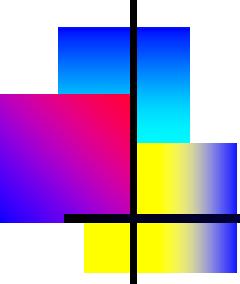


教材：数据库实用教程（第四版）

《数据库原理》课程

清华大学出版社

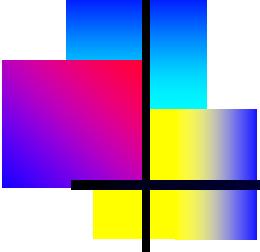
2024年12月30日



《数据库原理》

第七章 数据库设计

清华大学出版社
2024年12月30日



第7章 数据库设计

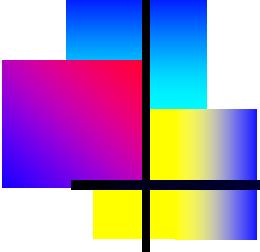
教学内容：

基本概念：数据库设计中规划、需求分析、概念设计、逻辑设计、物理设计及运行与维护各个阶段的目标、方法和应注意的事项。

总体要求：详细了解数据库应用系统设计的全过程。

教学重点：概念设计中 ER模型设计方法，逻辑设计中 ER模型向关系模型的转换方法。

本章特点：实用性较强。



§ 1 数据库设计概述

数据库设计是指对于一个给定的应用环境，提供一个确定最优数据模型与处理模式的逻辑设计，以及一个确定数据库存储结构与存取方法的物理设计，建立起既能反映现实世界信息和信息联系，满足用户数据要求和加工要求，又能被某个数据库管理系统所接受，同时能实现系统目标，并有效存取数据的数据库。

一、软件工程与软件生存期

软件生存期是指从软件的规划、研制、实现、投入运行后的维护，直到它被新的软件所取代而停止使用的整个期间。软件生存期通常分为六个阶段：

规划阶段

需求分析阶段

设计阶段

程序编制阶段

调试阶段

运行维护阶段：

二、数据库工程与数据库生存期

以数据库为基础的信息系统通常称为数据库应用系统，具有对信息的采集、组织、加工、抽取和传播等功能。数据库应用系统的开发是一项软件工程但又具有自己的特点，所以称为数据库工程。

一项数据库工程按内容可划分为两部分：

作为系统核心的数据库系统的设计与实现，

相应的应用软件及其他软件（如通讯软件）的设计与实现。

数据库系统生存期： 数据库应用系统从开始规划、设计、实现、维护到最后被新的系统取代而停止使用的整个期间。

数据库系统的软件生存期通常分为七个阶段：

规划阶段

需求分析阶段

概念设计阶段

逻辑设计阶段

物理设计阶段

实现和运行维护。

三、数据库设计的输入和输出

数据库设计过程的输入包括四部分内容：

总体信息需求；

处理需求；

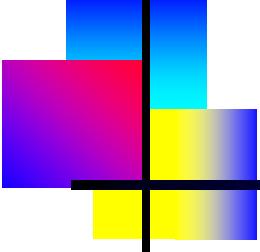
DBMS的特征；

硬件和OS特征。

数据库设计过程的输出主要有两部分内容：

完整的数据库结构；

基于数据库结构和处理需求的应用程序的设计原则。



§ 2 数据库设计的步骤

一、数据库设计的步骤

数据库设计的步骤是指数据库系统的软件生存期中：

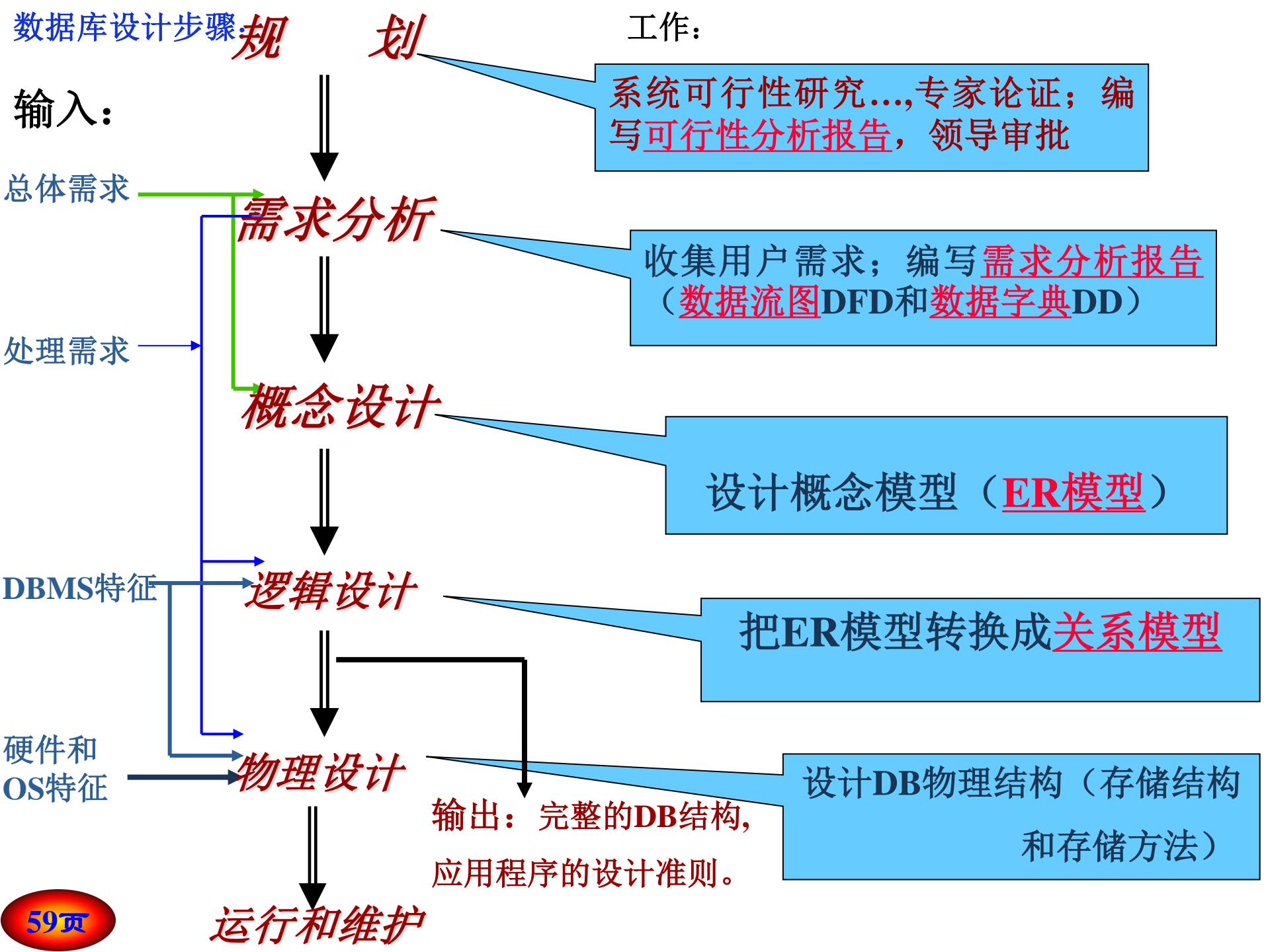
规划

需求分析

概念设计

逻辑设计

物理设计



二、规划阶段的任务和工作

规划阶段的主要任务是进行建立数据库的必要性及可行性分析，确定数据库系统在组织中和信息系统中的地位，以及各个数据库之间的联系。

规划工作完成后应写出详尽的可行性分析报告和数据库系统规划纲要，内容包括：信息范围、信息来源、人力资源、设备资源、软件及支持工具资源、开发成本估算、开发进度计划、现行系统向新系统过渡计划等。

三、需求分析

需求分析阶段要对系统的整个应用情况作全面的详细的调查，确定企业组织的目标，收集支持系统总的设计目标的基础数据和对这些数据的要求，确定用户的需求，并把这些要求写成用户和数据库设计者都能接受的文档。

