数据库考点汇总

基本概念：三级模式-两级映像、数据库设计

数据库模型：E-R模型、关系模型、关系代数（结合SQL语言）

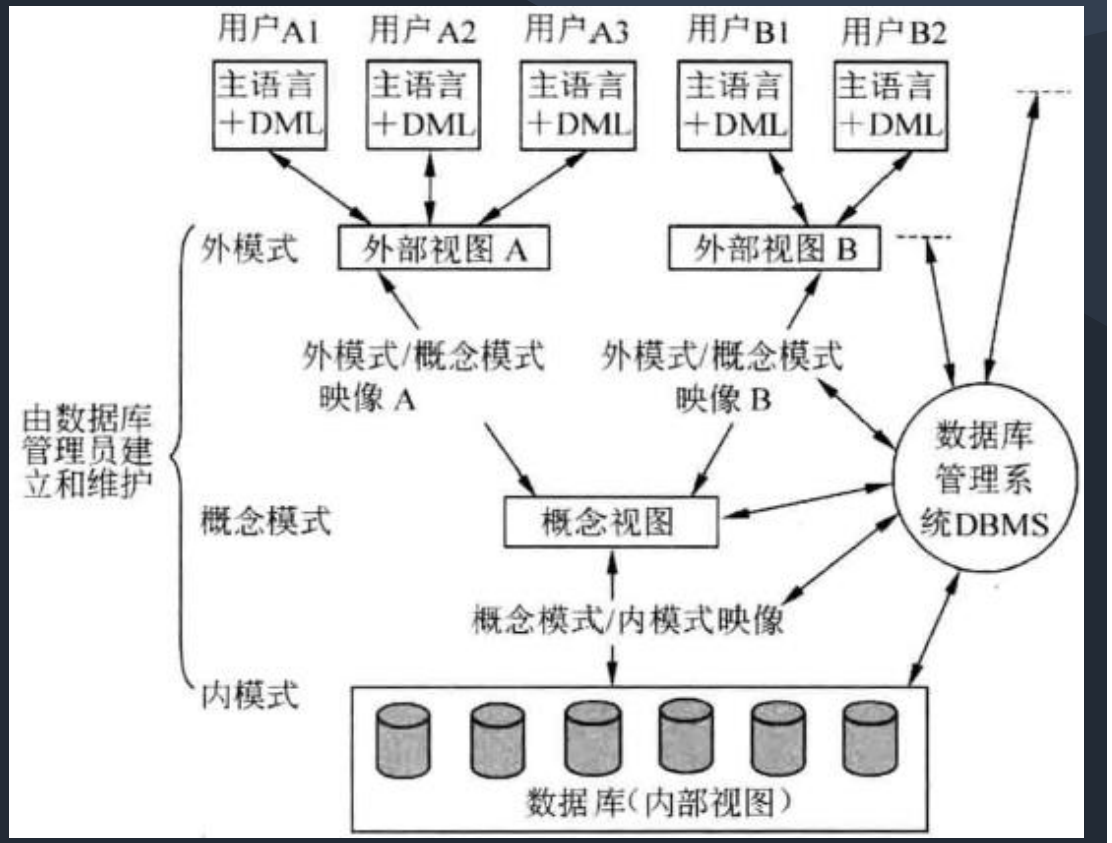
规范化：函数依赖、键与约束、范式、模式分解

事务并发：并发三种问题、三级封锁协议

数据库新技术：数据库安全与备份、反规范化、分布式数据库、缓存数据库、数据库集群，NoSql

