

今天你背完了吗？

反正我没背完

db1

## 小结

- 数据管理技术发展的三个阶段
- 数据模型
- 数据库系统结构

- 1.信息是一种已经被加工为特定形式的数据
- 2.从大量原始数据中抽取和推导出有价值信息的加工过程称为数据处理
- 3.数据库技术的软件实现就是数据库管理系统DBMS
- 4.数据库系统是基于数据库管理系统建立的具有特定数据处理功能的系统
- 5.数据库中存储的数据是集成的和共享的
- 6.在描述数据时，不止要描述数据本身，还要描述数据之间的联系，使整个组织的数据结构化
- 7.数据结构化是数据库的主要特征之一，是数据库与文件系统的根本区别
- 8.数据独立性：1.数据的物理独立性：数据的存储结构（物理结构）改变时，数据的逻辑结构可以不变，从而应用程序也不必改变；2.数据的逻辑独立性：数据的逻辑结构改变时，应用程序可以不变。
- 9.数据库提供了两方面的映像（转换）功能：1.存储结构（内模式）和全局逻辑结构（模式），即模式/内模式映像，保证物理独立性；2.全局逻辑结构（模式）和局部逻辑结构（外模式），外模式/模式映像，保证逻辑独立性。
- 10.数据库系统具有统一的数据控制功能包括：1.数据的安全性控制（保护数据防止不合法的使用导致数据的泄密、更改和破坏）2.数据的完整性控制（正确性相容性）3.并发控制（对多用户的并发操作进行控制、协调，保护数据的完整性）4.数据库恢复（将数据库从某一状态恢复到某一已知的正确状态）
- 11.概念模型用于信息世界建模，是现实世界到信息世界的抽象；数据模型用于机器世界，按计算机系统的观点对数据建模。
- 12.数据模型的三要素：1.数据结构（静态特性）；2.数据操作（动态特性）；3.完整性约束
- 13.数据结构由描述对象、描述对象之间联系组成，是数据静态特性的描述，是刻画数据模型最重要的方面
- 14.数据操作包括操作和操作规则，是数据动态特性的描述，数据库主要有检索和更新（插、删、改）两大类操作

15.完整性约束是完整性规则的集合，完整性规则是给定的数据模型中数据及其联系的所有制约和依存规则，用以保证数据的正确、相容。

16.常见的数据模型有三种：1.层次模型（用树结构表示数据之间的联系）；2.网状模型（用图结构）；3.关系模型（用二维表）

17.层次模型特点：1.有且只有一个节点没有双亲节点，称为根节点；2.根节点意外的所有节点都有且只有一个双亲节点。

18.网状模型特点：1.允许一个以上的节点没有双亲；2.一个节点可以有多个双亲。

19.模式也称为概念模式和逻辑模式，是数据库中全体数据的逻辑结构和特性的描述，是三级模式的核心，用模式DDL(Data Description Language模式描述语言)进行定义

20.外模式也称为子模式或者用户模式，是个别用户的数据视图，即与某一应用有关的数据的逻辑表示，用外模式DDL进行定义

21.内模式也称为存储模式，是数据在数据库系统内部的表示，即对数据的物理结构和存储方式的描述，用内模式DDL进行定义

22.三级模式结构优点：1.保证数据独立性；2.简化用户接口，方便用户使用；3.利于数据共享；4.利于数据的安全保密

23.DBMS的功能：1.数据库定义功能；2.数据库存取功能；3.数据库运行功能；4.数据组织、存储和管理；5.数据库的建立和维护功能；

24.数据库系统的构成：数据库(DB)，数据库管理系统(DBMS)，应用系统，数据库管理员(DBA)

