张家界航空工业职业技术学院
ZHANGJIAJIE INSTITUTE OF AERONAUTICAL ENGINEERING

人工智能技术应用

专业技能考核题库

专 业 名 称:	人工智能技术应用
专 业 代 码:	510209
适 用 年 级:	2021级
所 属 学 院:	信息技术学院
专业负责人:	邓卫红
制(修)订时间:	2022年4月

目录

张家界航空工业职业技术学院
人工智能技术应用专业技能考核题库

一、专业基本技能模块

模块一 程序设计

1.试题编号：1-1 任务实现1

（1）任务描述

任务一：从键盘读入三个不相同的数，把这三个数由小到大输出(20分)。

要求：使用分支结构语句实现。

任务二：使用循环语句打印出如下图案（30分）。

*

***

*****

****

要求：使用循环结构语句实现。

任务三：从键盘输入x，根据以下情形求y的值（30分）：

y=0;（当x<=0时）

y=2x+1;（当0<x<5时）

y=x2-1;（当x>=5时）

要求：使用多分支条件语句实现。

（2）作品提交要求见本模块附录1

（3）实施条件要求见本模块附录2

（4）评价标准见本模块附录3

2.试题编号：1-2 任务实现2

（1）任务描述

任务一：输入一个百分制分数，输出其对应的五级制成绩，包括：优（90分以上，含90）、良（80-90分，含80，不含90）、中（70-80分，含70不含80）、及格（60-70分，含60不含70）、不及格（60分以下，不含60）(20分)

要求：使用多分支条件语句实现。

任务二：输出阶梯形式的9*9乘法口诀表，如图所示（30分）。

1*1=1								
1*2=2	2*2=4							
1*3=3	2*3=6	3*3=9						
1*4=4	2*4=8	3*4=12	4*4=16					
1*5=5	2*5=10	3*5=15	4*5=20	5*5=25				
1*6=6	2*6=12	3*6=18	4*6=24	5*6=30				
1*7=7	2*7=14	3*7=21	4*7=28	5*7=35	6*7=42	7*7=49		
1*8=8	2*8=16	3*8=24	4*8=32	5*8=40	6*8=48	7*8=56	8*8=64	
1*9=9	2*9=18	3*9=27	4*9=36	5*9=45	6*9=54	7*9=63	8*9=72	9*9=81

乘法口诀表

要求：使用循环结构语句实现。

任务三：输入某人的收入，计算个人应缴的税额。如图所示（30分）。

级数	全月应纳税所得额	适用税率%	速算扣除数（元）
1	不超过500元的	5	0
2	超过500元至2000元的部分	10	25
3	超过2000元至5000元的部分	15	125
4	超过5000元至20000元的部分	20	375
5	超过20000元至40000元的部分	25	1375
6	超过40000元至60000元的部分	30	3375
7	超过60000元至80000元的部分	35	6375
8	超过80000元至100000元的部分	40	10375
9	超过100000元的部分	45	15375

税率表

要求：使用多分支条件语句实现。

（2）作品提交要求见本模块附录1

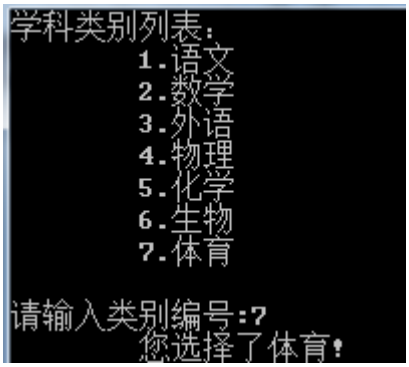
（3）实施条件要求见本模块附录2

（4）评价标准见本模块附录3

3.试题编号：1-3 任务实现3

（1）任务描述

任务一：请模拟实现一个课程菜单选择功能，如图所示。(20分)



要求：使用switch语句实现。

任务二：输入一个百分制的成绩t，将其转换成对应的等级然后输出，具体转换规则如下：90~100为A80~；89为B；70~79为C；60~69为D0~59为E。(30分)

要求：使用多分支条件语句实现。

任务三：请完成以下编程工作：①定义一个动物抽象类Animal，有动物名称，动物打招呼的方法。②定义它的两个子类Dog和Cat，该类继承动物类。③分别实现它们打招呼的方式。(30分)

要求：①使用抽象类和抽象方法。

②使用类的继承。

③在主函数(或主方法)中实例化对象，并让对象实现操作。

（2）作品提交要求见本模块附录1

（3）实施条件要求见本模块附录2

（4）评价标准见本模块附录3

4.试题编号：1-4 任务实现4

（1）任务描述

任务一：验证用户输入的数字是否在25-50范围内，如果输入错误或不在25-50范围内就要求用户重新输入。(20分)

要求：利用循环结构完成。

任务二：用户可以无限次的输入数字，请统计用户输入的数字中正数的个数，负数的个数，0的个数。直到用户输入999就结束程序，输出统计结果。(30分)

要求：使用循环结构语句实现。

任务三：求n的阶乘，如果输入的数不在范围之内则要求用户重新输入。(30分)

要求：使用循环结构语句实现。

（2）作品提交要求见本模块附录1

（3）实施条件要求见本模块附录2

（4）评价标准见本模块附录3

5.试题编号：1-5 任务实现5

（1）任务描述

任务一：编写程序实现：商店卖西瓜，20斤以上的每斤0.85元；重于15斤轻于等于20斤的，每斤0.90元；重于10斤轻于等于15斤的，每斤0.95元；重于5斤轻于等于10斤的，每斤1.00元；轻于或等于5斤的，每斤1.05元。输入西瓜的重量和顾客所付钱数，输出应付货款和应找钱数。(20分)

要求：使用分支结构语句实现。

任务二：已知xyz+yzz=532，其中x、y、z均为一位数，编写一个程序求出x、y、z分别代表什么数字。(30分)

要求：使用分支、循环结构语句实现。

任务三：从键盘输入一个整数N，打印出有N*2-1行的菱形。（30分）

```
   *
  ***
 *****
*******
 *****
  ***
   *
```

例如输入整数4，则屏幕输出如下菱形。

现要求输入整数为7，在屏幕中输出相应的菱形。要求：用循环结构语句实现。

（2）作品提交要求见本模块附录1

（3）实施条件要求见本模块附录2

（4）评价标准见本模块附录3

6.试题编号：1-6 任务实现6

（1）任务描述

任务一：有1、2、3三个数字，能组成哪些互不相同且无重复数字的三位数。(30分)

要求：使用循环结构语句实现。

任务二：输入10个学生的单科成绩，求出其中的最高分、最低分、平均分以及超过平均分的人数(30分)

要求：使用数组定义实现。

任务三：使用循环语句打印出如下图案。(30分)

****

*****

***

*

要求：使用循环结构语句实现。

（2）作品提交要求见本模块附录1

（3）实施条件要求见本模块附录2

（4）评价标准见本模块附录3

7.试题编号：1-7 任务实现7

（1）任务描述

任务一：根据如下要求计算机票优惠率，并输出。(20分)

输入：用户依次输入月份和需要订购机票的数量，分别保存到整数变量month和sum中。

计算规则如下：

航空公司规定在旅游的旺季7～9月份，如果订票数超过20张，票价优惠15%，20张以下，优惠5%；在旅游的淡季1～5月份、10月份、11月份，如果订票数超过20张，票价优惠30%，20张以下，优惠20%；其他情况一律优惠10%。

输出：根据输入月份和需要订购机票的数量，输出优惠率。

要求：使用分支结构实现上述程序功能。

任务二：使用冒泡排序法对数组中的整数按升序进行排序，如下所示：

原始数组：a[]={1,9,3,7,4,2,5,0,6,8}排序后：a[]={0,1,2,3,4,5,6,7,8,9}(30分)

要求：综合使用分支、循环结构语句实现，直接输出结果不计分。

任务三：输入一个年度，判断是否是闰年。例如，2000是闰年，1900不是闰年，1904是闰年。(30分)

要求：使用分支结构语句实现。

提示：以下两个条件，只要满足任意一个，即是闰年：①能整除4且不能整除100；②能整除400。

（2）作品提交要求见本模块附录1

（3）实施条件要求见本模块附录2

（4）评价标准见本模块附录3

8.试题编号：1-8 任务实现8

（1）任务描述

任务一：输出杨辉三角形，如下图所示：(20分)

*

***

*****

****

****

****

****

要求：使用循环结构语句实现，直接输出结果不计分。

任务二：编程实现判断一个字符串是否是“回文串”。所谓“回文串”是指一个字符串的第一位与最后一位相同，第二位与倒数第二位相同。例如：“159951”、“19891”是回文串，而“2011”不是。(30分)

要求：用带有一个输入参数的函数(或方法)实现，返回值类型为布尔类型。

任务三：任意输入十个数据，打印出改十个数据最大值、最小值。(30分)

要求：①定义一个大小为10的整形数组a；

②从键盘输入10个整数，放置到数组a中；

③输出数组a中的最大值、最小值。

（2）作品提交要求见本模块附录1

（3）实施条件要求见本模块附录2

（4）评价标准见本模块附录3

9.试题编号：1-9 任务实现9

（1）任务描述

任务一：输入n(n<100)个整数，找出其中最小的数，将它与最先输入的数交换后输出这些数。(20分)

要求：用数组解决任务。

任务二：从键盘输入三条边A，B，C的边长，请编程判断能否组成一个三角形。(30分)

要求：A,B,C<1000，如果三条边长A,B,C能组成三角形的话，输出YES，否则NO。

任务三：对于给定的一个字符串，统计其中数字字符出现的次数。(30分)

（2）作品提交要求见本模块附录1

（3）实施条件要求见本模块附录2

（4）评价标准见本模块附录3

10.试题编号：1-10 任务实现10

（1）任务描述

任务一：某运输队为超市运送暖瓶500箱，每箱装有6个暖瓶。已知每10个暖瓶的运费为5元，损坏一个不但不给运费还要赔10元，运后结算时，运输队共得1350元的运费。编程输出损坏暖瓶的个数。(20分)

要求：用循环语句实现

任务二：编写程序，从键盘接收一个只包含英文字母的字符串，对字符串中的字母进行大小写互转（大写字母转成小写，小写字母转成大写）。（30分）

例如键盘输入：abcABC输出：ABCabc

要求：使用循环和判断语句实现。

任务三：对于给定的一个字符串，统计其中数字字符出现的次数。（30分）

要求： 字符串只能由数字和字符组成。

（2）作品提交要求见本模块附录1

（3）实施条件要求见本模块附录2

（4）评价标准见本模块附录3

程序设计模块附录

附录1作品提交

①请建立以“考生号_题号”命名的成果文件夹，所有提交文件均放在该目录下。例如：144115040001_T1_1；

②分别将每个任务的代码以成员函数的形式封装到类中，并且在main函数中调用该成员函数；

③在成果文件夹中创建三个文件夹task1、task2、task3，将三个任务的源代码、编译后的文件及对应成员函数的程序流程图截图分别保存至相应文件夹；

④将成果文件夹压缩打包，按照要求上传至服务器。

⑤考核时间为180分钟。

附录2实施条件

表1考点提供的主要设备及软件表

序号	场地、设备、软件名称	规格/技术参数、用途	备注
1	人工智能实训机房	测试场地	保证参考人员有足够间距
2	计算机	CPU酷睿i5以上，内存4G以上，win7/win10/linux操作系统	用于软件开发和软件部署，每人一台
3	Pycharm2018.2以上、IntelliJ IDEA 2018.2以上、Eclipse4.7或以上	软件开发	参考人员自选一种开发工具
4	MSDN或者JDK帮助文档中文版	帮助文档	参考人员可以使用帮助文档

附录3评价标准

评分标准一：实操文档（10分）

表2 实操文档评分细则表

序号	评分项	分值	评分细则
1	实操文档有无	2分	有实操文档得分，无实操文档扣2分。
2	文档任务截图	4分	有操作过程截图得分，无操作过程截图扣4分。
3	文档任务截图标注	4分	有文档任务截图标注说明和画框得分，无标注和画框扣4分。

评分标准二：依据题的任务，完成任务功能（80分）

表3 项目功能评分细则表

序号	评分项	分值	评分细则
1	任务实现	80分	试题按任务分值评分;未按要求提交正确格式的源文件，扣5分;程序中出现了没有使用的变量扣1分；程序中出现了无用的循环、分支、循序结构扣1分，扣完为止。

评分标准三：职业素质（10分）

表4 职业素质评分细则表

序号	评分项	分值	评分细则
1	代码书写格式规范	3分	代码缩进不规范扣1分、方法划分不规范扣1分、语句结构不规范扣1分（如一行编写两个语句）、使用空行不规范扣1分，扣完为止。
2	注释规范	2分	整个项目没有注释扣2分、有注释，但注释不规范扣1分，扣完为止。
3	类名、变量名、方法名命名规范	5分	命名规范，为满分。类名、变量名或方法名命名不规范或没有实际意义的每个扣1分，扣完为止。

模块二 数据库设计与开发

1.试题编号：2-1《教务管理系统》项目教材订购管理模块

（1）任务描述

《教材订购管理》模块的 E-R图如图2.1.1 所示，逻辑数据模型如图2.1.2所示，物理数据模型如图 2.1.3 所示，数据表字段名定义见表2.1.1。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：

