束（30 分）

创建 T2-1.sql 文件并完成以下功能：

- 根据图 3.16.2、和表 3.16.1，创建数据表 employee、department、salary。
- 根据物理数据原型，创建数据关系表。（15 分）

将创建表 employee、department、salary 的语句以及执行结果截图，命名分别为 T2-1.jpg、T2-2.jpg、T2-3.jpg，将 T2-1.sql、T2-1.jpg、T2-2.jpg、T2-3.jpg 四个文件保存到考生目录 T2 下。

### 任务三：数据操作（50 分）

考生目录 T3 下创建 T3-1.sql 文件并用 SQL 语句执行以下操作：

- 修改表结构，在部门表中添加一个“部门简介”字段，remarks 类型为 text，显示表结构并将结果截图，命名为 T3-1.jpg，保存到考生目录 T3 下。（5 分）
- 在每个表中插入以下测试数据；

向 employee 表插入数据：

| 雇员编号 | 姓名 | 性别 | 职称    | 出生日期       | 部门编号 |
|------|----|----|-------|------------|------|
| 1001 | 张三 | 男  | 高级工程师 | 1975-01-01 | 111  |
| 1002 | 李四 | 女  | 助工    | 1985-01-01 | 111  |
| 1003 | 王五 | 男  | 工程师   | 1978-11-11 | 222  |
| 1004 | 赵六 | 男  | 工程师   | 1979-01-01 | 222  |

向 department 表插入数据：

| 部门编号 | 部门名称 |
|------|------|
| 111  | 生产部  |
| 222  | 销售部  |
| 333  | 人事部  |

向 salary 表插入数据：

| 雇员编号 | 基本工资 | 职务工资 | 扣除  |
|------|------|------|-----|
| 1001 | 2200 | 1100 | 200 |
| 1002 | 1200 | 200  | 100 |
| 1003 | 1900 | 700  | 200 |
| 1004 | 1950 | 700  | 150 |

查询 department、employee 两个表中的数据并截图，命名分别为 T3-2.jpg、T3-3.jpg，保存到考生目录 T3 下。（5 分）

- 将工资表中雇员编号为 1004 的员工将她的基本工资改为 5700 元，职务工资为 600，修改后查询数据并将结果截图，命名为 T3-4.jpg，保存到考生目录 T3 下。（5 分）
- 查询出每个雇员的雇员编号，姓名，职称，实发工资（基本工资+职务工资-扣除），并将结果截图，命名为 T3-5.jpg，保存到考生目录 T3 下。（10 分）
- 查询姓“李”员工的信息，并将结果截图，命名为 T3-6.jpg，保存到考生目录 T3 下。（5 分）
- 统计各类职称的人数，显示“职称”、“人数”两列数据并将结果截图，命名为 T3-7.jpg，保存到考生目录 T3 下。（5 分）
- 统计各部门的部门名称、基本工资总和、平均基本工资，若部门中没有员工则不显示，并将结果截图，命名为 T3-8.jpg，保存到考生目录 T3 下。（5 分）
- 创建视图 project\_view，视图的数据为统计所有员工基本工资的总和，使用语句显

示视图 `project_view` 的结构并截图，命名为 `T3-9.jpg`，保存到考生目录 `T3` 下。（10 分）

### 良好的习惯和职业素养（10 分）

- 答题过程中保持安静，有问题举手示意，有礼貌。（3 分）
- 数据库名命名规范，数据表名命名规范，字段名命名规范。（3 分）
- 答题完毕离场时整理好工位。（4 分）

## （2）实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。

计算机安装了 Windows 操作系统，MySQL 数据库软件以及 Office 办公软件。

## （3）考核时量及要求

- 考核时长 3 小时；
- 创建答题文件夹创建以“学号\_题号”命名的文件夹，存放所有答题文件；
- 创建答题文件，包括：SQL 脚本文件，如 `J3-1.sql`

## （4）评分细则

表 13-2 考核评价细则表

| 评价项            | 分值   | 评分细则                                                                                                    |
|----------------|------|---------------------------------------------------------------------------------------------------------|
| 数据库创建          | 10 分 | 创建考生目录 <code>T1</code> ，文件 <code>T1-1.sql</code> 、 <code>T1-1.jpg</code> ，得 1.5 分                       |
|                |      | 使用语句创建了数据库，得 3 分                                                                                        |
|                |      | 使用语句显示已经创建的数据库的名字，得 3 分                                                                                 |
|                |      | 使用新创建的数据库，得 2.5 分                                                                                       |
| 创建数据表及表间的关系及约束 | 30 分 | 创建考生目录 <code>T2</code> ，文件 <code>T2-1.sql</code> 、 <code>T2-1.jpg</code> 、 <code>T2-2.jpg</code> ，得 2 分 |
|                |      | 表 <code>department</code> 创建的字段名、字段类型、主键正确，并正确显示表结构及主键约束，得 8 分                                          |
|                |      | 表 <code>employee</code> 创建的字段名、字段类型、主键正确、外键正确，并正确显示表结构及主键约束，得 10 分                                      |
|                |      | 表 <code>salary</code> 创建的字段名、字段类型、主键正确，并正确显示表结构及主键和外键约束，得 10 分                                          |
| 数据操作           | 50 分 | 修改表结构，在部门表中添加一个“部门简介”字段， <code>remarks</code> 类型为 <code>text</code> ，显示表结构，得 5 分                        |
|                |      | 表 <code>employee</code> 和 <code>department</code> 正确插入数据，得 5 分                                          |
|                |      | 将工资表中雇员编号为 1004 的员工将她的基本工资改为 5700 元，职务工资为 600，修改后查询数据，得 5 分                                             |
|                |      | 查询出每个雇员的雇员编号，姓名，职称，实发工资（基本工资+职务工资-扣除），得 10 分                                                            |

|            |                       |     |                                                                                       |
|------------|-----------------------|-----|---------------------------------------------------------------------------------------|
|            |                       |     | 查询姓“李”员工的信息，得 5 分                                                                     |
|            |                       |     | 统计各类职称的人数，显示“职称”、“人数”两列数据，得 5 分                                                       |
|            |                       |     | 统计各部门的部门名称、基本工资总和、平均基本工资，若部门中没有员工则不显示，得 5 分                                           |
|            |                       |     | 创建视图 <b>project_view</b> ，视图的数据为统计所有员工基本工资的总和，使用语句显示视图 <b>project_view</b> 的结构，得 10 分 |
| 良好的习惯和职业素养 | 代码书写格式规范              | 4 分 | 脚本缩进、换行规范，得 4 分                                                                       |
|            | 答题过程中保持安静，有问题举手示意，有礼貌 | 3 分 | 遵守考场纪律、有礼貌，得 3 分                                                                      |
|            | 答题完毕离场时整理好工位          | 3 分 | 鼠标、键盘、凳子等摆放整齐，得 3 分                                                                   |

## 项目 7：图书管理系统数据库设计

### 14.试题编号：J1-7-1：“图书管理系统借阅管理模块”数据库设计

#### (1) 任务描述

《借阅管理》模块的 E-R 图如图 14-1 所示，物理数据模型如图 14-2 所示，数据表字段名定义见表 14-1。请按以下设计完成数据库创建、数据表创建和数据操作任务：

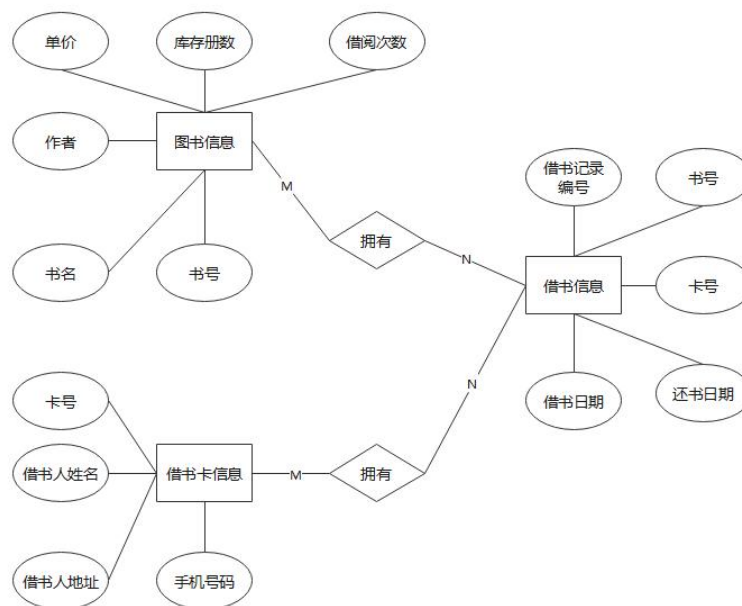


图 14-1E-R 图

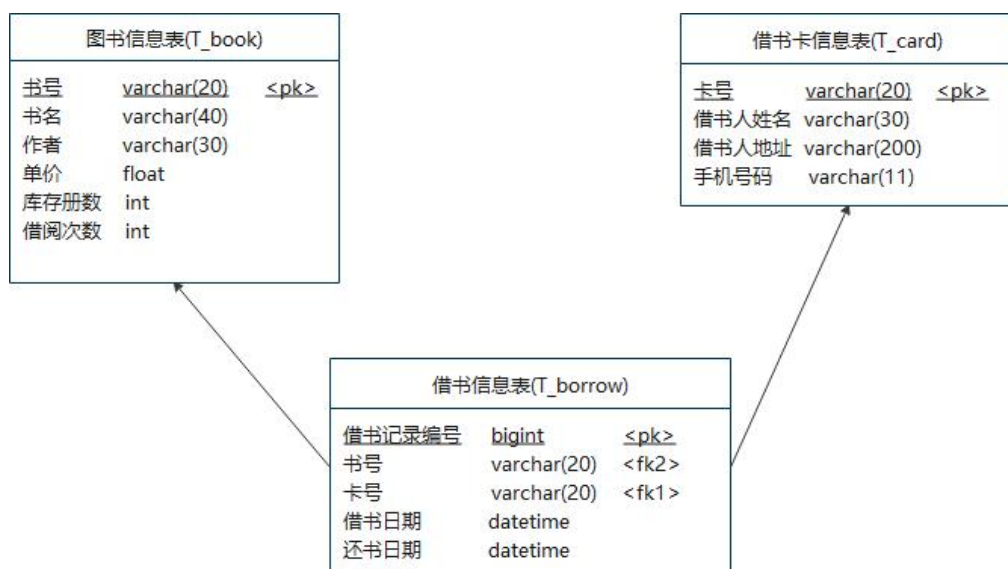


图 14-2 物理数据原型

表 14-1 字段名定义表

| 字段名       | 字段说明 | 字段名         | 字段说明   |
|-----------|------|-------------|--------|
| book_no   | 书号   | card_name   | 借书人姓名  |
| book_name | 书名   | adress      | 借书人地址  |
| author    | 作者   | mobile      | 手机号码   |
| price     | 单价   | borrow_id   | 借书记录编号 |
| qty       | 库存册数 | borrow_date | 借书日期   |
| loan_qty  | 借阅次数 | return_date | 还书日期   |
| card_no   | 卡号   |             |        |

### 任务一：创建数据库（10 分）

创建数据库 BookDB。新建 T1-1.sql 文件，在该文件中编写语句创建指定的数据库，使用语句显示已经创建的数据库，并使用语句使用 BookDB 数据库。将上述操作截图，命名为 T1-1.jpg 文件，将 T1-1.sql、T1-1.jpg 保存到考生目录 T1 下。

### 任务二：创建数据表及表间的关系及约束（30 分）

创建 T2-1.sql 文件并完成以下功能：

- 根据图 14-1、14-2 和表 14-1，创建数据表 T\_card、T\_book、T\_borrow。（15 分）
- 根据物理数据原型，创建数据表之间的关系。（15 分）

将创建表 T\_card、T\_book、T\_borrow 的语句以及执行结果截图，命名分别为 T2-1.jpg、T2-2.jpg、T2-3.jpg，将 T2-1.sql、T2-1.jpg、T2-2.jpg、T2-3.jpg 四个文件保存到考生目录 T2 下。

### 任务三：数据操作（50 分）

考生目录 T3 下创建 T3-1.sql 文件并用 SQL 语句执行以下操作：

➤ 在每个表中插入以下测试数据；

向 T\_card 表插入数据：

| 卡号    | 借书人姓名 | 借书人地址 | 手机号码        |
|-------|-------|-------|-------------|
| c1001 | 张三    | 天心区   | 17674782499 |
| c1002 | 张斌    | 雨花区   | 12300000000 |
| c1003 | 肖天    | 芙蓉区   | 17600000009 |
| c1004 | 王勇    | 星沙    | 15116242000 |
| c1005 | 王家其   | 天心    | 15250995038 |
| c1006 | 陈龙    | 开福区   | 17373102888 |

向 T\_book 表插入数据：

| 书号   | 书名    | 作者   | 单价   | 库存册数 | 借阅次数 |
|------|-------|------|------|------|------|
| b001 | 大数据概论 | 李四   | 56.3 | 50   | 13   |
| b002 | 西游记   | 吴承恩  | 32.5 | 100  | 53   |
| b003 | 红楼梦   | 曹雪芹  | 78.6 | 56   | 45   |
| b004 | 战争与和平 | 托尔斯泰 | 36.5 | 67   | 53   |
| b005 | 巴黎圣母院 | 雨果   | 57.3 | 4    | 13   |
| b006 | 童年    | 高尔基  | 36.5 | 2    | 53   |

向 T\_borrow 表插入数据：

| 借书记录编号 | 书号   | 卡号    | 借书日期       | 还书日期       |
|--------|------|-------|------------|------------|
| 12001  | b001 | c1001 | 2006-08-11 | 2006-09-21 |
| 12002  | b002 | c1002 | 2018-01-01 | 2018-01-16 |
| 12003  | b003 | c1004 | 2019-08-11 | 2019-09-21 |
| 12004  | b003 | c1006 | 2020-01-01 | 2020-01-10 |
| 12005  | b004 | c1001 | 2007-08-30 | 2007-09-21 |
| 12006  | b004 | c1001 | 2018-07-01 |            |

查询 T\_card、T\_book、T\_borrow 三个表中的数据并截图，命名分别为 T3-1.jpg、T3-2.jpg、

T3-3.jpg，保存到考生目录 T3 下。（5 分）

- 在 T\_book 表中增加一个字段“备注”（字段名：remark、字段类型：varchar(20)）放在现有字段的最后，增加字段后显示表结构并截图，命名分别为 T3-4.jpg，保存到考生目录 T3 下。（5 分）
- 查询出日期为 2019-10-31 以后借出的图书信息，包括书名和作者，并将结果截图，命名为 T3-5.jpg，保存到考生目录 T3 下。（5 分）
- 查询出没有还书（还书日期为空）的借书人姓名，并将结果截图，命名为 T3-6.jpg，

保存到考生目录 T3 下。(5 分)

- 查询出库存数量小于 5 册的图书的书名,并将结果截图,命名为 T3-7.jpg,保存到考生目录 T3 下。(5 分)
- 统计出学生卡号“c1001”2018 年借阅的图书的书名,并将结果截图,命名为 T3-8.jpg,保存到考生目录 T3 下。(5 分)
- 查询借书人地址中包含“天心”的借书人姓名,并将结果截图,命名为 T3-9.jpg,保存到考生目录 T3 下。(5 分)
- 编写视图 ProjInfo\_view1 查询借书人的姓名,手机号码和地址,使用语句显示视图 ProjInfo\_view1 中的数据并截图,命名为 T3-10.jpg,保存到考生目录 T3 下。(10 分)
- 删除上题中创建视图 ProjInfo\_view1,使用 show tables 语句查看所有数据库表,验证视图 ProjInfo\_view1 已经删除并将结果截图,命名为 T3-11.jpg,保存到考生目录 T3 下。(5 分)

#### 良好的习惯和职业素养 (10 分)

- 答题过程中保持安静,有问题举手示意,有礼貌。(3 分)
- 数据库名命名规范,数据表名命名规范,字段名命名规范。(3 分)
- 答题完毕离场时整理好工位。(4 分)

### (2) 实施条件

安排在教学实训机房,计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。

计算机安装了 Windows 操作系统,MySQL 数据库软件以及 Office 办公软件。

### (3) 考核时量及要求

- 考核时长 3 小时;
- 创建答题文件夹创建以“学号\_题号”命名的文件夹,存放所有答题文件;
- 创建答题文件,包括:SQL 脚本文件,如 J3-1.sql

### (4) 评分细则

表 14-2 考核评价细则表

| 评价项   | 分值   | 评分细则                                     |
|-------|------|------------------------------------------|
| 数据库创建 | 10 分 | 创建考生目录 T1, 文件 T1-1.sql、T1-1.jpg, 得 1.5 分 |
|       |      | 使用语句创建了数据库, 得 3 分                        |
|       |      | 使用语句显示已经创建的数据库的名字, 得 3 分                 |
|       |      | 使用新创建的数据库, 得 2.5 分                       |

|                |                       |      |                                                                                   |
|----------------|-----------------------|------|-----------------------------------------------------------------------------------|
| 创建数据表及表间的关系及约束 |                       | 30 分 | 创建考生目录 T2，文件 T2-1.sql、T2-1.jpg、T2-2.jpg，得 2 分                                     |
|                |                       |      | 表 T_card 创建的字段名、字段类型、主键正确，并正确显示表结构及主键约束，得 8 分                                     |
|                |                       |      | 表 T_book 创建的字段名、字段类型、主键正确、外键正确，并正确显示表结构及主键约束，得 8 分                                |
|                |                       |      | 表 T_borrow 创建的字段名、字段类型、主键正确，并正确显示表结构及主键和外键约束，得 12 分                               |
| 数据操作           |                       | 50 分 | 表 T_card 和 T_book 和 T_borrow 正确插入数据，得 5 分                                         |
|                |                       |      | 在 T_book 表中增加一个字段“备注”(字段名:remark、字段类型:varchar(20))放在现有字段的最后，增加字段后显示表结构，得 5 分      |
|                |                       |      | 查询出日期为 2019-10-31 以后借出的图书信息,包括书名和作者，得 5 分                                         |
|                |                       |      | 查询出没有还书(还书日期为空)的借书人姓名，得 5 分                                                       |
|                |                       |      | 查询出库存数量小于 5 册的图书的书名，得 5 分                                                         |
|                |                       |      | 统计出学生卡号“c1001”2018 年借阅的图书的书名，得 5 分                                                |
|                |                       |      | 查询借书人地址中包含“天心”的借书人姓名，得 5 分                                                        |
|                |                       |      | 编写视图 ProjInfo_view1 查询借书人的姓名,手机号码和地址，使用语句显示视图 ProjInfo_view1 中的数据，得 10 分          |
|                |                       |      | 删除上题中创建视图 ProjInfo_view1，使用 show tables 语句查看所有数据库表,验证视图 ProjInfo_view1 已经删除，得 5 分 |
| 良好的习惯和职业素养     | 代码书写格式规范              | 4 分  | 脚本缩进、换行规范，得 4 分                                                                   |
|                | 答题过程中保持安静，有问题举手示意，有礼貌 | 3 分  | 遵守考场纪律、有礼貌，得 3 分                                                                  |
|                | 答题完毕离场时整理好工位          | 3 分  | 鼠标、键盘、凳子等摆放整齐，得 3 分                                                               |

## 15.试题编号：J1-7-2：“图书订购系统管理模块”数据库设计

### (1) 任务描述

《订购管理》模块的 E-R 图如图 15-1 所示，物理数据模型如图 15-2 所示，数据表字段名定义见表 15-1。请按以下设计完成数据库创建、数据表创建和数据操作任务：

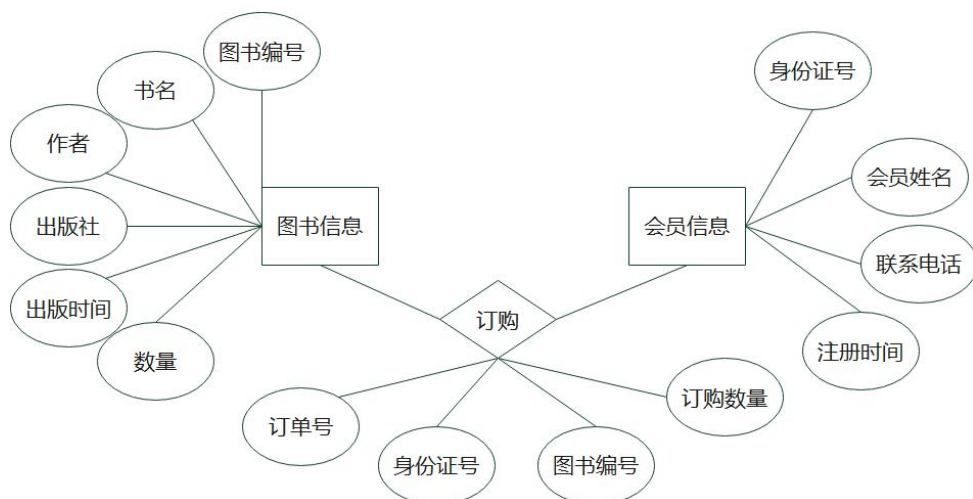


图 15-1 E-R 图

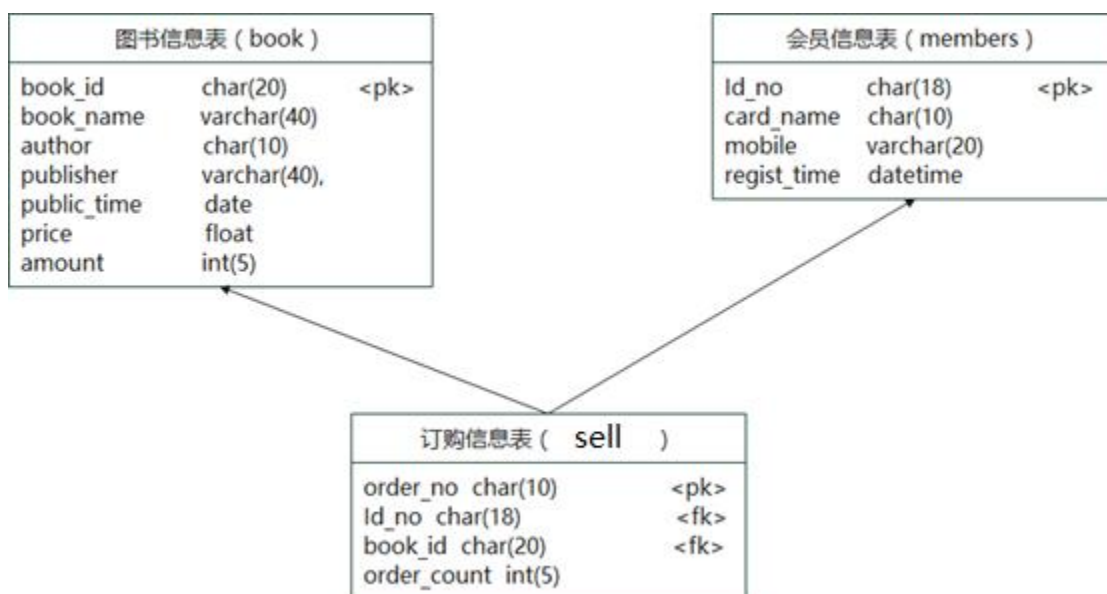


图 15-2 物理数据原型

表 15-1 字段名定义表

| 字段名         | 字段说明 | 字段名         | 字段说明 |
|-------------|------|-------------|------|
| book_id     | 图书编号 | card_name   | 会员姓名 |
| book_name   | 书名   | mobile      | 联系电话 |
| author      | 作者   | regist_time | 注册时间 |
| price       | 单价   | order_no    | 订单号  |
| public_time | 出版时间 | order_count | 数量   |
| amount      | 数量   | Id_no       | 身份证号 |
| publisher   | 出版社  |             |      |

### 任务一：创建数据库（10 分）

创建数据库 BookDB。新建 T1-1.sql 文件，在该文件中编写语句创建指定的数据库，使用语句显示已经创建的数据库，并使用语句使用 BookDB 数据库。将上述操作截图，命名为

T1-1.jpg 文件，将 T1-1.sql、T1-1.jpg 保存到考生目录 T1 下。

## 任务二：创建数据表及表间的关系及约束（30 分）

创建 T2-1.sql 文件并完成以下功能：

- 根据图 15-1、15-2 和表 15-1，创建数据表 book、members、sell。（15 分）
- 根据物理数据原型，创建数据表之间的关系。（15 分）

将创建表 book、members、sell 的语句以及执行结果截图，命名分别为 T2-1.jpg、T2-2.jpg、

T2-3.jpg，将 T2-1.sql、T2-1.jpg、T2-2.jpg、T2-3.jpg 四个文件保存到考生目录 T2 下。

## 任务三：数据操作（50 分）

考生目录 T3 下创建 T3-1.sql 文件并用 SQL 语句执行以下操作：

- 在每个表中插入以下测试数据；

向 book 表插入数据：

| 图书编号                 | 书名          | 作者  | 出版社     | 出版时间       | 价格   | 数量 |
|----------------------|-------------|-----|---------|------------|------|----|
| ISBN7-115-12683-6/tp | 计算机基础       | 刘仁权 | 人民邮电出版社 | 2020-01-11 | 28.9 | 56 |
| ISBN7-302-05701-x/tp | 网页程序设计      | 王珊珊 | 清华大学出版社 | 2019-09-11 | 59.9 | 65 |
| ISBN7-301-06342-3/tp | 网络数据库       | 王伟  | 北京大学出版社 | 2018-09-08 | 54.2 | 58 |
| ISBN 978-7-301-13355 | MYSQL 数据库   | 李刚  | 北京大学出版社 | 2020-09-21 | 48.9 | 45 |
| ISBN7-301-12683-6/tp | Hadoop 平台搭建 | 张华生 | 湘潭大学出版社 | 2021-05-01 | 45.6 | 16 |
| ISBN7-115-12689-6    | 网络操作系统      | 周德威 | 大连理工出版社 | 2019-10-15 | 46.2 | 15 |

向 members 表插入数据：

| 身份证号               | 会员姓名 | 联系电话        | 注册时间                |
|--------------------|------|-------------|---------------------|
| 430103198608201963 | 张三   | 51985523    | 2007-01-23 08:15:45 |
| 110108198003012023 | 李小冰  | 13651111081 | 2007-01-18 08:57:18 |
| 110108198205010022 | 王林   | 12501234123 | 2007-01-12 08:12:30 |
| 110101198001013013 | 张凯   | 13611320000 | 2007-01-15 09:12:23 |
| 110102198103010012 | 赵宏宇  | 13601234123 | 2007-03-04 18:23:45 |

向 sell 表插入数据：

| 订单号 | 身份证号 | 图书编号 | 数量 |
|-----|------|------|----|
|-----|------|------|----|

|   |                    |                      |    |
|---|--------------------|----------------------|----|
| 1 | 430103198608201963 | ISBN7-301-12683-6/tp | 16 |
| 2 | 430103198608201963 | ISBN7-115-12689-6    | 23 |
| 3 | 110108198003012023 | ISBN7-301-12683-6/tp | 25 |
| 4 | 110108198205010022 | ISBN7-115-12689-6    | 18 |
| 5 | 110101198001013013 | ISBN7-115-12683-6/tp | 15 |
| 6 | 430103198608201963 | ISBN7-301-06342-3/tp | 11 |
| 7 | 110102198103010012 | ISBN7-302-05701-x/tp | 9  |

查询 book、members、sell 三个表中的数据并截图,命名分别为 T3-1.jpg、T3-2.jpg、T3-3.jpg,保存到考生目录 T3 下。(5 分)

- 查询 book 表中 2019 年出版的图书书名和作者,并将结果进行截图,命名为 T3-4.jpg,保存到考生目录 T3 下。(5 分)
- 查询单次订购图书总数量最多的会员的姓名,并将结果进行截图,命名为 T3-5.jpg,保存到考生目录 T3 下。(10 分)
- 查询图书信息表中,图书的总数量,并将结果进行截图,命名为 T3-6.jpg,保存到考生目录 T3 下。(5 分)
- 查询订购了图书的书名、会员名、订购数量,并将结果截图,命名为 T3-7.jpg,保存到考生目录 T3 下。(10 分)
- 查询图书信息表中,价格最高的图书的书名,并将结果截图,命名为 T3-8.jpg,保存到考生目录 T3 下。(5 分)
- 查询会员信息表中,1980 年出生的会员的姓名,并将结果截图,命名为 T3-9.jpg,保存到考生目录 T3 下。(10 分)

#### 良好的习惯和职业素养 (10 分)

- 答题过程中保持安静,有问题举手示意,有礼貌。(3 分)
- 数据库名命名规范,数据表名命名规范,字段名命名规范。(3 分)
- 答题完毕离场时整理好工位。(4 分)

## (2) 实施条件

安排在教学实训机房,计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。

计算机安装了 Windows 操作系统，MySQL 数据库软件以及 Office 办公软件。

### (3) 考核时量及要求

- 考核时长 3 小时；
- 创建答题文件夹创建以“学号\_题号”命名的文件夹，存放所有答题文件；
- 创建答题文件，包括：SQL 脚本文件，如 J3-1.sql

### (4) 评分细则

表15-2 考核评价细则表

| 评价项            |                       | 分值   | 评分细则                                            |
|----------------|-----------------------|------|-------------------------------------------------|
| 数据库创建          |                       | 10 分 | 创建考生目录 T1，文件 T1-1.sql、T1-1.jpg，得 1.5 分          |
|                |                       |      | 使用语句创建了数据库，得 3 分                                |
|                |                       |      | 使用语句显示已经创建的数据库的名字，得 3 分                         |
|                |                       |      | 使用新创建的数据库，得 2.5 分                               |
| 创建数据表及表间的关系及约束 |                       | 30 分 | 创建考生目录 T2，文件 T2-1.sql、T2-1.jpg、T2-2.jpg，得 2 分   |
|                |                       |      | 表 members 创建的字段名、字段类型、主键正确，并正确显示表结构及主键约束，得 8 分  |
|                |                       |      | 表 book 创建的字段名、字段类型、主键正确，并正确显示表结构及主键约束，得 8 分     |
|                |                       |      | 表 sell 创建的字段名、字段类型、主键正确，并正确显示表结构及主键和外键约束，得 12 分 |
| 数据操作           |                       | 50 分 | 表 members 和 book 和 sell 正确插入数据，得 5 分            |
|                |                       |      | 查询 book 表中 2019 年出版的图书书名和作者，得 5 分               |
|                |                       |      | 查询单次订购图书总数量最多的会员的姓名，得 10 分                      |
|                |                       |      | 查询图书信息表中，图书的总数量，得 5 分                           |
|                |                       |      | 查询订购了图书的书名、会员名、订购数量，得 10 分                      |
|                |                       |      | 查询图书信息表中，价格最高的图书的书名，得 5 分                       |
|                |                       |      | 查询会员信息表中，1980 年出生的会员的姓名，得 10 分                  |
| 良好的习惯和职业素养     | 代码书写格式规范              | 4 分  | 脚本缩进、换行规范，得 4 分                                 |
|                | 答题过程中保持安静，有问题举手示意，有礼貌 | 3 分  | 遵守考场纪律、有礼貌，得 3 分                                |
|                | 答题完毕离场时整理好工位          | 3 分  | 鼠标、键盘、凳子等摆放整齐，得 3 分                             |

模块二 大数据平台部署

项目 1：Hadoop 伪分布式模式部署

1.试题编号：J2-1-1：美团大数据平台 Hadoop 伪分布式模式部署

(1) 任务描述

美团网购平台需为消费者提供非常好的精品消费指南及消费折扣,通过结合大数据高性能的分析，帮助消费者提前查找某地区的美食、酒店、演出等数据并完成数据展示。

现在要进行大数据分析并完成展示，你作为该公司大数据平台运维工程师，需在单机上安装伪分布式 Hadoop 环境，在该模式下，Hadoop 守护进程运行在一台机器上，模拟一个小规模的集群，伪分布式规划如下表：

表 1-1 伪分布式规划

| 节点<br>角色 | 虚拟机名<br>称 | 主机名    | 节点<br>IP | 运行进程                                              |
|----------|-----------|--------|----------|---------------------------------------------------|
| 单一<br>节点 | Hadoop01  | Node01 | 自动<br>分配 | NameNode<br><br>DataNode<br><br>SecondaryNameNode |

本环节需要使用虚拟机软件 VMware 自行创建虚拟机，并完成 Hadoop 伪分布式模式部署，所需安装包均位于“D:\大数据平台搭建与运维”下，主要任务如下：

任务一：创建虚拟机与设置远程登录（25 分）

- 创建虚拟机过程中，根据机房所提供的操作系统镜像文件选择相应操作系统和版本，截取客户机操作系统和版本选择后的界面，图片保存到考生文件夹下，并命名为“1-1-1 创建虚拟机 1.png”。然后根据题目要求虚拟机命名为 Hadoop01 并存放位置为 D:\VM，若 D 盘下\VM 路径不存在请先创建；截取虚拟机命名设置和虚拟机存放路径设置，图片保存到考生文件夹下，并命名为“1-1-2 创建虚拟机 2.png”。（8 分）
- 虚拟机创建过程中，根据题目要求设置主机名为 Node01，并打开网络连接，将设置主机名界面截图；虚拟机创建完成后，使用相关命令查看主机名，将命令及查看效果截图；

图片保存到考生文件夹下，并分别命名为“1-2-1 设置主机名.png”和“1-2-2 查看主机名.png”。（6 分）

- 使用相关命令查看虚拟机 IP 地址，截取命令及查看结果，图片保存到考生文件夹下，并命名为“1-3 查看 ip.png”。（5 分）
- 使用 xshell 远程连接虚拟机，会话名称要求设置为虚拟机的名称，截取设置完成后的新建会话属性界面，连接成功后在 xshell 上显示虚拟机登陆状态并截取登陆状态，图片保存到考生文件夹下，并分别命名为“1-4 远程登录 1.png”和“1-4 远程登录 2.png”。（6 分）

### **任务二：JDK 的安装和环境变量配置（15 分）**

- 使用 xftp 工具将 JDK 安装包上传到/export/software（此路径需自行创建）文件夹下，使用相关命令查看 JDK 安装包是否上传成功，并截取所使用的命令及查看效果，图片保存到考生文件夹下，并命名为“2-1 上传 jdk.png”。（3 分）
- 将 JDK 安装包解压到/export/servers（此路径需自行创建）目录下，并将 JDK 安装包重命名为 jdk，使用相关命令查看 JDK 安装包是否成功解压并重命名，截取重命名命令及查看效果，图片保存到考生文件夹下，并命名为“2-2 解压 jdk.png”。（3 分）
- 配置 JDK 环境变量，截取配置内容，图片保存到考生文件夹下，并命名为“2-3 配置 jdk 环境变量.png”。（6 分）
- 使用相关指令验证 JDK 安装是否成功，将命令及结果截图并保存，图片保存到考生文件夹下，并命名为“2-4 验证 jdk.png”。（3 分）

### **任务三：Hadoop 的安装和环境变量配置（15 分）**

- 使用 xftp 工具将 Hadoop 安装包上传到/export/software 文件夹下，截取/export/software 文件夹下的文件，图片保存到考生文件夹下，并命名为“3-1 上传 hadoop.png”。（3 分）
- 将 Hadoop 安装包解压到/export/servers 目录下，使用相关命令查看 Hadoop 安装包是否成功并截取查看效果，图片保存到考生文件夹下，并命名为“3-2 解压 hadoop.png”。（3 分）
- 配置 Hadoop 环境变量，截取配置内容，图片保存到考生文件夹下，并命名为“3-3 配

置 `hadoop` 环境变量.png”。(6 分)

- 使用相关指令验证 Hadoop 安装是否成功, 将命令及结果截图并保存, 图片保存到考生文件夹下, 并命名为 “3-4 验证 `hadoop`.png”。(3 分)

#### **任务四: 修改相关配置文件并启动 Hadoop (35 分)**

- Hadoop 的运行需要依赖 JDK, 而 `hadoop-env.sh` 文件为 Hadoop 的运行环境配置文件, 请提交该配置文件的修改内容截图, 图片保存到考生文件夹下, 并命名为 “4-1 配置 `hadoop` 运行环境.png”。(5 分)
- `core-site.xml` 文件为 Hadoop 的核心配置文件, 用于定义系统级别的参数, 设定 `namenode` 的主机名及端口, 请提交该配置文件的修改内容截图, 图片保存到考生文件夹下, 并命名为 “4-2 修改 `hadoop` 核心配置文件.png”。(8 分)
- `hdfs-site.xml` 文件为 HDFS 核心配置文件, 如文件副本的个数、块大小及是否使用强制权限等, 设定 HDFS 存储文件的副本个数默认为 1, 请提交该配置文件的修改内容截图, 图片保存到考生文件夹下, 并命名为 “4-3 修改 `hdfs` 核心配置文件.png”。(8 分)
- 执行相关命令格式化 NameNode, 截取格式化成功界面, 图片保存到考生文件夹下, 并命名为 “4-4-1 格式化 `namenode`.png”; 然后启动 `hdfs` 集群, 并使用 `jps` 进行查看, 截取启动命令与 `jps` 查看结果, 图片保存到考生文件夹下, 并命名为 “4-4-2 启动 Hadoop 集群.png”。(8 分)
- 使用相关命令在 HDFS 根目录下创建 `test` 目录, 并将 Hadoop 安装目录下的 `NOTICE.txt` 文档上传到 HDFS 根目录下的 `test` 目录下, 截取 web 界面文档上传情况, 图片保存到考生文件夹下, 并命名为 “4-5 上传文件.png”。(6 分)

#### **良好的习惯和职业素养 (10 分)**

- 答题过程中保持安静。(3 分)
- 有问题举手示意, 有礼貌。(3 分)
- 答题完毕离场时整理好工位, 不关闭计算机, 并按要求显示答案保存界面。(4 分)

## (2) 实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。

计算机安装有 Vmware workstation15, XShell, JDK1.8, 并有 Hadoop2.7.4 安装包和 CentOS7 64 位镜像文件，安装 Office 办公软件。

## (3) 考核时量及要求

- 考核时长 180 分钟；
- 请在桌面建立以“学号\_题号”命名的成果文件夹，所有提交文件均放在该目录下，操作过程的截图需按照题目要求，文件夹命名例如：201965110001\_J3-1-1。

## (4) 评分细则

表 1-2 任务一评分细则（25 分）

| 序号 | 评分项         | 分值  | 评分细则                                                                                                                                                                |
|----|-------------|-----|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 创建新的虚拟机     | 8 分 | 和参考答案一致满分；<br>创建版本为 CentOS 7 64 位，得 2 分；<br>考生文件夹中有 1-1-1 创建虚拟机 1.png，得 1 分；<br>虚拟机命名为 Hadoop01，得 2 分；<br>虚拟机存放路径为 D:\VM，得 2 分；<br>考生文件夹中有 1-1-2 创建虚拟机 2.png，得 1 分。 |
| 2  | 安装虚拟机并查看主机名 | 6 分 | 和参考答案一致满分；<br>主机名设置正确并打开网络连接，得 2 分；<br>考生文件夹中有 1-2-1 设置主机名.png 文件，得 1 分；<br>正确查看主机名（有 hostname 命令），得 2 分；<br>考生文件夹中有 1-2-2 查看主机名.png 文件，得 1 分。                      |
| 3  | 查看 ip 地址    | 5 分 | 和参考答案一致满分；                                                                                                                                                          |

|   |       |     |                                                                                                                                                               |
|---|-------|-----|---------------------------------------------------------------------------------------------------------------------------------------------------------------|
|   |       |     | <p>有 ip addr 或 ip addr show 命令，得 3 分；</p> <p>考生文件夹中有 1-3 查看 ip.png 文件，得 2 分。</p>                                                                              |
| 4 | 连接虚拟机 | 6 分 | <p>和参考答案一致满分；</p> <p>正确连接虚拟机，会话名称和 ip 地址均要求正确，得 2 分；</p> <p>考生文件夹中有 1-4 远程登录 1.png 文件，得 1 分；</p> <p>显示登录状态，得 2 分；</p> <p>考生文件夹中有 1-4 远程登录 2.png 文件，得 1 分。</p> |

表 1-3 任务二评分细则（15 分）

| 序号 | 评分项            | 分值  | 评分细则                                                                                           |
|----|----------------|-----|------------------------------------------------------------------------------------------------|
| 1  | 上传 JDK 安装包     | 3 分 | <p>和参考答案一致满分；</p> <p>查看 jdk 安装包，得 2 分；</p> <p>考生文件夹中有 2-1 上传 jdk.png，得 1 分。</p>                |
| 2  | 解压 JDK 安装包并重命名 | 3 分 | <p>和参考答案一致满分；</p> <p>重命名 JDK 安装包并正确查看，得 2 分；</p> <p>考生文件夹中有 2-2 解压 jdk.png 文件，得 1 分。</p>       |
| 3  | 配置环境变量         | 6 分 | <p>和参考答案一致满分；</p> <p>正确配置 jdk 环境变量，得 4 分；</p> <p>考生文件夹中有 2-3 配置 jdk 环境变量.png 文件，得 2 分。</p>     |
| 4  | 验证 JDK         | 3 分 | <p>和参考答案一致满分；</p> <p>有 java -version 命令，得 1.5 分；</p> <p>考生文件夹中有 2-4 验证 jdk.png 文件，得 1.5 分。</p> |

表 1-4 任务三评分细则（15 分）

| 序号 | 评分项         | 分值  | 评分细则                                                                                             |
|----|-------------|-----|--------------------------------------------------------------------------------------------------|
| 1  | 上传Hadoop安装包 | 3 分 | <p>和参考答案一致满分；</p> <p>查看 Hadoop 安装包上传情况，得 1.5 分；</p> <p>考生文件夹中有 3-1 上传 hadoop.png，得 1.5 分。</p>    |
| 2  | 解压Hadoop安装包 | 3 分 | <p>和参考答案一致满分；</p> <p>查看 Hadoop 安装包解压情况，得 1.5 分；</p> <p>考生文件夹中有 3-2 解压 hadoop.png 文件，得 1.5 分。</p> |

|   |           |     |                                                                                                    |
|---|-----------|-----|----------------------------------------------------------------------------------------------------|
| 3 | 配置环境变量    | 6 分 | 和参考答案一致满分；<br>正确配置 <b>hadoop</b> 环境变量，得 4 分；<br>考生文件夹中有 3-3 配置 <b>hadoop</b> 环境变量.png 文件，得 2 分。    |
| 4 | 验证 Hadoop | 3 分 | 和参考答案一致满分；<br>有 <b>hadoop version</b> 命令，得 1.5 分；<br>考生文件夹中有 3-4 验证 <b>hadoop</b> .png 文件，得 1.5 分。 |

