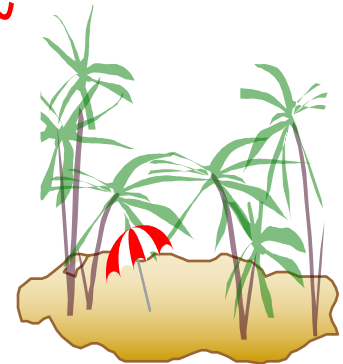


第四部分 数据库设计(Database Design)

section 1 需求分析 (requirements)

- 信息科学与工程学院 杨金民
- 2017.11



目录提纲

1. 数据库设计的具体含义是什么？
2. 数据库设计要解决什么问题？
3. 数据库设计面临什么挑战？
4. 数据库设计为什么是一门专业知识；
5. 数据库设计中的问题：
 - 如何来设计？设计方法，流程，环节，要素；
 - 对提交的一个设计，如何来判断它是否合理？

设计动机：正常运行时的数据正确性

是通过数据库设计来解决的，数据库设计要回答的三个问题：

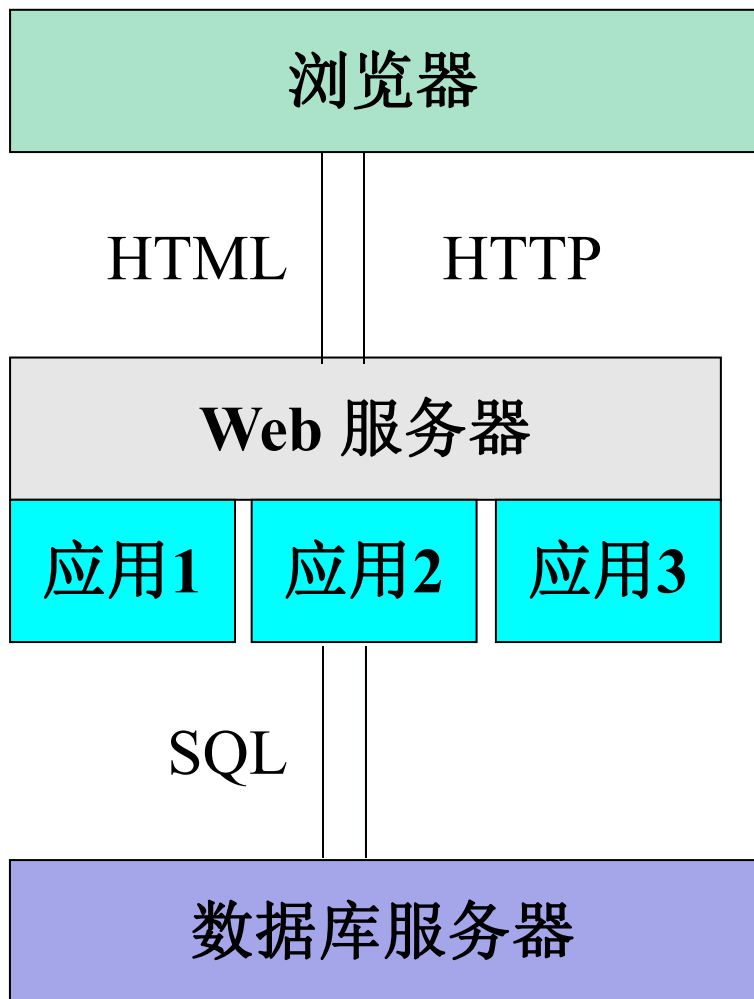
- 1) **覆盖性问题**：确定一个单位有哪些数据项（字段）？
- 2) **划分问题**：确定哪些字段构成一个表？
- 3) **关系问题**：确定表之间有什么关系？

挑战：

- 1) **业务表单**和数据库中**表**的**不一致**；
 - 业务表单中通常包含的是**综合信息**；
 - 而数据库中的一个表只能存储**单一类别的信息**；
- 2) 见到的和听到的是**局部**，而数据库设计要站在**全局**来考虑；

如果直接把业务表单中的数据项组成一个表，会带来一系列的数据正确性问题；

信息系统架构



//1) **加载数据库访问驱动程序:**

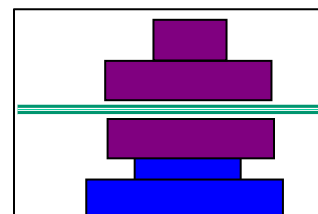
```
Class.forName("com.mysql.jdbc.Driver");
```

//2) **建立与数据库的链接:**

```
Connection connection =
```

```
DriverManager.getConnection("jdbc:mysql://  
192.168.105.100:3306/education","root","admin");
```