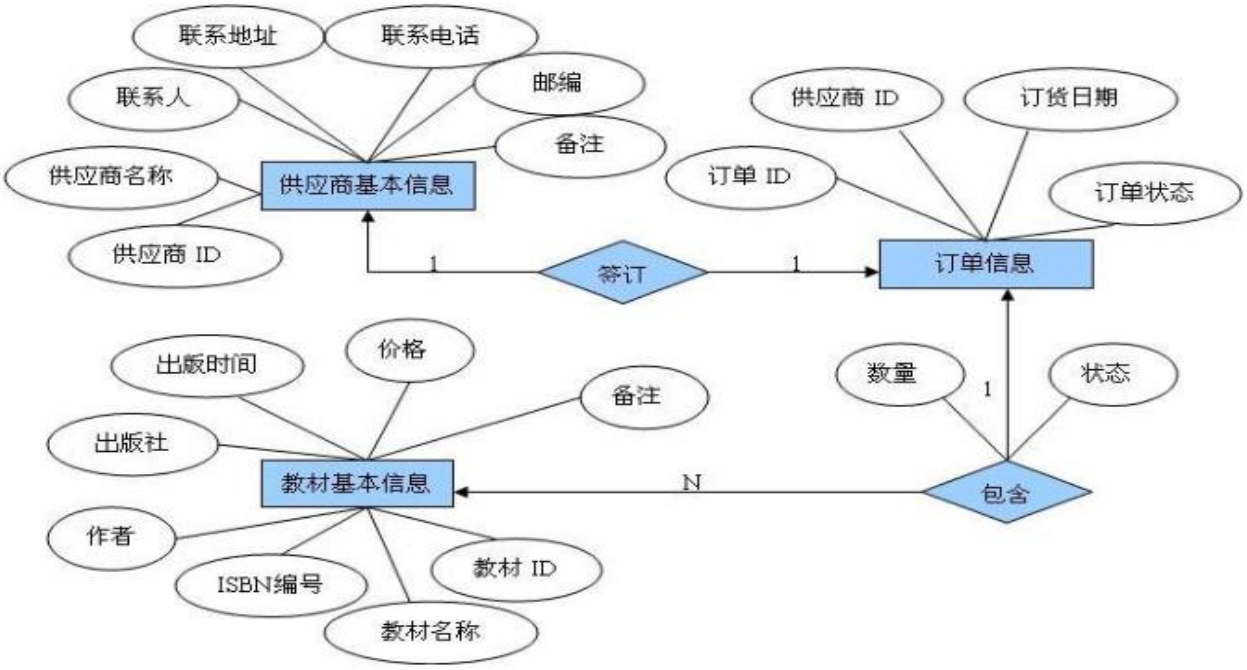


图2.1.1E-R图

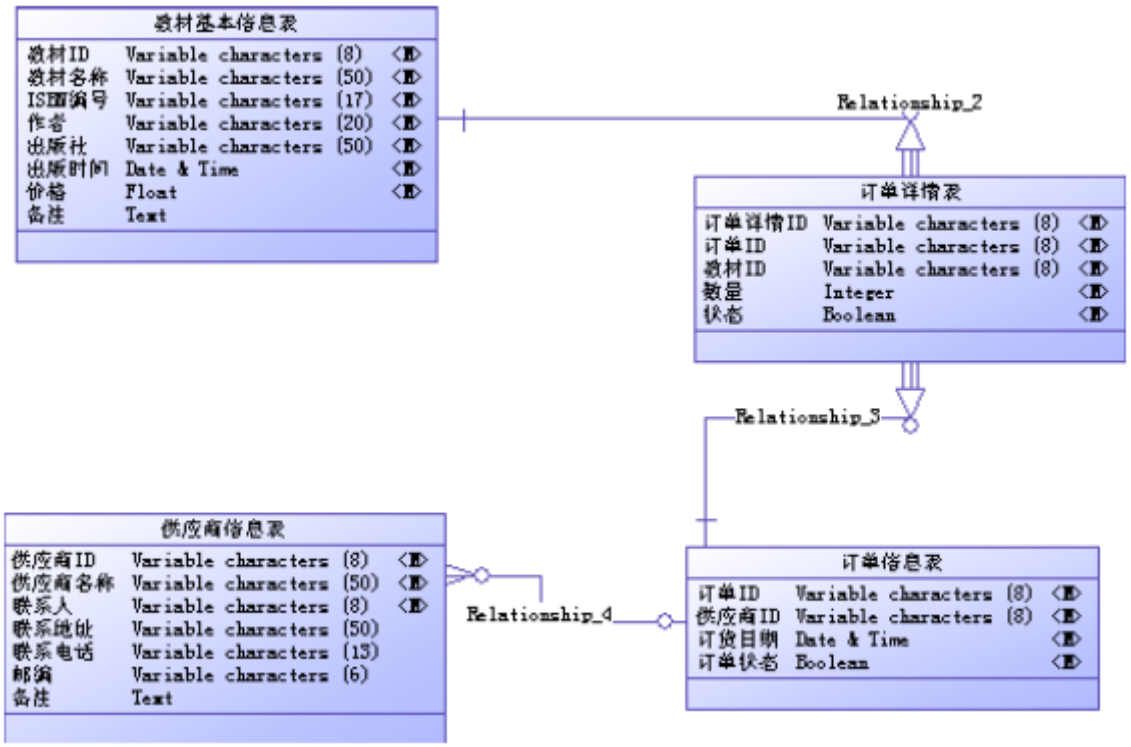


图2.1.2逻辑数据模型图

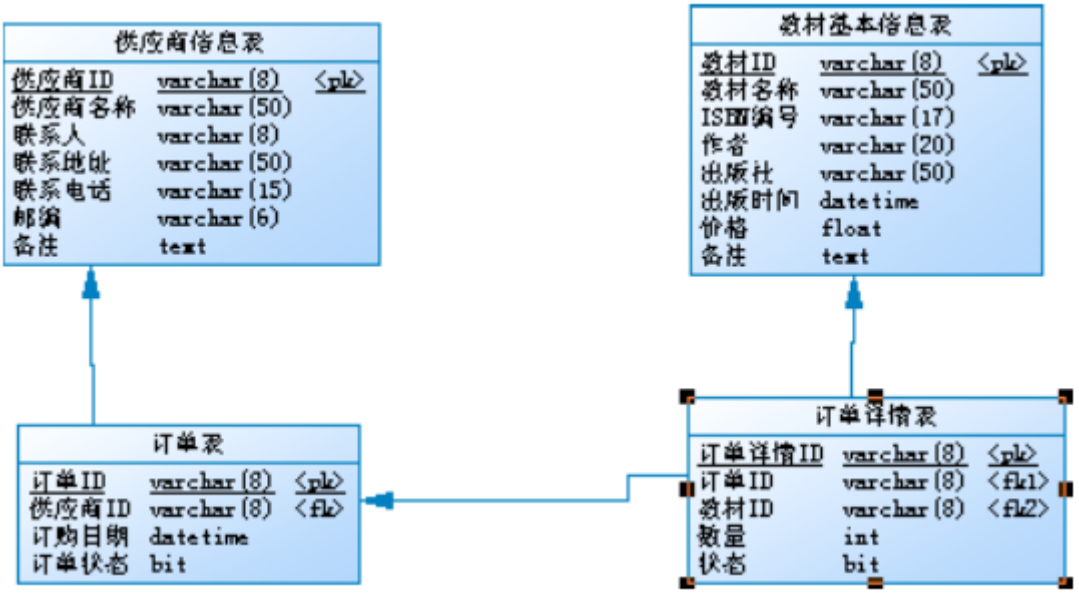


图2.1.3物理数据原型图

表2.1.1字段名定义表

字段名	字段说明	字段名	字段说明
book_id	教材 id	supplier_name	供应商名称
book_name	教材名称	supplier_people	联系人
book_isbn	教材 ISBN 编号	supplier_address	联系地址
book_author	作者	supplier_phone	联系电话
book_publisher	出版社	supplier_postcode	邮编
book_price	价格	supplier_remark	备注
book_rkm	备注	orderdet_id	订单详情 id
order_id	订单 id	orderdet_status	订单详情状态
order_datetime	订购时间	book_datetime	出版时间
order_status	订单状态	orderdet_num	数量
supplier_id	供应商 id		

任务一：创建数据库（10 分）

创建数据库HNIUEAM。

任务二：创建数据表（25 分）

根据图 2.1.2 和表 2.1.1，创建数据表 T_Supplier、 T_BookInfo、 T_Order。

任务三：创建数据表间的关系及约束（15 分）

根据物理数据原型，创建数据关系表。

任务四：数据操作（25 分）

用SQL语句完成如下操作：

①.向T_Supplier表插入数据：“BC0001,windows程序设计,0257-9413,刘立,电子工业出版社代理商,2010-11-10,42,无”；

②.查询出供应商名称为“电子工业出版社代理商”的订单编号及订单状态；

③.查询教材名称为“windows程序设计”的订购日期；

④.创建视图查询供应商名为“电子工业出版社代理商”所订购的教材的详细信息(包括名称，ISBN编号，作者，出版社，出版时间，价格，数量)；

⑤.创建存储过程，当订单详情表中相应订单的状态为“1”时,修改订单表的订单状态为“1”。

（2）作品提交要求见本模块附录1

（3）实施条件要求见本模块附录2

（4）评价标准见本模块附录3

2.试题编号：2-2《图书管理信息系统》 项目

（1）任务描述

《图书管理信息系统》中借书管理子模块的 E-R 图如图 2.2.1 所示，逻辑数据模型如图2.2.2所示，物理数据模型如图2.2.3所示，数据表字段名定义见表2.2.1。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：

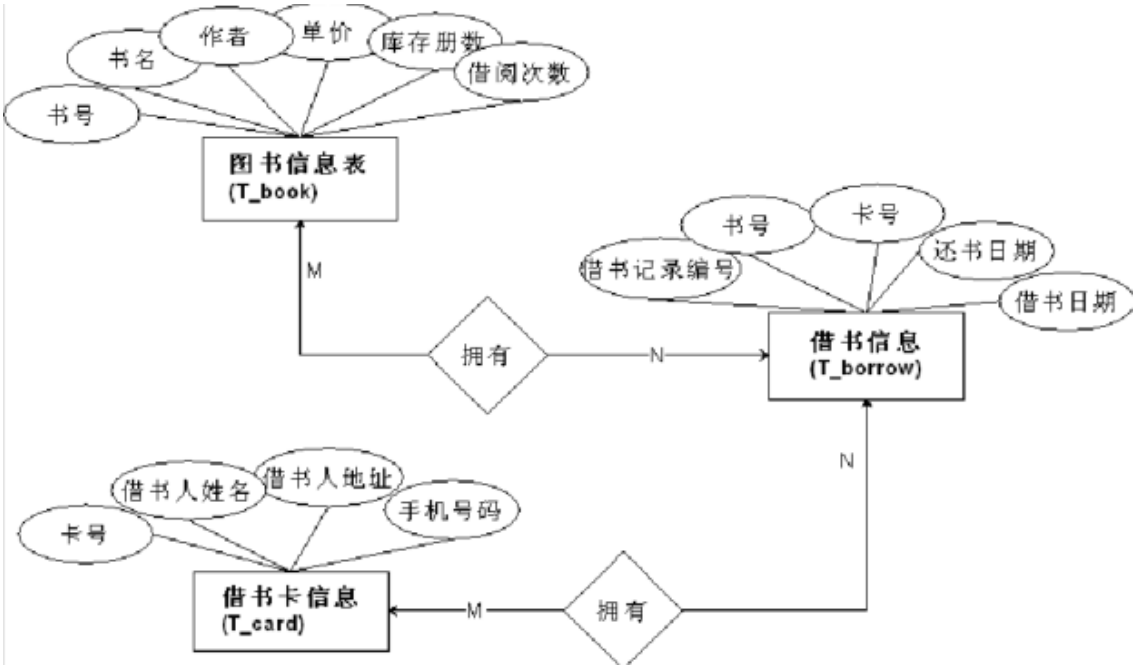


图2.2.1E-R图

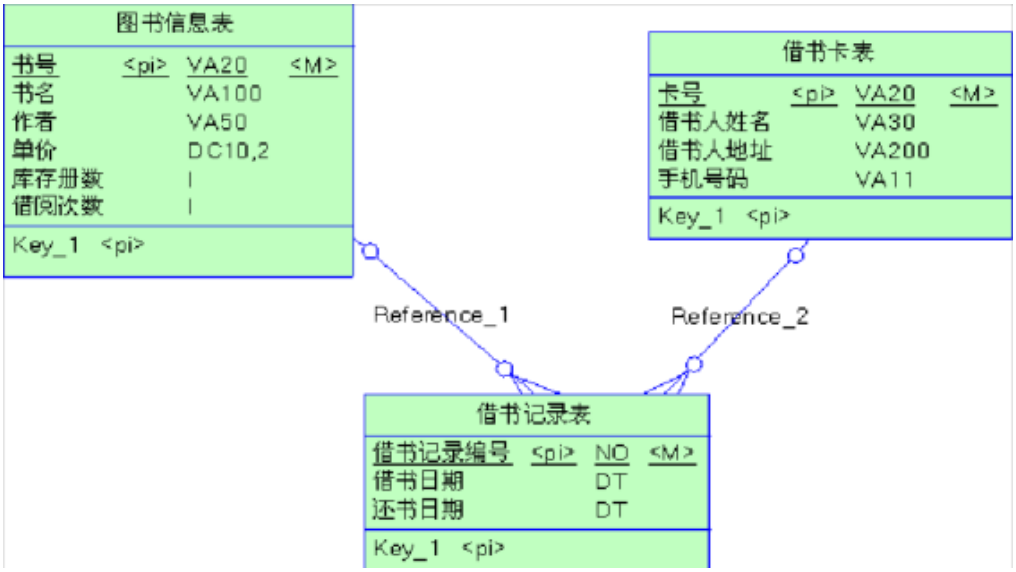


图2.2.2逻辑数据模型图

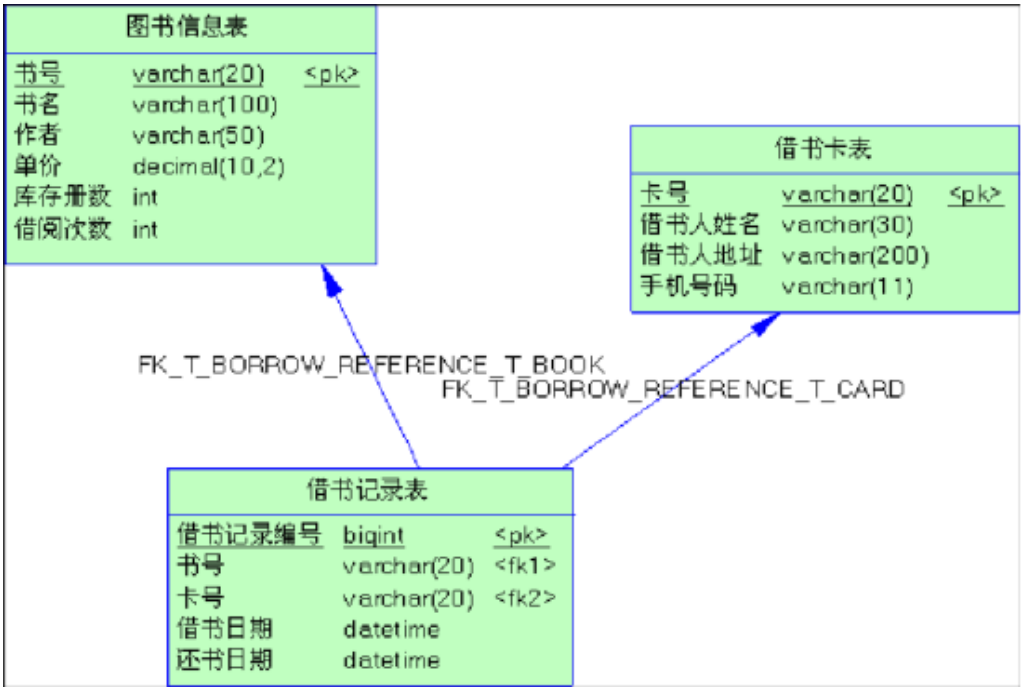


图2.2.3物理数据原型图

表2.2.1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。

任务二：创建数据表（25 分）

根据图2.2.2和表2.2.1，创建数据表T_card、T_book、T_borrow。

任务三：创建数据表间的关系及约束（15 分）

根据物理数据原型，创建数据关系表。

任务四：数据操作（25 分）

用SQL语句完成如下操作：

①.向每个表插入3条测试数据；

②.将“李”姓作者的所有图书单价下调10%；

③.查询出日期在2010-10-31至2010-11-31之间借出的图书信息；

④.查询出手机号为“135”开头的所有借书人姓名；

⑤.创建视图查询库存数量小于10册的图书信息；

（2）作品提交要求见本模块附录1

（3）实施条件要求见本模块附录2

（4）评价标准见本模块附录3

3.试题编号：2-3《学生管理信息系统》项目

（1）任务描述

《学生管理信息系统》中成绩管理子模块的E-R 图如图2.3.1所示，逻辑数据模型如图2.3.2所示， 物理数据模型如图2.3.3所示，数据表字段名定义见表2.3.1。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：

