

数据库系统(DNS)复习概要

数据库系统分为三部分：知识体系及概念、三种主要的运算、大数据知识部分。

知识体系及概念

1. 数据库系统=相互关联的数据的集合（数据库）+访问这些数据的程序。

数据库系统主要有如下几个方面的内容：

(1) 数据视图；

(2) 数据模型：用于描述数据、数据关系等的工具的集合；通常分为 4 类：

{ 关系模型：工具－表，用于记录；关系模型－表的集合；
E－R模型：实体－联系模型基于对现实世界的一组认识，
提供一个找出在数据库中表示的实体以及实体间如何关联的方法；
对象－关系模型：E－R 模型的扩展版；
半结构化数据类型：允许相同类型的数据项含有不同的数据定义。

(3) 数据库语言 = { 数据库定义语言(DDL)：定义数据库模式；
+
数据库操纵语言(DML) { 查询
更新：添加、删除和修改。

2. 数据库系统设计的目的：

- (1) 为用户提供方便高效使用数据信息的方法；
- (2) 为用户提供数据关系的抽象视图；
- (3) 保证数据的独立性。

3. 数据的独立性=物理独立性+逻辑独立性

物理独立性：应用程序与存储器中的数据保持独立关系；

逻辑独立性：应用程序与数据库的逻辑结构保持独立关系；

数据和程序的独立，简化了程序，减少了应用程序的修改和维护。

4. 文件处理系统，DNS 诞生之前的“数据库系统”，就是数据和程序的简单堆积。

有以下缺点：

- (1) 数据的冗余和不一致；
- (2) 数据访问困难；
- (3) 数据孤立；
- (4) 完整性问题：数据库中的值必须满足某些一致性约束；
- (5) 原子性问题：事件/指令发生故障时，可以恢复；
- (6) 并发访问异常；
- (7) 安全性问题。

5. 事务：数据库中一些操作的集合称为单元，单一逻辑工作单元的操作集合称作事务。

事务具有以下性质：

- (1) 原子性：一个事务的执行只有执行成功和不执行(程序发生故障时，数据要复原)；
- (2) 隔离性：并发时，一个事务的程序不会被其他事务干扰；
- (3) 一致性：数据库中的数据在执行一个事务前后都必须保持一致；
- (4) 持久性：一个事务完成后，它对数据库的改变必须是永久的。

6. 事务的隔离性即操作系统中的互斥，为**并发操作**提供了最基本的保证。具有以下优点：
- (1) 提高吞吐量和资源利用率；
 - (2) 减少等待时间。

并发通过并发控制机制来实现程序并发，其目的是通过调度实现事务的可串行化。

7. 事务的**隔离性级别**有如下几个：

- (1) **可串行化**；
- (2) **可重复读**：只允许读已提交的数据且可以重复；
- (3) **已提交读**：只允许读已提交的数据但不可以重复；
- (4) **未提交读**：允许读取未提交的数据。(最低的一致性级别)。

8. 实现并发且不允许脏写的方法：**锁、时间戳和快照隔离**。

9. 恢复系统可保证事务的原子性和持久性，主要要做到事务程序发生故障时，**数据不丢失且快速恢复**。

10. 形式化关系查询语言

$$\left\{ \begin{array}{l} \text{过程化语言：关系代数} \\ \text{非过程化语言：元组关系、域关系} \end{array} \right\} \left\{ \begin{array}{l} \text{一元运算：更名、投影、选择} \\ \text{二元运算：并、笛卡儿积、差} \\ \text{附加运算：交、赋值、自然链接、外链接、聚合} \end{array} \right.$$

第二部分：大数据相关概念

1. **大数据 (big data)**：指无法在一定时间范围内用常规软件工具进行捕捉、管理和处理的数据集合，是需要新处理模式才能处理，具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产。
2. **适应于大数据的技术**：大规模并行处理 (MPP) 数据库，数据挖掘电网，分布式文件系统，分布式数据库，云计算平台，互联网，和可扩展的存储系统。
3. 在维克托·迈尔-舍恩伯格及肯尼斯·库克耶编写的《大数据时代》中大数据指不用随机分析法 (抽样调查) 这样捷径，而采用所有数据进行分析处理。大数据的 **5V 特点** (IBM 提出)：Volume (大量)、Velocity (高速)、Variety (多样)、Value (低价值密度)、Veracity (真实性)。
4. **作用**：
 - (1) 对大数据的处理分析正成为新一代信息技术融合应用的结点。移动互联网、物联网、社交网络、数字家庭、电子商务等是新一代信息技术的应用形态，这些应用不断产生大数据。云计算为这些海量、多样化的大数据提供存储和运算平台。通过对不同来源数据的管理、处理、分析与优化，将结果反馈到上述应用中，将创造出巨大的经济和社会价值。大数据具有催生社会变革的能量。但释放这种能量，需要严谨的数据治理、富有洞见的数据分析和激发管理创新的环境；
 - (2) 大数据是信息产业持续高速增长的新引擎；
 - (3) 大数据利用将成为提高核心竞争力的关键因素；
 - (4) 大数据时代科学研究的方法手段将发生重大改变。例如，抽样调查是社会科学的基本研究方法。在大数据时代，可通过实时监测、跟踪研究对象在互联网上产生的海量行为数据，进行挖掘分析，揭示出规律性的东西，提出研究结论和对策。
5. **大数据时代存储所面对的问题**：
 - (1) 容量；