对接

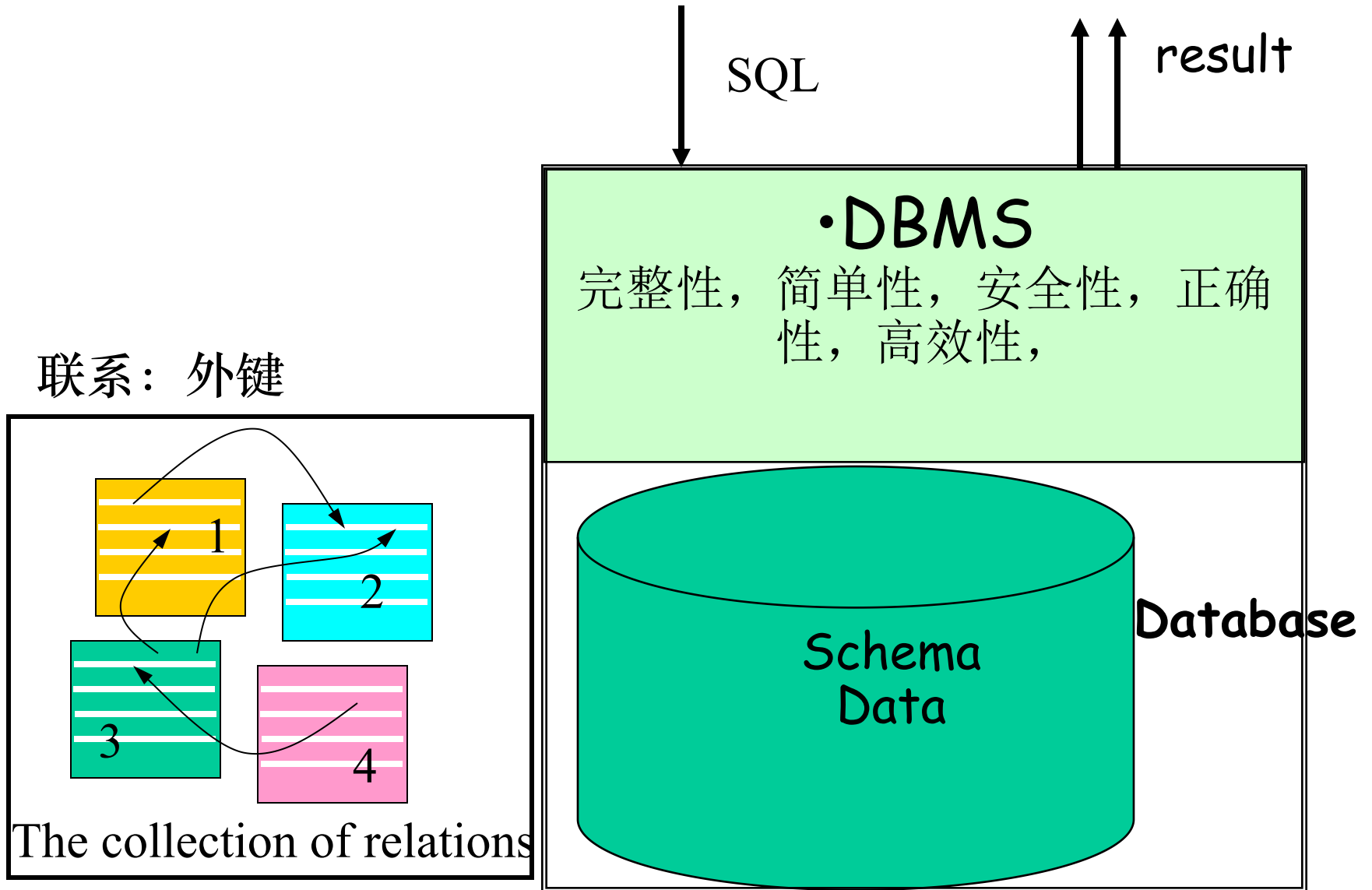


ODBC/JDBC
驱动程序

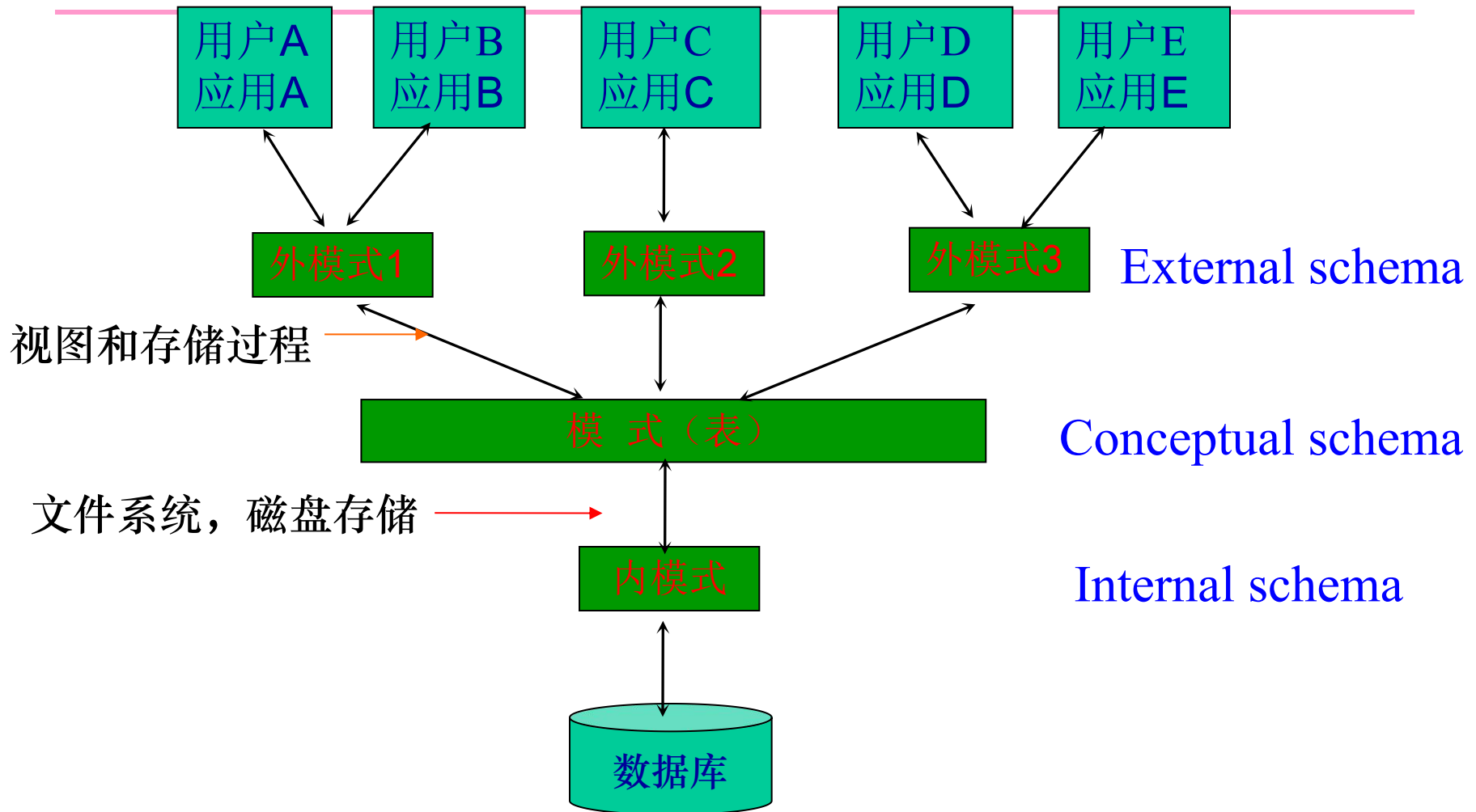
DBMS 1

Database 1

数据库特性，对数据库的访问 SQL



DBMS 的三级模式 (Schema) 架构



三级模式结构及二级映像实现了数据库系统的数据独立性

软件危机

- 软件的使用过程中，遇到的困惑和问题：
- 软件的**Bugs**和**faults**；
- **新的需求**不能**轻量级地满足**：快捷，简单，低成本；
- 通常的情况是补充和修改后，原有的功能被影响，其它部分出现问题。被沦陷进**问题重重，没完没了的梦魇**中；
- **数据正确性**问题（更新异常），**数据一致性**问题；
- 数据的**伸展性**问题。数据的推算、衍生，与合成，以免数据数据不正确，数据不一致问题，残缺不全；
- **性能问题，系统扩展问题，升级问题**；

设计水平问题

❖ 最突出得表现：**孤立地从局部来看待问题**，而不是用**联系的观点通盘看待问题**；

- 设计时考虑不周。例子：选课，一般情况是选择一次；往往忽视了重修。当单独来考虑重修和设计重修时，往往又忽视了选课和重修的关系。当某天用户提出要关联选课和重修时，这时要改动数据库，一旦改动了数据库，由于原有的应用和数据库的紧耦合性，原有应用执行失败；