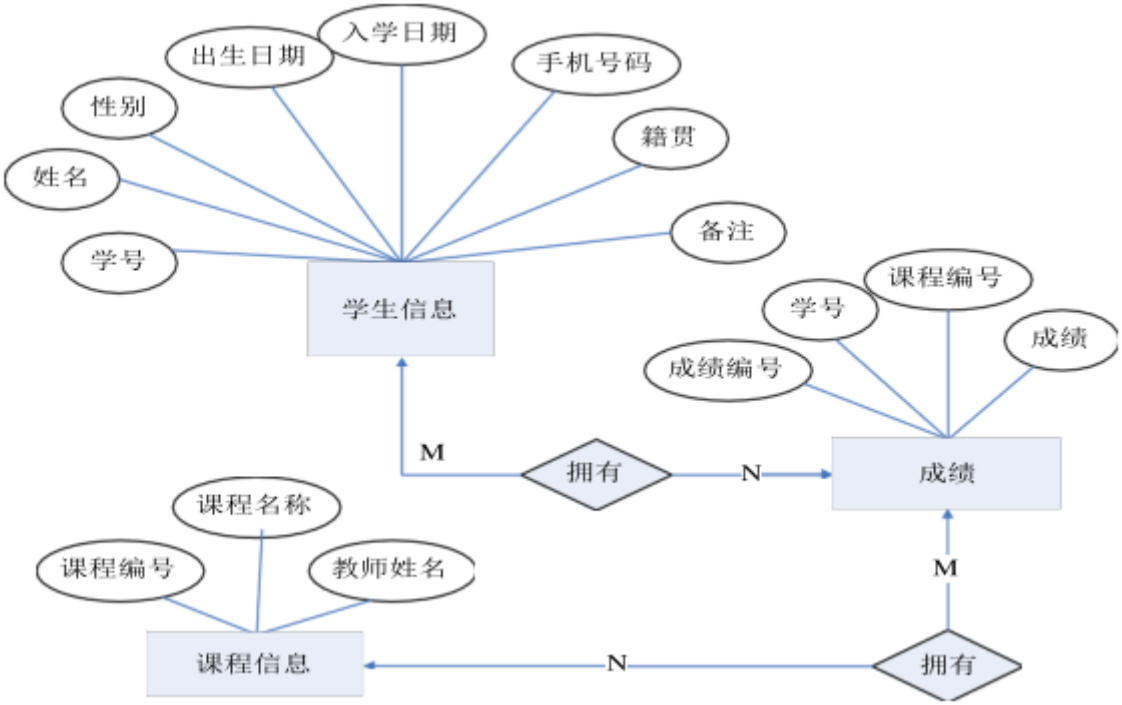


图2.3.1E-R图

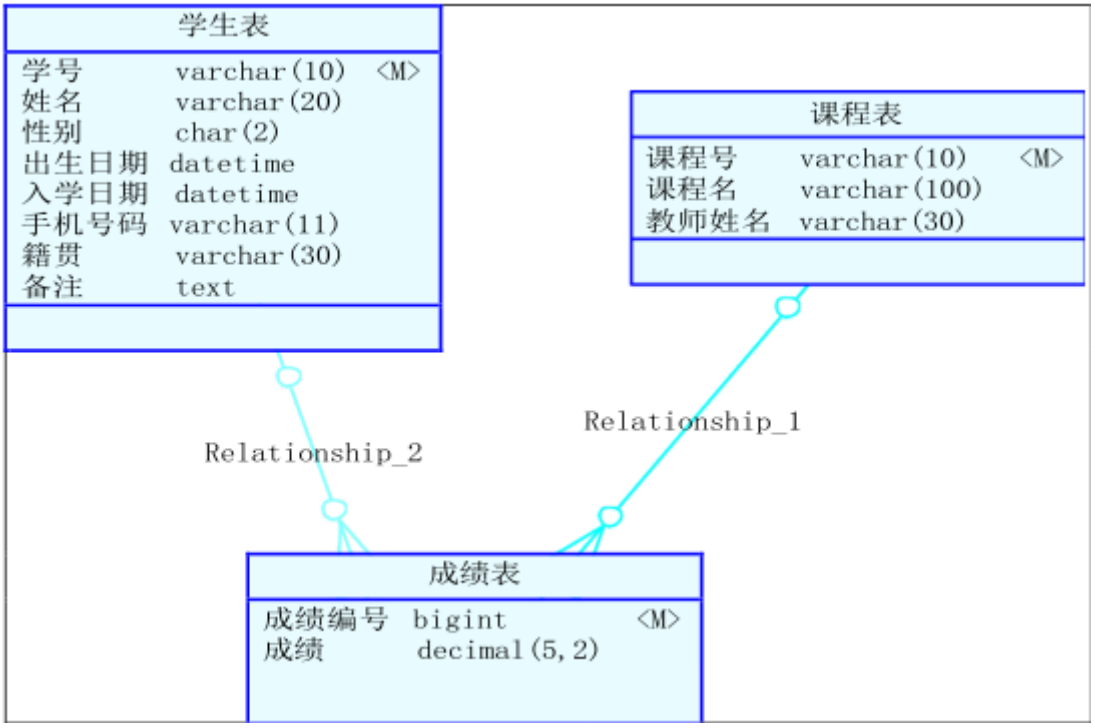


图2.3.2逻辑数据模型图

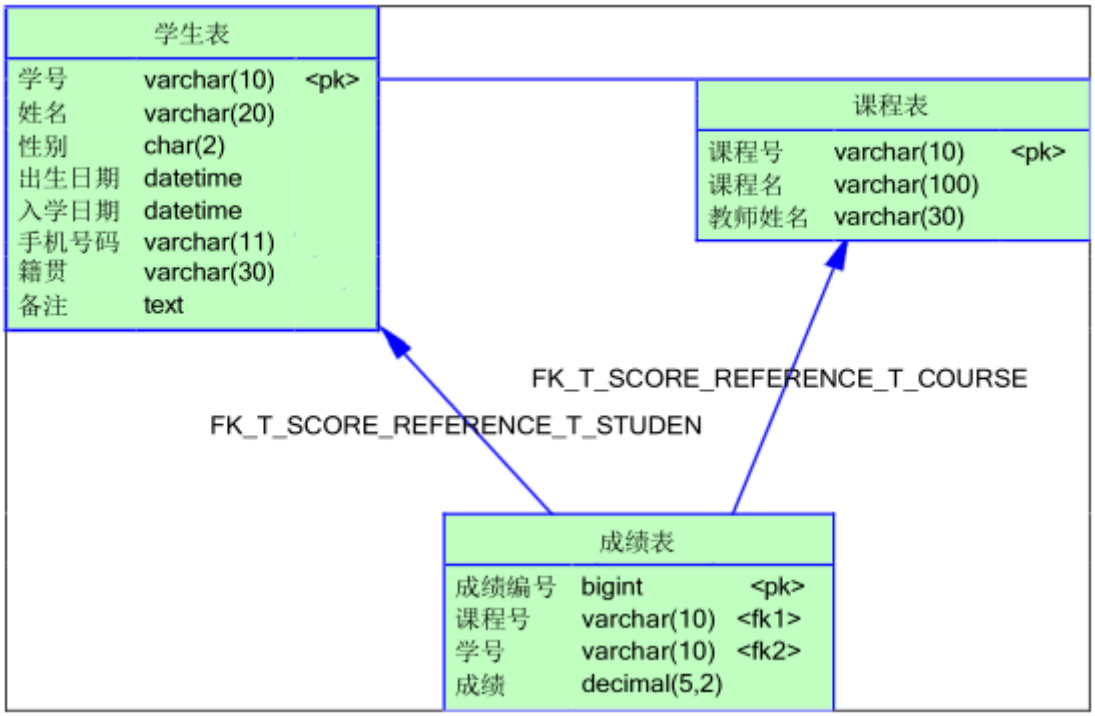


图2.3.3物理数据模型图

表2.3.1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 StudentDB。

任务二：创建数据表（25 分）

根据图2.3.2 和表2.3.1，创建数据表T_student、T_course、T_score。

任务三：创建数据表间的关系及约束（15分）

根据物理数据原型，创建数据关系。

任务四：数据操作（25 分）

用SQL语句完成如下操作：

①.向每个表插入3条测试数据；

②.删除所有选修“日语”的同学的选课记录；

③.查询出“数据库原理” 这门课的最高成绩；

④.查询出所有选修了“数据库原理”课程的学生学号、姓名和籍贯；

⑤.创建视图，查询指定课程名称的平均成绩。

（2）作品提交要求见本模块附录1

（3）实施条件要求见本模块附录2

（4）评价标准见本模块附录3

4.试题编号：2-4《人力资源管理系统》项目

（1）任务描述

《人力资源管理系统》中人员管理子模块的 E-R 图如图2.4.1 所示，逻辑数据模型如图2.4.2所示，物理数据模型如图2.4.3所示，数据表字段名定义见表2.4.1。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：

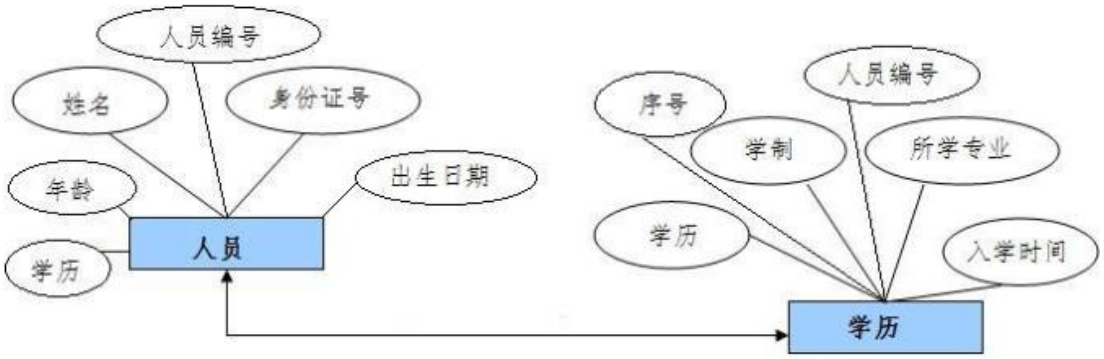


图2.4.1E-R图

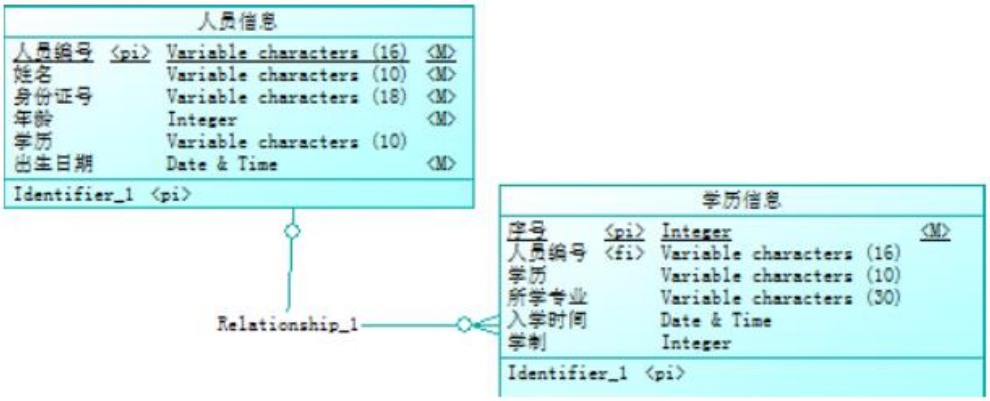


图2.4.2逻辑数据模型图

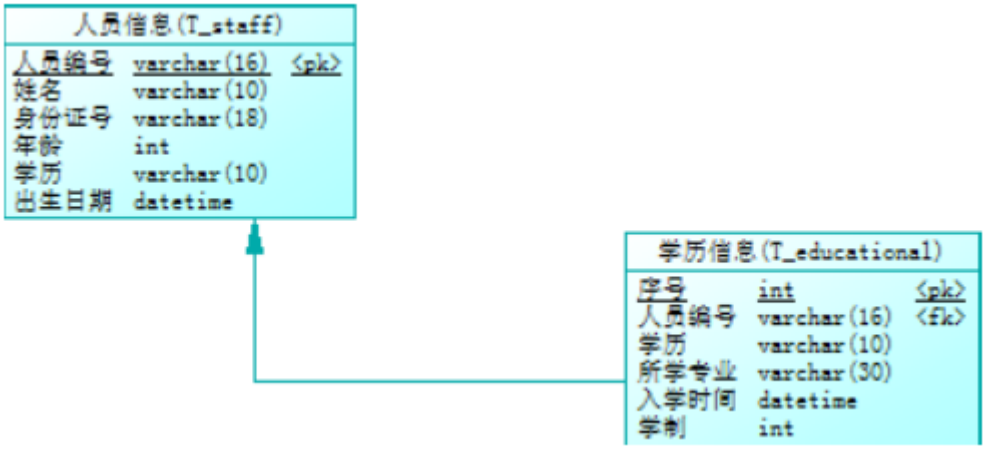


图2.4.3物理数据模型图

表2.4.1字段名定义表

字段名	字段说明	字段名	字段说明
staff_no	人员编号	id	序号(自动增长)
name	姓名	degree	学历
ic_card	身份证号	major	所学专业
age	年龄	reg_time	入学时间
bithday	出生日期	length_of_schooling	学制

任务一：创建数据库（10 分）

创建数据库ResourcesDB。

任务二：创建数据表（25 分）

根据图2.4.2 和表 2.4.1，创建数据表T_staff、T_educational。

任务三：创建数据表间的关系及约束（15 分）

①.为表设置主键，主键命名为“pk_<表名>_<主键标识>”；

②.根据逻辑数据模型，创建数据表之间的关系，关系命名为“fk_<表名>_<主表名>_<外键标识>”；

任务四：数据操作（25 分）

用SQL语句完成如下操作：

①.向每个表插入2条测试数据；

②.查询出T_staff表中大于平均年龄的人员名单；

③.查询出入学时间在 2015-9-1 之后的所有人员名单；

④.查询出学习“大数据技术”专业的所有人员名单；

⑤.创建存储过程，根据入学时间和学制计算每个人的毕业年份数。

（2）作品提交要求见本模块附录1

（3）实施条件要求见本模块附录2

（4）评价标准见本模块附录3

5.试题编号：2-5《员工工资管理》项目

（1）任务描述

《员工工资管理》的 E-R图如图2.5.1所示，逻辑数据模型如图2.5.2所示，物理数据模型如图 2.5.3所示，数据表字段名定义见表2.5.1 。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：

