数据库系统

目录

[数据库系统 1](#_Toc524289781)

[1 数据库管理系统的类型 3](#_Toc524289782)

[2 数据库模式与范式 3](#_Toc524289783)

[2.1 数据库的结构与模式 3](#_Toc524289784)

[2.2 数据模型 3](#_Toc524289785)

[2.3 关系代数 3](#_Toc524289786)

[2.4 数据的规范化 13](#_Toc524289787)

[2.5 反规范化 17](#_Toc524289788)

[3 数据库设计 17](#_Toc524289789)

[3.1 数据库设计的方法 17](#_Toc524289790)

[3.2 数据库设计的基本步骤 18](#_Toc524289791)

[3.3 需求分析 19](#_Toc524289792)

[3.4 概念结构设计 20](#_Toc524289793)

[3.5 逻辑结构设计 20](#_Toc524289794)

[3.6 物理结构设计 20](#_Toc524289795)

[4 事务管理 20](#_Toc524289796)

[4.1 并发控制 23](#_Toc524289797)

[4.2 故障与恢复 24](#_Toc524289798)

[5 备份与恢复 24](#_Toc524289799)

[6 数据集成 24](#_Toc524289800)

[7 分布式数据库系统 25](#_Toc524289801)

[7.1 分布式数据库的概念 25](#_Toc524289802)

[7.2 分布式数据库的架构 26](#_Toc524289803)

[8 数据仓库 26](#_Toc524289804)

[8.1 数据仓库的概念 26](#_Toc524289805)

[8.2 数据仓库的结构 26](#_Toc524289806)

[8.3 数据仓库的实现方法 26](#_Toc524289807)

[9 数据挖掘 26](#_Toc524289808)

[9.1 数据挖掘的概念 26](#_Toc524289809)

[9.2 数据挖掘的功能 26](#_Toc524289810)

[9.3 数据挖掘常用技术 26](#_Toc524289811)

[9.4 数据挖掘的流程 26](#_Toc524289812)

[10 NoSQL 26](#_Toc524289813)

[11 大数据 26](#_Toc524289814)

# 数据库管理系统的类型

# 数据库模式与范式

## 数据库的结构与模式

* 给定关系模式R（A，B，C，D，E）、S（D，E，F，G）和π1,2,4,6（R ?S），经过自然连接和投影运算后的属性列数分别为（9）。

2016年(9)

A.9和4

B.7和4

C.9和7

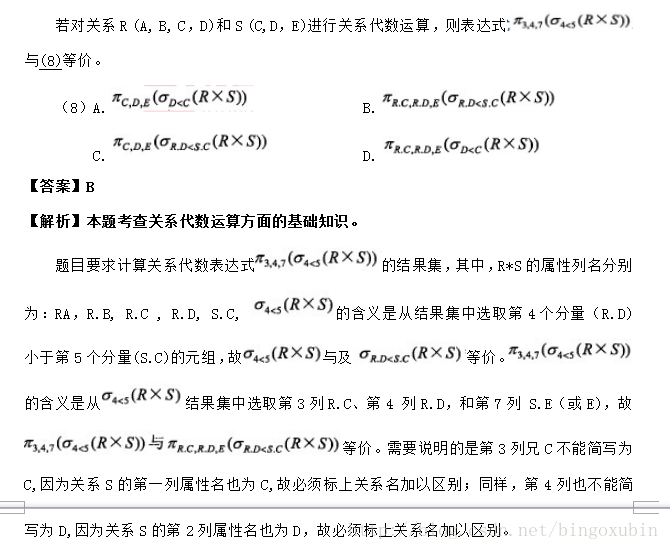
D.7和7

【答案】B

【解析】 R与S进行自然连接后，结果属性集为：A,B,C,D,E,F,G。 投影操作后，结果为：A,B,D,F。

## 数据模型

## 关系代数



* 某数据库中有员工关系E (员工号，姓名，部门，职称，月薪)；产品关系P (产品号，产品名称，型号，尺寸，颜色)；仓库关系W (仓库号，仓库名称，地址，负责人)库存关系1(仓库号，产品号，产品数量)。a.若数据库设计中要求：①仓库关系W中的“负责人”引用员工关系的员工号②库存关系I中的“仓库号，产品号”唯一标识I中的每一个记录③员工关系E中的职称为“工程师”的月薪不能低于3500元 则①②©依次要满足的完整性约束是(6)。b.若需得到每种产品的名称和该产品的总库存量，则对应的查询语句为：SELECT产品名称，SUM (产品数量)FROM P, I WHERE P.产品号=I.产品号(7);

2010年(6)

A.实体完整性、参照完整性、用户定义完整性

B.参照完整性、实体完整性、用户定义完整性

C.用户定义完整性、实体完整性、参照完整性

D.实体完整性、用户定义完整性、参照完整性

2010年(7)