- 数据要有伸缩性。能够基于现有数据衍生出各种新的数据视图，服务于不断发展的业务；例如，教师的工作量；

软件可靠性的体现

- 程序的**鲁棒性**：

- 1) 函数调用是否成功；
- 2) 结果是否为空；
- 3) 不为空时，才允许读结果；

- 数据的**正确性**；**完整性**；**安全性**；操作的**简单性**；数据处理的高**效性**；

- 系统的**可维护性**：系统的**模块化**，**层次化**，接口的**标准化**：**邦联性**，**可组合性**；

数据库设计面临的挑战

➤ 统计表明：

- ① 80 - 90% 的项目不能满足性能要求；
- ② 80%的项目延期提交，超出预算；
- ③ 40%的项目以失败或者放弃而告终；
- ④ 只有10-20%的项目取得成功；

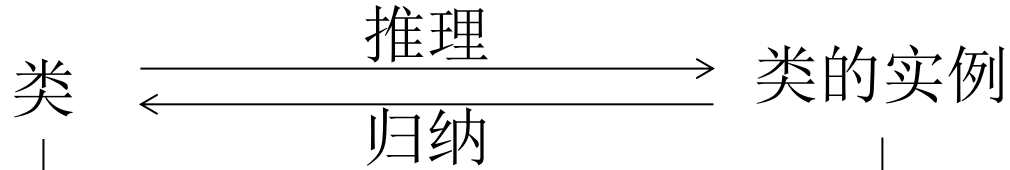
➤ 数据库设计是地基，至关重要；

➤ 不成功的原因：

- 1) 没有深刻领会**数据库的特征与特性**；
- 2) 对**业务需求**，**业务特性与特征**的把握不到位/全面/深刻；
- 3) 未严格按照数据库**设计方法学**，流程，环节，要素；
- 4) 没有全面掌握数据库设计的**技能**；

关系型数据库的特点

概念层面



表现层面

表的模式(Schema)

表的数据(data)

一个类对应一张表

一行数据对应一个实例

一张表对应一个类

一个实例对应一行数据

回答了数据的单位概念，数据的量的概念

一个数据只存一份

数据的区分概念?

