

a) 数据库系统基础概念

数据 (data), 数据库(database), 数据库管理系统 (DBMS), 数据库系统 (DBS)

一、 数据

1 数据的定义: 数据是描述事物的符号记录, 是数据库中存储的基本对象。

2 数据的种类: 文本、图形、图像、音频、视频、学生的档案记录、货物的运输情况等。

3 数据的特点: 数据与其语义是不可分的, 数据的含义称为数据的语义。

举例: 93 是一个数据

语义 1: 学生某门课的成绩

语义 2: 某人的体重

语义 3: 计算机系 2003 级学生人数

学生档案中的学生记录

(李明, 男, 197205, 江苏南京市, 计算机系, 1990)

语义: 学生姓名、性别、出生年月、籍贯、所在院系、 入学时间

解释: 李明是个大学生, 1972 年 5 月出生, 江苏南京市人, 1990 年考入计算机系

二、 数据库

1 数据库的定义: 数据库(Database,简称 DB)是长期储存在计算机内、有组织的、可共享的大量数据的集合。

2 数据库的基本特征: 数据按一定的数据模型组织、描述和储存

可为各种用户共享

冗余度较小

数据独立性较高

易扩展

三、 数据库管理系统 (DBMS)

1 数据库管理系统定义: 位于用户与操作系统之间的一层数据管理软件, 是基础软件, 是一个大型复杂的软件系统。

2 DBMS 的用途: 科学地组织和存储数据、高效地获取和维护数据。

3 DBMS 的主要功能: **数据定义功能**

提供数据定义语言(DDL)

定义数据库中的数据对象

**数据组织、存储和管理**

分类组织、存储和管理各种数据

确定组织数据的文件结构和存取方式

实现数据之间的联系

提供多种存取方法提高存取效率

**数据操纵功能**

提供数据操纵语言(DML)

实现对数据库的基本操作 (查询、插入、删除和修改)

**数据库的事务管理和运行管理**

数据库在建立、运行和维护时由 DBMS 统一管理和控制

保证数据的安全性、完整性、多用户对数据的并发使用

发生故障后的系统恢复

**数据库的建立和维护功能(实用程序)**

数据库初始数据装载转换

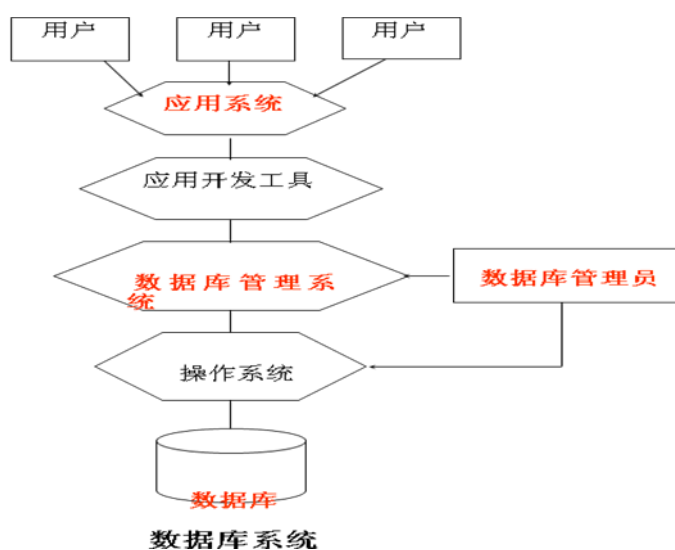
数据库转储  
介质故障恢复  
数据库的重组织  
性能监视分析等

#### 其它功能

DBMS 与网络中其它软件系统的通信  
两个 DBMS 系统的数据转换  
异构数据库之间的互访和互操作

#### 四、数据库系统（DBS）

- 1 数据库系统的定义：在计算机系统中引入数据库后的系统构成。
- 2 数据库系统的构成（理解）：数据库，数据库管理系统（及其开发工具），应用系统，数据库管理员。



#### 3 数据库系统的特点：数据结构化

数据的共享性高，冗余度低，易扩充  
数据独立性高  
数据由 DBMS 统一管理和控制

注：各个特点的具体介绍请看课件(第一张 ppt)