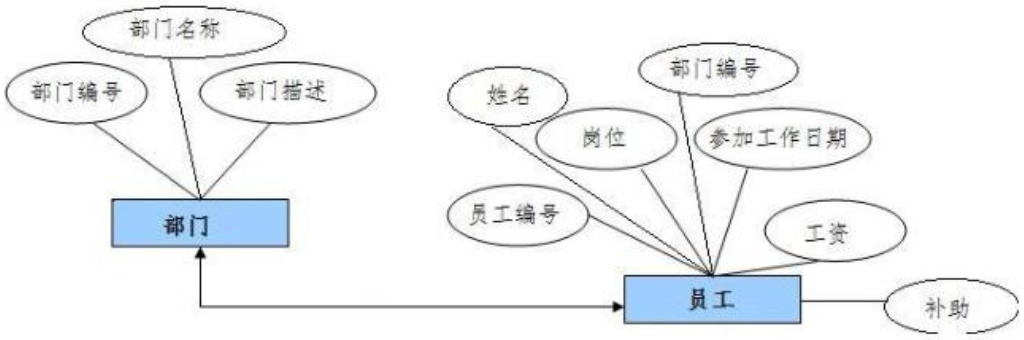


图2.5.1E-R图

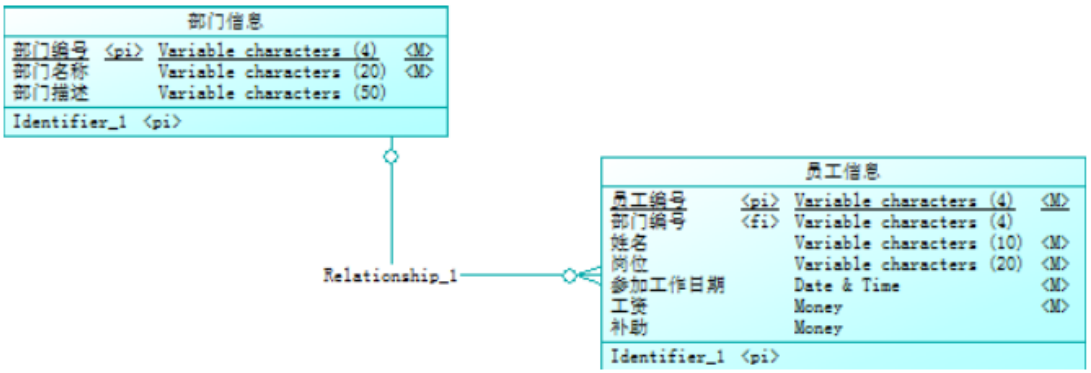


图2.5.2逻辑数据模型图

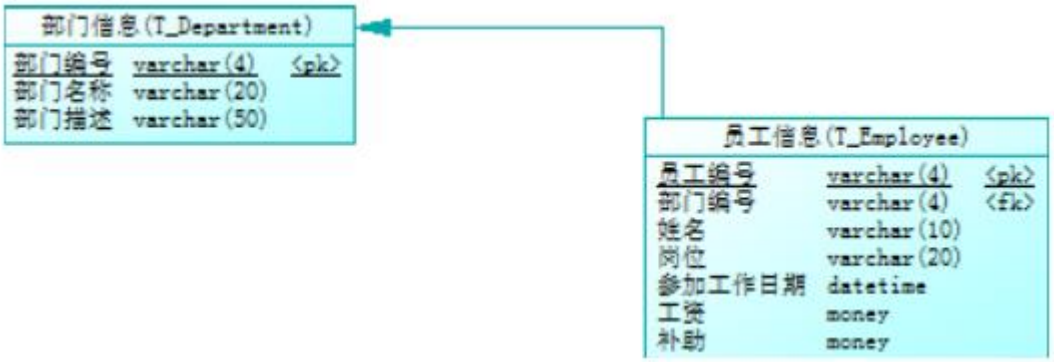


图2.5.3物理数据模型图

表2.5.1字段名定义表

字段名	字段说明	字段名	字段说明
dep_no	部门编号	post	岗位
dep_name	部门名称	work_time	参加工作日期
dep_desc	部门描述	salary	工资
emp_no	员工编号	bonus	补助
name	姓名		

任务一：创建数据库（10 分）

创建数据库SalaryDB。

任务二：创建数据表（25 分）

根据图2.24.2 和表2.24.1，创建数据表T_Department、T_Employee。

任务三：创建数据表间的关系及约束（15 分）

①.创建主键（两个表均设置）；

②.根据逻辑数据模型，创建数据表之间的关系，关系命名为“fk_<表名>_<主表名>_<外键标识>”；

任务四：数据操作（25 分）

用SQL语句完成如下操作：

①.向每个表插入2条测试数据；

②.查询出所有已有的岗位，要求取出重复项；

③.查询出每个部门每种岗位的平均工资和最高工资。

④.创建视图，显示所有没有补助的员工的姓名；

⑤.创建存储过程，显示平均工资低于3500的部门编号、平均工资、最高工资，要求以平均工资升序排序。

（2）作品提交要求见本模块附录1

（3）实施条件要求见本模块附录2

（4）评价标准见本模块附录3

6.试题编号：2-6《自学考试网》 项目

（1）任务描述

《自学考试网》的E-R 图如图2.6.1所示，逻辑数据模型如图2.6.2所示，物理数据模型如图2.6.3所示，数据表字段名定义见表2.6.1。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：

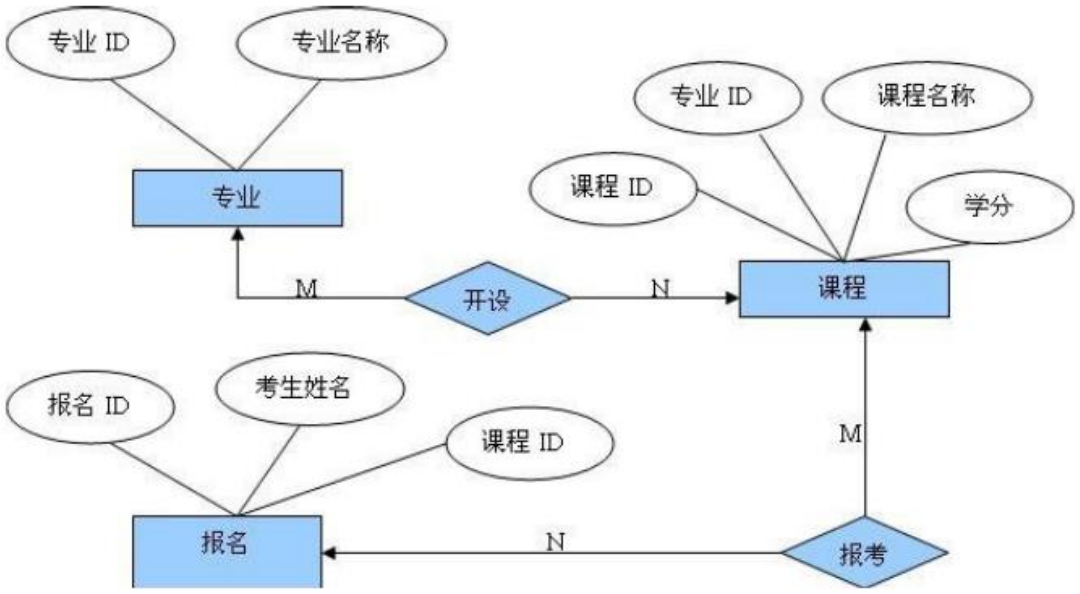


图2.6.1E-R图

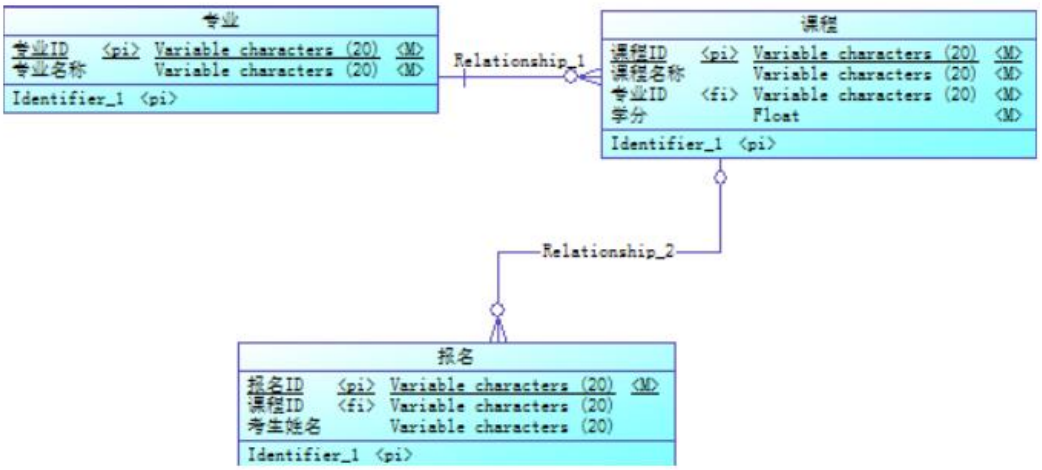


图2.6.2逻辑数据模型图

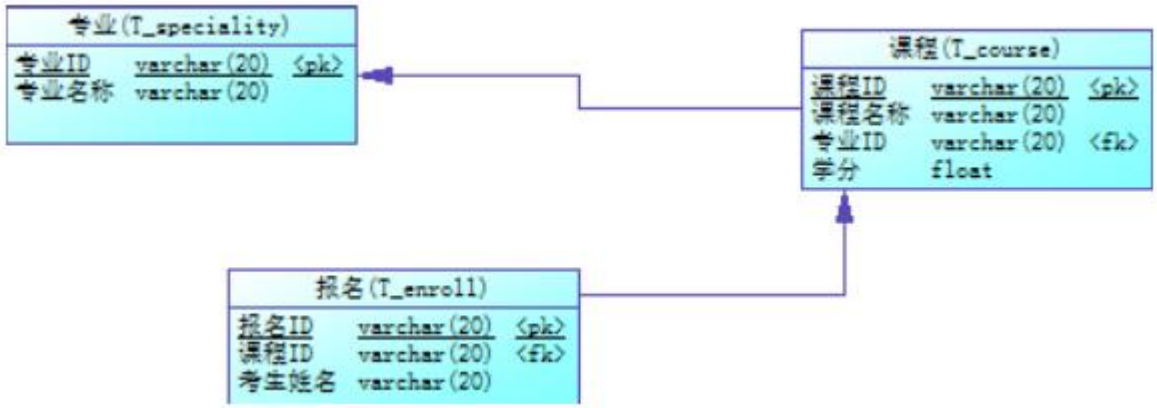


图2.6.3物理数据模型图

表2.6.1字段名定义表

字段名	字段说明	字段名	字段说明
id<pk>	专业 ID	mark	课程学分
name	专业名称	id<pk>	报名 ID
id<pk>	课程 ID	course_id	课程 ID
specialityid	专业 ID	name	考生姓名
name	课程名称		

任务一：创建数据库（10 分）

创建数据库SelfStudy。

任务二：创建数据表（25 分）

根据图2.6.2和表2.6.1，创建数据表T_speciality、T_course、T_enroll。

任务三：创建数据表间的关系及约束（15 分）

①.创建主键（三个表均设置）；

②.创建数据表之间的关系，关系命名为“fk_<表名>_<主表名>_<外键标识>”；

任务四：数据操作（25 分）

利用数据管理工具在表中插入以下数据， 用作测试。

表2.6.2 T_speciality表测试数据

id	name
001	大数据技术
002	计算机网络技术
003	人工智能技术应用

表2.6.3 T_course表测试数据

id	speciality_id	name	mark
001	001	Python程序设计	3
002	001	网页设计	3
003	001	爬虫应用技术与开发	3

表2.6.4 T_enroll表测试数据

id	course_id	name
001	001	王明
002	002	王明
003	003	王明

用SQL语句完成如下操作：

①.在T_course表插入数据：“004，001，高等数学，3”；

②.查询“大数据技术”专业开设的课程；

③.查询“大数据技术”专业有哪些考生报名；

④.查询出报考课程为“网页制作”的考生；

⑤.创建可查询考生姓名，报考课程名称的视图；

⑥.创建存储过程，查询报考某门课程（以课程名称为参数）的考生。

（2）作品提交要求见本模块附录1

（3）实施条件要求见本模块附录2

（4）评价标准见本模块附录3

7.试题编号：2-7《图书管理信息系统》项目

（1）任务描述

《图书管理信息系统》的 E-R 图如图2.7.1 所示，逻辑数据模型如图 2.7.2 所示，物理数据模型如图2.7.3 所示，数据表字段名定义见表2.7.1。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：

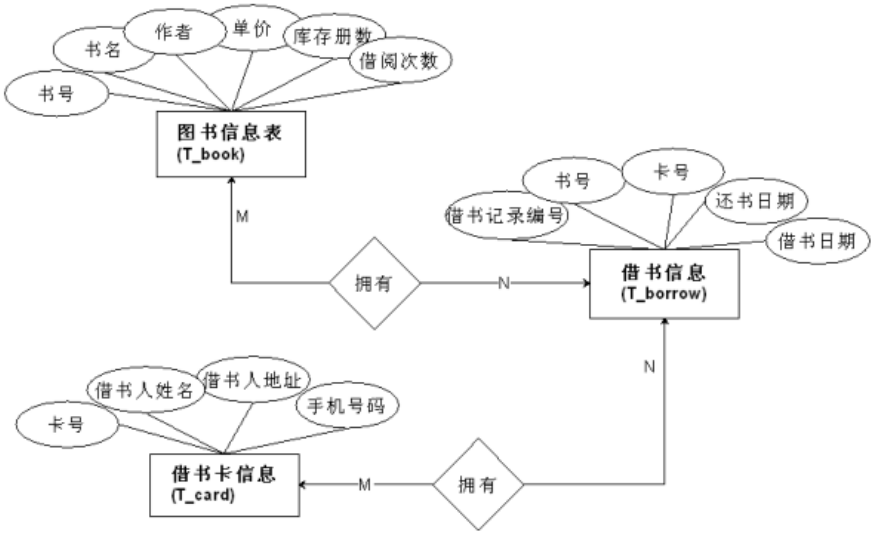


图2.7.1E-R图

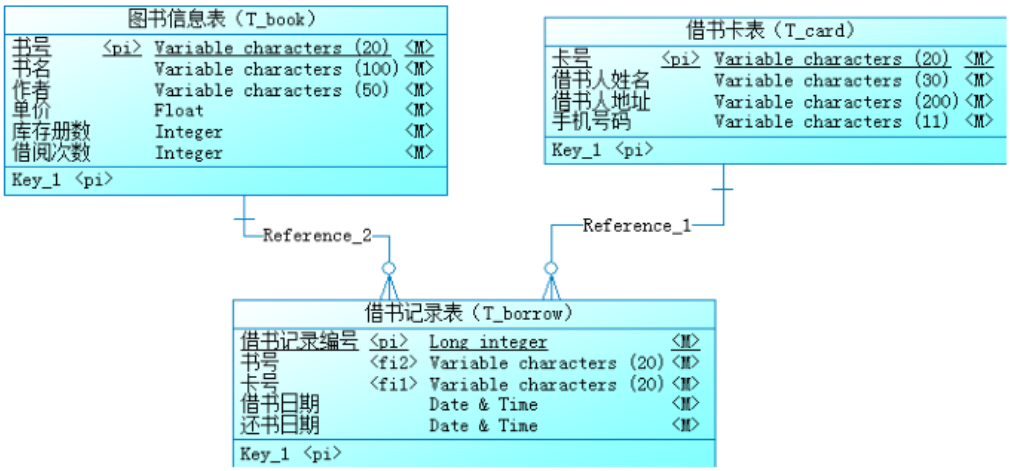


图2.7.2逻辑数据模型图

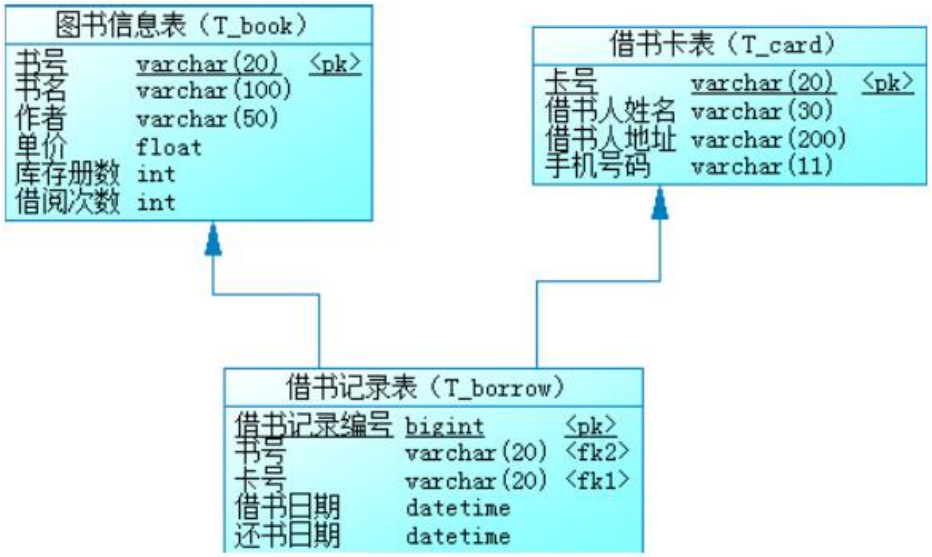


图2.7.3物理数据模型图

表2.7.1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。

任务二：创建数据表（25 分）

根据图2.7.2 和表 2.7.1，创建数据表T_card、T_book、T_borrow。

任务三：创建数据表间的关系及约束（15 分）

根据物理数据原型，创建数据关系。

任务四：数据操作（25 分）

用SQL语句查询出如下数据：

①.在T_book 表中插入数据：“9787302245339，Access数据库技术与应用，陈世红，27.20，50”；

②.查询出日期为2010-10-31以后借出的图书信息；

③.查询出没有还书的借书人姓名；

④.创建视图查询借书人的姓名，手机号码和地址；

⑤.查询出库存数量小于5册的图书信息；

（2）作品提交要求见本模块附录1

（3）实施条件要求见本模块附录2

（4）评价标准见本模块附录3

8.试题编号：2-8《银行信贷管理系统》项目

（1）任务描述

《银行信贷管理系统》的E-R 图如图2.8.1 所示，逻辑数据模型、物理数据模型如图2.8.2和图2.8.3所示。 数据表字段名定义见表 2.8.1。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：

