

数据库 (2)

- 数据库管理系统
 - 是一种负责数据库的定义、建立、操作、管理和维护的软件系统
 - 主要职能
 - 数据库的定义和建立
 - 数据库的操作
 - 数据库的控制
 - 数据库的维护
 - 故障恢复和数据通信
- 数据库模式
 - 三级模式
 - 外模式
 - 描述用户看到或使用的那部分数据的逻辑结构
 - 一个数据库可以有多个外模式
 - 一个应用程序只能使用一个外模式
 - 概念模式
 - 描述整个数据库中数据库的逻辑结构，描述现实世界中的实体及其性质与联系，定义记录、数据项、数据的完整性约束条件及记录之间的联系
 - 包含访问控制、保密定义和完整性检查以及概念 / 物理之间的映射
 - 一个数据库只有一个概念模式
 - 内模式
 - 最低层表示，不同与物理层，它假设外存是一个无限的线性地址空间
 - 定义存储记录的类型、存储域的表示、存储记录的物理顺序
 - 是数据在数据库内部的表示方式
 - 一个数据库只有一个内模式
 - 概念模式是数据库的中心和关键
 - 三个级别
 - 用户级数据库
 - 对应外模式，是用户看到和使用的数据库，又称用户视图
 - 概念级数据库
 - 对应概念模式，是所有用户视图的最小并集，是DBA看到的数据库
 - 物理级数据库
 - 对应内模式，描述数据的实际存储组织，最接近物理存储
 - 两级独立性
 - 物理独立性
 - 用户的应用程序与存储在磁盘上的数据库中的数据是相互独立的
 - 逻辑独立性
 - 用户的应用程序与数据库中的逻辑结构是相互独立的

- 逻辑独立性比物理独立性更难实现
- 数据模型
 - 概念数据模型（实体联系模型）E-R模型
 - 基本数据模型（结构数据模型）
 - 层次模型，用树形结构表示实体类型及实体间联系
 - 网状模型，用有向图表示实体类型及实体间联系
 - 关系模型，用表格结构表达实体集，用外键表示实体间联系
 - 面向对象模型，用面向对象观点来描述现实世界实体的逻辑组织、对象间限制和联系等
- 事务
 - 原子性，事务是数据库的逻辑工作单位，事务的所有操作要么全做要么不做
 - 一致性，事务的执行使数据库从一个一致性状态到另一个一致性状态
 - 隔离性，一个事务的执行不能被其它事务干扰
 - 持续性，一个事务一旦提交，它对数据库的改变必须是永久的
- 范式
 - 1NF，若关系模式R的每个关系r的属性值都是不可分的原子值
 - 2NF，R是1NF，每个非主属性完全函数依赖于候选键
 - 3NF，R是1NF，每个非主属性都不传递依赖于R的候选码
 - BCNF，R是1NF，每个属性都不传递依赖于R的候选键
- 并发控制
 - 封锁类型
 - 排他型封锁（X封锁）
 - 共享型封锁（S封锁）
 - 封锁协议
 - 一级封锁协议
 - 修改时加入X锁，结束时释放
 - 防止丢失更新，保证克恢复
 - 不能保证可重复读和不读脏数据
 - 二级封锁协议
 - 一级封锁协议加上读取前加S锁，读完即可释放
 - 防止丢失更新和防止读脏数据
 - 不能保证可重复读
 - 三级封锁协议
 - 一级封锁协议加上读取前加S锁，结束时释放
 - 防止丢失更新和防止读脏数据于数据重复读
 - 两段锁协议
 - 扩展阶段，读写首先申请封锁
 - 收缩阶段，释放一个封锁后不能申请和获得其它封锁
 - 有可能发生死锁
- 数据库完整性
 - 实体完整性

- 主码的任一属性不能为空
- 参照完整性
 - R中的外码只能对S中的主码引用，不能说S中主码没有的值
- 用户自定义完整性
 - 针对某一具体关系数据库的约束条件，它反映某一具体应用所涉及的数据必须满足的语义要求
- 触发器
 - 完整性保护措施
 - 自动进行某些操作以消除违法完整性约束条件所引起的负面影响
 - 还有安全性保护功能
- 数据库性能
 - SQL性能优化
 - 数据库性能优化
 - 查询优化
 - 把数据、日志、索引放到不同的设备上
 - 纵向横向分割表
 - 根据查询条件建立优化索引，优化访问方式，限制结果集的数据量
 - 用OR的语句可以分解成多个查询，使用UNION连接
 - 在SELECT语句中用WHERE限制返回的行数
 - 注意使用DISTINCT
 - 在IN后面的列表中，将最频繁出现的值放在前面
 - 在GROUP BY和HAVING之前剔除多余的行
 - 尽量将数据的处理放在服务器上，减少网络开销
 - 不要在一个语句中多次使用同一函数
- 数据库工程
 - 数据库设计
 - 规划
 - 必要性和可行性分析
 - 需求分析
 - 信息收集
 - 信息分析整理
 - 信息评审
 - 概念设计
 - 对需求说明书提供的数据和要求进行抽象于综合处理，构造反映用户环境的数据及其相互联系的概念模型
 - 概念模型应真实、充分反映现实世界中事物和事物之间的联系
 - 概念模型应简洁、明晰，独立于机器、容易理解
 - 概念模型应易于变动
 - 概念模型应很容易向关系、层次或网状等数据模型转换
 - 逻辑设计
 - 目的是把上一阶段生成的基本E-R图，转换为与具体DBMS所支持的数据模型相符合的逻辑结构