A.ORDERBY产品名称

B.ORDERBY产品数量

C.GROUPBY产品名称

D.GROUPBY产品数量

【答案】B C 【解析】本题考查关系的完整性约束和SQL的基本知识及应用。

(6)关系模型的完整性规则是对关系的某种约束条件。关系模型中可以有三类完整性约束：实体完整性、参照完整性和用户定义的完整性。实体完整性规定基本关系的主属性不能取空值。由于①仓库关系W中的“负责人”引用员工关系的员工号，所以应满足参照完整性约束；②库存关系I中的“仓库号，产品号”唯一标识I中的每一个记录，所以应满足实体完整性约束；③职称为“工程师”的月薪不能低于3500元，是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，所以应满足用户定义完整性约束。因此，试题(6)的正确答案为B。

2010年(7)

SQL查询是数据库中非常重要的内容。该SQL查询要求对查询结果进行分组，即具有相同名称的产品的元组为一组，然后计算每组的库存数量。由此可排除A、B和D，所以试题(7)的正确答案为C。

* 给定学生S (学号，姓名，年龄，入学时间，联系方式）和选课SC (学号，课程号， 成绩）关系，若要查询选修了1号课程的学生学号、姓名和成绩，则该查询与关系代数表达式（8)等价。

2011年(8)

A.

1B422C9F75214A3BA700136373C2E380.png

B.

BC74995477344656A9D437C72F7EB795.png

C.

3B201A6E55CD4C50955685EA2BC7437F.png

D.

CD783782B89D44F099FA0F11B66C18F8.png

【答案】B 【解析】本题考查关系代数运算方面的基础知识。

对于试题（8)，题目要求“查询选修了1号课程的学生学号和姓名”，因此先进行S与SC关系的自然连接，即选取S.学号=SC.学号的元组并去掉右边的重复属性“学号”，生成的新关系为（学号，姓名，年龄，入学时间，联系方式，课程号，成绩），共有7

个属性列。 选项A是错误的，因为自然连接后的第6个属性为课程号，其选取运算的实际含义为“学号=课程号”同时“成绩=1”，与题意不符。

选项B是正确的，因为该关系表达式的含义为：进行S与SC关系的自然连接，选取S.学号=sc.学号的元组并去掉右边的重复属性“学号”，再选取“课程号=1”的元组，最后进行学号、姓名和成绩的投影运算。

同理可以分析选项C和D都是错误的。

* 某企业工程项目管理数据库的部分关系模式如下所示，其中带实下划线的表示主键，虚下划线的表示外键。 供应商(供应商号，名称，地址，电话，账号） 项目(项目号，负责人，开工日期） 零件(零件号，厂规格，单价） 供应(项目号,零件号,供应商号，供应量） 员工（员工号，姓名，性别出生日期，职位，联系方式） 其中供应关系是（5）的联系。若一个工程项目可以有多个员工参加，每个员工可以参加多个项目，则项目和员工之间是（6）联系。对项目和员工关系进行设计时，(7)设计成一个独立的关系模式。

2011年(5)

A.2个实体之间的l:n

B.2个实体之间的n:m

C.3个实体之间的l:n:m

D.3个实体之间的k:n:m

2011年(6)

A.1:1

B.l:n

C.n:m

D.n:l

2011年(7)

A.多对多的联系在向关系模型转换时必须

B.多对多的联系在向关系模型转换时无须

C.只需要将一端的码并入多端，所以无须

D.不仅需要将一端的码并入多端，而且必须

【答案】D C A 【解析】本题考查关系模式和E-R图的概念和性质。

对于试题（6)、（7)，多对多的联系必须转换为一个独立的关系模式。下面分析诊疗科和医师之间的联系：根据E-R模型中一对多联系向关系模式转换规则可知，一个一对多的联系既可以转换为一个独立的关系模式，也可以与多端的关系模式合并。如果与多端的关系模式合并的话，需要将一端的码和联系上的属性合并到多端的关系模式中。由于本题将诊疗科的主键合并到了医师关系模式中，因此诊疗科和医师之间应该是一个一对多的联系。

* 给定员工关系EMP(EmpID，Ename，sex，age，tel，DepID),其属性含义分别为：员工号、姓名、性别、年龄、电话、部门号；部门关系DEP(DepID，Dname，Dtel，DEmpID)，其属性含义分别为：部门号、部门名、电话，负责人号。若要求DepID参照部门关系DEP的主码DepID,则可以在定义EMP时用（7)进行约束。 若要査询开发部的负责人姓名、年龄，则正确的关系代数表达式为（8)。

2013年(7)

A.Primary Key (DepID) On DEP (DepID)

B.Primary Key (DepID) On EMP (DepID)

C.Foreign Key (DepID) References DEP (DepID)

D.Foreign Key (DepID) References EMP (DepID)

2013年(8)

A.π2,4（σ8‘开发部'（EMP×DEP ））

B.π2，4（σ1=9（EMP σ2=’开发部‘ DEP））

C.π2，3（EMP×σ2=’开发部‘（DEP））

D.π2,3（π1,2,4,6（EMP ）σ2='开发部（DEP）'）

【答案】C B 【解析】本题考查关系代数运算方面的基础知识。

(7)员工关系中的DepID是一个外键，为了保证数据的正确性，通过参照完整性加以约束。SQL语言通过使用保留字Foreign

Key定义外键，References指明外码对应于哪个表的主码。参照完整性定义格式如下： Foreign

Key(属性名）References表名（属性名）

可见，若要求DepID参照部门关系DEP的主码DepID，则可以在定义EMP时用“Foreign Key (DepID)