#### b) 数据模型（完全掌握并能够应用）

数据模型：在数据库中用数据模型这个工具来抽象、表示和处理现实世界中的数据和信息。

数据模型应满足三方面要求：能比较真实地模拟现实世界，容易为人所理解，便于在计算机上实现

数据模型分类：第一类是概念模型，第二类是逻辑模型和物理模型。

概念模型：也称信息模型，它是按用户的观点来对数据和信息建模，用于数据库设计。

逻辑模型：主要包括网状模型、层次模型、关系模型、面向对象模型等，按计算机系统的观点对数据建模，用于 DBMS 实现。

物理模型：是对数据最底层的抽象，描述数据在系统内部的表示方式和存取方法，在磁盘或磁带上的存储方式和存取方法。

数据模型的组成要素：数据结构，数据操作，完整性约束条件。

## 1 概念模型

概念模型的**用途**：概念模型用于信息世界的建模；是现实世界到机器世界的一个中间层次；是数据库设计的有力工具；数据库设计人员和用户之间进行交流的语言

对概念模型的**基本要求**：较强的语义表达能力；能够方便、直接地表达应用中的各种语义知识；简单、清晰、易于用户理解

**重要概念**：

### (1) 实体 (Entity)

客观存在并可相互区别的事物称为实体。

可以是具体的人、事、物或抽象的概念。

### (2) 属性 (Attribute)

实体所具有的某一特性称为属性。

一个实体可以由若干个属性来刻画。

### (3) 码 (Key)

唯一标识实体的属性集称为码。

### (4) 域 (Domain)

属性的取值范围称为该属性的域。

### (5) 实体型 (Entity Type)

用实体名及其属性名集合来抽象和刻画同类实体称为实体型

### (6) 实体集 (Entity Set)

同一类型实体的集合称为实体集

### (7) 联系 (Relationship)

现实世界中事物内部以及事物之间的联系在信息世界中反映为实体内部的联系和实体之间的联系。

实体内部的联系通常是指组成实体的各属性之间的联系。

实体之间的联系通常是指不同实体集之间的联系。

两个实体间的联系： $1:1$ ， $1:m$ ， $m:n$ 。具体的定义请看 ppt（第一张 p63）。

概念模型的一种表示方法：实体—联系方法(E-R 方法)