1. 需求分析的步骤有四步

分析用户活动，产生用户活动图（即用户的业务流程图）；

确定系统范围，产生系统范围图（即确定人机界面）；

分析用户活动所涉及的数据，产生数据流图（数据的流向及加工）；

分析系统数据，产生数据字典。

2. 需求分析阶段的输入和输出

输入是总体信息需求和处理需求。

输出是需求说明书包括系统的数据流程和数据字典。

3. 数据流图和数据字典

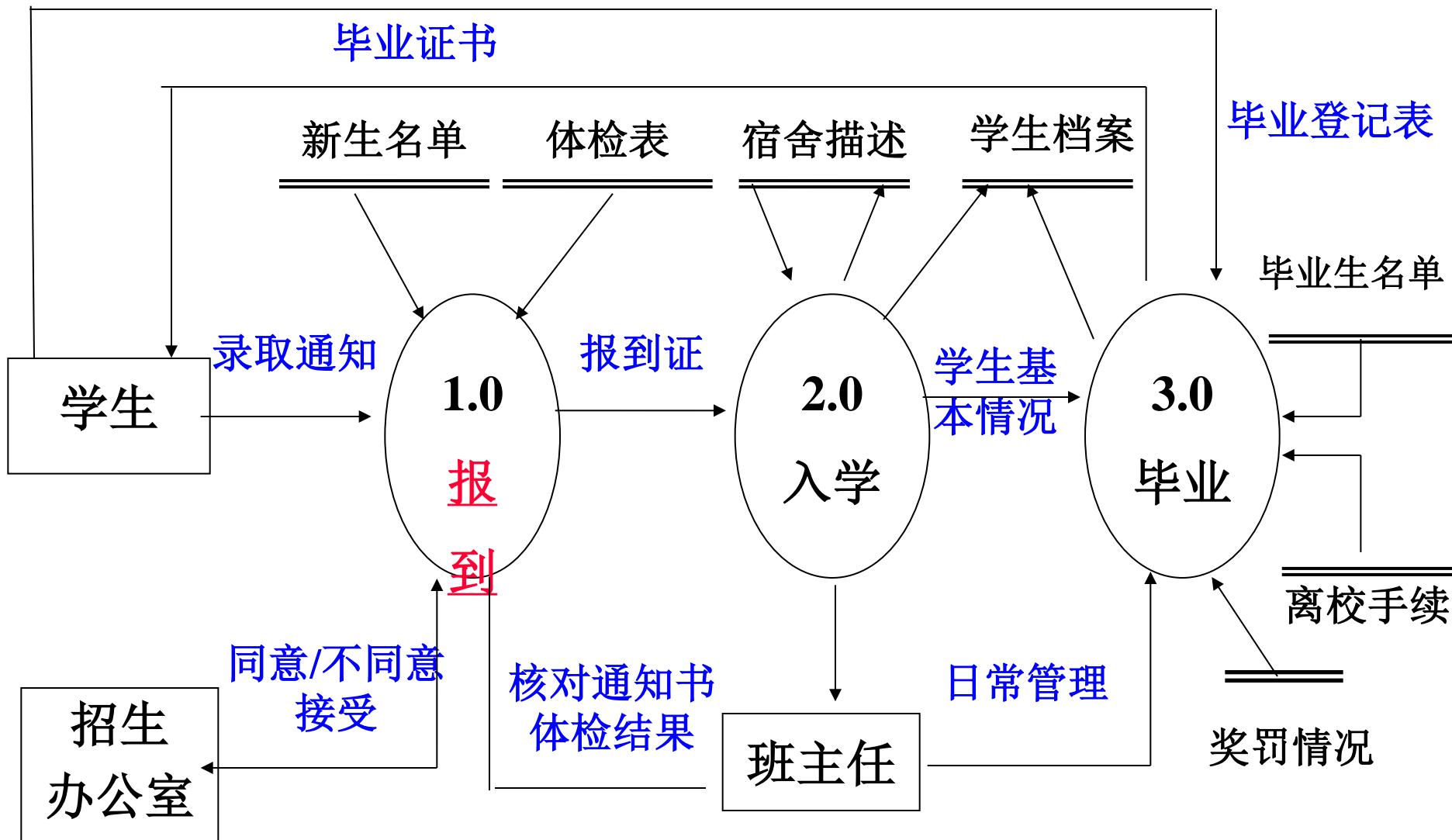
数据流图DFD有四个基本成分：

用箭头表示数据流、用圆圈表示加工或处理，

用双线段表示文件、用方框表示外部实体。

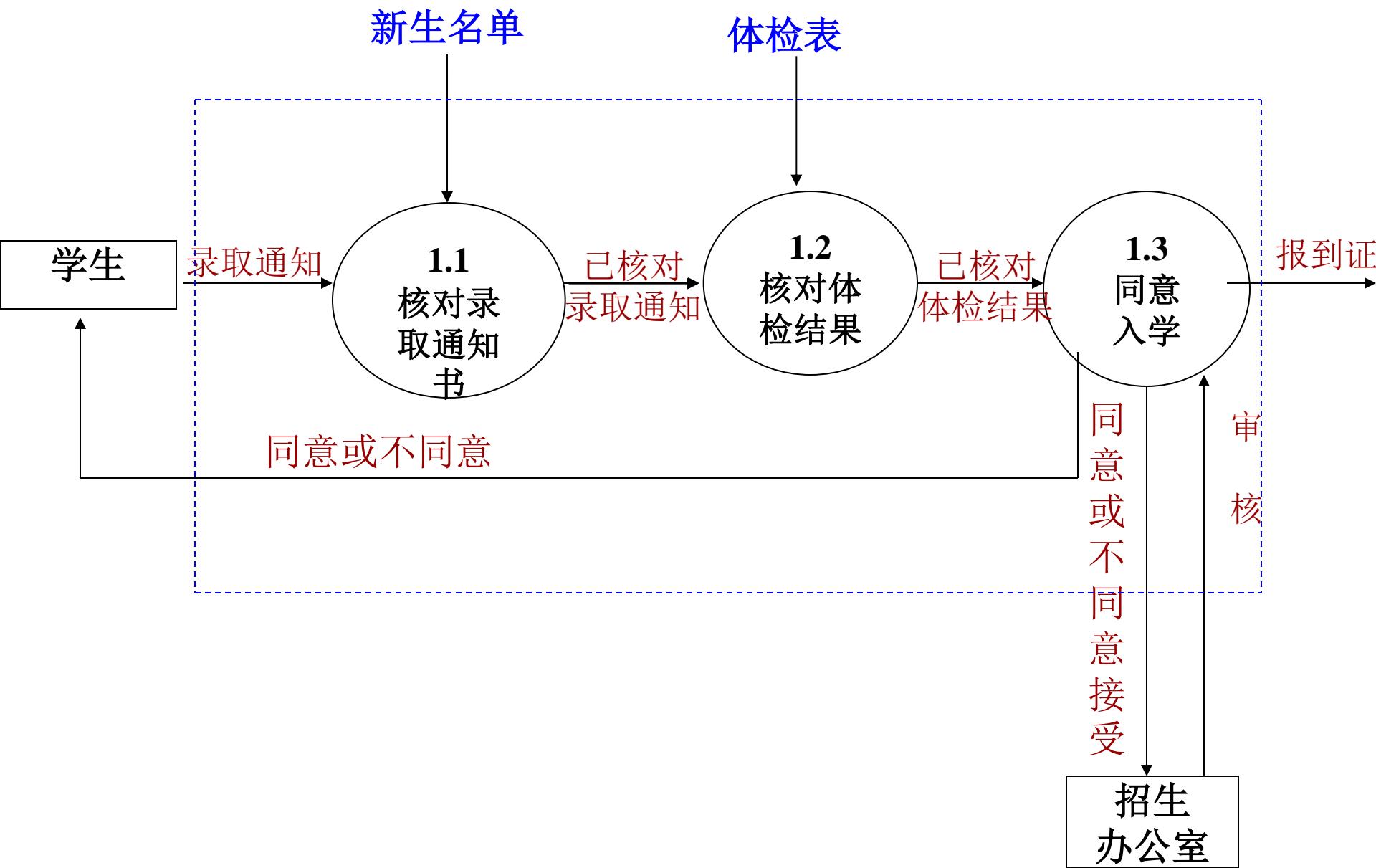
学籍管理数据流图

毕业登记表



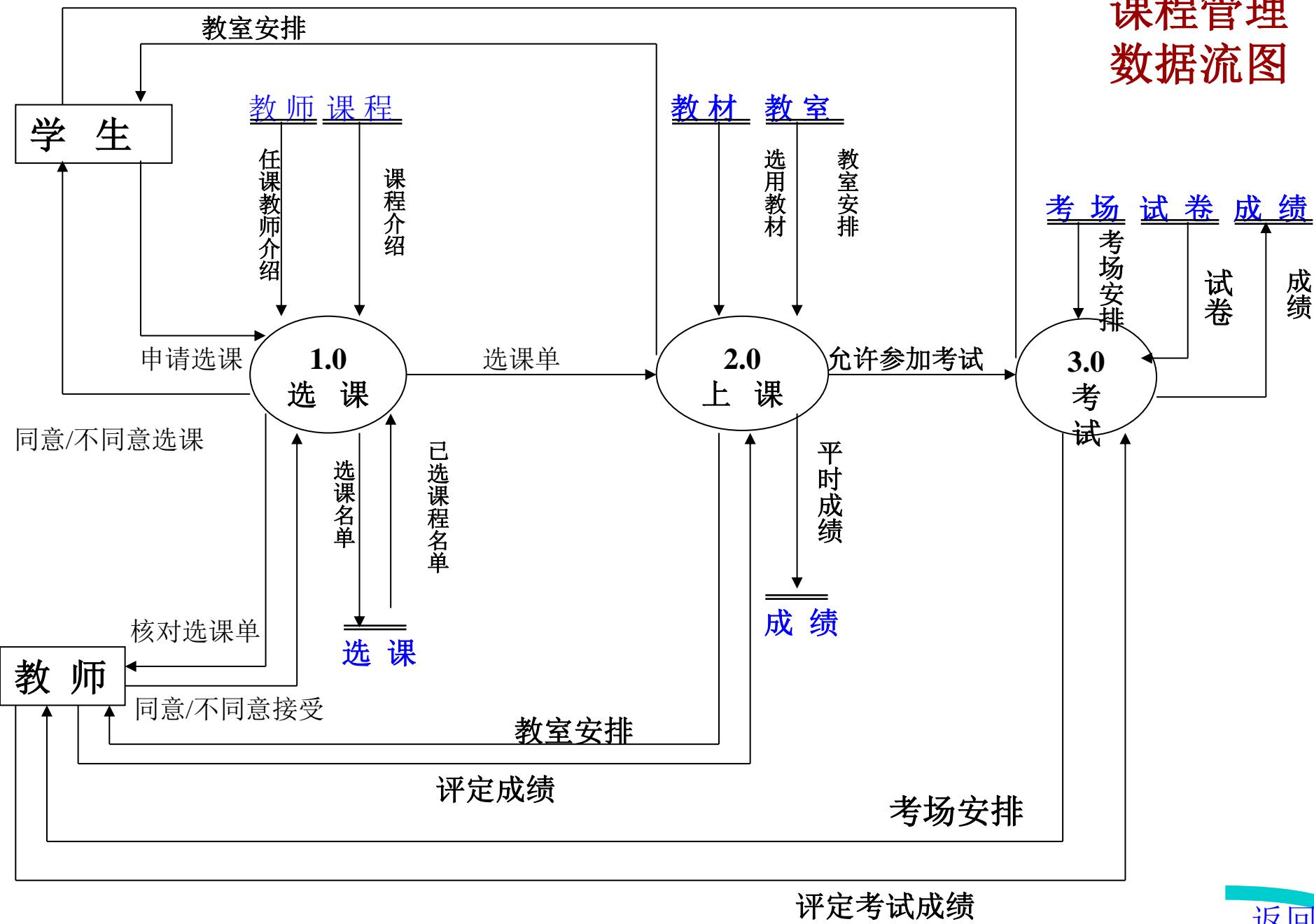
数据流图DFD它可形象地表示数据流与各业务活动的关系，是需求分析的工具和分析结果的描述手段。