References DEP (DepID)”进行约束。 2013年(8)

试题（8)要求“查询开发部的负责人姓名、年龄”的关系代数表达式，选项B是先进行δ2=’开发部（DEP）运算，即在DEP关系中选择部门名Dname=‘开发部’的元组；然后将EMP关系与其进行EMP.DepID=DEP.DepID的自然连接，并去掉右边的重复属性“DEP.DepID”，自然连接后的属性列为（EmpID,Ename,sex,age,tel,DepID,Dname,Dtel,DEmpID)；再此基础上进行δ1=9运算，即进行员工号EmpID等于部门负责人号DEmpID的选取运算；最后进行属性列2(Ename)和属性列4(age)的投影运算。

* 假设关系模式R(U，F),属性集U={A，B，C}，函数依赖集F={A→B，B→C}。若将其分解为p={Rl(U1，F1),R2(U2,F2)},其中U1={A,B},U2={A,C}。那么，关系模式R、R1、R2分别达到了（5）；分解D(6)。

2013年(5)

A.1NF、2NF、3NF

B.INF、3NF、3NF

C.2NF、2NF、3NF D.2NF、3NF、3NF

2013年(6)

A.有损连接但保持函数依赖

B.既无损连接又保持函数依赖

C.有损连接且不保持函数依赖

D.无损连接但不保持函数依赖

【答案】D B 【解析】本题考查关系数据库方面的基本知识。

(5)由关系模式R的函数依赖集F={A→B，B→C}可以得出A→C，存在传递依赖，但不存在非主属性对码的部分函数依赖，故R为2NF。又由于分解后的关系模式R1的函数依赖集F1={A→B}，关系模式R2的函数依赖集F2={A→C}，因此R1、R2分别达到了3NF。

(6)因为F=F1∪F2,所以分解p保持函数依赖。又由于关系模式R(U，F)的一个分解p={R1(U1,F1),R2(U2,F2)}具有无损连接的充分必要的条件是：U1∩U2→U1-U2∈F+或U1∩U2→U2-U1∈F+。分解p是否无损连接分析如下：

∵AB∩AC=A，AB-AC=B，AC-AB=C ∴A→B∈F+，A→C∈F+ ∴根据无损连接的充分必要的条件可知p为无损连接。

* 若关系模式R和S分别为：R(A,B,C,D)、S(B,C,E,F)，则关系R与S自然联结运算后的属性列有（6）个，与表达方式π1,3,5,6（σ3<6（RS））等价的SQL语句为： SELECT （7） FROM R, S WHERE （8）

2014年(6)

A.4

B.6

C.7

D.8

2014年(7)

A.A,R.C,E,F

B.A,C,S.B,S.E

C.A,C,S.B,S.C

D.R.A,R.C,S.B,S.C

2014年(8)

A.R.B=S.BAND R.C=S.CAND RC.B

B.R.B=S.BAND R.C=S.CAND R.C.F

C.R.B=S.BOR R.C=S.COR R.C.B

D.R.B=S.BOR R.C=S.COR R.C.F

【答案】B A B 【解析】本题考查关系代数运算与SQL查询方面的基础知识。

试题（6)的正确答案为选项B。自然连接RS是指R与S关系中相同属性列名的等值连接运算后，再去掉右边重复的属性列名S.B、S.C，所以经RS运算后的属性列名为：R.A、R.B、R.C、R.D、S.E和S.F，共有6个属性列。

试题（7)的正确答案为选项A。π1,3,5,6(σ3<6(RS))的含义是从RS结果集中选取R.C

* 设关系模式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）中的合并规则是指（5）为F所蕴涵。

2014年(5)

A.若A→B，B→C，则A→C

B.若，则X→Y

C.若A→B，A→C，则A→BC

D.若A→B，CB，则A→C

【答案】C 【解析】本题考査函数依赖推理规则。

函数依赖的公理系统（即Armstrong公理系统）为：设关系模式R(U，F),其中U为属性集，F是U上的一组函数依赖，那么有如下推理规则：

A1自反律：若，则X→Y为F所蕴涵。 A2增广律：若X→Y为F所蕴涵，且，则XZ→YZ为F所蕴涵。

A3传递律：若X→Y，Y→Z为F所蕴涵，则X→Z为F所蕴涵。 根据上述三条推理规则又可推出下述三条推理规则：

A4合并规则：若X→Y，X→Z，则X→YZ为F所蕴涵。 A5伪传递率：若X→Y，WY→Z，则XW4Z为F所蕴涵。

A6分解规则：若X→Y，，则X→Z为F所蕴涵。

选项A符合规则为A3，即传递规则；选项B符合规则为A1，即为自反规则；选项C符合规则为A4，即为合并规则；选项D符合规则为A6，即为分解规则。

* 若关系R、S如下图所示，则关系R与S进行自然连接运算后的元组个数和属性列数分别为（7）；关系代数表达式π1,4(σ3=6(R×S))与关系代数表达式（8）等价。

2015年(7)

A.6和6

B.4和6

C.3和6

D.3和4

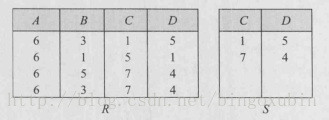
2015年(8)