表的特点：从列来看，即从类的概念来看，凡是要用到的属性，都要包含；
从行来看，所有实例对象都在表中；

Student

Name	studentNo	sex	nation	dept_no	birthday	phone
周山	2008043101	男	汉	420	2000/12/01	15107312332
汪兵	2008043214	女	回	420	1998/06/12	15116247890
刘军	2009043331	男	藏	410	1998/08/01	139073196999
张珊	2009043332	女	藏	410	1998/09/09	139073196967

Name	studentNo	sex	nation	dept_no	birthday	phone
周山	2008043101	男	汉	420	2000/12/01	15107312332
汪兵	2008043214	女	回	420	1998/06/12	15116247890

Name	studentNo
周山	2008043101
汪兵	2008043214
刘军	2009043331
张珊	2009043332

从单个用户来看，他仅只须要部分列，部分行。

数据库中的表：严格按类概念，分表存储；

但是用户希望的业务表，要由分布在多个表中列组合而成；

Student

Name	studentNo	sex	dept_no
周山	2008043101	男	420
汪兵	2008043214	女	420
张珊	2008043332	女	410

Course

Name	courseNo	textbook	syllabus
database	430601	数据库	introduction
操作系统	430803	操作统	all
java	430317	语言	section

Enroll

studentNo	courseNo	semester	class	grade
2008043332	430803	2010/01	A	90
2008043101	430803	2009/01	A	56
2008043101	430317	2008/02	B	56
2008043332	430601	2004/02	A	77

成绩表Score

Name	studentNo	course	grade
张珊	2008043332	操作系统	90
周山	2008043101	操作系统	56
周山	2008043101	java	56
张珊	2008043332	database	77

业务表单和数据库中表的不一致性

学生表单

学院名称：软件学院 学院编号：01

地址：湖南大学软件大楼

姓名	学号	性别	出生日期
周强	2008043101	男	1990/12/14
汪涵	2008043214	男	1992/02/21
张珊	2008043332	女	1988/07/09
刘丽	2008043315	女	1988/01/29

院长：林亚平

该表单包含的信息：

学生信息；
学院信息；

不正确的数据库设计

student-department

Name	studentNo	sex	birthdate	deptName	deptNo	deptDean	Address
张强	2008043101	男	1990/12/14	软件学院	01	林亚平	软件大楼
汪涵	2008043214	女	1992/02/21	金融学院	03	杨胜刚	红叶楼
张珊	2008043332	女	1988/07/09	会计学院	04	黄立红	逸夫楼
刘丽	2008043315	女	1988/01/29	会计学院	04	黄立红	逸夫楼
李娜	2008043358	女	1988/11/13	软件学院	01	林亚平	软件大楼
张军	2008043375	男	1988/09/09	软件学院	01	林亚平	软件大楼
谢莎	2008043398	女	1988/09/09	软件学院	01	林亚平	软件大楼

带来的问题1：数据冗余问题

student-department

Name	studentNo	sex	birthdate	deptName	deptNo	deptDean	Address
周强	2008043101	男	1990/12/14	软件学院	01	林亚平	软件大楼
汪涵	2008043214	女	1992/02/21	金融学院	03	杨胜刚	红叶楼
张珊	2008043332	女	1988/07/09	会计学院	04	黄立红	逸夫楼
刘丽	2008043315	女	1988/01/29	会计学院	04	黄立红	逸夫楼
李娜	2008043358	女	1988/11/13	软件学院	01	林亚平	软件大楼
张军	2008043375	男	1988/09/09	软件学院	01	林亚平	软件大楼
谢莎	2008043398	女	1988/09/09	软件学院	01	林亚平	软件大楼

问题： 1) 数据冗余问题：例如，软件学院信息多个地方重复；

解决方法：数据库设计

带来的问题2：无形中的数据丢失

student-department

Name	studentNo	sex	birthdate	deptName	deptNo	deptDean	Address
周强	2008043101	男	1990/12/14	软件学院	01	林亚平	软件大楼
汪涵	2008043214	女	1992/02/21	金融学院	03	杨胜刚	红叶楼
张珊	2008043332	女	1988/07/09	会计学院	04	黄立红	逸夫楼
刘丽	2008043315	女	1988/01/29	软件学院	01	林亚平	软件大楼
李娜	2008043358	女	1988/11/13	软件学院	01	林亚平	软件大楼
张军	2008043375	男	1988/09/09	会计学院	04	黄立红	逸夫楼
谢莎	2008043398	女	1988/09/09	软件学院	01	林亚平	软件大楼

