

## 07-10-Application

Created on 20220605.

Last modified on 2022 年 6 月 5 日.



# 目录



# Chapter 1 Introduction

1.1 与自然科学相关

1.2 与社会科学相关

1.3 与数据相关

1.3.1 科学计算（计算科学）

1.4 工程与科学计算方面，对数值模拟进行的多学科的应用研究



## Chapter 2 代数（符号）计算





# Chapter 3 计算生物学（生物信息学）

## 3.1 计算神经科学



## Chapter 4 计算化学



# Chapter 5 计算数学

## 5.1 计算数论

## 5.2 数值分析



# Chapter 6 计算物理

## 6.1 计算流体力学

## 6.2 有限元分析





# Chapter 7 计算机辅助工程

## 7.1 CAD/CAM 技术的理论研究、CAD/CAM 系统的软件开发平台研制



# Chapter 8 计算社会学

## 8.1 社区信息学



# Chapter 9    计算经济学

## 9.1    电子商务

## 9.2    计算金融



## Chapter 10 数字人文（人文计算）





# Chapter 11 计算机历史

## 11.1 计算机硬件的历史

## 11.2 计算机科学史



# Chapter 12 数据库

## 12.1 基本概念

数据库, database, 按照数据结构来组织、存储和管理数据的仓库。MySQL, 需要熟练。关系型数据库。程序 + 数据 = 软件。应用程序是操作和查询数据库服务器, 数据库服务器响应和提供数据。数据库: 增删改查。数据库管理系统, dbms。MySQL 免费, 小项目, 一百万以下的数据。甲骨文 ORACLE, 可以提供支持, 大项目。

老的: sybase, 银行、电信系统。db2。

分布式: HBASE, mongoDB。

嵌入式 SQLite SQL Server Postgre SQL 数据库系统 (Database System) 包括数据库 (Database), 数据库管理系统 (Database Management System), 应用开发工具, 管理员及用户。

## 12.2 数据库

B-S 体系中, 系统安装、修改、维护都只在服务器端。

### 12.2.1 三级模式

物理数据库 (对应一个文件) 《-》内模式 (数据的存储形式) 《-》概念模式 (数据划分为表) 《-》外模式 (各种用户视图) 《-》用户

### 12.2.2 数据库设计

需求分析: 数据处理要求、应用要求, 输出数据流图、数据字典、数据说明书;

概念结构设计, 输出 ER 模型【重点】, 用户数据模型;

逻辑结构设计: 转换规则、规范化理论, 输出关系模式, 视图、完整性约束、应用处理说明书;

物理设计: 结合 DBMS 特性、硬件特性、OS 特性等。

## ER 模型

实体，方框表示，如学生、课程；属性，椭圆表示，如学号、姓名；联系，菱形，如学生—M-选课-N—课程。

集成：画局部，合成整体的。集成会存在属性冲突、命名冲突、结构冲突等。

关系 O 对应的三个实体之间的联系数量分别为 A,B,C，则最少转换为 4 个关系模式（每个实体 1 个，关系至少一个）。

## 关系代数

关系代数表达式的等价转换；业务场景对应的关系代数表达式。

并交差笛卡尔积投影选择连接

### 12.2.3 规范化理论

解决数据冗余、更新异常、插入异常、删除异常等。

## 函数依赖

部分函数依赖：如当主键为（学号，课程）的复合键时，可以确定姓名；但是主键的一部分—学号，可以确定姓名。传递函数依赖：A->B->C，有 A->C

## 键

超键：唯一标识元组，可以存在冗余属性。候选键：唯一标识元组，不存在冗余属性。主键：任选一个候选键。外键：其他关系的主键。关联各表。

## 范式

1NF，属性值不可分，原子值；2NF，由 1NF 消除非主属性对候选键的部分依赖得到；3NF，由 2NF 消除非主属性对候选键的传递依赖得到；BCNF，由 3NF 消除主属性对候选键的传递依赖得到。越高级，数据表拆分越细。粒度小，性能下降。