A.πA,D(σC=D(R×S))

B.πA,R,D(σS.C=R.D(R×S))

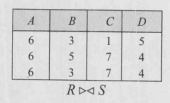
C.πA,R,D(σR.C=S.D(R×S))

D.πA,R,D(σS.C=S.D(R×S))



【答案】D C 【解析】本题考查关系运算方面的基础知识。

(7)根据自然连接要求，两个关系中进行比较的分量必须是相同的属性组，并且在结果中将重复属性列去掉，故R▷◁S后的属性列数为4。同时，自然连接是一种特殊的等值连接，即及关系中的C、D属性与S关系中的C、D属性进行等值连接，然后去掉复属性列，其结果为：



可见R▷◁后的元组个数为3。因此试题(7)的正确答案是D。

(8)关系代数表达式π1.4(σ3=6(R×S))中，R×S的6个属性列为：R.A、R.B、R.C、R.D、S.C和S.D，σ3=6(R×S)表示R与S关系进行笛卡儿积运算后，选取第三个属性R.C等于第六个属性S.D的元组；π1.4(σ3=6(R×S))表示从σ3=6(R×S)的结果中投影第一个和第四个属性列，即投影R.A和R.D属性列，因此试题(8)的正确答案是C。

* 给定关系R（A1，A2，A3，A4）上的函数依赖集F={A1→A2A5，A2→A3A4，A3→A2}，R的候选关键字为（10）。函数依赖（11）∈F+。

2016年(10)

A.A1

B.A1A2

C.A1A3

D.A1A2A3

2016年(11)

A.A5→A1A2

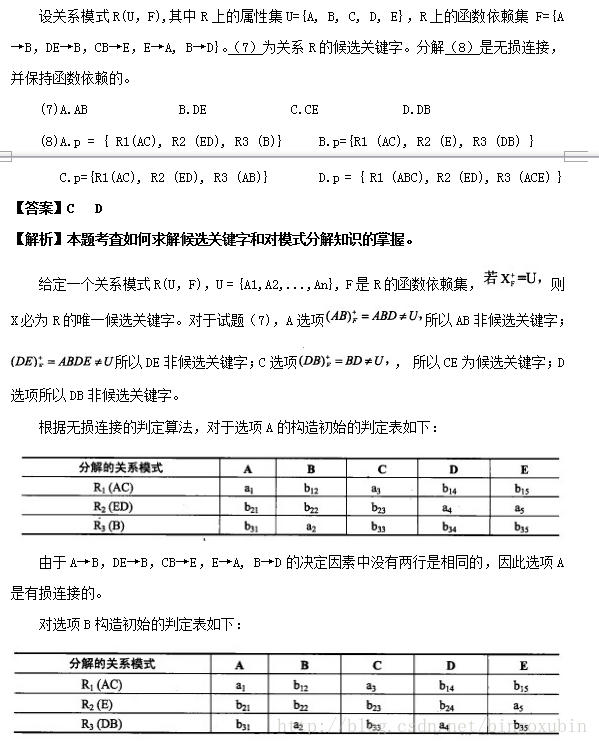
B.A4→A1A2

C.A3→A2A4

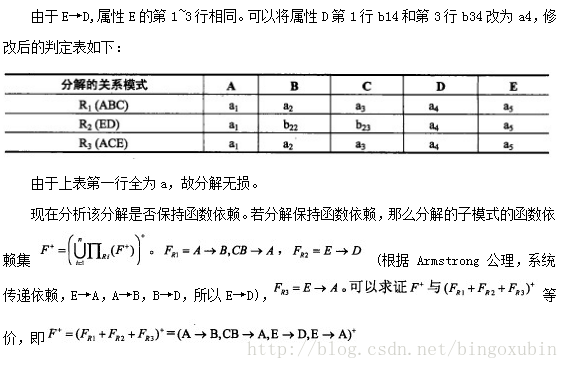
D.A2→A1A5

【答案】A C 【解析】 (10)通过绘制函数依赖图可以了解到，从A1出发，可以遍历全图，所以候选关键字为A1。(11)F+是代表，F函数依赖集的闭包，通俗一点，就是从F函数依赖集能推导出来的依赖关系。A3→A2A4是正确的。首先F中有直接的依赖关系：A3→A2，此外，又由A3→A2与A2→A4得出A3→A4。A3→A2与A3→A4合并为A3→A2A4。

## 数据的规范化







* 某商场商品数据库的商品关系模式P (商品代码，商品名称，供应商，联系方式，库存量)，函数依赖集F={商品代码一商品名称，（商品代码，供应商）一库存量，供应商一联系方式。商品关系模式P达到（7）；该关系模式分解成（8）后，具有无损连接的特性，并能够保持函数依赖。

2012年(7)

A.INF

B.2NF

C.3NF

D.BCNF

2012年(8)

A.P1 (商品代码，联系方式），P2 (商品名称，供应商，库存量）

B.P1 (商品名称，联系方式)，P2 (商品代码，供应商，库存量）

C.P1 (商品代码，商品名称，联系方式），P2 (供应商，库存量）

D.P1 (商品代码，商品名称)，P2 (商品代码，供应商，库存量)， P3 (供应商，联系方式)

