

第七章 数据库设计

7.1 数据库设计概述

数据库设计定义：数据库设计是指对于一个给定的应用环境，构造（设计）优化的数据库逻辑模式和物理结构，并据此建立数据库及其应用系统，使之能够有效地存储和管理数据，满足各种用户的应用需求，包括信息管理要求（应该存储和管理哪些数据对象）和数据操作要求（对数据对象进行哪些操作）

7.1.1 数据库设计的特点

1、数据库建设的基本规律

数据库设计的特点：三分技术，七分管理，十二分基础数据

管理包括数据库建设项目管理和企业的业务管理

基础数据包括数据的手续、整理、组织和不断更新

2、结构（数据）设计和行为（处理）设计相结合

数据库设计应该和应用系统设计相结合，即整个设计过程中要把数据库结构设计和数据的数据处理设计密切结合起来

传统的软件工程重行为设计，早期的数据库设计重结构设计

7.1.2 数据库设计方法

早期数据库设计主要采用手工与经验相结合的方法，即手工试凑法

缺点：与技术人员的经验与水平有关，设计质量难以保证

规范设计法：基本思想：过程迭代和逐步求精

典型方法：新奥尔良方法，基于 ER 模型方法、3NF 的设计方法、面向对象设计方法、统一建模语言（UML）方法

7.1.3 数据库设计的基本步骤

数据库设计分 6 个阶段：

1、需求分析阶段

了解分析用户需求，包括数据与处理

数据字典、全系统中数据项、数据结构、数据流、数据存储的描述

2、概念结构设计

设计的关键，通过对用户需求进行综合、归纳与抽象形成一个独立于具体 DBMS 的概念模型

概念模型（ER 图）、数据字典

3、逻辑结构设计

将概念结构转换为某个 DBMS 支持的数据模型，并优化

4、物理结构设计

为逻辑数据模型选取最适合应用环境的物理结构（包括存储结构与存储方法）

5、数据库实施

根据逻辑结构和物理结构建立数据库，编写调试应用程序

6、数据库运行和维护

不断运行、评估、修改

需求分析和概念设计独立于 DBMS，逻辑结构设计和物理结构设计与 DBMS 密切相关

性能监测、转储/恢复、数据库重组和重构

7.1.4 数据库设计过程中的各级模式

需求分析阶段

综合各个用户的应用需求

概念设计阶段：

形成独立于及其的特点，独立于各个 DBMS 的 ER 图

逻辑设计阶段

先将 E-R 图转换成具体的数据库产品支持的数据模型，如关系模型，形成数据库逻辑模式，然后根据用户处理的要求、安全性的考虑，在基本表的基础上建立必要的师徒，形成外模式

物理设计

根据 DBMS 特点和处理的需要，进行物理存储安排，建立索引，形成数据库内模式

7.2 需求分析

7.2.1 需求分析的任务

调查重点是数据和崔丽

（1）信息要求

用户要从数据库获得的信息从而导出数据库存储的数据

（2） 处理要求

用户要完成的处理功能以及对处理性能的要求

（3） 安全性与完整性要求

7.2.2 需求分析的方法

调查用户需求的步骤：

- 1、调查组织机构情况
- 2、调查各部分业务活动情况（数据）
- 3、协助用户明确对新系统的各种要求，包括信息要求。处理要求、完整性与安全性要求
- 4、确定新系统的便捷（人机边界）