25.DBMS的组成：1.语言编译处理程序；2.系统运行控制程序；3.系统建立和维护程序；4.数据字典。

26.数据库系统的特点：1.数据结构化；2.数据共享度高、冗余度低且易扩充；3.数据独立性好；4.由DBMS统一管理和控制。

27.数据的完整性是为了防止数据库中存在不符合语义的数据，也就是防止数据库中存在不正确的数据，防范对象是不合语义的、不正确的数据；数据的安全性是保护数据库防止恶意破坏和非法存取，防范对象是非法用户和非法操作。

## db2

---

# 本章小节

---

- 关系模型的基本概念

- 关系模型的三要素

- 关系代数

- 从数学角度，基本关系代数运算有5种：并、差、乘、选择、投影
  - 从数据库角度，核心的关系代数运算为：选择、投影、连接（或自然连接）

- 元组关系演算与域关系演算

- 关系语言的组成与特点

28. **关系的性质**：1.列是同质的，即每一列中的分量出自相同的域，拥有相同的数据类型；2.不同列的域可以相通，但不同列必须有不同的列名；3.列的顺序无所谓，即不同列的次序可以更换；4.任意两个元组不能完全相同；5.行的顺序无所谓，即不同行的次序可以更换；6.每一分量必须是不可再分的数据，满足这一条件的关系称其满足第一范式（1NF）

29.关系模型的语义约束：1.实体完整性；2.参照完整性；3.用户定义的完整性

30.关系模型必须支持的约束条件：1.实体完整性；2.参照完整性

31.实体完整性：主码不能为空

32.参照完整性：参照关系的外码要么为空值，非空时一定对应目标关系的主码的某个值。参照关系和目标关系不一定不一样。

33.关系数据操作的特点是集合操作，关系数据操作的基础是关系运算，分为代数运算和逻辑运算两种。

34.从数学角度，基本关系代数类型有五种：1.并；2.差；3.笛卡尔积；4.选择；5.投影。

35.从数据库角度，核心的关系代数运算为选择、投影、连接。

36.关系运算中，把不产生无限关系和无穷验证的运算称为安全运算，其运算表达式称为安全表达式，对其所采取的限制称为安全约束。

37.关系代数是安全运算，关系演算不一定是安全运算。

38.数据定义语言DDL包括外模式DDL，模式DDL，内模式DDL，如"create,drop,alter"

39.数据操纵语言DML有三种基本操作方法：插入、删除、修改，如"update,delete,insert"

40.数据检索语言DQL有查询，如“select”

41.数据库有四种基本操作方式：查询、插入、删除、修改

42.数据操纵语言DML有联机交互方式和宿主语言方式两种，联机交互方式下的DML称为自含式语言，可独立使用，适用于终端直接查询；宿主语言方式下的DML称为嵌入式语言，依附于宿主语言，嵌入到高级语言的程序中。

43.数据控制语言DCL，完成数据库的安全性控制、完整性控制、并发控制，如"grant,revoke"

44.关系数据语言的特点包括：1.一体化；2.非过程化；3.面向集合的存取方式；4.既可独立使用又可以与主语言嵌套使用。

45.关系数据语言的核心是查询，故也称为查询语言。

46.关系运算是设计关系数据语言的基础，关系运算的分类也决定了关系语言的分类。

## db3

# 小结

- 概念

- 基本表与视图；关系数据库的三级模式结构

- SQL的特点

- SQL的数据定义、查询、更新、控制功能

- 视图的作用及SQL的视图操作

47.SQL的特点：1.综合统一（一体化）；2.高度非过程化（非过程化）；3.面向集合的操作方式；4.以同一种语法结构提供两种使用方式（独立或嵌套）；5.语言简洁，易学易用。

48.基本表是实际存在的，每个表在存储空间中用一个存储文件来表示；导出表是由基本表导出的表，包括视图和快照两种。

49.关系数据库的三级模式结构为：内模式(stored file)，模式(base table)，外模式(view)

