脚本生成数据库需求

1. 项目背景

在学习数据库系统原理时，涉及到很多表的使用，包括增、删、改、查、创建等等，想要练习需要有大量的测试数据用于练习，因此需要有一种方法快速的批量的生成数据。

1. 项目目标
2. 使用python编写函数或类，实现数据的生成。
3. 功能描述
4. Instructor表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 列名 | 数据类型 | 长度 | 字段说明 | 创建说明 |
| ID | Integer |  | 教师的ID，唯一识别教师 | 0-20000，随机数字，避免重复 |
| Name | Char | 20 | 教师姓名 | 文件存储姓名库，读取后解析到列表中 |
| Dept\_name | Char | 20 | 系院名称 | 文件存储20个院系名称，读取后解析到列表 |
| Salary | integer |  | 教师年薪年薪 | 10000-20000，随机整数，避免重复 |

1. Department表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 列名 | 数据类型 | 长度 | 字段说明 | 创建说明 |
| Dept\_name | Char | 20 | 系院名称 | 文件存储20个院系名称，读取后解析到列表 |
| Building | Char | 15 | 大楼名称 | 文件存储20个大楼名称，读取后解析到列表 |
| budget | Integer |  | 预算费用 | 200000-300000，间隔500，随机整数 |

1. Create all tables

create table classroom

(building varchar(15),

room\_number varchar(7),

capacity numeric(4,0),

primary key (building, room\_number)

);

create table department

(dept\_name varchar(20),

building varchar(15),

budget numeric(12,2) check (budget > 0),

primary key (dept\_name)

);

create table course

(course\_id varchar(8),

title varchar(50),

dept\_name varchar(20),

credits numeric(2,0) check (credits > 0),

primary key (course\_id),

foreign key (dept\_name) references department

on delete set null

);

create table instructor

(ID varchar(5),

name varchar(20) not null,

dept\_name varchar(20),

salary numeric(8,2) check (salary > 29000),

primary key (ID),

foreign key (dept\_name) references department

on delete set null

);

create table section

(course\_id varchar(8),

sec\_id varchar(8),

semester varchar(6)

check (semester in ('Fall', 'Winter', 'Spring', 'Summer')),

year numeric(4,0) check (year > 1701 and year < 2100),

building varchar(15),

room\_number varchar(7),

time\_slot\_id varchar(4),

primary key (course\_id, sec\_id, semester, year),

foreign key (course\_id) references course

on delete cascade,

foreign key (building, room\_number) references classroom

on delete set null

);

create table teaches

(ID varchar(5),

course\_id varchar(8),

sec\_id varchar(8),

semester varchar(6),

year numeric(4,0),

primary key (ID, course\_id, sec\_id, semester, year),

foreign key (course\_id,sec\_id, semester, year) references section

on delete cascade,

foreign key (ID) references instructor

on delete cascade

);

create table student

(ID varchar(5),

name varchar(20) not null,

dept\_name varchar(20),

tot\_cred numeric(3,0) check (tot\_cred >= 0),

primary key (ID),

foreign key (dept\_name) references department

on delete set null

);

create table takes

(ID varchar(5),

course\_id varchar(8),

sec\_id varchar(8),

semester varchar(6),

year numeric(4,0),

grade varchar(2),

primary key (ID, course\_id, sec\_id, semester, year),

foreign key (course\_id,sec\_id, semester, year) references section

on delete cascade,

foreign key (ID) references student

on delete cascade

);

create table advisor

(s\_ID varchar(5),

i\_ID varchar(5),

primary key (s\_ID),

foreign key (i\_ID) references instructor (ID)

on delete set null,

foreign key (s\_ID) references student (ID)

on delete cascade

);

create table time\_slot

(time\_slot\_id varchar(4),

day varchar(1),

start\_hr numeric(2) check (start\_hr >= 0 and start\_hr < 24),

start\_min numeric(2) check (start\_min >= 0 and start\_min < 60),

end\_hr numeric(2) check (end\_hr >= 0 and end\_hr < 24),

end\_min numeric(2) check (end\_min >= 0 and end\_min < 60),

primary key (time\_slot\_id, day, start\_hr, start\_min)

);

create table prereq

(course\_id varchar(8),

prereq\_id varchar(8),

primary key (course\_id, prereq\_id),

foreign key (course\_id) references course

on delete cascade,

foreign key (prereq\_id) references course

);

1. 问题记录
2. 从文件存储的姓名库、院系名称库中提取姓名、院系名称、建筑名称等到列表中。

编写了get\_name函数用于提取，在提取姓名时遇到问题，中英文混合情况下，如何提取每行第一个的英文姓名单词。

解决方案：使用re模块，利用正则表达式解析出姓名

=========================================

For line in f:

Obj=re.match(r’\w+’,line,re.I)

Try:

Print obj.group()

List.appen(obj.group())

Except:

Continue

=========================================

1. 在生成每一条数据的时候，需要用到random模块，里面的函数比较有用
2. 在大量循环时，sql=sql % param会产生错误，原因不明，怀疑是在自身匹配时无法及时处理，所以改成sql=base % param，不要通过自身format再将值给自身，会有问题。