DFD可作为自顶向下逐步细化描述对象的工具。顶层的每一个圆圈(加工处理)都可以进一步细化为第二层；第二层的每一个圆圈都可以进一步细化为第三层；...直到最底层的每一个圆圈已表示一个最基本的处理动作为止。DFD可以形象地表示数据流与个业务活动的关系，它是需求分析的工具和分析结果的描述手段。



课程管理 数据流图

成绩单



返回

数据字典：是对系统中数据的详尽描述，提供了对数据库数据描述的集中管理。数据字典是各类数据描述（称为元数据）的集合，通常包括**数据项、数据结构、数据流、数据存储和加工过程**等五个方面。数据字典的功能是存储和检索各种数据描述，为DBA提供有关的报告。

数据字典是在需求分析阶段建立，并在数据库设计过程中不断改进、充实和完善。

(1) 数据项: 数据项是数据的最小单位。

数据项描述={数据项名, 含义说明, 类型, 长度, 取值范围,
与其它数据项的逻辑关系}

(2) 数据结构: 数据结构反映了数据之间的组合关系。

数据结构描述={数据结构名, 含义说明, 组成, {数据项或数据结
构} }

(3) 数据流: 表示某一加工处理过程的输入或输出数据。

数据流描述={ 数据流名, 说明, 流入的加工名, 流出的加工名,

组成该数据流的数据结构或数据项 }

例：考场安排 数据流 描述如下：

数据流名：考场安排

说 明：按各课程所选学生数，选定教室，时间确定考场安排

来 源：考试

去 向：学生，教师

数据结构：考场安排——考试课程，考试时间，教学楼，教室编号

例：考试课程的数据结构：

数据结构名：考试课程

说 明：作为考场安排的组成部分，说明某门课程，上课教师，
选课学生人数。

组 成：课程号

教师号

选课人数

(4) 数据存储:

数据存储是处理过程中要存储的数据。

数据存储描述={数据存储名、说明、输入数据流、输出数据流、

数据量（每次存取多少数据）、存取方式

（是批处理，还是联机处理；是检索，还是

更新；是顺序存取，还是随机存取） }

(5) 处理过程:

处理过程描述={处理过程名、说明、输入数据流、输出数据

流，简要说明处理工作}

例：课程是数据存储，在数据字典中可对其作如下描述：

数据存储名：课 程

说 明：对每门课程的名称、学分、先行课程号和摘要的
描述。

输出数据流：课程介绍

数据描述： 课程号

课程名

学分数

先行课程号

摘要

数据量： 每年100种

存取方式： 随机存取

例：加工过程“选课”，在数据字典中描述如下：

处理过程：确定选课名单

说 明：对要选某门课程的每一个学生，根据已选修课程确定其是否可选该课程。再根据学生选课的人数选择适当的教室，制定选课单。

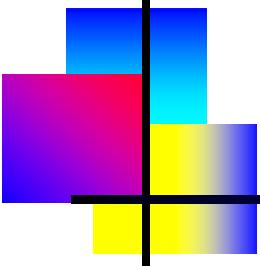
输入：学生选课

可选课程

已选课程

输出：选课单

程序提要：a、对所选课程在选课表中查找其是否已选此课程；
b、若未选过此课程，则在选课表中查找是否已选此课程的先行课程；
c、若a、b都满足，则在选课表中增加一条选课记录；
d、处理完全部学生的选课处理后，形成选课单。[返回](#)