问题： 2) 删除问题：例如，删除一个学院信息时，附带也把其学生信息也删除了；

解决方法：数据库设计

带来的问题3： 数据添加被抵制

student-department

Name	studentNo	sex	birthdate	deptName	deptNo	deptDean	Address
周强	2008043101	男	1990/12/14	软件学院	01	林亚平	软件大楼
汪涵	2008043214	女	1992/02/21	金融学院	03	杨胜刚	红叶楼
张珊	2008043332	女	1988/07/09	会计学院	04	黄立红	逸夫楼
刘丽	2008043315	女	1988/01/29	会计学院	04	黄立红	逸夫楼
李娜	2008043358	女	1988/11/13	软件学院	01	林亚平	软件大楼
张军	2008043375	男	1988/09/09	软件学院	01	林亚平	软件大楼
谢莎	2008043398	女	1988/09/09	软件学院	01	林亚平	软件大楼
				量子计算学院	59	毛先生	*****

问题： 3) 插入问题： 例如，新建一个学院信息时，由于还没有学生，出现插入不成功的问题；

解决方法： 数据库设计

带来的问题4：数据不一致

student-department

Name	studentNo	sex	birthdate	deptName	deptNo	deptDean	Address
周强	2008043101	男	1990/12/14	软件学院	01	林亚平	软件大楼
汪涵	2008043214	女	1992/02/21	金融学院	03	杨胜刚	红叶楼
张珊	2008043332	女	1988/07/09	会计学院	04	黄立红	逸夫楼
刘丽	2008043315	女	1988/01/29	会计学院	04	黄立红	逸夫楼
李娜	2008043358	女	1988/11/13	软件学院	01	林亚平	软件大楼
张军	2008043375	男	1988/09/09	软件学院	01	林亚平	软件大楼
谢莎	2008043398	女	1988/09/09	软件学院	01	林亚平	软件大楼

问题： 4) 修改问题：例如，修改软件学院的地址时，要修改多行记录，而不是一条记录；

解决方法：数据库设计

带来的问题5：数据结果错误

student-department

Name	studentNo	sex	birthdate	deptName	deptNo	deptDean	Address
周强	2008043101	男	1990/12/14	软件学院	01	林亚平	软件大楼
汪涵	2008043214	女	1992/02/21	金融学院	03	杨胜刚	红叶楼
张珊	2008043332	女	1988/07/09	会计学院	04	黄立红	逸夫楼
刘丽	2008043315	女	1988/01/29	会计学院	04	黄立红	逸夫楼
李娜	2008043358	女	1988/11/13	软件学院	01	林亚平	软件大楼
张军	2008043375	男	1988/09/09	软件学院	01	林亚平	软件大楼
谢莎	2008043398	女	1988/09/09	软件学院	01	林亚平	软件大楼

问题： 5) 统计问题：例如，统计有多少个学院时，由于信息重复出现使得统计困难；

解决方法：数据库设计

正确的数据库设计

student

name	studentNo	sex	birthdate	deptNo
周强	2008043101	男	1990/12/14	01
汪涵	2008043214	女	1992/02/21	03
张珊	2008043332	女	1988/07/09	04
刘丽	2008043315	女	1988/01/29	04
李娜	2008043358	女	1988/11/13	01
张军	2008043375	男	1988/09/09	01
谢莎	2008043398	女	1988/09/09	01

department

deptName	deptNo	Dean	Address
软件学院	01	林亚平	软件大楼
金融学院	03	杨胜刚	红叶楼
会计学院	04	黄立红	逸夫楼

引用



不存在上述5种问题全部自然消失；

业务表单：综合信息(一)

课程名称：高级数据库技术 课程编号：001

学 期：2010/01

姓名	学号	成绩
周强	2008043101	78
汪涵	2008043214	65
张珊	2008043332	81
刘丽	2008043315	92

上课老师：张三

该表单包含的信息：

学生信息；
课程信息；
老师信息；

学生选课信息；

教师教课信息；

业务表单：综合信息(二)

学院：软件学院

学生姓名：张三 学号：001

课程名称	课程编号	成绩	学分
数据库技术	2008043101	78	3
程序设计	2008043214	65	2
信息检索	2008043332	81	1
计算机网络	2008043315	92	2

该表单包含的信息：

学生信息；
课程信息；
学院信息；

学生选课信息；

不同表中记录之间的关系

faculty

Name	staffNo	sex	birthdate	deptNo
林亚平	101	男	1990/12/14	01
李娜	758	女	1988/11/13	01
谢莎	498	女	1988/09/09	01
杨胜刚	214	男	1992/02/21	03
刘丽	315	女	1988/01/29	04
张军	175	男	1988/09/09	01
黄立红	332	女	1988/07/09	04

department

deptName	deptNo	Dean	Address
软件学院	01	101	软件大楼
金融学院	03	214	红叶楼
会计学院	04	332	逸夫楼

1:1关系：一个院系对应一个院长，一个院长对应一个学院；

不同表中记录之间的关系

teacher

Name	staffNo	sex	birthdate	deptNo
林亚平	101	男	1990/12/14	01
李娜	758	女	1988/11/13	01
谢莎	498	女	1988/09/09	01
杨胜刚	214	男	1992/02/21	03
刘丽	315	女	1988/01/29	04
张军	175	男	1988/09/09	01
黄立红	332	女	1988/07/09	04

department

deptName	deptNo	DeanNo	Address
软件学院	01	101	软件大楼
金融学院	03	214	红叶楼
会计学院	04	332	逸夫楼

1:m关系：一个院系有多个教师；

不同表中记录之间的关系

student

Name	studentNo	sex	birthdate
周强	2008043101	男	1990/12/14
汪涵	2008043214	女	1992/02/21
张珊	2008043332	女	1988/07/09