【答案】A D 【解析】本题考查的是应试者关系数据库方面的基础知识。

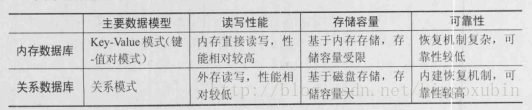
根据题意，零件P关系中的（商品代码，供应商）可决定的零件P关系的所有属性，所以零件P关系的主键为（商品代码，供应商）;又因为，根据题意（商品代码，供应商）一商品名称，而商品代码一商品名称，供应商一联系方式，可以得出商品名称和联系方式都部分依赖于码，所以，该关系模式属于1NF。关系模式P属于INF,1NF存在冗余度大、修改操作的不一致性、插入异常和删除异常四个问题。所以需要对模式分解，其中选项A、选项B和选项C的分解是有损且不保持函数依赖。例如，选项A中的分解P1的函数依赖集F1=0,分解P2的函数依赖集F2=0,丢失了F中的函数依赖，即不保持函数依赖。

## 反规范化

# 数据库设计

## 数据库设计的方法

知识点

1. 文件系统具有以下特点：•针对特定应用系统设计，难度较小； •数据冗余较大，可能在多个文件中复制相同的数据属性； •以应用系统为中心组织、管理数据；•符合特定应用系统要求的文件数据很难在不同的应用系统之间共享。
2. 关系型数据库具有以下特点。 •数据结构需要符合关系模式，设计难度较大；•遵守数据库范式，数据冗余较少； •以数据库为中心组织、管理数据； •数据独立于应用系统，很容易在不同的应用系统之间共享数据。
3. 内存数据库型数据库是将数据放在内存中直接操作的数据库，使用内存型数据库将极大地提高应用的性能，同时通过数据缓存、快速算法、并行操作等的改进，使内存型数据库相对于传统的关系型数据库数据处理性能提高10倍以上，同时内存型数据库的应用受到内存大小，数据恢复要求的限制。
4. 
5. SQL语句设计时，影响查询效率的设计原则是： •查询时尽量不要返回不需要的行、列；•需要进行多表连接査询时，尽量使用连接查询，避免使用子查询结构； •尽量避免采用NOTIN、NOTEXIST、LIKE等使用全表查询的操作；•尽量避免使用DISTINCT关键字。

试题

* 试题四【说明】某软件公司拟开发一套贸易综合管理系统，包括客户关系管理子系统和商品信息管理子系统两部分。客户关系管理子系统主要管理客户信息，并根据贸易业务需要频繁向客户发送相关的电子邮件、短信等提醒信息。商品信息管理子系统主要为客户提供商品信息在线查询功能，包括商品基本信息、实时库存与价格等。在对系统进行数据架构设计时，公司项目组的架构师王工主张采用文件系统进行数据管理，原因是目前公司客户和商品数量不大，且系统功能较为简单，采用文件系统进行数据管理简单直观，开发周期短。架构师李工则建议采用关系数据库进行数据管理，原因在于公司目前正处在高速扩张期，虽然目前的客户和商品数量不大，但随着公司快速发展，需要管理的数据必然飞速膨胀，采用关系数据库作为数据存储层，系统的扩展性更强，并能够对未来可能增加的复杂业务提供有效支持。经过讨论，项目组初步采纳了李工的意见，决定采用关系数据库存储客户数据，并针对业务特征对系统性能进行优化。

2015【问题1】

请从设计难度、数据冗余程度、数据架构、应用扩展性等4个方面对关系型数据库管理系统和文件系统两种数据存储方式进行比较，填写表4-1中(1)～(4)。

表4-1 关系型数据库管理系统和文件系统存储方式比较



(1) 数据结构需要符合关系模式，设计难度较大 (2) 可能在多个文件中复制相同的数据属性，数据冗余较大 (3)以应用系统为中心组织、管理数据 (4) 数据独立于应用系统，很容易在不同的应用系统之间共享数据

本题考查文件系统、关系型数据库、内存型数据库的主要特点。

此类题目要求考生认真阅读题目对现实问题的描述，依据系统的核心业务需求的特点(数据模型、读写性能、存储容量、可靠性)，正确选取不同的数据存储架构。并能够根据业务的具体情况分析影响数据库查询的主要原因。

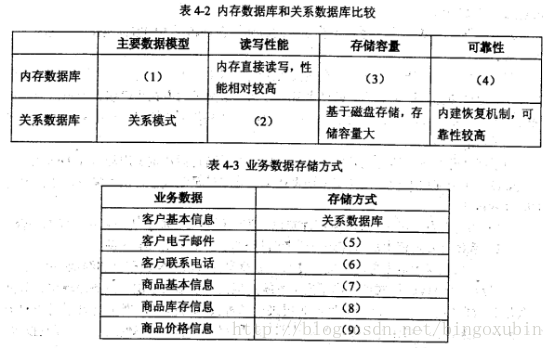
影响数据存储方式的主要考虑因素包括：设计难度、数据冗余程度、数据架构、应用扩展性；