调查方法：

- 1、跟班作业
- 2、开调查会
- 3、请专人介绍
- 4、询问
- 5、设计调查表请用户填写
- 6、查阅记录

分析表达用户需求方法：

结构化分析方法（SA）：从最上层的系统组织机构入手，自顶向下、逐层分解方式分析系统

7.2.3 数据字典

数据字典是关于数据库中数据的描述，即元数据，不是数据本身。在需求分析阶段建立。

数据字典内容：数据项、数据结构、数据流、数据存储、处理过程

对**数据项**和**数据结构**的定义来描述**数据流**、**数据存储**的逻辑内容

1、数据项

数据项是不可再分的数据单位

数据项描述={数据项名，数据项含义说明，别名，数据类型，长度，取值范围，取值含义，与其他数据项的逻辑关系，数据项之间的联系}

2、数据结构

数据结构反应了数据之间的组合关系，一个数据结构可以有若干个数据项组成，也可由若干和数据结构组成

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

数据流是数据结构在系统内传输的路径

数据流描述={数据流名, 说明, 数据流来源, 数据流去向, 组成: {数据结构}, 平均流量, 高峰期流量}

平均流量: 在单位时间(每天, 每周, 每月等)里传输次数

4、数据存储

数据存储是数据结构停留或保存的地方, 也就是数据流的来源和去向之一

数据存储秒数={数据存储名, 说明, 编号啊, 输入的数据流, 输出的数据流, 组成: {数据结构}, 数据量, 存取拼读, 存取方式}

存取方式: 批处理/联机处理; 检索/更新; 顺序检索/随机检索

5、处理过程

处理过程的具体处理逻辑一般用判定表或判定树类描述。数据字典中只需描述处理过程的说明性信息。

吹过程描述={处理过程名, 说明, 输入: {数据流}, 输出: {数据流}, 处理: {简要说明}}

简要说明: 说明该处理过程的功能及处理要求(处理频度要求等)

注意: (1) 需求分析设计人员应考虑到变化, 应使设计易于更改 (2) 用户参与

7.3 概念结构设计

7.3.1 概念模型

将需求分析得到的用户需求抽象为信息结构的过程就是概念结构设计

特点:

- (1) 真实充分反映现实世界
- (2) 易于理解, 方便和用户沟通
- (3) 易于更改
- (4) 易于向关系、网状、层次等数据模型转换

描述工具: E-R 图

7.3.2 ER 模型

1、实体之间的联系

- (1) 两个实体型之间的联系

一对一联系 (1: 1); 一对多联系 (1: n); 多对多联系 (m: n)

(2) 两个以上的实体型之间的联系

一对一、一对多、多对多

(3) 单个实体型内的联系

一对一、一对多、多对多

2、ER 图

实体型：用矩形表示，矩形框内写明实体名

属性：用椭圆形表示，并用无向边将其余相应的实体型连接起来

联系：用菱形表示。联系可以具有属性

7.3.5 概念结构设计

1、实体与属性的划分原则

遵循原则：现实世界事物尽量当做属性对待

两条准则：

(1) 属性不可分，不能再有需要描述的性质

(2) 属性不能与其他实体具有联系，ER 图表示的是实体之间的联系

2、ER 图的集成

(1) 合并，解决各分 ER 图之间的冲突，生成初步 ER 图

各个子系统的 ER 图之间存在的不一样的地方，称之为冲突

a) 属性冲突

属性域冲突，即属性值的类型。取值范围或取值集合不同

属性单位冲突

b) 命名冲突

同名异义

异名同义

可能发生在实体、联系一级上，也可能发生在属性一级上

c) 结构冲突

同一对象在不同应用中具有不同的抽象

同义实体在不同子系统的 ER 图中所包含的属性个数和属性排列次序不同

实体间的联系在不同的 ER 图中为不同的类型 (一对多，多对多等)

(2) 修改和重构, 消除不必要的冗余, 生成基本的 ER 图

冗余有冗余的数据和冗余的联系

冗余的数据是指可以由基本数据导出的数据

冗余的联系是指可以由其他联系导出的联系

消除冗余主要采用分析方法, 以数据字典和数据流图为依据

规范化理论消除冗余:

a) 确定 ER 图实体之间数据依赖

b) 求 FL 的最小覆盖 GL, 差集为 $D=FL-GL$

逐一考察 D 中的函数依赖, 看是否是冗余的 (D 中的函数依赖不一定冗余)

7.4 逻辑结构设计

逻辑结构的任务就是把概念结构设计阶段设计好的基本 ER 图转换为选用 DBMS 所支持的数据模型相符合的逻辑结构

7.4.1 ER 图向关系模型的转换

将 ER 图转换为关系模型: 将实体型、实体的属性和实体型之间的联系转化为关系模式

转换原则:

1、一个实体型转换为一个关系模式

关系的属性: 实体的属性 关系的码: 实体的码

2、实体型建的联系有以下不同情况

(1) 一个 1: 1 联系可以转换为一个独立的关系模式, 也可以与其他合并

(2) 一个 1: n 联系可以转换为一个独立的关系模式, 也可以与 n 端对应的关系模式合并

(3) 一个 m: n 联系转换为一个关系模式

关系的属性: 与该联系相连的各实体的码以及联系本身的属性

关系的码: 各实体码的组合

(4) 三个或三个以上实体间的一个多元联系转换为一个关系模式

关系的属性: 同上 关系的码: 同上

(5) 具有相同码的关系模式可以合并

7.4.2 数据模型的优化

一般数据模型还需要向特定 DBMS 的模型进行转换

主要依据：以规范化理论为指导

优化方法：

- (1) 确定数据依赖
- (2) 对各个关系模式之间的数据依赖作最小化处理，消除冗余的联系
- (3) 对关系模式逐一分析，确定各关系分别属于第几依赖
- (4) 分析关系模式是否合适，是否需要合并或分解

并不是规范化程度越高的关系就越优

- (5) 对关系模式进行必要分解。有垂直分解和水平分解两种方法

水平分解：把基本关系的**元组**分为若干子集合，定义每一个子集合是一个关系

如何分解：把经常使用的 20% 的数据分解出来，形成一个子关系。水平分解为若干子关系，尽量使每个事务存取对应一个子关系

垂直分解：把关系模式的属性分解为若干子集合，形成若干子关系模式。

原则：把经常在一起使用的属性从 R 中分解出来，不损失关系模式的语义

优点：提高效率 缺点：一些事务不得不执行连接操作，降低效率

适用范围：总效率得到提高

7.4.3 设计用户子模式

设计用户的外模式

- 1、使用更符合用户习惯的别名
- 2、对不同级别的用户定义不同的视图
- 3、简化用户对系统的使用

7.5 物理结构设计

物理结构：数据库在物理设备上的存储结构与存取方法称为数据库的物理结构

定义：为给定的逻辑数据模型选取一个最适合应用要求的物理结构的过程

物理结构设计步骤：1、确定数据库的物理结构 2、对物理结构进行评价

7.5.1 数据库物理设计的内容和方法

1、准备工作

充分了解应用环境，分析要运行的事务，了解 DBMS 提供的存取方法和存储结构

2、选择物理数据库设计所需参数

(1) 数据库查询事务

查询关系，查询条件所涉及的属性，连接条件所涉及的属性，查询的投影属性

(2) 数据库更新事务

被更新关系，更新操作条件涉及的属性，修改操作改变的属性值

(3) 每个事务运行的频率和性能要求

关系数据库物理设计的内容：存取方法（建立存取路径），存储结构

7.5.2 关系模式存取方法的选择

DBMS 常用存取方法

1、B+树索引存取方法

选择索引存取方法：哪些属性建立索引/组合索引，哪些索引要设计为唯一索引

- (1) 若一个（或一组）属性经常在查询条件中出现，则建立索引（组合索引）
- (2) 如果一个属性经常用最大值最小值，则在属性上建立索引
- (3) 如果一个（或一组）经常在连接条件中出现，则建立索引

2、hash 索引存取方法的选择

如果一个关系的属性主要出现在等值连接或是等值比较中，并满足

- (1) 该关系大小可预知，而且不变
- (2) 该关系大小动态改变，但 DBMS 提供 Hash 存取方法

3、聚簇存取方法的选择

定义：为了提高某个（组）属性的查询速度，把这些属性上具有相同值的元组集中存放在连续的物理块中成为聚簇，属性成为聚簇码

一个数据库可以建立多个聚簇索引，一个关系只能加入一个聚簇索引

聚簇索引适用条件：很少对基表进行增删改查，很少对其中变长列进行修改操作

P236

7.5.3 确定数据库的存储结构

确定数据库物理结构主要指数据的存放日志和存储结构

1、确定数据的存放位置

基本原则：易变部分与稳定部分分开存放；经常存取部分与寻去频率较低部分分开存放

2、确定系统配置