## 模式分解

表的拆分。保持函数依赖：冗余的依赖不需要保留。

无损分解：可以还原。

### 12.2.4 并发

事务机制：一系列操作作为一个整体，原子性、一致性、隔离性、持续性。

PV 操作可实现资源互斥使用。

## 并发问题

丢失更新，T1 与 T2 都读取 A，分别将 A-5，写回。有可能某个进程后写回导致 A 只是-5

不可重读读取：T1 读取数据计算后，在 T1 校验计算结果时，T2 进程修改数据，导致 T1 校验失败。

错误数据读出：T1 修改了数据，T2 读到了修改的数据，T1 进行回滚恢复了数据，导致 T2 的数据无意义。

## 封锁协议

S 封锁，X 封锁。一级：事务 T 修改数据 R 之间必须加 X 锁。防止丢失修改。二级：一级，且 T 读取 R 之间 S 锁，读完释放 S 锁，防止丢失修改、读脏数据。三级：一级，且 T 读取 R 之间 S 锁，事务完成再释放 S 锁，防止丢失修改、读脏数据、数据重复读。两段锁协议：可串行，可能死锁。

更新 = 读 + 写。

### 12.2.5 其他

#### 完整性约束

提高数据可靠性。

实体完整性约束：主键不能为空、重复。参照完整性约束：外键，如员工的部门，应该是对应的部门表中的主键或还未分配部门时为空。用户自定义完整性约束：如性别、年龄的限制

#### 安全

用户标识和鉴定，账户、校验存取控制，用户的操作权限密码存储和传输；视图和保护，视图授权；审计，记录用户对数据库的操作。

#### 数据备份

静态备份，关闭数据库，复制。快速，易归档；动态备份，运行时复制。有选择性备份、恢复某个表，灵活，出错导致问题更大。

备份策略，如连续 7 天：完增增增差增增完全备份增量备份，与上次完全备份的差异增量备份，与上次备份的差异。

海量，转储，

增删改查，先写日志文件，然后再实际处理。

## 故障与恢复

可预期故障：程序中预先设置 rollback 语句；不可预期：如算术溢出、违反存储保护等，可以由 DBMS 的恢复子系统通过日志，撤销对数据库的修改；系统故障：检查点法；介质故障：使用日志。

## 数据仓库

数据仓库特点：面向主题，而一般数据库面向业务组织数据。集成的，相对稳定的，一般不修改与删除。反映历史变化

数据源进行抽取、清理、装载、刷新，得到数据仓库。OLAP 服务器，提供查询、报表、分析、数据挖掘等工具。部门级的数据仓库整合，得到企业级的数据仓库。

## 数据挖掘

方法：决策树、神经网络、遗传算法、关联规则挖掘算法

分类：关联分析，数据之间的隐藏关系分析；序列模式分析，分析数据之间的因果关系；分类分析，记录做标记，按标记分类；聚类分析

## 反规范化技术

规范化程度低，存在冗余；高了导致数据表多，查询效率低。

牺牲空间和规范化程度来提高查询效率。

技术：增加派生性冗余列；增加派生性冗余列；重新组织表；分割表

## 大数据

多种不同类型进行联合分析。

数据量大，PB 级或以上需要快速处理数据多样性数据有价值

cookies 售卖；百度广告推送。

深度分析，关联分析，回归分析

集群平台。

大数据处理系统：高可扩展性；高性能；高度容错；支持异构环境；分析延迟小；开放接口易用；成本低；向下兼容。

## 12.3 分布式数据库

## 12.4 对象数据库

## 12.5 关系数据库

## 12.6 MySQL 基本操作

### 12.6.1 数据库基础操作

### 12.6.2 表

表由行 row 和列 column 组成关系模型（二维表）：不同的属性叫字段；每行数据是一个记录；每个记录都不相同的字段叫主键。

做表，先确定字段，确定每个字段的类型。

### 12.6.3 表间联系

表 a 是学生表，有学生的学号姓名班级等信息，表 b 是课程表，是所有开课课程的序号、课名称、老师等信息，表 a 和 b 没有公共元素，表 c 是选课表，有学号和所选课程号、成绩等信息。表 c 联系起表 a 和 b。需要建 5 个键，3 个表要有 3 个主键，同时表 c 要有 2 个外键投射到表 a 和表 b

一对一的表，常用在登录表和详细表。方便查询。

### 12.6.4 SQL Language

Structured Query Language, 结构化查询语言。包括数据定义语言（DDL），数据操作语言（DML）。

#### DDL 数据定义语言

#### DML 数据操作语言

增删改

INSERT INTO 表名（字段列表）VALUES ('XX','XX'),...,'(XX','XX)'; 日期和字符串要加引号，int 格式不需要引号。注意顺序，数值，数量对应好。