表 1-5 任务四评分细则（35 分）

| 序号 | 评分项                                 | 分值  | 评分细则                                                                                                                                                                                                  |
|----|-------------------------------------|-----|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 修改<br><b>hadoop-env.sh</b> 配置<br>文件 | 5 分 | 和参考答案一致满分；<br><b>hadoop-env.sh</b> 配置正确（路径配置正确），得 3 分；<br>考生文件夹中有 4-1 配置 <b>hadoop</b> 运行环境.png，得 2 分。                                                                                                |
| 2  | 修改<br><b>core-site.xml</b> 配置<br>文件 | 8 分 | 和参考答案一致满分；<br>正确配置 <b>core-site.xml</b> 文件，得 6 分；<br>考生文件夹中有 4-2 修改 <b>hadoop</b> 核心配置文件.png 文件，得 2 分。                                                                                                |
| 3  | 修改<br><b>hdfs-site.xml</b> 配置<br>文件 | 8 分 | 和参考答案一致满分；<br>正确配置 <b>hdfs-site.xml</b> 文件，得 6 分；<br>考生文件夹中有 4-3 修改 <b>hdfs</b> 核心配置文件.png 文件，得 2 分。                                                                                                  |
| 4  | 格式化<br>NameNode                     | 8 分 | 和参考答案一致满分；<br>格式化成功，得 3 分；<br>考生文件夹中有 4-4-1 格式化 <b>namenode</b> .png 文件，得 1 分；<br>有 <b>start-dfs.sh</b> 和 <b>jps</b> 命令，得 2 分；<br>进程显示与题目一致，得 1 分；<br>考生文件夹中有 4-4-2 启动 <b>Hadoop</b> 集群.png 文件，得 1 分。 |
| 5  | <b>HDFS</b> 根目录<br>下创建目录并上传<br>文件   | 6 分 | 和参考答案一致满分；<br><b>HDFS</b> 下创建 <b>test</b> 目录，得 2 分；<br>上传文档到 <b>test</b> 目录下，得 2 分；<br>考生文件夹中有 4-5 上传文件.png 文件，得 2 分。                                                                                 |

表 1-6 职业素质评分细则（10 分）

| 序号 | 评分项  | 分<br>值 | 评分细则                         |
|----|------|--------|------------------------------|
| 1  | 操作规范 | 3 分    | 答题过程中保持安静，不损坏机器；             |
| 2  | 文明考试 | 3 分    | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌； |
| 3  | 工位整理 | 4 分    | 答题完毕立场时整理好工位，不关闭计算机，并按       |

|  |  |  |             |
|--|--|--|-------------|
|  |  |  | 要求显示答案保存界面。 |
|--|--|--|-------------|

**2.试题编号：J2-1-2：去哪儿旅游网大数据平台 Hadoop 伪分布式模式部署**

**(1) 任务描述**

去哪儿旅游网平台需为旅行者提供非常好的精品旅游指南及旅游折扣,通过结合大数据高性能的分析，帮助旅行者提前查找某地区的美食、酒店、机票等数据并完成数据展示。

现在要进行大数据分析并完成展示，你作为该公司大数据平台运维工程师，需在单机上安装伪分布式 Hadoop 环境，在该模式下，Hadoop 守护进程运行在一台机器上，模拟一个小规模集群，伪分布式规划如下表：

表 2-1 伪分布式规划

| 节点<br>角色 | 虚拟机名<br>称 | 主机名    | 节点<br>IP | 运行进程                                              |
|----------|-----------|--------|----------|---------------------------------------------------|
| 单一<br>节点 | Hadoop01  | Node01 | 自动<br>分配 | NameNode<br><br>DataNode<br><br>SecondaryNameNode |

本环节需要使用虚拟机软件 VMware 自行创建虚拟机，并完成 Hadoop 伪分布式模式部署，所需安装包均位于“D:\大数据平台搭建与运维”下，主要任务如下：

**任务一：虚拟机的安装与主机名的设置（20 分）**

- 创建虚拟机过程中，根据机房所提供的操作系统镜像文件选择相应操作系统和版本，截取客户机操作系统和版本选择后的界面，图片保存到考生文件夹下，并命名为“1-1-1 创建虚拟机 1.png”。然后根据题目要求虚拟机命名为 Hadoop01 并存放位置为 D:\VM，若 D 盘下\VM 路径不存在请先创建；截取虚拟机命名设置和虚拟机存放路径设置，图片保存到考生文件夹下，并命名为“1-1-2 创建虚拟机 2.png”。（8 分）
- 虚拟机创建过程中，根据题目要求设置主机名为 Node01，并打开网络连接，将设置主机名界面截图；虚拟机创建完成后，使用相关命令查看主机名，将命令及查看效果截图；图片保存到考生文件夹下，并分别命名为“1-2-1 设置主机名.png”和“1-2-2 查看主机名.png”。（6 分）

- 查看 ssh 服务是否安装，将命令和结果截图，图片保存到考生文件夹下，并命名为“1-3-1 查看 ssh 安装情况.png”；查看 ssh 服务是否启动（若未启动先使用相关命令启动 ssh 服务再查看），将命令和结果截图，图片保存到考生文件夹下，并命名为“1-3-2 查看 ssh 启动情况.png”。（6 分）

#### **任务二：远程登录和免密登录设置（20 分）**

- 使用相关命令查看虚拟机 IP 地址，截取命令及查看结果，图片保存到考生文件夹下，并命名为“2-1 查看 ip.png”。（4 分）
- 使用 xshell 远程连接虚拟机，会话名称要求设置为虚拟机的名称，截取设置完成后的新建会话属性界面，连接成功后在 xshell 上显示虚拟机登陆状态并截取登陆状态，图片保存到考生文件夹下，并分别命名为“2-2-1 远程登录 1.png”和“2-2-2 远程登录 2.png”。（6 分）
- 在虚拟机 Hadoop01 上生成密钥文件，将命令及结果截图，图片保存到考生文件夹下，并命名为“2-3 生成密钥.png”。（5 分）
- 执行相关指令将公钥复制到本机，将命令截图，图片保存到考生文件夹下，并命名为“2-4 复制公钥.png”。（5 分）

#### **任务三：JDK、Hadoop 的安装与环境变量配置（20 分）**

- 使用 xftp 工具将 JDK、Hadoop 安装包上传到/export/software（此路径需自行创建）文件夹下，使用相关命令查看 JDK、Hadoop 安装包是否上传成功，并截取所使用的命令及查看效果，图片保存到考生文件夹下，并命名为“3-1 上传安装包.png”。（4 分）
- 将 JDK、Hadoop 安装包解压到/export/servers（此路径需自行创建）目录下，并将 JDK 安装包重命名为 jdk，使用相关命令查看 JDK、Hadoop 安装包是否成功解压，截取重命名命令及查看效果，图片保存到考生文件夹下，并命名为“3-2 解压安装包.png”。（4 分）
- 配置 JDK、Hadoop 环境变量，截取配置内容，图片保存到考生文件夹下，并命名为“3-3 配置环境变量.png”。（8 分）
- 使用相关指令验证 JDK、Hadoop 安装是否成功，将命令及结果截图并保存，图片保存

到考生文件夹下，并命名为“3-4 验证 jdk 与 hadoop.png”。（4 分）

#### 任务四：修改相关配置文件与启动集群（30 分）

- Hadoop 的运行需要依赖 JDK，而 `hadoop-env.sh` 文件为 Hadoop 的运行环境配置文件，请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“4-1 配置 hadoop 运行环境.png”。（4 分）
- `core-site.xml` 文件为 Hadoop 的核心配置文件，用于定义系统级别的参数，设定 `namenode` 的主机名及端口，请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“4-2 修改 hadoop 核心配置文件.png”。（8 分）
- `hdfs-site.xml` 文件为 HDFS 核心配置文件，如文件副本的个数、块大小及是否使用强制权限等，设定 HDFS 存储文件的副本个数默认为 1，请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“4-3 修改 hdfs 核心配置文件.png”。（8 分）
- 执行相关命令格式化 `NameNode`，截取格式化成功界面，图片保存到考生文件夹下，并命名为“4-4 格式化 namenode.png”。（6 分）
- 使用相关指令启动 `NameNode` 和 `DataNode` 进程，并使用 `jps` 指令进行查看，截取 `jps` 结果，图片保存到考生文件夹下，并命名为“4-5 启动进程.png”。（4 分）

#### 良好的习惯和职业素养（10 分）

- 答题过程中保持安静。（3 分）
- 有问题举手示意，有礼貌。（3 分）
- 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。（4 分）

### （2）实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。计算机安装有 Vmware workstation15，XShell，JDK1.8，并有 Hadoop2.7.4 安装包和 CentOS7 64 位镜像文件，安装 Office 办公软件。

### （3）考核时量及要求

- 考核时长 180 分钟；
- 请在桌面建立以“学号\_题号”命名的成果文件夹，所有提交文件均放在该目录下，操

作过程的截图需按照题目要求，文件夹命名例如：201965110001\_J3-1-1。

#### (4) 评分细则

表 2-2 任务一评分细则（20 分）

| 序号 | 评分项          | 分值  | 评分细则                                                                                                                                                                       |
|----|--------------|-----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 创建新的虚拟机      | 8 分 | 和参考答案一致满分；<br>创建版本为 CentOS 7 64 位，得 2 分；<br>考生文件夹中有 1-1-1 创建虚拟机 1.png，得 1 分；<br>虚拟机命名为 Hadoop01，得 2 分；<br>虚拟机存放路径为 D:\VM，得 2 分；<br><br>考生文件夹中有 1-1-2 创建虚拟机 2.png，得 1 分。    |
| 2  | 设置主机名并查看     | 6 分 | 和参考答案一致满分；<br>主机名设置正确并打开网络连接，得 2 分；<br>考生文件夹中有 1-2-1 设置主机名.png 文件，得 1 分；<br>正确查看主机名（有 hostname 命令），得 2 分；<br>考生文件夹中有 1-2-2 查看主机名.png 文件，得 1 分。                             |
| 3  | 查看并启动 ssh 服务 | 6 分 | 和参考答案一致满分；<br>查看 ssh 服务是否安装命令正确，得 2 分；<br>考生文件夹中有 1-3-1 查看 ssh 安装情况.png 文件，得 1 分；<br>查看 ssh 服务是否启动命令正确（若未启动，则使用启动命令），得 2 分；<br><br>考生文件夹中有 1-3-2 查看 ssh 启动情况.png 文件，得 1 分。 |

表 2-3 任务二评分细则（20 分）

| 序号 | 评分项   | 分值  | 评分细则                                                                                                                                                   |
|----|-------|-----|--------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 查看 ip | 4 分 | 和参考答案一致满分；<br>有 ip addr 或 ip addr show 命令查看 ip 地址，得 2 分；<br><br>考生文件夹中有 2-1 查看 ip.png，得 2 分。                                                           |
| 2  | 远程登录  | 6 分 | 和参考答案一致满分；<br>新建远程连接，会话名称和 ip 地址均要求正确，得 2 分；<br>考生文件夹中有 2-2-1 远程登录 1.png 文件，得 1 分。<br>xshell 成功显示登录状态，得 2 分；<br><br>考生文件夹中有 2-2-2 远程登录 2.png 文件，得 1 分。 |

|   |        |     |                                                                 |
|---|--------|-----|-----------------------------------------------------------------|
|   |        |     | 分。                                                              |
| 3 | 生成秘钥文件 | 5 分 | 和参考答案一致满分；<br>正确生成秘钥文件，得 4 分；<br>考生文件夹中有 2-3 生成秘钥.png 文件，得 1 分。 |
| 4 | 复制公钥   | 5 分 | 和参考答案一致满分；<br>正确复制公钥到本机，得 4 分；<br>考生文件夹中有 2-4 复制公钥.png 文件，得 1 分 |

表 2-4 任务三评分细则（20 分）

| 序号 | 评分项             | 分值  | 评分细则                                                                                                                   |
|----|-----------------|-----|------------------------------------------------------------------------------------------------------------------------|
| 1  | 上传安装包           | 4 分 | 和参考答案一致满分；<br>查看 JDK 和 Hadoop 安装包上传情况，得 2 分；<br>考生文件夹中有 3-1 上传安装包.png，得 2 分。                                           |
| 2  | 解压到指定路径         | 4 分 | 和参考答案一致满分；<br>重命名 JDK，得 1 分；<br>查看 JDK 和 Hadoop 安装包解压情况，得 1.5 分；<br>考生文件夹中有 3-2 解压安装包.png 文件，得 1.5 分。                  |
| 3  | 配置环境变量          | 8 分 | 和参考答案一致满分；<br>正确配置环境变量，得 6 分；<br>考生文件夹中有 3-3 配置环境变量.png 文件，得 2 分。                                                      |
| 4  | 验证 JDK 和 Hadoop | 4 分 | 和参考答案一致满分；<br>有 java -version 命令，得 1.5 分；<br>有 hadoop version 命令，得 1.5 分；<br>考生文件夹中有 3-4 验证 jdk 与 hadoop.png 文件，得 1 分。 |

表 2-5 任务四评分细则（30 分）

| 序号 | 评分项                   | 分值  | 评分细则                                                                                     |
|----|-----------------------|-----|------------------------------------------------------------------------------------------|
| 1  | 修改 hadoop-env.sh 配置文件 | 4 分 | 和参考答案一致满分；<br>hadoop-env.sh 配置正确（路径配置正确），得 2 分；<br>考生文件夹中有 4-1 配置 hadoop 运行环境.png，得 2 分。 |
| 2  | 修改 core-site.xml 配置文件 | 8 分 | 和参考答案一致满分；<br>正确配置 core-site.xml 文件，得 6 分；<br>考生文件夹中有 4-2 修改 hadoop 核心配置文                |

|   |                                     |     |                                                                                                        |
|---|-------------------------------------|-----|--------------------------------------------------------------------------------------------------------|
|   |                                     |     | 件.png 文件，得 2 分。                                                                                        |
| 3 | 修改 hdfs-site.xml 配置文件               | 8 分 | 和参考答案一致满分；<br>正确配置 hdfs-site.xml 文件，得 6 分；<br>考生文件夹中有 4-3 修改 hdfs 核心配置文件.png 文件，得 2 分。                 |
| 4 | 格式化 NameNode                        | 6 分 | 和参考答案一致满分；<br>格式化成功，得 4 分；<br>考生文件夹中有 4-4 格式化 namenode.png 文件，得 2 分。                                   |
| 5 | 指令启动 NameNode 和 DataNode 并使用 jps 查看 | 4 分 | 和参考答案一致满分；<br>启动 NameNode 和 DataNode 正确，得 2 分；<br>使用 jps 查看进程，得 1 分；<br>考生文件夹中有 4-5 启动进程.png 文件，得 1 分。 |

表 2-6 职业素质评分细则（10 分）

| 序号 | 评分项  | 分值  | 评分细则                              |
|----|------|-----|-----------------------------------|
| 1  | 操作规范 | 3 分 | 答题过程中保持安静，不损坏机器；                  |
| 2  | 文明考试 | 3 分 | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌；      |
| 3  | 工位整理 | 4 分 | 答题完毕立场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |

### 3.试题编号：J2-1-3：赶集网大数据平台 Hadoop 伪分布式模式部署

#### （1）任务描述

赶集网平台需为应聘者匹配合适的岗位，通过结合大数据高性能的分析，帮助应聘者提前查找某地区某种岗位的薪资水平、福利待遇、岗位要求等数据并完成数据展示。

现在要进行大数据分析并完成展示，你作为该公司大数据平台运维工程师，需在单机上安装伪分布式 Hadoop 环境，在该模式下，Hadoop 守护进程运行在一台机器上，模拟一个小规模的集群，伪分布式规划如下表：

表 3-1 伪分布式规划

| 节点 | 虚拟机名 | 主机名 | 节点 | 运行进程 |
|----|------|-----|----|------|
|----|------|-----|----|------|

| 角色       | 称        |        | IP       |                   |
|----------|----------|--------|----------|-------------------|
| 单一<br>节点 | Hadoop01 | Node01 | 自动分<br>配 | NameNode          |
|          |          |        |          | DataNode          |
|          |          |        |          | SecondaryNameNode |

本环节需要使用虚拟机软件 VMware 自行创建虚拟机，并完成 Hadoop 伪分布式模式部署，所需安装包均位于“D:\大数据平台搭建与运维”下，主要任务如下：

#### 任务一：虚拟机的安装与 SSH 服务配置（20 分）

- 创建虚拟机过程中，根据机房所提供的操作系统镜像文件选择相应操作系统和版本，截取客户机操作系统和版本选择后的界面，图片保存到考生文件夹下，并命名为“1-1-1 创建虚拟机 1.png”。然后根据题目要求虚拟机命名为 Hadoop01 并存放位置为 D:\VM，若 D 盘下\VM 路径不存在请先创建；截取虚拟机命名设置和虚拟机存放路径设置，图片保存到考生文件夹下，并命名为“1-1-2 创建虚拟机 2.png”。（4 分）
- 虚拟机创建过程中，根据题目要求设置主机名为 Node01，并打开网络连接，将设置主机名界面截图；虚拟机创建完成后，使用相关命令查看主机名，将命令及查看效果截图；图片保存到考生文件夹下，并分别命名为“1-2-1 设置主机名.png”和“1-2-2 查看主机名.png”。（6 分）
- 查看 ssh 服务是否安装，将命令和结果截图，图片保存到考生文件夹下，并命名为“1-3-1 查看 ssh 安装情况.png”；查看 ssh 服务是否启动（若未启动先使用相关命令启动 ssh 服务再查看），将命令和结果截图，图片保存到考生文件夹下，并命名为“1-3-2 查看 ssh 启动情况.png”。（4 分）
- 使用 xshell 远程连接虚拟机，会话名称要求设置为虚拟机的名称，截取设置完成后的新建会话属性界面，连接成功后在 xshell 上显示虚拟机登陆状态并截取登陆状态，图片保存到考生文件夹下，并分别命名为“1-4-1 远程登录 1.png”和“1-4-2 远程登录 2.png”。（3 分）
- 根据要求修改主机 hosts 文件，配置主机名与 ip 地址的映射，截取修改后的 hosts 文件内容；图片保存到考生文件夹下，并命名为“1-5 修改 hosts 文件.png”。（3 分）

## 任务二：JDK、Hadoop 的安装与环境变量配置（20 分）

- 使用 xftp 工具将 JDK、Hadoop 安装包上传到/export/software（此路径需自行创建）文件夹下，使用相关命令查看 JDK、Hadoop 安装包是否上传成功，并截取所使用的命令及查看效果，图片保存到考生文件夹下，并命名为“2-1 上传安装包.png”。（4 分）
- 将 JDK、Hadoop 安装包解压到/export/servers（此路径需自行创建）目录下，并将 JDK 安装包重命名为 jdk，使用相关命令查看 JDK、Hadoop 安装包是否成功解压，截取重命名命令及查看效果，图片保存到考生文件夹下，并命名为“2-2 解压安装包.png”。（4 分）
- 配置 JDK、Hadoop 环境变量，截取配置内容，图片保存到考生文件夹下，并命名为“2-3 配置环境变量.png”。（8 分）
- 使用相关指令验证 JDK、Hadoop 安装是否成功，将命令及结果截图并保存，图片保存到考生文件夹下，并命名为“2-4 验证 jdk 与 hadoop.png”。（4 分）

## 任务三：修改 Hadoop 相关配置文件（20 分）

- Hadoop 的运行需要依赖 JDK，而 hadoop-env.sh 文件为 Hadoop 的运行环境配置文件，请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“3-1 配置 hadoop 运行环境.png”。（4 分）
- core-site.xml 文件为 Hadoop 的核心配置文件，用于定义系统级别的参数，设定 namenode 的主机名及端口，请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“3-2 修改 hadoop 核心配置文件.png”。（8 分）
- hdfs-site.xml 文件为 HDFS 核心配置文件，如文件副本的个数、块大小等，设定 HDFS 存储文件的副本个数默认为 1、SecondaryNameNode 地址和端口，请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“3-3 修改 hdfs 核心配置文件.png”。（8 分）

## 任务四：启动与测试 Hadoop（30 分）

- 执行相关命令格式化 NameNode，截取格式化成功界面，图片保存到考生文件夹下，并命名为“4-1 格式化 namenode.png”。（8 分）

- 使用相关指令启动 hdfs 集群，并使用 jps 进行查看，截取启动命令与 jps 查看结果，图片保存到考生文件夹下，并命名为“4-2 启动 Hadoop 集群.png”。（6 分）
- 使用相关命令关闭防火墙，并截图命名为“4-3-1 关闭防火墙.png”；通过浏览器查看 HDFS 集群情况并截取 web 界面，图片保存到考生文件夹下，并命名为“4-3-2 查看 web 界面.png”。（4 分）
- 使用相关命令在 HDFS 根目录下创建 test 目录，截取 web 界面目录创建情况，图片保存到考生文件夹下，并命名为“4-4 创建目录.png”。（6 分）
- 使用 hdfs shell 命令将 Hadoop 安装目录下的 NOTICE.txt 文档上传到 HDFS 根目录下的 test 目录下，截取 web 界面上文档上传情况，图片保存到考生文件夹下，并命名为“4-5 上传文件.png”。（6 分）

#### 良好的习惯和职业素养（10 分）

- 答题过程中保持安静。（3 分）
- 有问题举手示意，有礼貌。（3 分）
- 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。（4 分）

### （2）实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。计算机安装有 Vmware workstation15, XShell, JDK1.8, 并有 Hadoop2.7.4 安装包和 CentOS7 64 位镜像文件，安装 Office 办公软件。

### （3）考核时量及要求

- 考核时长 180 分钟；
- 请在桌面建立以“学号\_题号”命名的成果文件夹，所有提交文件均放在该目录下，操作过程的截图需按照题目要求，文件夹命名例如：201965110001\_J3-1-1。

### （4）评分细则

表 3-2 任务一评分细则（20 分）

| 序号 | 评分项     | 分值  | 评分细则                                     |
|----|---------|-----|------------------------------------------|
| 1  | 创建新的虚拟机 | 4 分 | 和参考答案一致满分；<br>创建版本为 CentOS 7 64 位，得 1 分； |

|   |              |     |                                                                                                                                                                                                     |
|---|--------------|-----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|   |              |     | <p>考生文件夹中有 1-1-1 创建虚拟机 1.png,得 1 分;<br/>虚拟机命名为 Hadoop01 并存放路径为 D:\VM, 得 1 分;</p> <p>考生文件夹中有 1-1-2 创建虚拟机 2.png,得 1 分。</p>                                                                            |
| 2 | 设置主机名并查看     | 6 分 | <p>和参考答案一致满分;</p> <p>主机名设置正确并打开网络连接, 得 2 分;</p> <p>考生文件夹中有 1-2-1 设置主机名.png 文件, 得 1 分;</p> <p>正确查看主机名 (有 hostname 命令), 得 2 分;</p> <p>考生文件夹中有 1-2-2 查看主机名.png 文件, 得 1 分。</p>                          |
| 3 | 查看并启动 ssh 服务 | 4 分 | <p>和参考答案一致满分;</p> <p>查看 ssh 服务是否安装命令正确, 得 1 分;</p> <p>考生文件夹中有 1-3-1 查看 ssh 安装情况.png 文件, 得 1 分;</p> <p>查看 ssh 服务是否启动命令正确 (若未启动, 则使用启动命令), 得 1 分;</p> <p>考生文件夹中有 1-3-2 查看 ssh 启动情况.png 文件, 得 1 分。</p> |
| 4 | 远程登录         | 3 分 | <p>和参考答案一致满分;</p> <p>新建远程连接, 会话名称和 ip 地址均要求正确, 得 1 分;</p> <p>xshell 成功显示登录状态, 得 1 分;</p> <p>考生文件夹中有 1-4-1 远程登录 1.png 和 1-4-2 远程登录 2.png 文件, 得 1 分。</p>                                              |
| 5 | 配置主机名与 ip 映射 | 3 分 | <p>和参考答案一致满分;</p> <p>正确打开/etc/hosts, 配置主机名与 ip 地址映射正确, 得 2 分;</p> <p>考生文件夹中有 1-5 修改 hosts 文件.png 文件, 得 1 分</p>                                                                                      |

表 3-3 任务二评分细则 (20 分)

| 序号 | 评分项      | 分值  | 评分细则                                                                                          |
|----|----------|-----|-----------------------------------------------------------------------------------------------|
| 1  | 上传到指定路径下 | 4 分 | <p>和参考答案一致满分;</p> <p>查看 JDK 和 Hadoop 安装包上传情况, 得 2 分;</p> <p>考生文件夹中有 2-1 上传安装包.png, 得 2 分。</p> |
| 2  | 解压到指定路径下 | 4 分 | <p>和参考答案一致满分;</p> <p>重命名 JDK, 得 1 分;</p> <p>查看 JDK 和 Hadoop 安装包解压情况, 得 1.5 分;</p>             |

|   |                   |     |                                                                                                                        |
|---|-------------------|-----|------------------------------------------------------------------------------------------------------------------------|
|   |                   |     | 考生文件夹中有 2-2 解压安装包.png 文件，得 1.5 分。                                                                                      |
| 3 | 配置环境变量            | 8 分 | 和参考答案一致满分；<br>正确配置环境变量，得 6 分；<br>考生文件夹中有 2-3 配置环境变量.png 文件，得 2 分。                                                      |
| 4 | 能够正确显示 Hadoop 的版本 | 4 分 | 和参考答案一致满分；<br>有 java -version 命令，得 1.5 分；<br>有 hadoop version 命令，得 1.5 分；<br>考生文件夹中有 2-4 验证 jdk 与 hadoop.png 文件，得 1 分。 |

表 3-4 任务三评分细则（20 分）

| 序号 | 评分项                   | 分值  | 评分细则                                                                                     |
|----|-----------------------|-----|------------------------------------------------------------------------------------------|
| 1  | 修改 hadoop-env.sh 配置文件 | 4 分 | 和参考答案一致满分；<br>hadoop-env.sh 配置正确（路径配置正确），得 2 分；<br>考生文件夹中有 3-1 配置 hadoop 运行环境.png，得 2 分。 |
| 2  | 修改 core-site.xml 配置文件 | 8 分 | 和参考答案一致满分；<br>正确配置 ip 地址及端口号，得 6 分；<br>考生文件夹中有 3-2 修改 hadoop 核心配置文件.png 文件，得 2 分。        |
| 3  | 修改 hdfs-site.xml 配置文件 | 8 分 | 和参考答案一致满分；<br>正确配置 hdfs-site.xml 文件，得 6 分；<br>考生文件夹中有 3-3 修改 hdfs 核心配置文件.png 文件，得 2 分。   |

表 3-5 任务四评分细则（30 分）

| 序号 | 评分项                  | 分值  | 评分细则                                                                        |
|----|----------------------|-----|-----------------------------------------------------------------------------|
| 1  | 格式化 NameNode         | 8 分 | 和参考答案一致满分；<br>格式化成功，得 6 分；<br>考生文件夹中有 4-1 格式化 namenode.png 文件，得 2 分。        |
| 2  | 启动 hdfs 集群并使用 jps 查看 | 6 分 | 和参考答案一致满分；<br>成功启动 hdfs 集群，得 4 分；<br>考生文件夹中有 4-2 启动 Hadoop 集群.png 文件，得 2 分。 |
| 3  | 查看 web 界面            | 4 分 | 和参考答案一致满分；<br>关闭防火墙命令正确，得 2 分；                                              |

|   |            |     |                                                                                    |
|---|------------|-----|------------------------------------------------------------------------------------|
|   |            |     | 考生文件夹中有 4-3-1 关闭防火墙.png 和 4-3-2 查看 web 界面.png 文件，得 2 分。                            |
| 4 | web 界面创建目录 | 6 分 | 和参考答案一致满分；<br>HDFS 根目录下创建目录正确，得 4 分；<br>考生文件夹中有 4-4 创建目录.png 文件，得 2 分。             |
| 5 | web 界面上传文档 | 6 分 | 和参考答案一致满分；<br>正确上传文档到 HDFS 根目录下的 test 目录下，得 4 分；<br>考生文件夹中有 4-5 上传文件.png 文件，得 2 分。 |

表 3-6 职业素质评分细则（10 分）

| 序号 | 评分项  | 分值  | 评分细则                              |
|----|------|-----|-----------------------------------|
| 1  | 操作规范 | 3 分 | 答题过程中保持安静，不损坏机器；                  |
| 2  | 文明考试 | 3 分 | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌；      |
| 3  | 工位整理 | 4 分 | 答题完毕立场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |

#### 4.试题编号：J2-1-4：大麦网大数据平台 Hadoop 伪分布式模式部署

##### （1）任务描述

大麦网平台需为消费者提供便捷的在线订购票服务，通过结合大数据高性能的分析，帮助消费者提前查找某地区的演唱会、话剧、音乐节等数据并完成数据展示。

现在要进行大数据分析并完成展示，你作为该公司大数据平台运维工程师，需在单机上安装伪分布式 Hadoop 环境，在该模式下，Hadoop 守护进程运行在一台机器上，模拟一个小规模集群，伪分布式规划如下表：

表 4-1 伪分布式规划

| 节点角色 | 虚拟机名称    | 主机名    | 节点 IP | 运行进程                                      |
|------|----------|--------|-------|-------------------------------------------|
| 单一节点 | Hadoop01 | Node01 | 自动分配  | NameNode<br>DataNode<br>SecondaryNameNode |

本环节需要使用虚拟机软件 VMware 自行创建虚拟机，并完成 Hadoop 伪分布式模式部

署，所需安装包均位于“D:\大数据平台搭建与运维”下，主要任务如下：

**任务一：创建虚拟机与设置远程登录（20 分）**

- 创建虚拟机并根据题目要求设置主机名为 Node01，打开网络连接，将设置主机名界面截图；虚拟机创建完成后，使用相关命令查看主机名，将命令及查看效果截图；图片保存到考生文件夹下，并分别命名为“1-1-1 设置主机名.png”和“1-1-2 查看主机名.png”。（6 分）
- 查看 ssh 服务是否安装，将命令和结果截图，图片保存到考生文件夹下，并命名为“1-2-1 查看 ssh 安装情况.png”；查看 ssh 服务是否启动（若未启动先使用相关命令启动 ssh 服务），将命令和结果截图，图片保存到考生文件夹下，并命名为“1-2-2 查看 ssh 启动情况.png”。（6 分）
- 使用 xshell 远程连接虚拟机，会话名称要求设置为虚拟机的名称，截取设置完成后的新建会话属性界面，连接成功后在 xshell 上显示虚拟机登陆状态并截取登陆状态，图片保存到考生文件夹下，并分别命名为“1-3-1 远程登录 1.png”和“1-3-2 远程登录 2.png”。（6 分）
- 根据要求修改主机 hosts 文件，配置主机名与 ip 地址的映射，截取 hosts 文件修改后的内容；图片保存到考生文件夹下，并命名为“1-4 修改 hosts 文件.png”。（2 分）

**任务二：JDK、Hadoop 的安装与环境变量配置（20 分）**

- 使用 xftp 工具将 JDK、Hadoop 安装包上传到/export/software（此路径需自行创建）文件夹下，使用相关命令查看 JDK、Hadoop 安装包是否上传成功，并截取所使用的命令及查看效果，图片保存到考生文件夹下，并命名为“2-1 上传安装包.png”。（4 分）
- 将 JDK、Hadoop 安装包解压到/export/servers（此路径需自行创建）目录下，并将 JDK 安装包重命名为 jdk，使用相关命令查看 JDK、Hadoop 安装包是否成功解压，截取重命名命令及查看效果，图片保存到考生文件夹下，并命名为“2-2 解压安装包.png”。（4 分）
- 配置 JDK、Hadoop 环境变量，截取配置内容，图片保存到考生文件夹下，并命名为“2-3 配置环境变量.png”。（6 分）

- 使用相关指令验证 JDK、Hadoop 安装是否成功，将命令及结果截图并保存，图片保存到考生文件夹下，并命名为“2-4 验证 jdk 与 hadoop.png”。(6 分)

### 任务三：修改 Hadoop 相关配置文件 (20 分)

- Hadoop 的运行需要依赖 JDK，而 `hadoop-env.sh` 文件为 Hadoop 的运行环境配置文件，请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“3-1 配置 hadoop 运行环境.png”。(4 分)
- `core-site.xml` 文件为 Hadoop 的核心配置文件，用于定义系统级别的参数，设定 `namenode` 的主机名及端口、存放临时文件的目录，请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“3-2 修改 hadoop 核心配置文件.png”。(8 分)
- `hdfs-site.xml` 文件为 HDFS 核心配置文件，如文件副本的个数、块大小及是否使用强制权限等，设定 HDFS 存储文件的副本个数默认为 1、`SecondaryNameNode` 地址和端口，请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“3-3 修改 hdfs 核心配置文件.png”。(8 分)

### 任务四：启动与测试 Hadoop (30 分)

- 格式化 `NameNode` 后，使用相关指令启动 `hdfs` 集群，并使用 `jps` 进行查看，截取启动命令与 `jps` 查看结果，图片保存到考生文件夹下，并命名为“4-1 启动 Hadoop 集群.png”。(5 分)
- 使用相关命令关闭防火墙，并截图命名为“4-2-1 关闭防火墙.png”；通过浏览器查看 HDFS 集群情况并截取 web 界面，图片保存到考生文件夹下，并命名为“4-2-2 查看 web 界面.png”。(8 分)
- 使用相关命令在 HDFS 根目录下创建 `input` 目录，截取 web 界面目录创建情况，图片保存到考生文件夹下，并命名为“4-3 创建目录.png”。(4 分)
- 使用 `hdfs shell` 命令将 Hadoop 安装目录下的 `README.txt` 文档上传到 HDFS 根目录下的 `input` 目录下，截取 web 界面上文档上传情况，图片保存到考生文件夹下，并命名为“4-4 上传文件.png”。(5 分)
- 使用 `wordcount` 模块对刚上传到 HDFS 的文档进行词频统计，截取词频统计结果，图片

保存到考生文件夹下，并命名为“4-5 词频统计.png”。（8 分）

### 良好的习惯和职业素养（10 分）

- 答题过程中保持安静。（3 分）
- 有问题举手示意，有礼貌。（3 分）
- 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。（4 分）

### （2）实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。

计算机安装有 Vmware workstation15, XShell, JDK1.8, 并有 Hadoop2.7.4 安装包和 CentOS7 64 位镜像文件，安装 Office 办公软件。

### （3）考核时量及要求

- 考核时长 180 分钟；
- 请在桌面建立以“学号\_题号”命名的成果文件夹，所有提交文件均放在该目录下，操作过程的截图需按照题目要求，文件夹命名例如：201965110001\_J3-1-1。

### （4）评分细则

表 4-2 任务一评分细则（20 分）

| 序号 | 评分项           | 分值  | 评分细则                                                                                                                                                                        |
|----|---------------|-----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 创建新的虚拟机并查看主机名 | 6 分 | 和参考答案一致满分；<br>创建虚拟机正确设置主机名并打开网络连接，得 2 分；<br>考生文件夹中有 1-1-1 设置主机名.png，得 1 分；<br>正确查看主机名，得 2 分；<br><br>考生文件夹中有 1-1-2 查看主机名.png，得 1 分。                                          |
| 2  | 查看并启动 ssh 服务  | 6 分 | 和参考答案一致满分；<br>查看 ssh 服务是否安装命令正确，得 2 分；<br>考生文件夹中有 1-2-1 查看 ssh 安装情况.png 文件，得 1 分；<br>查看 ssh 服务是否启动命令正确（若未启动，则先使用启动命令），得 2 分；<br><br>考生文件夹中有 1-2-2 查看 ssh 启动情况.png 文件，得 1 分。 |
| 3  | 远程登录          | 6 分 | 和参考答案一致满分；<br>新建远程连接，会话名称和 ip 地址均要求正确，得 2 分；<br><br>xshell 成功显示登录状态，得 2 分；                                                                                                  |

|   |              |     |                                                                                               |
|---|--------------|-----|-----------------------------------------------------------------------------------------------|
|   |              |     | 考生文件夹中有 1-3-1 远程登录 1.png 和 1-3-2 远程登录 2.png 文件，得 2 分。                                         |
| 4 | 配置主机名与 ip 映射 | 2 分 | 和参考答案一致满分；<br>正确打开/etc/hosts，配置主机名与 ip 地址映射正确，得 1 分；<br>考生文件夹中有 1-4 修改 hosts 文件.png 文件，得 1 分。 |

表 4-3 任务二评分细则（20 分）

| 序号 | 评分项               | 分值  | 评分细则                                                                                                               |
|----|-------------------|-----|--------------------------------------------------------------------------------------------------------------------|
| 1  | 上传到指定路径下          | 4 分 | 和参考答案一致满分；<br>查看 JDK 和 Hadoop 安装包上传情况，得 2 分；<br>考生文件夹中有 2-1 上传安装包.png，得 2 分。                                       |
| 2  | 解压到指定路径下          | 4 分 | 和参考答案一致满分；<br>重命名 JDK，得 1 分；<br>查看 JDK 和 Hadoop 安装包解压情况，得 1.5 分；<br>考生文件夹中有 2-2 解压安装包.png 文件，得 1.5 分。              |
| 3  | 配置环境变量            | 6 分 | 和参考答案一致满分；<br>正确配置环境变量，得 4 分；<br>考生文件夹中有 2-3 配置环境变量.png 文件，得 2 分。                                                  |
| 4  | 能够正确显示 Hadoop 的版本 | 6 分 | 和参考答案一致满分；<br>有 java -version 命令，得 2 分；<br>有 hadoop version 命令，得 2 分；<br>考生文件夹中有 2-4 验证 jdk 与 hadoop.png 文件，得 2 分。 |

表 4-4 任务三评分细则（20 分）

| 序号 | 评分项 | 分值 | 评分细则 |
|----|-----|----|------|
|----|-----|----|------|

|   |                       |     |                                                                                          |
|---|-----------------------|-----|------------------------------------------------------------------------------------------|
| 1 | 修改 hadoop-env.sh 配置文件 | 4 分 | 和参考答案一致满分；<br>hadoop-env.sh 配置正确（路径配置正确），得 3 分；<br>考生文件夹中有 3-1 配置 hadoop 运行环境.png，得 1 分。 |
| 2 | 修改 core-site.xml 配置文件 | 8 分 | 和参考答案一致满分；<br>正确配置 ip 地址及端口号；<br>考生文件夹中有 3-2 修改 hadoop 核心配置文件.png 文件，得 2 分。              |
| 3 | 修改 hdfs-site.xml 配置文件 | 8 分 | 和参考答案一致满分；<br>正确配置 hdfs-site.xml 文件，得 6 分；<br>考生文件夹中有 3-3 修改 hdfs 核心配置文件.png 文件，得 2 分。   |

表 4-5 任务四评分细则（30 分）

| 序号 | 评分项                    | 分值  | 评分细则                                                                                        |
|----|------------------------|-----|---------------------------------------------------------------------------------------------|
| 1  | 格式化NameNode并启动 hdfs 集群 | 5 分 | 和参考答案一致满分；<br>格式化成功，得 2 分；<br>成功启动 hdfs 集群，得 2 分；<br>考生文件夹中有 4-1 启动 Hadoop 集群.png 文件，得 1 分。 |
| 2  | 查看 web 界面              | 8 分 | 和参考答案一致满分；<br>关闭防火墙命令正确，得 4 分；<br>考生文件夹中有 4-2-1 关闭防火墙.png 和 4-2-2 查看 web 界面.png 文件，得 4 分。   |
| 3  | web 界面创建目录             | 4 分 | 和参考答案一致满分；<br>HDFS 根目录下创建目录正确，得 2 分；<br>考生文件夹中有 4-3 创建目录.png 文件，得 2 分。                      |
| 4  | web 界面上传文档             | 5 分 | 和参考答案一致满分；<br>正确上传文档到 HDFS 根目录下的 input 目录下，得 3 分；<br>考生文件夹中有 4-4 上传文件.png 文件，得 2 分。         |
| 5  | 词频统计                   | 8 分 | 和参考答案一致满分；<br>正确统计文档词频，得 6 分；<br>考生文件夹中有 4-5 词频统计.png 文件，得 2 分。                             |

表 4-6 职业素质评分细则（10 分）

| 序号 | 评分项  | 分值  | 评分细则                              |
|----|------|-----|-----------------------------------|
| 1  | 操作规范 | 3 分 | 答题过程中保持安静，不损坏机器；                  |
| 2  | 文明考试 | 3 分 | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌；      |
| 3  | 工位整理 | 4 分 | 答题完毕立场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |

## 项目 2：Hadoop 完全分布式模式部署

### 5.试题编号：J2-2-1：豆瓣网 Hadoop 完全分布式模式部署

#### （1）任务描述

豆瓣网作为一家影视作品评分平台，现在需要通过各个评分网站查询到每部影视作品的详细信息，帮助使用者查找到影视作品名、主演、上映时间和评分等数据内容并完成数据展示。

现在要进行大数据分析并完成展示，你作为该公司大数据平台运维工程师，需在 Hadoop 伪分布式模式基础上部署完全分布式 Hadoop 环境，JDK 安装包、Hadoop 安装包所在路径为：/export/software，JDK 和 Hadoop 解压包所在路径为：/export/servers。完全分布式规划如下表：

表 5-1 完全分布式规划

| 节点角色 | 虚拟机名     | IP 地址               | 主机名    | 运行进程                                         |
|------|----------|---------------------|--------|----------------------------------------------|
| 主节点  | Hadoop01 | 192.168.233.13<br>0 | master | NameNode<br>ResoureManager                   |
| 从节点  | Hadoop02 | 192.168.233.13<br>1 | slave1 | DataNode<br>NodeManager<br>SecondaryNameNode |
|      | Hadoop03 | 192.168.233.13<br>2 | slave2 | DataNode<br>NodeManager                      |

主要任务如下：

#### 任务一：规划完全分布式模式下的 Hadoop 集群（30 分）

- 对已有虚拟机 Hadoop01（用户名：root 密码：123456）进行克隆，得到 Hadoop02 和 Hadoop03 虚拟机，截取克隆后得到的三台虚拟机；图片保存到考生文件夹下，并命名为“1-1 克隆虚拟机.png”。（4 分）
- 开启三台虚拟机并按集群规划修改主机名，使用相关命令显示两台从节点修改后的主机

名，截取命令及显示效果，图片保存到考生文件夹下，并命名为“1-2-1 修改主机名 1”、“1-2-2 修改主机名 2”。（6 分）

- 在三台主机上均配置集群各主机的主机名和 IP 地址映射关系，截取任意一台主机的 IP 映射配置内容，图片保存到考生文件夹下，并命名为“1-3 配置 IP 映射.png”。（4 分）
- 根据题目要求设置虚拟机的 IP，修改虚拟机网卡配置文件，将 slave1 主机修改后的内容截图，图片保存到考生文件夹下，并命名为“1-4 修改 ip 地址.png”。（10 分）
- 重启网络服务后，使用相关命令查看各主机修改后的 IP 地址，截取 slave1 的查看效果，图片保存到考生文件夹下，并命名为“1-5 查看修改后的 ip 地址.png”。（6 分）

## 任务二：远程登录和免密登录设置（20 分）

- 先查看 ssh 服务是否安装，将命令和结果截图，图片保存到考生文件夹下，并命名为“2-1-1 查看 ssh 安装情况.png”，再查看 ssh 服务是否启动（若未启动先使用相关命令启动 ssh 服务），将命令和结果截图，图片保存到考生文件夹下，并命名为“2-1-2 查看 ssh 启动情况.png”。（5 分）
- 使用 XShell 远程连接三台主机，将连接成功后的界面截图，图片保存到考生文件夹下，并命名为“2-2 远程登录.png”。（2 分）
- 在 master 上生成密钥文件，将命令截图，图片保存到考生文件夹下，并命名为“2-3-1 生成密钥文件.png”；将 master 的公钥复制到 master、slave1、slave2，将复制到 slave1 的操作命令截图，图片保存到考生文件夹下，并命名为“2-3-2 复制密钥.png”。（3 分）
- 在 slave1 上生成密钥文件，将命令截图，图片保存到考生文件夹下，并命名为“2-4-1 生成密钥文件.png”；将 slave1 的公钥复制至 master、slave1、slave2，将复制到 master 的操作命令截图，图片保存到考生文件夹下，并命名为“2-4-2 复制密钥.png”。（3 分）
- 在 slave2 上生成密钥文件，将命令截图，图片保存到考生文件夹下，并命名为“2-5-1 生成密钥文件.png”；将 slave2 的公钥复制至 master、slave1、slave2，将复制到 master 的操作命令截图，图片保存到考生文件夹下，并命名为“2-5-2 复制密钥.png”。（3 分）
- 测试免密登录是否成功，在 master 节点上使用相关命令免密登录到 slave1、然后 slave1 免密登录 slave2、slave2 免密登录 master，截取命令和结果，图片保存到考生文件夹下，