50.并、差、交运算的操作对象必须是相容的，是同类关系，即具有相同数量的属性列，每个属性列的域也必须是相同的。

51.视图的作用：1.简化用户操作；2.使用户能过从多种角度看待同一数据；3.提供了一定的逻辑独立性；4.能够对数据进行安全保护。

52.空值与另一个值的算术运算结果为NULL，与另一个值的比较运算结果为UNKNOWN，逻辑运算结果直接从字面理解即可

53.嵌入式SQL的意义是：嵌入式SQL把SQL语句嵌入到高级语言中，把SQL的最佳特性与程序设计语言的最佳特性结合起来，使SQL功能更强，灵活性更强。

54.对于嵌入式SQL，DBMS多采用预编译的方式处理，把嵌入在程序中的SQL语句翻译成高级语言（主语言）代源码，然后按主语言的通常方式进行编译、连接形成可执行代码。

55.动态SQL允许在程序运行过程中临时组装SQL语句。

## db4

# 小结

## • 函数依赖

- 定义，三种类型函数依赖，函数依赖的公理系统，函数依赖集的闭包，属性关于函数依赖集的闭包，最小函数依赖集

## • 范式

- 1NF,2NF,3NF,BCNF



## • 多值依赖与第四范式

## • 模式分解的理论

- 模式分解遵循的原则，到3NF和BCNF分解算法

## • 候选码的求解理论和算法

56.数据依赖是指，一个关系内部的属性之间相互依赖又相互制约的关系，分为函数依赖和多值依赖两种。

57.码的性质：1.唯一性：唯一地表示关系中的元组；2.最小性：去掉主码中的任意一个属性，则主码失去标识的唯一性。

58.Armstrong公理是有效的，完备的，有效性指：由F根据Armstrong推出的每个函数依赖一定在F所包含的函数依赖中，完备性指：F重的任意一个函数依赖都可以由F根据Armstrong公理导出。

59.最小依赖集满足的三个条件：1.右部单属性化；2.没有多余的FD；3.每个FD左部没有多余属性

60.一个低一级范式的关系模式，通过模式分解可以转化成若干个高级范式的关系模式的集合，这一过程称为规范化。

61.当一个关系只包含原子值这一约束时称为1NF

62.满足1NF的关系中，每个非主属性完全依赖于码时称为2NF（即1NF消除非主属性对码的部分函数依赖）

63.满足2NF的关系中，每个非主属性都不传递依赖于码时称为3NF

64.BCNF关系模式有如下三个性质：1.所有非主属性都完全函数依赖于某个候选码；2.所有主属性都完全函数依赖于每个不包含它的候选码；3.没有任何属性完全依赖于非码的任何一组属性

65.多值依赖：X, Y, Z为U的子集，且 $Z=U-X-Y$ ，关系模式R中多值依赖成立，当且仅当对于任意的关系r，给定的(x,z)值有一组y的值，这组值仅仅取决于x而与z无关。(就是说我交换了两个元组Y上的值后不影响Z上的取值)

66.多值依赖于函数依赖的区别：1.多值依赖的有效性与属性集的范围有关；2. $X \twoheadrightarrow Y$ 成立，无法断言 $X \rightarrow Y$ 成立，其中 $Y' \in Y$ （比较显然 考虑其判定条件即可）

67.满足BCNF的关系中，不存在非平凡的非函数依赖的多值依赖则称为4NF（即只存在函数依赖，或者平凡的多值依赖）

68.规范化的设计原则为“一事一地”，规范化的实质是概念的单一化。

69.无损分解（无损连接性）的判定准则是： $U_1 \wedge U_2 \rightarrow U_1 - U_2$ 或者 $U_1 \wedge U_2 \rightarrow U_2 - U_1$ （即R1R2的公共属性必须构成二者之一的候选码）

70.投影分解应当遵守的原则：1.保持无损连接性；2.保持函数依赖。

71.若分解要求保持函数依赖，则模式分解总可以达到3NF，但不一定能达到BCNF；若分解要求保持无损连接性，则模式分解总可以达到4NF或者更高；若分解既要保持函数依赖，又要保持无损连接性，则可以达到3NF，但不一定能达到BCNF。