- 输入
 - 独立于DBMS的概念模式
 - 处理需求
 - 约束条件
 - DBMS特性
 - 输出
 - DBMS可处理的模式
 - 子模式
 - 应用程序设计指南
 - 物理设计指南
- 物理设计
 - 对一个给定的逻辑数据模型选取一个最适合饮用环境的物理结构的过程
 - 步骤
 - 设计存储记录结构
 - 确定数据存储安排
 - 设计访问方法
 - 进行完整性和安全性分析设计
 - 程序设计
- 数据库建模
 - E-R图要素
 - 实体：矩形框
 - 属性：椭圆形
 - 实体间联系：菱形框
 - 1:1
 - 1:n
 - n:n
 - E-R图集成
 - 一次集成（难度大）
 - 逐步集成
 - 冲突
 - 属性冲突
 - 命名冲突
 - 结构冲突
 - E-R图向关系模式转换
 - 一个实体转换为一个关系模式
 - 一个1:1联系可以转换为一个独立的关系模式，也可以与任意一端对应的关系模式合并
 - 一个1:n联系可以转换为一个独立的关系模式，也可以与任意n端对应的关系模式合并
 - 一个m: n联系转换为一个独立的关系模式，与该联系相连的各实体的码以及联系本身的属性均转换为关系的属性
 - 三个以上实体间的一个多元联系可以转换为一个独立的关系模式

- 常见数据库管理系统

- Oracle
- Sybase
- Informix
- SQL Server
- DB2

- 并行数据库系统

- 高性能
- 高可用性
- 可扩充性
- 结构
 - SM结构
 - 多个处理机、一个共享内存和多个磁盘存储器
 - 实现简单、负载均衡
 - 成本高，可扩充性差，可用性不是很好
 - SD结构
 - 多个独立内存的处理机和多个磁盘构成
 - 成本低、可扩充性好、可用性强、负载均衡
 - 实现复杂、潜在性能问题
 - SN结构
 - 多个处理节点组成，各节点有独立的处理机、内存、磁盘
 - 成本低、可扩展性、高可用性
 - 实现复杂、负载难以均衡

- 数据仓库与数据挖掘

- 联机事务处理（PLTP）与联机分析处理（OLAP）

OLTP	OLAP
● 操作人员，低层管理人员	决策人员，高层管理人员
● 日常操作处理	分析决策
● 面向应用	面向主题
● 当前数据	历史数据，多维集成
● 数十条记录	上百万记录
● 简单事务	复杂查询
● 上千用户	上百用户
● 100M ~ 100G	100G ~ 100T

- 数据仓库

- 面向主题的、集成的、相对稳定的、随时间变化的数据集合，用于支持管理决策
- 特征
 - 面向主题
 - 集成的
 - 相对稳定的
 - 随时间变化
- 分类

- 企业仓库
- 数据集市
- 虚拟仓库
- 结构
- 结构
 - 数据源
 - 数据准备区
 - 数据仓库数据库
 - 数据集市 / 知识挖掘库
 - 管理工具和应用工具
- 数据挖掘
 - 从大量的、不完全的、有噪声的、模糊的、随机的实际应用数据中，提取隐含在其中的、人们事先不知道的，但又潜在有用的信息和知识的过程
 - 功能
 - 自动预测趋势和行为
 - 关联分析
 - 聚类
 - 概念描述
 - 偏差检测
 - 常见技术
 - 决策树
 - 神经网络
 - 遗传算法
 - 关联规则挖掘算法
 - 分析方法
 - 关联分析，发现不同事件至今的关联性
 - 序列分析，发现一定时间间隔内接连发生的事件
 - 分类分析，分析具有类别的样本的特点，得到决定样本属于各种类别的规则或方法
 - 聚类分析，根据物以类聚的原理，将本身没有类别的样本聚集成不同的组
 - 预测，根据样本的已知特征估算某个连续类型的变量的取值的过程
 - 时间序列分析，随时间变化的事件序列，目的是预测未来发展趋势，或者寻找相似发展模式或者是发现周期发展规律
 - 流程
 - 问题定义
 - 建立数据挖掘库
 - 分析数据
 - 调整数据
 - 模型化
 - 评价和解释

幕布 - 思维概要整理工具