并命名为“2-6 测试免密登录.png”。(4 分)

### 任务三：集群主节点的配置 (40 分)

- `hadoop-env.sh` 文件为配置 Hadoop 的环境变量，提供了 Hadoop 中 `JAVA_HOME` 的运行环境。请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“3-1 修改 `hadoop-env.sh` 文件.png”。(4 分)
- `core-site.xml` 文件为 Hadoop 的核心配置文件，用于定义系统级别的参数，设定 `namenode` 的主机名及端口、存放临时文件的目录请设定为 `/export/servers/hadoop-2.7.4/tmp`，请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“3-2 修改 `hadoop` 核心配置文件.png”。(8 分)
- `hdfs-site.xml` 文件为 HDFS 核心配置文件，如文件副本的个数、块大小及是否使用强制权限等，设定 HDFS 存储文件的副本个数为 2、`SecondaryNameNode` 地址和端口，请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“3-3 修改 `hdfs` 核心配置文件.png”。(8 分)
- `mapred-site.xml` 文件是 MapRduce 的核心配置文件，用于指定 MapRduce 运行时的框架，需要对已有 `mapred-site.xml.template` 文件进行复制并重命名得到该文件，请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“3-4 修改 MapReduce 核心文件.png”。(8 分)
- `yarn-site.xml` 文件是 YARN 框架的核心配置文件，需要指定 YARN 集群的管理者，请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“3-5 修改 `yarn` 核心配置文件.png”。(8 分)
- `slaves` 文件用于记录 Hadoop 集群所有从节点（HDFS 的 `DataNode` 和 YARN 的 `NodeManager` 所在主机），根据题目要求进行配置并提交该配置文件的内容截图，图片保存到考生文件夹下，并命名为“3-6 修改 `slaves` 文件.png”。(4 分)

### 良好的习惯和职业素养 (10 分)

- 答题过程中保持安静。(3 分)
- 有问题举手示意，有礼貌。(3 分)

- 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。（4分）

## （2）实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。

计算机安装有 Vmware workstation15, XShell, JDK1.8, 并有 Hadoop2.7.4 安装包和 CentOS7 64 位镜像文件，安装 Office 办公软件。

## （3）考核时量及要求

- 考核时长 180 分钟；
- 请在桌面建立以“学号\_题号”命名的成果文件夹，所有提交文件均放在该目录下，操作过程的截图需按照题目要求，文件夹命名例如：201965110001\_J3-1-1。

## （4）评分细则

表 5-2 任务一评分细则（20 分）

| 序号 | 评分项          | 分值   | 评分细则                                                                                           |
|----|--------------|------|------------------------------------------------------------------------------------------------|
| 1  | 正确克隆虚拟机      | 4 分  | 和参考答案一致满分；<br>正确克隆两台虚拟机命名为 Hadoop02 和 Hadoop03，得 2 分；<br>考生文件夹中有 1-1 克隆虚拟机.png 文件，得 2 分。       |
| 2  | 正确修改主机名      | 6 分  | 和参考答案一致满分；<br>两台从节点修改主机名正确，得 3 分；<br>考生文件夹中有 1-2-1 修改主机名 1.png 文件和 1-2-2 修改主机名 2.png 文件，得 3 分。 |
| 3  | 正确配置 IP 映射   | 4 分  | 和参考答案一致满分；<br>正确配置 ip 映射关系，得 2 分；<br>考生文件夹中有 1-3 配置 IP 映射.png 文件，得 2 分。                        |
| 4  | 正确修改 IP 地址文件 | 10 分 | 和参考答案一致满分；<br>正确打开虚拟机网卡配置文件，得 2 分；<br>配置文件内容正确，得 6 分；<br>考生文件夹中有 1-4 修改 ip 地址.png 文件，得 2 分。    |
| 5  | 重启网络服务正确     | 6 分  | 和参考答案一致满分；<br>重启网络服务命令正确，得 3 分；                                                                |

|  |  |  |                                                                                         |
|--|--|--|-----------------------------------------------------------------------------------------|
|  |  |  | <p>有 ip addr 或 ip addr show 命令，得 1 分；</p> <p>考生文件夹中有 1-5 查看修改后的 ip 地址.png 文件，得 2 分。</p> |
|--|--|--|-----------------------------------------------------------------------------------------|

表 5-3 任务二评分细则（20 分）

| 序号 | 评分项            | 分值  | 评分细则                                                                                                                                                                                               |
|----|----------------|-----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 查看 ssh 服务      | 5 分 | <p>和参考答案一致满分；</p> <p>查看 ssh 服务是否安装命令正确，得 1.5 分；</p> <p>考生文件夹中有 2-1-1 查看 ssh 安装情况.png 文件，得 1 分；</p> <p>查看 ssh 服务是否启动命令正确（若未启动，则先使用启动命令），得 1.5 分；</p> <p>考生文件夹中有 2-1-2 查看 ssh 启动情况.png 文件，得 1 分。</p> |
| 2  | 远程登录           | 2 分 | <p>和参考答案一致满分；</p> <p>考生文件夹中有 2-2 远程登录.png 文件，得 2 分。</p>                                                                                                                                            |
| 3  | master 上生成密钥文件 | 3 分 | <p>和参考答案一致满分；</p> <p>成功生成密钥，得 1 分；</p> <p>公钥复制命令正确，得 1 分；</p> <p>考生文件夹中有 2-3-1 生成密钥文件.png 和 2-3-2 复制密钥.png 文件，得 1 分。</p>                                                                           |
| 4  | slave1 上生成密钥文件 | 3 分 | <p>和参考答案一致满分；</p> <p>成功生成密钥，得 1 分；</p> <p>公钥复制命令正确，得 1 分；</p> <p>考生文件夹中有 2-4-1 生成密钥文件.png 和 2-4-2 复制密钥.png 文件，得 1 分。</p>                                                                           |
| 5  | slave2 上生成密钥文件 | 3 分 | <p>和参考答案一致满分；</p> <p>成功生成密钥，得 1 分；</p> <p>公钥复制命令正确，得 1 分；</p> <p>考生文件夹中有 2-5-1 生成密钥文件.png 和 2-5-2 复制密钥.png 文件，得 1 分。</p>                                                                           |
| 6  | 测试免密登录         | 4 分 | <p>和参考答案一致满分；</p> <p>成功使用免密登录进入三个节点，得 2 分；</p> <p>考生文件夹中有 2-6 测试免密登录.png 文件，得 2 分。</p>                                                                                                             |

表 5-4 任务三评分细则（50 分）

| 序号 | 评分项                     | 分值  | 评分细则                                                                                                                                                        |
|----|-------------------------|-----|-------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 修改 hadoop-env.sh 配置文件   | 4 分 | 和参考答案一致满分；<br>hadoop-env.sh 配置正确（路径配置正确），得 2 分；<br>考生文件夹中有 3-1 修改 hadoop-env.sh 文件.png，得 2 分。                                                               |
| 2  | 修改 core-site.xml 配置文件   | 8 分 | 和参考答案一致满分；<br>正确配置 namenode 主机名及端口，得 3 分；<br>正确配置临时文件的目录，得 3 分；<br>考生文件夹中有 3-2 修改 hadoop 核心配置文件.png 文件，得 2 分。                                               |
| 3  | 修改 hdfs-site.xml 配置文件   | 8 分 | 和参考答案一致满分；<br>正确配置 hdfs-site.xml 文件，得 6 分；<br>考生文件夹中有 3-3 修改 hdfs 核心配置文件.png 文件，得 2 分。                                                                      |
| 4  | 修改 mapred-site.xml 配置文件 | 8 分 | 和参考答案一致满分；<br>成功复制 mapred-site.xml.template 并重命名为 mapred-site.xml，得 2 分；<br>正确配置 mapred-site.xml 文件内容，得 4 分；<br>考生文件夹中有 3-4 修改 MapReduce 核心文件.png 文件，得 2 分。 |
| 5  | 修改 yarn-site.xml 配置文件   | 8 分 | 和参考答案一致满分；<br>正确配置 yarn-site.xml 文件，得 6 分；<br>考生文件夹中有 3-5 修改 yarn 核心配置文件.png 文件，得 2 分。                                                                      |
| 6  | 修改 slaves 配置文件          | 4 分 | 和参考答案一致满分；<br>正确配置 slaves 文件内容，得 2 分；<br>考生文件夹中有 3-6 修改 slaves 文件.png 文件，得 2 分。                                                                             |

表 5-5 职业素质评分细则（10 分）

| 序号 | 评分项  | 分值  | 评分细则                         |
|----|------|-----|------------------------------|
| 1  | 操作规范 | 3 分 | 答题过程中保持安静，不损坏机器；             |
| 2  | 文明考试 | 3 分 | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌； |

|   |      |     |                                   |
|---|------|-----|-----------------------------------|
| 3 | 工位整理 | 4 分 | 答题完毕立场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |
|---|------|-----|-----------------------------------|

## 6.试题编号：J2-2-2：起点网 Hadoop 完全分布式模式部署

### （1）任务描述

起点网是一家网络文学综合服务平台。现在需要通过各个小说网站查询到每部网络小说的详细信息，帮助使用者查找到小说名、作者、月票和推荐票等数据内容并完成数据展示。

现在要进行大数据分析并完成展示，你作为该公司大数据平台运维工程师，需在 Hadoop 伪分布式模式基础上部署完全分布式 Hadoop 环境，JDK 安装包、Hadoop 安装包所在路径为：/export/software，JDK 和 Hadoop 解压包所在路径为：/export/servers。完全分布式规划如下表：

表 6-1 完全分布式规划

| 节点角色 | 虚拟机名     | IP 地址           | 主机名    | 运行进程                                         |
|------|----------|-----------------|--------|----------------------------------------------|
| 主节点  | Hadoop01 | 192.168.233.130 | master | NameNode<br>ResoureManager                   |
| 从节点  | Hadoop02 | 192.168.233.131 | slave1 | DataNode<br>NodeManager<br>SecondaryNameNode |
|      | Hadoop03 | 192.168.233.132 | slave2 | DataNode<br>NodeManager                      |

主要任务如下：

#### 任务一：规划完全分布式模式下的 Hadoop 集群（30 分）

- 通过对已有虚拟机 Hadoop01（用户名：root 密码：123456）进行克隆，得到 Hadoop02 和 Hadoop03 虚拟机，截取克隆后得到的三台虚拟机；图片保存到考生文件夹下，并命名为“1-1 克隆虚拟机.png”。（4 分）
- 按集群规划修改主机名，分别截取修改后两台从节点的主机名，并命名为“1-2-1 修改

主机名 1”、“1-2-2 修改主机名 2”，图片保存到考生文件夹下。（4 分）

- 根据题目要求在相关文件下配置 IP 映射，要求三台主机均进行 IP 映射配置，并截取主节点的 IP 映射配置内容，图片保存到考生文件夹下，并命名为“1-3 配置 IP 映射.png”。（4 分）
- 根据题目要求设置三台主机的 IP 地址，修改虚拟机网卡配置文件，将 slave1 的修改内容截图，图片保存到考生文件夹下，并命名为“1-4 修改 ip 地址.png”。（8 分）
- 完成上述操作后重启网络服务，使用相关命令查看 IP 配置是否生效，截取命令及结果，图片保存到考生文件夹下，并命名为“1-5 查看修改后的 ip.png”。（6 分）
- 使用相关命令检测三台主机之间是否相互连通，截取主节点 ping 两台从节点的测试情况，图片保存到考生文件夹下，并命名为“1-6 连通测试.png”。（4 分）

## 任务二：修改主节点配置文件（40 分）

- Xshell 连接三台主机后，在主节点上修改相关配置文件，hadoop-env.sh 文件为配置 Hadoop 的环境变量，提供了 Hadoop 中 JAVA\_HOME 的运行环境。请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“2-1 修改 hadoop-env.sh 文件.png”。（4 分）
- core-site.xml 文件为 Hadoop 的核心配置文件，用于定义系统级别的参数，设定 namenode 的主机名及端口、存放临时文件的目录请设定为 /export/servers/hadoop-2.7.4/tmp，请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“2-2 修改 hadoop 核心配置文件.png”。（8 分）
- hdfs-site.xml 文件为 HDFS 核心配置文件，如文件副本的个数、块大小及是否使用强制权限等，设定 HDFS 存储文件的副本个数为 2、SecondaryNameNode 地址和端口，请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“2-3 修改 hdfs 核心配置文件.png”。（8 分）
- mapred-site.xml 文件是 MapReduce 的核心配置文件，用于指定 MapReduce 运行时的框架，需要对已有 mapred-site.xml.template 文件进行复制并重命名得到该文件，请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“2-4 修改 MapReduce 核心

文件.png”。(8分)

- yarn-site.xml 文件是 YARN 框架的核心配置文件，需要指定 YARN 集群的管理者，请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“2-5 修改 yarn 核心配置文件.png”。(8分)
- slaves 文件用于记录 Hadoop 集群所有从节点（HDFS 的 DataNode 和 YARN 的 NodeManager 所在主机），根据题目要求进行配置并提交该配置文件的内容截图，图片保存到考生文件夹下，并命名为“2-6 修改 slaves 文件.png”。(4分)

### 任务三：分发配置文件与测试 Hadoop 集群（20分）

- 由于所修改的配置文件都存在于 Hadoop 安装目录下的 etc/hadoop 目录下，因此将主节点上的 Hadoop 安装目录下的 etc/hadoop 目录分发到两个从节点即可，将分发至从节点 slave1 的命令截图，图片保存到考生文件夹下，并命名为“3-1 分发 hadoop 配置文件.png”。(4分)
- 在主节点上执行相关命令格式化 NameNode，截取格式化成功界面，图片保存到考生文件夹下，并命名为“3-2 格式化 namenode.png”。(8分)
- 使用相关指令分别启动 hdfs 和 yarn 集群，并使用 jps 进行查看，截取启动命令与 master 上的 jps 查看结果，图片保存到考生文件夹下，并命名为“3-3-1 启动 Hadoop 集群.png”；截取 slave1 上 jps 查看结果，图片保存到考生文件夹下，并命名为“3-3-2 查看 slave1 启动情况.png”。(8分)

### 良好的习惯和职业素养（10分）

- 答题过程中保持安静。(3分)
- 有问题举手示意，有礼貌。(3分)
- 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。(4分)

## （2）实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。计算机安装有 Vmware workstation15, XShell, JDK1.8, 并有 Hadoop2.7.4 安装包和 CentOS7 64 位镜像文件，安装 Office 办公软件。

### (3) 考核时量及要求

- 考核时长 180 分钟；
- 请在桌面建立以“学号\_题号”命名的成果文件夹，所有提交文件均放在该目录下，操作过程的截图需按照题目要求，文件夹命名例如：201965110001\_J3-1-1。

### (4) 评分细则

表 6-2 任务一评分细则（30 分）

| 序号 | 评分项          | 分值  | 评分细则                                                                                                          |
|----|--------------|-----|---------------------------------------------------------------------------------------------------------------|
| 1  | 正确克隆虚拟机      | 4 分 | 和参考答案一致满分；<br>正确克隆两台虚拟机命名为 Hadoop02 和 Hadoop03，得 2 分；<br>考生文件夹中有 1-1 克隆虚拟机.png 文件，得 2 分。                      |
| 2  | 正确修改主机名      | 4 分 | 和参考答案一致满分；<br>两台从节点修改主机名正确，得 2 分；<br>考生文件夹中有 1-2-1 修改主机名 1.png 文件和 1-2-2 修改主机名 2.png 文件，得 2 分。                |
| 3  | 正确配置 IP 映射   | 4 分 | 和参考答案一致满分；<br>正确配置 ip 映射关系，得 2 分；<br>考生文件夹中有 1-3 配置 IP 映射.png 文件，得 2 分。                                       |
| 4  | 正确修改 IP 地址文件 | 8 分 | 和参考答案一致满分；<br>正确打开虚拟机网卡配置文件，得 2 分；<br>配置文件内容正确，得 4 分；<br>考生文件夹中有 1-4 修改 ip 地址.png 文件，得 2 分。                   |
| 5  | 重启网络服务正确     | 6 分 | 和参考答案一致满分；<br>重启网络命令正确，得 3 分；<br>有 ip addr 或 ip addr show 命令，得 1 分；<br>考生文件夹中有 1-5 查看修改后的 ip 地址.png 文件，得 2 分。 |
| 6  | 连通测试         | 4 分 | 和参考答案一致满分；<br>三台主机之间连通成功，得 2 分；<br>考生文件夹中有 1-6 连通测试.png 文件，得 2 分。                                             |

表 6-3 任务二评分细则（40 分）

| 序号 | 评分项                  | 分值  | 评分细则                                            |
|----|----------------------|-----|-------------------------------------------------|
| 1  | 修改 hadoop-env.sh 配置文 | 4 分 | 和参考答案一致满分；<br>hadoop-env.sh 配置正确（路径配置正确），得 2 分； |

|   |                                      |     |                                                                                                                                                                                                                  |
|---|--------------------------------------|-----|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|   | 件                                    |     | 考生文件夹中有 2-1 修改 <code>hadoop-env.sh</code> 文件.png，得 2 分。                                                                                                                                                          |
| 2 | 修改 <code>core-site.xml</code> 配置文件   | 8 分 | 和参考答案一致满分；<br>正确配置 <code>namenode</code> 主机名及端口，得 3 分；<br>正确配置临时文件的目录，得 3 分；<br>考生文件夹中有 2-2 修改 <code>hadoop</code> 核心配置文件.png 文件，得 2 分。                                                                          |
| 3 | 修改 <code>hdfs-site.xml</code> 配置文件   | 8 分 | 和参考答案一致满分；<br>正确配置 <code>hdfs-site.xml</code> 文件，得 6 分；<br>考生文件夹中有 2-3 修改 <code>hdfs</code> 核心配置文件.png 文件，得 2 分。                                                                                                 |
| 4 | 修改 <code>mapred-site.xml</code> 配置文件 | 8 分 | 和参考答案一致满分；<br>成功复制 <code>mapred-site.xml.template</code> 并重命名为 <code>mapred-site.xml</code> ，得 2 分；<br>正确配置 <code>mapred-site.xml</code> 文件内容，得 4 分；<br>考生文件夹中有 2-4 修改 <code>MapReduce</code> 核心文件.png 文件，得 2 分。 |
| 5 | 修改 <code>yarn-site.xml</code> 配置文件   | 8 分 | 和参考答案一致满分；<br>正确配置 <code>yarn-site.xml</code> 文件，得 6 分；<br>考生文件夹中有 2-5 修改 <code>yarn</code> 核心配置文件.png 文件，得 2 分。                                                                                                 |
| 6 | 修改 <code>slaves</code> 配置文件          | 4 分 | 和参考答案一致满分；<br>正确配置 <code>slaves</code> 文件内容，得 2 分；<br>考生文件夹中有 2-6 修改 <code>slaves</code> 文件.png 文件，得 2 分。                                                                                                        |

表 6-4 任务三评分细则（20 分）

| 序号 | 评分项                                         | 分值  | 评分细则                                                                                         |
|----|---------------------------------------------|-----|----------------------------------------------------------------------------------------------|
| 1  | 分发配置文件                                      | 4 分 | 和参考答案一致满分；<br>分发配置文件到从节点命令正确，得 3 分；<br>考生文件夹中有 3-1 分发 <code>hadoop</code> 配置文件.png 文件，得 1 分。 |
| 2  | 格式化 <code>NameNode</code>                   | 8 分 | 和参考答案一致满分；<br>格式化成功，得 6 分；<br>考生文件夹中有 3-2 格式化 <code>namenode.png</code> 文件，得 2 分。            |
| 3  | 启动 <code>hdfs</code> 和 <code>yarn</code> 集群 | 8 分 | 和参考答案一致满分；<br>启动命令正确且 <code>master</code> 和 <code>slave1</code> 上运行的进程正确，                    |

|  |  |  |                                                                               |
|--|--|--|-------------------------------------------------------------------------------|
|  |  |  | 得 6 分；<br>考生文件夹中有 3-3-1 启动 Hadoop 集群.png 和 3-3-2 查看 slave1 启动情况.png 文件，得 2 分。 |
|--|--|--|-------------------------------------------------------------------------------|

表 6-5 职业素质评分细则（10 分）

| 序号 | 评分项  | 分值  | 评分细则                              |
|----|------|-----|-----------------------------------|
| 1  | 操作规范 | 3 分 | 答题过程中保持安静，不损坏机器；                  |
| 2  | 文明考试 | 3 分 | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌；      |
| 3  | 工位整理 | 4 分 | 答题完毕立场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |

## 7.试题编号：J2-2-3：夸克小说网 Hadoop 完全分布式模式部署

### （1）任务描述

夸克小说作为一家小说阅读平台。现在需要通过各个小说网站查询到每部小说的详细信息，帮助使用者查找到书名、作者、收藏和评价等数据内容并完成数据展示。

现在要进行大数据分析并完成展示，你作为该公司大数据平台运维工程师，需在 Hadoop 伪分布式模式基础上部署完全分布式 Hadoop 环境，JDK 安装包、Hadoop 安装包所在路径为：/export/software，JDK 和 Hadoop 解压包所在路径为：/export/servers。完全分布式规划如下表：

表 7-1 完全分布式规划

| 节点角色 | 虚拟机名     | IP 地址           | 主机名    | 运行进程                                         |
|------|----------|-----------------|--------|----------------------------------------------|
| 主节点  | Hadoop01 | 192.168.233.130 | master | NameNode<br>ResoureManager                   |
| 从节点  | Hadoop02 | 192.168.233.131 | slave1 | DataNode<br>NodeManager<br>SecondaryNameNode |
|      | Hadoop03 | 192.168.233.132 | slave2 | DataNode<br>NodeManager                      |

主要任务如下：

**任务一：规划完全分布式模式下的 Hadoop 集群（15 分）**

- 通过对已有虚拟机 Hadoop01（用户名：root 密码：123456）进行克隆，得到 Hadoop02 和 Hadoop03 虚拟机，截取克隆后得到的三台虚拟机；图片保存到考生文件夹下，并命名为“1-1 克隆虚拟机.png”。（5 分）
- 按集群规划修改主机名，分别截取修改后两台从节点的主机名，并命名为“1-2-1 修改主机名 1”、“1-2-2 修改主机名 2”，图片保存到考生文件夹下。（5 分）
- 根据题目要求设置虚拟机的 IP，修改虚拟机网卡配置文件，将 slave1 的修改内容截图，图片保存到考生文件夹下，并命名为“2-3 修改 ip 地址.png”。（5 分）

**任务二：集群主节点的配置（40 分）**

- Hadoop 的运行需要依赖 JDK，而 hadoop-env.sh 文件为 Hadoop 的运行环境配置文件，请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“2-1 配置 hadoop 运行环境.png”。（4 分）
- core-site.xml 文件为 Hadoop 的核心配置文件，用于定义系统级别的参数，设定 namenode 的主机名及端口、存放临时文件的目录，请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“2-2 修改 hadoop 核心配置文件.png”。（8 分）
- hdfs-site.xml 文件为 HDFS 核心配置文件，如文件副本的个数、块大小及是否使用强制权限等，设定 HDFS 存储文件的副本个数默认为 3、SecondaryNameNode 地址和端口，请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“2-3 修改 hdfs 核心配置文件.png”。（8 分）
- mapred-site.xml 文件是 MapRduce 的核心配置文件，用于指定 MapRduce 运行时的框架，请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“2-4 修改 MapReduce 核心文件.png”。（8 分）
- yarn-site.xml 文件是 YARN 框架的核心配置文件，需要指定 YARN 集群的管理者，请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“2-5 修改 yarn 核心配置文件.png”。（8 分）

- slaves 文件用于记录集群所有从节点主机名，请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“2-6 修改 slaves 文件.png”。（2 分）

### 任务三：分发配置文件与测试 Hadoop 集群（35 分）

- 由于所修改的配置文件都存在于 Hadoop 安装目录下的 etc/hadoop 目录下，因此将主节点上的 Hadoop 安装目录下的 etc/hadoop 目录分发到两个从节点即可，将分发至从节点 slave1 的命令截图，图片保存到考生文件夹下，并命名为“3-1 分发 hadoop 配置文件.png”。（5 分）
- 在主节点上执行相关命令格式化 NameNode，截取格式化成功界面，图片保存到考生文件夹下，并命名为“3-2 格式化 namenode.png”。（8 分）
- 使用相关指令分别启动 hdfs 和 yarn 集群，并使用 jps 进行查看，截取启动命令与 master 上的 jps 查看结果，图片保存到考生文件夹下，并命名为“3-3-1 启动 Hadoop 集群.png”；截取 slave1 上 jps 查看结果，图片保存到考生文件夹下，并命名为“3-3-2 查看 slave1 启动情况.png”。（8 分）
- 使用相关命令在 HDFS 根目录下创建 input 目录，截取 web 界面目录创建情况，图片保存到考生文件夹下，并命名为“3-4 创建目录.png”。（6 分）
- 使用 hdfs shell 命令将 Hadoop 安装目录下的 README.txt 文档上传到 HDFS 根目录下的 input 目录下，截取 web 界面上文档上传情况，图片保存到考生文件夹下，并命名为“3-5 上传文件.png”。（8 分）

### 良好的习惯和职业素养（10 分）

- 答题过程中保持安静。（3 分）
- 有问题举手示意，有礼貌。（3 分）
- 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。（4 分）

## （2）实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。计算机安装有 Vmware workstation15，XShell，JDK1.8，并有 Hadoop2.7.4 安装包和 CentOS7 64 位镜像文件，安装 Office 办公软件。

### (3) 考核时量及要求

- 考核时长 180 分钟；
- 请在桌面建立以“学号\_题号”命名的成果文件夹，所有提交文件均放在该目录下，操作过程的截图需按照题目要求，文件夹命名例如：201965110001\_J3-1-1。

### (4) 评分细则

表 7-2 任务一评分细则（15 分）

| 序号 | 评分项          | 分<br>值 | 评分细则                                                                                           |
|----|--------------|--------|------------------------------------------------------------------------------------------------|
| 1  | 正确克隆虚拟机      | 5 分    | 和参考答案一致满分；<br>正确克隆两台虚拟机命名为 Hadoop02 和 Hadoop03，得 3 分；<br>考生文件夹中有 1-1 克隆虚拟机.png 文件，得 2 分。       |
| 2  | 正确修改主机名      | 5 分    | 和参考答案一致满分；<br>两台从节点修改主机名正确，得 3 分；<br>考生文件夹中有 1-2-1 修改主机名 1.png 文件和 1-2-2 修改主机名 2.png 文件，得 2 分。 |
| 4  | 正确修改 IP 地址文件 | 5 分    | 和参考答案一致满分；<br>正确打开虚拟机网卡配置文件，得 1 分；<br>配置文件内容正确，3 分；<br>考生文件夹中有 2-3 修改 ip 地址.png 文件，得 1 分。      |

表 7-3 任务二评分细则（40 分）

| 序号 | 评分项                   | 分<br>值 | 评分细则                                                                                                          |
|----|-----------------------|--------|---------------------------------------------------------------------------------------------------------------|
| 1  | 修改 hadoop-env.sh 配置文件 | 4 分    | 和参考答案一致满分；<br>hadoop-env.sh 配置正确（路径配置正确），得 2 分；<br>考生文件夹中有 2-1 修改 hadoop-env.sh 文件.png，得 2 分。                 |
| 2  | 修改 core-site.xml 配置文件 | 8 分    | 和参考答案一致满分；<br>正确配置 namenode 主机名及端口，得 3 分；<br>正确配置临时文件的目录，得 3 分；<br>考生文件夹中有 2-2 修改 hadoop 核心配置文件.png 文件，得 2 分。 |

|   |                         |     |                                                                                                                                                             |
|---|-------------------------|-----|-------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 3 | 修改 hdfs-site.xml 配置文件   | 8 分 | 和参考答案一致满分；<br>正确配置 hdfs-site.xml 文件，得 6 分；<br>考生文件夹中有 2-3 修改 hdfs 核心配置文件.png 文件，得 2 分。                                                                      |
| 4 | 修改 mapred-site.xml 配置文件 | 8 分 | 和参考答案一致满分；<br>成功复制 mapred-site.xml.template 并重命名为 mapred-site.xml，得 2 分；<br>正确配置 mapred-site.xml 文件内容，得 4 分；<br>考生文件夹中有 2-4 修改 MapReduce 核心文件.png 文件，得 2 分。 |
| 5 | 修改 yarn-site.xml 配置文件   | 8 分 | 和参考答案一致满分；<br>正确配置 yarn-site.xml 文件，得 6 分；<br>考生文件夹中有 2-5 修改 yarn 核心配置文件.png 文件，得 2 分。                                                                      |
| 6 | 修改 slaves 配置文件          | 4 分 | 和参考答案一致满分；<br>正确配置 slaves 文件内容，得 2 分；<br>考生文件夹中有 2-6 修改 slaves 文件.png 文件，得 2 分。                                                                             |

表 7-4 任务三评分细则（35 分）

| 序号 | 评分项               | 分值  | 评分细则                                                                                                                           |
|----|-------------------|-----|--------------------------------------------------------------------------------------------------------------------------------|
| 1  | 分发配置文件            | 5 分 | 和参考答案一致满分；<br>分发配置文件到从节点命令正确，得 3 分；<br>考生文件夹中有 3-1 分发 hadoop 配置文件.png 文件，得 2 分。                                                |
| 2  | 格式化 NameNode      | 8 分 | 和参考答案一致满分；<br>格式化成功，得 6 分；<br>考生文件夹中有 3-2 格式化 namenode.png 文件，得 2 分。                                                           |
| 3  | 启动 hdfs 和 yarn 集群 | 8 分 | 和参考答案一致满分；<br>启动集群命令正确且 master 和 slave1 上运行的进程正确，得 6 分；<br>考生文件夹中有 3-3-1 启动 Hadoop 集群.png 和 3-3-2 查看 slave1 启动情况.png 文件，得 2 分。 |
| 4  | web 界面下创建目录       | 6 分 | 和参考答案一致满分；<br>HDFS 根目录下创建目录正确，得 4 分；<br>考生文件夹中有 3-4 创建目录.png 文件，得 2 分。                                                         |
| 5  | web 界面上上传文档       | 8 分 | 和参考答案一致满分；<br>HDFS 根目录下的 input 下上传文档成功，得 6 分；<br>考生文件夹中有 3-5 上传文件.png 文件，得 2 分。                                                |

表 7-5 职业素质评分细则（10 分）

| 序号 | 评分项  | 分<br>值 | 评分细则                              |
|----|------|--------|-----------------------------------|
| 1  | 操作规范 | 3 分    | 答题过程中保持安静，不损坏机器；                  |
| 2  | 文明考试 | 3 分    | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌；      |
| 3  | 工位整理 | 4 分    | 答题完毕立场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |

## 8.试题编号：J2-2-4：中关村在线平台 Hadoop 完全分布式模式部署

### （1）任务描述

中关村在线平台通过各个电商平台网站查询到各类电子商品的详细信息，帮助使用者找到品牌、生产厂商、上市时间和价格等数据内容并完成数据展示。

现在要进行大数据分析并完成展示，你作为该公司大数据平台运维工程师，需在 Hadoop 伪分布式模式基础上部署完全分布式 Hadoop 环境，JDK 安装包、Hadoop 安装包所在路径为：/export/software，JDK 和 Hadoop 解压包所在路径为：/export/servers。完全分布式规划如下表：

表 8-1 完全分布式规划

| 节点角色 | 虚拟机名     | IP 地址               | 主机名    | 运行进程                                                  |
|------|----------|---------------------|--------|-------------------------------------------------------|
| 主节点  | Hadoop01 | 192.168.233.13<br>0 | node01 | NameNode<br>ResoureManager<br>NodeManager<br>DataNode |
| 从节点  | Hadoop02 | 192.168.233.13<br>1 | node02 | DataNode<br>NodeManager                               |
|      | Hadoop03 | 192.168.233.13<br>2 | node03 | DataNode<br>NodeManager<br>SecondaryNameNode          |

主要任务如下：

### 任务一：规划完全分布式模式下的 Hadoop 集群（30 分）

- 对已有虚拟机 Hadoop01（用户名：root 密码：123456）进行克隆，得到 Hadoop02 和 Hadoop03 虚拟机，截取克隆后得到的三台虚拟机；图片保存到考生文件夹下，并命名为“1-1 克隆虚拟机.png”。（4 分）
- 开启三台虚拟机并按集群规划修改主机名，使用相关命令显示两台从节点修改后的主机名，截取命令及显示效果，图片保存到考生文件夹下，并命名为“1-2-1 主机名 1”、“1-2-2 主机名 2”。（6 分）
- 在三台主机上均配置集群各主机的主机名和 IP 地址映射关系，截取任意一台主机的 IP 映射配置内容，图片保存到考生文件夹下，并命名为“1-3 配置 IP 映射.png”。（4 分）
- 根据题目要求设置虚拟机的 IP，修改虚拟机网卡配置文件，将 node01 主机的修改内容截图，图片保存到考生文件夹下，并命名为“1-4 修改 ip 地址.png”。（10 分）
- 重启网络服务后，使用相关命令查看各主机修改后的 IP 地址，截取 node02 的查看效果，图片保存到考生文件夹下，并命名为“1-5 查看修改后的 ip 地址.png”。（6 分）

### 任务二：远程登录和免密登录设置（20）

- 先查看 ssh 服务是否安装，将命令和结果截图，图片保存到考生文件夹下，并命名为“2-1-1 查看 ssh 安装情况.png”，再查看 ssh 服务是否启动（若未启动先使用相关命令启动 ssh 服务），将命令和结果截图，图片保存到考生文件夹下，并命名为“2-1-2 查看 ssh 启动情况.png”。（4 分）
- 使用 XShell 远程连接三台主机，将连接成功后的界面截图，图片保存到考生文件夹下，并命名为“2-2 远程登录.png”。（2 分）
- 在 node01 上生成密钥文件，将命令截图，图片保存到考生文件夹下，并命名为“2-3-1 生成密钥文件.png”；将 node01 的公钥复制到 node01、node02、node03，将复制到 node02 的操作命令截图，图片保存到考生文件夹下，并命名为“2-3-2 复制密钥.png”。（4 分）
- 在 node02 上生成密钥文件，将命令截图，图片保存到考生文件夹下，并命名为“2-4-1 生成密钥文件.png”；将 node02 的公钥复制至 node01、node02、node03，将复制到 node01 的操作命令截图，图片保存到考生文件夹下，并命名为“2-4-2 复制密钥.png”。（4 分）

- 在 node03 上生成密钥文件，将命令截图，图片保存到考生文件夹下，并命名为“2-5-1 生成密钥文件.png”；将 node03 的公钥复制至 node01、node02、node03，将复制到 node01 的操作命令截图，图片保存到考生文件夹下，并命名为“2-5-2 复制密钥.png”。（4 分）
- 测试免密登陆是否成功，在 node01 节点上使用相关命令免密登录到 node02、然后 node02 免密登录 node03、node03 免密登录 node01，截取命令和结果，图片保存到考生文件夹下，并命名为“2-6 测试免密登录.png”。（2 分）

### 任务三：集群主节点的配置（40）

- hadoop-env.sh 文件为配置 Hadoop 的环境变量，提供了 Hadoop 中 JAVA\_HOME 的运行环境。请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“3-1 修改 hadoop-env.sh 文件.png”。（4 分）
- core-site.xml 文件为 Hadoop 的核心配置文件，用于定义系统级别的参数，设定 namenode 的主机名及端口、存放临时文件的目录请设定为 /export/servers/hadoop-2.7.4/tmp，请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“3-2 修改 hadoop 核心配置文件.png”。（8 分）
- hdfs-site.xml 文件为 HDFS 核心配置文件，如文件副本的个数、块大小及是否使用强制权限等，设定 HDFS 存储文件的副本个数为 3、SecondaryNameNode 地址和端口，请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“3-3 修改 hdfs 核心配置文件.png”。（8 分）
- mapred-site.xml 文件是 MapReduce 的核心配置文件，用于指定 MapReduce 运行时的框架，需要对已有 mapred-site.xml.template 文件进行复制并重命名得到该文件，请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“3-4 修改 MapReduce 核心文件.png”。（10 分）
- yarn-site.xml 文件是 YARN 框架的核心配置文件，需要指定 YARN 集群的管理者，请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“3-5 修改 yarn 核心配置文件.png”。（8 分）
- slaves 文件用于记录 Hadoop 集群所有从节点（HDFS 的 DataNode 和 YARN 的

NodeManager 所在主机), 根据题目要求进行配置并提交该配置文件的内容截图, 图片保存到考生文件夹下, 并命名为“3-6 修改 slaves 文件.png”。(4 分)

### 良好的习惯和职业素养 (10 分)

- 答题过程中保持安静。(3 分)
- 有问题举手示意, 有礼貌。(3 分)
- 答题完毕离场时整理好工位, 不关闭计算机, 并按要求显示答案保存界面。(4 分)

## (2) 实施条件

安排在教学实训机房, 计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。计算机安装有 Vmware workstation15, XShell, JDK1.8, 并有 Hadoop2.7.4 安装包和 CentOS7 64 位镜像文件, 安装 Office 办公软件。

## (3) 考核时量及要求

- 考核时长 180 分钟;
- 请在桌面建立以“学号\_题号”命名的成果文件夹, 所有提交文件均放在该目录下, 操作过程的截图需按照题目要求, 文件夹命名例如: 201965110001\_J3-1-1。

## (4) 评分细则

表 8-2 任务一评分细则 (30 分)

| 序号 | 评分项        | 分值  | 评分细则                                                                                           |
|----|------------|-----|------------------------------------------------------------------------------------------------|
| 1  | 正确克隆虚拟机    | 4 分 | 和参考答案一致满分;<br>正确克隆两台虚拟机命名为 Hadoop02 和 Hadoop03, 得 2 分;<br>考生文件夹中有 1-1 克隆虚拟机.png 文件, 得 2 分。     |
| 2  | 正确修改主机名    | 6 分 | 和参考答案一致满分;<br>两台从节点修改主机名正确, 得 4 分;<br>考生文件夹中有 1-2-1 改主机名 1.png 文件和 1-2-2 改主机名 2.png 文件, 得 2 分。 |
| 3  | 正确配置 IP 映射 | 4 分 | 和参考答案一致满分;<br>正确配置 ip 映射关系, 得 2 分;<br>考生文件夹中有 1-3 配置 IP 映射.png 文件, 得 2 分。                      |

|   |              |      |                                                                                                                     |
|---|--------------|------|---------------------------------------------------------------------------------------------------------------------|
| 4 | 正确修改 IP 地址文件 | 10 分 | 和参考答案一致满分；<br>正确打开网卡配置文件，得 2 分；<br>配置文件内容正确，得 6 分；<br><br>考生文件夹中有 1-4 修改 ip 地址.png 文件，得 2 分。                        |
| 5 | 重启网络服务正确     | 6 分  | 和参考答案一致满分；<br>重启网络服务命令正确，得 3 分；<br>有 ip addr 或 ip addr show 命令，得 1 分；<br><br>考生文件夹中有 1-5 查看修改后的 ip 地址.png 文件，得 2 分。 |

表 8-3 任务二评分细则（20 分）

| 序号 | 评分项            | 分值  | 评分细则                                                                                                                                                                        |
|----|----------------|-----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 查看 ssh 服务      | 4 分 | 和参考答案一致满分；<br>查看 ssh 服务是否安装命令正确，得 1 分；<br>考生文件夹中有 2-1-1 查看 ssh 安装情况.png 文件，得 1 分；<br>查看 ssh 服务是否启动命令正确（若未启动，则先使用启动命令），得 1 分；<br><br>考生文件夹中有 2-1-2 查看 ssh 启动情况.png 文件，得 1 分。 |
| 2  | 远程登录           | 2 分 | 和参考答案一致满分；<br>考生文件夹中有 2-2 远程登录.png 文件，得 2 分。                                                                                                                                |
| 3  | node01 上生成密钥文件 | 4 分 | 和参考答案一致满分；<br>成功生成密钥，得 1 分；<br>公钥复制命令正确，得 1 分；<br><br>考生文件夹中有 2-3-1 生成密钥文件.png 和 2-3-2 复制密钥.png 文件，得 2 分。                                                                   |
| 4  | node02 上生成密钥文件 | 4 分 | 和参考答案一致满分；<br>成功生成密钥，得 1 分；<br>公钥复制命令正确，得 1 分；<br><br>考生文件夹中有 2-4-1 生成密钥文件.png 和 2-4-2 复制密钥.png 文件，得 2 分。                                                                   |
| 5  | node03 上生成密钥文件 | 4 分 | 和参考答案一致满分；<br>成功生成密钥，得 1 分；<br>公钥复制命令正确，得 1 分；<br><br>考生文件夹中有 2-5-1 生成密钥文件.png 和 2-5-2                                                                                      |

|   |        |     |                                                                         |
|---|--------|-----|-------------------------------------------------------------------------|
|   |        |     | 复制秘钥.png 文件，得 2 分。                                                      |
| 6 | 测试免密登录 | 2 分 | 和参考答案一致满分；<br>成功使用免密登录进入三个节点，得 1 分；<br>考生文件夹中有 2-6 测试免密登录.png 文件，得 1 分。 |

表 8-4 任务三评分细则（40 分）

| 序号 | 评分项                     | 分值  | 评分细则                                                                                                                                                        |
|----|-------------------------|-----|-------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 修改 hadoop-env.sh 配置文件   | 4 分 | 和参考答案一致满分；<br>hadoop-env.sh 配置正确（路径配置正确），得 2 分；<br>考生文件夹中有 3-1 修改 hadoop-env.sh 文件.png，得 2 分。                                                               |
| 2  | 修改 core-site.xml 配置文件   | 8 分 | 和参考答案一致满分；<br>正确配置 namenode 主机名及端口，得 3 分；<br>正确配置临时文件的目录，得 3 分；<br>考生文件夹中有 3-2 修改 hadoop 核心配置文件.png 文件，得 2 分。                                               |
| 3  | 修改 hdfs-site.xml 配置文件   | 8 分 | 和参考答案一致满分；<br>正确配置 hdfs-site.xml 文件，得 6 分；<br>考生文件夹中有 3-3 修改 hdfs 核心配置文件.png 文件，得 2 分。                                                                      |
| 4  | 修改 mapred-site.xml 配置文件 | 8 分 | 和参考答案一致满分；<br>成功复制 mapred-site.xml.template 并重命名为 mapred-site.xml，得 2 分；<br>正确配置 mapred-site.xml 文件内容，得 4 分；<br>考生文件夹中有 3-4 修改 MapReduce 核心文件.png 文件，得 2 分。 |
| 5  | 修改 yarn-site.xml 配置文件   | 8 分 | 和参考答案一致满分；<br>正确配置 yarn-site.xml 文件，得 6 分；<br>考生文件夹中有 3-5 修改 yarn 核心配置文件.png 文件，得 2 分。                                                                      |
| 6  | 修改 slaves 配置文件          | 4 分 | 和参考答案一致满分；<br>正确配置 slaves 文件内容，得 2 分；<br>考生文件夹中有 3-6 修改 slaves 文件.png 文件，得 2 分。                                                                             |