案例考点：ER图+关系模式设计、反规范化技术、缓存数据库等新技术。

三级模式-两级映像



◆内模式：管理如何存储物理的数据，对应具体物理存储文件。

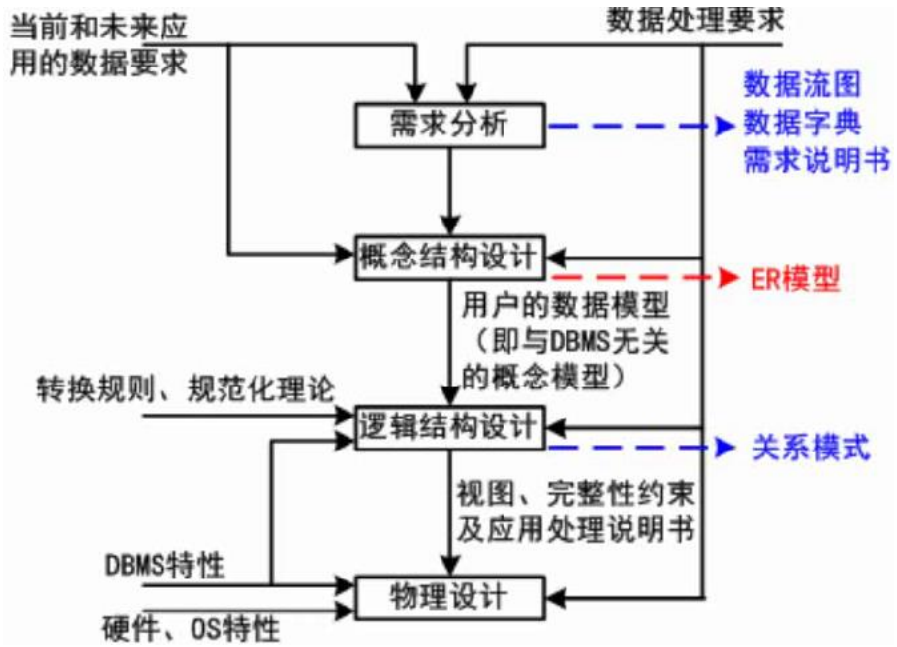
◆模式：又称为概念模式，就是我们通常使用的基本表，根据应用、需求将物理数据划分成一张张表。