联系本身：

用菱形表示，菱形框内写明联系名，并用无向边分别与有关实体连接起来，同时在无向边旁标上联系的类型（ $1:1$ 、 $1:n$  或  $m:n$ ）。

**学会根据题目要求画实体关系图。**

例：用 E-R 图表示某个工厂物资管理的概念模型

实体

仓库： 仓库号、面积、电话号码

零件： 零件号、名称、规格、单价、描述

供应商： 供应商号、姓名、地址、电话号码、帐号

项目： 项目号、预算、开工日期

职工： 职工号、姓名、年龄、职称

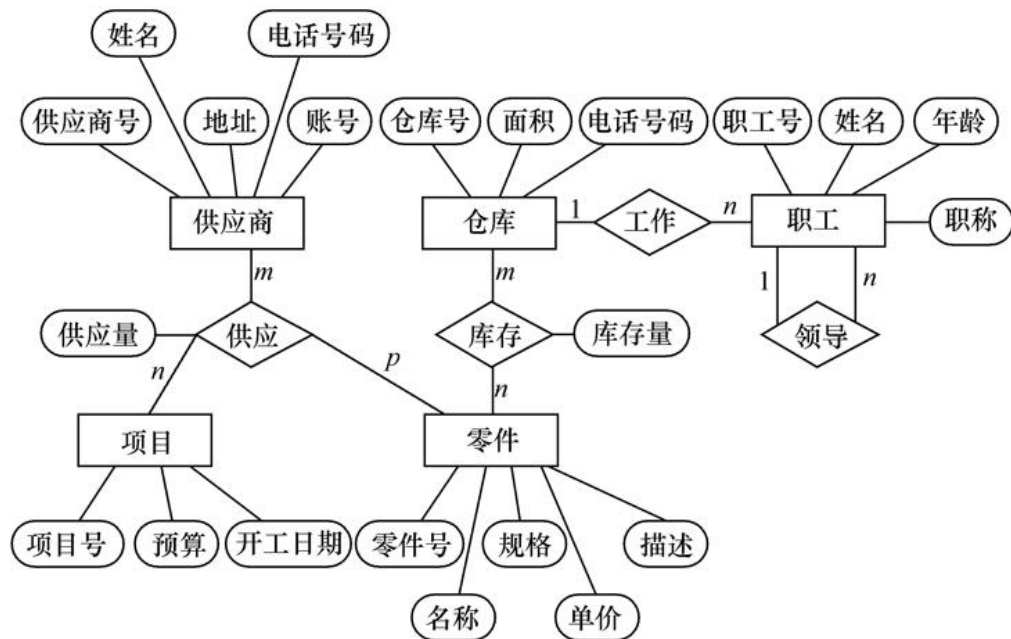
实体之间的联系如下：

(1) 一个仓库可以存放多种零件，一种零件可以存放在多个仓库中。仓库和零件具有多对多的联系。用库存量来表示某种零件在某个仓库中的数量。

(2) 一个仓库有多个职工当仓库保管员，一个职工只能在一个仓库工作，仓库和职工之间是一对多的联系。职工实体型中具有一对多的联系

(3)职工之间具有领导-被领导关系。即仓库主任领导若干保管员。

(4)供应商、项目和零件三者之间具有多对多的联系



(c) 完整的实体-联系图

最常用的数据模型:

层次模型(Hierarchical Model)

网状模型(Network Model)

关系模型(Relational Model)

面向对象模型(Object Oriented Model)

对象关系模型(Object Relational Model)

数据模型重点掌握关系模型（其他的老师没怎么讲）

### 关系数据模型的数据结构

在用户观点下，关系模型中数据的逻辑结构是一张二维表，它由行和列组成。

**关系 (Relation):** 一个关系对应通常说的一张表。

**元组 (Tuple):** 表中的一行即为一个元组。

**属性 (Attribute):** 表中的一列即为一个属性，给每一个属性起一个名称即属性名。

**主码 (Key):** 表中的某个属性组，它可以唯一确定一个元组。

**域 (Domain):** 属性的取值范围。

**分量:** 元组中的一个属性值。

**关系模式:** 对关系的描述，例如

关系名 (属性 1, 属性 2, ..., 属性 n)

学生 (学号, 姓名, 年龄, 性别, 系, 年级)

关系必须是规范化的，满足一定的规范条件

最基本的规范条件：关系的每一个分量必须是一个不可分的数据项，不允许表中还有表。

关系术语	一般表格的术语
关系名	表名
关系模式	表头（表格的描述）
关系	（一张）二维表
元组	记录或行
属性	列
属性名	列名
属性值	列值
分量	一条记录中的一个列值
非规范关系	表中有表（大表中嵌有小表）

### 关系数据模型的操纵与完整性约束（必考内容！）

数据操作是集合操作，操作对象和操作结果都是关系，**关系数据模型的操作主要包括：**

查询

插入

删除

更新

数据操作是集合操作，操作对象和操作结果都是关系，即若干元组的集合。

存取路径对用户隐蔽，用户只要指出“干什么”，不必详细说明“怎么干”。

#### 关系的完整性约束条件

**实体完整性：**若属性  $A$  是基本关系  $R$  的主属性，则属性  $A$  不能取空值。

**参照完整性：**若属性（或属性组） $F$  是基本关系  $R$  的外码它与基本关系  $S$  的主码  $K_s$  相对应（基本关系  $R$  和  $S$  不一定是不同的关系），则对于  $R$  中每个元组在  $F$  上的值必须为：或者取空值（ $F$  的每个属性值均为空值）

或者等于  $S$  中某个元组的主码值。

**用户定义的完整性：**针对某一具体关系数据库的约束条件，反映某一具体应用所涉及的数据必须满足的语义要求。

关系数据模型的**存储结构**：实体及实体间的联系都用表来表示，表以文件形式存储，有的 DBMS 一个表对应一个操作系统文件，有的 DBMS 自己设计文件结构。

#### 关系数据模型的优缺点

优点：

建立在严格的数学概念的基础上；概念单一；关系模型的存取路径对用户透明，从而具有更高的数据独立性，更好的安全保密性，也简化了程序员的工作和数据库开发建立的工作。（教材 P30）