文件系统具有以下特点：•针对特定应用系统设计，难度较小； •数据冗余较大，可能在多个文件中复制相同的数据属性； •以应用系统为中心组织、管理数据；•符合特定应用系统要求的文件数据很难在不同的应用系统之间共享。 关系型数据库具有以下特点。 •数据结构需要符合关系模式，设计难度较大；

•遵守数据库范式，数据冗余较少； •以数据库为中心组织、管理数据； •数据独立于应用系统，很容易在不同的应用系统之间共享数据。

2015【问题2】

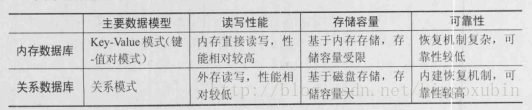
对系统的核心业务需求进行认真分析后，公司的资深架构师张工提出一种内存数据库和关系数据库的混合存储架构，其核心思想是将需要频繁读写的数据存入内存数据库，而将相对固定不变的数据存入关系数据库。请首先分析比较内存数据库和关系数据库在数据模型、读写性能、存储容量、可靠性等方面的差异，填写表4-2中(1)～(4)的空白，并根据张工的思路指定各种业务数据的存储方式，填写表4-3中(5)～(9)中的空白。



(1) Key-Value模式(键-值对模式) (2) 外存读写，性能相对较低 (3) 基于内存存储， 存储容量受限 (4) 恢复机制复杂，可靠性较低 (5) 内存数据库 (6) 内存数据库 (7) 关系数据库 (8) 内存数据库 (9) 内存数据库

内存数据库型数据库是将数据放在内存中直接操作的数据库，使用内存型数据库将极大地提高应用的性能，同时通过数据缓存、快速算法、并行操作等的改进，使内存型数据库相对于传统的关系型数据库数据处理性能提高10倍以上，同时内存型数据库的应用受到内存大小，数据恢复要求的限制。

关系型数据库和内存数据的主要特点如下：



根据贸易综合管理系统的需求描述，需要频繁向客户发送相关的电子邮件、短信等提醒信息，并实时更新商品库存信息和商品价格信息，因此混合存储架构中，将客户电子邮件、客户联系电话、商品库存信息、商品价格信息等数据存入内存数据库：客户基本信息，商品基本信息相对稳定、访问频率较低，存入关系型数据库。

2015【问题3】

系统开发完成进行压力测试时，发现在较大数据量的情况下，部分业务查询响应时间过长，经过分析发现其主要原因是部分SQL查询语句效率低下。请判断表4-4中的SQL语句设计策略哪些可能会提升查询效率，哪些可能会降低查询效率，在(1)~(4)中填入“提升”或“降低”。

表4-4 SQL设计策略对性能的影响



(1) 提升 (2) 降低 (3) 降低 (4) 提升

SQL语句设计时，影响查询效率的设计原则是： •查询时尽量不要返回不需要的行、列；•需要进行多表连接査询时，尽量使用连接查询，避免使用子查询结构； •尽量避免采用NOTIN、NOTEXIST、LIKE等使用全表查询的操作；•尽量避免使用DISTINCT关键字。

* 设有职务工资关系P (职务，最低工资，最高工资），员工关系EMP (员工号，职务，工资），要求任何一名员工，其工资值必须在其职务对应的工资范围之内，实现该需求的方法是（6）.

2009年(6)

A.建立“EMP.职务”向“P.职务”的参照完整性约束

B.建立“P.职务”向“EMP.职务”的参照完整性约束

C.建立EMP上的触发器程序审定该需求

D.建立P上的触发器程序审定该需求

【答案】C 【解析】本题考查对数据完整性约束方面基础知识的掌握。

完整性约束分为实体完整性约束、参照完整性约束和用户自定义完整性约束三类。其中实体完整性约束可以通过 PrimaryKey指定，参照完整性约束通过Foreign Key指定，某些简单的约束可以通过Check、Assertion等实现。针对复杂的约束，系统提供了触发器机制，通过用户编程实现。本题中的约束条件只能通过编写职工表上的触发器，在对工资进行修改或插入新记录时触发，将新工资值与工资范围表中职工职务对应的工资范围比对，只有在范围内才提交，否则回滚。

## 数据库设计的基本步骤

* 在数据库设计的需求分析阶段应完成包括（5）在内的文档。

2009年(5)

A.E-R图

B.关系模式

C.数据字典和数据流图

D.任务书和设计方案

【答案】C 【解析】本题考查数据库设计方面的相关知识。

需求分析阶段的任务是对现实世界要处理的对象（组织、部门和企业等）进行详细调查，在了解现行系统的概况，确定新系统功能的过程中收集支持系统目标的基础数据及处理方法。需求分析是在用户调查的基础上，通过分析，逐步明确用户对系统的需求，包括数据需求和围绕这些数据的业务处理需求，以及对数据安全性和完整性方面的要求。在需求分析阶段应完成的文档是数据字典和数据流图。

* 在数据库设计的(5)阶段进行关系规范化。

2010年(5)

A.需求分析

B.概念设计

C.逻辑设计

D.物理设计

【答案】C 【解析】本题考查数据库设计的基础知识。

数据库设计分为用户需求分析、概念设计、逻辑设计和物理设计4个主要阶段。将抽象的概念模型转化为与选用的DBMS产品所支持的数据模型相符合的逻辑模型，它是物理设计的基础。包括模式初始设计、子模式设计、应用程序设计、模式评价以及模式求精。