## db5

# 小结

- 数据库设计的阶段以及各阶段的任务
- E-R法以及系统概念结构设计自底向上的方法
- 逻辑结构与物理结构设计

72.数据库设计的六个阶段：1.需求分析；2.概念结构设计；3.逻辑结构设计；4.物理结构设计（存储结构和存取方法）；5.数据库实施；6.数据库运行和维护。

73.需求分析的目标：调查的重点是“数据”和“处理”，包括：1.处理要求；2.信息要求；3.安全性和完整性要求

74.需求分析阶段的数据字典可以看作是数据元素表，数据库实施阶段建立起的数据字典是数据库系统的重要组成部分

75.应用程序的概念结构设计方法：1.自底向上；2.自顶向下；3.逐步扩张；4.混合策略

76.局部ER图设计步骤：1.选择局部应用；2.利用数据抽象机制建立实体模型；3.确定实体之间的联系类型，形成E-R图

77.建立实体模型的关键是确定实体及其属性



78.数据抽象机制分为三类：1.分类：定义某一概念作为现实世界中一组对象的类型；2.聚集：定义某一类型的组成成分；3.概括：定义类型之间的子集联系

79.实体模型的调整原则：1.属性必须是不可再分的数据项；2.属性与其他实体不可以有联系；3.属性与其对应的实体必须保持1:1或者n:1的关系

80.合并ER图的冲突包括：1.属性冲突；2.命名冲突；3.结构冲突

81.关系模式的分解：1.水平分解（28法则）；2.垂直分解（把常用的属性分解出去形成一个单独的子模式 必须保持无损连接性和函数依赖）

82.数据库物理设计中，常用的关系存取方法有三种：1.索引方法；2.聚集方法；3.HASH方法

## db6

---

### 小结

---

- 物理存储系统
- 数据存储结构/物理结构
- 缓存管理
- 索引

83.块/页是存储分配和数据交换的单位

84.索引文件由索引项（索引记录）构成，包括两个域：1.索引域（搜索码）；2.指针

85.索引有两种基本类型：1.排序索引；2.哈希索引

86.聚集索引又称为主索引，非聚集索引又称为辅助索引

87.稠密索引为每个搜索码值都建立一个索引项，稀疏索引只有当文件记录以索引域排序时才可以使用，非聚集索引都是稠密索引

88.B树叶子结点存记录的指针，B+树叶子结点存的是记录而不是记录的指针

## db7

---

# 小结

- 查询处理的步骤
- 查询操作的实现算法
- 查询的代数优化和物理优化的概念和原则
- 查询优化的一般步骤

89.关系查询处理分四个阶段：1.查询分析；2.查询检查；3.查询优化；4.查询执行

90.对于驻留在磁盘上的大型数据库，从磁盘访问数据的I/O代价通常是最重要的代价

91.实现查询操作的算法--选择：1.全表扫描法；2.索引扫描法

92.实现查询操作的算法--排序：1.内存能够完全容纳的关系采用快速排序等算法；2.内存不能够完全容纳的关系采用外排序-归并算法

93.实现查询操作的算法--连接：1.嵌套-循环法；2.索引链接法；3.排序-合并法；4.Hash-join法

94.嵌套循环法中，缓冲区块数为k，外表分配k-1块，内标分配1块，则访问块的次数为： $b_r + \frac{b_r}{k-1} * b_s$ ，其中br为外表块数，bs为内标块数，由此可见应当选较小的表做外表。