表 8-5 职业素质评分细则（10 分）

| 序号 | 评分项  | 分值  | 评分细则                              |
|----|------|-----|-----------------------------------|
| 1  | 操作规范 | 3 分 | 答题过程中保持安静，不损坏机器；                  |
| 2  | 文明考试 | 3 分 | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌；      |
| 3  | 工位整理 | 4 分 | 答题完毕立场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |

## 9.试题编号：J2-2-5：京东 Hadoop 完全分布式模式部署

### （1）任务描述

京东作为一家电子商务平台。现在需要通过查询各家店铺和各类商品的详细情况，帮助使用者查找到店铺名、商品名、品牌和销售量等数据内容并完成数据展示。

现在要进行大数据分析并完成展示，你作为该公司大数据平台运维工程师，需在 Hadoop 伪分布式模式基础上部署完全分布式 Hadoop 环境，JDK 安装包、Hadoop 安装包所在路径为：/export/software，JDK 和 Hadoop 解压包所在路径为：/export/servers。完全分布式规划如下表：

表 9-1 完全分布式规划

| 节点角色 | 虚拟机名     | IP 地址           | 主机名    | 运行进程                                                  |
|------|----------|-----------------|--------|-------------------------------------------------------|
| 主节点  | Hadoop01 | 192.168.233.130 | node01 | NameNode<br>ResoureManager<br>NodeManager<br>DataNode |
| 从节点  | Hadoop02 | 192.168.233.131 | node02 | DataNode<br>NodeManager                               |
|      | Hadoop03 | 192.168.233.132 | node03 | DataNode<br>NodeManager<br>SecondaryNameNode          |

主要任务如下：

### 任务一：规划完全分布式模式下的 Hadoop 集群（30 分）

- 对已有虚拟机 Hadoop01（用户名：root 密码：123456）进行克隆，得到 Hadoop02 和 Hadoop03 虚拟机，截取克隆后得到的三台虚拟机；图片保存到考生文件夹下，并命名为“1-1 克隆虚拟机.png”。（4 分）
- 开启三台虚拟机并按集群规划修改主机名，使用相关命令显示两台从节点修改后的主机名，截取命令及显示效果，图片保存到考生文件夹下，并命名为“1-2-1 主机名 1”、“1-2-2 主机名 2”。（4 分）
- 根据题目要求在相关文件下配置 IP 映射，要求三台主机均进行 IP 映射配置，并截取主节点的 IP 映射配置内容，图片保存到考生文件夹下，并命名为“1-3 配置 IP 映射.png”。（4 分）
- 根据题目要求设置三台主机的 IP 地址，修改虚拟机网卡配置文件，将 node02 的修改内容截图，图片保存到考生文件夹下，并命名为“1-4 修改 ip 地址.png”。（8 分）
- 完成上述操作后重启网络服务，使用相关命令查看 IP 配置是否生效，截取命令及结果，图片保存到考生文件夹下，并命名为“1-5 查看 ip.png”。（6 分）
- 使用相关命令检测三台主机之间是否相互连通，截取主节点 ping 两台从节点的测试情况，图片保存到考生文件夹下，并命名为“1-6 连通测试.png”。（4 分）

## 任务二：修改主节点配置文件（40 分）

- Xshell 连接三台主机后，在主节点上修改相关配置文件，hadoop-env.sh 文件为配置 Hadoop 的环境变量，提供了 Hadoop 中 JAVA\_HOME 的运行环境。请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“2-1 修改 hadoop-env.sh 文件.png”。（4 分）
- core-site.xml 文件为 Hadoop 的核心配置文件，用于定义系统级别的参数，设定 namenode 的主机名及端口、存放临时文件的目录请设定为 /export/servers/hadoop-2.7.4/tmp，请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“2-2 修改 hadoop 核心配置文件.png”。（8 分）
- hdfs-site.xml 文件为 HDFS 核心配置文件，如文件副本的个数、块大小及是否使用强制权限等，设定 HDFS 存储文件的副本个数为 3、SecondaryNameNode 地址和端口，

请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“2-3 修改 hdfs 核心配置文件.png”。（8 分）

- mapred-site.xml 文件是 MapRduce 的核心配置文件，用于指定 MapRduce 运行时的框架，需要对已有 mapred-site.xml.template 文件进行复制并重命名得到该文件，请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“2-4 修改 MapReduce 核心文件.png”。（8 分）

- yarn-site.xml 文件是 YARN 框架的核心配置文件，需要指定 YARN 集群的管理者，请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“2-5 修改 yarn 核心配置文件.png”。（8 分）

- slaves 文件用于记录 Hadoop 集群所有从节点（HDFS 的 DataNode 和 YARN 的 NodeManager 所在主机），根据题目要求进行配置并提交该配置文件的内容截图，图片保存到考生文件夹下，并命名为“2-6 修改 slaves 文件.png”。（4 分）

### **任务三：分发配置文件与测试 Hadoop 集群（20 分）**

- 由于所修改的配置文件都存在于 Hadoop 安装目录下的 etc/hadoop 目录下，因此将主节点上的 Hadoop 安装目录下的 etc/hadoop 目录分发到两个从节点即可，将分发至从节点 node02 的命令截图，图片保存到考生文件夹下，并命名为“3-1 分发 hadoop 配置文件.png”。（4 分）

- 在主节点上执行相关命令格式化 NameNode，截取格式化成功界面，图片保存到考生文件夹下，并命名为“3-2 格式化 namenode.png”。（8 分）

- 使用相关指令分别启动 hdfs 和 yarn 集群，并使用 jps 进行查看，截取启动命令与 node01 上的 jps 查看结果，图片保存到考生文件夹下，并命名为“3-3-1 启动 Hadoop 集群.png”；截取 node02 上 jps 查看结果，图片保存到考生文件夹下，并命名为“3-3-2 查看 node02 启动情况.png”。（8 分）

### **良好的习惯和职业素养（10 分）**

- 答题过程中保持安静。（3 分）
- 有问题举手示意，有礼貌。（3 分）

- 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。（4分）

## （2）实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。

计算机安装有 Vmware workstation15, XShell, JDK1.8, 并有 Hadoop2.7.4 安装包和 CentOS7 64 位镜像文件，安装 Office 办公软件。

## （3）考核时量及要求

- 考核时长 180 分钟；
- 请在桌面建立以“学号\_题号”命名的成果文件夹，所有提交文件均放在该目录下，操作过程的截图需按照题目要求，文件夹命名例如：201965110001\_J3-1-1。

## （4）评分细则

表 9-2 任务一评分细则（30 分）

| 序号 | 评分项          | 分值  | 评分细则                                                                                           |
|----|--------------|-----|------------------------------------------------------------------------------------------------|
| 1  | 正确克隆虚拟机      | 4 分 | 和参考答案一致满分；<br>正确克隆两台虚拟机命名为 Hadoop02 和 Hadoop03，得 2 分；<br>考生文件夹中有 1-1 克隆虚拟机.png 文件，得 2 分。       |
| 2  | 正确修改主机名      | 4 分 | 和参考答案一致满分；<br>两台从节点修改主机名正确，得 2 分；<br>考生文件夹中有 1-2-1 修改主机名 1.png 文件和 1-2-2 修改主机名 2.png 文件，得 2 分。 |
| 3  | 正确配置 IP 映射   | 4 分 | 和参考答案一致满分；<br>正确配置 ip 映射关系，得 2 分；<br>考生文件夹中有 1-3 配置 IP 映射.png 文件，得 2 分。                        |
| 4  | 正确修改 IP 地址文件 | 8 分 | 和参考答案一致满分；<br>正确打开虚拟机网卡配置文件，得 2 分；<br>配置文件内容正确，得 4 分；<br>考生文件夹中有 1-4 修改 ip 地址.png 文件，得 2 分。    |
| 5  | 重启网络服务正确     | 6 分 | 和参考答案一致满分；<br>重启网络服务命令正确，得 3 分；                                                                |

|   |      |     |                                                                       |
|---|------|-----|-----------------------------------------------------------------------|
|   |      |     | 有 ip addr 或 ip addr show 命令，得 1 分；<br>考生文件夹中有 1-5 查看 ip.png 文件，得 2 分。 |
| 6 | 连通测试 | 4 分 | 和参考答案一致满分；<br>三台主机之间连通成功，得 2 分；<br>考生文件夹中有 1-6 连通测试.png 文件，得 2 分      |

表 9-3 任务二评分细则（40 分）

| 序号 | 评分项                     | 分值  | 评分细则                                                                                                                                                        |
|----|-------------------------|-----|-------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 修改 hadoop-env.sh 配置文件   | 4 分 | 和参考答案一致满分；<br>hadoop-env.sh 配置正确（路径配置正确），得 2 分；<br>考生文件夹中有 2-1 修改 hadoop-env.sh 文件.png，得 2 分。                                                               |
| 2  | 修改 core-site.xml 配置文件   | 8 分 | 和参考答案一致满分；<br>正确配置 namenode 主机名及端口，得 3 分；<br>正确配置临时文件的目录，得 3 分；<br>考生文件夹中有 2-2 修改 hadoop 核心配置文件.png 文件，得 2 分。                                               |
| 3  | 修改 hdfs-site.xml 配置文件   | 8 分 | 和参考答案一致满分；<br>正确配置 hdfs-site.xml 文件，得 6 分；<br>考生文件夹中有 2-3 修改 hdfs 核心配置文件.png 文件，得 2 分。                                                                      |
| 4  | 修改 mapred-site.xml 配置文件 | 8 分 | 和参考答案一致满分；<br>成功复制 mapred-site.xml.template 并重命名为 mapred-site.xml，得 2 分；<br>正确配置 mapred-site.xml 文件内容，得 4 分；<br>考生文件夹中有 2-4 修改 MapReduce 核心文件.png 文件，得 2 分。 |
| 5  | 修改 yarn-site.xml 配置文件   | 8 分 | 和参考答案一致满分；<br>正确配置 yarn-site.xml 文件，得 6 分；<br>考生文件夹中有 2-5 修改 yarn 核心配置文件.png 文件，得 2 分。                                                                      |
| 6  | 修改 slaves 配置文件          | 4 分 | 和参考答案一致满分；<br>正确配置 slaves 文件内容，得 2 分；<br>考生文件夹中有 2-6 修改 slaves 文件.png 文件，得 2 分。                                                                             |

表 9-4 任务三评分细则（20 分）

| 序号 | 评分项               | 分值  | 评分细则                                                                                                                           |
|----|-------------------|-----|--------------------------------------------------------------------------------------------------------------------------------|
| 1  | 分发配置文件            | 4 分 | 和参考答案一致满分；<br>分发配置文件到从节点命令正确，得 3 分；<br>考生文件夹中有 3-1 分发 hadoop 配置文件.png 文件，得 1 分。                                                |
| 2  | 格式化 NameNode      | 8 分 | 和参考答案一致满分；<br>格式化成功，得 6 分；<br>考生文件夹中有 3-2 格式化 namenode.png 文件，得 2 分。                                                           |
| 3  | 启动 hdfs 和 yarn 集群 | 8 分 | 和参考答案一致满分；<br>启动集群命令正确且 node01 和 node02 上运行的进程正确，得 6 分；<br>考生文件夹中有 3-3-1 启动 Hadoop 集群.png 和 3-3-2 查看 node02 启动情况.png 文件，得 2 分。 |

表 9-5 职业素质评分细则（10 分）

| 序号 | 评分项  | 分值  | 评分细则                              |
|----|------|-----|-----------------------------------|
| 1  | 操作规范 | 3 分 | 答题过程中保持安静，不损坏机器；                  |
| 2  | 文明考试 | 3 分 | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌；      |
| 3  | 工位整理 | 4 分 | 答题完毕立场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |

## 10.试题编号：J2-2-6：西瓜视频 Hadoop 完全分布式模式部署

### （1）任务描述

西瓜视频作为一家视频平台，现在需要查询不同创作者和各类视频的详细情况，帮助用户查找视频名，创作者，视频播放量和评价等数据内容并完成数据展示。

现在要进行大数据分析并完成展示，你作为该公司大数据平台运维工程师，现在要进行大数据分析并完成展示，你作为该公司大数据平台运维工程师，需在 Hadoop 伪分布式模式基础上部署完全分布式 Hadoop 环境，JDK 安装包、Hadoop 安装包所在路径为：

/export/software，JDK 和 Hadoop 解压包所在路径为：/export/servers。完全分布式规划如下

表：

表 10-1 完全分布式规划

| 节点角色 | 虚拟机名     | IP 地址               | 主机名    | 运行进程                                                  |
|------|----------|---------------------|--------|-------------------------------------------------------|
| 主节点  | Hadoop01 | 192.168.233.13<br>0 | node01 | NameNode<br>ResoureManager<br>NodeManager<br>DataNode |
| 从节点  | Hadoop02 | 192.168.233.13<br>1 | node02 | DataNode<br>NodeManager                               |
|      | Hadoop03 | 192.168.233.13<br>2 | node03 | DataNode<br>NodeManager<br>SecondaryNameNode          |

主要任务如下：

**任务一：规划完全分布式模式下的 Hadoop 集群（15 分）**

- 对已有虚拟机 Hadoop01（用户名：root 密码：123456）进行克隆，得到 Hadoop02 和 Hadoop03 虚拟机，截取克隆后得到的三台虚拟机；图片保存到考生文件夹下，并命名为“1-1 克隆虚拟机.png”。（5 分）
- 开启三台虚拟机并按集群规划修改主机名，使用相关命令显示两台从节点修改后的主机名，截取命令及显示效果，图片保存到考生文件夹下，并命名为“1-2-1 主机名 1”、“1-2-2 主机名 2”。（5 分）
- 根据题目要求设置三台主机的 IP 地址，修改虚拟机网卡配置文件，将 node02 的修改内容截图，图片保存到考生文件夹下，并命名为“1-3 修改 ip 地址.png”。（5 分）

**任务二：集群主节点的配置（40 分）**

- Hadoop 的运行需要依赖 JDK，而 hadoop-env.sh 文件为 Hadoop 的运行环境配置文件，请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“2-1 配置 hadoop 运行环境.png”。（4 分）
- core-site.xml 文件为 Hadoop 的核心配置文件，用于定义系统级别的参数，设定 namenode 的主机名及端口、存放临时文件的目录，请提交该配置文件的修改内容截图，

图片保存到考生文件夹下，并命名为“2-2 修改 hadoop 核心配置文件.png”。（8 分）

- **hdfs-site.xml** 文件为 HDFS 核心配置文件，如文件副本的个数、块大小及是否使用强制权限等，设定 HDFS 存储文件的副本个数默认为 3、SecondaryNameNode 地址和端口，请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“2-3 修改 hdfs 核心配置文件.png”。（8 分）
- **mapred-site.xml** 文件是 MapRduce 的核心配置文件，用于指定 MapRduce 运行时的框架，请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“2-4 修改 MapReduce 核心文件.png”。（8 分）
- **yarn-site.xml** 文件是 YARN 框架的核心配置文件，需要指定 YARN 集群的管理者，请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“2-5 修改 yarn 核心配置文件.png”。（8 分）
- **slaves** 文件用于记录集群所有从节点主机名，请提交该配置文件的修改内容截图，图片保存到考生文件夹下，并命名为“2-6 修改 slaves 文件.png”。（4 分）

### 任务三：分发配置文件与测试 Hadoop 集群（35 分）

- 由于所修改的配置文件都存在于 Hadoop 安装目录下的 **etc/hadoop** 目录下，因此将主节点上的 Hadoop 安装目录下的 **etc/hadoop** 目录分发到两个从节点即可，将分发至从节点 **node02** 的命令截图，图片保存到考生文件夹下，并命名为“3-1 分发 hadoop 配置文件.png”。（5 分）
- 在主节点上执行相关命令格式化 **NameNode**，截取格式化成功界面，图片保存到考生文件夹下，并命名为“3-2 格式化 namenode.png”。（8 分）
- 使用相关指令分别启动 **hdfs** 和 **yarn** 集群，并使用 **jps** 进行查看，截取启动命令与 **node01** 上的 **jps** 查看结果，图片保存到考生文件夹下，并命名为“3-3-1 启动 Hadoop 集群.png”；截取 **node02** 上 **jps** 查看结果，图片保存到考生文件夹下，并命名为“3-3-2 查看 node02 启动情况.png”。（8 分）
- 使用相关命令在 HDFS 根目录下创建 **test** 目录，截取 web 界面目录创建情况，图片保存到考生文件夹下，并命名为“3-4 创建目录.png”。（6 分）

- 使用 `hdfs shell` 命令将 Hadoop 安装目录下的 NOTICE.txt 文档上传到 HDFS 根目录下的 test 目录下，截取 web 界面上文档上传情况，图片保存到考生文件夹下，并命名为“3-5 上传文件.png”。（8 分）

### 良好的习惯和职业素养（10 分）

- 答题过程中保持安静。（3 分）
- 有问题举手示意，有礼貌。（3 分）
- 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。（4 分）

## （2）实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。计算机安装有 Vmware workstation15, XShell, JDK1.8, 并有 Hadoop2.7.4 安装包和 CentOS7 64 位镜像文件，安装 Office 办公软件。

## （3）考核时量及要求

- 考核时长 180 分钟；
- 请在桌面建立以“学号\_题号”命名的成果文件夹，所有提交文件均放在该目录下，操作过程的截图需按照题目要求，文件夹命名例如：201965110001\_J3-1-1。

## （4）评分细则

表 10-2 任务一评分细则（15 分）

| 序号 | 评分项          | 分值  | 评分细则                                                                                           |
|----|--------------|-----|------------------------------------------------------------------------------------------------|
| 1  | 正确克隆虚拟机      | 5 分 | 和参考答案一致满分；<br>正确克隆两台虚拟机命名为 Hadoop02 和 Hadoop03，得 3 分；<br>考生文件夹中有 1-1 克隆虚拟机.png 文件，得 2 分。       |
| 2  | 正确修改主机名      | 5 分 | 和参考答案一致满分；<br>两台从节点修改主机名正确，得 3 分；<br>考生文件夹中有 1-2-1 修改主机名 1.png 文件和 1-2-2 修改主机名 2.png 文件，得 2 分。 |
| 3  | 正确修改 IP 地址文件 | 5 分 | 和参考答案一致满分；<br>正确打开虚拟机网卡配置文件，得 1 分；<br>配置文件内容正确，得 3 分；                                          |

|  |  |  |                                    |
|--|--|--|------------------------------------|
|  |  |  | 考生文件夹中有 1-3 修改 ip 地址.png 文件，得 1 分。 |
|--|--|--|------------------------------------|

表 10-3 任务二评分细则（40 分）

| 序号 | 评分项                     | 分值  | 评分细则                                                                                                                                                        |
|----|-------------------------|-----|-------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 修改 hadoop-env.sh 配置文件   | 4 分 | 和参考答案一致满分；<br>hadoop-env.sh 配置正确（路径配置正确），得 2 分；<br>考生文件夹中有 2-1 修改 hadoop-env.sh 文件.png，得 2 分。                                                               |
| 2  | 修改 core-site.xml 配置文件   | 8 分 | 和参考答案一致满分；<br>正确配置 namenode 主机名及端口，得 3 分；<br>正确配置临时文件的目录，得 3 分；<br>考生文件夹中有 2-2 修改 hadoop 核心配置文件.png 文件，得 2 分。                                               |
| 3  | 修改 hdfs-site.xml 配置文件   | 8 分 | 和参考答案一致满分；<br>正确配置 hdfs-site.xml 文件，得 6 分；<br>考生文件夹中有 2-3 修改 hdfs 核心配置文件.png 文件，得 2 分。                                                                      |
| 4  | 修改 mapred-site.xml 配置文件 | 8 分 | 和参考答案一致满分；<br>成功复制 mapred-site.xml.template 并重命名为 mapred-site.xml，得 2 分；<br>正确配置 mapred-site.xml 文件内容，得 4 分；<br>考生文件夹中有 2-4 修改 MapReduce 核心文件.png 文件，得 2 分。 |
| 5  | 修改 yarn-site.xml 配置文件   | 8 分 | 和参考答案一致满分；<br>正确配置 yarn-site.xml 文件，得 6 分；<br>考生文件夹中有 2-5 修改 yarn 核心配置文件.png 文件，得 2 分。                                                                      |
| 6  | 修改 slaves 配置文件          | 4 分 | 和参考答案一致满分；<br>正确配置 slaves 文件内容，得 2 分；<br>考生文件夹中有 2-6 修改 slaves 文件.png 文件，得 2 分。                                                                             |

表 10-4 任务三评分细则（35 分）

| 序号 | 评分项 | 分值 | 评分细则 |
|----|-----|----|------|
|----|-----|----|------|

|   |                   |     |                                                                                                                                |
|---|-------------------|-----|--------------------------------------------------------------------------------------------------------------------------------|
| 1 | 分发配置文件            | 5 分 | 和参考答案一致满分；<br>分发配置文件到从节点命令正确，得 3 分；<br>考生文件夹中有 3-1 分发 hadoop 配置文件.png 文件，得 2 分。                                                |
| 2 | 格式化 NameNode      | 8 分 | 和参考答案一致满分；<br>格式化成功，得 6 分；<br>考生文件夹中有 3-2 格式化 namenode.png 文件，得 2 分                                                            |
| 3 | 启动 hdfs 和 yarn 集群 | 8 分 | 和参考答案一致满分；<br>启动集群命令正确且 node01 和 node02 上运行的进程正确，得 6 分；<br>考生文件夹中有 3-3-1 启动 Hadoop 集群.png 和 3-3-2 查看 node02 启动情况.png 文件，得 2 分。 |
| 4 | web 界面下创建目录       | 6 分 | 和参考答案一致满分；<br>HDFS 根目录下创建目录正确，得 4 分；<br>考生文件夹中有 3-4 创建目录.png 文件，得 2 分                                                          |
| 5 | web 界面上上传文档       | 8 分 | 和参考答案一致满分；<br>文档成功上传至 HDFS 根目录下的 test 目录下，得 6 分；<br>考生文件夹中有 3-5 上传文件.png 文件，得 2 分                                              |

表 10-5 职业素质评分细则（10 分）

| 序号 | 评分项  | 分值  | 评分细则                              |
|----|------|-----|-----------------------------------|
| 1  | 操作规范 | 3 分 | 答题过程中保持安静，不损坏机器；                  |
| 2  | 文明考试 | 3 分 | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌；      |
| 3  | 工位整理 | 4 分 | 答题完毕立场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |

## 模块三 Python 数据采集技术

### 项目 1：数字内容服务行业数据采集与可视化

#### 1. 试题编号：J3-1-1：时光网电影票房数据采集与可视化

##### (1) 任务描述

电影业中，票房已经成为衡量一部电影是否成功的重要指标之一，可以用观众人数或门票收入来计算，一般直接用来衡量一部电影的热播程度。“电影票房”网是一个提供电影票房数据的网站。某企业想要了解电影票房等相关信息，因此需要进行数据采集的相关工作，具体任务如下：

##### 任务一：实现会员卡号生成器关键算法并绘制流程图（30 分）

某企业数据采集项目组在了解时光网的过程中，发现该网站的会员卡生成规律是，会员卡号由 6 位组成，前 3 位是 610，后面的依次是 001, 002, 003...100, 所有会员卡初始化密码为"000000"，现要求项目组按要求生成 100 个会员卡的卡号，输出格式如下：

```
100个会员卡号及密码是：{610001: '000000', 610002: '000000', 610003: '000000', 610004: '000000', 610005: '000000', 610006: '000000', 610007: '000000', 610008: '000000', 610009: '000000', 610010: '000000', 610011: '000000', 610012: '000000', 610013: '000000', 610014: '000000', 610015: '000000', 610016: '000000', 610017: '000000', 610018: '000000', 610019: '000000', 610020: '000000', 610021: '000000', 610022: '000000', 610023: '000000', 610024: '000000', 610025: '000000', 610026: '000000', 610027: '000000', 610028: '000000', 610029: '000000', 610030: '000000', 610031: '000000', 610032: '000000', 610033: '000000', 610034: '000000', 610035: '000000', 610036: '000000', 610037: '000000', 610038: '000000', 610039: '000000', 610040: '000000', 610041: '000000', 610042: '000000', 610043: '000000', 610044: '000000', 610045: '000000', 610046: '000000', 610047: '000000', 610048: '000000', 610049: '000000', 610050: '000000', 610051: '000000', 610052: '000000', 610053: '000000', 610054: '000000', 610055: '000000', 610056: '000000', 610057: '000000', 610058: '000000', 610059: '000000', 610060: '000000', 610061: '000000', 610062: '000000', 610063: '000000', 610064: '000000', 610065: '000000', 610066: '000000', 610067: '000000', 610068: '000000', 610069: '000000', 610070: '000000', 610071: '000000', 610072: '000000', 610073: '000000', 610074: '000000', 610075: '000000', 610076: '000000', 610077: '000000', 610078: '000000', 610079: '000000', 610080: '000000', 610081: '000000', 610082: '000000', 610083: '000000', 610084: '000000', 610085: '000000', 610086: '000000', 610087: '000000', 610088: '000000', 610089: '000000', 610090: '000000', 610091: '000000', 610092: '000000', 610093: '000000', 610094: '000000', 610095: '000000', 610096: '000000', 610097: '000000', 610098: '000000', 610099: '000000', 610100: '000000'}
```

具体要求如下：

- 绘制会员卡号生成器程序的流程图，要求流程图流程完整，代表流程的图形正确，能够体现循环结构语句的流程。流程图以图片的形式提交，保存到考生文件夹下，并命名为 test1.png（10 分）
- 使用循环语句，并正确设定循环。（4 分）
- 正确生成卡号。（3 分）
- 为每个卡号配对正确的初始密码，并成对保存卡号和密码在字典中。（3 分）
- 能够按题目要求的格式正确输出结果。（5 分）
- 关键代码有写注释，并且表达清晰。（3 分）
- 程序原文件命名为 test2.py，并保存到考生文件夹中 T1 目录下。（2 分）

##### 任务二：数据采集（60 分）

数据采集项目组接到对时光网的电影票房的数据采集任务，要求新建“电影信息.py”，并在代码中实现以下功能或者使用 scrapy 框架实现：

- 获得网页源代码，截取相应的代码，命名为“T1-1.jpg”。访问网站（http://localhost:8080/%E6%97%B6%E5%85%89%E7%BD%91/topList.json），使用 Python 中的 requests 模块向该网站发送请求并获得网页的源代码。要求能够定义请求头、发送请求、查看网页响应状态码、获得网页源代码，截取相应的代码，命名为“T1-1.jpg”。

（12 分）

- 解析网页数据。选用适当的解析方法（re 模块、lxml 模块或者 BeautifulSoup 模块）对网页源代码进行解析，解析的数据包括：电影名、导演、主演、评分、上映时间、主要奖项。要求解析数据的前 5 条进行打印显示，并将结果进行截图，命名为“T1-2.jpg”。

（32 分）

样例如下：

| 电影名    | 导演         | 主演       | 评分  | 上映时间            | 主要奖项 |
|--------|------------|----------|-----|-----------------|------|
| 肖申克的救赎 | “弗兰克·德拉邦特” | “蒂姆·罗宾斯” | 9.2 | 1994 年 9 月 23 日 | 最佳影片 |

- 数据存储。采集的数据前 100 条保存为“电影信息.csv”文件。（8 分）
- 代码注释。在代码中有关键代码的注释。（4 分）
- 将“电影信息.csv”文件、“电影信息.py”文件、“T1-1.jpg”、“T1-2.jpg”文件保存到考生文件夹 T1 目录下。（4 分）

**良好编程习惯和职业素养（10 分）**

- 答题过程中保持安静。（3 分）
- 有问题举手示意，有礼貌。（3 分）
- 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。（4 分）

**（2）实施条件**

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。计算机安装有 Python、PyCharm 集成开发环境、Echarts、VSCode，Python 中安装 requests、

re、BS4、numpy、pandas、matplotlib、pycharts 等模块，安装 Office 办公软件，安装绘制流程图的软件-亿图图示。

### (3) 考核时量及要求

- 考核时长 180 分钟；
- 请在桌面建立以“学号\_题号”命名的成果文件夹，所有提交文件均放在该目录下。  
文件夹命名例如：201965110001\_J2\_1；
- 将各任务源文件夹分别保存为 task1、task2、task3，运行结果分别截图。

### (4) 评分细则

表 1-1 任务一评分细则（30 分）

| 序号 | 评分项       | 分值   | 评分细则                                                                                                              |
|----|-----------|------|-------------------------------------------------------------------------------------------------------------------|
| 1  | 开发环境的正确使用 | 2 分  | 未按要求提交正确格式的源文件，扣 2 分。                                                                                             |
| 2  | 流程图设计合理   | 10 分 | 未按要求提交正确命名的流程图截图，扣 2 分；<br>流程图中绘图使用的符号不当，每个错误符号扣 1 分，扣完为止；<br>流程图中需体现循环语句，未体现扣 4 分；<br>流程图未能完全体现题目的处理逻辑，酌情扣 1-5 分 |
| 3  | 程序设计合理    | 13 分 | 变量名命名不规范或者未注释，每处扣 1 分；<br>关键代码未写注释，每处扣 1 分；<br>未完全使用 for 的双重循环语句，酌情扣 3-6 分<br>未正确生成卡号，扣 3 分<br>未正确配置卡号密码，扣 3 分    |
| 4  | 功能实现      | 5 分  | 未按题目要求输出结果，扣 2 分<br>输出结果不正确，扣 3 分<br>程序代码运行无法按题目要求显示结果，扣 5 分                                                      |

表 1-2 任务二 评分细则（60 分）

| 序号 | 评分项     | 分值   | 评分细则                                                                                                                                                                |
|----|---------|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 获得网页源代码 | 12 分 | T1-1.jpg 中和参考答案一致满分 12 分；<br>T1 文件夹中有 T1-1.jpg，得 2 分；<br>T1-1.jpg 中定义了请求头，得 4 分；<br>T1-1.jpg 中发送请求，得 2 分；<br>T1-1.jpg 中查看网页响应状态码，得 2 分；<br>T1-1.jpg 中获得网页源代码，得 2 分。 |
| 2  | 解析网页数据  | 32 分 | T1-2.jpg 中和参考答案一致满分 32 分；                                                                                                                                           |

|   |            |     |                                                                                                                                                                                                       |
|---|------------|-----|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|   |            |     | T1 文件夹中有 T1-2.jpg，得 2 分；<br>T1-2.jpg 中打印的电影名正确，得 4 分；<br>T1-2.jpg 中打印的导演正确，得 8 分；<br>T1-2.jpg 中打印的主演正确，得 8 分；<br>T1-2.jpg 中打印的评分正确，得 4 分；<br>T1-2.jpg 中打印的上映时间正确，得 4 分；<br>T1-2.jpg 中打印的主要奖项正确，得 4 分。 |
| 3 | 数据存储       | 8 分 | “电影信息.csv”文件中按要求存储正确的数据，得 8 分。                                                                                                                                                                        |
| 4 | 源代码命名正确并保存 | 4 分 | T1 文件夹中有“电影信息.py”文件，得 4 分。                                                                                                                                                                            |
| 5 | 代码注释       | 4 分 | 代码中有注释，得 4 分。                                                                                                                                                                                         |

表 1-3 职业素养评分细则（10 分）

| 序号 | 评分项  | 分值  | 评分细则                              |
|----|------|-----|-----------------------------------|
| 1  | 操作规范 | 3 分 | 答题过程中保持安静，不损坏机器                   |
| 2  | 文明考试 | 3 分 | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌；      |
| 3  | 工位整理 | 4 分 | 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |

## 2.试题编号：J3-1-2：豆瓣小说数据采集与可视化

### （1）任务描述

豆瓣（douban）是一个社区网站，创立于 2005 年 3 月 6 日。该网站以书影音起家，提供关于书籍、电影、音乐等作品的信息，无论描述还是评论都由用户提供。“豆瓣读书”中的小说标签提供了小说的相关信息。请以豆瓣网站中的小说数据为目标进行数据采集的工作，

具体任务如下：

### 任务一：密码破解的关键算法并绘制流程图（30 分）

实习生小张在熟悉数据采集任务的过程中，发现有一些社区网站用户的密码进行了加密保存。例如：现有一个字符串，该字符串为加密后的一个密码，具体情况是，该字符串为 "ixo678"，其中 o 是字母；现猜测其加密码规则是：在真实密码的 ASCII 码的值上加上了 5，现要求将该字符串进行解密，得到真正的密码。

具体要求如下：

- 绘制字符串解密程序的流程图，要求流程图流程完整，代表流程的图形正确，能够体现循环结构语句的流程。流程图以图片的形式提交，保存到考生文件夹下，并命名为 test1.png（10 分）
- 使用 for 语句，并正确设定循环。（5 分）
- 正确使用 ord 函数或 chr 函数进行密码转换。（5 分）
- 能够按题目要求输出真正的密码。（5 分）
- 关键代码有写注释，并且表达清晰。（3 分）
- 程序原文件命名为 test2.py，并保存到考生文件夹中 T1 目录下。（2 分）

### 任务二：数据采集（60 分）

新建“豆瓣小说.py”，并在代码中实现以下功能，或者使用 scrapy 框架实现：

- 获得网页源代码，截取相应的代码，命名为“T1-1.jpg”。访问网站  
([http://localhost:8080/%E8%B1%86%E7%93%A3%E5%B0%8F%E8%AF%B4/%E8%B1%86%E7%93%A3%E5%9B%BE%E4%B9%A6%E6%A0%87%E7%AD%BE\\_%20%E5%B0%8F%E8%AF%B4.html](http://localhost:8080/%E8%B1%86%E7%93%A3%E5%B0%8F%E8%AF%B4/%E8%B1%86%E7%93%A3%E5%9B%BE%E4%B9%A6%E6%A0%87%E7%AD%BE_%20%E5%B0%8F%E8%AF%B4.html))，使用 Python 中的 requests 模块向该网站发送请求并获得网页的源代码。要求能够定义请求头、发送请求、查看网页响应状态码、获得网页源代码，截取相应的代码，命名为“T1-1.jpg”。（12 分）
- 解析网页数据。选用适当的解析方法（re 模块、lxml 模块或者 BeautifulSoup 模块）对网页源代码进行解析，解析的数据包括：书名、作者、出版社、出版日期、价格，评分，评价数。要求解析数据的前 5 条进行打印显示，并将结果进行截图，命名为

“T1-2.jpg”。(32 分)

样例如下:

| 书名 | 作者、出版社、出版日期、价格               | 评分  | 评价数    |
|----|------------------------------|-----|--------|
| 活着 | 余华/作家出版社/2012-8-1/20.00<br>元 | 9.3 | 227720 |

- 数据存储。采集的数据保存为“豆瓣小说.csv”文件。(8 分)
- 代码注释。在代码中有关键代码的注释。(4 分)
- 将“豆瓣小说.csv”文件、“豆瓣小说.py”文件、“T1-1.jpg”、“T1-2.jpg”文件保存到考生文件夹 T1 目录下。(4 分)

#### 良好编程习惯和职业素养 (20 分)

- 答题过程中保持安静。(6 分)
- 有问题举手示意,有礼貌。(6 分)
- 答题完毕离场时整理好工位,不关闭计算机,并按要求显示答案保存界面。(8 分)

### (2) 实施条件

安排在教学实训机房,计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。计算机安装有 Python、PyCharm 集成开发环境、Echarts、VSCode, Python 中安装 requests、re、BS4、numpy、pandas、matplotlib、pycharts 等模块,安装 Office 办公软件,安装绘制流程图的软件-亿图图示。

### (3) 考核时量及要求

- 考核时长 180 分钟;
- 请在桌面建立以“学号\_题号”命名的成果文件夹,所有提交文件均放在该目录下。  
文件夹命名例如:201965110001\_J2\_1;
- 将各任务源文件夹分别保存为 task1、task2、task3,运行结果分别截图

### (4) 评分细则

表 2-1 任务一评分细则 (30 分)

| 序号 | 评分项 | 分值 | 评分细则 |
|----|-----|----|------|
|----|-----|----|------|

|   |           |      |                                                                                                                       |
|---|-----------|------|-----------------------------------------------------------------------------------------------------------------------|
| 1 | 开发环境的正确使用 | 2 分  | 未按要求提交正确格式的源文件，扣 2 分。                                                                                                 |
| 2 | 流程图设计合理   | 10 分 | 未按要求提交正确命名的流程图截图，扣 2 分；<br>流程图中绘图使用的符号不当，每个错误符号扣 1 分，扣完为止；<br>流程图中需体现循环语句，未体现扣 4 分；<br>流程图未能完全体现题目的处理逻辑，酌情扣 1-5 分。    |
| 3 | 程序设计合理    | 13 分 | 变量名命名不规范或者未注释，每处扣 1 分；<br>关键代码未写注释，每处扣 1 分；<br>未正确使用 <code>ord()</code> 函数，扣 2 分<br>未正确使用 <code>chr()</code> 函数，扣 2 分 |
| 4 | 功能实现      | 5 分  | 未按题目要求输出真正的密码，扣 2 分<br>程序代码运行无法显示结果，扣 5 分                                                                             |

表 2-2 任务二评分细则（60 分）

| 序号 | 评分项        | 分值   | 评分细则                                                                                                                                                                                                 |
|----|------------|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 获得网页源代码    | 12 分 | T1-1.jpg 中和参考答案一致满分 12 分；<br>T1 文件夹中有 T1-1.jpg，得 2 分；<br>T1-1.jpg 中定义了请求头，得 4 分；<br>T1-1.jpg 中发送请求，得 2 分；<br>T1-1.jpg 中查看网页响应状态码，得 2 分；<br>T1-1.jpg 中获得网页源代码，得 2 分。                                  |
| 2  | 解析网页数据     | 32 分 | T1-2.jpg 中和参考答案一致满分 32 分；<br>T1 文件夹中有 T1-2.jpg，得 2 分；<br>T1-2.jpg 中打印的电影名正确，得 6 分；<br>T1-2.jpg 中打印的评分正确，得 6 分；<br>T1-2.jpg 中打印的类型正确，得 6 分；<br>T1-2.jpg 中打印的主演正确，得 6 分；<br>T1-2.jpg 中打印的上映时间正确，得 6 分。 |
| 3  | 数据存储       | 8 分  | “猫眼电影.csv”文件中按要求存储正确的数据，得 8 分。                                                                                                                                                                       |
| 4  | 源代码命名正确并保存 | 4 分  | T1 文件夹中有“猫眼电影.py”文件，得 4 分。                                                                                                                                                                           |
| 5  | 代码注释       | 4 分  | 代码中有注释，得 4 分。                                                                                                                                                                                        |

表 2-3 职业素养评分细则（10 分）

| 序号 | 评分项  | 分值  | 评分细则                              |
|----|------|-----|-----------------------------------|
| 1  | 操作规范 | 3 分 | 答题过程中保持安静，不损坏机器                   |
| 2  | 文明考试 | 3 分 | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌；      |
| 3  | 工位整理 | 4 分 | 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |

### 3.试题编号：J3-1-3：猫眼电影数据采集与可视化

#### （1）任务描述

猫眼电影是美团旗下的一家集媒体内容、在线购票、用户互动社交、电影衍生品销售等服务的一站式电影互联网平台，也是人们了解电影信息的重要渠道。某影视公司想要了解最新的电影信息以此为依据来制定自己的企业发展计划，请帮助该公司完成电影数据的采集、处理和展示的工作，具体任务如下：

#### 任务一：实现电影评分排序功能关键算法并绘制流程图（30 分）

现获得猫眼电影网站上的一个样本数据，要求对该样本数据中的评分进行升序排列并输出结果。样本数据是："流浪地球 2,9.3;我和我的家乡,9.1;灰姑娘,8.5;爱情公寓,6.8;妖猫传,7.5"，结果输出示例如下：

第5名是：《爱情公寓》，评分为6.8  
 第4名是：《妖猫传》，评分为7.5  
 第3名是：《灰姑娘》，评分为8.5  
 第2名是：《我和我的家乡》，评分为9.1  
 第1名是：《流浪地球2》，评分为9.3

具体要求如下：

- 绘制电影评分排序程序的流程图，要求流程图流程完整，代表流程的图形正确，能

够体现循环结构语句的流程。流程图以图片的形式提交，保存到考生文件夹下，并命名为 test1.png（10 分）

- 正确保存样本数据为字符串类型。（2 分）
- 将样本数据进行处理，保存到字典中。（5 分）
- 根据题目要求，按电影的评分进行升序排列，要求使用 sorted 方法实现。（5 分）
- 能够按题目要求的格式正确输出结果。（3 分）
- 关键代码有写注释，并且表达清晰。（3 分）
- 程序原文件命名为 test2.py，并保存到考生文件夹中 T1 目录下。（2 分）

## 任务二：数据采集（60 分）

新建“猫眼电影.py”，并在代码中实现以下功能，或者使用 scrapy 框架实现：

- 获得网页源代码，截取相应的代码，命名为“T1-1.jpg”。访问网站  
(http://localhost:8080/%E7%8C%AB%E7%9C%BC%E7%94%B5%E5%BD%B1/%E7%BB%8F%E5%85%B8%E5%BD%B1%E7%89%87\_%E7%94%B5%E5%BD%B1%E5%A4%A7%E5%85%A8\_%E7%BB%8F%E5%85%B8%E9%AB%98%E6%B8%85%E7%94%B5%E5%BD%B1-%E7%8C%AB%E7%9C%BC%E7%94%B5%E5%BD%B1.html)，使用 Python 中的 requests 模块向该网站发送请求并获得网页的源代码。要求能够定义请求头、发送请求、查看网页响应状态码、获得网页源代码，截取相应的代码，命名为“T1-1.jpg”。（12 分）
- 解析网页数据。选用适当的解析方法（re 模块、lxml 模块或者 BeautifulSoup 模块）对网页源代码进行解析，解析的数据包括：电影名、评分、类型、国家、上映时间。要求解析数据的前 5 条进行打印显示，并将结果进行截图，命名为“T1-2.jpg”。（32 分）

样例如下：

| 电影名 | 评分  | 类型       | 主演        | 上映时间       |
|-----|-----|----------|-----------|------------|
| 八佰  | 9.2 | 剧情/战争/历史 | 王千源/张译/姜武 | 2020-08-21 |

- 数据存储。采集的数据保存为“猫眼电影.csv”文件。（8 分）

- 代码注释。在代码中有关键代码的注释。（4分）
- 将“猫眼电影.csv”文件、“猫眼电影.py”文件、“T1-1.jpg”、“T1-2.jpg”文件保存到考生文件夹 T1 目录下。（4分）

### 良好编程习惯和职业素养（10分）

- 答题过程中保持安静。（3分）
- 有问题举手示意，有礼貌。（3分）
- 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。（4分）