缺点：

存取路径对用户透明导致查询效率往往不如非关系数据模型；为提高性能，必须对用户的查询请求进行优化，增加了开发 DBMS 的难度。

### **数据库系统结构（要求理解）**

从数据库管理系统角度看，数据库系统通常采用三级模式结构，是数据库系统内部的系统结构。

从数据库最终用户角度看（数据库系统外部的体系结构），数据库系统的结构分为：单用户结构，主从式结构，分布式结构，客户 / 服务器，浏览器 / 应用服务器 / 数据库服务器多层结构等。

数据库系统模式的概念：

#### **型(Type)**

对某一类数据的结构和属性的说明。

#### **值(Value)**

是型的一个具体赋值。

例如

学生记录型：

（学号，姓名，性别，系别，年龄，籍贯）

一个记录值：

（900201，李明，男，计算机，22，江苏）

#### **模式（Schema）**

数据库逻辑结构和特征的描述

是型的描述

反映的是数据的结构及其联系

模式是相对稳定的

#### **实例（Instance）**

模式的一个具体值

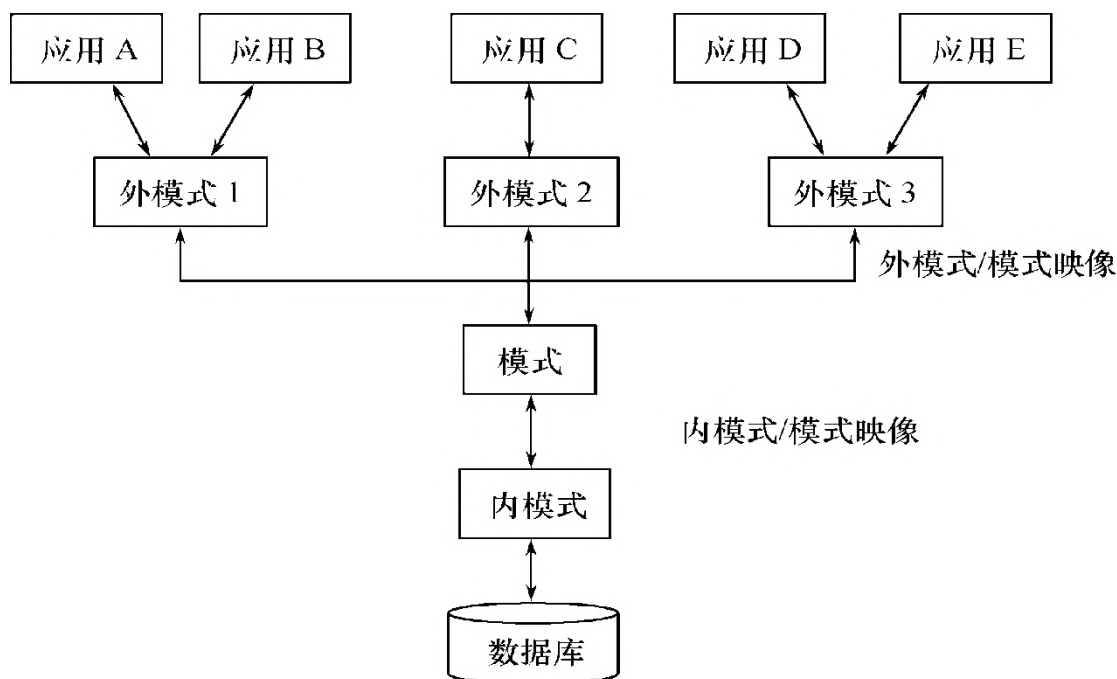
反映数据库某一时刻的状态

同一个模式可以有很多实例

实例随数据库中的数据更新而变动

### **数据库系统的三级模式结构（主要考察各层的含义及特点，相互之间的关系）：**

模式（Schema），外模式（External Schema）内模式（Internal Schema）



**模式：**（也称逻辑模式）

数据库中全体数据的逻辑结构和特征的描述。

所有用户的公共数据视图，综合了所有用户的需求。

一个数据库只有一个模式

**模式的地位：**是数据库系统模式结构的中间层。

与数据的物理存储细节和硬件环境无关。

与具体的应用程序、开发工具及高级程序设计语言无关。

**模式的定义：**

数据的逻辑结构（数据项的名字、类型、取值范围等）。

数据之间的联系。

数据有关的安全性、完整性要求。

**外模式：**（也称子模式或用户模式）

数据库用户（包括应用程序员和最终用户）使用的局部数据的逻辑结构和特征的描述。

数据库用户的数据视图，是与某一应用有关的数据的逻辑表示。

**外模式的地位：**介于模式与应用之间。

模式与外模式的关系：一对多。

外模式通常是模式的子集。

一个数据库可以有多个外模式。反映了不同的用户的应用需求、看待数据的方式、对数据保密的要求。

对模式中同一数据，在外模式中的结构、类型、长度、保密级别等都可以不同

外模式与应用的关系：一对多。

同一外模式也可以为某一用户的多个应用系统所使用。

但一个应用程序只能使用一个外模式。