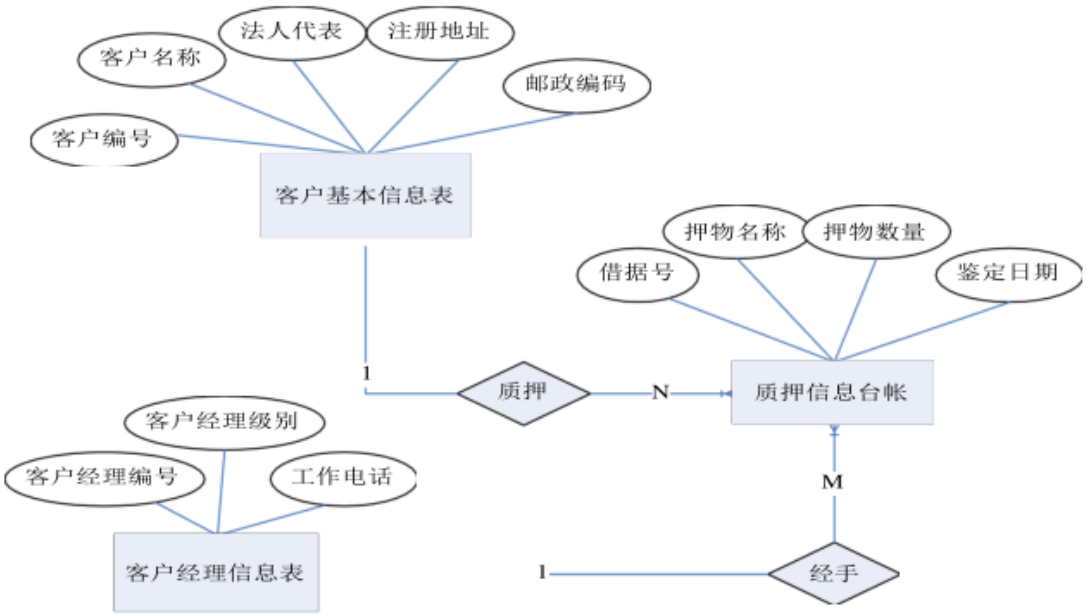


图2.8.1E-R图

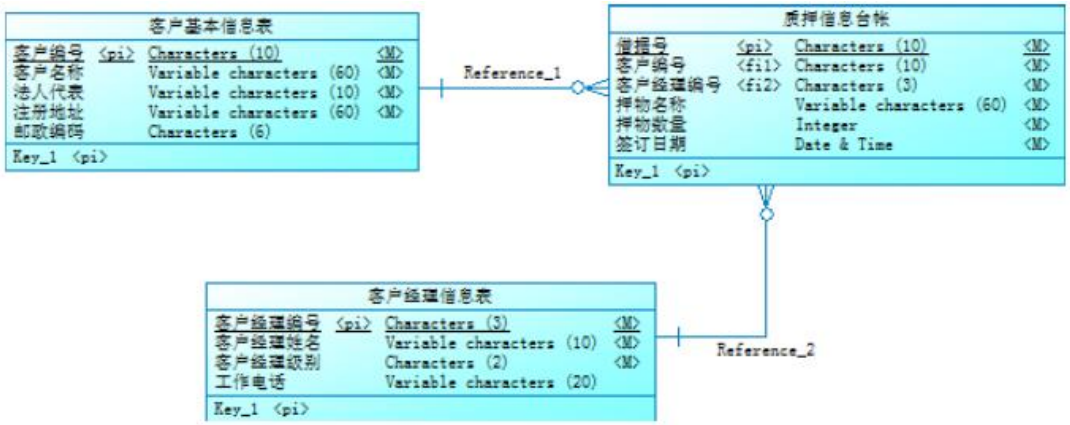


图2.8.2逻辑数据模型图

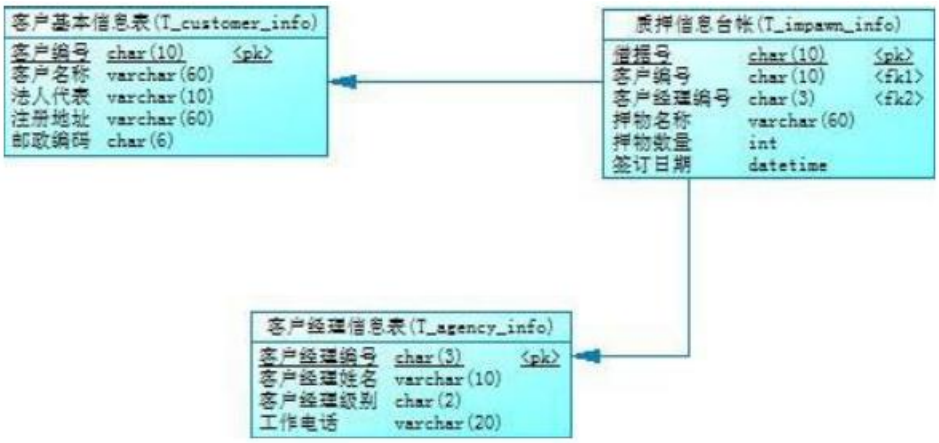


图2.8.3物理数据模型图

表2.8.1字段名定义表

字段名	字段说明	字段名	字段说明
agency_id	客户经理编号	reg_address	注册地址
agency_name	客户经理姓名	post_code	邮政编码
agency_level	客户经理级别	borrow_id	借据号
cust_id	客户编号	pawn_goods_name	押物名称
cust_name	客户名称	pawn_goods_num	押物数量
legal_name	法人代表	contract_date	签订日期
agency_phone	工作电话		

任务一：创建数据库（10 分）

创建数据库 BankCreditLoanDB。

任务二：创建数据表（25 分）

根据图2.4.2和表2.4.1，创建数据表T_customer_info、T_impawn_info、T_agency_info。

任务三：创建数据表间的关系及约束（15 分）

①.为表设置主键，主键命名为“pk_<表名>_<主键标识>”；

②.根据逻辑数据模型，创建数据表之间的关系，关系命名为“fk_<表名>_<主表名>_<外键标识>”；

③.要求邮政编码由6位数字组成。

任务四：数据操作（25 分）

用SQL语句执行以下操作：

①.分别向三个表中插入一条测试数据，其中客户经理编号为“001”；

②.查询“XX 公司”质押的物品及数量（说明：“XX 公司”为插入测试数据中的公司名称）；

③.统计每个客户经理所经手的质押业务数，查询结果集应包含字段：客户经理姓名、质押业务数；

④.创建存储过程P_customer_info，删除指定客户编号的客户基本信息，同时也删除该客户在质押信息台帐中的所有记录。

（2）作品提交要求见本模块附录1

（3）实施条件要求见本模块附录2

（4）评价标准见本模块附录3

9.试题编号：2-9《建设工程监管信息系统》项目系统权限管理模块

（1）任务描述

《系统权限管理》模块的 E-R 图如图 2.9.1 所示，逻辑数据模型如图 2.9.2 所示，物理数据模型如图 2.9.3 所示，数据表字段名定义见表 2.9.1。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：

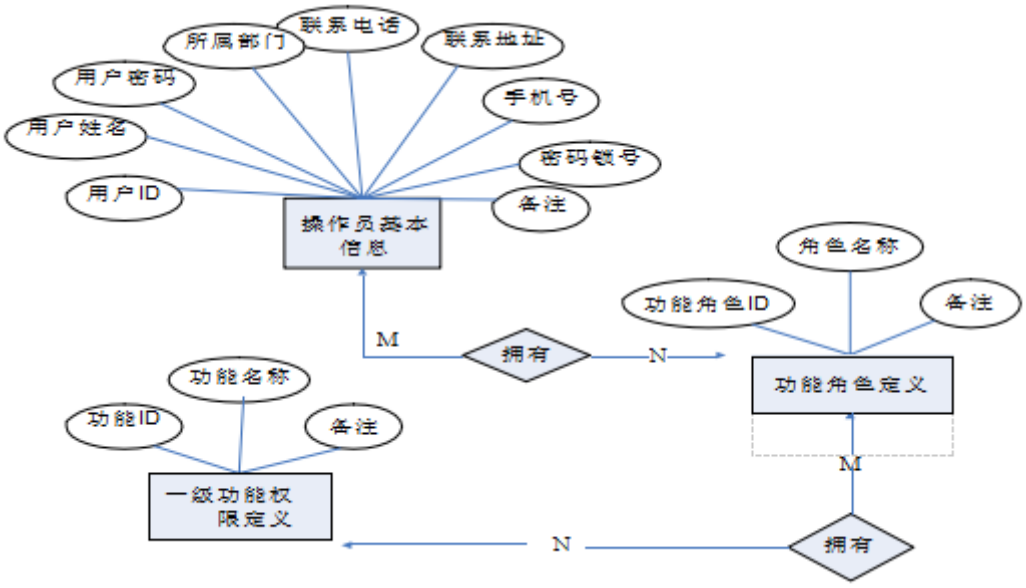


图2.9.1E-R图

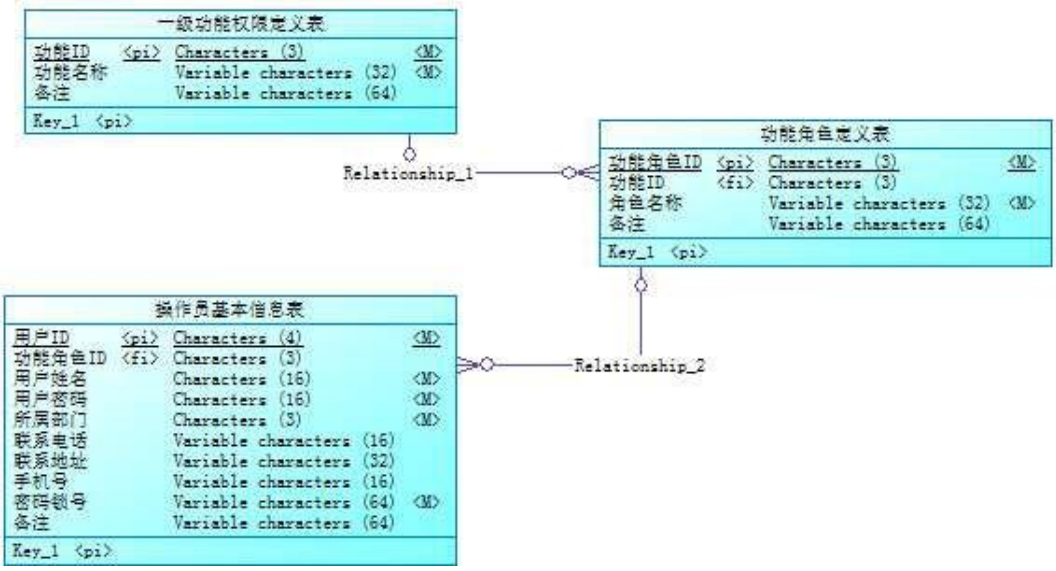


图2.9.2逻辑数据模型图

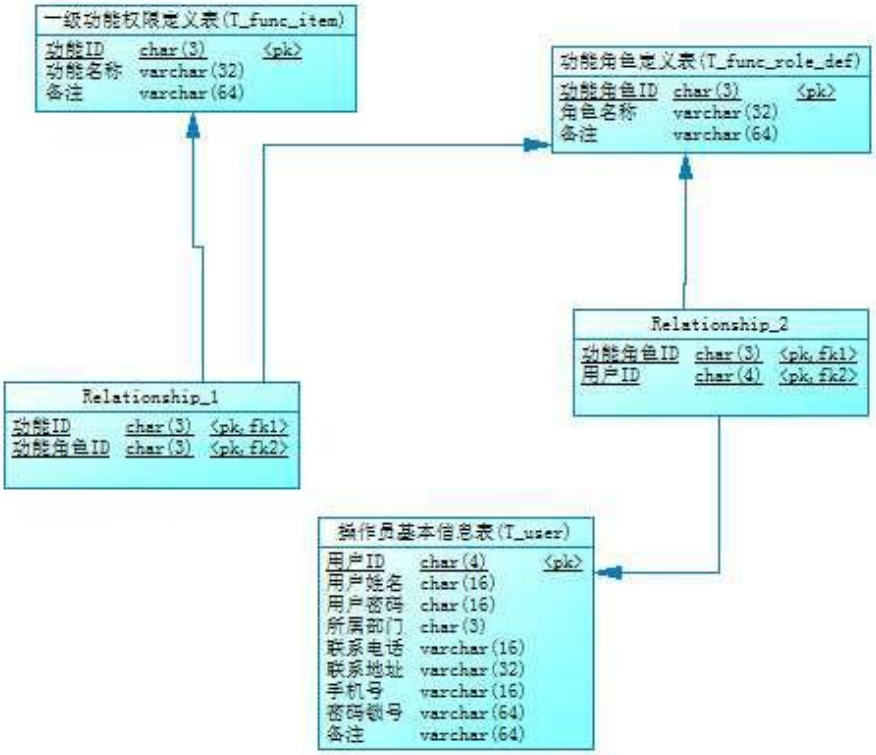


图2.9.3物理数据模型图

表2.9.1字段名定义表

字段名	字段说明	字段名	字段说明
func_id	功能 id	user_passwd	用户密码
func_name	功能名称	dept_id	所属部门
func_role_id	功能角色 id	telephone	联系电话
func_role_name	角色名称	address	联系地址
user_id	用户 id	handphone	手机号
user_name	用户姓名	usb_no	密码锁号
reserve	备注		

任务一：创建数据库（10 分）

创建数据库 ConstructionDB。

任务二：创建数据表（25 分）

根据图 2.1.2 和表 2.1.1，创建数据表 T_user、T_func_item、T_func_role_def 及两个关系表（关系表的名字自拟）。

任务三：创建数据表间的关系及约束（15 分）

根据物理数据原型，创建数据关系表。

任务四：数据操作（25 分）

用 SQL 语句完成如下操作：

①.在T_user 表插入数据“：id01，刘德华，123，KBB，5678900，湖南长沙，13899005678，ldh123，admin”；

②.查询出所属部门为“KBB”的操作员的基本信息；

③.查询出姓名为“刘德华”的操作员具有哪些功能权限；

④.查询出“投标责任人”角色所拥有的功能；

5.创建视图查询操作员的姓名，密码和所属部门；

6.创建存储过程，查询指定操作员所具有的功能权限。

（2）作品提交要求见本模块附录 1

（3）实施条件要求见本模块附录 2

（4）评价标准见本模块附录 3

10. 试题编号：2-10《某电子商务网站》项目产品管理模块

任务描述

《产品管理》模块的 E-R 图如图 2.10.1 所示，逻辑数据模型如图 2.10.2 所示，物理数据模型如图 2.10.3 所示，数据表字段名定义见表 2.10.1。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：