- (2) 延迟；
- (3) 安全；
- (4) 成本；
- (5) 灵活性；
- (6) 数据的累积和一致性问题。

第三部分：SQL 查询操作、关系查询语言(关系代数)、E-R 模型、函数依赖和范式判定

数据库系统考核制试题

- 1、数据库查询语言 关系代数、SQL；
- 2、E-R 模型和范式判定；
- 3、大数据

4、

简述数据库管理系统的主要功能

5、

什么是关系数据库系统中的事务？请简要给出事务的性质

6、

十、什么是 **NoSQL** 数据库？请结合自己的理解，谈一谈数据库兴起的原因（7 分）

7、

在关系数据库管理系统中，用于实现数据完整性保护的措施有哪些？

8、

在关系数据库管理系统中，什么是数据独立性？数据库管理系统是如何实现数据独立性的？

9、

数据库系统是一个多用户共享系统，必须确保用户事务并发执行的正确性。问：不正确的事务并发执行可能产生哪几种类型的错误？请简要给出每一种错误的产生原因。

答案

4、数据库管理系统(DBMS)的主要功能：

(1) **数据定义功能**。DBMS 提供相应数据语言来定义 (DDL) 数据库结构，它们是刻画数据库框架，并被保存在数据字典中。

(2) **数据存取功能**。DBMS 提供数据操纵语言 (DML)，实现对数据库数据的基本存取操作：检索，插入，修改和删除。

(3) **数据库运行管理功能**。DBMS 提供数据控制功能，即是数据的安全性、完整性和并发控制等对数据库运行进行有效地控制和管理，以确保数据正确有效。

(4) **数据库的建立和维护功能**。包括数据库初始数据的装入，数据库的转储、恢复、重组织，系统性能监视、分析等功能。

(5) **数据库的传输**。DBMS 提供处理数据的传输，实现用户程序与 DBMS 之间的通信，通常与操作系统协调完成。

5、数据库中一些操作的集合称为单元，单一的逻辑单元称为事务。事务具有原子性、隔离性、一致性和持久性。

6、NoSQL 数据库是一种与关系数据库截然不同的数据库管理系统，它的数据储存格式可以是松散的、通常不支持 Join 操作，并且支持横向扩展，也可以称之为非关系数据库。

云计算时代或者大数据时代对数据库技术提出了新的需求，主要体现在：

- (1) 海量级数据；
- (2) 大规模集群管理，分布式应用可以更加简单地部署、应用和管理；
- (3) 低延迟读写速度；
- (4) 建设以及运营成本。

云计算时代或者大数据时代，传统的关系型数据库暴露了很多的缺点：

- (1) 支持容量有限；
- (2) 高并发读写速度慢，当数据量达到一定级别时，关系型数据库地系统逻辑非常复杂，很容易发生死锁等问题，严重影响读写速度；
- (3) 扩展性差，传统地关系型数据库很难横向扩展；
- (4) 建设和运营成本高。

NoSQL 数据库是一种应大数据时代而生地数据库，相对于传统地关系型数据库，NoSQL 数据库非常关注对数据高并发读写和海量数据的存储，在架构和数据模型方面作了简化，在扩展和并发方面做了增强。其常用模型有以下三种：

- (1) 列式数据库 (Column-Oriented)。储存数据时围绕着“列”，而不是传统那样围绕着行；
- (2) 键-值 (Key-Value)。一个 key 对应一个 value；
- (3) 文档 (Document)。

7、数据库的完整性是指数据的正确性、一致性和相容性，确保数据库种的数据可以成功和正确地更新，防止错误地数据进入数据库造成无效操作。

数据完整性包括：实体完整性、域完整性、参照完整性和用户定义完整性。

实现数据完整性的措施主要包括：

- (1) 约束。定义了可输入表或者表的单个列种数据的限制条件，约束独立于表结构；

- (2) 默认值。在没有为某列指定数据时，使用默认指定数值；默认值可以是常数也可以是公式；
- (3) 规则。规则时对已存入数据的规定和限制；
- (4) 属性约束。包括非空值约束、域约束、检查子句等；
- (5) 全局约束。

8、数据的独立性包含逻辑独立性和物理独立性，其中，物理独立性是指程序与存储器中的数据相互独立；逻辑独立性是指程序与数据间的逻辑结构保持独立。

数据库管理系统通过**三级模式结构和两级映射**实现数据的独立性。三级模式结构包括：外模式、概念模式和内模式。两级映射是指当整个系统要求改变模式时（增加记录类型、增加数据项，由 DBMS 对各个**外模式 / 概念模式**的映像做相应改变，从而保证了**数据的逻辑独立性**；当数据的存储结构改变时，由 DBMS 对**概念模式 / 内模式**的映像做相应改变，从而保证了**数据的物理独立性**。

9、不正确的并发导致数据不一致性，主要体现在**丢失修改、读“脏”数据、不可重复读和产生“幽灵”数据**：

- (1) 丢失修改，两个事务一起读入并修改某一数据，后写入的修改了先写入数据；
- (2) 脏数据，指不正确的数据。事务 A 要读取事务 B 写入的数据，B 成功写入但未成功提交，数据恢复到 B 执行前，但数据被 A 读取，产生不正确的数据；
- (3) 不可重复读，前一个事务修改了数据，导致后面的事务不能在读写该数据；