**外模式的用途**

保证数据库安全性的一个有力措施。

每个用户只能看见和访问所对应的外模式中的数据。

**内模式**（也称存储模式）

是数据物理结构和存储方式的描述。

是数据在数据库内部的表示方式：记录的存储方式（顺序存储，按照 B 树结构存储，按 hash 方法存储）；索引的组织方式；数据是否压缩存储；数据是否加密；数据存储记录结构的规定；

一个数据库只有一个内模式。

**两层映像（主要考察名称、含义及作用）：**外模式 / 模式映像，模式 / 内模式映像。

**外模式 / 模式映像含义：**每一个外模式，数据库系统都有一个外模式 / 模式映像，定义外模式与模式之间的对应关系。

**外模式 / 模式映像作用：**保证数据的逻辑独立性；

**模式 / 内模式映像含义：**数据库中模式 / 内模式映像是唯一的，定义了数据全局逻辑结构与存储结构之间的对应关系。

**模式 / 内模式映像作用：**保证数据的物理独立性。

关系代数：并、差、交、笛卡尔积、投影、选择、连接、除。

参照教材 P52-P60 和 ppt（第二章 关系数据库(续)）。

要求掌握关系代数的运算方法，要求能够根据查询要求写出使用关系代数表达的查询表达式或运算结果。

关系演算考的可能性不大，老师给的考试范围没提到，而且上课也没怎么讲。

### c)关系模型数据操作

1. 掌握 DDL 语句，包括 CREATE, DROP, ALTER；会创建基本的数据对象包括基本表，视图，索引。
2. 掌握 DML 语句，包括 INSERT, UPDATE, DELETE；会基本的应用。
3. 掌握 DQL 语句，要求综合掌握和灵活运用所有查询语句。

知识点在教材第三章 P84-P126。

教材第五章系统的讲了完整性，复习完整性时可以参考，其中的触发器老师没讲，考试范围里也没涉及到。

第六章主要看教材 P170-P182 里的内容，其中多值依赖和 4NF 不要求掌握。

这一章的考试要求有：

- 1.掌握 1NF, 2NF, 3NF, BCNF 的定义，理解各个范式的含义及特点。
- 2.掌握函数依赖并根据问题描述写出相应的函数依赖，能根据范式定义判断关系模式的范式级别或进行关系模式的规范化处理。
- 3.能对所生成的逻辑模型进行分析，能判断所属的范式，并将其规范化到某一级别的范式（最高 BCNF）。

### 第十章：理解数据库事务处理机制（数据库的恢复）

- 1 事务的定义：一个数据库操作序列，这些操作要么全做要么全不做，是一个不可分割的工作单位，是并发控制的基本单位。
- 2 事务的特性：原子性，一致性，隔离性，持续性。
- 3 数据库的恢复：把数据库从错误状态恢复到某一已知的正确状态。