course

Name	courseNo	textbook	syllabus
database	430601	数据库	introduction
operating system	430803	操作统	all
java	430317	语言	section

take_course

studentNo	courseNo	semester	grade
2008043332	430803	2010/01	90
2008043435	430317	2009/01	56
2008043101	430317	2008/02	82
2008043332	430007	2004/02	77

n:m关系：一门课程可以有多个学生来选择，
一个学生可以选择多门课程；

关系型数据库的实质

表与表之间存在联系，这就是为什么叫关系型数据库。

正是不同表中记录之间存在关系，沿着这种关系就能够把它们综合起来，构建出业务表单，满足业务需求；

设计合理性的验证标准

3个范式(Normal Form):

第一范式：一个表的任一个字段的取值必须是原子类型

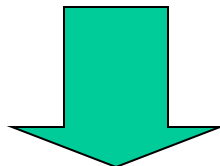
第二范式：一个表中除了主关键字段外的任何一个字段必须完全函数依赖于主关键字段；

第三范式：一个表中除了主关键字段外的任何一个字段不允许间接依赖于主关键字段；

基于范式一对表进行规范化例子

textbook

title	Author-set	Publisher (name, branch)	Keyword-set
compiler	{smith, John}	(Oxford, Beijing)	{word, sentence, syntax}
network	{Jack, Smith}	(Amason, Changsha)	(transport, fault-tolerance)
database	{Jim, Tom, Phillipe}	(Greatwall, Changsha)	{relation, record, foreign key}



基于范式一对表进行规范化

title	Author
compiler	smith
compiler	John
network	Jack
network	Smith
database	Jim
database	Tom
database	Phillipe

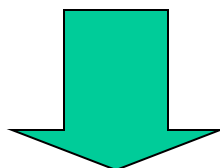
title	Keyword
compiler	word
compiler	sentence
compiler	syntax
network	transport
network	fault-tolerance
database	relation
database	record
database	foreign key

title	publisher	branch
compiler	Oxford	Beijing
network	Amason	Shanghai
database	Greatwall	Changsha

基于范式二对表进行规范化例子

student-department

Name	studentNo	sex	birthdate	deptName	deptNo	deptDean	Address
张强	2008043101	男	1990/12/14	软件学院	01	林亚平	软件大楼
汪涵	2008043214	女	1992/02/21	金融学院	03	杨胜刚	红叶楼
张珊	2008043332	女	1988/07/09	会计学院	04	黄立红	逸夫楼
刘丽	2008043315	女	1988/01/29	会计学院	04	黄立红	逸夫楼
李娜	2008043358	女	1988/11/13	软件学院	01	林亚平	软件大楼
张军	2008043375	男	1988/09/09	软件学院	01	林亚平	软件大楼
谢莎	2008043398	女	1988/09/09	软件学院	01	林亚平	软件大楼



基于范式对表进行规范化例子

student

Name	studentNo	sex	birthdate	deptNo
周强	2008043101	男	1990/12/14	01
汪涵	2008043214	女	1992/02/21	03
张珊	2008043332	女	1988/07/09	04
刘丽	2008043315	女	1988/01/29	04
李娜	2008043358	女	1988/11/13	01
张军	2008043375	男	1988/09/09	01
谢莎	2008043398	女	1988/09/09	01

department

deptName	deptNo	Dean	Address
软件学院	01	林亚平	软件大楼
金融学院	03	杨胜刚	红叶楼
会计学院	04	黄立红	逸夫楼

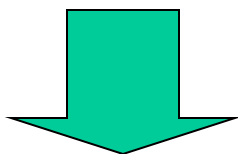
引用



基于范式三对表进行规范化例子

teacher

Name	teacherNo	sex	birthdate	rank	salary	deptNo
周强	2008043101	男	1990/12/14	副教授	5000	01
汪涵	2008043214	女	1992/02/21	教授	6000	03
张珊	2008043332	女	1988/07/09	讲师	3000	04
刘丽	2008043315	女	1988/01/29	院士	9000	04
李娜	2008043358	女	1988/11/13	教授	6000	01
张军	2008043375	男	1988/09/09	讲师	3000	01
王五	2008043313	男	1988/09/09	副教授	5000	01
赵六	2008043232	男	1988/09/09	讲师	3000	01
谢莎	2008043657	女	1988/09/09	副教授	5000	01