## （2）实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。计算机安装有 Python、PyCharm 集成开发环境、Echarts、VSCode，Python 中安装 requests、re、BS4、numpy、pandas、matplotlib、pyecharts 等模块，安装 Office 办公软件，安装绘制流程图的软件-亿图图示。

## （3）考核时量及要求

- 考核时长 180 分钟；
- 请在桌面建立以“学号\_题号”命名的成果文件夹，所有提交文件均放在该目录下。

文件夹命名例如：201965110001\_J2\_1；

将各任务源文件夹分别保存为 task1、task2、task3，运行结果分别截图。

## （4）评分细则

表 3-1 任务一评分细则（30 分）

| 序号 | 评分项       | 分值   | 评分细则                                                                                                               |
|----|-----------|------|--------------------------------------------------------------------------------------------------------------------|
| 1  | 开发环境的正确使用 | 2 分  | 未按要求提交正确格式的源文件，扣 2 分。                                                                                              |
| 2  | 流程图设计合理   | 10 分 | 未按要求提交正确命名的流程图截图，扣 2 分；<br>流程图中绘图使用的符号不当，每个错误符号扣 1 分，扣完为止；<br>流程图中需体现循环语句，未体现扣 4 分；<br>流程图未能完全体现题目的处理逻辑，酌情扣 1-5 分。 |
| 3  | 程序设计合理    | 13 分 | 变量名命名不规范或者未注释，每处扣 1 分；<br>关键代码未写注释，每处扣 1 分；<br>初始样本数据未保存为字符串类型，扣 2 分<br>对样本数据进行处理后未保存为字典类型，扣 5 分<br>未按要求进行排序，扣 5 分 |

|   |      |     |                                                              |
|---|------|-----|--------------------------------------------------------------|
| 4 | 功能实现 | 5 分 | 未按题目要求输出结果，扣 2 分<br>输出结果不正确，扣 3 分<br>程序代码运行无法按题目要求显示结果，扣 5 分 |
|---|------|-----|--------------------------------------------------------------|

表 3-2 任务二评分细则（30 分）

| 序号 | 评分项        | 分值   | 评分细则                                                                                                                                                                                                        |
|----|------------|------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 获得网页源代码    | 12 分 | T1-1. jpg 中和参考答案一致满分 12 分；<br>T1 文件夹中有 T1-1. jpg，得 2 分；<br>T1-1. jpg 中定义了请求头，得 4 分；<br>T1-1. jpg 中发送请求，得 2 分；<br>T1-1. jpg 中查看网页响应状态码，得 2 分；<br>T1-1. jpg 中获得网页源代码，得 2 分。                                   |
| 2  | 解析网页数据     | 32 分 | T1-2. jpg 中和参考答案一致满分 32 分；<br>T1 文件夹中有 T1-2. jpg，得 2 分；<br>T1-2. jpg 中打印的电影名正确，得 6 分；<br>T1-2. jpg 中打印的评分正确，得 6 分；<br>T1-2. jpg 中打印的类型正确，得 6 分；<br>T1-2. jpg 中打印的主演正确，得 6 分；<br>T1-2. jpg 中打印的上映时间正确，得 6 分。 |
| 3  | 数据存储       | 8 分  | “猫眼电影.csv”文件中按要求存储正确的数据，得 8 分。                                                                                                                                                                              |
| 4  | 源代码命名正确并保存 | 4 分  | T1 文件夹中有“猫眼电影.py”文件，得 4 分。                                                                                                                                                                                  |
| 5  | 代码注释       | 4 分  | 代码中有注释，得 4 分。                                                                                                                                                                                               |

表 3-3 职业素养评分细则（10 分）

| 序号 | 评分项  | 分值  | 评分细则                              |
|----|------|-----|-----------------------------------|
| 1  | 操作规范 | 3 分 | 答题过程中保持安静，不损坏机器                   |
| 2  | 文明考试 | 3 分 | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌；      |
| 3  | 工位整理 | 4 分 | 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |

#### 4.试题编号：J3-1-4：淘书网图书信息的采集与可视化

##### （1）任务描述

淘书网是国内图书品种最全的图书网站之一，是国内特价稀缺图书品种涵盖最多的网上书店。现需要对图书信息进行了解，具体任务如下：

**任务一：实现图书购买记录的关键算法并绘制流程图（30 分）**

某人到淘书网上购买了以下物品，现需要对购物金额进行统计。清单如下图所示：

| 图书名称    | 价格    |
|---------|-------|
| 视觉 001  | 26.5  |
| 美学散步    | 44    |
| 中华经典十二部 | 18    |
| 百年孤独套装  | 148.8 |
| 效率人生    | 24    |
| 古代汉语词典  | 102   |

以上数据保存为字典，《中化经典十二部》的金额统计出错，请修改为 28，修改需要通过代码体现，不能在定义时直接进行修改。输出显示为：您购买 n 本图书，共计 n 元。保留 2 位小数。输出格式如下：

您购买6本图书，共计373.30元

具体要求如下：

- 绘制图书购买记录程序的流程图，要求流程图流程完整，代表流程的图形正确。流程图以图片的形式提交，保存到考生文件夹下，并命名为 test1.png（10 分）
- 正确定义字典，并体现保存了图书购买记录原始数据。（3 分）
- 正确修改“《中化经典十二部》”的价格。（3 分）
- 正确统计合计金额。（4 分）
- 能够按题目要求的格式正确输出结果，合计金额保留 2 位小数。（5 分）
- 关键代码有写注释，并且表达清晰。（3 分）
- 程序原文件命名为 test2.py，并保存到考生文件夹中 T1 目录下。（2 分）

**任务一：数据采集（30 分）**

新建“淘书网.py”，并在代码中实现以下功能，或者使用 scrapy 框架实现：

- 获得网页源代码，截取相应的代码，命名为“T1-1.jpg。访问网站  
(<http://localhost:8080/%E6%B7%98%E4%B9%A6%E7%BD%91/%E6%96%87%E5%AD%A6%20-%20%E6%B7%98%E4%B9%A6%E7%BD%91.html>)，使用 Python 中的 requests 模块向该网站发送请求并获得网页的源代码。要求能够定义请求头、

发送请求、查看网页响应状态码、获得网页源代码，截取相应的代码，命名为“T1-1.jpg”。（12 分）

- 解析网页数据。选用适当的解析方法（re 模块、lxml 模块或者 BeautifulSoup 模块）对网页源代码进行解析，解析的数据包括：书名、出版社、出版日期、简介、原价、淘书价、折扣。要求解析数据的前 5 条进行打印显示，并将结果进行截图，命名为“T1-2.jpg”。（32 分）

样例如下：

| 书名    | 出版社       | 出版日期       | 简介                                                                      | 原价    | 淘书价   | 折扣   |
|-------|-----------|------------|-------------------------------------------------------------------------|-------|-------|------|
| 最美的诗歌 | 合肥工业大学出版社 | 2011-12-01 | 中国是诗的国度，中国文人有着诗化人生，中国人有着诗意生活。这是一种大美，是值得今人珍视的宝贵人文资源，也渗透到关于未来理想生活的愿景之中... | 26.00 | 20.30 | 73 折 |

- 数据存储。采集的数据保存为“淘书网.csv”文件。（8 分）
- 代码注释。在代码中有关键代码的注释。（4 分）
- 将“淘书网.csv”文件、“淘书网.py”文件、“T1-1.jpg”、“T1-2.jpg”文件保存到考生文件夹 T1 目录下。（4 分）

#### 良好编程习惯和职业素养（10 分）

- 答题过程中保持安静。（3 分）
- 有问题举手示意，有礼貌。（3 分）
- 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。（4 分）

#### （2）实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。计算机安装有 Python、PyCharm 集成开发环境、Echarts、VSCode，Python 中安装 requests、re、BS4、numpy、pandas、matplotlib、pyecharts 等模块，安装 Office 办公软件，安装绘制流程图的软件-亿图图示。

### (3) 考核时量及要求

- 考核时长 180 分钟；
- 请在桌面建立以“学号\_题号”命名的成果文件夹，所有提交文件均放在该目录下。  
文件夹命名例如：201965110001\_J2\_1；
- 将各任务源文件夹分别保存为 task1、task2、task3，运行结果分别截图。

### (4) 评分细则

表 4-1 任务一评分细则（30 分）

| 序号 | 评分项       | 分值   | 评分细则                                                                                                   |
|----|-----------|------|--------------------------------------------------------------------------------------------------------|
| 1  | 开发环境的正确使用 | 2 分  | 未按要求提交正确格式的源文件，扣 2 分。                                                                                  |
| 2  | 流程图设计合理   | 10 分 | 未按要求提交正确命名的流程图截图，扣 2 分；<br>流程图中绘图使用的符号不当，每个错误符号扣 1 分，<br>扣完为止；<br>流程图未能完全体现题目的处理逻辑，酌情扣 1-5 分。          |
| 3  | 程序设计合理    | 13 分 | 变量名命名不规范或者未注释，每处扣 1 分；<br>关键代码未写注释，每处扣 1 分；<br>未正确定义字典及保存数据，扣 3 分<br>未正确修改数据，扣 3 分<br>未正确统计出合计金额，扣 4 分 |
| 4  | 功能实现      | 5 分  | 未按题目要求输出结果，扣 2 分<br>输出结果不正确，扣 3 分<br>程序代码运行无法按题目要求显示结果，扣 5 分                                           |

表 4-2 任务二评分细则（30 分）

| 序号 | 评分项     | 分值   | 评分细则                                                                                                                                                               |
|----|---------|------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 获得网页源代码 | 12 分 | T1-1.jpg 中和参考答案一致满分 6 分；<br>T1 文件夹中有 T1-1.jpg，得 2 分；<br>T1-1.jpg 中定义了请求头，得 4 分；<br>T1-1.jpg 中发送请求，得 2 分；<br>T1-1.jpg 中查看网页响应状态码，得 2 分；<br>T1-1.jpg 中获得网页源代码，得 2 分。 |

|   |            |      |                                                                                                                                                                                                                                                               |
|---|------------|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2 | 解析网页数据     | 32 分 | T1-2.jpg 中和参考答案一致满分 32 分；<br>T1 文件夹中有 T1-2.jpg，得 2 分；<br>T1-2.jpg 中打印的书名正确，得 6 分；<br>T1-2.jpg 中打印的出版社正确，得 4 分；<br>T1-2.jpg 中打印的出版日期正确，得 4 分；<br>T1-2.jpg 中打印的简介正确，得 4 分；<br>T1-2.jpg 中打印的原价正确，得 4 分；<br>T1-2.jpg 中打印的淘书价正确，得 4 分；<br>T1-2.jpg 中打印的折扣正确，得 4 分。 |
| 3 | 数据存储       | 8 分  | “淘书网.csv”文件中按要求存储正确的数据，得 8 分。                                                                                                                                                                                                                                 |
| 4 | 源代码命名正确并保存 | 4 分  | T1 文件夹中有“淘书网.py”文件，得 4 分。                                                                                                                                                                                                                                     |
| 5 | 代码注释       | 4 分  | 代码中有注释，得 4 分。                                                                                                                                                                                                                                                 |

表 4-3 职业素养评分细则（10 分）

| 序号 | 评分项  | 分值  | 评分细则                              |
|----|------|-----|-----------------------------------|
| 1  | 操作规范 | 3 分 | 答题过程中保持安静，不损坏机器                   |
| 2  | 文明考试 | 3 分 | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌；      |
| 3  | 工位整理 | 4 分 | 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |

## 5.试题编号：J3-1-5：当当图书好评数据采集与可视化

### （1）任务描述

某图书馆为了满足读者需求，要从当当网上购买一批新书，但是图书选择却很困难。因此急切需要了解最新的图书信息，具体的任务如下：

#### 任务三：实现图书售货统计的关键算法并绘制流程图（30 分）

现需统计当当网上某一分类近三天的售货情况，要求按天求出当天的售货本数以及售货金额，请按要求进行格式进行结果输出。售货情况如下所示：

| 日期     | 图书名称 | 售出数量 | 图书单价 |
|--------|------|------|------|
| 11月24日 | 三字经  | 15   | 9.9  |
|        | 百家姓  | 25   | 6.1  |
|        | 千字文  | 10   | 16.4 |
| 11月5日  | 三字经  | 25   | 9.9  |
|        | 弟子规  | 5    | 12.3 |
|        | 孙子兵法 | 15   | 6    |
|        | 易经   | 10   | 50.3 |
| 11月26日 | 乡土中国 | 10   | 11.6 |
|        | 三字经  | 20   | 9.9  |
|        | 百家姓  | 15   | 6.1  |

结果输出要求如下所示：

```

11月24日
  三字经： 数量：15  单价：9.9
  百家姓： 数量：25  单价：6.1
  千字文： 数量：10  单价：16.4
  11月24日卖出图书50本，小计：465.0元
11月25日
  三字经： 数量：25  单价：9.9
  弟子规： 数量：5   单价：12.3
  孙子兵法： 数量：15  单价：6.0
  易经：    数量：10  单价：50.3
  11月25日卖出图书55本，小计：902.0元
11月26日
  乡土中国： 数量：10  单价：11.6
  三字经： 数量：20  单价：9.9
  百家姓： 数量：15  单价：6.1
  11月26日卖出图书45本，小计：405.5元
  
```

具体要求如下：

- 绘制当当网图书售货统计程序的流程图，要求流程图流程完整，代表流程的图形正确，能够体现循环结构语句的流程。流程图以图片的形式提交，保存到考生文件夹下，并命名为 test1.png（10分）
- 按要求创建4个字典保存售货信息中的内容。（4分）
- 使用循环语句，并正确设定循环。（2分）

- 正确统计每日售货本数及小计。（4 分）
- 能够按题目要求的格式正确输出结果，指定数据保留 1 位小数。（5 分）
- 关键代码有写注释，并且表达清晰。（3 分）
- 程序原文件命名为 test2.py，并保存到考生文件夹中 T1 目录下。（2 分）

**任务二：数据采集（60 分）**

新建“当当图书.py”，并在代码中实现以下功能，或者使用 scrapy 框架实现：

- 获得网页源代码，截取相应的代码，命名为“T1-1.jpg”。访问网站  
(http://localhost:8080/%E5%BD%93%E5%BD%93%E5%9B%BE%E4%B9%A6/%E3%80%90%E6%98%9F%E5%9B%BE%E4%B9%A6%E3%80%91%E7%95%85%E9%94%80%E6%A6%9C-%E8%BF%9130%E6%97%A55%E6%98%9F%E5%9B%BE%E4%B9%A6%E6%A6%9C-%E5%BD%93%E5%BD%935%E6%98%9F%E5%9B%BE%E4%B9%A6%E6%8E%92%E8%A1%8C%E6%A6%9C.html)，使用 Python 中的 requests 模块向该网站发送请求并获得网页的源代码。要求能够定义请求头、发送请求、查看网页响应状态码、获得网页源代码，截取相应的代码，命名为“T1-1.jpg”。  
(12 分)
- 解析网页数据。选用适当的解析方法（re 模块、lxml 模块或者 BeautifulSoup 模块）对网页源代码进行解析，解析的数据包括：书名、作者、评论数、五星好评、出版日期、原价、现价、折扣。要求解析数据的前 5 条进行打印显示，并将结果进行截图，命名为“T1-2.jpg”。（32 分）

样例如下：

| 图书名称 | 作者 | 评论数   | 五星评分次数 | 出版日期     | 原价  | 现价 | 折扣    |
|------|----|-------|--------|----------|-----|----|-------|
| 傅雷家书 | 傅雷 | 23830 | 23320  | 2020/8/1 | 198 | 99 | 5.0 折 |

- 数据存储。采集的数据保存为“当当图书.csv”文件。（8 分）
- 代码注释。在代码中有关键代码的注释。（4 分）
- 将“当当图书.csv”文件、“当当图书.py”文件、“T1-1.jpg”、“T1-2.jpg”文件保存到考生文件夹 T1 目录下。（4 分）

### 良好编程习惯和职业素养（10分）

- 答题过程中保持安静。（3分）
- 有问题举手示意，有礼貌。（3分）
- 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。（4分）

### （2）实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。计算机安装有 Python、PyCharm 集成开发环境、Echarts、VSCode，Python 中安装 requests、re、BS4、numpy、pandas、matplotlib、pyecharts 等模块，安装 Office 办公软件，安装绘制流程图的软件-亿图图示。

### （3）考核时量及要求

- 考核时长 180 分钟；
- 请在桌面建立以“学号\_题号”命名的成果文件夹，所有提交文件均放在该目录下。

文件夹命名例如：201965110001\_J2\_1；

- 将各任务源文件夹分别保存为 task1、task2、task3，运行结果分别截图。

### （4）评分细则

表 5-1 任务一评分细则（30 分）

| 序号 | 评分项       | 分值   | 评分细则                                                                                                               |
|----|-----------|------|--------------------------------------------------------------------------------------------------------------------|
| 1  | 开发环境的正确使用 | 2 分  | 未按要求提交正确格式的源文件，扣 2 分。                                                                                              |
| 2  | 流程图设计合理   | 10 分 | 未按要求提交正确命名的流程图截图，扣 2 分；<br>流程图中绘图使用的符号不当，每个错误符号扣 1 分，扣完为止；<br>流程图中需体现循环语句，未体现扣 4 分；<br>流程图未能完全体现题目的处理逻辑，酌情扣 1-5 分。 |
| 3  | 程序设计合理    | 13 分 | 变量名命名不规范或者未注释，每处扣 1 分；<br>关键代码未写注释，每处扣 1 分；<br>未正确设定 4 个字典，每个扣 1 分<br>未正确统计每日售货本数及小计，扣 4 分                         |
| 4  | 功能实现      | 5 分  | 未按题目要求输出结果，扣 2 分<br>输出结果不正确，扣 3 分<br>程序代码运行无法按题目要求显示结果，扣 5 分                                                       |

表 5-2 任务二评分细则（60 分）

| 序号 | 评分项        | 分值   | 评分细则                                                                                                                                                                                                                                                                                           |
|----|------------|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 获得网页源代码    | 12 分 | T1-1.jpg 中和参考答案一致满分 12 分；<br>T1 文件夹中有 T1-1.jpg，得 2 分；<br>T1-1.jpg 中定义了请求头，得 4 分；<br>T1-1.jpg 中发送请求，得 2 分；<br>T1-1.jpg 中查看网页响应状态码，得 2 分；<br>T1-1.jpg 中获得网页源代码，得 2 分。                                                                                                                            |
| 2  | 解析网页数据     | 32 分 | T1-2.jpg 中和参考答案一致满分 32 分；<br>T1 文件夹中有 T1-2.jpg，得 2 分；<br>T1-2.jpg 中打印的图书名称正确，得 4 分；<br>T1-2.jpg 中打印的作者正确，得 4 分；<br>T1-2.jpg 中打印的评论数正确，得 4 分；<br>T1-2.jpg 中打印的五星评分次数正确，得 4 分；<br>T1-2.jpg 中打印的出版日期正确，得 4 分；<br>T1-2.jpg 中打印的原价正确，得 4 分；<br>T1-2.jpg 中打印的现价正确，得 4 分；<br>T1-2.jpg 中打印的折扣正确，得 2 分。 |
| 3  | 数据存储       | 8 分  | “当当图书.csv”文件中按要求存储正确的数据，得 8 分。                                                                                                                                                                                                                                                                 |
| 4  | 源代码命名正确并保存 | 4 分  | T1 文件夹中有“当当图书.py”文件，得 4 分。                                                                                                                                                                                                                                                                     |
| 5  | 代码注释       | 4 分  | 代码中有注释，得 4 分。                                                                                                                                                                                                                                                                                  |

表 5-3 职业素养评分细则（10 分）

| 序号 | 评分项  | 分值  | 评分细则                         |
|----|------|-----|------------------------------|
| 1  | 操作规范 | 3 分 | 答题过程中保持安静，不损坏机器              |
| 2  | 文明考试 | 3 分 | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌； |

|   |      |     |                                   |
|---|------|-----|-----------------------------------|
| 3 | 工位整理 | 4 分 | 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |
|---|------|-----|-----------------------------------|

## 项目 2：社会服务行业数据采集与可视化

### 6.试题编号：J3-2-1：美食杰数据采集与可视化

#### (1) 任务描述

随着社会的发展，人民生活水平显著提高，对饮食的要求也越来越高。而年轻一代做饭基本要靠网络，请以美食杰网站为目标，获取相关的菜谱信息，具体任务如下所示：

#### 任务一：实现数据采集小组分组关键算法并绘制流程图（30 分）

现有数据采集项目组要对美食杰网站上的数据进行采集，为了更好的对数据进行分析，将数据采集项目组分 4 组按不同的数据特征进行采集，现有 12 名程序员，要求随机分配到各组中，每组有且只能有 3 名人员。人员名单如下：'石佳','陈浩拓','肖湘吉','丁庆玲','陈娅','陈雨杭','高家纯','刘腾','刘小叶','唐志荣','倪思理','周子涵'。格式输出示例如下（每组名单每次运行随机显示）：

```
数据采集第1组名单：刘小叶  陈浩拓  刘腾
数据采集第2组名单：高家纯  唐志荣  石佳
数据采集第3组名单：陈娅   肖湘吉  陈雨杭
数据采集第4组名单：周子涵  倪思理  丁庆玲
```

具体要求如下：

- 绘制数据采集小组分组程序的流程图，要求流程图流程完整，代表流程的图形正确，能够体现循环结构语句的流程。流程图以图片的形式提交，保存到考生文件夹下，并命名为 test1.png（10 分）
- 使用循环语句，并正确设定循环。（3 分）
- 正确使用随机函数。（2 分）
- 正确完成小组成员的分配。（3 分）
- 数据采集小组名单保存在嵌套列表中。（2 分）

- 显示格式要求：直接打印嵌套列表。能够按题目要求正确输出结果。（5分）
- 关键代码有写注释，并且表达清晰。（3分）
- 程序原文件命名为 test2.py，并保存到考生文件夹中 T1 目录下。（2分）

**任务二：数据采集（60 分）**

新建“美食杰.py”，并在代码中实现以下功能，或者使用 scrapy 框架实现：

- 获得网页源代码，截取相应的代码，命名为“T1-1.jpg”。访问美食杰网站  
(http://localhost:8080/%E7%BE%8E%E9%A3%9F%E6%9D%B0/%E8%8F%9C%E8%B0%B1%E5%A4%A7%E5%85%A8\_%E5%AE%B6%E5%BA%AD%E8%8F%9C%E8%B0%B1\_%E5%AE%B6%E5%BA%AD%E5%AE%9E%E7%94%A8%E8%8F%9C%E8%B0%B1%E5%A4%A7%E5%85%A8.html)，使用 Python 中的 requests 模块向该网站发送请求并获得网页的源代码。要求能够定义请求头、发送请求、查看网页响应状态码、获得网页源代码，截取相应的代码，命名为”T1-1.jpg”。（12分）
- 解析网页数据。选用适当的解析方法（re 模块、lxml 模块或者 BeautifulSoup 模块）对网页源代码进行解析，解析的数据包括：菜名、作者、评论数、人气数、步数、烹饪时间、工艺，菜谱信息。要求解析数据的前 5 条进行打印显示，并将结果进行截图，命名为“T1-2.jpg”。（32分）

样例如下：

| 菜名  | 作者  | 评论数 | 人气数 | 步数 | 烹饪时间  | 工艺      |
|-----|-----|-----|-----|----|-------|---------|
| 烤排骨 | 笑甜点 | 0   | 1   | 6  | 60 分钟 | 烤 / 家常味 |

- 数据存储。采集的数据保存为 “美食杰.csv”文件。（8分）
- 代码注释。在代码中有关键代码的注释。（4分）
- 将“美食杰.csv”文件、“美食杰.py”文件、“T1-1.jpg”、“T1-2.jpg”文件保存到考生文件夹 T1 目录下。（4分）

**良好编程习惯和职业素养（10 分）**

- 答题过程中保持安静。（3分）
- 有问题举手示意，有礼貌。（3分）

- 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。（4分）

## （2）实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。  
计算机安装有 Python、PyCharm 集成开发环境、Echarts、VSCode，Python 中安装 requests、re、BS4、numpy、pandas、matplotlib、pyecharts 等模块，安装 Office 办公软件，安装绘制流程图的软件-亿图图示。

## （3）考核时量及要求

- 考核时长 180 分钟；
- 请在桌面建立以“学号\_题号”命名的成果文件夹，所有提交文件均放在该目录下。  
文件夹命名例如：201965110001\_J2\_1；
- 将各任务源文件夹分别保存为 task1、task2、task3，运行结果分别截图。

## （4）评分细则

表 6-1 任务一评分细则（30 分）

| 序号 | 评分项       | 分值   | 评分细则                                                                                                               |
|----|-----------|------|--------------------------------------------------------------------------------------------------------------------|
| 1  | 开发环境的正确使用 | 2 分  | 未按要求提交正确格式的源文件，扣 2 分。                                                                                              |
| 2  | 流程图设计合理   | 10 分 | 未按要求提交正确命名的流程图截图，扣 2 分；<br>流程图中绘图使用的符号不当，每个错误符号扣 1 分，扣完为止；<br>流程图中需体现循环语句，未体现扣 4 分；<br>流程图未能完全体现题目的处理逻辑，酌情扣 1-5 分。 |
| 3  | 程序设计合理    | 13 分 | 变量名命名不规范或者未注释，每处扣 1 分；<br>关键代码未写注释，每处扣 1 分；<br>未正确使用随机数，扣 2 分<br>未正确完成小组成员的分配，扣 3 分<br>未正确保存小组成员的数据，扣 2 分          |
| 4  | 功能实现      | 5 分  | 未按题目要求输出结果，扣 2 分<br>输出结果不正确，扣 3 分<br>程序代码运行无法按题目要求显示结果，扣 5 分                                                       |

表 6-2 任务二评分细则（60 分）

| 序号 | 评分项 | 分值 | 评分细则 |
|----|-----|----|------|
|----|-----|----|------|

|   |            |      |                                                                                                                                                                                                                                                                                                   |
|---|------------|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1 | 获得网页源代码    | 12 分 | <p>T1-1.jpg 中和参考答案一致满分 12 分；</p> <p>T1 文件夹中有 T1-1.jpg，得 2 分；</p> <p>T1-1.jpg 中定义了请求头，得 4 分；</p> <p>T1-1.jpg 中发送请求，得 2 分；</p> <p>T1-1.jpg 中查看网页响应状态码，得 2 分；</p> <p>T1-1.jpg 中获得网页源代码，得 2 分。</p>                                                                                                    |
| 2 | 解析网页数据     | 32 分 | <p>T1-2.jpg 中和参考答案一致满分 32 分；</p> <p>T1 文件夹中有 T1-2.jpg，得 2 分；</p> <p>T1-2.jpg 中打印的菜名正确，得 6 分；</p> <p>T1-2.jpg 中打印的作者正确，得 4 分；</p> <p>T1-2.jpg 中打印的评论数正确，得 4 分；</p> <p>T1-2.jpg 中打印的人气正确，得 4 分；</p> <p>T1-2.jpg 中打印的步数正确，得 4 分；</p> <p>T1-2.jpg 中打印的时间正确，得 4 分；</p> <p>T1-2.jpg 中打印的工艺正确，得 4 分。</p> |
| 3 | 数据存储       | 8 分  | “美食杰.csv”文件中按要求存储正确的数据，得 8 分。                                                                                                                                                                                                                                                                     |
| 4 | 源代码命名正确并保存 | 4 分  | T1 文件夹中有“美食杰.py”文件，得 4 分。                                                                                                                                                                                                                                                                         |
| 5 | 代码注释       | 4 分  | 代码中有注释，得 4 分。                                                                                                                                                                                                                                                                                     |

表 6-3 职业素养评分细则（10 分）

| 序号 | 评分项  | 分值  | 评分细则                              |
|----|------|-----|-----------------------------------|
| 1  | 操作规范 | 3 分 | 答题过程中保持安静，不损坏机器                   |
| 2  | 文明考试 | 3 分 | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌；      |
| 3  | 工位整理 | 4 分 | 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |

## 7.试题编号：J3-2-2：湖南人才网数据采集与可视化

### （1）任务描述

湖南人才网创立于 2002 年初，凭借着雄厚的技术实力、专业的客户服务，以用人单位需求为导向，为企业纳才、人才求职提供高效快捷的交流平台。某应届毕业生想要了解本专业的就业请求，请将网页数据爬取下来并提供该学生。具体任务如下：

#### 任务一：实现数据转换及统计的算法并绘制流程图（30 分）

现得到部分招聘信息的数据，保存的数据类型为字符串，现要求将这些字符串对应保存到相应的列表中，并且根据 str3 提供的数据求出岗位的平均工资。数据如下：

str1="数据分析师,金融数据分析助理;链家数据分析师,数据分析测评工程师。数据统计分析专员"

str2="福建省三福百货有限公司;深圳市洋葱电子商务有限公司,广州市链家房地产代理有限公司,广州固诚信息咨询服务有限公司,广州思迈特软件有限公司"

str3="10000,6000,8000,8000,12000"

结果输出格式示例如下：

```
str1转换成列表显示为： ['数据分析师', '金融数据分析助理', '链家数据分析师', '数据分析测评工程师', '数据统计分析专员']
str2转换成列表显示为： ['福建省三福百货有限公司', '深圳市洋葱电子商务有限公司', '广州市链家房地产代理有限公司', '广州固诚信息咨询服务有限公司', '广州思迈特软件有限公司']
str3转换成列表显示为： [10000, 6000, 8000, 8000, 12000]
岗位平均工资为： 8800.0
```

具体要求如下：

- 绘制数据转换及统计程序的流程图，要求流程图流程完整，代表流程的图形正确。

流程图以图片的形式提交，保存到考生文件夹下，并命名为 test1.png（10 分）

- 使用字符串的相关函数处理字符串并保存为列表。（6 分）
- 正确统计平均工资，输出时保留 1 位小数。（5 分）
- 按题目要求正确输出结果。（4 分）
- 关键代码有写注释，并且表达清晰。（3 分）
- 程序原文件命名为 test2.py，并保存到考生文件夹中 T1 目录下。（2 分）

#### 任务二：数据采集（60 分）

新建“湖南人才网.py”，并在代码中实现以下功能，或者使用 scrapy 框架实现：

- 获得网页源代码，截取相应的代码，命名为“T1-1.jpg”。访问网站  
(http://localhost:8080/%E6%B9%96%E5%8D%97%E4%BA%BA%E6%89%8D%E7%BD%91/%E5%85%A8%E6%96%87%E6%90%9C%E7%B4%A2-%20%E6%B9%96%E5%8D%97%E4%BA%BA%E6%89%8D%E7%BD%91%20-%20www.hnrcsc.com.html)，使用 Python 中的 requests 模块向该网站发送请求并获得网页的源代码。要求能够定义请求头、发送请求、查看网页响应状态码、获得网页源代码，截取相应的代码，命名为”T1-1.jpg”。（12 分）
- 解析网页数据。选用适当的解析方法（re 模块、lxml 模块或者 BeautifulSoup 模块）对网页源代码进行解析，解析的数据包括：职位名称、公司名称、工作地点、更新日期、薪资范围、所属行业、企业性质、招聘人数。要求解析数据的前 5 条进行打印显示，并将结果进行截图，命名为“T1-2.jpg”。（32 分）

样例如下：

| 职位名称       | 公司名称         | 企业性质   | 更新日期       | 薪资范围      | 所属行业  | 招聘人数 |
|------------|--------------|--------|------------|-----------|-------|------|
| 设备租赁事业部总经理 | 宏林建设工程集团有限公司 | 有限责任公司 | 2020-08-28 | 15000 元以上 | 建筑工程业 | 1    |

- 数据存储。采集的数据保存为 “湖南人才网.csv”文件。（8 分）
- 代码注释。在代码中有关键代码的注释。（4 分）
- 将“湖南人才网.csv”文件、“湖南人才网.py”文件、“T1-1.jpg”、“T1-2.jpg”文件保存到考生文件夹 T1 目录下。（4 分）

**良好编程习惯和职业素养（10 分）**

- 答题过程中保持安静。（3 分）
- 有问题举手示意，有礼貌。（3 分）

- 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。（4分）

## （2）实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。计算机安装有 Python、PyCharm 集成开发环境、Echarts、VSCode，Python 中安装 requests、re、BS4、numpy、pandas、matplotlib、pyecharts 等模块，安装 Office 办公软件，安装绘制流程图的软件-亿图图示。

## （3）考核时量及要求

- 考核时长 180 分钟；
- 请在桌面建立以“学号\_题号”命名的成果文件夹，所有提交文件均放在该目录下。  
文件夹命名例如：201965110001\_J2\_1；
- 将各任务源文件夹分别保存为 task1、task2、task3，运行结果分别截图。

## （4）评分细则

表 7-1 任务一评分细则（30 分）

| 序号 | 评分项       | 分值   | 评分细则                                                                                      |
|----|-----------|------|-------------------------------------------------------------------------------------------|
| 1  | 开发环境的正确使用 | 2 分  | 未按要求提交正确格式的源文件，扣 2 分。                                                                     |
| 2  | 流程图设计合理   | 10 分 | 未按要求提交正确命名的流程图截图，扣 2 分；<br>流程图中绘图使用的符号不当，每个错误符号扣 1 分，扣完为止；<br>流程图未能完全体现题目的处理逻辑，酌情扣 1-5 分。 |
| 3  | 程序设计合理    | 13 分 | 变量名命名不规范或者未注释，每处扣 1 分；<br>关键代码未写注释，每处扣 1 分；<br>未正确处理字符串，每处扣 2 分                           |
| 4  | 功能实现      | 5 分  | 未按题目要求输出结果，每处扣 1 分<br>输出结果不正确，扣 3 分<br>程序代码运行无法按题目要求显示结果，扣 5 分                            |

表 7-2 任务二评分细则（60 分）

| 序号 | 评分项 | 分值 | 评分细则 |
|----|-----|----|------|
|----|-----|----|------|

|   |            |      |                                                                                                                                                                                                                                                                                                                |
|---|------------|------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1 | 获得网页源代码    | 12 分 | <p>T1-1.jpg 中和参考答案一致满分 12 分；</p> <p>T1 文件夹中有 T1-1.jpg，得 2 分；</p> <p>T1-1.jpg 中定义了请求头，得 4 分；</p> <p>T1-1.jpg 中发送请求，得 2 分；</p> <p>T1-1.jpg 中查看网页响应状态码，得 2 分；</p> <p>T1-1.jpg 中获得网页源代码，得 2 分。</p>                                                                                                                 |
| 2 | 解析网页数据     | 32 分 | <p>T1-2.jpg 中和参考答案一致满分 32 分；</p> <p>T1 文件夹中有 T1-2.jpg，得 2 分；</p> <p>T1-2.jpg 中打印的职位名称正确，得 6 分；</p> <p>T1-2.jpg 中打印的公司名称正确，得 4 分；</p> <p>T1-2.jpg 中打印的企业性质正确，得 4 分；</p> <p>T1-2.jpg 中打印的更新日期正确，得 4 分；</p> <p>T1-2.jpg 中打印的工资范围正确，得 4 分；</p> <p>T1-2.jpg 中打印的所属行业正确，得 4 分；</p> <p>T1-2.jpg 中打印的招聘人数正确，得 4 分。</p> |
| 3 | 数据存储       | 8 分  | “湖南人才网.csv”文件中按要求存储正确的数据，得 8 分。                                                                                                                                                                                                                                                                                |
| 4 | 源代码命名正确并保存 | 4 分  | T1 文件夹中有“湖南人才网.py”文件，得 4 分。                                                                                                                                                                                                                                                                                    |
| 5 | 代码注释       | 4 分  | 代码中有注释，得 4 分。                                                                                                                                                                                                                                                                                                  |

表 7-3 职业素养评分细则（10 分）

| 序号 | 评分项  | 分值  | 评分细则                         |
|----|------|-----|------------------------------|
| 1  | 操作规范 | 3 分 | 答题过程中保持安静，不损坏机器              |
| 2  | 文明考试 | 3 分 | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌； |

|   |      |     |                                   |
|---|------|-----|-----------------------------------|
| 3 | 工位整理 | 4 分 | 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |
|---|------|-----|-----------------------------------|

## 8.试题编号：J3-2-3：链家长沙楼盘数据采集与可视化

### (1) 任务描述

某房地产开发商新开发了楼盘，想要根据长沙其他楼盘的情况进行定价，请帮助该企业完成以下工作：

#### 任务一：实现手机号计数功能关键算法并绘制流程图（30 分）

现获取了某公司楼盘销售人员的手机号码，保存为字符串，各手机号码用空格隔开，合法的手机号为 11 位的数字，统计其中所有合法的手机号码数量。

该字符串是：18711389426 18711389427 13511142256 136p1234567 1375478741  
13912568741 13100023x

具体要求如下：

- 绘制手机号计数程序的流程图，要求流程图流程完整，代表流程的图形正确，能够体现循环结构语句的流程。流程图以图片的形式提交，保存到考生文件夹下，并命名为 test1.png（10 分）
- 正确定义某公司人员的手机号码字符串。（2 分）
- 对手机号码字符串完成按空格切分操作。（2 分）
- 正确判断每个手机号码是否是 11 位的数字。（4 分）
- 统计合法的手机号码个数。（2 分）
- 输出合法的手机号码数量，按题目要求输出正确结果。（5 分）
- 关键代码有写注释，并且表达清晰。（3 分）
- 程序原文件命名为 test2.py，并保存到考生文件夹中 T1 目录下。（2 分）

#### 任务二：数据采集（60 分）

新建“链家.py”，并在代码中实现以下功能，或者使用 scrapy 框架实现：

- 获得网页源代码，截取相应的代码，命名为“T1-1.jpg。访问链家网站

([http://localhost:8080/%E9%93%BE%E5%AE%B6%E9%95%BF%E6%B2%99%E6%A5%BC%E7%9B%98/%E9%95%BF%E6%B2%99%E6%A5%BC%E7%9B%98\\_%E6%96%B0%E5%BC%80%E7%9B%98%E6%A5%BC%E7%9B%98\\_%E6%88%BF%E4%BB%B7%E4%BF%A1%E6%81%AF\\_%E9%95%BF%E6%B2%99%E6%A5%BC%E7%9B%98\(%E9%95%BF%E6%B2%99%E9%93%BE%E5%AE%B6%E6%96%B0%E6%88%BF\).html](http://localhost:8080/%E9%93%BE%E5%AE%B6%E9%95%BF%E6%B2%99%E6%A5%BC%E7%9B%98/%E9%95%BF%E6%B2%99%E6%A5%BC%E7%9B%98_%E6%96%B0%E5%BC%80%E7%9B%98%E6%A5%BC%E7%9B%98_%E6%88%BF%E4%BB%B7%E4%BF%A1%E6%81%AF_%E9%95%BF%E6%B2%99%E6%A5%BC%E7%9B%98(%E9%95%BF%E6%B2%99%E9%93%BE%E5%AE%B6%E6%96%B0%E6%88%BF).html)), 使用 Python 中的 requests 模块向该网站发送请求并获得网页的源代码。要求能够定义请求头、发送请求、查看网页响应状态码、获得网页源代码, 截取相应的代码, 命名为“T1-1.jpg”。(12 分)

- 解析网页数据。选用适当的解析方法 (re 模块、lxml 模块或者 BeautifulSoup 模块) 对网页源代码进行解析, 解析的数据包括: 区域、小区名称、建面、价格、地址。要求解析数据的前 5 条进行打印显示, 并将结果进行截图, 命名为“T1-2.jpg”。(32 分)

样例如下:

| 区域 | 小区名称  | 价格                             | 建筑面积                  | 地址           |
|----|-------|--------------------------------|-----------------------|--------------|
| 雨花 | 中交牛津街 | 30000 元/m <sup>2</sup><br>(均价) | 41-120 m <sup>2</sup> | 万家丽南路与振华路交汇处 |

- 数据存储。采集的数据保存为“链家长沙楼盘.csv”文件。(8 分)
- 代码注释。在代码中有关键代码的注释。(4 分)
- 将“链家长沙楼盘.csv”文件、“链家.py”文件、“J4-2-1.jpg”文件保存到考生文件夹 T1 目录下。(4 分)

#### 良好编程习惯和职业素养 (10 分)

- 答题过程中保持安静。(3 分)
- 有问题举手示意, 有礼貌。(3 分)
- 答题完毕离场时整理好工位, 不关闭计算机, 并按要求显示答案保存界面。(4 分)

## (2) 实施条件

安排在教学实训机房, 计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。

计算机安装有 Python、PyCharm 集成开发环境、Echarts、VSCode，Python 中安装 requests、re、BS4、numpy、pandas、matplotlib、pyecharts 等模块，安装 Office 办公软件，安装绘制流程图的软件-亿图图示。

**(3) 考核时量及要求**

- 考核时长 180 分钟；
- 请在桌面建立以“学号\_题号”命名的成果文件夹，所有提交文件均放在该目录下。

文件夹命名例如：201965110001\_J2\_1；

将各任务源文件夹分别保存为 task1、task2、task3，运行结果分别截图。

**(4) 评分细则**

表 8-1 任务一评分细则（30 分）

| 序号 | 评分项       | 分值   | 评分细则                                                                                                                          |
|----|-----------|------|-------------------------------------------------------------------------------------------------------------------------------|
| 1  | 开发环境的正确使用 | 2 分  | 未按要求提交正确格式的源文件，扣 2 分。                                                                                                         |
| 2  | 流程图设计合理   | 10 分 | 未按要求提交正确命名的流程图截图，扣 2 分；<br>流程图中绘图使用的符号不当，每个错误符号扣 1 分，扣完为止；<br>流程图中需体现循环语句，未体现扣 4 分；<br>流程图未能完全体现题目的处理逻辑，酌情扣 1-5 分。            |
| 3  | 程序设计合理    | 13 分 | 变量名命名不规范或者未注释，每处扣 1 分；<br>关键代码未写注释，每处扣 1 分；<br>未正确对手机号码字符串进行切分操作，扣 2 分<br>未正确判断每个手机号码是否是 11 位数字，扣 4 分<br>未正确统计合法的手机号码个数，扣 2 分 |
| 4  | 功能实现      | 5 分  | 未按题目要求输出结果，扣 2 分<br>输出结果不正确，扣 3 分<br>程序代码运行无法按题目要求显示结果，扣 5 分                                                                  |

表 8-2 任务二评分细则（60 分）

| 序号 | 评分项        | 分值   | 评分细则                                                                                                                                                                                                         |
|----|------------|------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 获得网页源代码    | 12 分 | T1-1. jpg 中和参考答案一致满分 12 分；<br>T1 文件夹中有 T1-1. jpg，得 2 分；<br>T1-1. jpg 中定义了请求头，得 4 分；<br>T1-1. jpg 中发送请求，得 2 分；<br>T1-1. jpg 中查看网页响应状态码，得 2 分；<br>T1-1. jpg 中获得网页源代码，得 2 分。                                    |
| 2  | 解析网页数据     | 32 分 | T1-2. jpg 中和参考答案一致满分 32 分；<br>T1 文件夹中有 T1-2. jpg，得 2 分；<br>T1-2. jpg 中打印的区域正确，得 6 分；<br>T1-2. jpg 中打印的小区名称正确，得 6 分；<br>T1-2. jpg 中打印的价格正确，得 6 分；<br>T1-2. jpg 中打印的建筑面积正确，得 6 分；<br>T1-2. jpg 中打印的地址正确，得 6 分。 |
| 3  | 数据存储       | 8 分  | “链家长沙楼盘.csv”文件中按要求存储正确的数据，得 8 分。                                                                                                                                                                             |
| 4  | 源代码命名正确并保存 | 4 分  | T1 文件夹中有“链家长沙楼盘.py”文件，得 4 分。                                                                                                                                                                                 |
| 5  | 代码注释       | 4 分  | 代码中有注释，得 4 分。                                                                                                                                                                                                |