逻辑设计阶段的任务是将概念模型设计阶段得到的基本E-R图转换为与选用的DBMS产品所支持的数据模型相符合的逻辑结构。如采用基于E-R模型的数据库设计方法，该阶段就是将所设计的E-R模型转换为某个DBMS所支持的数据模型；如采用用户视图法，则应进行模式的规范化，列出所有的关键字以及用数据结构图描述表集合中的约束与联系，汇总各用户视图的设计结果，将所有的用户视图合成一个复杂的数据库系统。

## 需求分析

* 在数椐库设计的需求分析阶段应当形成（5)，这些文档可以作为（6)阶段的设计依据。

2012年(5)

A.程序文档、数据字典和数据流图

B.需求说明文档、 程序文档和数据流图

C.需求说明文档、 数据字典和数据流图

D.需求说明文档、 数据字典和程序文档

2012年(6)

A.逻辑结构设计

B.概念结构设计

C.物理结构设计

D.数据库运行和维护

【答案】C B 【解析】本题考查数据库系统基本概念方面的基础知识。

数据库设计主要分为用户需求分析、概念结构、逻辑结构和物理结构设计四个阶段。其中，在用户需求分析阶段中，数据库设计人员采用一定的辅助工具对应用对象的功能、性能、限制等要求所进行的科学分析，并形成需求说明文档、数据字典和数据流程图。用户需求分析阶段形成的相关文档用以作为概念结构设计的设计依据。

## 概念结构设计

## 逻辑结构设计

* 假设某证券公司的股票交易系统中有正在运行的事务，此时，若要转储该交易系统数据库中的全部数据，则应采用（12）方式。

2016年(12)

A.静态全局转储

B.动态全局转储

C.静态增量转储

D.动态增量转储

【答案】B 【解析】 从题目中“系统中有正在运行的事务”可知应采用动态方式，从题目中“全部数据”可知应是全局转储，所以应采用：动态全局转储。

## 物理结构设计

# 事务管理

* 在数据库系统中，“事务”是访问数据库并可能更新各种数据项的一个程序执行单元。为了保证数据完整性，要求数据库系统维护事务的原子性、一致性、隔离性和持久性。针对事务的这4种特性，考虑以下的架构设计场景：假设在某一个时刻只有一个活动的事务，为了保证事务的原子性，对于要执行写操作的数据项，数据库系统在磁盘上维护数据库的一个副本，所有的写操作都在数据库副本上执行，而保持原始数据库不变，如果在任一时刻操作不得不中止，系统仅需要删除副本，原数据库没有受到任何影响。这种设计策略称为（44)。事务的一致性要求在没有其他事务并发执行的情况下，事务的执行应该保证数据库的一致性。数据库系统通常采用（45)机制保证单个事务的一致性。事务的隔离性保证操作并发执行后的系统状态与这些操作以某种次序顺序执行（即可串行化执行）后的状态是等价的。两阶段锁协议是实现隔离性的常见方案，该协议(46)。持久性保证一旦事务完成，该事务对数据库所做的所有更新都是永久的，如果事务完成后系统出现故障，则需要通过恢复机制保证事务的持久性。假设在日志中记录所有对数据库的修改操作，将一个事务的所有写操作延迟到事务提交后才执行，则在日志中(47)，当系统发生故障时，如果某个事务已经开始，但没有提交，则该事务应该（48)。

2012年(44)

A.主动冗余

B.影子拷贝

C.热备份

D.多版本编程

2012年(45)

A.逻辑正确性检查

B.物理正确性检查

C.完整性约束检查

D.唯一性检查

2012年(46)

A.能够保证事务的可串行化执行，可能发生死锁

B.不能保证事务的可串行化执行，不会发生死锁

C.不能保证事务的可串行化执行，可能发生死锁

D.能够保证事务的可串行化执行，不会发生死锁

2012年(47)

A.无需记录“事务开始执行”这一事件

B.无需记录“事务已经提交”这一事件

C.无需记录数据项被事务修改后的新值

D.无需记录数据项被事务修改前的原始值

2012年(48)

A.重做

B.撤销

C.什么都不做

D.抛出异常后退出

【答案】B C A D C 【解析】本题主要考查数据库系统架构设计知识。

在数据库系统中，“事务”是访问并可能更新各种数据项的一个程序执行单元。为了保证数据完整性，要求数据库系统维护事务的原子性、一致性、隔离性和持久性。

题干中第1个架构设计场景描述了数据库设计中为了实现原子性和持久性的最为简单的策略：“影子拷贝”。该策略假设在某一个时刻只有一个活动的事务，首先对数据库做副本（称为影子副本)，并在磁盘上维护一个dp\_pointer指针，指向数据库的当前副本。对于要执行写操作的数据项，数据库系统在磁盘上维护数据库的一个副本，所有的写操作都在数据库副本上执行，而保持原始数据库不变，如果在任一时刻操作不得不中止，

系统仅需要删除新副本，原数据库副本没有受到任何影响。

题干中的第2个架构设计场景主要考查考生对事务一致性实现机制的理解。事务的一致性要求在没有其他事务并发执行的情况下，事务的执行应该保证数据库的一致性。数据库系统通常采用完整性约束检查机制保证单个事务的一致性。