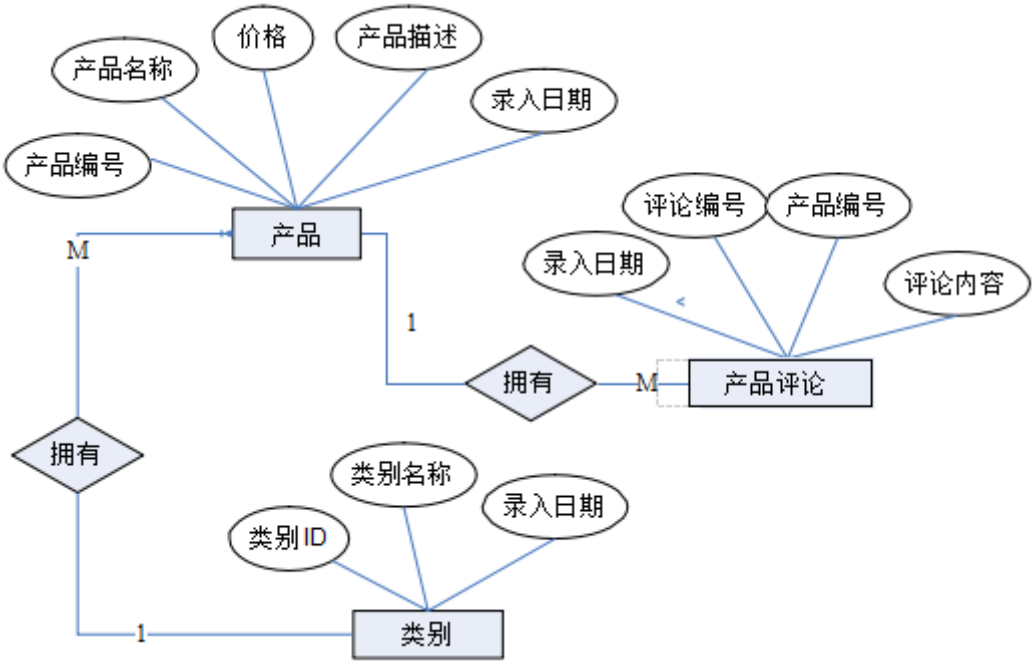


图2.10.1E-R图

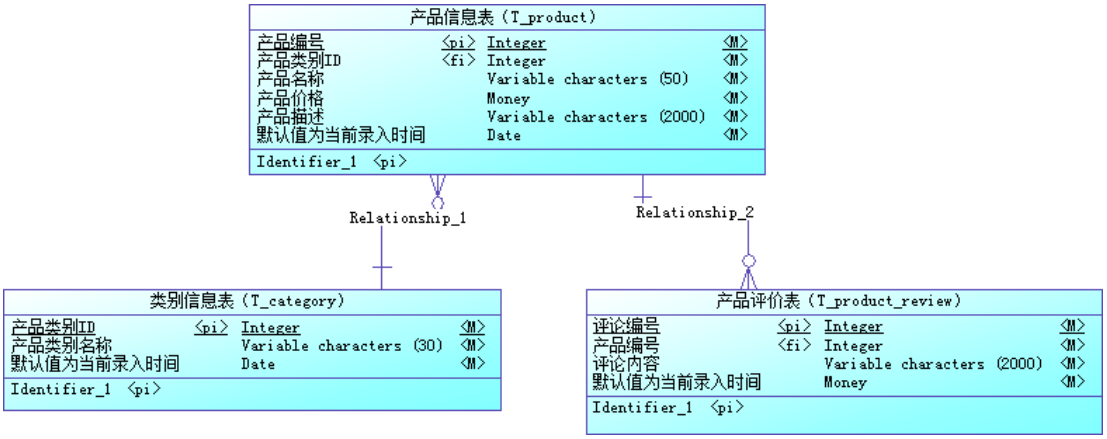


图2.10.2逻辑数据模型图

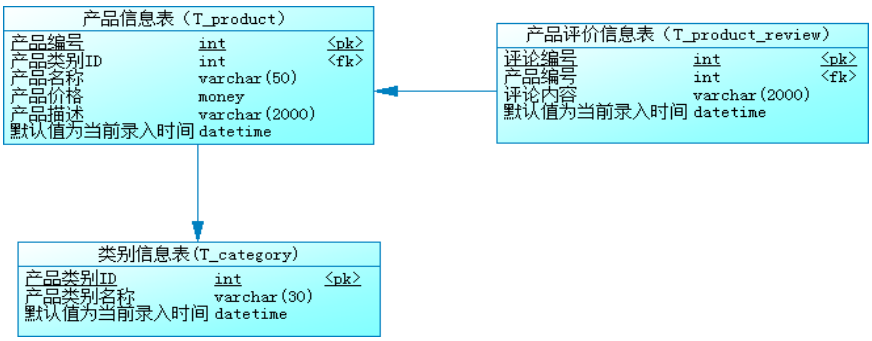


图2.10.3物理数据原型图

表2.10.1字段名定义表

字段名	字段说明	字段名	字段说明
category_id 标识列	产品类别 ID	remark	产品描述
category_name	产品类别名称	register_date	默认值为当前录入时间
register_date	默认值为当前录入时间	review_id 标识列	评论编号
product_id	产品编号	product_id 标识列	产品编号
category_id	产品类别 ID	review	评论内容
product_name	产品名称	register_date	默认值为当前录入时间
price	产品价格		

任务一：创建数据库（10 分）

创建数据库 ProductDB。

任务二：创建数据表（25 分）

根据图2.10.2 和表2.10.1，创建数据表 T_category、T_product_review、T_product，其中产品表的产品ID(product_id)列设置为标识列，自动从 1 开始增长。

任务三：创建数据表间的关系及约束（15 分）

①.创建主键（三个表均设置）；

②.产品价格列（Price）只能输入 1-1000 之间的数；

③.录入时间列(Register_date)默认值为当前录入时间（三个表均设置）。

任务四：数据操作（25 分）

①.用 SQL 语句查询出如下数据：在三个表分别中录入 3 条测试数据（样本数据包含下面题目中使用的数据）；

②.查询某类别下所有产品；

③.查询产品价格在 300-500 元之间的产品；

④.查询录入日期在 2011 年 3 月到 6 月之间的产品数据；

⑤.查询产品价格在 90-200 元之间的所有评论；

⑥.查询评论数在 1-3 条的所有产品。

（2）作品提交要求见本模块附录 1

（3）实施条件要求见本模块附录 2

（4）评价标准见本模块附录 3

数据库设计模块附录

附录1作品提交

答案以“答题文件”的形式提交。请按以下要求创建答题文件夹和答题文件：

①创建答题文件夹

创建以”考生号_题号”命名的文件夹，存放所有答题文件，例如：“340103**********_2_1\”

②创建答题文件

▪ SQL脚本文件

创建project.sql文件，如：“340103**********_2_1\project.sql，存放SQL脚本代码。

▪ 数据库文件

创建db子文件夹，如:“340103**********_2_1\db\”，存放数据库备份文件，它用于教师阅卷时还原数据库。

③提交答题文件

将“考生号_题号”文件夹打包，形成“考生号_题号.rar”文件，如：“340103**********_2_1.rar”，将该文件按要求进行上传。

④考核时量

考核时长为180分钟。

附录2实施条件

所需的软硬件设备如下表。

表1考点提供的主要设备及软件表

序号	设备、软件名称	规格/技术参数、用途	备注
1	人工智能实训机房	测试场地	保证参考人员有足够间距
2	计算机	CPU酷睿i5以上，内存4G以上，win7/win10操作系统	用于软件开发和软件部署，每人一台
3	Office		编写文档
4	SQLServer2008或以上、Oracle10g或以上、MySQL5.5或以上	数据库管理系统	参考人员任选一种数据库管理系统

附录3评价标准

表2考核评价细则表

评价项	分值	评分细则
数据库创建	10分	没有成功创建数据库，扣5-8分。
数据表创建	25分	数据表创建不成功每一项扣3-5分，字段创建不符合要求每一项扣2-3分，扣完为止。
约束及关系创建	15分	约束创建不成功每一项扣3-5分，关系创建不符合要求每一项扣5分，扣完为止。
数据访问	25分	没有正确写出SQL语句每一项扣4-5分，扣完为止。

数据库管理系统配置与使用		5分	数据库服务器与管理工具配置不正确，无法连接数据库扣5分。
文档规范	数据库名命名规范	2分	数据库名命名不规范扣2分。
	数据表名命名规范	3分	数据表命名不规范每张表扣1分，扣完为止。
	字段名命名规范	5分	字段名命名不规范每项扣0.5分，扣完为止。

表3职业素质评分细则表

序号	评分项	分值	评分 细则
1	代码书写格式规范	3分	代码缩进不规范扣1分、方法划分不规范扣1分、语句结构不规范扣1分（如一行编写两个语句）、使用空行不规范扣1分，扣完为止。
2	注释规范	2分	整个项目没有注释扣2分、有注释，但注释不规范扣1分，扣完为止。
5	运行正确	5分	所写代码无法正常运行扣5分。

二、岗位核心技能模块

模块一 爬虫应用技术与开发

1.试题编号：3-1 爬取知乎的数据挖掘话题网页

（1）任务描述

对知乎指定的网页进行爬取

网页:https://www.zhihu.com/topic/19553534/hot

任务一：依赖库准备（10 分）

①.导入requests库

②.导入re库

任务二：爬取指定网页的数据（70 分）

①.usl书写正确

②.写入正则式

③.爬取的内容用text类型赋予新变量

④.使用全局正则搜索

⑤.输出二进制类型数据

⑥.输出当前编码形式

（2）作品提交要求见本模块附录 1

（3）实施条件要求见本模块附录 2

（4）评价标准见本模块附录 3

2.试题编号：3-2 爬取知乎的发现模块

（1）任务描述

对知乎指定的网页进行爬取

网页:https://www.zhihu.com/explore

任务一：依赖库准备（10 分）

①.导入requests库

②.导入re库

任务二：爬取指定网页的数据（70 分）

①.usl书写正确

②.写入正则式

③.爬取的内容用text类型赋予新变量

④.使用全局正则搜索

⑤.输出二进制类型数据

⑥.输出当前编码形式

（2）作品提交要求见本模块附录 1

（3）实施条件要求见本模块附录 2

（4）评价标准见本模块附录 3

3.试题编号：3-3 爬取知乎的人工智能模块

（1）任务描述

对知乎指定的网页进行爬取

网页:https://www.zhihu.com/topic/19551275/hot

任务一：依赖库准备（10 分）

①.导入requests库

②.导入re库

任务二：爬取指定网页的数据（70 分）

①.usl书写正确

②.写入正则式

③.爬取的内容用text类型赋予新变量

④.使用全局正则搜索

⑤.输出二进制类型数据

⑥.输出当前编码形式

（2）作品提交要求见本模块附录 1

（3）实施条件要求见本模块附录 2

（4）评价标准见本模块附录 3

4.试题编号：3-4 爬取知乎的计算机科学模块

（1）任务描述

对知乎指定的网页进行爬取

网页:https://www.zhihu.com/topic/19580349/hot

任务一：依赖库准备（10 分）

①.导入requests库

②.导入re库

任务二：爬取指定网页的数据（70 分）

①.usl书写正确

②.写入正则式

③.爬取的内容用text类型赋予新变量

④.使用全局正则搜索

⑤.输出二进制类型数据

⑥.输出当前编码形式

（2）作品提交要求见本模块附录 1

（3）实施条件要求见本模块附录 2

（4）评价标准见本模块附录 3

5.试题编号：3-5 爬取知乎的机器学习模块

（1）任务描述

对知乎指定的网页进行爬取

网页:https://www.zhihu.com/topic/19559450/hot

任务一：依赖库准备（10 分）

①.导入requests库

②.导入re库

任务二：爬取指定网页的数据（70 分）

①.usl书写正确

②.写入正则式

③.爬取的内容用text类型赋予新变量

④.使用全局正则搜索

⑤.输出二进制类型数据

⑥.输出当前编码形式

（2）作品提交要求见本模块附录 1

（3）实施条件要求见本模块附录 2
（4）评价标准见本模块附录 3

6.试题编号：3-6 爬取知乎的人工智能算法模块

（1）任务描述

对知乎指定的网页进行爬取

网页:https://www.zhihu.com/topic/19691108/hot

任务一：依赖库准备（10 分）

①.导入requests库

②.导入re库

任务二：爬取指定网页的数据（70 分）

①.usl书写正确

②.写入正则式

③.爬取的内容用text类型赋予新变量

④.使用全局正则搜索

⑤.输出二进制类型数据

⑥.输出当前编码形式

（2）作品提交要求见本模块附录 1
（3）实施条件要求见本模块附录 2
（4）评价标准见本模块附录 3

7.试题编号：3-7 爬取知乎的深度学习模块

（1）任务描述

对知乎指定的网页进行爬取

网页:https://www.zhihu.com/topic/19813032/hot

任务一：依赖库准备（10 分）

①.导入requests库

②.导入re库

任务二：爬取指定网页的数据（70 分）

①.usl书写正确

②.写入正则式

③.爬取的内容用text类型赋予新变量

④.使用全局正则搜索

⑤.输出二进制类型数据

⑥.输出当前编码形式

（2）作品提交要求见本模块附录 1
（3）实施条件要求见本模块附录 2
（4）评价标准见本模块附录 3

8.试题编号：3-8 爬取知乎的 BERT 模块

（1）任务描述

对知乎指定的网页进行爬取

网页:https://www.zhihu.com/topic/20743626/hot

任务一：依赖库准备

①.导入requests库

②.导入re库

任务二：爬取指定网页的数据（70 分）

①.usl书写正确

②.写入正则式

③.爬取的内容用text类型赋予新变量

④.使用全局正则搜索

⑤.输出二进制类型数据

⑥.输出当前编码形式

（2）作品提交要求见本模块附录 1

（3）实施条件要求见本模块附录 2

（4）评价标准见本模块附录 3

9.试题编号：3-9 爬取知乎的自然语言处理模块

（1）任务描述

对知乎指定的网页进行爬取

网页:https://www.zhihu.com/topic/19560026/hot

任务一：依赖库准备（10 分）

①.导入requests库

②.导入re库

任务二：爬取指定网页的数据（70 分）

①.usl书写正确

②.写入正则式

③.爬取的内容用text类型赋予新变量

④.使用全局正则搜索

⑤.输出二进制类型数据

⑥.输出当前编码形式

（2）作品提交要求见本模块附录 1

（3）实施条件要求见本模块附录 2

（4）评价标准见本模块附录 3

10.试题编号：3-10 爬取知乎的机器翻译模块

（1）任务描述

对知乎指定的网页进行爬取

网页:https://www.zhihu.com/topic/19616892/hot

任务一：依赖库准备（10 分）

①.导入requests库

②.导入re库

任务二：爬取指定网页的数据（70 分）

①.usl书写正确

②.写入正则式

③.爬取的内容用text类型赋予新变量

④.使用全局正则搜索

⑤.输出二进制类型数据

⑥.输出当前编码形式

（2）作品提交要求见本模块附录 1

（3）实施条件要求见本模块附录 2

（4）评价标准见本模块附录 3

爬虫应用技术与开发模块附录

附录1作品提交

答案以“答题文件”的形式提交。请按以下要求创建答题文件夹和答题文件：

①创建答题文件夹

创建以”考生号_题号”命名的文件夹，存放所有答题文件，例如：“340103**********_3_1\”。

②创建答题文件

a.项目源文件

创建任务子文件夹，如：“340103**********_3_1\task\”，存放任务所有结果。

b.页面截图文件

在任务子文件夹中，存放截图.doc文件，它用于保存安装，配置，启动，运行执行过程中的屏幕截图，每张截图中每个关键配置或结果等，必须用红色矩形框标识出来并加以文字说明。