表 8-3 职业素养评分细则（10 分）

| 序号 | 评分项  | 分值  | 评分细则                              |
|----|------|-----|-----------------------------------|
| 1  | 操作规范 | 3 分 | 答题过程中保持安静，不损坏机器                   |
| 2  | 文明考试 | 3 分 | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌；      |
| 3  | 工位整理 | 4 分 | 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |

## 9.试题编号：J3-2-4：土流网数据采集与可视化

### （1）任务描述

土流网土地资讯频道提供及时、全面的土地市场行情与土地政策信息，包括农业用地、工商用地、住宅用地、土地发布、土地转让、土地出租、土地估价、服务网点、土地数据等信息。某农产品公司计划租赁 400 亩农地，建一个蔬菜厂，想要了解最新的土地数据行情以及政策法规信息，请帮助该公司完成土流网数据的采集、处理和展示的工作，具体任务如下：

#### 任务一：实现土地信息输入的关键算法并绘制流程图（30 分）

土流网的用户若要发布土地的相关资讯，则必须输入该土地的介绍，网站规定土地的介绍不能大于 100 字符，也不能为空，请编写一个程序，模拟土地介绍输入的过程。结果输出格式示例如下：

请输入本块土地的介绍：

您输入的文字为空，请重新输入：长沙县北山镇明山村商业用地及地面建筑物全部。土地证号长国用（2003）字第3039号。商户出让至2043年3月25日。使用面积玖佰零五平方米（折合1.4亩）。产权证号：长房权证北山字第00006864号，砖木结构仓库用途，建筑面积99.53㎡，房屋1栋1层

您输入的长度已超过规定，请重新输入：长沙县北山镇明山村商业用地及地面建筑物全部。土地证号长国用（2003）字第3039号。商户出让至2043年3月25日。

本块土地的介绍是：长沙县北山镇明山村商业用地及地面建筑物全部。土地证号长国用（2003）字第3039号。商户出让至2043年3月25日。

具体要求如下：

- 绘制土地信息输入的流程图，要求流程图流程完整，代表流程的图形正确。流程图以图片的形式提交，保存到考生文件夹下，并命名为 test1.png（10 分）
- 正确输入一个字符串，并保存。（2 分）
- 判断该字符串长度是否不超过 100 个字符（含空格），若超过则要求重新输入。（4 分）
- 判断该字符串是否为空，若是则要求重新输入。（4 分）
- 能够按题目要求的格式正确输出结果。（5 分）
- 关键代码有写注释，并且表达清晰。（3 分）
- 程序原文件命名为 test2.py，并保存到考生文件夹中 T1 目录下。（2 分）

任务二：任务一：数据采集（30 分）

新建“土流网.py”，并在代码中实现以下功能，或者使用 scrapy 框架实现：

- 获得网页源代码，截取相应的代码，命名为“T1-1.jpg”。访问网站（<http://localhost:8080/%E5%9C%9F%E6%B5%81%E7%BD%91/%E5%86%9C%E6%9D%91%E7%94%A8%E5%9C%B0%E6%B5%81%E8%BD%AC%E4%BB%B7%E6%A0%BC%E4%BF%A1%E6%81%AF%20-%20%E4%BA%A4%E6%98%93%E5%B9%B3%E5%8F%B0%20-%20%E5%9C%9F%E6%B5%81%E7%BD%91.html>），使用 Python 中的 requests 模块向该网站发送请求并获得网页的源代码。要求能够定义请求头、发送请求、查看网页响应状态码、获得网页源代码，截取相应的代码，命名为”T1-1.jpg”。（12 分）
- 解析网页数据。选用适当的解析方法（re 模块、lxml 模块或者 BeautifulSoup 模块）

对网页源代码进行解析，解析的数据包括：名称、年限、转让形式、价格、土地类型、用途、面积、更新时间。要求解析数据的前 5 条进行打印显示，并将结果进行截图，命名为“T1-2.jpg”（32 分）

样例如下：

| 名称                         | 年限   | 转让形式 | 价格             | 土地类型      | 用途 | 面积    | 更新时间               |
|----------------------------|------|------|----------------|-----------|----|-------|--------------------|
| 长沙浏阳市<br>300.0 亩其它<br>耕地出租 | 10 年 | 出租   | 1000 元/<br>亩/年 | 耕地 <br>其它 | 耕地 | 300 亩 | 2021/8/20<br>18:35 |

- 数据存储。采集的数据保存为“土流网.csv”文件。（8 分）
- 代码注释。在代码中有关键代码的注释。（4 分）
- 将“土流网.csv”文件、“土流网.py”文件、“T1-1.jpg”、“T1-2.jpg”文件保存到考生文件夹 T1 目录下。（4 分）

#### 良好编程习惯和职业素养（10 分）

- 答题过程中保持安静。（3 分）
- 有问题举手示意，有礼貌。（3 分）
- 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。（4 分）

### （2）实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。计算机安装有 Python、PyCharm 集成开发环境、Echarts、VSCode，Python 中安装 requests、re、BS4、numpy、pandas、matplotlib、pyecharts 等模块，安装 Office 办公软件，安装绘制流程图的软件-亿图图示。

### （3）考核时量及要求

- 考核时长 180 分钟；
- 请在桌面建立以“学号\_题号”命名的成果文件夹，所有提交文件均放在该目录下。  
文件夹命名例如：201965110001\_J2\_1；
- 将各任务源文件夹分别保存为 task1、task2、task3，运行结果分别截图。

#### (4) 评分细则

表 9-1 任务一评分细则（30 分）

| 序号 | 评分项       | 分值   | 评分细则                                                                                                                                                           |
|----|-----------|------|----------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 开发环境的正确使用 | 2 分  | 未按要求提交正确格式的源文件，扣 2 分。                                                                                                                                          |
| 2  | 流程图设计合理   | 10 分 | 未按要求提交正确命名的流程图截图，扣 2 分；<br>流程图中绘图使用的符号不当，每个错误符号扣 1 分，扣完为止；<br>流程图未能完全体现题目的处理逻辑，酌情扣 1-5 分。                                                                      |
| 3  | 程序设计合理    | 13 分 | 变量名命名不规范或者未注释，每处扣 1 分；<br>关键代码未写注释，每处扣 1 分；<br>未判断输入的字符串是否不超过 100 个字符，扣 2 分<br>未实现若字符串超过 100 个字符则重新输入功能，扣 2 分<br>未判断输入的字符串是否为空，扣 2 分<br>未实现若字符串为空则重新输入功能，扣 2 分 |
| 4  | 功能实现      | 5 分  | 未按题目中格式输出结果，扣 2 分<br>程序代码运行无法显示结果，扣 5 分                                                                                                                        |

表 9-2 任务二评分细则（60 分）

| 序号 | 评分项        | 分值   | 评分细则                                                                                                                                                                                                                                                                                                  |
|----|------------|------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 获得网页源代码    | 12 分 | T1-1. jpg 中和参考答案一致满分 12 分；<br>T1 文件夹中有 T1-1. jpg，得 2 分；<br>T1-1. jpg 中定义了请求头，得 4 分；<br>T1-1. jpg 中发送请求，得 2 分；<br>T1-1. jpg 中查看网页响应状态码，得 2 分；<br>T1-1. jpg 中获得网页源代码，得 2 分。                                                                                                                             |
| 2  | 解析网页数据     | 32 分 | T1-2. jpg 中和参考答案一致满分 32 分；<br>T1 文件夹中有 T1-2. jpg，得 2 分；<br>T1-2. jpg 中打印的名称正确，得 4 分；<br>T1-2. jpg 中打印的年限正确，得 4 分；<br>T1-2. jpg 中打印的转让形式正确，得 4 分；<br>T1-2. jpg 中打印的价格正确，得 4 分；<br>T1-2. jpg 中打印的土地类型正确，得 4 分；<br>T1-2. jpg 中打印的用途正确，得 4 分；<br>T1-2. jpg 中打印的面积正确，得 4 分；<br>T1-2. jpg 中打印的更新时间正确，得 2 分。 |
| 3  | 数据存储       | 8 分  | “土流网.csv”文件中按要求存储正确的数据，得 8 分。                                                                                                                                                                                                                                                                         |
| 4  | 源代码命名正确并保存 | 4 分  | T1 文件夹中有“土流网.py”文件，得 4 分。                                                                                                                                                                                                                                                                             |
| 5  | 代码注释       | 4 分  | 代码中有注释，得 4 分。                                                                                                                                                                                                                                                                                         |

表 9-3 职业素养评分细则（10 分）

| 序号 | 评分项  | 分值  | 评分细则                              |
|----|------|-----|-----------------------------------|
| 1  | 操作规范 | 3 分 | 答题过程中保持安静，不损坏机器                   |
| 2  | 文明考试 | 3 分 | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌；      |
| 3  | 工位整理 | 4 分 | 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |

### 项目 3：资讯服务行业数据采集与可视化

#### 10.试题编号：J3-3-1：中国大学 Top100 数据采集与可视化

##### （1）任务描述

中国大学网是学生高考择校、出国留学、社会各界了解分析国内外高校的有效渠道。高考在即，请帮助考生们快速了解国内高校，具体任务如下：

##### 任务一：实现大学排名功能关键算法并绘制流程图（30 分）

某位考生针对自己的情况，从中国大学网站上取得了部分大学的数据，要求对这些数据中的大学按评分进行排名，具体数据如下表所示：

| 序号 | 大学名称     | 评分    |
|----|----------|-------|
| 1  | 中国科学技术大学 | 526.4 |
| 2  | 上海交通大学   | 625.9 |
| 3  | 复旦大学     | 556.7 |
| 4  | 清华大学     | 852.5 |
| 5  | 北京大学     | 746.7 |
| 6  | 南京大学     | 566.1 |
| 7  | 浙江大学     | 649.2 |

结果输出格式示例如下：

| 排名 | 大学名称     | 评分    |
|----|----------|-------|
| 1  | 清华大学     | 852.5 |
| 2  | 北京大学     | 746.7 |
| 3  | 浙江大学     | 649.2 |
| 4  | 上海交通大学   | 625.9 |
| 5  | 南京大学     | 566.1 |
| 6  | 复旦大学     | 556.7 |
| 7  | 中国科学技术大学 | 526.4 |

具体要求如下：

- 绘制大学排名程序的流程图，要求流程图流程完整，代表流程的图形正确，能够体现循环结构语句的流程。流程图以图片的形式提交，保存到考生文件夹下，并命名为 test1.png（10 分）
- 正确保存初始数据为字典类型。（5 分）
- 根据题目要求，按大学的评分进行排名，要求使用 sorted 方法实现。（7 分）
- 能够按题目要求的格式正确输出结果。（3 分）
- 关键代码有写注释，并且表达清晰。（3 分）
- 程序原文件命名为 test2.py，并保存到考生文件夹中 T1 目录下。（2 分）

## 任务二：数据采集（60 分）

新建“中国大学.py”，并在代码中实现以下功能，或者使用 scrapy 框架实现：

- 获得网页源代码，截取相应的代码，命名为“T1-1.jpg”。访问网站  
(<http://localhost:8080/%E4%B8%AD%E5%9B%BD%E5%A4%A7%E5%AD%A6%E6%8E%92%E5%90%8D/bcur.json>)，使用 Python 中的 requests 模块向该网站发送请求并获得网页的源代码。要求能够定义请求头、发送请求、查看网页响应状态码、获得网页源代码，截取相应的代码，命名为“T1-1.jpg”。（12 分）
- 解析网页数据。选用适当的解析方法（re 模块、lxml 模块或者 BeautifulSoup 模块）对网页源代码进行解析，解析的数据包括：大学排名，大学名称，所在省市。要求解析数据的前 5 条进行打印显示，并将结果进行截图，命名为“T1-2.jpg”。（32 分）

样例如下：

| 排名 | 学校名称 | 省份 | 总分 |
|----|------|----|----|
|----|------|----|----|

|   |      |     |      |
|---|------|-----|------|
| 1 | 清华大学 | 北京市 | 95.9 |
|---|------|-----|------|

- 数据存储。采集的数据保存为“中国大学.csv”文件。（8分）
- 代码注释。在代码中有关键代码的注释。（4分）
- 将“中国大学.csv”文件、“中国大学.py”文件、“T1-1.jpg”、“T1-2.jpg”文件保存到考生文件夹 T1 目录下。（4分）

#### 良好编程习惯和职业素养（10分）

- 答题过程中保持安静。（3分）
- 有问题举手示意，有礼貌。（3分）
- 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。（4分）

### （2）实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。计算机安装有 Python、PyCharm 集成开发环境、Echarts、VSCode，Python 中安装 requests、re、BS4、numpy、pandas、matplotlib、pyecharts 等模块，安装 Office 办公软件，安装绘制流程图的软件-亿图图示。

### （3）考核时量及要求

- 考核时长 180 分钟；
- 请在桌面建立以“学号\_题号”命名的成果文件夹，所有提交文件均放在该目录下。  
文件夹命名例如：201965110001\_J2\_1；
- 将各任务源文件夹分别保存为 task1、task2、task3，运行结果分别截图。

### （4）评分细则

表 10-1 任务一评分细则（30 分）

| 序号 | 评分项       | 分值   | 评分细则                                                                                      |
|----|-----------|------|-------------------------------------------------------------------------------------------|
| 1  | 开发环境的正确使用 | 2 分  | 未按要求提交正确格式的源文件，扣 2 分。                                                                     |
| 2  | 流程图设计合理   | 10 分 | 未按要求提交正确命名的流程图截图，扣 2 分；<br>流程图中绘图使用的符号不当，每个错误符号扣 1 分，扣完为止；<br>流程图未能完全体现题目的处理逻辑，酌情扣 1-5 分。 |
| 3  | 程序设计合理    | 13 分 | 变量名命名不规范或者未注释，每处扣 1 分；<br>关键代码未写注释，每处扣 1 分；                                               |

|   |      |     |                                                              |
|---|------|-----|--------------------------------------------------------------|
|   |      |     | 初始数据未保存为字典类型，扣 5 分<br>未按要求进行排序，酌情扣 1-7 分                     |
| 4 | 功能实现 | 5 分 | 未按题目要求输出结果，扣 2 分<br>输出结果不正确，扣 3 分<br>程序代码运行无法按题目要求显示结果，扣 5 分 |

表 10-2 任务二评分细则（60 分）

| 序号 | 评分项        | 分值   | 评分细则                                                                                                                                                                          |
|----|------------|------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 获得网页源代码    | 12 分 | T1-1. jpg 中和参考答案一致满分 12 分；<br>T1 文件夹中有 T1-1. jpg，得 2 分；<br>T1-1. jpg 中定义了请求头，得 4 分；<br>T1-1. jpg 中发送请求，得 2 分；<br>T1-1. jpg 中查看网页响应状态码，得 2 分；<br>T1-1. jpg 中获得网页源代码，得 2 分。     |
| 2  | 解析网页数据     | 32 分 | T1-2. jpg 中和参考答案一致满分 32 分；<br>T1 文件夹中有 T1-2. jpg，得 2 分；<br>T1-2. jpg 中打印的排名正确，得 8 分；<br>T1-2. jpg 中打印的学校名称正确，得 8 分；<br>T1-2. jpg 中打印的省份正确，得 8 分；<br>T1-2. jpg 中打印的总分正确，得 6 分。 |
| 3  | 数据存储       | 8 分  | “中国大学.csv”文件中按要求存储正确的数据，得 8 分。                                                                                                                                                |
| 4  | 源代码命名正确并保存 | 4 分  | T1 文件夹中有“中国大学.py”文件，得 4 分。                                                                                                                                                    |
| 5  | 代码注释       | 4 分  | 代码中有注释，得 4 分。                                                                                                                                                                 |

表 4 职业素养评分细则（10 分）

| 序号 | 评分项  | 分值  | 评分细则                              |
|----|------|-----|-----------------------------------|
| 1  | 操作规范 | 3 分 | 答题过程中保持安静，不损坏机器                   |
| 2  | 文明考试 | 3 分 | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌；      |
| 3  | 工位整理 | 4 分 | 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |

## 11.试题编号：J3-3-2：北京 2020 年天气数据采集与可视化

### （1）任务描述

天气预报（测）或气象预报（测）是使用现代科学技术对未来某一地点地球大气层的状态进行预测，主要是使用收集大量的数据（气温、湿度、风向和风速、气压等等），然后使用目前对大气过程的认识（气象学）来确定未来空气变化，也是天气预报是气象工作为国民经济和国防建设服务的重要手段。现需对 2020 年北京市的天气数据进行分析，具体任务如下：

#### 任务一：实现天气数据核对功能关键算法并绘制流程图（30 分）

现有北京市 2020 年 10 月的部分天气数据，在进行数据分析过程中发现数据有误，请按要求对数据进行核对和修改。有误的数据如下：

```
list1=["日期","天气状况","气温","风力风向"]
```

```
list2=["2020 年 10 月 1 日","25℃/12℃","西北风 4-5 级"]
```

```
list3=["2020 年 10 月 2 日","晴/阴","阴天","24℃/10℃","西北风 2-3 级"]
```

```
list4=["2020 年 10 月 2 日","阴/晴","25℃/12℃","西南风 4-5 级"]
```

不允许在程序中直接改变 4 个列表创建时的数据，需要通过列表的增、删、改、查等操作对数据进行修正，并按要求计算每天的温差。正确数据如下表所示：

| 日期              | 天气状况 | 气温      | 风力风向      |
|-----------------|------|---------|-----------|
| 2020 年 10 月 1 日 | 晴/晴  | 25℃/12℃ | 西北风 4-5 级 |
| 2020 年 10 月 2 日 | 晴/阴  | 24℃/10℃ | 西北风 2-3 级 |
| 2020 年 10 月 3 日 | 阴/晴  | 25℃/12℃ | 西南风 4-5 级 |

结果输出格式示例：

```
-----完善后的数据如下-----
['日期', '天气状况', '气温', '风力风向']
['2020年10月1日', '晴/晴', '25℃/12℃', '西北风4-5级']
['2020年10月2日', '晴/阴', '24℃/10℃', '西北风2-3级']
['2020年10月3日', '阴/晴', '25℃/12℃', '西南风4-5级']
2020年10月1日北京市的温差是：13℃
2020年10月2日北京市的温差是：14℃
2020年10月3日北京市的温差是：13℃
```

具体要求如下：

- 绘制天气数据核对程序的流程图，要求流程图流程完整，代表流程的图形正确。流程图以图片的形式提交，保存到考生文件夹下，并命名为 test1.png（10 分）
- 定义原始数据列表。（4 分）
- 按正确数据对程序中的数据进行修改。（3 分）
- 计算每天的温差。（3 分）
- 能够按题目要求的格式正确输出结果。（5 分）
- 关键代码有写注释，并且表达清晰。（3 分）
- 程序原文件命名为 test2.py，并保存到考生文件夹中 T1 目录下。（2 分）

## 任务二：数据采集（60 分）

新建“历史天气.py”，并在代码中实现以下功能，或者使用 scrapy 框架实现：

- 获得网页源代码，截取相应的代码，命名为“T1-1.jpg”。访问网站  
([http://localhost:8080/%E5%8E%86%E5%8F%B2%E5%A4%A9%E6%B0%94/%E5%8C%97%E4%BA%AC%E5%8E%86%E5%8F%B2%E5%A4%A9%E6%B0%94%E9%A2%84%E6%8A%A5%E6%9F%A5%E8%AF%A2\\_2020%E5%B9%B410%E6%9C%88%E4%BB%BD%E5%8C%97%E4%BA%AC%E5%A4%A9%E6%B0%94%E8%AE%B0%E5%BD%95\\_%E5%8C%97%E4%BA%AC2020%E5%B9%B410%E6%9C%88%E4%BB%BD%E5%A4%A9%E6%B0%94%E6%83%85%E5%86%B5\\_%E5%A4%A9%E6%B0%94%E5%90%8E%E6%8A%A5.html](http://localhost:8080/%E5%8E%86%E5%8F%B2%E5%A4%A9%E6%B0%94/%E5%8C%97%E4%BA%AC%E5%8E%86%E5%8F%B2%E5%A4%A9%E6%B0%94%E9%A2%84%E6%8A%A5%E6%9F%A5%E8%AF%A2_2020%E5%B9%B410%E6%9C%88%E4%BB%BD%E5%8C%97%E4%BA%AC%E5%A4%A9%E6%B0%94%E8%AE%B0%E5%BD%95_%E5%8C%97%E4%BA%AC2020%E5%B9%B410%E6%9C%88%E4%BB%BD%E5%A4%A9%E6%B0%94%E6%83%85%E5%86%B5_%E5%A4%A9%E6%B0%94%E5%90%8E%E6%8A%A5.html))，使用 Python 中的 requests 模块向该网站发送请求并获得网页的源代码。要求能够定义请求头、发送请求、查看网页响应状态码、获得网页源代码，截取相应的代码，命名为“T1-1.jpg”。（12 分）
- 解析网页数据。选用适当的解析方法（re 模块、lxml 模块或者 BeautifulSoup 模块）对网页源代码进行解析，解析的数据包括：日期、天气状况、气温、风力风向。要求解析数据的前 5 条进行打印显示，并将结果进行截图，命名为“T1-2.jpg”。（32 分）

样例如下：

|    |      |    |      |
|----|------|----|------|
| 日期 | 天气状况 | 气温 | 风力风向 |
|----|------|----|------|

|                 |     |         |           |
|-----------------|-----|---------|-----------|
| 2020 年 10 月 1 日 | 晴/晴 | 25℃/12℃ | 西北风 4-5 级 |
|-----------------|-----|---------|-----------|

- 数据存储。采集的数据保存为“历史天气.csv”文件。（8 分）
- 代码注释。在代码中有关键代码的注释。（4 分）
- 将“历史天气.csv”文件、“历史天气.py”文件、“T1-1.jpg”、“T1-2.jpg”文件保存到考生文件夹 T1 目录下。（4 分）

### 良好编程习惯和职业素养（10 分）

- 答题过程中保持安静。（3 分）
- 有问题举手示意，有礼貌。（3 分）
- 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。（4 分）

## （2）实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。计算机安装有 Python、PyCharm 集成开发环境、Echarts、VSCode，Python 中安装 requests、re、BS4、numpy、pandas、matplotlib、pyecharts 等模块，安装 Office 办公软件，安装绘制流程图的软件-亿图图示。

## （3）考核时量及要求

- 考核时长 180 分钟；
- 请在桌面建立以“学号\_题号”命名的成果文件夹，所有提交文件均放在该目录下。  
文件夹命名例如：201965110001\_J2\_1；
- 将各任务源文件夹分别保存为 task1、task2、task3，运行结果分别截图。

## （4）评分细则

表 11-1 任务一评分细则（30 分）

| 序号 | 评分项       | 分值   | 评分细则                                                                                      |
|----|-----------|------|-------------------------------------------------------------------------------------------|
| 1  | 开发环境的正确使用 | 2 分  | 未按要求提交正确格式的源文件，扣 2 分。                                                                     |
| 2  | 流程图设计合理   | 10 分 | 未按要求提交正确命名的流程图截图，扣 2 分；<br>流程图中绘图使用的符号不当，每个错误符号扣 1 分，扣完为止；<br>流程图未能完全体现题目的处理逻辑，酌情扣 1-5 分。 |
| 3  | 程序设计合理    | 13 分 | 变量名命名不规范或者未注释，每处扣 1 分；<br>关键代码未写注释，每处扣 1 分；                                               |

|   |      |     |                                                              |
|---|------|-----|--------------------------------------------------------------|
|   |      |     | 未全部对数据进行修改，漏一处扣 1 分<br>未正确计算每天的温差，每天扣 1 分                    |
| 4 | 功能实现 | 5 分 | 未按题目要求输出结果，扣 2 分<br>输出结果不正确，扣 3 分<br>程序代码运行无法按题目要求显示结果，扣 5 分 |

表 11-2 任务二评分细则（60 分）

| 序号 | 评分项        | 分值   | 评分细则                                                                                                                                                                            |
|----|------------|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 获得网页源代码    | 12 分 | T1-1. jpg 中和参考答案一致满分 12 分；<br>T1 文件夹中有 T1-1. jpg，得 2 分；<br>T1-1. jpg 中定义了请求头，得 4 分；<br>T1-1. jpg 中发送请求，得 2 分；<br>T1-1. jpg 中查看网页响应状态码，得 2 分；<br>T1-1. jpg 中获得网页源代码，得 2 分。       |
| 2  | 解析网页数据     | 32 分 | T1-2. jpg 中和参考答案一致满分 32 分；<br>T1 文件夹中有 T1-2. jpg，得 2 分；<br>T1-2. jpg 中打印的日期正确，得 8 分；<br>T1-2. jpg 中打印的天气情况正确，得 8 分；<br>T1-2. jpg 中打印的气温正确，得 8 分；<br>T1-2. jpg 中打印的风力风向正确，得 6 分。 |
| 3  | 数据存储       | 8 分  | “历史天气.csv”文件中按要求存储正确的数据，得 8 分。                                                                                                                                                  |
| 4  | 源代码命名正确并保存 | 4 分  | T1 文件夹中有“历史天气.py”文件，得 4 分。                                                                                                                                                      |
| 5  | 代码注释       | 4 分  | 代码中有注释，得 4 分。                                                                                                                                                                   |

表 11-3 职业素养评分细则（10 分）

| 序号 | 评分项  | 分值  | 评分细则                              |
|----|------|-----|-----------------------------------|
| 1  | 操作规范 | 3 分 | 答题过程中保持安静，不损坏机器                   |
| 2  | 文明考试 | 3 分 | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌；      |
| 3  | 工位整理 | 4 分 | 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |

## 12. 试题编号：J3-3-3：糗事百科数据采集与可视化

### （1）任务描述

糗事百科是以糗友真实糗事为主题的笑话网站，话题轻松休闲，在年轻人中十分流行。

请以此网站为目标，完成以下任务：

**任务一：实现小孩报酬的关键算法并绘制流程图（30 分）**

数据采集工程师小张在糗事百科上发现一个小段子，内容是：一个小孩给一个富翁做事，商量的报酬是第一天 2 角，第二天是第一天的 2 倍，第 30 天富翁净身出户，所有的家产全部归小孩所有。这个小孩具体拿了多少钱呢？结果输出格式示例如下：

第30天小孩拿到的全部报酬是：214748364.6元

具体要求如下：

- 绘制小孩报酬程序的流程图，要求流程图流程完整，代表流程的图形正确，能够体现循环结构语句的流程。流程图以图片的形式提交，保存到考生文件夹下，并命名为 test1.png（10 分）
- 正确设定一个函数实现递归算法，带一个参数的传递。（5 分）
- 正确设定递归的终止条件和调用函数自身的表达式。（5 分）
- 输出小孩最终的报酬，能够按题目要求正确输出结果。（5 分）
- 关键代码有写注释，并且表达清晰。（3 分）
- 程序原文件命名为 test2.py，并保存到考生文件夹中 T1 目录下。（2 分）

**任务二：数据采集（60 分）**

新建“糗事百科.py”，并在代码中实现以下功能，或者使用 scrapy 框架实现：

- 获得网页源代码，截取相应的代码，命名为“T1-1.jpg”。访问网站  
([http://localhost:8080/%E7%B3%97%E4%BA%8B%E7%99%BE%E7%A7%91/%E5%B9%BD%E9%BB%98%E7%AC%91%E8%AF%9D%E5%A4%A7%E5%85%A8\\_%E7%88%86%E7%AC%91%E7%AC%91%E8%AF%9D\\_%E7%AC%91%E7%A0%B4%E4%BD%A0%E7%9A%84%E8%82%9A%E5%AD%90%E7%9A%84%E6%90%9E%E7%AC%91%E6%AE%B5%E5%AD%90%20-%20%E7%B3%97%E4%BA%8B%E7%99%BE%E7%A7%91.html](http://localhost:8080/%E7%B3%97%E4%BA%8B%E7%99%BE%E7%A7%91/%E5%B9%BD%E9%BB%98%E7%AC%91%E8%AF%9D%E5%A4%A7%E5%85%A8_%E7%88%86%E7%AC%91%E7%AC%91%E8%AF%9D_%E7%AC%91%E7%A0%B4%E4%BD%A0%E7%9A%84%E8%82%9A%E5%AD%90%E7%9A%84%E6%90%9E%E7%AC%91%E6%AE%B5%E5%AD%90%20-%20%E7%B3%97%E4%BA%8B%E7%99%BE%E7%A7%91.html))，使用 Python 中的 requests 模块向该网站发送请求并获得网页的源代码。要求能够定义请求头、发送请求、查看网页响应状态码、获

得网页源代码，截取相应的代码，命名为“T1-1.jpg”。（12 分）

- 解析网页数据。选用适当的解析方法（re 模块、lxml 模块或者 BeautifulSoup 模块）对网页源代码进行解析，解析的数据包括：用户名、性别（粉色为女、蓝色为男）、年龄、点赞数（好笑）、段子内容、评论数据。要求解析数据的前 5 条进行打印显示，并将结果进行截图，命名为“T1-2.jpg”。（32 分）

样例如下：

| 用户名      | 性别 | 年龄 | 点赞数  | 段子内容                                            | 评论数 |
|----------|----|----|------|-------------------------------------------------|-----|
| 糗糗 queen | 男  | 26 | 1293 | 你们一般都是跟朋友，同学，同事，陌生人撞衫吧，我就比较厉害了，昨天在街上跟一宠物狗撞衫了... | 36  |

- 数据存储。采集的数据保存为“糗事百科.csv”文件。（8 分）
- 代码注释。在代码中有关键代码的注释。（4 分）
- 将“糗事百科.csv”文件、“糗事百科.py”文件、“T1-1.jpg”、“T1-2.jpg”文件保存到考生文件夹 T1 目录下。（4 分）

### 良好编程习惯和职业素养（10 分）

- 答题过程中保持安静。（3 分）
- 有问题举手示意，有礼貌。（3 分）
- 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。（4 分）

## （2）实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。计算机安装有 Python、PyCharm 集成开发环境、Echarts、VSCode，Python 中安装 requests、re、BS4、numpy、pandas、matplotlib、pyecharts 等模块，安装 Office 办公软件，安装绘制流程图的软件-亿图图示。

## （3）考核时量及要求

- 考核时长 180 分钟；

- 请在桌面建立以“学号\_题号”命名的成果文件夹，所有提交文件均放在该目录下。

文件夹命名例如：201965110001\_J2\_1；

- 将各任务源文件夹分别保存为 task1、task2、task3，运行结果分别截图。

#### (4) 评分细则

表 12-1 任务一评分细则（30 分）

| 序号 | 评分项       | 分值   | 评分细则                                                                                                                |
|----|-----------|------|---------------------------------------------------------------------------------------------------------------------|
| 1  | 开发环境的正确使用 | 2 分  | 未按要求提交正确格式的源文件，扣 2 分。                                                                                               |
| 2  | 流程图设计合理   | 10 分 | 未按要求提交正确命名的流程图截图，扣 2 分；<br>流程图中绘图使用的符号不当，每个错误符号扣 1 分，扣完为止；<br>流程图中需体现函数的调用，未体现扣 4 分；<br>流程图未能完全体现题目的处理逻辑，酌情扣 1-5 分。 |
| 3  | 程序设计合理    | 13 分 | 变量名命名不规范或者未注释，每处扣 1 分；<br>关键代码未写注释，每处扣 1 分；<br>未设定函数实现了题目功能，扣 3 分<br>设定了函数但函数未设定参数，扣 1 分<br>未正确设定递归，扣 5 分           |
| 4  | 功能实现      | 5 分  | 未按题目要求输出结果，扣 2 分<br>输出结果不正确，扣 3 分<br>程序代码运行无法按题目要求显示结果，扣 5 分                                                        |

表 12-2 任务二评分细则（60 分）

| 序号 | 评分项        | 分值   | 评分细则                                                                                                                                                                                                                                       |
|----|------------|------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 获得网页源代码    | 12 分 | T1-1. jpg 中和参考答案一致满分 12 分；<br>T1 文件夹中有 T1-1. jpg，得 2 分；<br>T1-1. jpg 中定义了请求头，得 4 分；<br>T1-1. jpg 中发送请求，得 2 分；<br>T1-1. jpg 中查看网页响应状态码，得 2 分；<br>T1-1. jpg 中获得网页源代码，得 2 分。                                                                  |
| 2  | 解析网页数据     | 32 分 | T1-2. jpg 中和参考答案一致满分 32 分；<br>T1 文件夹中有 T1-2. jpg，得 2 分；<br>T1-2. jpg 中打印的用户名正确，得 4 分；<br>T1-2. jpg 中打印的性别正确，得 4 分；<br>T1-2. jpg 中打印的年龄正确，得 4 分；<br>T1-2. jpg 中打印的点赞数正确，得 4 分；<br>T1-2. jpg 中打印的段子内容正确，得 6 分；<br>T1-2. jpg 中打印的评论数正确，得 4 分。 |
| 3  | 数据存储       | 8 分  | “糗事百科.csv”文件中按要求存储正确的数据，得 8 分。                                                                                                                                                                                                             |
| 4  | 源代码命名正确并保存 | 4 分  | T1 文件夹中有“糗事百科.py”文件，得 4 分。                                                                                                                                                                                                                 |
| 5  | 代码注释       | 4 分  | 代码中有注释，得 4 分。                                                                                                                                                                                                                              |

表 12-3 职业素养评分细则（10 分）

| 序号 | 评分项  | 分值  | 评分细则                              |
|----|------|-----|-----------------------------------|
| 1  | 操作规范 | 3 分 | 答题过程中保持安静，不损坏机器                   |
| 2  | 文明考试 | 3 分 | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌；      |
| 3  | 工位整理 | 4 分 | 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |

### 13.试题编号：J3-3-4：胡润富豪榜数据采集与可视化

#### （1）任务描述

胡润富豪榜一般指胡润百富榜。胡润百富是追踪记录中国企业家群体变化的权威机构，是由出生于欧洲卢森堡的英国注册会计师胡润先生于 1999 年创立的。是中国推出第一份财富排行榜，也是现在国内财经榜单里影响最大的一个榜单。现需要了解最新的富豪排行，具体任务如下：

#### 任务一：企业家的年龄的关键算法并绘制流程图（30 分）

某位企业家智力早熟，11 岁就上了大学，很早就接手家族企业。一次，他参加某个重

要会议，年轻的脸孔引人注目。于是有人询问他的年龄，他回答说：“我年龄的立方是个 4 位数。我年龄的 4 次方是个 6 位数。这 10 个数字正好包含了从 0 到 9 这 10 个数字，每个都恰好出现 1 次。”请你编程计算，他当时到底有多大年龄。

具体要求如下：

- 绘制企业家年龄程序的流程图，要求流程图流程完整，代表流程的图形正确，能够体现循环结构语句的流程。流程图以图片的形式提交，保存到考生文件夹下，并命名为 test1.png（10 分）
- 使用循环语句，并正确设定循环。（4 分）
- 正确设定判断正确年龄的条件。（6 分）
- 输出企业家的正确年龄，能够按题目要求的输出结果。（5 分）
- 关键代码有写注释，并且表达清晰。（3 分）
- 程序原文件命名为 test2.py，并保存到考生文件夹中 T1 目录下。（2 分）

**任务二：数据采集（60 分）**

新建“胡润富豪榜.py”，并在代码中实现以下功能，或者使用 scrapy 框架实现：

- 获得网页源代码，截取相应的代码，命名为“T1-1.jpg”。访问网站  
(http://localhost:8080/%E8%83%A1%E6%B6%A6%E5%AF%8C%E8%B1%AA%E6%A6%9C/HsRankDetailsList.json)，使用 Python 中的 requests 模块向该网站发送请求并获得网页的源代码。要求能够定义请求头、发送请求、查看网页响应状态码、获得网页源代码，截取相应的代码，命名为“T1-1.jpg”。（12 分）
- 解析网页数据。选用适当的解析方法（re 模块、lxml 模块或者 BeautifulSoup 模块、json 模块）对网页源代码进行解析，解析的数据包括：姓名、性别、年龄、财富、关联企业、行业类别。要求解析数据的前 5 条进行打印显示，并将结果进行截图，命名为“T1-2.jpg”。（32 分）

样例如下：

| 姓名     | 性别 | 年龄 | 财富      | 关联企业 | 行业类别  |
|--------|----|----|---------|------|-------|
| 埃隆·马斯克 | 先生 | 50 | 12800 亿 | 特斯拉  | 新能源汽车 |

- 数据存储。采集的数据的前 100 条保存为“胡润富豪榜.csv”文件。（8 分）

- 代码注释。在代码中有关键代码的注释。（4分）
- 将“胡润富豪榜.csv”文件、“胡润富豪榜.py”文件、“T1-1.jpg”、“T1-2.jpg”文件保存到考生文件夹 T1 目录下。（4分）

### 良好编程习惯和职业素养（10分）

- 答题过程中保持安静。（3分）
- 有问题举手示意，有礼貌。（3分）
- 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。（4分）

## （2）实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。计算机安装有 Python、PyCharm 集成开发环境、Echarts、VSCode，Python 中安装 requests、re、BS4、numpy、pandas、matplotlib、pyecharts 等模块，安装 Office 办公软件，安装绘制流程图的软件-亿图图示。

## （3）考核时量及要求

- 考核时长 180 分钟；
- 请在桌面建立以“学号\_题号”命名的成果文件夹，所有提交文件均放在该目录下。  
文件夹命名例如：201965110001\_J2\_1；
- 将各任务源文件夹分别保存为 task1、task2、task3，运行结果分别截图。

## （4）评分细则

表 13-1 任务一评分细则（30 分）

| 序号 | 评分项       | 分值   | 评分细则                                                                                                               |
|----|-----------|------|--------------------------------------------------------------------------------------------------------------------|
| 1  | 开发环境的正确使用 | 2 分  | 未按要求提交正确格式的源文件，扣 2 分。                                                                                              |
| 2  | 流程图设计合理   | 10 分 | 未按要求提交正确命名的流程图截图，扣 2 分；<br>流程图中绘图使用的符号不当，每个错误符号扣 1 分，扣完为止；<br>流程图中需体现循环语句，未体现扣 4 分；<br>流程图未能完全体现题目的处理逻辑，酌情扣 1-5 分。 |
| 3  | 程序设计合理    | 13 分 | 变量名命名不规范或者未注释，每处扣 1 分；<br>关键代码未写注释，每处扣 1 分；<br>未正确设定企业家年龄判断的条件，扣 6 分                                               |
| 4  | 功能实现      | 5 分  | 未按题目要求输出结果，扣 2 分<br>输出结果不正确，扣 3 分                                                                                  |

|  |  |  |                         |
|--|--|--|-------------------------|
|  |  |  | 程序代码运行无法按题目要求显示结果，扣 5 分 |
|--|--|--|-------------------------|

表 13-2 任务二评分细则（60 分）

| 序号 | 评分项        | 分值   | 评分细则                                                                                                                                                                                                                              |
|----|------------|------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 获得网页源代码    | 12 分 | T1-1.jpg 中和参考答案一致满分 12 分；<br>T1 文件夹中有 T1-1.jpg，得 2 分；<br>T1-1.jpg 中定义了请求头，得 4 分；<br>T1-1.jpg 中发送请求，得 2 分；<br>T1-1.jpg 中查看网页响应状态码，得 2 分；<br>T1-1.jpg 中获得网页源代码，得 2 分。                                                               |
| 2  | 解析网页数据     | 32 分 | T1-2.jpg 中和参考答案一致满分 32 分；<br>T1 文件夹中有 T1-2.jpg，得 2 分；<br>T1-2.jpg 中打印的姓名正确，得 4 分；<br>T1-2.jpg 中打印的性别正确，得 4 分；<br>T1-2.jpg 中打印的年龄正确，得 4 分；<br>T1-2.jpg 中打印的财富正确，得 4 分；<br>T1-2.jpg 中打印的关联企业正确，得 6 分；<br>T1-2.jpg 中打印的行业类别正确，得 4 分。 |
| 3  | 数据存储       | 8 分  | “胡润富豪榜.csv”文件中按要求存储正确的数据，得 8 分。                                                                                                                                                                                                   |
| 4  | 源代码命名正确并保存 | 4 分  | T1 文件夹中有“胡润富豪榜.py”文件，得 4 分。                                                                                                                                                                                                       |
| 5  | 代码注释       | 4 分  | 代码中有注释，得 4 分。                                                                                                                                                                                                                     |

表 13-3 职业素养评分细则（10 分）

| 序号 | 评分项  | 分值  | 评分细则                              |
|----|------|-----|-----------------------------------|
| 1  | 操作规范 | 3 分 | 答题过程中保持安静，不损坏机器                   |
| 2  | 文明考试 | 3 分 | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌；      |
| 3  | 工位整理 | 4 分 | 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |

---

## 模块四 数据处理与可视化

### 项目 1：数字内容服务行业数据处理与可视化

#### 1. 试题编号：J4-1-1：时光网电影票房数据处理与可视化

##### (1) 任务描述

电影业中，票房已经成为衡量一部电影是否成功的重要指标之一，可以用观众人数或门票收入来计算，一般直接用来衡量一部电影的热播程度。“电影票房”网是一个提供电影票房数据的网站。某企业想要了解电影票房等相关信息，因此需要进行数据处理及可视化的相关工作，具体任务如下：

##### 任务一：数据处理和分析（60 分）

使用 Python 读取“T1\_movies.csv”文件，完成如下数据处理和分析操作：

- 删除电影名这列数据中的时间信息，如“让子弹飞（2010）”转换成“让子弹飞”，将“累计票房”这列数据只保留票房信息，如“累计票房 65355.3 万”转换成“65355.3”，完成之后打印出前五行数据，将代码和结果截图，命名为 T1-1.jpg，并保存到考生文件夹的 T1 目录。（16 分）
- 将导演这列数据只保留第一个导演的中文名称，如“徐静蕾 Jinglei Xu 安韩瑾 Hanjin An”转换成“徐静蕾”，将类型中的数据进行拆分，有多个类型则只保留第一个。例如，“剧情/战争/历史”改为“剧情”，并输出数据的前五行，将代码和结果截图，命名为 T1-2.jpg，并保存到考生文件夹的 T1 目录。（16 分）
- 上映时间这列数据只保留年份。例如“2020-08-21”改为“2020”，并输出数据的前五行，将代码和结果截图，命名为 T1-3.jpg，并保存到考生文件夹的 T1 目录。（12 分）
- 将处理好的数据保存为“movies\_清洗后.csv”，并保存到考生文件夹的 T1 目录。（8 分）
- 将源代码文件命名为 T1.py，保存到考生文件夹的 T1 目录。（4 分）
- 代码注释。在代码中有关键代码的注释。（4 分）