95.排序合并法只适用于等值连接或者自然连接，访问块的次数为 $b_r + b_s$

96.Hash join法只适用于等值连接或者自然连接

97.表达式执行的可选方法包括：1.物化方法；2.流水线方法

98.查询的执行总开销=I/O代价+CPU代价+内存代价，最重要的是I/O代价

99.查询优化的目标是选择一个高效执行的查询处理策略，使查询代价最小，即访问磁盘的块数最少

100.按照优化的层次，查询优化分为代数优化（改变代数表达式中操作的次序和组合）和物理优化（存取路径和底层操作算法的选择）

101.代数优化指通过对关系代数表达式的等价变换来提高查询效率

102.查询树的启发式优化：1.选择运算尽早执行（减少中间关系--减少元组数量）；2.投影运算尽早执行（减少中间关系--减少属性数量）；3.把投影运算和选择运算同时进行，把投影同其前或其后的双目运算结合起来（减少扫描关系的数目）；4.把某些选择运算同它前面的笛卡尔积结合起来形成连接运算；5.找出公共子表达式，把公共子表达式的结果写入中间文件避免重复使用（中间结果复用）

103.物理优化的常用方法：1.基于规则的启发式优化方法；2.基于代价估算的优化方法；3.两者结合的优化方法

104.选择操作的启发式规则：1.对于小关系用全表扫描法；2.对于大关系用索引扫描法

105.连接操作的启发式规则：1.如果两个表都已经按照连接属性排序就用排序-合并法；2.如果有表在连接属性上建立了索引用索引连接法；3.如果在连接属性上未排序也未建立索引用Hash join法；4.否则用嵌套-循环法，将小表做为外表。



106.查询优化的一般步骤：1.把查询转换成语法树，如关系代数语法树；2.利用查询树的启发式优化转换成优化后的标准形态；3.利用基于启发式规则的物理优化，选择底层存取路径。生成查询计划，利用基于代价的物理优化，选择代价最小的。

## db8

# 小 结

- 事务的定义以及特性
- 数据库恢复的原理
- 故障的种类以及恢复策略
- 检查点技术
- 并发操作可能导致的问题
- 并发控制的原理以及三级封锁协议
- 多粒度锁概念
- 活锁与死锁的处理
- 可串行化调度的概念及两段锁协议

107.事务（Transaction）是用户定义的数据库操作序列，这些操作要么都做要么都不做，是一个不可分割的工作单位。

108.事务的特性：1.原子性（Atomicity）事务中包含的操作要么都做，要么都不做；2.一致性（Consistency）事务执行的结果一定是使数据库从一个一致性状态变到另一个一致性状态；3.隔离性（Isolation）一个事物的执行不能被其他事务干扰。即一个事物内部的操作和使用的数据对于其他并发事务是隔离的，并发执行的事物之间不能相互干扰；4.持久性（Durability）一个事物一旦提交以后，他对数据库的影响必须是永久的。其他操作或故障不应对其执行结果有任何影响。