4.恢复的实现技术：数据转储和登录日志文件。

#### 第十一章：理解并发控制原理

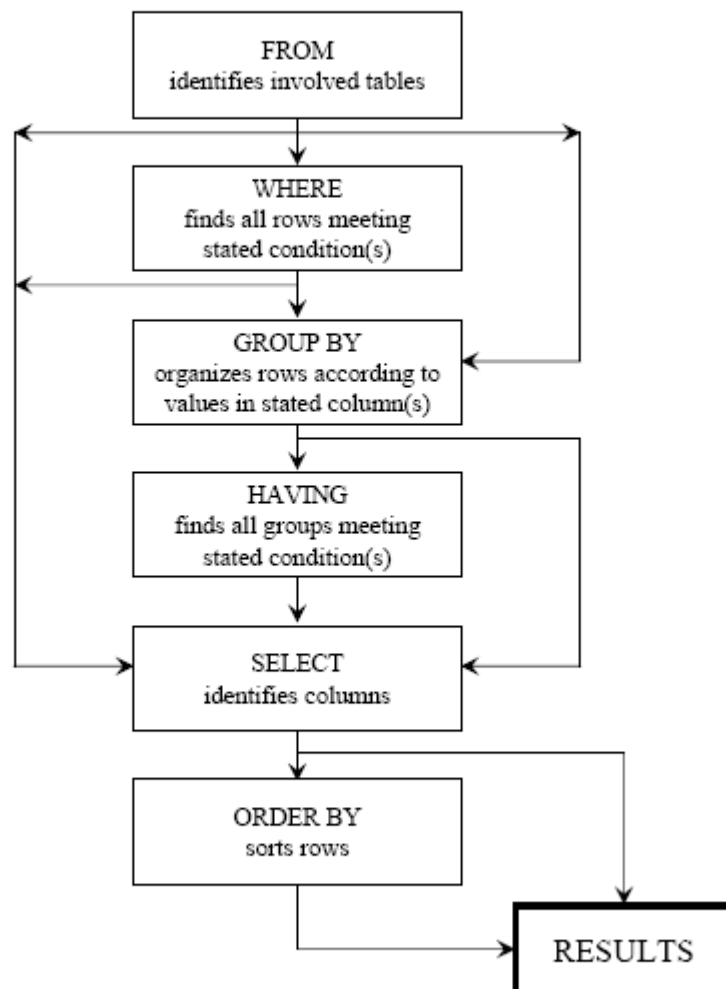
1 并发控制就是要用正确的方式调度并发操作，使一个用户事务的执行不受其他事务的干扰，从而避免造成数据的不一致性。

2 并发控制机制的任务：对并发操作进行正确调度；保证事务的隔离性；保证数据库的一致性。

3 并发控制的主要技术：封锁(Locking)，时间戳(Timestamp)，乐观控制法。

还有这个图也很重要：

## SQL statement processing order



Source: "Modern Database Management" by F. McFadden, J. Hoffer, and M. Prescott,

**参考习题：**另外附件中有课后习题答案，多看看课后的简答题和有关范式，关系代数的题

### 一、（每小题 4 分，共 12 分）

用 SQL 语言定义

- 1) 学生关系 S，包括学号 SNo、姓名 SN、年龄 SA、系别 SD；
- 2) 课程关系 C，包括课程号 CNo、课程名 CN、学分 CC；
- 3) 学生选课关系 SC，包括 SNo、CNo 和成绩 G。

注意：说明主键码和外键码（如果有的话）。

1) CREATE TABLE S{

SNo INT PRIMARY KEY,

SN CHAR(30),

SA INT,

SD CHAR(20)

};

2) CREATE TABLE C{

CNo INT PRIMARY KEY,

CN CHAR(20),

CC INT

};

3) CREATE TABLE SC{

SNo INT,

CNo INT,

G INT,

PRIMARY KEY(SNo, CNo),

FOREIGN KEY (SNo) REFERENCES S (SNo),

FOREIGN KEY (CNo) REFERENCES C (CNo)

};