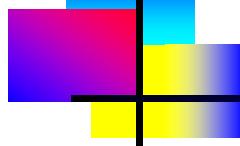
四、概念设计

1、概念设计的目标：是产生反映企业组织信息需求的

数据库概念结构，即概念模式。概念模式独立于数据库逻辑结构，也独立于支持数据库的DBMS。

概念设计是整个数据库设计的关键。





2、概念设计的方法

设计概念结构通常有四种方法：

自顶向下:即首先定义全局概念结构的框架。然后逐步细化。

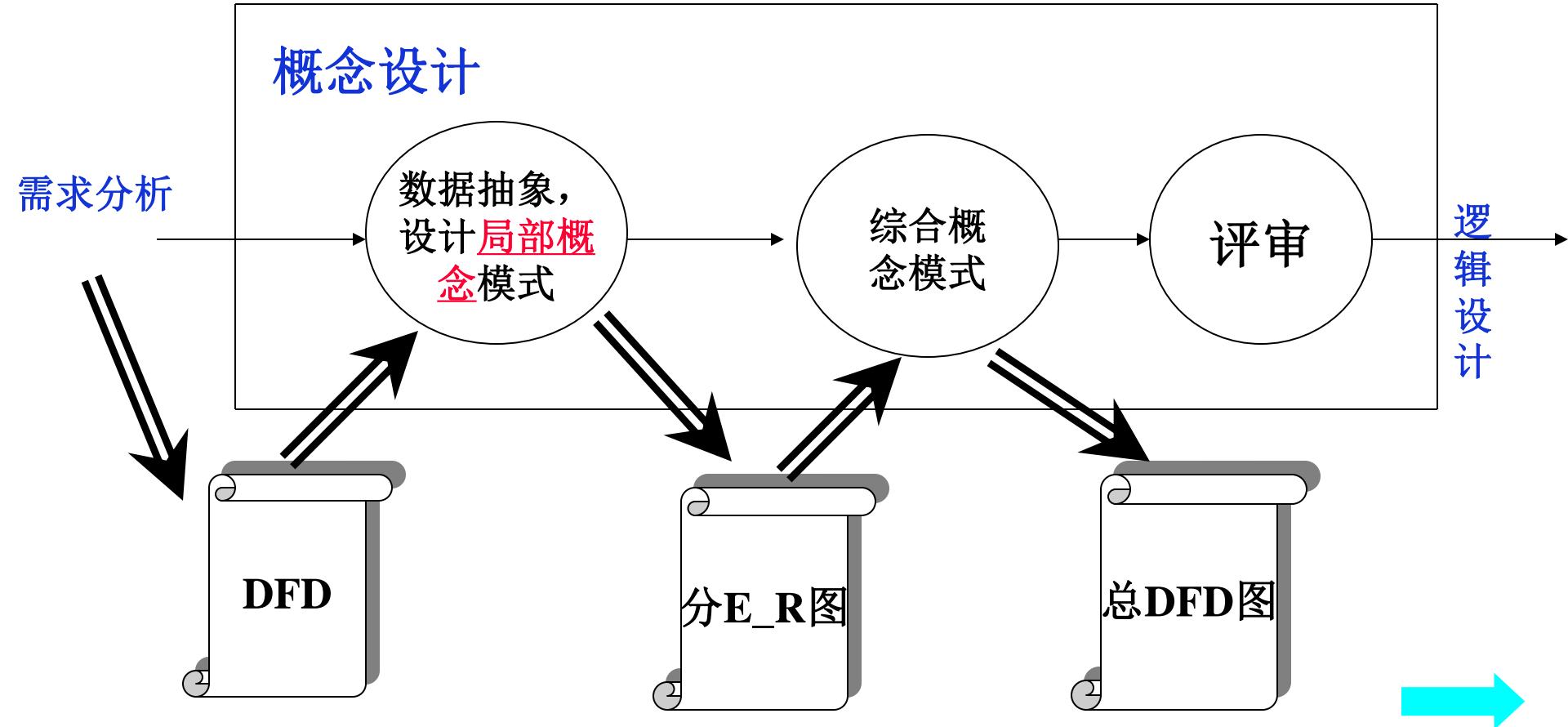
自底向上:即首先定义各局部应用概念结构。然后将它们集成起来，得到全局概念结构。

逐步扩张:首先定义最重要的核心概念结构,然后将它们向外扩充，以滚雪球的方式逐步生成其他概念结构，直至总体概念结构。

混合策略: 即将自顶向下和自底向上相结合，用自顶向下策略设计一个全局概念结构的框架,以它为骨架集成自底向上策略中设计的各局部概念结构。

无论哪种方式，都以E_R模型为工具来描述概念结构。

3、概念设计的步骤(分三步)：进行数据抽象, 设计局部概念模式;
将局部概念模式综合成全局概念模式;
评审。



数据抽象

抽象是对实际的人、物、事或概念的人为处理，抽取人们关心的共同特性，忽略非本质的细节，将这些特性用各种概念精确地加以描述，这些概念组成了某种模型。

抽象有两种形式：①系统状态的抽象，即抽象对象，
②系统转换的抽象，即抽象运算。

概念设计的目的就是要定义抽象对象的关系结构。

对象的形式就有两种形式：聚集和概括。

(1) 聚集

聚集（Aggregation）的数学意义就是笛卡尔积的概念。通过聚集，形成对象之间的一个联系对象。

聚集层次表示：“是……的一部分”（is part of）的关系。

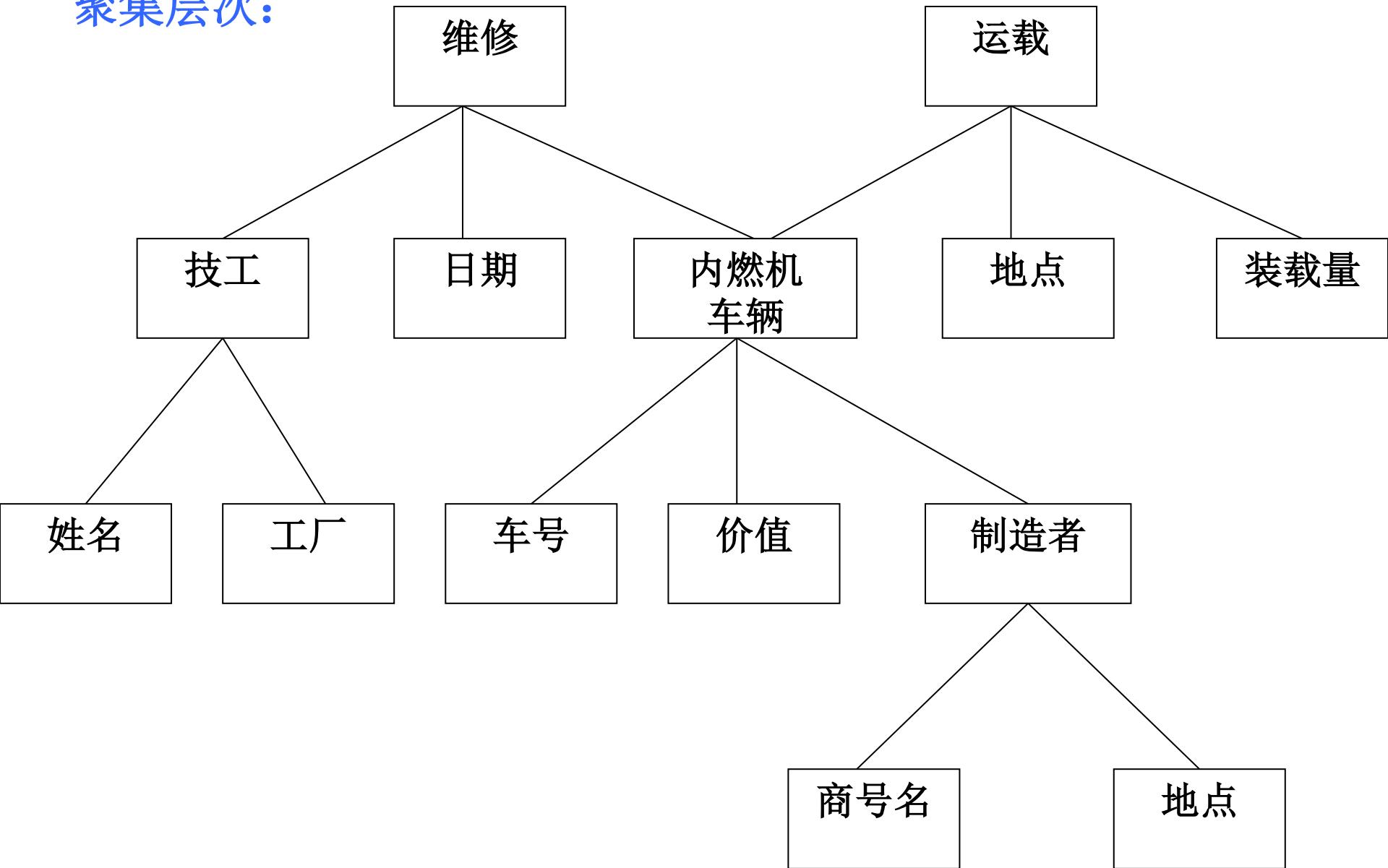
(2) 概括

概括 (Generalization) 是从一类对象形成一个对象。对于一类对象 {01, 02, ..., 0n} 可以概括成对象0，那么0i称为0的其中一个。例如一类对象 {汽车、卡车、自行车、摩托车、...} 可以概括为一个对象 “公路车辆”。概括层次表示 “是.....一种” (is a) 的关系。

(3) 数据抽象层次

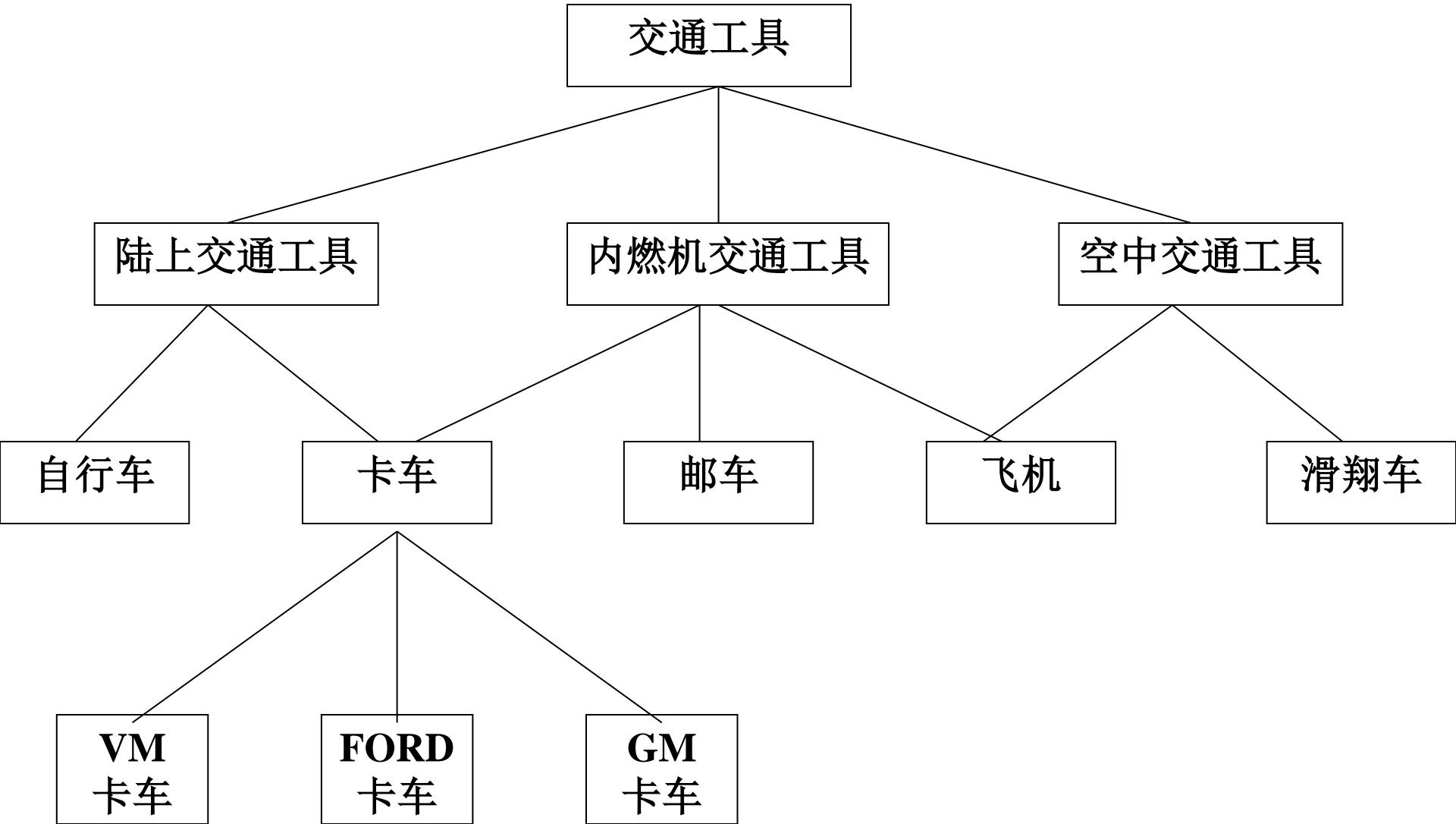
一个聚集对象可能是某类对象的概括，此时它也是一个概括对象。一个概括对象也可能是对象联系的聚集，此时，它也可以是聚集对象。一般说来，每个对象既可以是聚集对象，又可以是概括对象。当反复利用概括和聚集进行数据抽象时，就可以形成对象的层次关系。

聚集层次:



返回

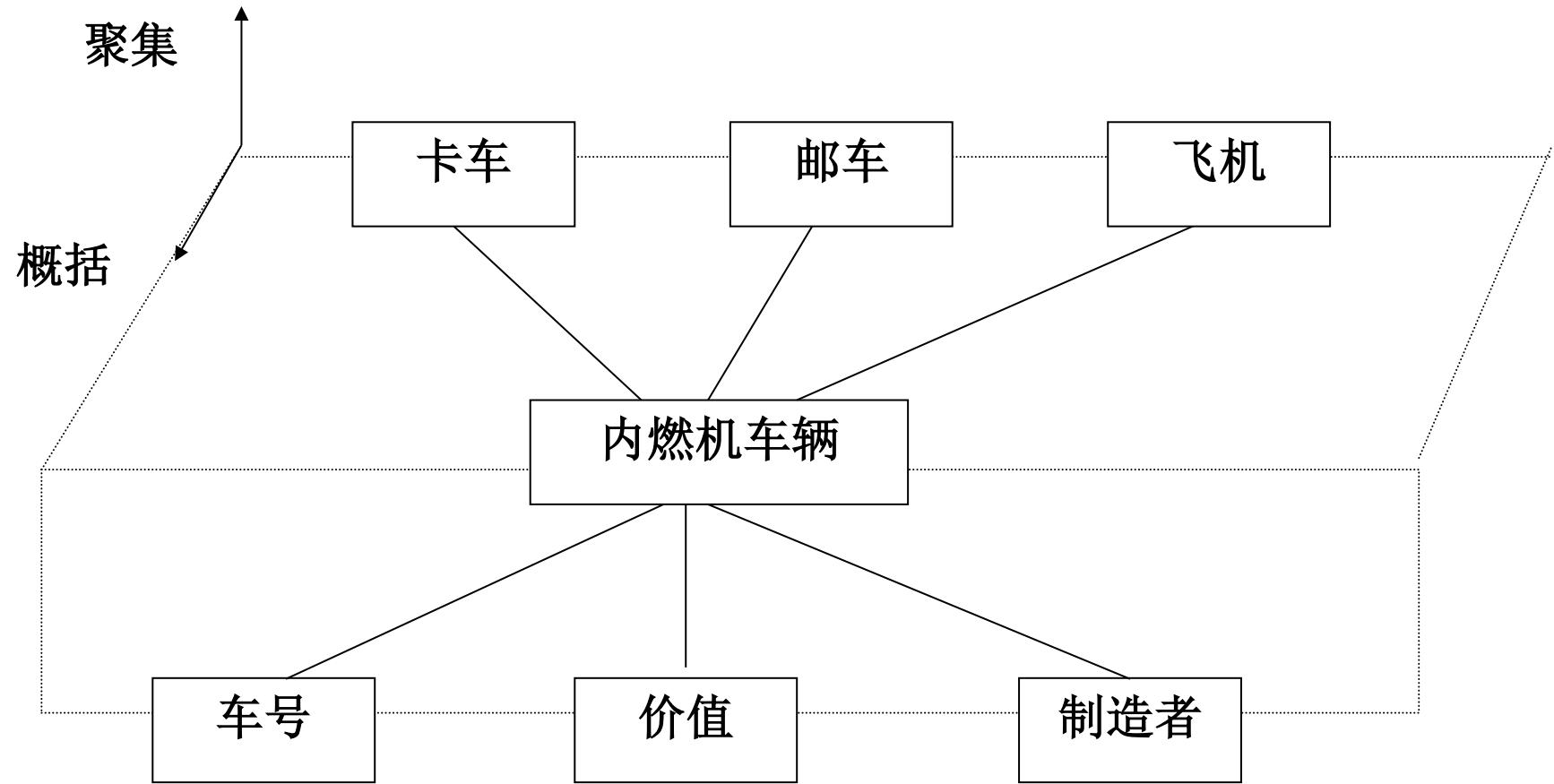
概括层次：



表示“是.....一种”(is a)的关系

返回

聚集层次和概括层次的合并:



ER模型的操作

在利用ER模型进行数据库概念设计的过程中，常常需要对ER图进行种种变换。这些变换又称ER模型的操作，包括实体类型、联系类型和属性的分裂、合并和增删等等。

(1) 实体类型的分裂：一个实体类型可以根据需要分裂成若干个实体类型。分裂方式有垂直分割和水平分割两种。

a. **垂直分割：**垂直分割系指把一个实体类型的属性分成若干组，然后按组形成若干实体类型。

b. **水平分割：**水平分割系指把一个实体类型分裂为互不相交的子类（即得到原实体类型的一个分割）。

(2) 实体类型的合并

实体类型合并是实体类型分裂的逆过程，相应地，也有水平合并

和垂直合并两种（一般要求被合并的实体应具有相同的键）。

在实体类型水平分裂时，原有的联系类型也要相应分裂，反之，在水平合并时，联系类型是否改变或分裂要视分裂实际情况而定。相应地，垂直合并时，也可能导致新联系类型的产生。

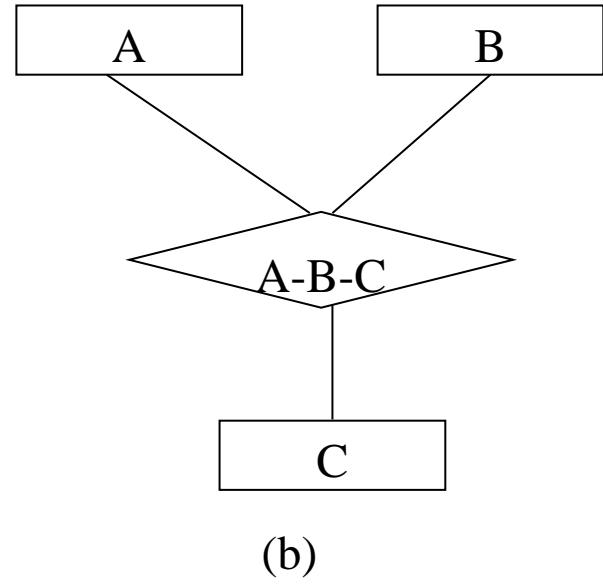
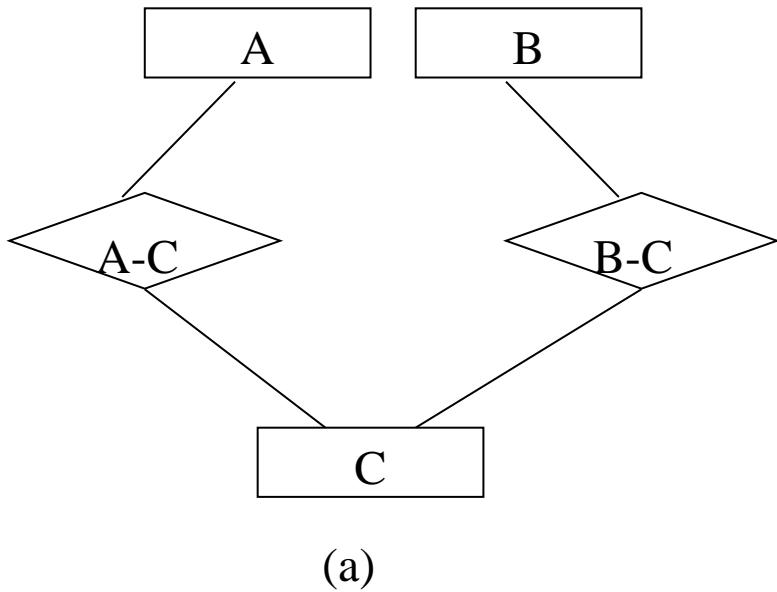
（3）联系类型的分裂

一个联系类型可分裂成几个新联系类型。新联系类型可能和原联系类型不同。

（4）联系类型的合并

联系类型的合并是分裂操作的逆过程。必须注意，合并的联系类型必须是定义在相同的实体类型组合中，否则是不合法的合并，下图的合并就是不合法的合并。

不合法的合并:



必须注意， 合并的联系类型必须是定义在相同的实体类
型组合中， 否则是不合法的合并。

4. 采用ER方法的概念设计步骤

采用ER方法的概念设计步骤分为三步：

第一步：设计局部ER模式：①确定局部结构范围划分；

②实体定义；

③联系定义。

第二步：设计全局ER模式：①确定公共实体类型；

②局部ER模式的合并；

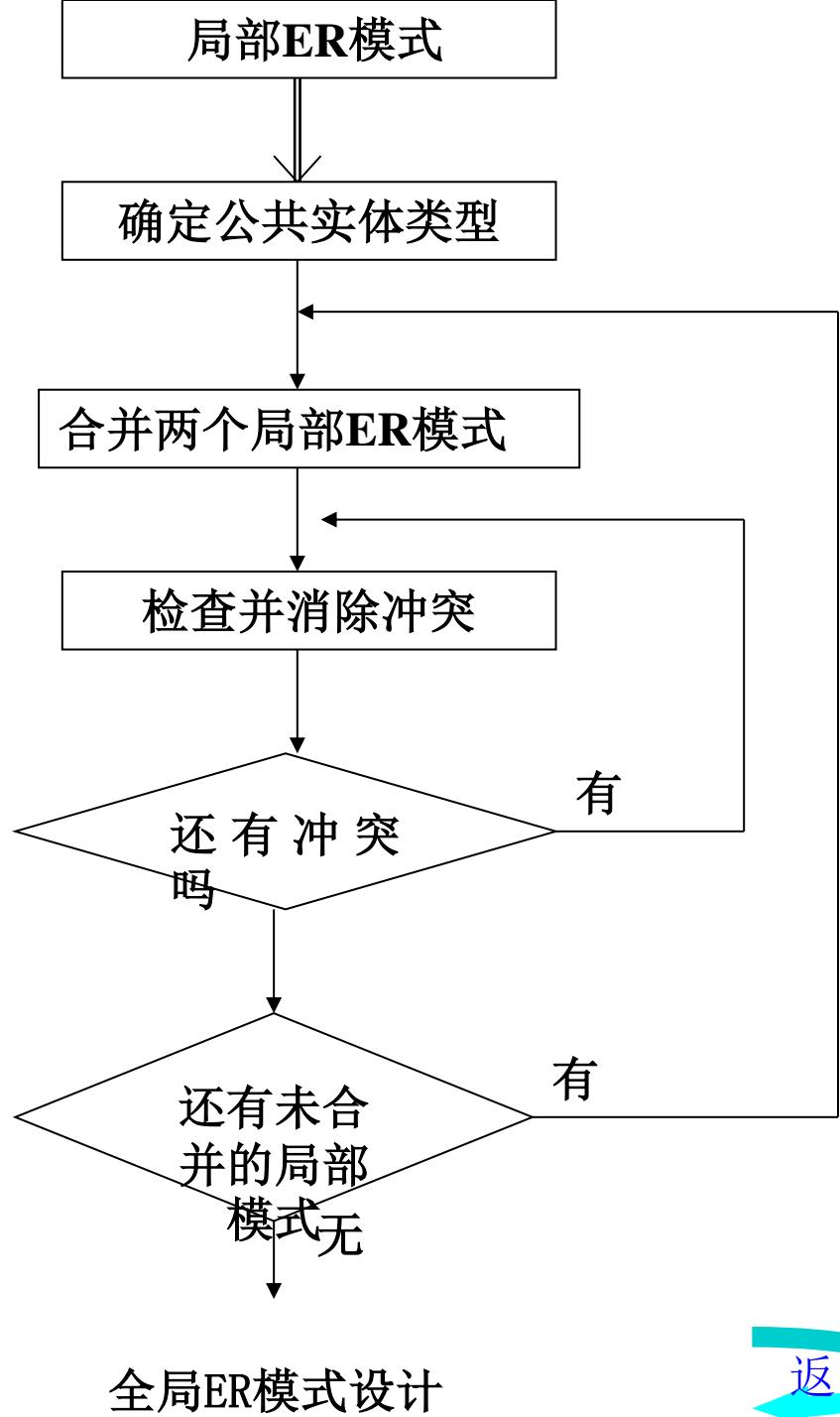
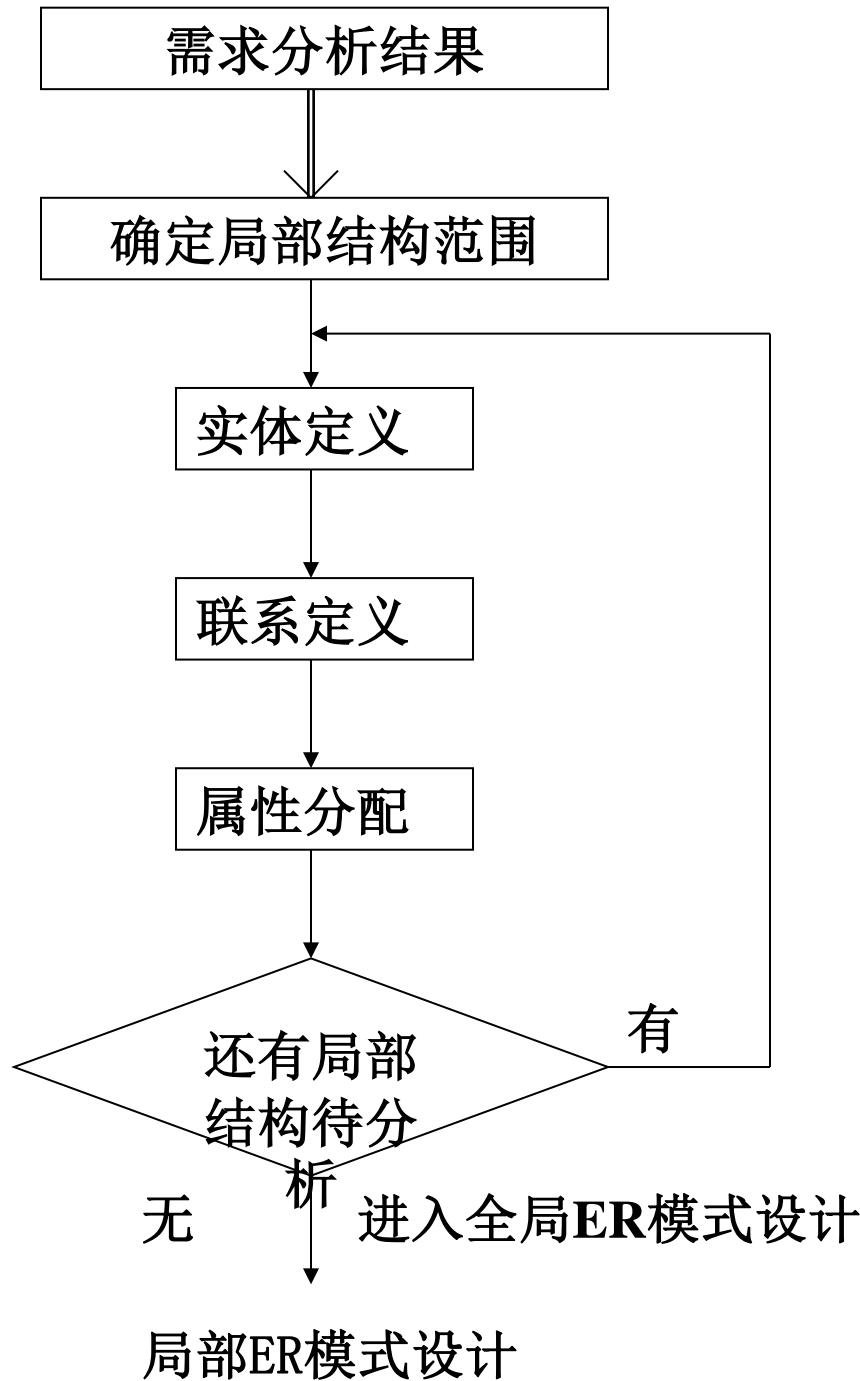
③消除冲突。