109.利用数据库的并发控制机制和数据库恢复机制可以保证事务的特性不遭到破坏

110.1.原子性由恢复机制实现；2.一致性由恢复机制和完整性约束实现；3.隔离性通过并发控制实现；4.持久性通过恢复机制实现

111.造成原子性破坏的原因是故障；造成一致性破坏的原因是原子性被破坏；造成隔离性被破坏的原因是事务的并发执行产生错误；造成持久性被破坏的原因是原子性被破坏

112.事务是数据库恢复和并发控制的基本单位

113.事务开始Begin Transaction；事务正常结束Commit；事务不正常结束Rollback

114.数据库恢复是指将数据库从错误状态恢复到某一已知的正确状态

115.数据库恢复是通过数据库管理系统的恢复子系统完成的

116.故障主要有四种：1.事务内部的故障；2.系统故障；3.介质故障；4.计算机病毒

117.数据库恢复的基本原理为冗余

118.日志是用来记录事务对数据库进行更新操作的文件，主要有以记录为单位和以数据块为单位两种格式

119.事务故障和系统故障恢复必须使用日志文件

120.动态转储必须使用日志文件，日志文件和后援副本一起使用

121.必须先写日志文件在写数据库

122.事务故障的恢复使用UNDO（反向扫描日志文件进行逆操作）

123.系统故障的恢复使用UNDO+REDO（故障发生前已经提交的事务REDO 故障发生时未提交的事务UNDO）

124.介质故障的恢复：1.装入某一后援副本使数据库恢复至最近一次转储时的一致状态；2.装入转储以后的日志文件副本重做已经完成的事务

125.事务并发执行的优点：1.提高系统吞吐量；2.减少平均响应时间

126.并发操作导致的数据不一致有三种：1.丢失更新（读入统一数据进行修改）；2.“脏”数据的读出（事务撤销后 修改也撤销 但另一个事物读了的事修改后的脏数据）；3.不能重复读（T1读取数据后 T2更新数据 T1再读取内容不一致）

127.并发控制采用的主要方法是封锁机制

128.排它锁（X锁）：加锁事务可以读写，其他事务不可以上任何锁

129.共享锁（S锁）：加锁事务只能读，其他事务可以加S锁但不能加X锁

130.一级封锁协议：事务T在修改数据R之前必须对其增加X锁，直到事务结束后再释放（防止丢失更新）

131.二级封锁协议：一级封锁协议加上事务T在读取数据R之前必须对其增加S锁，读完后即可释放（防止“脏”数据的读出）

132.三级封锁协议：一级封锁协议加上事务T在读取数据R之前必须对其增加S锁，知道事务结束后再释放（防止不可重复读）

133.封锁对象的大小称为封锁粒度（一般与并发度和封锁开销负相关）

134.显示封锁：应事务的要求直接加到当前的数据对象的封锁；隐式封锁：该数据对象没有独立加锁，由于其上级节点加锁导致该数据对象加上了锁

135.常见的意向锁有三种：1.意向共享锁（IS）；2.意向排他锁（IX）；3.意向共享排他锁（SIX 先加S再加IX）

136.死锁的预防有两种方法：1.一次封锁法；2.顺序封锁法

137.死锁的检测有两种方法：1.超时法；2.等待图法

138.死锁恢复的通常策略是选择一个死锁代价小的事务将其撤销

139.多个事务的并发结果是正确的，当且仅当其结果与按某一次序串行执行时的结果相同，称这种调度策略为可串行化调度

140.可串行化是事务并发调度正确性的准则

141.两段锁协议的两个阶段：1.扩展阶段（获得封锁）；2.收缩阶段（释放封锁）

142.两段锁协议的内容：1.对数据进行读写操作前要获得对该数据的封锁；2.释放一个封锁后，事务不在获得别的封锁

143.若所有事务均遵守两段锁协议，则这些事务的所有并发调度都是可串行化的

144.遵守两段锁协议的事务仍有可能发生死锁

## db9

### 小 结

- 数据库安全性控制与完整性控制的基本概念
- 安全性控制常用技术，包括用户标识与鉴别，自主存取控制，强制存取控制，以及视图、审计等
- 完整性约束条件的种类、表示及实现机制

145.数据库的安全性是指保护数据库以防止不合法的使用造成的数据泄露、更改和破坏（用户）

146.保障数据库安全的主要方法：1.用户标识与鉴别；2.多层存取控制；3.视图机制；4.审计；5.数据加密。其中用户标识与鉴别是DBMS提供的最外层安全保护机制