◆外模式：对应数据库中的视图这个级别，将表进行一定的处理后再提供给用户使用

◆外模式—模式映像：是表和视图之间的映射，存在于概念级和外部级之间，若表中数据发生了修改，只需要修改此映射，而无需修改应用程序。

◆模式—内模式映像：是表和数据的物理存储之间的映射，存在于概念级和内部级之间，若修改了数据存储方式，只需要修改此映射，而不需要去修改应用程序。

数据库设计



（1）需求分析：即分析数据存储的要求，产出物有数据流图、数据字典、需求说明书。获得用户对系统的三个要求：信息要求、处理要求、系统要求。

（2）概念结构设计：就是设计E-R图，也即实体-联系图。工作步骤包括：选择局部应用、逐一设计分E-R图、E-R图合并。

分E-R 图进行合并时，它们之间存在的冲突主要有以下3 类。

◆属性冲突。同一属性可能会存在于不同的分E-R 图中。

◆命名冲突。相同意义的属性，在不同的分E-R 图上有着不同的命名，或是名称相同的属性在不同的分E-R 图中代表着不同的意义。

◆结构冲突。同一实体在不同的分E-R 图中有不同的属性，同一对象在某一分E-R 图中被抽象为实体而在另一分E-R 图中又被抽象为属性。

（3）逻辑结构设计：将E-R图，转换成关系模式。工作步骤包括：确定数据模型、将E-R 图转换成为指定的数据模型、确定完整性约束和确定用户视图。

（4）物理设计：步骤包括确定数据分布、存储结构和访问方式。

（5）数据库实施阶段。根据逻辑设计和物理设计阶段的结果建立数据库，编制与调试应用程序，组织数据入库，并进行试运行。

（6）数据库运行和维护阶段。数据库应用系统经过试运行即可投入运行，但该阶段需要不断地对系统进行评价、调整与修改。

考试真题

5.采用三级模式结构的数据库系统中，如果对一个表创建聚簇索引，那么改变的是数据库的（　）。
A.外模式　　B.模式　　C.内模式　　D.用户模式

6.假设系统中有正在运行的事务，若要转储全部数据库，则应采用　（　）方式。
A.静态全局转储　B.动态增量转储　C.静态增量转储　D.动态全局转储

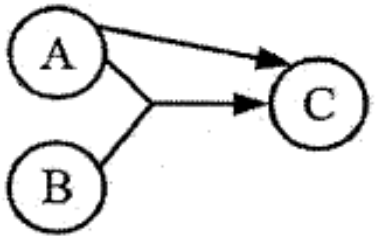
函数依赖

◆给定一个X，能唯一确定一个Y，就称X确定Y，或者说Y依赖于X，例如Y=X*X函数。

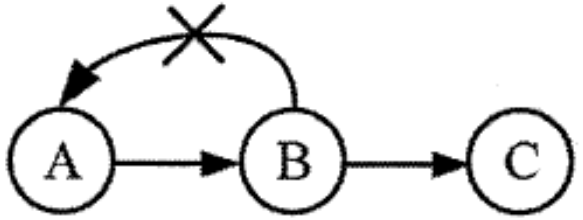
函数依赖又可扩展以下两种规则：

◆部分函数依赖：A可确定C，(A,B)也可确定C，(A,B)中的一部分（即A）可以确定C，称为部分函数依赖。

◆传递函数依赖：当A和B不等价时，A可确定B，B可确定C，则A可确定C，是传递函数依赖；若A和B等价，则不存在传递，直接就可确定C。



部分函数依赖



传递函数依赖

◆超键：能唯一标识此表的属性的组合。

◆候选键：超键中去掉冗余的属性，剩余的属性就是候选键。

◆主键：任选一个候选键，即可作为主键。

◆外键：其他表中的主键。

◆主属性：候选键内的属性为主属性，其他属性为非主属性。

函数依赖

◆函数依赖的公理系统(Armstrong)
设关系模式R<U，F>，U是关系模式R的属性全集，F是关系模式R的一个函数依赖集。对于R<U，F>来说有以下的：
◆自反律：若Y⊆X⊆U，则X→Y为F所逻辑蕴含

◆增广律：若X→Y为F所逻辑蕴含，且Z⊆U，则XZ→YZ为F所逻辑蕴含

◆传递律：若X→Y和Y→Z为F所逻辑蕴含，则X→Z为F所逻辑蕴含

◆合并规则：若X→Y，X→Z，则X→YZ为F所蕴涵

◆伪传递率：若X→Y，WY→Z，则XW→Z为F所蕴涵

◆分解规则：若X→Y， Z⊆Y ， 则X→Z为F所蕴涵

范式

◆第一范式1NF：关系中的每一个分量必须是一个不可分的数据项
◆实例：用一个单一的关系模式学生来描述学校的教务系统：学生(学号,学生姓名,系号,系主任姓名,课程号,成绩)
◆依赖关系（学号->学生姓名，学号->系名，系名->系主任姓名，（学号，课程号）->成绩）

学号	学生姓名	所在系	系主任姓名	课程号	成绩
201102	张明	计算机系	章三	04	70
201103	王红	计算机系	章三	05	60
201103	王红	计算机系	章三	04	80
201103	王红	计算机系	章三	06	87
201104	李青	机械系	王五	09	79
...	...	...	...	...	...