---

## 任务二：数据可视化（30 分）

根据题目需求读取相关数据，使用 echarts 或 pyecharts 完成如下题目：

- 使用折线图展示每年上映的电影数量变化趋势，标题为“每年上映电影数量”，标题颜色为红色，标题字体大小为 20 像素，居中显示。要求将运行结果截图保存为 T2-1.jpg，源文件命名为 T2-1.html，保存到考生文件夹的 T2 目录。（10 分）
- 使用环形图展示各类型电影数量占比情况，要求内圆半径设为“50%”，外圆半径设为“75%”，标题为“各类型电影数量占比”。要求将运行结果截图保存为 T2-2.jpg，源文件命名为 T2-2.html，保存到考生文件夹的 T2 目录。（10 分）。
- 使用柱状图展示票房最高的前 5 项电影，标题为“票房 TOP5 电影名称”。要求将运行结果截图保存为 T2-3.jpg，源文件命名为 T2-3.html，保存到考生文件夹的 T2 目录。（10 分）

## 良好编程习惯和职业素养（10 分）

- 答题过程中保持安静。（3 分）
- 有问题举手示意，有礼貌。（3 分）
- 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。（4 分）

## （2）实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。计算机安装有 Python、PyCharm 集成开发环境、Echarts、VSCode，Python 中安装 numpy、pandas、matplotlib、pyecharts 等模块，安装 Office 办公软件。

## （3）考核时量及要求

- 考核时长 180 分钟；
- 请在桌面建立以“学号\_题号”命名的成果文件夹，所有提交文件均放在该目录下。  
文件夹命名例如：201965110001\_J2\_1；
- 将各任务源文件分别保存在考生文件夹下的对应目录，运行结果分别截图。

## （4）评分细则

表 1 任务一评分细则（60 分）

| 序号 | 评分项        | 分值   | 评分细则                                                                                                                                                              |
|----|------------|------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 电影名和票房处理正确 | 16 分 | 和参考答案一致满分；<br>电影名处理正确，得 6 分；<br>票房数据处理正确，得 6 分；<br>有 <code>print(data.head())</code> 或者 <code>print(data.head(5))</code> 代码，得 2 分；<br>T1 文件夹中有 T1-1.jpg 文件，得 2 分。 |
| 2  | 导演和类型处理正确  | 16 分 | 和参考答案一致满分；<br>导演名称处理正确，得 6 分；<br>类型处理正确，得 6 分；<br>有 <code>print(data.head())</code> 或者 <code>print(data.head(5))</code> 代码，得 2 分；<br>T1 文件夹中有 T1-2.jpg 文件，得 2 分。  |
| 3  | 上映时间处理正确   | 12 分 | 和参考答案一致满分；<br>上映时间处理正确，得 8 分；<br>有 <code>print(data.head())</code> 或者 <code>print(data.head(5))</code> 代码，得 2 分；<br>T1 文件夹中有 T1-3.jpg 文件，得 2 分。                   |
| 4  | 清洗后数据正确保存  | 8 分  | T1 文件夹中有 movies_清洗后.csv 文件，得 8 分。                                                                                                                                 |
| 5  | 源代码命名正确并保存 | 4 分  | T1 文件夹中有 T1.py 文件，得 4 分                                                                                                                                           |
| 6  | 代码注释       | 4 分  | 代码中有注释，得 4 分                                                                                                                                                      |

表 2 任务二评分细则（30 分）

| 序号 | 评分项                | 分值   | 评分细则                                                                                                           |
|----|--------------------|------|----------------------------------------------------------------------------------------------------------------|
| 1  | 折线图展示正确、源代码命名正确并保存 | 10 分 | 运行结果和参考答案一致满分；<br>标题正确，得 2 分；<br>折线图展示正确，得 6 分；<br>T2 文件夹中有 T2-1.html 文件，得 1 分；<br>T2 文件夹中有 T2-1.jpg 文件，得 1 分。 |
| 2  | 环形图展示正确、源代码命名正确并保存 | 10 分 | 运行结果和参考答案一致满分；<br>标题正确，得 2 分；<br>环形图展示正确，得 6 分；<br>T2 文件夹中有 T2-2.jpg 文件，得 1 分；<br>T2 文件夹中有 T2-2.html 文件，得 1 分。 |
| 3  | 柱状图展示正确、源代码命名正确并保存 | 10 分 | 运行结果和参考答案一致满分；<br>标题正确，得 2 分；<br>柱状图展示正确，得 6 分；<br>T2 文件夹中有 T2-3.jpg 文件，得 1 分；<br>T2 文件夹中有 T2-3.html 文件，得 1 分。 |

表 3 职业素养评分细则（10 分）

| 序号 | 评分项  | 分值  | 评分细则                              |
|----|------|-----|-----------------------------------|
| 1  | 操作规范 | 3 分 | 答题过程中保持安静，不损坏机器                   |
| 2  | 文明考试 | 3 分 | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌；      |
| 3  | 工位整理 | 4 分 | 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |

## 2.试题编号：J4-1-2：豆瓣小说数据处理与可视化

### （1）任务描述

豆瓣（douban）是一个社区网站，创立于 2005 年 3 月 6 日。该网站以书影音起家，提供关于书籍、电影、音乐等作品的信息，无论描述还是评论都由用户提供。“豆瓣读书”中的小说标签提供了小说的相关信息。请以豆瓣网站中的小说数据为目标进行数据处理、展示的工作，具体任务如下：

#### 任务一：数据处理与分析（60 分）

使用 Python 读取“T1\_豆瓣小说.csv”文件，完成如下数据处理和分析操作：

- 从“作者、出版社、出版日期、价格”的信息中抽取“出版日期”出来成为单列，完成之后打印出前五五行数据，将代码和结果截图，命名为 T1-1.jpg，并保存到考生文件夹的 T1 目录。（12 分）
- 将“出版日期”改为“出版年份”，对应的数据也只保留年份，如“2012-8-1”修改为“2012”，并将出版年份这列数据改为 int 类型，完成之后打印出前五五行数据，将代码和结果截图，命名为 T1-2.jpg，并保存到考生文件夹的 T1 目录。（20 分）
- 统计出近 10 年（取数据中前 10 个年份）出版小说的平均评分（按出版年份分组统计），要求保留到 2 位小数，并把年份以及对应的平均评分输出到控制台和文件“豆瓣小说\_平均评分.csv”，将代码截图，命名为 T1-3.jpg，并将截图和文件保存到考生文件夹的 T1 目录。（20 分）
- 将源代码文件命名为 T1.py，保存到考生文件夹的 T1 目录。（4 分）
- 代码注释。在代码中有关键代码的注释。（4 分）

#### 任务二：数据可视化（30 分）

根据题目需求读取相关数据，使用 echarts 或 pyecharts 完成如下题目：

- 
- 使用条形图展示前 10 项较高评分的图书名称，标题为“评分 TOP10 的图书”，标题颜色为红色，标题字体大小为 20 像素，居中显示。要求将运行结果截图保存为 T2-1.jpg，源文件命名为 T2-1.html，保存到考生文件夹的 T2 目录。（10 分）
  - 使用散点图展示评分与评价数之间的关系，标题为“评分和评价数量之间的关系”，要求将运行结果截图保存为 T2-2.jpg，源文件命名为 T2-2.html，保存到考生文件夹的 T2 目录。（10 分）
  - 使用折线图展示近 10 年出版小说的平均评分情况，标题为“近 10 年小说平均评分”，要求将运行结果截图保存为 T2-3.jpg，源文件命名为 T2-3.html，保存到考生文件夹的 T2 目录。（10 分）

#### 良好编程习惯和职业素养（10 分）

- 答题过程中保持安静。（3 分）
- 有问题举手示意，有礼貌。（3 分）
- 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。（4 分）

### （2）实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。计算机安装有 Python、PyCharm 集成开发环境、Echarts、VSCode，Python 中安装 numpy、pandas、matplotlib、pyecharts 等模块，安装 Office 办公软件。

### （3）考核时量及要求

- 考核时长 180 分钟；
- 请在桌面建立以“学号\_题号”命名的成果文件夹，所有提交文件均放在该目录下。  
文件夹命名例如：201965110001\_J2\_1；
- 将各任务源文件分别保存在考生文件夹下的对应目录，运行结果分别截图。

### （4）评分细则

表 1 任务一评分细则（60 分）

| 序号 | 评分项             | 分值   | 评分细则                                                                                                                                                                           |
|----|-----------------|------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 出版日期数据处理正确      | 12 分 | 和参考答案一致满分；<br>新增出版日期列，得 8 分；<br>有 <code>print(data.head())</code> 或者 <code>print(data.head(5))</code> 代码，得 2 分；<br>T1 文件夹中有 T1-1.jpg 文件，得 2 分。                                 |
| 2  | 出版年份数据处理正确      | 20 分 | 和参考答案一致满分；<br>重命名正确，得 6 分；<br>提取年份正确，得 6 分；<br>类型转换正确，得 4 分；<br>有 <code>print(data.head())</code> 或者 <code>print(data.head(5))</code> 代码，得 2 分；<br>T1 文件夹中有 T1-2.jpg 文件，得 2 分。 |
| 3  | 统计小说平均评分正确并正确保存 | 20 分 | 和参考答案一致满分；<br>按照出版年份分组求评分的平均值，得 6 分；<br>按照出版年份降序排列，得 6 分；<br>按要求输出，得 4 分；<br>T1 文件夹中有豆瓣小说_平均评分.csv 文件，得 2 分；<br>T1 文件夹中有 T1-3.jpg 文件，得 2 分。                                    |
| 4  | 源代码命名正确并保存      | 4 分  | T1 文件夹中有 T1.py 文件，得 4 分                                                                                                                                                        |
| 5  | 代码注释            | 4 分  | 代码中有注释，得 4 分                                                                                                                                                                   |

表 2 任务二评分细则（30 分）

| 序号 | 评分项                | 分值   | 评分细则                                                                                                           |
|----|--------------------|------|----------------------------------------------------------------------------------------------------------------|
| 1  | 条形图展示正确、源代码命名正确并保存 | 10 分 | 运行结果和参考答案一致满分；<br>标题正确，得 2 分；<br>条形图展示正确，得 6 分；<br>T2 文件夹中有 T2-1.html 文件，得 1 分；<br>T2 文件夹中有 T2-1.jpg 文件，得 1 分。 |
| 2  | 散点图展示正确、源代码命名正确并保存 | 10 分 | 运行结果和参考答案一致满分；<br>标题正确，得 2 分；<br>散点图展示正确，得 6 分；<br>T2 文件夹中有 T2-2.jpg 文件，得 1 分；<br>T2 文件夹中有 T2-2.html 文件，得 1 分。 |
| 3  | 折线图展示正确、源代码命名正确并保存 | 10 分 | 运行结果和参考答案一致满分；<br>标题正确，得 2 分；<br>折线图展示正确，得 6 分；<br>T2 文件夹中有 T2-3.jpg 文件，得 1 分；<br>T2 文件夹中有 T2-3.html 文件，得 1 分。 |

表 3 职业素养评分细则（10 分）

| 序号 | 评分项  | 分值  | 评分细则                              |
|----|------|-----|-----------------------------------|
| 1  | 操作规范 | 3 分 | 答题过程中保持安静，不损坏机器                   |
| 2  | 文明考试 | 3 分 | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌；      |
| 3  | 工位整理 | 4 分 | 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |

### 3.试题编号：J4-1-3：猫眼电影数据处理与可视化

#### （1）任务描述

猫眼电影是美团旗下的一家集媒体内容、在线购票、用户互动社交、电影衍生品销售等服务的一站式电影互联网平台，也是人们了解电影信息的重要渠道。某影视公司想要了解最新的电影信息以此为依据来制定自己的企业发展计划，请帮助该公司完成电影数据的处理和展示的工作，具体任务如下：

#### 任务一：数据处理与分析（60 分）

使用 Python 的 pandas 读取“T1\_猫眼电影.csv”文件，完成如下数据处理和分析操作：

- 删除数据中的缺失值，并将类型中的数据进行拆分，有多个类型则只保留第一个。  
例如，“剧情/战争/历史”改为“剧情”，并输出数据的前五行，将代码和结果截图，命名为 T1-1.jpg，并保存到考生文件夹的 T1 目录。（10 分）
- 将上映时间只保留年份，例如“2020-08-21”改为“2020”，并输出数据的前五行，将代码和结果截图，命名为 T1-2.jpg，并保存到考生文件夹的 T1 目录。（6 分）
- 打印数据的行数和列数，删除评分中值为“暂无评分”的数据行，并将评分这列的数据类型转换为 float 类型，再次打印数据的行数和列数，将代码和结果截图，命名为 T1-3.jpg，并保存到考生文件夹的 T1 目录。（12 分）
- 将处理好的数据保存为“猫眼电影\_清洗后.csv”，并保存到考生文件夹的 T1 目录。（4 分）
- 统计出每年评分最高电影的上映时间、电影名和评分信息，将结果输出到控制台和文件“每年评分最高电影.csv”中，将代码和结果截图（可以截取最后 5 条数据），命名为 T1-5.jpg，和文件一起并保存到考生文件夹的 T1 目录。（20 分）

---

➤ 将源代码文件命名为 T1.py，保存到考生文件夹的 T1 目录。（4 分）

➤ 代码注释。在代码中有关键代码的注释。（4 分）

### 任务二：数据可视化（30 分）

根据题目需求读取相关数据，使用 echarts 或 pyecharts 完成如下题目：

➤ 使用折线图展示每年上映的电影数量变化趋势，标题为“每年上映电影数量”，标题颜色为红色，标题字体大小为 20 像素，居中显示。要求将运行结果截图保存为

T2-1.jpg，源文件命名为 T2-1.html，保存到考生文件夹的 T2 目录。（15 分）

➤ 使用环形图展示各类型电影数量占比情况，要求内圆半径设为“50%”，外圆半径设为“75%”，标题为“各类型电影数量占比”。要求将运行结果截图保存为 T2-2.jpg，

源文件命名为 T2-2.html，保存到考生文件夹的 T2 目录。（15 分）

### 良好编程习惯和职业素养（10 分）

➤ 答题过程中保持安静。（3 分）

➤ 有问题举手示意，有礼貌。（3 分）

➤ 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。（4 分）

### （2）实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。

计算机安装有 Python、PyCharm 集成开发环境、Echarts、VSCode，Python 中安装 numpy、pandas、matplotlib、pyecharts 等模块，安装 Office 办公软件。

### （3）考核时量及要求

➤ 考核时长 180 分钟；

➤ 请在桌面建立以“学号\_题号”命名的成果文件夹，所有提交文件均放在该目录下。

文件夹命名例如：201965110001\_J2\_1；

➤ 将各任务源文件分别保存在考生文件夹下的对应目录，运行结果分别截图。

### （4）评分细则

表 1 任务一评分细则（60 分）

| 序号 | 评分项                 | 分值   | 评分细则                                                                                                                                                                  |
|----|---------------------|------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 类型处理正确              | 10 分 | 和参考答案一致满分；<br>成功删除数据中的缺失值，得 2 分；<br>类型数据拆分正确，得 4 分；<br>有 <code>print(data.head())</code> 或者 <code>print(data.head(5))</code> 代码，得 2 分；<br>T1 文件夹中有 T1-1.jpg 文件，得 2 分。 |
| 2  | 上映时间处理正确            | 6 分  | 和参考答案一致满分；<br>上映时间处理正确，得 2 分；<br>有 <code>print(data.head())</code> 或者 <code>print(data.head(5))</code> 代码，得 2 分；<br>T1 文件夹中有 T1-2.jpg 文件，得 2 分。                       |
| 3  | 评分数据处理正确            | 12 分 | 和参考答案一致满分；<br>打印行数和列数，得 2 分；<br>删除评分中值为“暂无评分”的数据行，得4分；<br>将评分列转为 float 型，得 2 分；<br>再次打印行数和列数，得 2 分；<br>T1 文件夹中有 T1-3.jpg 文件，得 2 分。                                    |
| 4  | 清洗后数据正确保存           | 4 分  | T1 文件夹中有猫眼电影_清洗后.csv 文件，4 分。                                                                                                                                          |
| 5  | 正确统计每年评分最高电影信息处理并保存 | 20 分 | 和参考答案一致满分；<br>正确统计每年评分最高电影的信息，得 12 分；<br>正确输出，得 4 分；<br>正确保存，得 2 分；<br>T1 文件夹中有 T1-5.jpg 文件，得 2 分。                                                                    |
| 6  | 源代码命名正确并保存          | 4 分  | T1 文件夹中有 T1.py 文件，得 4 分                                                                                                                                               |
| 7  | 代码注释                | 4 分  | 代码中有注释，得 4 分                                                                                                                                                          |

表 2 任务二评分细则（30 分）

| 序号 | 评分项                | 分值   | 评分细则                                                                                                            |
|----|--------------------|------|-----------------------------------------------------------------------------------------------------------------|
| 1  | 折线图展示正确、源代码命名正确并保存 | 15 分 | 运行结果和参考答案一致满分；<br>标题正确，得 3 分；<br>折线图展示正确，得 10 分；<br>T2 文件夹中有 T2-1.html 文件，得 1 分；<br>T2 文件夹中有 T2-1.jpg 文件，得 1 分。 |
| 2  | 环形图展示正确、源代码命名正确并保存 | 15 分 | 运行结果和参考答案一致满分；<br>标题正确，得 3 分；<br>环形图展示正确，得 10 分；<br>T2 文件夹中有 T2-1.html 文件，得 1 分；<br>T2 文件夹中有 T2-1.jpg 文件，得 1 分。 |

表 3 职业素养评分细则（10 分）

| 序号 | 评分项  | 分值  | 评分细则                              |
|----|------|-----|-----------------------------------|
| 1  | 操作规范 | 3 分 | 答题过程中保持安静，不损坏机器                   |
| 2  | 文明考试 | 3 分 | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌；      |
| 3  | 工位整理 | 4 分 | 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |

#### 4.试题编号：J4-1-4：淘书网图书信息的采集与可视化

##### (1) 任务描述

淘书网是国内图书品种最全的图书网站之一，是国内特价稀缺图书品种涵盖最多的网上书店。现需要对图书信息进行了解，具体任务如下：

##### 任务一：数据处理与分析（60 分）

使用 Python 的 pandas 读取“T1\_淘书网.csv”文件，完成如下数据处理和分析操作：

- 将“出版日期”改为“出版年份”，对应的数据也只保留年份，如“2012-09-01”修改为“2012”，并将出版年份这列数据改为 int 类型，完成之后打印出前五行业数据，将代码和结果截图，命名为 T1-1.jpg，并保存到考生文件夹的 T1 目录。（16 分）
- 将折扣这列数据用百分比的形式呈现，例如“43 折”修改为“43%”，完成之后打印出前五行业数据，将代码和结果截图，命名为 T1-2.jpg，并保存到考生文件夹的 T1 目录。（12 分）
- 将处理好的数据保存为“淘书网\_清洗后.csv”，并保存到考生文件夹的 T1 目录。（8 分）
- 统计出近 10 年（取数据中前 10 个年份）出版图书的数量（按出版年份分组统计），并把年份以及对应的图书数量输出到控制台和“淘书网\_图书数量.csv”，将代码截图，命名为 T1-4.jpg，并将截图和文件保存到考生文件夹的 T1 目录。（16 分）
- 将源代码文件命名为 T1.py，保存到考生文件夹的 T1 目录。（4 分）
- 代码注释。在代码中有关键代码的注释。（4 分）

##### 任务二：数据可视化（30 分）

根据题目需求读取相关数据，使用 echarts 或 pyecharts 完成如下题目：

- 使用条形图显示前 10 个折扣最高的图书名称，标题为“折扣最高图书 TOP 10”。要

---

求将运行结果截图保存为 T2-1.jpg，源文件命名为 T2-1.html，保存到考生文件夹的 T2 目录。（15 分）

- 使用折线图展示每年出版的图书数量变化趋势，标题为“每年出版图书数量”，标题颜色为红色，标题字体大小为 20 像素，居中显示。要求将运行结果截图保存为 T2-2.jpg，源文件命名为 T2-2.html，保存到考生文件夹的 T2 目录。（15 分）

### 良好编程习惯和职业素养（10 分）

- 答题过程中保持安静。（3 分）
- 有问题举手示意，有礼貌。（3 分）
- 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。（4 分）

## （2）实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。计算机安装有 Python、PyCharm 集成开发环境、Echarts、VSCode，Python 中安装 numpy、pandas、matplotlib、pyecharts 等模块，安装 Office 办公软件。

## （3）考核时量及要求

- 考核时长 180 分钟；
- 请在桌面建立以“学号\_题号”命名的成果文件夹，所有提交文件均放在该目录下。  
文件夹命名例如：201965110001\_J2\_1；
- 将各任务源文件分别保存在考生文件夹下的对应目录，运行结果分别截图。

## （4）评分细则

表 1 任务一评分细则（60 分）

| 序号 | 评分项                 | 分值   | 评分细则                                                                                                                                                                               |
|----|---------------------|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 出版日期数据处理正确          | 16 分 | 和参考答案一致满分；<br>重命名正确，得 4 分；<br>提取年份数据正确，得 4 分；<br>数据类型转换正确，得 4 分；<br>有 <code>print(data.head())</code> 或者 <code>print(data.head(5))</code> 代码，得 2 分；<br>T1 文件夹中有 T1-1.jpg 文件，得 2 分。 |
| 2  | 折扣列数据处理正确           | 12 分 | 和参考答案一致满分；<br>折扣数据处理正确，得 8 分；<br>有 <code>print(data.head())</code> 或者 <code>print(data.head(5))</code> 代码，得 2 分；<br>T1 文件夹中有 T1-2.jpg 文件，得 2 分。                                    |
| 3  | 清洗后数据正确保存           | 8 分  | T1 文件夹中有淘书网_清洗后.csv 文件。                                                                                                                                                            |
| 4  | 正确统计近 10 年出版图书数量并保存 | 16 分 | 和参考答案一致满分；<br>按照出版年份分组求图书数量，得 4 分；<br>按照出版年份降序排列，得 4 分；<br>按要求输出，得 2 分；<br>T1 文件夹中有淘书网_图书数量.csv 文件，得 4 分；<br>T1 文件夹中有 T1-4.jpg 文件，得 2 分。                                           |
| 5  | 源代码命名正确并保存          | 4 分  | T1 文件夹中有 T1.py 文件，得 4 分                                                                                                                                                            |
| 6  | 代码注释                | 4 分  | 代码中有注释，得 4 分                                                                                                                                                                       |

表 2 任务二评分细则（30 分）

| 序号 | 评分项                | 分值   | 评分细则                                                                                                            |
|----|--------------------|------|-----------------------------------------------------------------------------------------------------------------|
| 1  | 条形图展示正确、源代码命名正确并保存 | 15 分 | 运行结果和参考答案一致满分；<br>标题正确，得 3 分；<br>条形图展示正确，10 分；<br>T2 文件夹中有 T2-1.jpg 文件，得 1 分；<br>T2 文件夹中有 T2-1.html 文件，得 1 分。   |
| 2  | 折线图展示正确、源代码命名正确并保存 | 15 分 | 运行结果和参考答案一致满分；<br>标题正确，得 3 分；<br>折线图展示正确，得 10 分；<br>T2 文件夹中有 T2-2.jpg 文件，得 1 分；<br>T2 文件夹中有 T2-2.html 文件，得 1 分。 |

表 3 职业素养评分细则（10 分）

| 序号 | 评分项  | 分值  | 评分细则                              |
|----|------|-----|-----------------------------------|
| 1  | 操作规范 | 3 分 | 答题过程中保持安静，不损坏机器                   |
| 2  | 文明考试 | 3 分 | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌；      |
| 3  | 工位整理 | 4 分 | 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |

---

## 5.试题编号：J4-1-5：小说排行榜数据处理与可视化

### （1）任务描述

随着手机的普及，手机看小说也越来越普遍。某公司基于这一现象想要开发一款小说推荐软件，因此需要了解小说的下载数据。请将小说信息下载下来，并进行相关的处理和展示，具体的任务如下：

#### 任务一：数据处理与分析（60 分）

使用 Python 的 pandas 读取“T1\_电子书.csv”文件，完成如下数据处理和分析操作：

- 将文件大小这一列数据的单位进行处理，如“522KB”处理成“522”，并将该列数据类型转换为 float，输出数据的前五行，将代码和结果截图，命名为 T1-1.jpg，并保存到考生文件夹的 T1 目录（16 分）；
- 打印数据的行数和列数，删除总下载次数低于 1000（低关注度）的数据，再次打印数据的行数和列数，将代码和结果截图，命名为 T1-2.jpg，并保存到考生文件夹的 T1 目录。（16 分）
- 将处理好的数据保存为“电子书\_清洗后.csv”，并保存到考生文件夹的 T1 目录。（8 分）
- 打印今日下载次数最多的 10 本小说信息，将代码和结果截图，命名为 T1-4.jpg，并保存到考生文件夹的 T1 目录（12 分）；
- 将源代码文件命名为 T1.py，保存到考生文件夹的 T1 目录。（4 分）
- 代码注释。在代码中有关键代码的注释。（4 分）

#### 任务二：数据可视化（30 分）

根据题目需求读取相关数据，使用 echarts 或 pyecharts 完成如下题目：

- 使用条形图显示今日下载次数最多的 10 本小说名称，标题为“日下载次数 TOP10 图书”。要求将运行结果截图保存为 T2-1.jpg，源文件命名为 T2-1.html，保存到考生文件夹的 T2 目录。（15 分）
- 使用折线图展示总下载次数最多的 10 本小说，标题为“总下载次数 TOP10 图书”。要求将运行结果截图保存为 T2-2.jpg，源文件命名为 T2-2.html，保存到考生文件夹

---

的 T2 目录。(15 分)

#### 良好编程习惯和职业素养 (10 分)

- 答题过程中保持安静。(3 分)
- 有问题举手示意, 有礼貌。(3 分)
- 答题完毕离场时整理好工位, 不关闭计算机, 并按要求显示答案保存界面。(4 分)

#### (2) 实施条件

安排在教学实训机房, 计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。计算机安装有 Python、PyCharm 集成开发环境、Echarts、VSCode, Python 中安装 numpy、pandas、matplotlib、pyecharts 等模块, 安装 Office 办公软件。

#### (3) 考核时量及要求

- 考核时长 180 分钟;
- 请在桌面建立以“学号\_题号”命名的成果文件夹, 所有提交文件均放在该目录下。

文件夹命名例如: 201965110001\_J2\_1;

- 将各任务源文件分别保存在考生文件夹下的对应目录, 运行结果分别截图。

#### (4) 评分细则

表 1 任务一评分细则 (60 分)

| 序号 | 评分项          | 分值   | 评分细则                                                                                                                                                                 |
|----|--------------|------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 文件大小数据处理正确   | 16 分 | 和参考答案一致满分；<br>文件大小数据处理正确，得 6 分；<br>数据类型转换正确，得 6 分；<br>有 <code>print(data.head())</code> 或者 <code>print(data.head(5))</code> 代码，得 2 分；<br>T1 文件夹中有 T1-1.jpg 文件，得 2 分。 |
| 2  | 正确删除总下载次数数据  | 16 分 | 和参考答案一致满分；<br>打印数据的行数和列数，得 2 分；<br>正确删除总下载次数低于 1000 的数据，得 10 分；<br>再次正确打印数据的行数和列数，得 2 分；<br>T1 文件夹中有 T1-2.jpg 文件，得 2 分。                                              |
| 3  | 清洗后数据正确保存    | 8 分  | T1 文件夹中有电子书_清洗后.csv 文件，得 8 分。                                                                                                                                        |
| 4  | 今日下载次数数据处理正确 | 12 分 | 和参考答案一致满分；<br>利用排序函数正确获取今日下载次数最多的 10 本小说信息，得 8 分；<br>将结果正确的输出，得 2 分；<br>T1 文件夹中有 T1-4.jpg 文件，得 2 分。                                                                  |
| 5  | 源代码命名正确并保存   | 4 分  | T1 文件夹中有 T1.py 文件，得 4 分                                                                                                                                              |
| 6  | 代码注释         | 4 分  | 代码中有注释，得 4 分                                                                                                                                                         |

表 2 任务二评分细则（30 分）

| 序号 | 评分项                | 分值   | 评分细则                                                                                                            |
|----|--------------------|------|-----------------------------------------------------------------------------------------------------------------|
| 1  | 条形图展示正确、源代码命名正确并保存 | 15 分 | 运行结果和参考答案一致满分；<br>标题正确，得 3 分；<br>条形图展示正确，得 10 分；<br>T2 文件夹中有 T2-1.jpg 文件，得 1 分；<br>T2 文件夹中有 T2-1.html 文件，得 1 分。 |
| 2  | 折线图展示正确、源代码命名正确并保存 | 15 分 | 运行结果和参考答案一致满分；<br>标题正确，得 3 分；<br>折线图展示正确，得 10 分；<br>T2 文件夹中有 T2-2.jpg 文件，得 1 分；<br>T2 文件夹中有 T2-2.html 文件，得 1 分。 |

表 3 职业素养评分细则（10 分）

| 序号 | 评分项  | 分值  | 评分细则                              |
|----|------|-----|-----------------------------------|
| 1  | 操作规范 | 3 分 | 答题过程中保持安静，不损坏机器                   |
| 2  | 文明考试 | 3 分 | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌；      |
| 3  | 工位整理 | 4 分 | 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |

---

## 6.试题编号：J4-1-6：当当图书好评数据处理与可视化

### （1）任务描述

某图书馆为了满足读者需求，要购买一批新书，但是图书选择却很困难。因此急切需要了解最新的图书信息，具体的任务如下：

#### 任务一：数据处理与分析（60 分）

使用 Python 的 pandas 读取“T1\_当当图书.csv”文件，进行如下处理：

- 删除图书名称中的备注部分，如“局外人(诺贝尔文学奖得主加缪代表作)”修改为“局外人”，输出数据的前五行，将代码和结果截图，命名为 T1-1.jpg，并保存到考生文件夹的 T1 目录。（12 分）
- 统计出 2020 年每个月出版图书的数量，并输出结果，将代码和结果截图，命名为 T1-2.jpg，并保存到考生文件夹的 T1 目录。（16 分）
- 统计五星好评数 TOP10 的图书信息，并输出结果，将代码和结果截图，命名为 T1-3.jpg，并保存到考生文件夹的 T1 目录。（16 分）
- 将处理好的数据保存为“当当图书\_清洗后.csv”，并保存到考生文件夹的 T1 目录。（8 分）
- 将源代码文件命名为 T1.py，保存到考生文件夹的 T1 目录。（4 分）
- 代码注释。在代码中有关键代码的注释。（4 分）

#### 任务二：数据可视化（30 分）

根据题目需求使用 Python 读取相关数据，使用 matplotlib 或 Tableau 绘制图表：

- 使用柱状图显示五星好评数 TOP5 的图书名称，标题为“五星好评数 TOP5 图书”。要求将运行结果截图保存为 T2-1.jpg，源文件命名为 T2-1.py，保存到考生文件夹的 T2 目录。（15 分）
- 使用折线图展示 2020 年和 2021 年每个月出版的图书数量，显示图例，标题为“2020 年和 2021 年出版的图书数量”。要求将运行结果截图保存为 T2-2.jpg，源文件命名为 T2-2.py，保存到考生文件夹的 T2 目录。（15 分）

#### 良好编程习惯和职业素养（10 分）

- 答题过程中保持安静。(3 分)
- 有问题举手示意,有礼貌。(3 分)
- 答题完毕离场时整理好工位,不关闭计算机,并按要求显示答案保存界面。(4 分)

## (2) 实施条件

安排在教学实训机房,计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。  
计算机安装有 Python、PyCharm 集成开发环境、Echarts、VSCode, Python 中安装 numpy、pandas、matplotlib、pyecharts 等模块,安装 Office 办公软件。

## (3) 考核时量及要求

- 考核时长 180 分钟;
- 请在桌面建立以“学号\_题号”命名的成果文件夹,所有提交文件均放在该目录下。  
文件夹命名例如:201965110001\_J2\_1;
- 将各任务源文件分别保存在考生文件夹下的对应目录,运行结果分别截图。

## (4) 评分细则

表 1 任务一评分细则 (60 分)

| 序号 | 评分项                  | 分值   | 评分细则                                                                                                                                              |
|----|----------------------|------|---------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 图书名称数据处理正确           | 12 分 | 和参考答案一致满分;<br>图书名称数据处理正确,得 8 分;<br>有 <code>print(data.head())</code> 或者 <code>print(data.head(5))</code> 代码,得 2 分;<br>T1 文件夹中有 T1-1.jpg 文件,得 2 分。 |
| 2  | 正确统计 2020 年每个月出版图书数量 | 16 分 | 和参考答案一致满分;<br>筛选 2020 年数据,得 6 分;<br>按月份统计出版图书数量,得 6 分;<br>正确输出结果,得 2 分;<br>T1 文件夹中有 T1-2.jpg 文件,得 2 分。                                            |
| 3  | 正确统计五星好评数 TOP10 图书信息 | 16 分 | 和参考答案一致满分;<br>转换五星好评数数据类型,得 4 分;<br>正确获取五星好评数前 10 数据信息,得 8 分;<br>正确输出结果,得 2 分;<br>T1 文件夹中有 T1-4.jpg 文件,得 2 分。                                     |
| 4  | 清洗后数据正确保存            | 8 分  | T1 文件夹中有当当图书_清洗后.csv 文件,得 8 分。                                                                                                                    |
| 5  | 源代码命名正确并保存           | 4 分  | T1 文件夹中有 T1.py 文件,得 4 分                                                                                                                           |
| 6  | 代码注释                 | 4 分  | 代码中有注释,得 4 分                                                                                                                                      |

表 2 任务二评分细则 (30 分)

| 序号 | 评分项                | 分值   | 评分细则                                                                                                                                           |
|----|--------------------|------|------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 柱状图展示正确、源代码命名正确并保存 | 15 分 | 运行结果和参考答案一致满分；<br>标题正确，得 3 分；<br>没有中文乱码，得 2 分；<br>柱状图展示正确，得 8 分；<br>T2 文件夹中有 T2-1.jpg 文件，得 1 分。<br>T2 文件夹中有 T2-1.py 文件，得 1 分。                  |
| 2  | 折线图展示正确、源代码命名并保存   | 15 分 | 运行结果和参考答案一致满分；<br>标题正确，得 3 分；<br>没有中文乱码，得 2 分；<br>折线图展示正确，得 6 分；<br>显示图例正确，得 2 分；<br>T2 文件夹中有 T2-2.jpg 文件，得 1 分；<br>T2 文件夹中有 T2-2.py 文件，得 1 分。 |

| 序号 | 评分项  | 分值  | 评分细则                              |
|----|------|-----|-----------------------------------|
| 1  | 操作规范 | 3 分 | 答题过程中保持安静，不损坏机器                   |
| 2  | 文明考试 | 3 分 | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌；      |
| 3  | 工位整理 | 4 分 | 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |

表 3 职业素养评分细则（10 分）

## 项目 2：社会服务行业数据处理与可视化

### 7.试题编号：J4-2-1：链家长沙小区数据处理与可视化

#### （1）任务描述

某青年想要在长沙购买住房，但是对长沙的小区却不了解，请帮助他获得长沙小区信息并进行分析和展示，具体的任务如下：

#### 任务一：数据处理与分析（60 分）

使用 Python 的 pandas 读取“链家长沙小区.csv”文件，完成如下数据处理和分析操作：

- 将参考价格这一列数据处理成以元为单位的数字，如“5456 元/m2”处理成“5456”，并将该列的数据类型转换成 float，输出数据的前五行，将代码和结果截图，命名为 T1-1.jpg，并保存到考生文件夹的 T1 目录（12 分）；
- 打印数据的行数和列数，将建成年份的单位去掉，如“1986 年”改为“1986”，建成年份为“未知”的数据删除，低于 2000 年（不包括 2000）的数据删除，再次打印数据

---

的行数和列数，将代码和结果截图，命名为 T1-2.jpg，并保存到考生文件夹的 T1 目录。(16 分)

- 将处理好的数据保存为“链家长沙\_清洗后.csv”，并保存到考生文件夹的 T1 目录。  
(8 分)
- 统计出长沙各个区域房子的平均价格，要求价格保留小数点后 2 位，将结果输出到控制台和文件“长沙各区域\_平均房价.csv”，将代码和结果截图，命名为 T1-4.jpg，并和输出文件一起保存到考生文件夹的 T1 目录。(16 分)
- 将源代码文件命名为 T1.py，保存到考生文件夹的 T1 目录。(4 分)
- 代码注释。在代码中有关键代码的注释。(4 分)

### 任务二：数据可视化 (30 分)

根据题目需求读取数据，使用 echarts 或 pyecharts 完成如下题目：

- 使用饼图展示各个区域在售房子数量占比情况，标题为“各区域在售房子数量占比”，要求将运行结果截图保存为 T2-1.jpg，源文件命名为 T2-1.html，保存到考生文件夹的 T2 目录。(15 分)
- 使用条形图展示各个区域 90 天内房屋成交总量，标题为“各区域在 90 天内房屋成交量”，要求将运行结果截图保存为 T2-2.jpg，源文件命名为 T2-2.html，保存到考生文件夹的 T2 目录。(15 分)

### 良好编程习惯和职业素养 (10 分)

- 答题过程中保持安静。(3 分)
- 有问题举手示意，有礼貌。(3 分)
- 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。(4 分)

### (2) 实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。

计算机安装有 Python、PyCharm 集成开发环境、Echarts、VSCode，Python 中安装 numpy、pandas、matplotlib、pyecharts 等模块，安装 Office 办公软件。

### (3) 考核时量及要求

- 考核时长 180 分钟；
- 请在桌面建立以“学号\_题号”命名的成果文件夹，所有提交文件均放在该目录下。  
文件夹命名例如：201965110001\_J2\_1；
- 将各任务源文件分别保存在考生文件夹下的对应目录，运行结果分别截图。

### (4) 评分细则

表 1 任务一评分细则（60 分）

| 序号 | 评分项           | 分值   | 评分细则                                                                                                                                                             |
|----|---------------|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 参考价格数据处理正确    | 12 分 | 和参考答案一致满分；<br>正确去除单位，得 4 分；<br>转换数据类型正确，得 4 分；<br>有 <code>print(data.head())</code> 或者 <code>print(data.head(5))</code> 代码，得 2 分；<br>T1 文件夹中有 T1-1.jpg 文件，得 2 分。 |
| 2  | 建成年份数据处理正确    | 16 分 | 和参考答案一致满分；<br>正确打印行数和列数，得 2 分；<br>正确去除建成年份单位，得 4 分；<br>正确删除题目要求的数据，得 6 分；<br>再次打印行数和列数，得 2 分；<br>T1 文件夹中有 T1-2.jpg 文件，得 2 分。                                     |
| 3  | 清洗后数据正确保存     | 8 分  | T1 文件夹中有链家长沙_清洗后.csv 文件，得 8 分。                                                                                                                                   |
| 4  | 正确统计各区域房子平均价格 | 16 分 | 和参考答案一致满分；<br>利用分组函数正确统计各区域房子的平均价格，得 8 分；<br>平均价格保留小数点后 2 位，得 2 分；<br>按要求输出，得 2 分；<br>T1 文件夹中有链家长沙_清洗后.csv 文件，得 2 分；<br>T1 文件夹中有 T1-4.jpg 文件，得 2 分。              |
| 5  | 源代码命名正确并保存    | 4 分  | T1 文件夹中有 T1.py 文件，得 4 分                                                                                                                                          |
| 6  | 代码注释          | 4 分  | 代码中有注释，得 4 分                                                                                                                                                     |

表 2 任务二评分细则（30 分）

| 序号 | 评分项                | 分值   | 评分细则                                                                                                            |
|----|--------------------|------|-----------------------------------------------------------------------------------------------------------------|
| 1  | 饼图展示正确、源代码命名正确并保存  | 15 分 | 运行结果和参考答案一致满分；<br>标题正确，得 3 分；<br>饼图展示正确，得 10 分；<br>T2 文件夹中有 T2-1.html 文件，得 1 分；<br>T2 文件夹中有 T2-1.jpg 文件，得 1 分。  |
| 2  | 条形图展示正确、源代码命名正确并保存 | 15 分 | 运行结果和参考答案一致满分；<br>标题正确，得 3 分；<br>条形图展示正确，得 10 分；<br>T2 文件夹中有 T2-2.html 文件，得 1 分；<br>T2 文件夹中有 T2-2.jpg 文件，得 1 分。 |

表 3 职业素养评分细则（10 分）

| 序号 | 评分项  | 分值  | 评分细则                              |
|----|------|-----|-----------------------------------|
| 1  | 操作规范 | 3 分 | 答题过程中保持安静，不损坏机器                   |
| 2  | 文明考试 | 3 分 | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌；      |
| 3  | 工位整理 | 4 分 | 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |

## 8. 试题编号：J4-2-2：美食杰数据处理与可视化

### （1）任务描述

随着社会的发展，人民生活水平显著提高，对饮食的要求也越来越高。而年轻一代做饭基本要靠网络，请以美食杰网站为目标，获取相关的菜谱信息，具体任务如下所示：

#### 任务一：数据处理与分析（60 分）

使用 Python 的 pandas 读取“T1\_美食杰菜谱.csv”文件，完成如下数据处理和分析操作：

- 将工艺这列数据拆分为工艺和口味两列数据，如“烤 / 家常味”对应“工艺：烤”和“口味：家常味”，并输出数据的前五行，将代码和结果截图，命名为 T1-1.jpg，并保存到考生文件夹的 T1 目录。（16 分）
- 打印数据的行数和列数，将烹饪时间这列中值为“数天”、“数小时”的行删除，再次打印数据的行数和列数，将代码和结果截图，命名为 T1-2.jpg，并保存到考生文件夹的 T1 目录。（12 分）
- 将烹饪时间这列数据处理成以分钟为单位的数据，并且只包含数字，如“60 分钟”处理成“60”，“1.5 小时”处理成“90”，完成之后将烹饪时间的数据类型转换成 int，并输出烹饪时间大于 60 的前五行数据，将代码和结果截图，命名为 T1-3.jpg，并

---

保存到考生文件夹的 T1 目录。(16 分)

- 将处理好的数据保存为“美食杰\_清洗后.csv”，并保存到考生文件夹的 T1 目录。(8 分)
- 将源代码文件命名为 T1.py，保存到考生文件夹的 T1 目录。(4 分)
- 代码注释。在代码中有关键代码的注释。(4 分)

## 任务二：数据可视化 (30 分)

根据题目需求读取相关数据，使用 echarts 或 pyecharts 完成如下题目：

- 使用折线图展示人气最高的 10 个菜谱，标题为“人气 TOP10 菜谱名称”，要求将运行结果截图保存为 T2-1.jpg，源文件命名为 T2-1.html，保存到考生文件夹的 T2 目录。(10 分)
- 使用饼图展示菜谱中各类工艺的平均步数占比情况，标题为“各工艺类型菜谱平均步数占比图”，居中显示，要求将运行结果截图保存为 T2-2.jpg，源文件命名为 T2-2.html，保存到考生文件夹的 T2 目录。(10 分)
- 使用柱状图展示各种口味的菜谱数量，标题为“各种口味菜谱数量”，要求将运行结果截图保存为 T2-3.jpg，源文件命名为 T2-3.html，保存到考生文件夹的 T2 目录。(10 分)

## 良好编程习惯和职业素养 (10 分)

- 答题过程中保持安静。(3 分)
- 有问题举手示意，有礼貌。(3 分)
- 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。(4 分)

## (2) 实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。

计算机安装有 Python、PyCharm 集成开发环境、Echarts、VSCode，Python 中安装 numpy、pandas、matplotlib、pyecharts 等模块，安装 Office 办公软件。