147.存取控制机制主要包含两个部分：1.用户权限定义；2.合法权限检查

148.用户权限定义和合法权限检查机制一起组成了DBMS的安全子系统

149.存取方法主要分为：1.自主存取控制（DAC）；2.强制存取控制（MAC）

150.SQL中可以授予用户两类权限：1.用户级权限（使用整个数据库权限的限定）；2.关系级权限（使用关系和视图权限的限定）

151.MAC中主体的敏感度标记(Label)称为许可证级别，客体的敏感度标记称为密级

152.MAC中仅当主体的许可证级别大于等于客体的密级时，该主体才能读取相应的客体；仅当主题的许可证级别等于客体的密级时，该主体才能写相应的客体

153.数据完整性是指数据的正确性和相容性

154.数据完整性控制是为了防止数据库中存在不符合语义的数据，防止错误信息的输入和输出

155.安全性防范的是非法用户和非法操作，完整性防范的是不合语义的数据

156.施加在数据库数据之上的语义约束条件称为完整性约束约束条件

157.完整性约束有两种：1.静态约束（反应数据库状态合理性的约束）；2.动态约束（反应数据库状态变迁的约束）

158.静态约束分为：1.静态列级约束；2.静态元组约束；3.静态关系约束。动态约束同理

159.数据库完整性控制包含三个功能：1.定义功能；2.检查功能；3.违约响应

160.完整性约束条件按照完整性约束的检查时机分为两种：1.立即执行约束；2.延迟执行约束

161.用户标识和认证是系统提供的最外层安全保护措施，标识是指系统采用一定的方式标识其用户或应用程序的名字或身份，认证是指系统在其用户或应用程序登录时判断其是否为合法的授权用户。

162.存取控制确保合法的用户按照指定的权限使用DBMS和访问数据，而非法用户或不具有相关权限的用户则不能

163.审计，把用户对数据库的所有操作都自动记录下来放入审计日志中，DBA可以利用审计跟踪的信息，重现导致数据库现有状况的一系列事件，找出非法存取数据的人、时间、内容等

164.数据加密，防止数据库中数据在存储和传输中失密的有效手段。加密的基本思想是根据一定的算法将原始数据(明文)变换为不可识别的格式(密文)，从而使得不知道解密算法的人无法获知数据的内容。

## db10

### 小 结

- 什么是分布式数据库系统，分布式数据库系统的特点
- 分布式数据库系统的模式结构
- 数据分片
- 分布透明性
- 分布式查询和优化
- 分布式事务处理技术

165.分布式数据库以数据分布为前提，强调场地自治性（局部应用）和自治场地之间的协作性（全局应用）

166.分布式数据库的特点：1.数据独立性；2.集中与自治相结合的控制结构；3.适当增加数据冗余；4.全局的一致性、可串行性和可恢复性

167.分布式数据库的数据独立性包括：1.逻辑独立性；2.物理独立性；3.分布独立性（分布透明性）

168.分布模式的映像类型决定了分布式数据库是冗余的还是非冗余的

169.数据分片的方式有四种：1.水平分片；2.垂直分片；3.导出分片；4.混合分片

170.数据分片的约束：1.完全性；2.不相交性；3.可重构性

171.分布透明性（分布独立性）包括：1.分片透明性；2.位置透明性；3.局部数据结构透明性

172.DDBMS的组成包含四个部分：1.LDBMS（局部场地上的数据库管理系统）；2.GDBMS（全局数据库管理系统）；3.GDD（全局数据字典）；4.CM（通信管理）