范式

◆第二范式：如果关系R属于1NF，且每一个非主属性完全函数依赖于任何一个候选码，则R属于2NF。通俗地说，2NF就是在1NF的基础上，表中的每一个非主属性不会依赖复合主键中的某一个列。按照定义，上面的学生表就不满足2NF，因为学号不能完全确定课程号和成绩(每个学生可以选多门课)。将学生表分解为：
◆学生(学号,学生姓名,系名,系主任)
◆选课(学号,课程号,成绩)。
每张表均属于2NF。

◆第三范式：在满足1NF的基础上，表中不存在非主属性对码的传递依赖。
◆继续上面的实例，学生关系模式就不属于3NF，因为学生无法直接决定系主任，是由学号->系名，再由系名->系主任，因此存在非主属性对主属性的传递依赖，
◆将学生表进一步分解为：
学生(学号，学生姓名，系编号)
系(系编号，系名，系主任)
选课(学号,课程号,成绩)
每张表都属于3NF。

考试真题

7.给定关系模式R（U，F），其中U为属性集，F是U上的一组函数依赖，那么函数依赖的公理系统（Armstrong 公理系统）中的分解规则是指（　）为F所蕴涵。

A.若X→Y，Y→Z，则X→Y　　B.若Y⊆X⊆U，则X→Y

C.若X→Y，Z⊆Y，则X→Z　　D.若X→Y，Y→Z，则X→YZ

5&6.某企业开发信息管理系统平台进行 E-R 图设计，人力部门定义的是员工实体具有属性：员工号、姓名、性别、出生日期、联系方式和部门,培训部门定义的培训师实体具有属性:培训师号，姓名和职称，其中职称={初级培训师，中级培训师，高级培训师},这种情况属于(5)，在合并E-R图时，解决这一冲突的方法是(6)

A.属性冲突　　B.结构冲突　　C.命名冲突　　D.实体冲突

A.员工实体和培训师实体均保持不变

B.保留员工实体、删除培训师实体

C.员工实体中加入职称属性，删除培训师实体

D将培训师实体所有属性并入员工实体，删除培训师实体

考试真题

试题（6）、（7）

给出关系 $R(U,F)$，$U=\{A,B,C,D,E\}$，$F=\{A\to B, D\to C, BC\to E, AC\to B\}$，求属性闭包的等式成立的是__（6）__。$R$ 的候选关键字为__（7）__。

（6）A. $(A)_F^+=U$　　B. $(B)_F^+=U$　　C. $(AC)_F^+=U$　　D. $(AD)_F^+=U$

（7）A. AD　　B. AB　　C. AC　　D. BC

5. 数据库的安全机制中，通过提供（5）供第三方开发人员调用进行数据更新，从而保证数据库的关系模式不被第三方所获取。

A. 索引　　B. 视图　　C. 存储过程　　D. 触发器

8. 分布式数据库系统除了包含集中式数据库系统的模式结构之外，还增加了几个模式级别，其中(8) 定义分布式数据库中数据的整体逻辑结构，使得数据使用方便，如同没有分布一样。

A. 分片模式　　B. 全局外模式

C. 分布模式　　D. 全局概念模式

关系代数

◆笛卡尔积：S1*S2，产生的结果包括S1和S2的所有属性列，并且S1中每条记录依次和S2中所有记录组合成一条记录，最终属性列为S1+S2属性列，记录数为S1*S2记录数。
◆投影：实际是按条件选择某关系模式中的某列，列也可以用数字表示。
◆选择：实际是按条件选择某关系模式中的某条记录。

关系S1

Sno	Sname	Sdept
No0001	Mary	IS
No0003	Candy	IS
No0004	Jam	IS

关系S2

Sno	Sname	Sdept
No0001	Mary	IS
No0008	Katter	IS
No0021	Tom	IS

S1 × S2（笛卡尔积）

Sno	Sname	Sdept	Sno	Sname	Sdept
No0001	Mary	IS	No0001	Mary	IS
No0001	Mary	IS	No0008	Katter	IS
No0001	Mary	IS	No0021	Tom	IS
No0003	Candy	IS	No0001	Mary	IS
No0003	Candy	IS	No0008	Katter	IS
No0003	Candy	IS	No0021	Tom	IS
No0004	Jam	IS	No0001	Mary	IS
No0004	Jam	IS	No0008	Katter	IS
No0004	Jam	IS	No0021	Tom	IS

(投影)