第三步：全局ER模式的优化原则：

相关实体类型的合并；

冗余属性的消除；

冗余联系的消除。

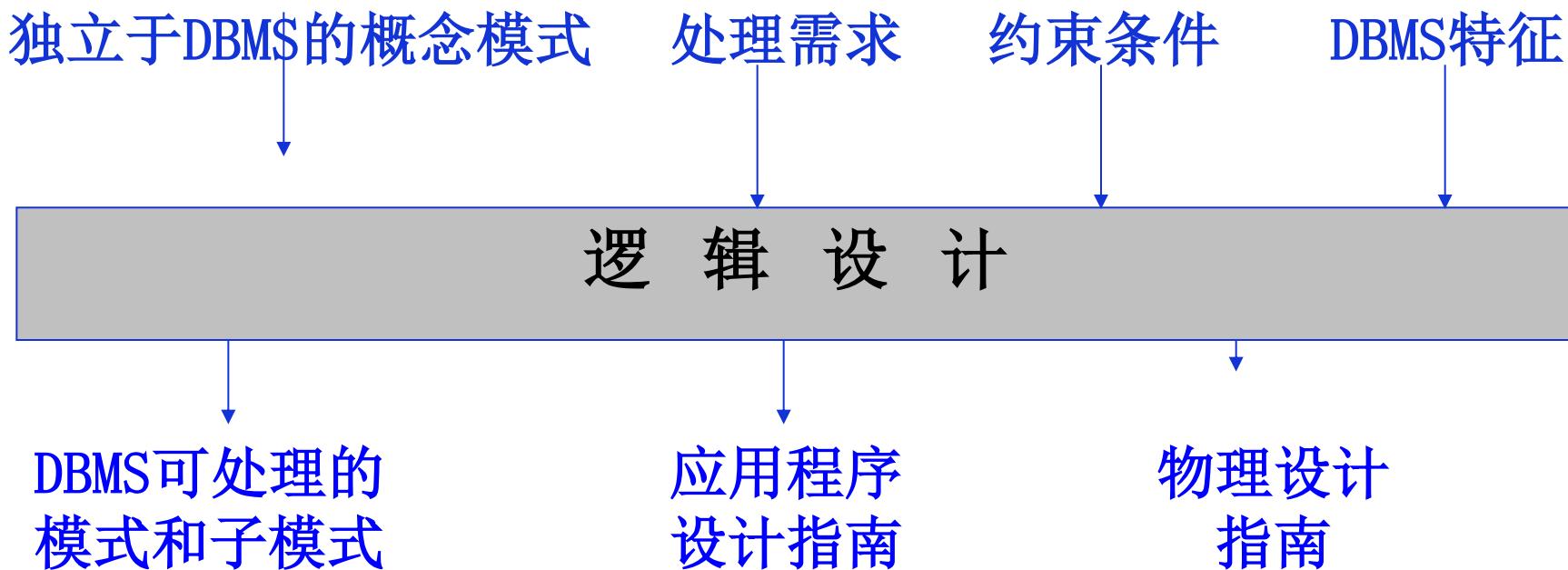


返回

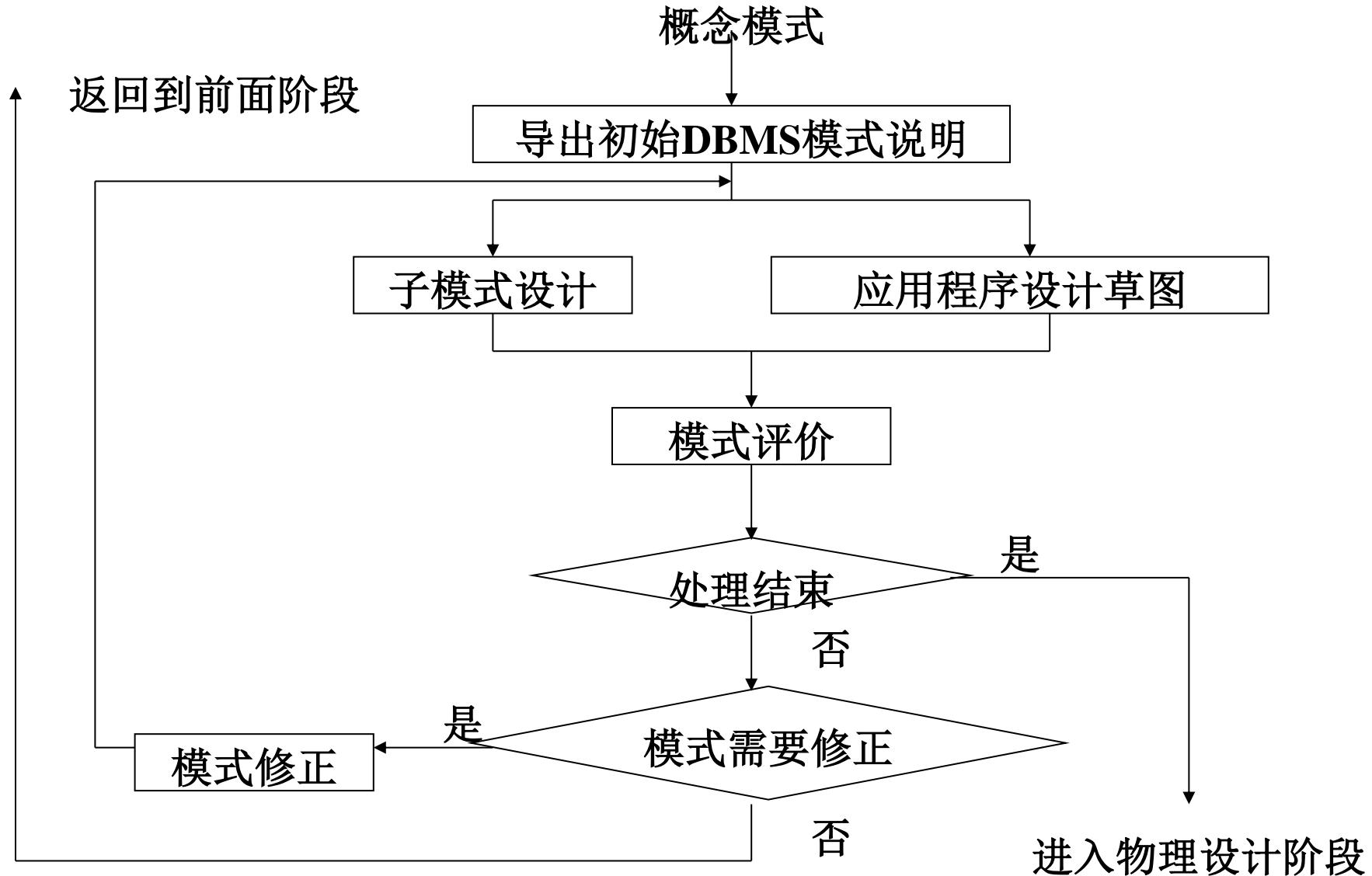
五、逻辑设计

逻辑设计的目的是把概念设计阶段设计好的全局ER模式转换成与选用的具体机器上的DBMS所支持的数据模型相符合的逻辑结构（包括数据库模式和外模式）。

1. 逻辑设计输入输出环境：



2. 逻辑设计的步骤:



3. ER模型向关系模型的转换

(1) ER模型转换为关系模型的一般规则:

① 实体类型的转换: 将每个实体类型转换成一个关系模式, 实体的属性即为关系模式的属性, 实体标识符即为关系模式的键。

② 联系类型的转换, 根据不同的情况做不同的处理:

a 若实体间的联系是1: 1...。

b 若实体间的联系是1: N...。

c 弱实体: 若实体间的联系是1: N的, 而且在N端实体类型为弱实体, 转换成的关系模式中将1端实体类型(父表)的键作为外键放在N端的弱实体(子表)中。弱实体的主键由父表的主键与弱实体本身候选键组成。也可以为弱实体建立新的独立的标识符ID。

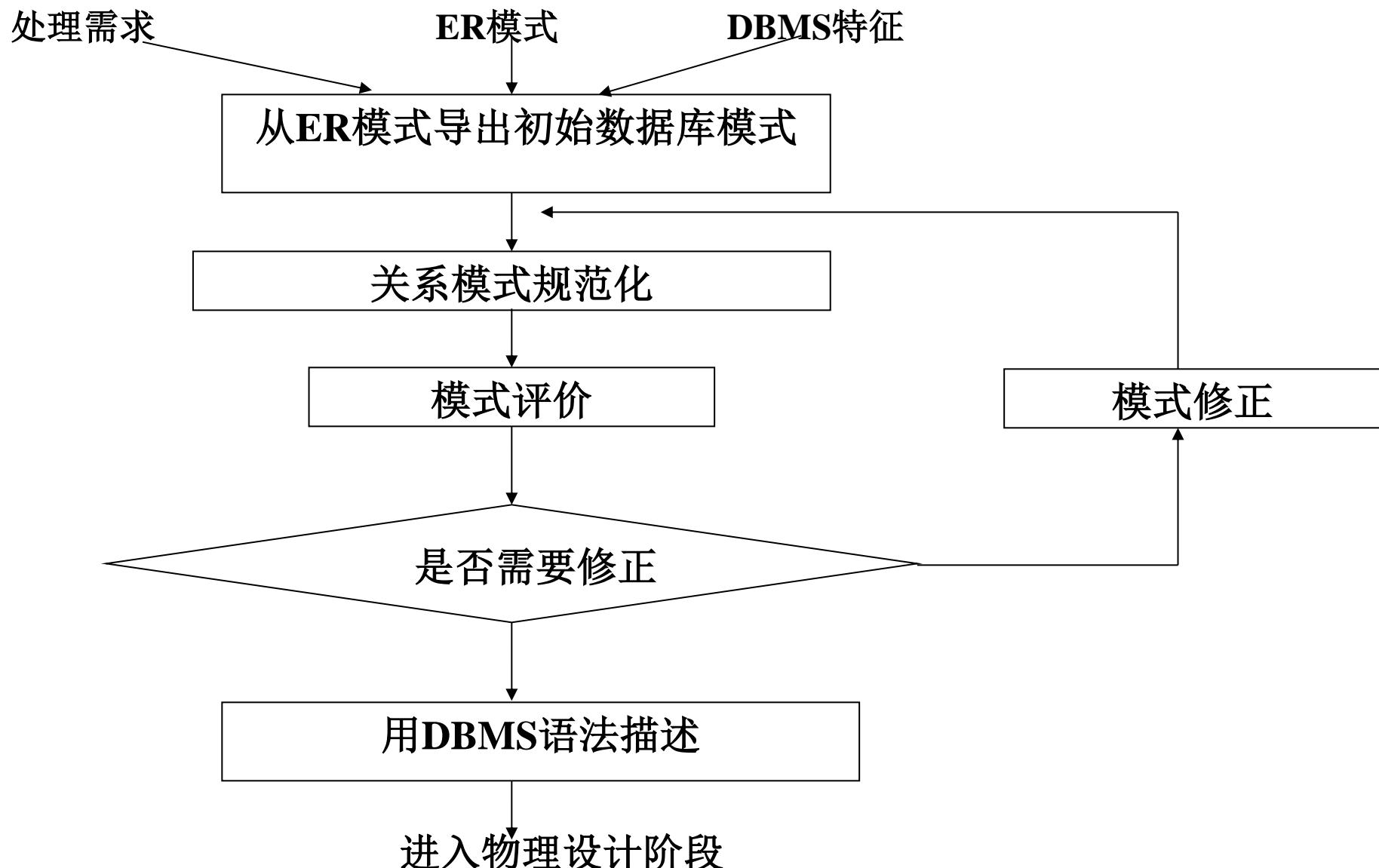
d 若实体间的联系是M: N的，则将联系类型也转换成关系模式，

其属性为两端实体类型的键加上联系类型的属性，而键为两端实体键的组合。

(2) 超类和子类的转换规则

将超类和子类各转换成一个关系模式，在子类转换成的关系模式（子表）中加入超类转换成关系模式（父表）的键，从而实现父表与子表的联系。由于父表与子表的主键相同，所以子表的主键也是外键。