173.DDBMS按照全局控制方式可分为：1.全局控制分散的DDBMS（这类结构也称为完全分不的DDBMS）；2.全局控制部分分散的DDBMS

174.DDBMS按照局部DBMS的类型可分为：1.同构型DDBMS；2.异构型DDBMS

175.分布式数据库的查询可分为三类：1.局部查询；2.远程查询；3.全局查询

176.分布式查询的处理过程为：1.查询分解；2.选择操作执行的次序；3.选择执行操作的方法

177.分布式查询的开销=I/O代价+CPU代价+通信代价，其中通讯代价是查询优化的首要目标

178.分布式查询优化可分为查询策略的 分布优化 和 局部优化，其中 分布优化 更重要。

179.通讯代价的一般计算公式： $TC(X) = C_0 + X * C_1$

180.连接分布式查询优化中经常采用半连接来缩减关系，进而节省传输开销

181.分布式数据库系统中，一个全局事务被划分为分布在许多节点上的子事务，分布事务的原子性是：组成该事务的所有子事务，要么一致地全部提交，要么一致地全部回滚

182.分布式事务管理主要包括：1.事务的恢复；2.并发控制

183.保证分布式事务的原子性最常使用两段提交协议

184.两段提交协议把一个分布事务的所有局部事务管理分为两类：1.协调者；2.参与者

185.两段提交协议的内容：第一阶段：协调者征求意见做决定；第二阶段：参与者执行决定

## Manipulate

---

1.使用元组关系演算表达五种基本关系代数类型。

2.安全关系演算和关系代数互推。

3.SQL语句的使用，如：DISTINCT,BETWEEN AND,ORDER BY ASC/DESC；表自身的连接(table X, table Y)；外连接(\* or +) 注意括号是实际存在的 表示一个空行 mysql里面使用left join等方式来实现)；ANY ALL；IN NOT IN；EXISTS NOT EXISTS(尤其是用not exists表达“任意和蕴含”很重要)；UNION、MINUS、INTERSECT；GROUP BY [HAVING]；

4.SQL库函数的使用，如：COUNT，SUM，AVG，MAX，MIN，COUNT默认按列计数COUNT(\*)按行计数

5.SQL数据定义：基本表的创建(CREATE TABLE)、删除(DROP TABLE)、修改(ALTER TABLE)；视图的创建(CREATE VIEW)、删除(DROP VIEW)；索引的创建(CREATE INDEX)、删除(DROP INDEX)

6.SQL数据更新：UPDATE(set)，INSERT(values)，DELETE

7.空值的检索:IS NULL,IS NOT NULL

8.SQL数据控制：GRANT TO(授权),REVOKE FROM(收回权限)

9.求属性关于函数依赖集的闭包算法：Armstrong

10.函数依赖集F的极小化算法：去掉F中的某个关系看是否满足条件

11.多值依赖的判定条件：元组t和元组s，其中t[X]=s[X]，若交换其在Y上的值仍然能在关系中找到对应的元组，则称X->->Y

12.无损连接性的判定算法：nk表，取交集

- 13.保持函数依赖性的判定算法：把所有F取并为G，看对于F中的每一个函数依赖是否属于G+
- 14.达到3NF保持函数依赖的分解算法：F最小化，U去掉多余属性，按左部相同进行分组，一定要注意分组时属性集的包含关系
- 15.达到3NF保持函数依赖和无损连接的分解算法：在算法14的基础上查分解是否包含主码，不包含就加进去
- 16.达到BCNF的无损连接分解算法：把左部非码的函数依赖剥掉
- 17.快速求解候选码的算法：L, R, N, LR
- 18.单属性依赖集合候选码的图论求解方法：关键属性集和独立回路，候选码为关键属性集数目乘以各个独立回路节点数
- 19.多属性依赖集合候选码的图论求解方法：关键属性集，从LR类累加属性求闭包，直到求完所有属性
- 20.画ER图，并消除冗余（写出最小依赖集）
- 21.ER图转换为关系模型（一般只有n:m联系才需要变成关系）
- 22.构造查询树（select投影 from笛卡尔积 where选择）
- 23.查询树优化（选择下沉，构造连接，投影下沉）
- 24.利用检查点技术进行数据恢复（检查点时的事务进UNDO 检查点后新发起的事务进UNDO 检查点后提交的事务进REDO）
- 25.绘出意向锁相容矩阵
- 26.可串行化调度的判定（交换不冲突操作的次序）
- 27.Grant和Revoke的详细使用(grant to public [With Grant Option],revoke from)
- 28.完整性规则的表述（D:Data;O:Operation;A:Assertion;C:Condition;P:Procedure）
- 29.绘出分布式数据库的模式结构