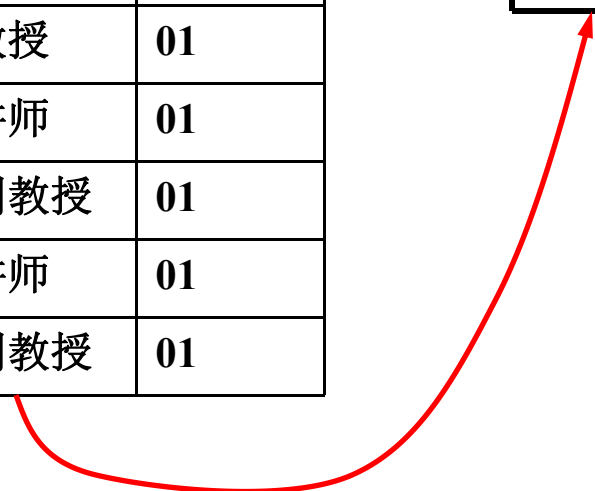
基于范式三对表进行规范化例子

teacher

Name	teacherNo	sex	birthdate	rank	deptNo
周强	2008043101	男	1990/12/14	副教授	01
汪涵	2008043214	女	1992/02/21	教授	03
张珊	2008043332	女	1988/07/09	讲师	04
刘丽	2008043315	女	1988/01/29	院士	04
李娜	2008043358	女	1988/11/13	教授	01
张军	2008043375	男	1988/09/09	讲师	01
王五	2008043313	男	1988/09/09	副教授	01
赵六	2008043232	男	1988/09/09	讲师	01
谢莎	2008043657	女	1988/09/09	副教授	01

salary_level

rank	salary
副教授	5000
教授	6000
讲师	3000
院士	9000



数据的延伸性

➤ 数据要有伸缩性。能够基于现有数据衍生出各种新的数据视图，服务于不断发展的业务；



➤ Hotel (hotelNo, name, address)

➤ Room (roomNo, hotelNo, type, price)

➤ Booking (hotelNo, guestNo, dateFrom, dateTo, roomNo)

➤ Guest (guestNo, name, address)

- 1) 每个月的营业收入是多少？
- 2) 各个宾馆的入住率是多少？
- 3) 各种房间的入住率是多少？
- 4) 公司的常客是哪些人？

专业知识

- ❖ 需求发现技巧;
- ❖ 不合理的数据库设计会引发如下问题:
 - ❖ 数据冗余;
 - ❖ 更新异常;
- ❖ 涉及必须基于业务特性, 迎合业务特性, 覆盖业务需求;
- ❖ 设计方法学:
 - ❖ 需求表述; ;
 - ❖ ER 建模;
 - ❖ 将ER model 映射成relation model;
 - ❖ 采用一个数据库管理系统, 实现一个数据库, 搭建起一个数据库服务器;

基于业务特性来设计数据库

对于房产中介公司的数据库，其中的客户看房记录表viewing:

PropertyNo	ClientNo	ViewDate	ViewTime	Comments
PA14	CR62	1-Jun-01	10:00	goods
PL94	CR74	1-Aug-01	16:30	preference
PG21	CR62	1-Jul-01	9:10	expensive

- **问题：**
- **1) viewTime 这个字段须要吗？即使用户并不关注它，但是如果公司业务规定不能两个客户同时去看同一个房源，那么这个字段就必要？**
- **2) propertyNo 和 clientNo 能作为该表的主键吗？**

数据库设计的第一步：获取需求

- 数据库设计者了解业务：
 - 1) things (entity (事实));
 - 2) activity/event(活动/事件)
 - 3) involved persons(相关人员);
 - 4) procedure(过程)
 - 5) record (记录);
 - 6) 业务特征与特性，发展态势;

了解业务的技巧

五个获取业务需求的途径：

- 1) 查看业务规章制度的文档;
- 2) 收集已有的业务表单;
- 3) 面对面沟通交流;
- 4) 观察业务的运转情况;
- 5) 分析, 推理, 研究, 归纳提炼;

了解系统, 需求, 偏好, 诉求;

获取需求中对数据要关注的重点

- 1) 标识性数据；
- 2) 数据间的关系，是否存在关联性和依赖性，独立性；
- 3) 数据的域，业务规则；
- 4) 数据量，增长特性，访问特性；

随堂测试

1. 数据库中的表有什么特征和特性?
2. 数据库设计要解决什么问题，这些问题具体体现在哪些方面?
3. 数据库设计面临什么挑战?
4. 数据库设计包括哪两个方面?
5. 获取业务需求的五项途径是什么?

需求分析

- **需求分析报告:** 以文本或者图形方式来陈述企业的业务, 以及其中的处理流程, 记录, 及其流转情况。 .
- **优点:** 是易于理解, 方便交流。在认识和把握企业特性, 业务, 过程, 记录, 数据等方面达成共识。以求全面, 准确的认识和理解企业。
- **不足:** 在结构和关系上模糊, 不精准, 让计算机处理难; ;
- 例如: 1) 6除2; 2) 中国队大胜美国队, 中国队大败美国队;

Modeling

对一个企业的数据进行结构化处理，准确地表达数据的内涵；

基本要素：

- 概念及其定义；
- 表达概念所用的符号；
- 有关处理的规则；

例如： $(6 \div 2 + 3) \times 4$ ；

数据库设计的过程

(门槛越低)

面向人

需求陈述

通过建模来结构化

概念数据库设计

ER Modeling

目标模型

逻辑数据库设计

验证

目标数据库

物理数据库设计

- 表
- 主键;
- 外键;

- ✓ 完整性;
- ✓ 安全;
- ✓ 性能;

面向计算机处理
(更多的专业知识)