题干中的第3个架构设计场景主要考查数据库的锁协议。两阶段锁协议是实现事务隔离性的常见方案，该协议通过定义锁的增长和收缩两个阶段约束事务的加锁和解锁过程，能够保证事务的串行化执行，但由于事务不能一次得到所有需要的锁，因此该协议会可能会导致死锁。

. 题干中的第4个架构设计场景主要考查数据库的恢复机制，主要描述了基于日志的延迟修改技术（deferred-modification

technique)的设计与恢复过程。该技术通过在日志中记录所有对数据库的修改操作，将一个事务的所有写操作延迟到事务提交后才执行，

曰志中需要记录“事务开始”和“事务提交”时间，还需要记录数据项被事务修改后的新值，无需记录数据项被事务修改前的原始值。当系统发生故障时，如果某个事务已经开始，但没有提交，则该事务对数据项的修改尚未体现在数据库中，因此无需做任何恢复动作。

## 并发控制

* 若系统中存在n个等待事务Ti（i=0,1，2，…，n-1），其中：T0正等待被T1锁住的数据项A1，T1正等待被T2锁住的数据项A2，…，Ti正等待被Ti+1锁住的数据项Ai+1，…，Tn-1正等待被T0锁住的数据项A0，则系统处于（5）状态。

2015年(5)

A.封锁

B.死锁

C.循环

D.并发处理

【答案】B 【解析】本题考查关系数据库事务处理方面的基础知识。

与操作系统一样，封锁的方法可能引起活锁和死锁。例如事务T1封锁了数据R，事务了T2请求封锁R，于是T2等待。T3也请求封锁R，当T1释放了R上的封锁之后系统首先批准了T3的请求，T2仍然等待。然后T4又请求封锁R，当：T3释放R上的封锁后系统又批准了T4的请求，……T2有可能长期等待，这就是活锁。避免活锁的简单方法是采用先来先服务的策略。即让封锁子系统按请求封锁的先后次序对事务排队。数据R上的锁一旦释放就批准申请队列中的第一个事务获得锁。

又如事务T1封锁了数据R1，T2封锁了数据R2，T3封锁了数据R3。然后T1又请求封锁R2，T2请求封锁R3，T3请求封锁R1。于是出现T1等待T2释放R2上的封锁，T2等待T3释放R3上的封锁，T3等待T1释放R1上的封锁。这就使得三个事务永远不能结束。即多个事务都请求封锁别的事务已封锁的数据，导致无法运行下去的现象称为死锁。

## 故障与恢复

# 备份与恢复

* 数据备份是信息系统运行管理时保护数据的重要措施。(17)可针对上次任何一 种备份进行，将上次备份后所有发生变化的数据进行备份，并将备份后的数据进行标记。

2011年(17)

A.增量备份

B.差异备份

C.完全备份

D.按需备份

【答案】A 【解析】本题主要考查对各种数据备份机制的理解。

根据题干描述，可以看出增量备份可针对上次任何一种备份进行，将上次备份后所有发生变化的数据进行备份，并将备份后的数据进行标记。

# 数据集成

* 某企业欲对内部的数据库进行数据集成。如果集成系统的业务逻辑较为简单，仅使用数据库中的单表数据即可实现业务功能，这时采用（18)方式进行数据交换与处理较为合适；如果集成系统的业务逻辑较为复杂，并需要通过数据库中不同表的连接操作获取数据才能实现业务功能，这时采用（19)方式进行数据交换与处理较为合适。

2011年(18)

A.数据网关

B.主动记录

C.包装器

D.数据映射

2011年(19)

A.数据网关

B.主动记录

C.包装器

D.数据映射

【答案】B D 【解析】本题主要考查数据集成的相关知识。

关键要判断在进行集成时，需要数据库中的单表还是多表进行数据整合。如果是单表即可完成整合，则可以将该表包装为记录，采用主动记录的方式进行集成；如果需要多张表进行数据整合，则需要采用数据映射的方式完成数据集成与处理。

# 分布式数据库系统

## 分布式数据库的概念

* 在分布式数据库中包括分片透明、复制透明、位置透明和逻辑透明等基本概念，其中：（6）是指局部数据模型透明，即用户或应用程序无需知道局部场地使用的是哪种数据模型。

2015年(6)

A.分片透明

B.复制透明

C.位置透明

D.逻辑透明

【答案】D 【解析】本题考查对分布式数据库基本概念的理解。

分片透明是指用户或应用程序不需要知道逻辑上访问的表具体是怎么分块存储的。复制透明是指采用复制技术的分布方法，用户不需要知道数据是复制到哪些节点，如何复制的位置透明是指用户无须知道数据存放的物理位置，逻辑透明，即局部数据模型透明，是指用户或应用程序无须知道局部场地使用的是哪种数据模型。

## 分布式数据库的架构

# 数据仓库

## 数据仓库的概念

## 数据仓库的结构

## 数据仓库的实现方法

# 数据挖掘

## 数据挖掘的概念

## 数据挖掘的功能

## 数据挖掘常用技术

## 数据挖掘的流程

# NoSQL

# 大数据