③提交答题文件

将”考生号_题号”文件夹打包，形成“考生号_题号.RAR”文件，如：“340103**********_3_1.rar”，将该文件按要求进行上传。

④考核时量

考核时间为180分钟。

附录 2 实施条件

所需的软硬件设备如下表。

表 1 考点提供的主要设备及软件表

序号	场地、设备、软件名称	规格/技术参数、用途	备注
1	人工智能实训机房	测试场地	保证参考人员有足够间距
2	计算机	CPU酷睿i5以上，内存4G以上，win7/win10/linux操作系统	用于软件开发和软件部署，每人一台
3	Pycharm2018.2以上	软件开发	参考人员自选一种开发工具

附录3评价标准

表2任务一评分细则（10分）

序号	评分项	分值	评分细则
1	选择requests工具	6分	开发环境选择不正确，无法启动开发环境扣 6 分。
2	配置pycharm库并测试	4分	依赖库配置不正确，依赖库报错。

表3任务二评分细则（70分）

序号	评分项	分值	评分细则
1	导入库正确	10 分	导入不正确扣10分

2	Ulr书写正确	15 分	没有成功新建数据库扣 5 分，没有成功创建表扣 3 分/处。插入数据出现不完整，不符合要求的情况扣 2 分/处，扣完为止。
3	正则式书写正确	15 分	正则式书写错误扣15分
4	爬取的内容赋予新变量	15 分	设计的类体现了数据和业务的分离，没有系统架构分层设计扣 2 分/处，扣完为止。
5	二进制输出	5 分	项目与数据库连接配置不正确，出现异常扣 2 分/处，扣完为止。
6	输出当前的编码形式	10 分	编码形式输出当前的形式否则扣10分

表 4 职业素质评分细则表（10分）

序号	评分项	分值	评分细则
1	代码书写格式规范	5 分	代码缩进不规范扣 1 分、方法划分不规范扣 1 分、语句结构不规范扣 1 分（如一行编写两个语句）、使用空行不规范扣 1 分，扣完为止。
2	注释规范	5 分	整个项目没有注释扣 2 分、有注释，但注释不规范扣 1 分，扣完为止。

表 5 实操文档评分细则表（10分）

序号	评分项	分值	评分细则
1	实操文档有无	2分	有实操文档得分，无实操文档扣2分。
2	文档任务截图	4分	有操作过程截图得分，无操作过程截图扣4分。
3	文档任务截图标注	4分	有文档任务截图标注说明和画框得分，无标注和画框扣4分

模块二 数据挖掘与机器学习

1.试题编号：4-1 泰坦尼克号信息调查数据挖掘模块

（1）任务描述

我们要通过对泰坦尼克号的titanic.csv文件里面的字段进行生存预测以及对各个活下来的因素进行分析：

字段名定义表

字段名	字段说明	字段名	字段说明
func_idPassengerId	乘客编号	Parch	父母/小孩的个数
Survived	是否生存	Ticket	船票信息
Pclass	船舱等级	Fare	票价
Name	名字	Cabin	船舱号
Sex	性别	Embarked	登船港口
Age	年龄	SibSp	兄弟姐妹/配偶的个数

任务一：数据信息处理（20 分）

①.读取给定的数据集

②.查看数据行列名

③.查看前五行

④.查看缺失值数量

任务二：数据分析（30 分）

①.使用SibSp和Parch生成一个新字段family_cnt，使用SibSp+Parch+1得到

②.将字段family_cnt分成三种信息（生成一个新字段family_type），数值为1，得到’sigle’，数值为2,3，得到'middle'，其他数值为'big'，

③.Embarked字段使用众数填充缺失值，Fare使用均值填充缺失数据，Cabin使用众数填充

④.将特征离散值进行独热处理

⑤.将特征连续值进行标准化处理

⑥.用均值填充Age列，创建一个新列Age_new

任务三：模型预测（30 分）

①.使用Pclass，Sex，family_type，cabin，family_cnt作为预测age的特征值，将没有缺失Age的数据作为训练集，缺失信息作为测试集

②.使用L1正则模型，配合网格搜索交叉验证，找到最优参数，参数自定

③.输出准确率和预测值

（2）作品提交要求见本模块附录 1

（3）实施条件要求见本模块附录 2

（4）评价标准见本模块附录 3

2.试题编号：4-2 银行信贷预测数据分析模块

（1）任务描述

该项目利用单模型：决策树、贝叶斯、SVM 等；集成模型：随机森林、梯度提升树等；评分卡模型： 逻辑回归；项目可输出： 评分卡；对数据集包含有关信贷申请人的信息。在全球范围内，银行使用这种数据集和信息数据类型来创建模型，以帮助决定接受/拒绝谁的贷款。在进行所有探索性数据分析、清理和处理我们可能（将）发现的所有异常之后，一个好/坏申请人的模式将暴露在机器学习模型中学习：

任务一：数据信息处理（20 分）

①.读取给定的数据集

②.查看数据行列名

③.查看前五行

④.查看缺失值数量

任务二：数据分析（30 分）

1.使用中文做列名

（1）用户的基本属性 user_info.txt，共 6 个字段，分别是 用户 id、性别、职业、教育程度、婚姻状态、户口类型，其中字段性别为 0 表示性别未知。例如：57189,1,2,4,3,2

字段名称	字段类型	说明
用户 id	整数	例：57189
性别	整数	0:未知， 1:男， 2:女
职业	整数	
教育程度	整数	
婚姻状态	整数	
户口类型	整数	

（2）银行流水记录 bank_detail.txt，共 5 个字段，分别是 用户 id、时间戳、交易类型、交易金额、工资收入标记，其中第 2 个字段，时间戳为 0 表示时间未知；第 3 个字段，交易类型有两个值，1 表示支出、0 表示收入；第 5 个字段，工资收入标记为 1 时，表示工资收入。例如：6951,5894316387,0,13.756664,0

字段名称	字段类型	说明
用户 id	整数	例：57189
时间戳	整数	0 表示时间未知 例：5894316387
交易类型	整数	1 表示支出、0 表示收入
交易金额	浮点数	
工资收入标记	整数	0 表示不是工资收入，1 表示工资收入

（3）用户浏览行为 browse_history.txt。共 4 个字段，分别为用户 id、时间戳、浏览行为数据、浏览子行为编号。例如 34724,5926003545,172,1

字段名称	字段类型	说明
用户 id	整数	例：57189
时间戳	整数	0 表示时间未知 例：5894316387
浏览行为数据	整数	
浏览子行为编号	整数	

（4）信用卡账单记录 bill_detail.txt，共 15 个字段，分别为

字段名称	字段类型	说明
用户 id	整数	
账单时间戳	整数	0 表示时间未知 例：5894316387

字段名称	字段类型	说明
银行 id	枚举类型	
上期账单金额	浮点数	
上期还款金额	浮点数	
信用卡额度	浮点数	
本期账单余额	浮点数	
本期账单最低还款额	浮点数	
消费笔数	整数	
本期账单金额	浮点数	
调整金额	浮点数	
循环利息	浮点数	
可用金额	浮点数	
预借现金额度	浮点数	
还款状态	枚举值	

（5）放款时间信息 loan_time.txt。共 2 个字段，用户 id 和放款时间。

字段名称	字段类型	说明
用户 id	整数	例：57189
放款时间	整数	0 表示时间未知 例：5894316387

（6）顾客是否发生逾期行为的记录 overdue.txt。共 2 个字段，为用户 id 和样本标签，样本标签为 1，表示逾期 30 天以上；样本标签为 0，表示逾期 10 天以内。

字段名称	字段类型	说明
用户 id	整数	例：57189
样本标签	整数	0 表示逾期 10 天以内， 1 表示逾期 30 天以上

任务三：模型预测（30 分）

①.使用上面的合并的表数据切分'上期账单金额', '上期还款金额','本期账单余额','本期账单最低还款额'作为的特征值，将'样本标签'的列数据作为标签，训练集和测试集的比例为7：3

②.使用随机森林模型，配合网格搜索交叉验证，找到最优参数，参数自定

③.输出准确率和预测值

3.试题编号：4-3 共享单车数据分析模块

（1）任务描述

某城市的共享单车 2011 年到 2012 年的数据集。该数据集包括了租车日期，租车季节，租车天气，租车气温，租车空气湿度等数据。本次将使用数据挖掘对这一数据集进行探索性分析

字段名	字段说明	字段名	字段说明
datetime	日期时间	humidity	湿度
season	季节	windspeed	风速

holiday	假期	casual	休闲用户
weather	天气	registered	注册用户
temp	温度	count	共计数
atemp	加过权重之后的温度	workingday	工作日

任务一：数据信息处理（20 分）

①.读取给定的数据集

②.查看数据行列名

③.查看前五行

④.查看缺失值数量

任务二：数据分析（30 分）

①. 把一些浮点字段变为整形

②. 把季节改成春夏秋冬

③. 指定要保留的时间数据（季节，月，日，小时，周，注册人数）然后画热力图

④. 画箱线图统计月份和骑行数量的关系

⑤. 画箱线图统计春夏秋冬和骑行数量的关系

⑥. 画箱线图统计季节和骑行数量的关系

任务三：模型预测（30 分）

①.用温度列预测注册用户，训练和测试集的比例7：3

②.使用L1模型，配合网格搜索交叉验证，找到最优参数，参数自定

③.输出准确率和预测值

（2）作品提交要求见本模块附录 1

（3）实施条件要求见本模块附录 2

（4）评价标准见本模块附录 3

4.试题编号：4-4 航空公司客户价值数据分析模块

（1）任务描述

借助航空公司数据，对客户进行分类,对不同类别的客户进行特征分析，比较不同类别客户的价值，对不同价值的客户进行个性化服务，指定相应的营销策略字段名定义表：

客户基本信息	MEMBER_NO	会员卡号
	FFP_DATE	入会时间
	GENDER	性别
	FFP_TIER	会员卡级别
	WORK_CITY	工作地城市
	WORK_PROVINCE	工作地所在省份
	WORK_COUNTRY	工作地所在国家
	AGE	年龄
乘客信息	FIRST_FLIGHT_DATE	第一次飞行日期
	LOAD_TIME	观测窗口的结束时间
	FLIGHT_COUNT	飞行次数
	SUM_YR_1	第一年总票价
	SUM_YR_2	第二年总票价
	SEG_KM_SUM	观测窗口总飞行公里数
	WEIGHTED_SEG_KM	观测窗口总加权飞行公里数（Σ舱位折扣×航段距离）
	LAST_FLIGHT_DATE	末次飞行日期
	AVG_FLIGHT_COUNT	观测窗口季度平均飞行次数
	BEGIN_TO_FIRST	观测窗口第一次乘机时间至MAX（观测窗口时段，入会时间）时长
	LAST_TO_END	最后一次乘机时间至观测窗口末端时长
	AVG_INTERVAL	平均乘机时间间隔
	MAX_INTERVAL	观测窗口内最大乘机间隔
	avg_discount	平均折扣率
	P1Y_Flight_Count	第一年乘机次数
	L1Y_Flight_Count	第二年乘机次数
	Ration_L1Y_Flight_Count	第二年的乘机次数比率
	Ration_P1Y_Flight_Count	第一年的乘机次数比率
积分信息	EXCHANGE_COUNT	积分兑换次数
	AVG_BP_SUM	观测窗口季度平均基本积分累计
	BP_SUM	观测窗口总基本积分
	EP_SUM_YR_1	第一年精英资格积分
	EP_SUM_YR_2	第二年精英资格积分
	ADD_POINTS_SUM_YR_1	观测窗口中第一年其他积分
	ADD_POINTS_SUM_YR_2	观测窗口中第二年其他积分
	P1Y_BP_SUM	第一年里程积分
	L1Y_BP_SUM	第二年里程积分
	EP_SUM	观测窗口总精英积分
	ADD_Point_SUM	观测窗口中其他积分
	Eli_Add_Point_Sum	非乘机积分总和
	L1Y_ELi_Add_Points	第二年非乘机积分总和
	Points_Sum	总累计积分
	L1Y_Points_Sum	第二年观测窗口总累计积分
	Ration_P1Y_BPS	第一年里程积分占最近两年积分比例
	Ration_L1Y_BPS	第二年里程积分占最近两年积分比例
	Point_NotFlight	非乘机的积分变动次数

任务一：数据信息处理（20 分）

①.读取给定的数据集

②.查看数据行列名

③.查看前五行

④.查看缺失值数量

任务二：数据分析（30 分）

①.切分列并创建新表

```
"FFP_DATE", "LOAD_TIME", "FLIGHT_COUNT",
"SUM_YR_1", "SUM_YR_2", "SEG_KM_SUM",
"AVG_INTERVAL", "MAX_INTERVAL", "avg_discount"
```

②.计算每公里支付价格

③.计算乘坐航班频率

④.将飞行次数，总里程列名改为中文

⑤.处理成员日期时长

⑥.计算时间间隔差值

任务三：模型预测（30 分）

①.使用'LEN_REL','FLIGHT_COUNT','avg_discount','SEG_KM_SUM','LAST_TO_END'作为聚类计算的特征

②.使用聚类模型，将目标样本分为5个群体

③.查看每个样本预测的群体的标签

（2）作品提交要求见本模块附录 1

（3）实施条件要求见本模块附录 2

（4）评价标准见本模块附录 3

5.试题编号：4-5 纽约爱彼迎Airbnb数据挖掘模块

（1）任务描述

自 2008 年以来，客人和房东利用 Airbnb 扩大了旅行的可能性，并提出了一种更独特、个性化的体验世界的方式。通过 Airbnb 提供的数百万个房源的数据分析是该公司的一个关键因素。这些数以百万计的房源产生了大量的数据其可以被分析并用于安全、商业决策、了解客户和供应商（房东）在平台上的行为和表现、指导营销举措、实施创新的附加服务等等。

id 挂牌编号

name 挂牌名字

host_id 主人编号

host_name 主人名字

neighbourhood_group 房屋所在区域

neighbourhood 房屋具体地区

latitude 经纬度

longitude 经纬度

room_type 房间类型

price 价格

minimum_nights 最少的预定夜数

number_of_reviews 评论数

last_review 最新评论

reviews_per_month 每月评论数

calculated_host_listings_count 主人拥有房屋的数量

availability_365 可供预订的天数

任务一：数据信息处理（20 分）

①.读取给定的数据集

②.查看数据行列名

③.查看前五行

④.查看缺失值数量

任务二：数据分析（30 分）

①. 删除房屋名称，主人姓名列

②. 填充 reviews_per_month 列，填充为 0

③. 分别对针对评论数以及年度可用天数进行分组组成新的数据

④. 建立房型与价格、评论数目、可用天数的表格并进行可视化

⑤.年度可用天数与房间类型，地区的联系表

⑥.价格与房间类型，地区的联系表

任务三：模型预测（30 分）

①.根据房间类型预测价格，训练和测试集的比例7：3

②.使用L1模型，配合网格搜索交叉验证，找到最优参数，参数自定

③.输出准确率和预测值

（2）作品提交要求见本模块附录 1

（3）实施条件要求见本模块附录 2

（4）评价标准见本模块附录 3

6.试题编号：4-6 IBM员工离职因素数据分析模块

（1）任务描述

IBM 员工离职原因数据及包括员工编号、年龄、受教育程度、离家距离、生活和工作的平衡、工作参与情况等信息。通过分析该数据集可以找出员工流失的因素，例如，工作角色和流失率的相关性；离家距离与流失率的相关性；平均月收入和受教育程度对流失率的影响”。

age：年龄字段

bussinessTravel ： 是否经常出差

DistanceFromHome： 距离

Education：教育程度

gender： 性别

joblevel： 工作等级

MonthlyIncome ： 员工月收入

numcompanies： 曾经工作过公司数量

这 8 个字段来进行预测 是否离职 attrition 字段表示 是否离职

任务一：数据信息处理（20 分）

①.读取给定的数据集

②.查看数据行列名

③.查看前五行

④.查看缺失值数量

任务二：数据分析（30 分）

①. 删除部门不是 研究部的行

②. 按列排序没啥用

③. 把 gender 字段 male 转成 0， female 转 0

④. 把 BussinessTravel 字段的 频繁出差转成 2 很少转 1，不出差转 0

⑤. 数据切分前 90 行为训练集

⑥. 数据切分后 10 行为测试集

任务三：模型预测（30 分）

①.用身高,体重,收入列预测attrition是否离职，

②.使用 KNN模型，配合网格搜索交叉验证，找到最优参数，参数自定

③.输出准确率和预测值

（2）作品提交要求见本模块附录 1

（3）实施条件要求见本模块附录 2

（4）评价标准见本模块附录 3

7.试题编号：4-7 信用卡欺诈检测模块

（1）任务描述

假设你受雇于帮助一家信用卡公司检测潜在的欺诈案件，你的工作是确保客户不会因未购买的商品而被收取费用。给你一个包含人与人之间交易的数据集，他们是欺诈与否的信息，并要求你区分它们。我们的最终目的是通过构建分类模型来对欺诈交易进行分类区分来解决上述情况