4. 关系数据库的逻辑设计：运用关系数据库模式设计理论，使设计过程形式化地进行，并且结果可以验证。



(1) 导出初始关系模式：按照ER模型向关系模型转换规则将即全局ER模式转换成初始关系模式。

(2) 运用模式设计理论，对初始关系模式进行规范化处理。

规范化过程分为两个步骤：

① 确定规范级别：首先考察数据依赖集合，在仅有函数依赖时，3NF或BCNF是适宜的标准，如还包括多值依赖时，应达到4NF。由于多值依赖语义的复杂性、非直观性，一般使用得并不多。

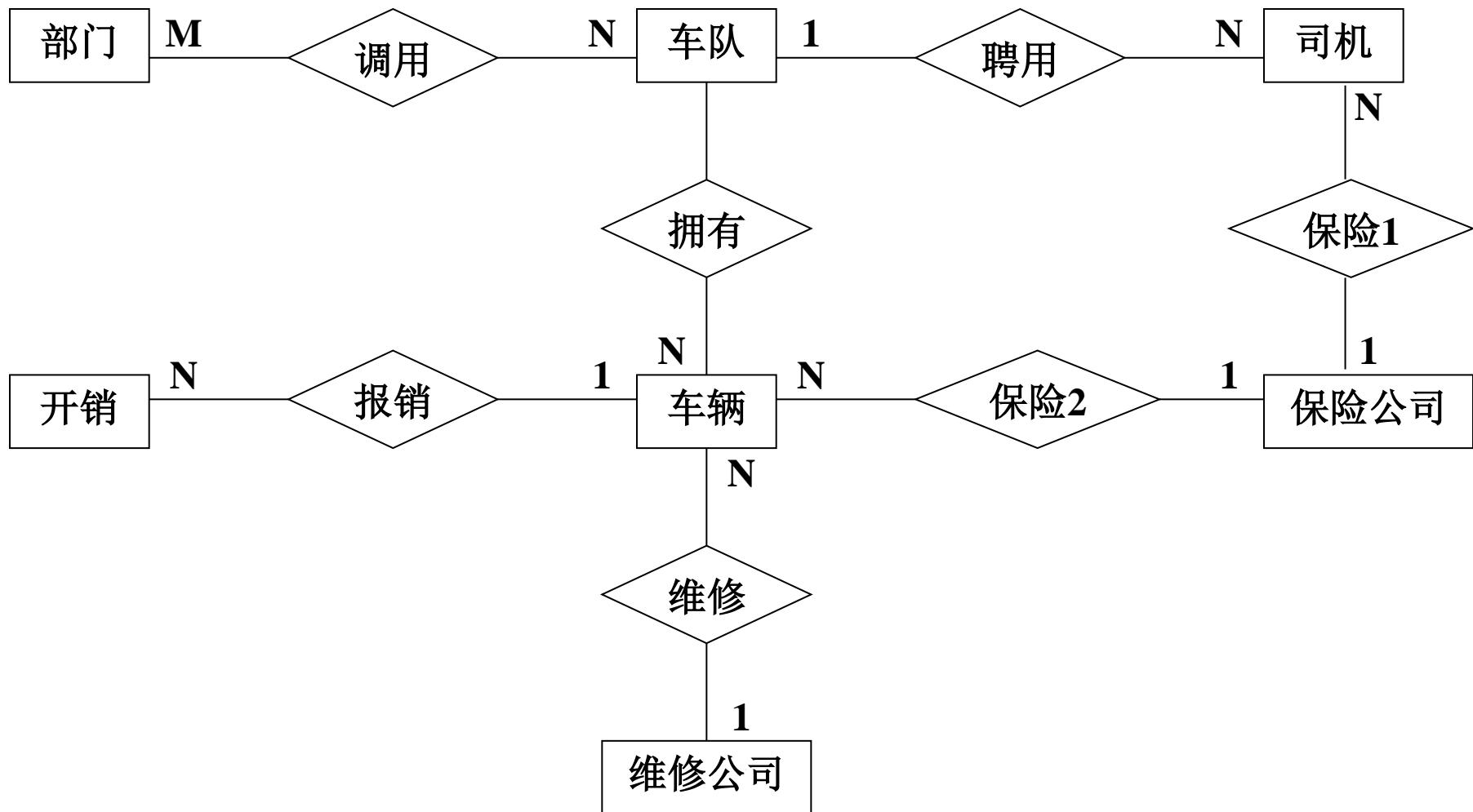
② 实施规范化处理：利用第5章的算法，逐一考察关系模式，判断它们是否满足规范要求。若不符合上一步所确定的规范级别，则利用相应的规范算法将关系模式规范化。在规范化综合或分解过程中，要特别注意保持依赖和无损联接要求。

(3) 模式评价：检查已给出的数据库模式是否完全满足用户的功能要求，是否具有较高的效率，并确定需要加以修正的部分。模式评价主要包括功能和性能两个方面。

(4) 模式修正：根据模式评价的结果，对已生成的模式集进行修正。修正的方式依赖于导致修正的原因，如果因为需求分析、概念设计的疏漏导致某些应用不能得到支持，则应相应增加新的关系模式或属性；如果因为性能考虑而要求修正，则可采用合并、分解或选用另外结构的方式进行。

在经过模式评价及修正的反复多次后，最终的数据库模式得以确定，全局逻辑结构设计即告结束。

例：公司车队信息系统的ER模型



该ER图有7个实体类型， 其结构如下：

部门（部门号, 名称, 负责人）

车队（车队号, 名称, 地址）

司机（司机号, 姓名, 执照号, 电话, 工资）

车辆（车牌号, 车型, 颜色, 载重）

保险公司（保险公司号, 名称, 地址）

维修公司（维修公司号, 名称, 地址）

开销（顺序号, 费用类型, 费用, 日期, 经手人

六、物理设计

数据库最终是要存储在物理设备上的。数据库在物理设备上的存储结构与存储方法称为数据库的物理结构，它依赖于给定计算机系统。数据库的物理结构主要指数据库的存储记录格式、存储记录安排和存取方法。物理设计可分五步完成，前三步涉及到物理结构设计，后两步涉及到约束和具体的程序设计。

- (1) **存储记录结构设计：**包括记录的组成、数据项的类型、长度，以及逻辑记录到存储记录的映射。
- (2) **确定数据存放位置：**可以把经常同时被访问的数据组合在一起，“记录聚簇（cluster）”技术能满足这个要求。
- (3) **存取方法的设计：**存取路径分为主存取路径与辅存取路径，前者用于主键检索，后者用于辅助键检索。

- (4) 完整性和安全性考虑：设计者应在完整性、安全性、有效性和效率方面进行分析，作出权衡。
- (5) 程序设计：在逻辑数据库结构确定后，应用程序设计就应当随之开始。物理数据独立性的目的是消除由于物理结构的改变而引起对应用程序的修改。当物理独立性未得到保证时，可能会发生对程序的修改。

七、数据库的实现

根据逻辑设计和物理设计的结果，在计算机系统上建立起实际数据库结构、装入数据、测试和试运行的过程称为数据库的实现阶段。实现阶段主要有三项工作。

(1) 建立实际数据库结构：对描述逻辑设计和物理设计结果的程序（即“源模式”），经 DBMS 编译成目标模式和执行后建立了实际的数据库结构。

- (2) 装入试验数据对应用程序进行调试。试验数据可以是实际数据，也可由手工生成或用随机数发生器生成。应使测试数据尽可能复盖现实世界的各种情况。
- (3) 装入实际数据，进入试运行状态。测量系统的性能指标，是否符合设计目标。如果不符合，则返回前面几步修改数据库的物理结构，甚至逻辑结构。

规范化理论是数据库设计的指南和工具，具体地讲可

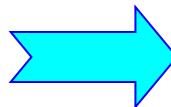
在以下三个方面起重要作用：

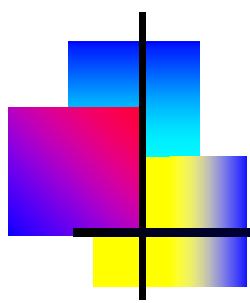
1. 在数据库需求分析阶段用数据依赖的概念来分析和表示各数据项之间的联系；
2. 在概念设计阶段,用规范化理论消除初步ER图中冗余的联系；
3. 在ER图向关系模型转换过程中,用模式分解的概念和算法指导逻辑设计。

八、数据库的运行和维护

数据库系统正式运行，标志着数据库设计与应用开发工作的结束和维护阶段的开始。运行维护阶段的主要任务有四项：

- (1) 维护数据库的安全性与完整性：检查系统安全性是否受到侵犯，及时调整授权和密码，实施系统转储与后备，发生故障后及时恢复。
- (2) 监测并改善数据库运行性能：对数据库的存储空间状况及响应时间进行分析评价，结合用户反应确定改进措施，实施再构造或再格式化。
- (3) 根据用户要求对数据库现有功能进行扩充。
- (4) 及时改正运行中发现的系统错误。要充分认识到，数据库系统只要在运行，就要不断地进行评价、调整、修改。如果应用变化太大，再组织工作已无济于事，那么表明原数据库应用系统生存期已结束，应该设计新的数据库应用系统了。





精读和习题要求

精读教材： P. 143~170

习 题7： P. 171 12、15、15