Sno	Sname
No0001	Mary
No0003	Candy
No0004	Jam

(选择)

Sno	Sname	Sdept
No0003	Candy	IS

关系代数

◆自然连接的结果显示全部的属性列，但是相同属性列只显示一次，显示两个关系模式中属性相同且值相同的记录。

设有关系R、S如下左图所示，自然连接结果如下右图所示：

A	B	C
a	b	c
b	a	d
c	d	e
d	f	g

（a）关系 R

A	C	D
a	c	d
d	f	g
b	d	g

（b）关系 S

A	B	C	D
a	b	c	d
b	a	d	g

$R \bowtie S$

SQL语句关键字

◆数据库查询select…from…where;
◆分组查询group by，分组时要注意select后的列名要适应分组，having为分组查询附加条件：
select sno,avg(score) from student group by sno having(avg(score)>60)
◆更名运算as： select sno as “学 号” from t1
◆字符串匹配like，%匹配多个字符串，_匹配任意一个字符串： select * from t1 where sname like 'a_'
◆数据库插入insert into…values()： insert into t1 values('a',66)
◆数据库删除delete from…where： delete t1 where sno=4
◆数据库修改update…set…where： update t1 set sname='aa' where sno=3
◆排序order by，默认为升序，降序要加关键字DESC： select * from t1 order by sno desc

◆ DISTINCT：过滤重复的选项，只保留一条记录。
◆ UNION ：出现在两个SQL语句之间，将两个SQL语句的查询结果取或运算，即值存在于第一句或第二句都会被选出。
◆ INTERSECT ：对两个SQL语句的查询结果做与运算，即值同时存在于两个语句才被选出。
◆ MIN、AVG、MAX：分组查询时的聚合函数

考试真题

8.给定关系R（A，B，C，D)和S(A，C，E，F)，以下（　）与σR.B>S.E(R⋈S)等价。

A.σ2>7(RxS)　　B.π1,2,3,4,7,8(σ1=5∧2>7∧3=6(RxS))

C.σ2>'7'(RxS)　　D.π1,2,3,4,7,8(σ1=5∧2>'7'∧3=6(RxS))

数据仓库中，数据（8）是指数据一旦进入数据仓库后，将被长期保留并定期加载和刷新，可以进行各种查 询操作，但很少对数据进行修改和删除操作。

（8）A.面向主题　　B.集成性　　C.相对稳定性　　D.反映历史变化

给定关系模式R(U，F)，其中：属性集 U={A1 ,A2,A3,A4,A5,A6}，函数依赖集F={A1→A2, A1→A3, A3→A4, A1A5→A6}。关系模式 R 的候选码为（ ），由于R存在非主属性对码的部分函数依赖，所以 R 属于（ ）。

A． A1A3　　B.A1A4　C. A1 A5　　D.A1A6

A． 1NF　　B.2NF　　C.3NF　　D.BCNF

分布式数据库两阶段提交协议中的两个阶段是指（12） 。

A． 加锁阶段、解锁阶段　　B.获取阶段、运行阶段

C.表决阶段、执行阶段　　D.扩展阶段、收缩阶段

考试真题

给定元组演算表达式R*={t | (∃u)(R(t)∧S(u)∧t[3]<u[2])} ，若关系 R、S如下图所示，则（）。

A	B	C
1	2	3
4	5	6
7	8	9
10	11	12

R

A	B	C
3	7	11
4	5	6
5	9	13
6	10	14

S

A. R*={(3, 7, 11), (5, 9, 13), (6, 10, 14)}
B. R*={(3, 7, 11), (4, 5, 6), (5, 9, 13), (6, 10, 14)}
C. R*={(1, 2, 3), (4, 5, 6), (7, 8, 9)}
D. R*={(1, 2, 3), (4, 5, 6), (7, 8, 9), (10, 11, 12)}