数据库设计的过程

三步曲：

为什么要三步，而不是一步？

概念数据库设计；从业务需求采用高级建模方法建模；

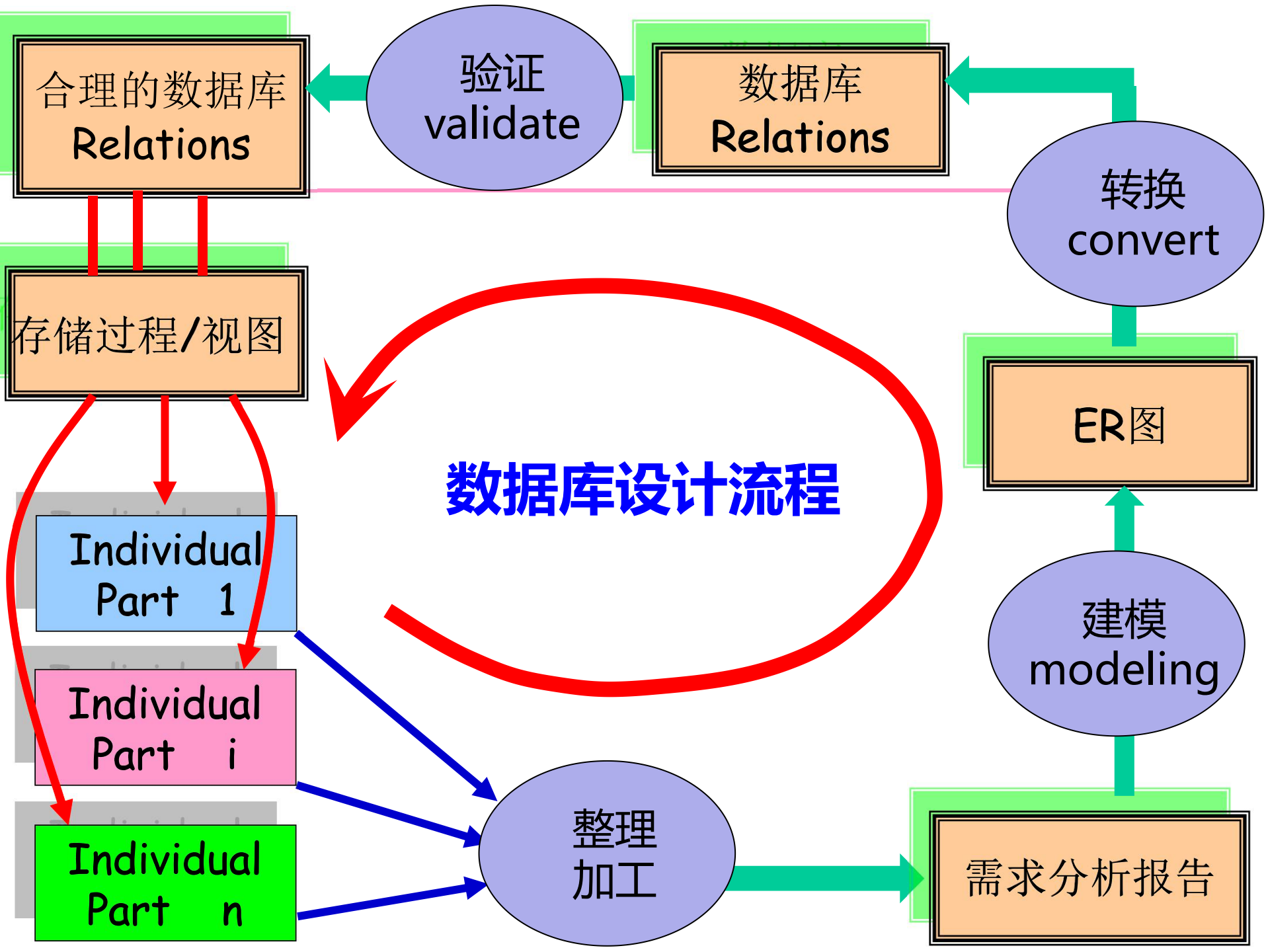


逻辑数据库设计；将高级模型转换为关系模型；



物理数据库设计；采用特定的DBMS产品实现数据库；

数据库设计的过程



规章制度
业务表单
现场观察
调查问卷
当面交流

识别/归纳/抽象
推理/提炼

建模
Modeling

ER图

需求分析报告

调查
分析
研究

整理
加工

业务详情：
● 事情 / 实体；
● 活动 / 事件；
● 过程和环节；
● 业务表单 / 记录；
● 人员；
● 特征和趋势；

Individual
Part 1

Individual
Part i

Individual
Part n

数据库设计要回答的三个问题

直接地说，有三个问题：

- 1) 覆盖性问题：确定一个单位有哪些数据项（字段）？
- 2) 划分问题：确定哪些字段构成一个表？
- 3) 关系问题：确定表之间有什么关系？

数据库设计的目标

- 没有异常；
- 输出的数据正确；
- 尽量少的冗余；
- 具有可伸缩性：只要数据库中存在的数据库项，任何业务需求都能满足，想要的数据库项都能组合出来；

数据库设计的挑战

挑战：

业务表单和数据库中表的不一致；

- 业务表单中通常包含的是综合信息；
- 而数据库中的一个表只能存储单一类的信息；

见到的和听到的是局部，而数据库设计要站在全局来考虑；

如果直接把业务表单中的数据项组成一个表，会带来一系列的数据正确性问题；