UPDATE 表名 SET 字段名 = 字段值 WHERE 条件 DELETE FROM 表名 WHERE 条件

数据查询最重要：（1）单表查询 SELECT DISTINCT 字段列表 FROM 表名 WHERE 条件  
ORDER BY 排序依据 GROUP BY 分组依据 HAVING 分组筛选条件

limit 语句

单表查询

特殊查询

## DQL 数据查询语言

数据查询

## DCL 数据控制语言

控制数据的使用权限

## 视图

SQL-> 外模式, 包括各个视图 -> 模式, 包括各个基本表, -> 内模式, 包括各个存储文件  
create view 名称 as select 选择的项目

## 索引

普通索引

唯一索引: 可以空

主键索引: 不能空

组合索引

create [unique] index 名称 on 表名 (字段名) 几十万行记录能感觉到几毫秒的速度提高。加入索引会提高查询速度, 增删的速度会降低。大表, 表中数据经常查询, 建索引。like 查询不支持索引

### 12.6.5 数据库设计

需求分析: 分析业务和数据处理要求

概要设计: 设计数据库和 E-R 模型图详细设计: E-R 图转换为多张表, 逻辑设计, 代码编写:

subsubsection 概念模型 实体, 客观世界中的对象实体的属性实体间的联系:n 对 m 的

subsubsection E-R 图 entity-relationship E-R 图: 矩形是实体, 属性是圆形, 联系是菱形比如, 教师-学生-课程; 比如, 论坛用户-发帖-跟帖-版块比如, 商店-顾客

subsubsection E-R 图到关系范式的转换 1 对 1 的关系: 部门-经理, 部门这个表中加入经理编号这个自造的字段, 把经理表中的主键加入。1 对多: 科室-医生, 把医生表加入科室编号。m 对 n: 学生-课程, 构造 3 个表, 中间的联系做一个表, 转化成 1 对多顾客-商品, 构建新表 (订购)

subsubsection 关系范式 评估表的质量, 评估规范标准。1NF, 2NF, 3NF, BCNF, 4NF, 5NF。范式越高, 质量越高, 代码难度越难写。项目快就范式低, 项目人力资源充足, 就提高范式通常



3 范式就可以。1NF，目标是确保每列的原子性，每一列不可分割。不满足 1NF 的数据库不是关系数据库。2NF，实体属性完全依赖主键，不能只依赖主键的一部分。如选课表（学号，姓名，课程号，成绩，系名，系主任），复合主键（学号 + 课程号），这不符合 2NF，因为这里如姓名，姓名之和主键的一部分（学号）有关系，不是和整个主键有关系。更好的做法是分成两个表选课表（学号，课程号，成绩），学生表（学号，姓名，系名，系主任）3NF，在 2NF 基础上，任意非主属性都不传递依赖于主关键字。如系主任之和系名有关，系名和学号有关。相当于把一些部件放在子装配体里面。应该把学生表分成两个表，学生表（学号，姓名，系代号），系别表（系代号，系名，系主任）；

冗余设计：在业务频繁的时候增加冗余。

水平分表与垂直分表：数据大时。水平分表，把数据放在不同表中，如按照学号奇偶划分，大表可以提高查询速度，表中数据有独立性时如记录数据中的不同时期等的的数据。缺点是会增加复杂度。垂直分表，把不常用的信息放在另外表中。管理有冗余，查询所有信息时需要 join 命令。

读写分离数据库备份三份，1 个 master 库用于更新 insert，update，delete，2 个 slave 库用于查询 select，read，

## 12.7 多数据库系统集成技术研究



# Chapter 13    数据处理

13.1    数据管理

13.2    数据挖掘

13.3    信息架构

13.4    信息管理

13.5    信息检索

13.6    知识管理

13.7    知识工程

13.8    多媒体, 超媒体

13.8.1    声音和音乐计算



## Chapter 14 人机交互



## Chapter 15 图像处理与科学可视化





# Chapter 16    Else

- 16.1    高性能存储系统，处理机同步通信机制
- 16.2    超常指令字（VLIW）系统结构
- 16.3    格点计算模型及体系结构的研究
- 16.4    波分复用 WDM 全光网中的路由及波长分配算法的研究
- 16.5    计算机在信息产业中的应用
- 16.6    计算机在制造产业中的应用
- 16.7    各个领域计算机应用软件的开发技术
- 16.8    过程工程
- 16.9    计算机应用工程化
- 16.10    以机器人足球为标准问题的 MAS 体系结构与合作规划
- 16.11    MAS 中的策略协作学习



# Chapter 17 CIMS 及其它先进制造技术

17.1 信息系统

17.2 大规模科学与工程计算

17.3 计算机图形学及可视化技术, 计算机辅助设计