案例真题

2022年11月试题4

阅读以下关于数据库缓存的叙述，在答题纸上回答问题1至问题3。
【说明】
某大型电商平台建立了一个在线 B2B 商店系统，并在全国多地建设了货物仓储中心，通过提前备货的方式来提高货物的运送效率。但是在运营过程中，发现会出现很多跨仓储中心调货从而延误货物运送的情况。为此，该企业计划新建立一个全国仓储货物管理系统，在实现仓储中心常规管理功能之外，通过对在线 B2B 商店系统中订单信息进行及时的分析和挖掘，并通过大数据分析预测各地仓储中心中各类货物的配置数量，从而提高运送效率，降低成本。
当用户通过在线 B2B 商店系统选购货物时，全国仓储货物管理系统会通过该用户所在地址、商品类别以及仓储中心的货物信息和地址，实时为用户订单反馈货物起运地（某仓储中心）并预测送达时间。反馈送达时间的响应时间应小于1秒。
为满足反馈送达时间功能的性能要求，设计团队建议在全国仓储货物管理系统中采用数据缓存集群的方式，将仓储中心基本信息、商品类别以及库存数量放置在内存的缓存中，而仓储中心的其它商品信息则存储在数据库系统。

案例真题

【问题1】(9分)
设计团队在讨论缓存和数据库的数据一致性问题时，李工建议采取数据实时同步更新方案，而张工则建议采用数据异步准实时更新方案。
请用200字以内的文字，简要介绍两种方案的基本思路，说明全国仓储货物管理系统应该采用哪种方案，并说明采取该方案的原因。

【问题2】(9分)
随着业务的发展，仓储中心以及商品的数量日益增加，需要对集群部署多个缓存节点，提高缓存的处理能力。李工建议采用缓存分片方法，把缓存的数据拆分到多个节点分别存储，减轻单个缓存节点的访问压力，达到分流效果。
缓存分片方法常用的有哈希算法和一致性哈希算法，李工建议采用一致性哈希算法来进行分片。请用200字以内的文字简要说明两种算法的基本原理，并说明李工采用一致性哈希算法的原因。

【问题3】(7分)
全国仓储货物管理系统开发完成，在运营一段时间后，系统维护人员发现大量黑客故意发起非法的商品送达时间查询请求，造成了缓存击穿。张工建议尽快采用布隆过滤器方法解决。请用200字以内的文字解释布隆过滤器的工作原理和优缺点。

案例真题

答案：【问题1】
实时方案：当数据库数据更新时，同时更新内存的缓存数据。
异步准实时更新方案：当数据库数据更新时，不立即更新缓存数据，而是将需要更新的操作记录成日志，再逐步排队完成更新。
本题中，建议采用准实时方案，理由是：题目中对性能有严格要求，要求1s内完成。实时同步方案最大的问题在于同步并发时的性能不可控。所以准实时方案才能确保该要求能实现。
【问题2】哈希分片：通过对key进行hash操作，可以把数据分配到不同实例，这类似于取余操作，余数相同的，放在一个实例上。
一致性哈希分片：哈希分片的改进，把存储结点和需要存储的数据都存放在一个hash环上，数据根据hash值在hash环上按正时针方向找到对应的数据存储结点上。
一致性哈希分片的方式在扩充缓存结点时，只需要对少量数据进行存储位置的更新，而哈希分片需要对几乎所有数据进行存储位置更新。
【问题3】布隆过滤器通过一个很长的二进制向量和一系列随机映射函数来记录与识别某个数据是否在一个集合中。如果数据不在集合中，能被识别出来，不需要到数据库中进行查找，所以能将数据库查询返回值为空的查询过滤掉。优点：
1、占用内存小
2、查询效率高
3、不需要存储元素本身，在某些对保密要求比较严格的场合有很大优势
缺点：
1、有一定的误判率，即存在假阳性，不能准确判断元素是否在集合中。
2、不能获取元素本身
3、一般情况下不能从布隆过滤器中删除元素

谢谢！