## 二、（每小题 4 分，共 12 分）

按题 1 的关系模式，用 SQL 语言

- 1) 于学生选课关系 SC，授予王云查询权限以及对 SNo 和 CNo 的插入权限；
- 2 删除学号为“9900205”的学生；
- 3 从学生选课关系 SC 中，删除王云（学生关系中可能有重名）的所有选课。

1) GRANT SELECT, INSERT (SNo, CNo) ON SC TO 王云;

2) DELETE FROM S

WHERE SNo=9900205

3) DELETE FROM SC

WHERE SNo IN

(SELECT SNo

FROM S

WHERE SN='王云'

);

## 三、（每小题 3 分，共 6 分）

结合题 1 的关系模式，

- 1) 用 SQL 语句查询每个学生的平均成绩，**输出学号和平均成绩**，
- 2) 按平均成绩排序（降序），若平均成绩相同，再按学号排序（升序）。

```
SELECT SNo,AVG(G)

FROM SC

GROUP BY SNo

ORDER BY AVG(G) DESC,SNo
```

#### 四、（每小题 4 分，共 8 分）

商店销售商品的数据库，涉及如下数据：

商店号、商店名、地址；

商品号、商品名、单价；

某商店销售某商品的数量和日期。

- 1) 设计 E/R 图；
- 2) 转换成关系模型。

(1)



- 2) 商店（商店号，商店名，地址）

商品（商品号，商品名，单价）

销售（商店号@，商品号@，数量，日期）（@表示外键）

六、（每小题 6 分，共 12 分）有关图书发行的关系模型如下：

书店 S（书店号 SNo，书店名 SN，地址 A）

图书 B（书号 BNo，书名 T，作者 N，价格 P）

图书馆 L（馆号 LNo，馆名 LN，城市 C）

图书发行 SBL（SNo，LNo，BNo，发行量 Q）

分别用 1) 关系代数和 2) SQL 语言表达如下查询：

“上海图书馆”（馆名）收藏图书的书名和出售该书的书店名。

$\pi_{T, SN}(\sigma_{LN = \text{'上海图书馆'}}(L \bowtie SBL \bowtie B \bowtie S))$ （有待优化）

2) SELECT T, SN

FROM L, SBL, B, S

WHERE L. LNo=SBL. LNo AND SBL. BNo=B. BNo

AND SBL. SNo=S. SNo

AND L. LN= '上海图书馆'

七、（第 1 小题 4 分，第 2 小题 8 分，共 12 分）

图书管理数据库关系模型如下：

图书 B（书号 BN，书名 T，作者 A）

学生 S（姓名 N，班级 C，借书证号 LN）

借书 L（LN，BN，日期 D）

查询：2002. 1. 1（20020101）以前借书的学生姓名和书名。

要求：1) 以笛卡尔积为基础表达查询；

1)

$\pi N, T(\sigma_{D < 20020101}(\pi_{BN, T, A, N, C, LN, D}(\sigma_{L.LN=S.LN \text{ AND } L.BN=B.BN}(\sigma_{B \neq S \neq L}))))$  (还有待优化)

#### 八、小题 6 分，共 12 分)

涉及到学生、教师和课程的关系模式 STC (SNo, SN, SA, TN, CN, G)，其中 6 个属性分别为学生的学号、姓名、年龄、教师的姓名、课程名以及学生的成绩。假设学生有重名，课程名也可能有重名。又假设每个教师只教一门课，但一门课可有几个教师开设。当某个学生选定某门课后，其上课教师就固定了。

1) 写出键码和函数依赖；

2) 分解关系模式使之属于 BCNF 范式。

1) 键码：{SNo, CN} 和 {SNo, TN}

函数依赖：SNo  $\rightarrow$  SN, SA (BC 范式违例)

TN  $\rightarrow$  CN (BC 范式违例)

SNo, CN  $\rightarrow$  TN, G

a) SNo, CN  $\rightarrow$  SN, SA

SNo, TN  $\rightarrow$  G

b) SNo, TN  $\rightarrow$  CN

c) SNo, TN  $\rightarrow$  SN, SA (a, b, c 为部分依赖，可不写)

2) STC1 (SNo, SN, SA)

STC2 (TN, CN)

STC3 (SNo, TN, G)