## (3) 考核时量及要求

- 考核时长 180 分钟；

- 请在桌面建立以“学号\_题号”命名的成果文件夹，所有提交文件均放在该目录下。

文件夹命名例如：201965110001\_J2\_1；

- 将各任务源文件分别保存在考生文件夹下的对应目录，运行结果分别截图。

#### (4) 评分细则

表 1 任务一评分细则（60 分）

| 序号 | 评分项           | 分值   | 评分细则                                                                                                                                                             |
|----|---------------|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 工艺数据处理正确      | 16 分 | 和参考答案一致满分；<br>新增工艺列数据，得 6 分；<br>新增口味列数据，得 6 分；<br>有 <code>print(data.head())</code> 或者 <code>print(data.head(5))</code> 代码，得 2 分；<br>T1 文件夹中有 T1-1.jpg 文件，得 2 分。 |
| 2  | 烹饪时间数据处理正确    | 12 分 | 和参考答案一致满分；<br>正确打印行数和列数，得 2 分；<br>正确删除题目要求的数据，得 6 分；<br>再次打印行数和列数，得 2 分；<br>T1 文件夹中有 T1-2.jpg 文件，得 2 分。                                                          |
| 3  | 烹饪时间中数据单位处理正确 | 16 分 | 和参考答案一致满分；<br>烹饪时间中分钟数据处理正确，得 4 分；<br>烹饪时间中小时数据处理正确，得 6 分；<br>正确转换烹饪时间数据类型，得 2 分；<br>正确输出前五数据，得 2 分；<br>T1 文件夹中有 T1-3.jpg 文件，得 2 分                               |
| 4  | 清洗后数据正确保存     | 8 分  | T1 文件夹中有美食杰_清洗后.csv 文件，得 8 分。                                                                                                                                    |
| 5  | 源代码命名正确并保存    | 4 分  | T1 文件夹中有 T1.py 文件，得 4 分                                                                                                                                          |
| 6  | 代码注释          | 4 分  | 代码中有注释，得 4 分                                                                                                                                                     |

表 2 任务二评分细则（30 分）

| 序号 | 评分项     | 分值   | 评分细则                                                                                                           |
|----|---------|------|----------------------------------------------------------------------------------------------------------------|
| 1  | 折线图展示正确 | 10 分 | 运行结果和参考答案一致满分；<br>标题正确，得 2 分；<br>折线图展示正确，得 6 分；<br>T2 文件夹中有 T2-1.html 文件，得 1 分；<br>T2 文件夹中有 T2-1.jpg 文件，得 1 分。 |
| 2  | 饼图展示正确  | 10 分 | 运行结果和参考答案一致满分；<br>标题正确，得 2 分；<br>饼图展示正确，得 6 分；<br>T2 文件夹中有 T2-2.html 文件，得 1 分；<br>T2 文件夹中有 T2-2.jpg 文件，得 1 分。  |
| 3  | 柱状图展示正确 | 10 分 | 运行结果和参考答案一致满分；<br>标题正确，得 2 分；<br>柱状图展示正确，得 6 分；<br>T2 文件夹中有 T2-3.html 文件，得 1 分；<br>T2 文件夹中有 T2-3.jpg 文件，得 1 分。 |

表 3 职业素养评分细则（10 分）

| 序号 | 评分项  | 分值  | 评分细则                              |
|----|------|-----|-----------------------------------|
| 1  | 操作规范 | 3 分 | 答题过程中保持安静，不损坏机器                   |
| 2  | 文明考试 | 3 分 | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌；      |
| 3  | 工位整理 | 4 分 | 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |

## 9.试题编号：J4-2-3：湖南人才网数据处理与可视化

### （1）任务描述

湖南人才网创立于 2002 年初，凭借着雄厚的技术实力、专业的客户服务，以用人单位需求为导向，为企业纳才、人才求职提供高效快捷的交流平台。某应届毕业生想要了解本专业的就业请求，请将网页数据爬取下来并提供该学生。

请完成以下任务。

#### 任务一：数据处理与分析（60 分）

使用 Python 的 pandas 读取“T1\_湖南人才网数据.csv”文件，完成如下数据处理和分析操作：

- 打印数据的行数和列数，将薪资范围为“面议”的数据删除，再次打印数据的行数

---

和列数，将代码和结果截图，命名为 T1-1.jpg，并保存到考生文件夹的 T1 目录。

(8 分)

- 将薪资范围拆分为两列，分别为最低工资和最高工资，完成之后剔除最高工资为空的数据行，将最低工资和最高工资（去除单位‘元’）转换成 int 类型。最后向控制台输出以下内容：打印数据的行数和列数，输出数据的前五行。将代码和结果截图，命名为 T1-2.jpg，并保存到考生文件夹的 T1 目录。(20 分)
- 将处理好的数据保存为“湖南人才网数据\_清洗后.csv”，并保存到考生文件夹的 T1 目录。(8 分)
- 统计出不同公司性质的招聘人数、最低工资和最高工资的平均值，要求保留 2 位小数，将结果输出到控制台和文件“不同公司性质招聘情况.csv”中，要求文件的列名为“公司性质”、“招聘人数”、“最低工资”和“最高工资”，将代码和结果截图，命名为 T1-4.jpg，和文件一并保存到考生文件夹的 T1 目录。(16 分)
- 将源代码文件命名为 T1.py，保存到考生文件夹的 T1 目录。(4 分)
- 代码注释。在代码中有关键代码的注释。(4 分)

## 任务二：数据可视化 (30 分)

使用 Python 分析数据并使用 matplotlib 或 Tableau 绘制图表：

- 使用饼图展示各类性质企业招聘人数的分布比例，要求每个扇区显示百分比和标签名称，标题为“不同企业性质招聘人数”，将运行结果截图保存为 T2-1.jpg，源文件命名为 T2-1.py，保存到考生文件夹的 T2 目录。(10 分)
- 使用折线图展示各类性质企业的最低工资和最高工资，标题为“不同企业性质工资”，要求显示图例，将运行结果截图保存为 T2-2.jpg，源文件命名为 T2-2.py，保存到考生文件夹的 T2 目录。(10 分)
- 使用柱状图展示薪资最高的 5 个工作岗位，标题为“工资 TOP5 职位”，将运行结果截图保存为 T2-3.jpg，源文件命名为 T2-3.py，保存到考生文件夹的 T2 目录。(10 分)

## 良好编程习惯和职业素养 (10 分)

- 
- 答题过程中保持安静。(3 分)
  - 有问题举手示意，有礼貌。(3 分)
  - 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。(4 分)

## (2) 实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。计算机安装有 Python、PyCharm 集成开发环境、Echarts、VSCode，Python 中安装 numpy、pandas、matplotlib、pyecharts 等模块，安装 Office 办公软件。

## (3) 考核时量及要求

- 考核时长 180 分钟；
- 请在桌面建立以“学号\_题号”命名的成果文件夹，所有提交文件均放在该目录下。  
文件夹命名例如：201965110001\_J2\_1；
- 将各任务源文件分别保存在考生文件夹下的对应目录，运行结果分别截图。

## (4) 评分细则

表 1 任务一评分细则 (60 分)

| 序号 | 评分项            | 分值   | 评分细则                                                                                                                                                                                                                             |
|----|----------------|------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 正确删除薪资范围中的指定数据 | 8 分  | 和参考答案一致满分；<br>正确打印行数和列数，得 2 分；<br>正确删除面议数据，得 2 分；<br>再次正确打印行数和列数，得 2 分；<br>T1 文件夹中有 T1-1.jpg 文件，得 2 分。                                                                                                                           |
| 2  | 正确处理薪资范围数据     | 20 分 | 和参考答案一致满分；<br>将薪资范围拆分为最低工资和最高工资，得 4 分；<br>正确删除最高工资为空的数据，得 6 分；<br>最高工资和最低工资数据类型转换正确，得 4 分；<br>正确打印行数和列数，得 2 分；<br>有 <code>print(data.head())</code> 或者 <code>print(data.head(5))</code> 代码，得 2 分；<br>T1 文件夹中有 T1-2.jpg 文件，得 2 分。 |
| 3  | 清洗后数据正确保存      | 8 分  | T1 文件夹中有湖南人才网数据_清洗后.csv 文件，得 8 分。                                                                                                                                                                                                |
| 4  | 正确统计数据         | 16 分 | 和参考答案一致满分；<br>利用分组函数正确统计数据，得 6 分；<br>保留 2 位小数，得 2 分；<br>按要求输出，得 4 分；<br>T1 文件夹中有不同公司性质招聘情况.csv 文件,得 2 分；<br>T1 文件夹中有 T1-4.jpg 文件，得 2 分。                                                                                          |
| 5  | 源代码命名正确并保存     | 4 分  | T1 文件夹中有 T1.py 文件，得 4 分                                                                                                                                                                                                          |
| 6  | 代码注释           | 4 分  | 代码中有注释，得 4 分                                                                                                                                                                                                                     |

表 2 任务二评分细则（30 分）

| 序号 | 评分项                | 分值   | 评分细则                                                                                                                                           |
|----|--------------------|------|------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 饼图展示正确、源代码命名正确并保存  | 10 分 | 运行结果和参考答案一致满分；<br>标题正确，得 1 分；<br>没有中文乱码，得 1 分；<br>饼图展示正确，得 6 分；<br>T2 文件夹中有 T2-1.py 文件，得 1 分；<br>T2 文件夹中有 T2-1.jpg 文件，得 1 分。                   |
| 2  | 折线图展示正确、源代码命名正确并保存 | 10 分 | 运行结果和参考答案一致满分；<br>标题正确，得 1 分；<br>没有中文乱码，得 1 分；<br>折线图展示正确，得 5 分；<br>显示图例正确，得 1 分；<br>T2 文件夹中有 T2-2.py 文件，得 1 分；<br>T2 文件夹中有 T2-2.jpg 文件，得 1 分。 |
| 3  | 柱状图展示正确、源代码命名正确并保存 | 10 分 | 运行结果和参考答案一致满分；<br>标题正确，得 1 分；<br>没有中文乱码，得 1 分；<br>柱状图展示正确，得 6 分；<br>T2 文件夹中 T2-3.py 文件，得 1 分；<br>T2 文件夹中有 T2-3.jpg 文件，得 1 分。                   |

表 3 职业素养评分细则（10 分）

| 序号 | 评分项  | 分值  | 评分细则                              |
|----|------|-----|-----------------------------------|
| 1  | 操作规范 | 3 分 | 答题过程中保持安静，不损坏机器                   |
| 2  | 文明考试 | 3 分 | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌；      |
| 3  | 工位整理 | 4 分 | 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |

## 10.试题编号：J4-2-4：酒店数据处理与可视化

### （1）任务描述

每年的国家法定节假日都是旅游的高峰，而酒店是旅游中不可缺少的。某公司想要开发一款软件，帮助用户推荐符合自己要求的酒店，而前期需要获取酒店的相关信息，具体的工作如下：

#### 任务一：数据处理与分析（60 分）

使用 Python 的 pandas 读取“T1\_去哪网数据.csv”文件，进行如下处理：

- 删除酒店名称中的备注部分，如“铭逸美域酒店(长沙含浦店)”修改为“铭逸美域酒店”，输出数据的前五行，将代码和结果截图，命名为 T1-1.jpg，并保存到考生

---

文件夹的 T1 目录。(10 分)

- 对地址进行处理，只保留“/”前面的数据，如“砂子塘·雨花亭/林科大”转换之后为“砂子塘·雨花亭”，输出数据的前五行，将代码和结果截图，命名为 T1-2.jpg，并保存到考生文件夹的 T1 目录。(10 分)
- 统计各城市酒店的平均评分，要求保留到 2 位小数，并把城市以及对应的平均评分输出到控制台和文件“各城市酒店\_平均评分.csv”，将代码截图，命名为 T1-3.jpg，并将截图和文件保存到考生文件夹的 T1 目录。(12 分)；
- 统计各城市酒店的平均价格，要求保留到 2 位小数，并把城市以及对应的平均价格输出到控制台和文件“各城市酒店\_平均价格.csv”，将代码截图，命名为 T1-4.jpg，并将截图和文件保存到考生文件夹的 T1 目录。(12 分)；
- 将处理之后的数据输出到“酒店数据\_清洗后.csv”，并保存到考生文件夹的 T1 目录。

(8 分)

- 将源代码文件命名为 T1.py，保存到考生文件夹的 T1 目录。(4 分)
- 代码注释。在代码中有关键代码的注释。(4 分)

## 任务二：数据可视化 (30 分)

使用 Python 分析数据并使用 matplotlib 或 Tableau 绘制图表：

- 使用柱状图显示单价前 7 酒店的价格，标题为“TOP 7 酒店价格”，要求将运行结果截图保存为 T2-1.jpg，源文件命名为 T2-1.py，保存到考生文件夹的 T2 目录。(10 分)
- 使用折线图显示单价前 7 酒店的评分，标题为“TOP 7 酒店评分”，X 轴为单价信息，Y 轴为评分信息，要求将运行结果截图保存为 T2-2.jpg，源文件命名为 T2-2.py，保存到考生文件夹的 T2 目录。(10 分)
- 使用饼图展示各个城市酒店的分布情况，要求每个扇区显示百分比和数据标签，标题为“各个城市的酒店分布情况”，将运行结果截图保存为 T2-3.jpg，源文件命名为 T2-3.py，保存到考生文件夹的 T2 目录。(10 分)

## 良好编程习惯和职业素养 (10 分)

- 
- 答题过程中保持安静。(3 分)
  - 有问题举手示意，有礼貌。(3 分)
  - 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。(4 分)

## (2) 实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。计算机安装有 Python、PyCharm 集成开发环境、Echarts、VSCode，Python 中安装 numpy、pandas、matplotlib、pyecharts 等模块，安装 Office 办公软件。

## (3) 考核时量及要求

- 考核时长 180 分钟；
- 请在桌面建立以“学号\_题号”命名的成果文件夹，所有提交文件均放在该目录下。  
文件夹命名例如：201965110001\_J2\_1；
- 将各任务源文件分别保存在考生文件夹下的对应目录，运行结果分别截图。

## (4) 评分细则

表 1 任务一评分细则 (60 分)

| 序号 | 评分项        | 分值   | 评分细则                                                                                                                                              |
|----|------------|------|---------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 酒店名称数据处理正确 | 10 分 | 和参考答案一致满分；<br>酒店名称数据处理正确，得 6 分；<br>有 <code>print(data.head())</code> 或者 <code>print(data.head(5))</code> 代码，得 2 分；<br>T1 文件夹中有 T1-1.jpg 文件，得 2 分。 |
| 2  | 地址数据处理正确   | 10 分 | 和参考答案一致满分；<br>地址数据处理正确，得 6 分；<br>有 <code>print(data.head())</code> 或者 <code>print(data.head(5))</code> 代码，得 2 分；<br>T1 文件夹中有 T1-2.jpg 文件，得 2 分。   |
| 3  | 正确统计平均评分   | 12 分 | 和参考答案一致满分；<br>利用分组函数正确统计平均评分，得 4 分；<br>保留 2 位小数，得 2 分；<br>按要求输出，得 2 分；<br>T1 文件夹中有各城市酒店_平均评分.csv 文件，得 2 分；<br>T1 文件夹中有 T1-3.jpg 文件，得 2 分。         |
| 4  | 正确统计平均价格   | 12 分 | 和参考答案一致满分；<br>利用分组函数正确统计平均值，得 4 分；<br>保留 2 位小数，得 2 分；<br>按要求输出，得 2 分；<br>T1 文件夹中有各城市酒店_平均价格.csv 文件，得 2 分；<br>T1 文件夹中有 T1-4.jpg 文件，得 2 分。          |
| 5  | 清洗后数据正确保存  | 8 分  | T1 文件夹中有酒店数据_清洗后.csv 文件，得 8 分。                                                                                                                    |
| 6  | 源代码命名正确并保存 | 4 分  | T1 文件夹中有 T1.py 文件，得 4 分                                                                                                                           |
| 7  | 代码注释       | 4 分  | 代码中有注释，得 4 分                                                                                                                                      |

表 2 任务二评分细则（30 分）

| 序号 | 评分项                | 分值   | 评分细则                                                                                                                          |
|----|--------------------|------|-------------------------------------------------------------------------------------------------------------------------------|
| 1  | 柱状图展示正确、源代码命名正确并保存 | 10 分 | 运行结果和参考答案一致满分；<br>标题正确，得 1 分；<br>没有中文乱码，得 1 分；<br>柱状图展示正确，得 6 分；<br>T2 文件夹中有 T2-1.py 文件，得 1 分；<br>T2 文件夹中有 T2-1.jpg 文件，得 1 分。 |
| 2  | 折线图展示正确、源代码命名正确并保存 | 10 分 | 运行结果和参考答案一致满分；<br>标题正确，得 1 分；<br>没有中文乱码，得 1 分；<br>折线图展示正确，得 6 分；<br>T2 文件夹中有 T2-2.py 文件，得 1 分；<br>T2 文件夹中有 T2-2.jpg 文件，得 1 分。 |
| 3  | 饼图展示正确、源代码命名正确并保存  | 10 分 | 运行结果和参考答案一致满分；<br>标题正确，得 1 分；<br>没有中文乱码，得 1 分；<br>饼图展示正确，得 6 分；<br>T2 文件夹中有 T2-3.py 文件，得 1 分；<br>T2 文件夹中有 T2-3.jpg 文件，得 1 分。  |

表 3 职业素养评分细则（10 分）

| 序号 | 评分项  | 分值  | 评分细则                              |
|----|------|-----|-----------------------------------|
| 1  | 操作规范 | 3 分 | 答题过程中保持安静，不损坏机器                   |
| 2  | 文明考试 | 3 分 | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌；      |
| 3  | 工位整理 | 4 分 | 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |

## 11. 试题编号：J4-2-5：土流网数据处理与可视化

### （1）任务描述

土流网土地资讯频道提供及时、全面的土地市场行情与土地政策信息，包括农业用地、工商用地、住宅用地、土地发布、土地转让、土地出租、土地估价、服务网点、土地数据等信息。某农产品公司计划租赁 400 亩农地，建一个蔬菜厂，想要了解最新的土地数据行情以及政策法规信息，请帮助该公司完成土流网数据的采集、处理和展示的工作，具体任务如下：

#### 任务一：数据处理与分析（60 分）

使用 Python 的 pandas 读取“T1\_土流网数据.csv”文件，做如下处理：

- 从名称这列数据中提取地址、面积、出租形式三列数据，如“宿州埇桥区 113 亩果园出租”分别对应“宿州埇桥区”、“113 亩”、“出租”三个值，然后删除名称这

---

列数据，并输出数据的前五行，将代码和结果截图，命名为 T1-1.jpg，并保存到考生文件夹的 T1 目录。（20 分）

- 将土地类型中的数据进行拆分，有多个类型则只保留第一个。例如，“耕地|其它耕地”改为“耕地”，并输出数据的前五行，将代码和结果截图，命名为 T1-2.jpg，并保存到考生文件夹的 T1 目录。（8 分）
- 将处理之后的数据输出到“土流网数据\_清洗后.csv”，并保存到考生文件夹的 T1 目录。（8 分）
- 统计各种类型土地的数量，并输出结果，将代码和结果截图，命名为 T1-4.jpg，并保存到考生文件夹的 T1 目录。（8 分）
- 统计各出租形式的土地数量，并输出结果，将代码和结果截图，命名为 T1-5.jpg，并保存到考生文件夹的 T1 目录。（8 分）
- 将源代码文件命名为 T1.py，保存到考生文件夹的 T1 目录。（4 分）
- 代码注释。在代码中有关键代码的注释。（4 分）

## 任务二：数据可视化（30 分）

使用 Python 分析数据并使用 matplotlib 或 Tableau 绘制图表：

- 使用饼图展示不同类型土地的分布比例，标题为“不同类型土地分布情况”，要求每个扇区显示百分比和数据标签，将运行结果截图保存为 T2-1.jpg，源文件命名为 T2-1.py，保存到考生文件夹的 T2 目录。（15 分）
- 使用混合图显示出租年限前 5 的面积和出租年限，横坐标为土地面积前 5 的地址信息，用柱状图展示面积，折线图展示出租年限，折线图的样式设置为“o-m”。显示图例，标题为“面积 TOP5 土地信息”，要求将运行结果截图保存为 T2-2.jpg，源文件命名为 T2-2.py，保存到考生文件夹的 T2 目录。（15 分）

## 良好编程习惯和职业素养（10 分）

- 答题过程中保持安静。（3 分）
- 有问题举手示意，有礼貌。（3 分）
- 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。（4 分）

---

## (2) 实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。计算机安装有 Python、PyCharm 集成开发环境、Echarts、VSCode，Python 中安装 numpy、pandas、matplotlib、pyecharts 等模块，安装 Office 办公软件。

## (3) 考核时量及要求

- 考核时长 180 分钟；
- 请在桌面建立以“学号\_题号”命名的成果文件夹，所有提交文件均放在该目录下。

文件夹命名例如：201965110001\_J2\_1；

- 将各任务源文件分别保存在考生文件夹下的对应目录，运行结果分别截图。

## (4) 评分细则

表 1 任务一评分细则（60 分）

| 序号 | 评分项           | 分值   | 评分细则                                                                                                                                                                                                      |
|----|---------------|------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 名称数据处理正确      | 20 分 | 和参考答案一致满分；<br>正确提取地址数据，得 4 分；<br>正确提取面积数据，得 4 分；<br>正确提取出租形式数据，得 4 分；<br>正确删除名称列，得 4 分；<br>有 <code>print(data.head())</code> 或者 <code>print(data.head(5))</code> 代码，得 2 分；<br>T1 文件夹中有 T1-1.jpg 文件，得 2 分。 |
| 2  | 土地类型数据处理正确    | 8 分  | 和参考答案一致满分；<br>土地类型数据处理正确，得 4 分；<br>有 <code>print(data.head())</code> 或者 <code>print(data.head(5))</code> 代码，得 2 分；<br>T1 文件夹中有 T1-2.jpg 文件，得 2 分。                                                         |
| 3  | 清洗后数据正确保存     | 8 分  | T1 文件夹中有土流网数据_清洗后.csv 文件，得 2 分。                                                                                                                                                                           |
| 4  | 正确统计各类型土地数量   | 8 分  | 和参考答案一致满分；<br>利用分组函数正确统计土地类型数量，得 4 分；<br>按要求输出，得 2 分；<br>T1 文件夹中有 T1-4.jpg 文件，得 2 分。                                                                                                                      |
| 5  | 正确统计各出租形式土地数量 | 8 分  | 和参考答案一致满分；<br>利用分组函数正确统计各出租形式土地数量，得 4 分；<br>按要求输出，得 2 分；<br>T1 文件夹中有 T1-5.jpg 文件，得 2 分。                                                                                                                   |
| 6  | 源代码命名正确并保存    | 4 分  | T1 文件夹中有 T1.py 文件，得 4 分                                                                                                                                                                                   |
| 7  | 代码注释          | 4 分  | 代码中有注释，得 4 分                                                                                                                                                                                              |

表 2 任务二评分细则（30 分）

| 序号 | 评分项                | 分值   | 评分细则                                                                                                                                                |
|----|--------------------|------|-----------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 饼图展示正确、源代码命名正确并保存  | 15 分 | 运行结果和参考答案一致满分；<br>标题正确，得 2 分；<br>没有中文乱码，得 1 分；<br>饼图展示正确，得 8 分；<br>显示百分比和数据标签正确，得 2 分；<br>T2 文件夹中有 T2-1.py 文件，得 1 分；<br>T2 文件夹中有 T2-1.jpg 文件，得 1 分。 |
| 2  | 混合图展示正确、源代码命名正确并保存 | 15 分 | 运行结果和参考答案一致满分；<br>标题正确，得 2 分；<br>没有中文乱码，得 1 分；<br>柱状图和折线图展示正确，得 8 分；<br>显示图例正确，得 2 分；<br>T2 文件夹中有 T2-2.py 文件，得 1 分；<br>T2 文件夹中有 T2-2.jpg 文件，得 1 分。  |

表 3 职业素养评分细则（10 分）

| 序号 | 评分项  | 分值  | 评分细则                              |
|----|------|-----|-----------------------------------|
| 1  | 操作规范 | 3 分 | 答题过程中保持安静，不损坏机器                   |
| 2  | 文明考试 | 3 分 | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌；      |
| 3  | 工位整理 | 4 分 | 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |

### 项目 3：资讯服务行业数据处理与可视化

#### 12.试题编号：J4-3-1：北京 2020 年天气数据处理与可视化

##### (1) 任务描述

天气预报（测）或气象预报（测）是使用现代科学技术对未来某一地点地球大气层的状态进行预测，主要是使用收集大量的数据（气温、湿度、风向和风速、气压等等），然后使用目前对大气过程的认识（气象学）来确定未来空气变化，也是天气预报是气象工作为国民经济和国防建设服务的重要手段。现需对 2020 年北京市的天气数据进行分析，具体任务如下：

##### 任务一：数据处理与分析（60 分）

使用 Python 的 pandas 读取“T1\_北京 2020 年天气数据.csv”文件，完成如下数据处理和分析操作：

- 将气温这列数据拆分为最高温度和最低温度两列数据，删除℃单位，如“21℃/14℃”对应“最高温度：21”和“最低温度：14”，并转换成 int 类型数据，删除气温这列数据，输出数据的前五行，将代码和结果截图，命名为 T1-1.jpg，并保存到考生文件夹的 T1 目录。（20 分）
- 对天气状况和风力风向数据进行处理，天气状况如“晴/晴”保留第一个“晴”，风力风向如“南风 1-2 级/南风 1-2 级”保留第一个“南风 1-2 级”，输出数据的前五行，将代码和结果截图，命名为 T1-2.jpg，并保存到考生文件夹的 T1 目录。（12 分）
- 将日期转换成日期型数据，统计北京每个月的最高温度和最低温度，将结果输出到控制台和文件“每个月的最高温度和最低温度.csv”中，要求文件的列名为“月份”、

---

“最高温度”和“最低温度”，将代码和结果截图，命名为 T1-3.jpg，和文件一并保存到考生文件夹的 T1 目录。（16 分）

- 将处理之后的数据输出到“北京 2020 年天气数据\_清洗后.csv”，并保存到考生文件夹的 T1 目录。（4 分）
- 将源代码文件命名为 T1.py，保存到考生文件夹的 T1 目录。（4 分）
- 代码注释。在代码中有关键代码的注释。（4 分）

## 任务二：数据可视化（30 分）

使用 Python 分析数据并使用 matplotlib 或 Tableau 绘制图表：

- 使用折线图展示北京每个月的最大温度和最低温度，标题为“北京每个月的最大温度和最低温度”，显示图例，图例为“最高温度”和“最低温度”，将运行结果截图保存为 T2-1.jpg，源文件命名为 T2-1.py，保存到考生文件夹的 T2 目录。（15 分）
- 使用饼图展示北京市一年当中各类天气状况出现的天数，要求每个扇区显示百分比和数据标签，标题为“北京市一年天气状况分布情况”，将运行结果截图保存为 T2-2.jpg，源文件命名为 T2-2.py，保存到考生文件夹的 T2 目录。（15 分）

## 良好编程习惯和职业素养（10 分）

- 答题过程中保持安静。（3 分）
- 有问题举手示意，有礼貌。（3 分）
- 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。（4 分）

## （2）实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。

计算机安装有 Python、PyCharm 集成开发环境、Echarts、VSCode，Python 中安装 numpy、pandas、matplotlib、pyecharts 等模块，安装 Office 办公软件。

## （3）考核时量及要求

- 考核时长 180 分钟；
- 请在桌面建立以“学号\_题号”命名的成果文件夹，所有提交文件均放在该目录下。  
文件夹命名例如：201965110001\_J2\_1；

- 将各任务源文件分别保存在考生文件夹下的对应目录，运行结果分别截图。

#### (4) 评分细则

表 1 任务一评分细则（60 分）

| 序号 | 评分项             | 分值   | 评分细则                                                                                                                                                                                                                         |
|----|-----------------|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 气温数据处理正确        | 20 分 | 和参考答案一致满分；<br>气温拆分处理正确并新增两列数据，得 4 分；<br>最高高度和最低温度删除单位，得 4 分；<br>正确转换最高温度和最低温度数据类型，得 4 分；<br>正确删除气温列，得 4 分；<br>有 <code>print(data.head())</code> 或者 <code>print(data.head(5))</code> 代码，得 2 分；<br>T1 文件夹中有 T1-1.jpg 文件，得 2 分。 |
| 2  | 天气状况和风力风向数据处理正确 | 12 分 | 和参考答案一致满分；<br>天气状况数据处理正确，得 4 分；<br>风力风向数据处理正确，得 4 分；<br>有 <code>print(data.head())</code> 或者 <code>print(data.head(5))</code> 代码，得 2 分；<br>T1 文件夹中有 T1-2.jpg 文件，得 2 分。                                                       |
| 3  | 正确统计每个月温度       | 16 分 | 和参考答案一致满分；<br>正确转换日期列数据类型，得 4 分；<br>正确统计每个月的最高和最低温度，得 6 分；<br>按要求输出，得 2 分；<br>T1 文件夹中有北京 2020 年天气数据_清洗后.csv 文件，得 2 分；<br>T1 文件夹中有 T1-3.jpg 文件，得 2 分                                                                          |
| 4  | 清洗后数据正确保存       | 4 分  | T1 文件夹中有北京 2020 年天气数据_清洗后.csv 文件，得 4 分。                                                                                                                                                                                      |
| 5  | 源代码命名正确并保存      | 4 分  | T1 文件夹中有 T1.py 文件，得 4 分                                                                                                                                                                                                      |
| 6  | 代码注释            | 4 分  | 代码中有注释，得 4 分                                                                                                                                                                                                                 |

表 2 任务二评分细则（30 分）

| 序号 | 评分项                | 分值   | 评分细则                                                                                                                                                          |
|----|--------------------|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 折线图展示正确、源代码命名正确并保存 | 15 分 | 运行结果和参考答案一致满分；<br>标题正确，得 2 分；<br>没有中文乱码，得 1 分；<br>正确显示负号，得 2 分；<br>折线图展示正确，得 6 分；<br>显示图例，得 2 分；<br>T2 文件夹中有 T2-1.py 文件，得 1 分；<br>T2 文件夹中有 T2-1.jpg 文件，得 1 分。 |
| 2  | 饼图展示正确、源代码命名正确并保存  | 15 分 | 运行结果和参考答案一致满分；<br>标题正确，得 2 分；<br>没有中文乱码，得 1 分；<br>饼图展示正确，得 10 分；<br>T2 文件夹中有 T2-2.py 文件，得 1 分；<br>T2 文件夹中有 T2-2.jpg 文件，得 1 分。                                 |

表 3 职业素养评分细则（10 分）

| 序号 | 评分项  | 分值  | 评分细则                              |
|----|------|-----|-----------------------------------|
| 1  | 操作规范 | 3 分 | 答题过程中保持安静，不损坏机器                   |
| 2  | 文明考试 | 3 分 | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌；      |
| 3  | 工位整理 | 4 分 | 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |

### 13.试题编号：J4-3-2：糗事百科数据处理与可视化

#### （1）任务描述

糗事百科是以糗友真实糗事为主题的笑话网站，话题轻松休闲，在年轻人中十分流行。

请以此网站为目标，完成以下任务：

#### 任务一：数据处理与分析（60 分）

使用 Python 的 pandas 读取相关文件，进行如下处理：

- 使用 pandas 的 read\_csv 函数读取“糗事百科数据.csv”文件中的“用户名,性别,年龄,点赞数,评论数”五列数据，输出数据的前五行，将代码和结果截图，命名为 T1-1.jpg，并保存到考生文件夹的 T1 目录。（12 分）
- 筛选出点赞数前 10 名的用户信息，将结果输出到控制台和文件“点赞 TOP10 数据.csv”中，将代码和结果截图，命名为 T1-2.jpg，和文件一并保存到考生文件夹的 T1 目录。（12 分）

- 
- 统计男女用户的平均年龄，要求保留 2 位小数，并输出结果，将代码和结果截图，命名为 T1-3.jpg，和文件一并保存到考生文件夹的 T1 目录。（12 分）
  - 统计男女用户人数信息(注意去重)，将结果输出到控制台和文件“男女用户人数.csv”中，要求文件的列名为“性别”和“人数”，将代码和结果截图，命名为 T1-4.jpg，和文件一并保存到考生文件夹的 T1 目录。（16 分）
  - 将源代码文件命名为 T1.py，保存到考生文件夹的 T1 目录。（4 分）。
  - 代码注释。在代码中有关键代码的注释。（4 分）

## 任务二：数据可视化（30 分）

使用 Python 分析数据并使用 matplotlib 或 Tableau 绘制图表：

- 使用饼图展示各个年龄段用户的分布比例，标题为“各年龄段用户分布情况”，要求每个扇区显示百分比，将运行结果截图保存为 T2-1.jpg，源文件命名为 T2-1.py，保存到考生文件夹的 T2 目录。（15 分）
- 使用柱状图展示点赞数前五的用户，标题为“点赞数 TOP5 用户”，将运行结果截图保存为 T2-2.jpg，源文件命名为 T2-2.py，保存到考生文件夹的 T2 目录。（15 分）

## 良好编程习惯和职业素养（10 分）

- 答题过程中保持安静。（3 分）
- 有问题举手示意，有礼貌。（3 分）
- 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。（4 分）

## （2）实施条件

安排在教学实训机房，计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。

计算机安装有 Python、PyCharm 集成开发环境、Echarts、VSCode，Python 中安装 numpy、pandas、matplotlib、pyecharts 等模块，安装 Office 办公软件。

## （3）考核时量及要求

- 考核时长 180 分钟；
- 请在桌面建立以“学号\_题号”命名的成果文件夹，所有提交文件均放在该目录下。  
文件夹命名例如：201965110001\_J2\_1；

- 将各任务源文件分别保存在考生文件夹下的对应目录，运行结果分别截图。

#### (4) 评分细则

表 1 任务一评分细则（60 分）

| 序号 | 评分项               | 分值   | 评分细则                                                                                                                                            |
|----|-------------------|------|-------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 正确获取文件内容          | 12 分 | 和参考答案一致满分；<br>正确读取文件内容，得 8 分；<br>有 <code>print(data.head())</code> 或者 <code>print(data.head(5))</code> 代码，得 2 分；<br>T1 文件夹中有 T1-1.jpg 文件，得 2 分。 |
| 2  | 点赞数前 10 的用户信息统计正确 | 12 分 | 和参考答案一致满分；<br>利用排序函数正确筛选点赞数前十的用户信息，得 6 分；<br>正确输出，得 2 分；<br>T1 文件夹中有点赞 TOP10 数据.csv 文件，得 2 分；<br>T1 文件夹中有 T1-2.jpg 文件，得 2 分。                    |
| 3  | 正确统计男女平均年龄        | 12 分 | 和参考答案一致满分；<br>利用分组函数正确统计男女平均年龄，得 6 分；<br>保留两位小数，得 2 分；<br>按要求输出，得 2 分；<br>T1 文件夹中有 T1-3.jpg 文件，得 2 分。                                           |
| 4  | 正确统计用户人数并保存       | 16 分 | 和参考答案一致满分；<br>利用分组函数正确统计男女人数信息，得 8 分；<br>按要求输出结果，得 2 分；<br>T1 文件夹中有男女用户人数.csv 文件，得 4 分；<br>T1 文件夹中有 T1-4.jpg 文件，得 2 分。                          |
| 5  | 源代码命名正确并保存        | 4 分  | T1 文件夹中有 T1.py 文件，得 4 分                                                                                                                         |
| 6  | 代码注释              | 4 分  | 代码中有注释，得 4 分                                                                                                                                    |

表 2 任务二评分细则（30 分）

| 序号 | 评分项                | 分值   | 评分细则                                                                                                                                                |
|----|--------------------|------|-----------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 饼图展示正确、源文件命名正确并保存  | 15 分 | 运行结果和参考答案一致满分；<br>标题正确，得 2 分；<br>没有中文乱码，得 1 分；<br>饼图展示正确，得 8 分；<br>显示百分比和数据标签正确，得 2 分；<br>T2 文件夹中有 T2-1.py 文件，得 1 分；<br>T2 文件夹中有 T2-1.jpg 文件，得 1 分。 |
| 2  | 柱状图展示正确、源文件命名正确并保存 | 15 分 | 运行结果和参考答案一致满分；<br>标题正确，得 2 分；<br>没有中文乱码，得 1 分；<br>柱状图展示正确，得 10 分；<br>T2 文件夹中有 T2-2.py 文件，得 1 分；<br>T2 文件夹中有 T2-2.jpg 文件，得 1 分。                      |

表 3 职业素养评分细则（10 分）

| 序号 | 评分项  | 分值  | 评分细则                              |
|----|------|-----|-----------------------------------|
| 1  | 操作规范 | 3 分 | 答题过程中保持安静，不损坏机器                   |
| 2  | 文明考试 | 3 分 | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌；      |
| 3  | 工位整理 | 4 分 | 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |

## 14.试题编号：J4-3-3：胡润富豪榜数据处理与可视化

### （1）任务描述

胡润富豪榜一般指胡润百富榜。胡润百富是追踪记录中国企业家群体变化的权威机构，是由出生于欧洲卢森堡的英国注册会计师胡润先生于 1999 年创立的。是中国推出第一份财富排行榜，也是现在国内财经榜单里影响最大的一个榜单。现需要了解最新的富豪排行，具体任务如下：

#### 任务一：数据处理与分析（60 分）

使用 Python 的 pandas 读取“T1\_胡润富豪榜.csv”文件，完成如下数据处理和分析操作：

- 将财富这一列数据处理成纯数字，如“147 亿”处理成“147”，并输出数据的前五行，将代码和结果截图，命名为 T1-1.jpg，并保存到考生文件夹的 T1 目录。（16 分）
- 统计出财富榜上男女的人数，输出结果，将代码和结果截图，命名为 T1-2.jpg，并保存到考生文件夹的 T1 目录；（16 分）

- 
- 按行业类别分组计算财富总和,输出财富总和前 10 的行业类别及对应的财富数据,将代码和结果截图,命名为 T1-3.jpg,并保存到考生文件夹的 T1 目录。(20 分)
  - 将源代码文件命名为 T1.py,保存到考生文件夹的 T1 目录。(4 分)
  - 代码注释。在代码中有关键代码的注释。(4 分)

## 任务二：数据可视化 (30 分)

根据题目需求读取数据,使用 echarts 或 pyecharts 完成如下题目:

- 使用折线图绘制前 10 名财富总和最多的行业类别,标题为“财富总和 TOP10 行业”,要求将运行结果截图保存为 T2-1.jpg,源文件命名为 T2-1.html,保存到考生文件夹的 T2 目录。(10 分)
- 使用饼图展示富豪榜中男女人数占比情况,标题为“富豪榜男女比例”,要求将运行结果截图保存为 T2-2.jpg,源文件命名为 T2-2.html,保存到考生文件夹的 T2 目录。(10 分)
- 按财富进行排序,使用横向条形图绘制前 15 名最有钱的人,标题为“全球 TOP15 富豪榜”,要求将运行结果截图保存为 T2-3.jpg,源文件命名为 T2-3.html,保存到考生文件夹的 T2 目录。(10 分)

## 良好编程习惯和职业素养 (10 分)

- 答题过程中保持安静。(3 分)
- 有问题举手示意,有礼貌。(3 分)
- 答题完毕离场时整理好工位,不关闭计算机,并按要求显示答案保存界面。(4 分)

## (2) 实施条件

安排在教学实训机房,计算机最好满足 CPU i5、内存 8G、硬盘 500G 以上配置要求。

计算机安装有 Python、PyCharm 集成开发环境、Echarts、VSCode,Python 中安装 numpy、pandas、matplotlib、pyecharts 等模块,安装 Office 办公软件。

## (3) 考核时量及要求

- 考核时长 180 分钟;
- 请在桌面建立以“学号\_题号”命名的成果文件夹,所有提交文件均放在该目录下。

文件夹命名例如：201965110001\_J2\_1；

- 将各任务源文件分别保存在考生文件夹下的对应目录，运行结果分别截图。

#### (4) 评分细则

表 1 任务一评分细则（60 分）

| 序号 | 评分项        | 分值   | 评分细则                                                                                                                                             |
|----|------------|------|--------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 财富数据处理正确   | 16 分 | 和参考答案一致满分；<br>财富数据处理正确，得 12 分；<br>有 <code>print(data.head())</code> 或者 <code>print(data.head(5))</code> 代码，得 2 分；<br>T1 文件夹中有 T1-1.jpg 文件，得 2 分。 |
| 2  | 正确统计男女人数   | 16 分 | 和参考答案一致满分；<br>利用分组函数正确统计男女人数，得 12 分；<br>正确输出，得 2 分；<br>T1 文件夹中有 T1-2.jpg 文件，得 2 分。                                                               |
| 3  | 正确计算财富总和   | 20 分 | 和参考答案一致满分；<br>正确转换财富数据类型，得 4 分；<br>利用分组函数正确计算总财富，得 8 分；<br>正确输出财富总和前 10 的信息，得 6 分；<br>T1 文件夹中有 T1-3.jpg 文件，得 2 分。                                |
| 4  | 源代码命名正确并保存 | 4 分  | T1 文件夹中有 T1.py 文件，得 4 分                                                                                                                          |
| 5  | 代码注释       | 4 分  | 代码中有注释，得 4 分                                                                                                                                     |

表 2 任务二评分细则（30 分）

| 序号 | 评分项                | 分值   | 评分细则                                                                                                           |
|----|--------------------|------|----------------------------------------------------------------------------------------------------------------|
| 1  | 折线图展示正确、源代码命名正确并保存 | 10 分 | 运行结果和参考答案一致满分；<br>标题正确，得 2 分；<br>折线图展示正确，得 6 分；<br>T2 文件夹中有 T2-1.html 文件，得 1 分；<br>T2 文件夹中有 T2-1.jpg 文件，得 1 分。 |
| 2  | 饼图展示正确、源代码命名正确并保存  | 10 分 | 运行结果和参考答案一致满分；<br>标题正确，得 2 分；<br>饼图展示正确，得 6 分；<br>T2 文件夹中有 T2-2.html 文件，得 1 分；<br>T2 文件夹中有 T2-2.jpg 文件，得 1 分。  |
| 3  | 条形图展示正确、源代码命名正确并保存 | 10 分 | 运行结果和参考答案一致满分；<br>标题正确，得 2 分；<br>条形图展示正确，得 6 分；<br>T2 文件夹中有 T2-3.html 文件，得 1 分；                                |

|  |  |  |                             |
|--|--|--|-----------------------------|
|  |  |  | T2 文件夹中有 T2-3.jpg 文件，得 1 分。 |
|--|--|--|-----------------------------|

表 3 职业素养评分细则（10 分）

| 序号 | 评分项  | 分值  | 评分细则                              |
|----|------|-----|-----------------------------------|
| 1  | 操作规范 | 3 分 | 答题过程中保持安静，不损坏机器                   |
| 2  | 文明考试 | 3 分 | 尊重考场工作人员，遵守考场纪律，有问题举手示意，有礼貌；      |
| 3  | 工位整理 | 4 分 | 答题完毕离场时整理好工位，不关闭计算机，并按要求显示答案保存界面。 |