’Class’是响应变量，如果发生被盗刷，则取值 1，否则为 0。

任务一：数据信息处理（20 分）

①.读取给定的数据集

②.查看数据行列名

③.查看前五行

④.查看缺失值数量

任务二：数据分析（30 分）

①.把数据按照索引排序
②.对 Amount 列进行数据标准化
③.提取类别所在列，class 列，其他列为特征，提取特征列
④.计算 class==1 的样本数量并取出对应的索引值并转换成数组
⑤.取出 class=0 的索引值
⑥.特征降维到

任务三：模型预测（30 分）
①.切分训练和测试集的比例7：3
②.使用逻辑回归模型，配合网格搜索交叉验证，找到最优参数，参数自定
③.输出准确率和预测值

（2）作品提交要求见本模块附录 1

（3）实施条件要求见本模块附录 2

（4）评价标准见本模块附录 3

8.试题编号：4-8 app客户流失及客户行为偏好分析模块

（1）任务描述

深入了解用户画像及行为偏好，挖掘出影响用户流失的关键因素，并通过算法预测客户访问的转化结果，从而更好地完善产品设计、提升用户体验！

数据说明

此次数据是携程用户一周的访问数据，为保护客户隐私，已经将数据经过了脱敏，和实际商品的订单量、浏览量、转化率等有一些差距，不影响问题的可解性。

字段	解释
sampleid	样本id
label	目标变量
d	访问日期
arrival	入住日期
iforderpv_24h	24小时内是否访问订单填写页
decisionhabit_user	决策习惯：以用户为单位观察决策习惯
historyvisit_7ordernum	近7天用户历史订单数
historyvisit_totalordernum	近1年用户历史订单数
hotelcr	当前酒店历史cr
ordercanceledprecent	用户一年内取消订单率
landhalfhours	24小时内登陆时长
ordercanncelednum	用户一年内取消订单数
commentnums	当前酒店点评数
starprefer	星级偏好
novoters	当前酒店评分人数
consuming_capacity	消费能力指数
historyvisit_avghotelnum	近3个月用户历史日均访问酒店数
cancelrate	当前酒店历史取消率
historyvisit_visit_detailpagenum	7天内访问酒店详情页数
delta_price1	用户偏好价格-24小时浏览最多酒店价格
price_sensitive	价格敏感指数
hoteluv	当前酒店历史uv
businessrate_pre	24小时历史浏览次数最多酒店商务属性指数
ordernum_oneyear	用户年订单数
cr_pre	24小时历史浏览次数最多酒店历史cr
avgprice	平均价格
lowestprice	当前酒店可定最低价
firstorder_bu	首单bu
customereval_pre2	24小时历史浏览酒店客户评分均值
delta_price2	用户偏好价格-24小时浏览酒店平均价格
commentnums_pre	24小时历史浏览次数最多酒店点评数
customer_value_profit	客户价值_近1年
commentnums_pre2	24小时历史浏览酒店点评数均值
cancelrate_pre	24小时内已访问次数最多酒店历史取消率
novoters_pre2	24小时历史浏览酒店评分人数均值
novoters_pre	24小时历史浏览次数最多酒店评分人数
ctrip_profits	客户价值
deltaprice_pre2_t1	24小时内已访问酒店价格与对手价差均值，t+1
lowestprice_pre	24小时内已访问次数最多酒店可订最低价
uv_pre	24小时历史浏览次数最多酒店历史uv
uv_pre2	24小时历史浏览酒店历史uv均值
lowestprice_pre2	24小时内已访问酒店可订最低价均值
lasthtlordergap	一年内距离上次下单时长
businessrate_pre2	24小时内已访问酒店商务属性指数均值
cityuvs	昨日访问当前城市同入住日期的app uv数
cityorders	昨日提交当前城市同入住日期的app订单数
lastpvgap	一年内距上次访问时长
cr	用户转化率
sid	会话id，sid=1可认为是新访
visitnum_oneyear	年访问次数
h	访问时间点

任务一：数据信息处理（20 分）
①.读取给定的数据集

②.查看数据行列名
③.查看前五行
④.查看缺失值数量

任务二：数据分析（30 分）
①.处理异常值，最低酒店定价有小于 0 的，有等于 1 的值，明显属于异常值，删除掉这些值
②.画出访问日期和入住日期的直方图
③.画出近 7 天用户历史订单数和近一年天用户历史订单数的直方图
④.画出当前酒店点评数和当前酒店评分人数的直方图
⑤.画出平均价格和用户年订单数的直方图
⑥.画出平均价格和当前酒店可定最低价的直方图

任务三：模型预测（30 分）
①.用当前酒店历史cr和当前酒店历史uv列预测当前酒店评分人数，训练和测试集的比例7：3
②.使用L1模型，配合网格搜索交叉验证，找到最优参数，参数自定
③.输出准确率和预测值

（2）作品提交要求见本模块附录 1

（3）实施条件要求见本模块附录 2

（4）评价标准见本模块附录 3

9.试题编号：4-9 电信用户流失预测模块

（1）任务描述

研究背景

①.做好“用户流失预测”可以降低营销成本。老生常谈，“新客户开发成本”是“老客户维护成本”的 5 倍。

②.获得更好的用户体验。并不是所有的增值服务都可以有效留住客户。

③.获得更高的销售回报。价格敏感型客户和非价格敏感性客户

提出问题

①.流失客户有哪些显著性特征？

②.当客户在哪些特征下什么条件下比较容易发生流失？

数据集描述

该数据是 datafountain 上的《电信客户流失数据》点此下载数据

该数据集有 21 个变量，7043 个数据点。变量可分为以下三个部分：用户属性、用户行为、研究对象

用户属性

customerID ：用户 ID。

gender：性别。（Female & Male）

SeniorCitizen ：老年人 （1 表示是，0 表示不是）

Partner ：是否有配偶 （Yes or No）

Dependents ：是否经济独立 （Yes or No）

tenure ：客户的职位（0-72，共 73 个职位）

用户行为

PhoneService ：是否开通电话服务业务 （Yes or No）

MultipleLines：是否开通了多线业务（Yes 、No or No phoneservice 三种）

InternetService：是否开通互联网服务 （No，DSL 数字网络，fiber optic 光纤网络 三种）

OnlineSecurity：是否开通网络安全服务（Yes，No，No internetserive 三种）

OnlineBackup：是否开通在线备份业务（Yes，No，No internetserive 三种）

DeviceProtection：是否开通了设备保护业务（Yes，No，No internetserive 三种）

TechSupport：是否开通了技术支持服务（Yes，No，No internetserive 三种）
StreamingTV：是否开通网络电视（Yes，No，No internetserive 三种）
StreamingMovies：是否开通网络电影（Yes，No，No internetserive 三种）
Contract：签订合同方式 （按月，一年，两年）
PaperlessBilling：是否开通电子账单（Yes or No）
PaymentMethod：付款方式（bank transfer，credit card，electronic check，mailed check）
MonthlyCharges：月费用
TotalCharges：总费用
研究对象
Churn：该用户是否流失（Yes or No）
任务一：数据信息处理（20 分）
①.读取给定的数据集
②.查看数据行列名
③.查看前五行
④.查看缺失值数量
任务二：数据分析（30 分）
①. 填充 TotalCharges 列中位数填充
②.’Churn’列重新编码“Yes”=1，“No“=0（map 函数）
③.用 gender 列和 PhoneService 列画出直方图
④.用 gender 列和 MultipleLines 列画出直方图
⑤.用 gender 列和 InternetService 画出直方图
⑥.用 gender 列和 OnlineSecurity 列画出直方图
任务三：模型预测（30 分）
①.用MonthlyCharges：月费用列预测Churn：该用户是否流失，训练和测试集的比例7：3
②.使用逻辑回归模型，配合网格搜索交叉验证，找到最优参数，参数自定
③.输出准确率和预测值
（2）作品提交要求见本模块附录 1
（3）实施条件要求见本模块附录 2
（4）评价标准见本模块附录 3

10.试题编号：4-10 Video Game Sales电子游戏销售分析模块

（1）任务描述
项目背景
vgsales 是由 vgchartz.com 的一个刮版生成的，是电子游戏行业综合销售数据，希望通过分析电子游戏行业在全球的发展概况，产生一份综合的游戏行业报告。
分析目的
从市场角度： 探究近几十年来电子游戏市场的发展趋势。
从平台角度： 探究用户最喜欢的游戏平台 top10 是什么，近些年的趋势有什么变化？
从类型角度： 探究用户最喜欢的游戏类型 top10 是什么，近些年的趋势有什么变化？
从发行商角度： 探究电子游戏发行商 top10 的销售情况以及近些年来的总体变化。（分别从销售额和发行量角度）
从排行榜角度： 对排行榜前 100 的电子游戏属性进行总结

文件名称	说明	包含特征	特征对应中文名称
vgsales.csv	电子游戏销售数据	Rank、Name、Platform、Year、Genre、Publisher、NA_Sales、EU_Sales、JP_Sales Other_Sales、Global_Sales	排名、游戏名、平台、发行年份、类型、发行商、NA销售额、EU销售额、JP销售额、其他地区销售额、总销售额

数据集的每一行表示一条用户行为，由排名、游戏名、平台、发行年份、类型、发行商、NA 销售额、EU 销售额、JP 销售额、其他地区销售额、总销售额组成，并以逗号分隔。关于数据集中每一列的详细描述如下：

排名 整数类型，序列化后的排名

游戏名 字符串，游戏名称

平台 字符串，该游戏发行平台名称

发行年份 浮点型，该游戏发行的日期

类型 字符串，该游戏的类型

发行商 字符串，该游戏的发行商名称

NA 销售额 浮点型，小数点后有效数字为 2 位，该游戏北美销售额(百万)

EU 销售额 浮点型，小数点后有效数字为 2 位， 该游戏欧洲销售额(百万)

JP 销售额 浮点型，小数点后有效数字为 2 位，该游戏日本销售额(百万)

其他地区销售额 浮点型，小数点后有效数字为 2 位，该游戏世界其他地区销售额(百万)

总销售额 浮点型 ，小数点后有效数字为 2 位，该游戏全球销售总额(百万)。

任务一：数据信息处理（20 分）

①.读取给定的数据集

②.查看数据行列名

③.查看前五行

④.查看缺失值数量

任务二：数据分析（30 分）

①.列字段重新命名

②.把缺失值填充为 0

③.画出 NA 销售额和总销售额的折线图

④.画出 EU 销售额和总销售额的折线图

⑤.画出 JP 销售额和总销售额的折线图

⑥.画出其他地区销售额和总销售额的折线图，并合并 NA 销售额，EU 销售额，JP 销售额，其他地区销售额为一张新表，名字为 New_game

任务三：模型预测（30 分）

①.用New_game表预测总销售额，训练和测试集的比例7：3

②.使用L1模型，配合网格搜索交叉验证，找到最优参数，参数自定

③.输出准确率和预测值

（2）作品提交要求见本模块附录 1

（3）实施条件要求见本模块附录 2

（4）评价标准见本模块附录 3

数据挖掘与机器学习附录

附录1作品提交

答案以“答题文件”的形式提交。请按以下要求创建答题文件夹和答题文件：

①创建答题文件夹

创建以”考生号_题号”命名的文件夹，存放所有答题文件，例如：“340103**********_4_1\”。

②创建答题文件

a.项目源文件

创建任务子文件夹，如：“340103**********_4_1\task\”，存放任务所有结果。

b.页面截图文件

在任务子文件夹中，存放截图.doc文件，它用于保存安装，配置，启动，运行执行过程中的屏幕截图，每张截图中每个关键配置或结果等，必须用红色矩形框标识出来并加以文字说明。

③提交答题文件

将”考生号_题号”文件夹打包，形成“考生号_题号.RAR”文件，如：“340103**********_4_1.rar”，将该文件按要求进行上传。

④考核时量

考核时间为180分钟。

附录 2 实施条件

表1 考点提供的主要设备及软件表

序号	场地、设备、软件名称	规格/技术参数、用途	规格/技术参数、用途备注
1	人工智能实训机房	测试场地	保证参考人员有足够间距
2	计算机	CPU 酷睿 i5 以上，内存 4G以上 Win7/win10/linux 操作系统	用于软件开发和软件部署，每 人一台
3	Pycharm软件，环境3.8以上	Python开发	参考人员环境选择和软件选择正确
4	Numpy,pandas,matplotlib Seaborn,sklearn	用于数据挖掘库	考试前把依赖库都安装完成

附录 3 评价标准

表 2 任务一数据信息处理（20 分）

序号	评分项	分值	评分细则
1	读取给定的数据集	5 分	要求开发的环境为python==3.8以上，否则扣 5 分。
2	查看数据行列名	5 分	查看行名5分，查看列名5分

3	查看前五行	5 分	没有正确输出扣5分
4	查看缺失值数量	5 分	缺失值数量错误扣 10 分

表 3 任务二数据分析（30 分）

序号	评分项	分值	评分细则
1	字段处理	30 分	字段处理错误一个扣5分，共6题

表 4 任务三模型预测（30 分）

序号	评分项	分值	评分细则
1	训练测试集选择正确	10 分	训练测试集合按要求的列名选择，否则扣10分
2	选择要求的模型训练	5 分	模型选择不对扣 5 分；
3	要求用网格交叉验证找参数	5 分	要求用交叉验证找参数否则扣5分
4	输出准确率和预测值	10分	准确率5分，预测值5分

表 5 职业素养（10 分）

序号	评分项	分值	评分细则
1	代码书写格式规范	5 分	代码使用驼峰命名法，未使用驼峰命名法扣1分，扣完为止
2	注释规范	3 分	整个项目没有注释扣 3 分、有注释，但注释不规范扣 1 分，扣完为止。
3	运行正确	2 分	代码无报错

表 6 实操文档（10 分）

序号	评分项	分值	评分细则
1	实操文档有无	2分	有实操文档得分，无实操文档扣2分。
2	文档任务截图	4分	有操作过程截图得分，无操作过程截图扣4分。
3	文档任务截图标注	4分	有文档任务截图标注说明和画框得分，无标注和画框扣4分