

数据库考点汇总

基本概念：三级模式-两级映像、数据库设计

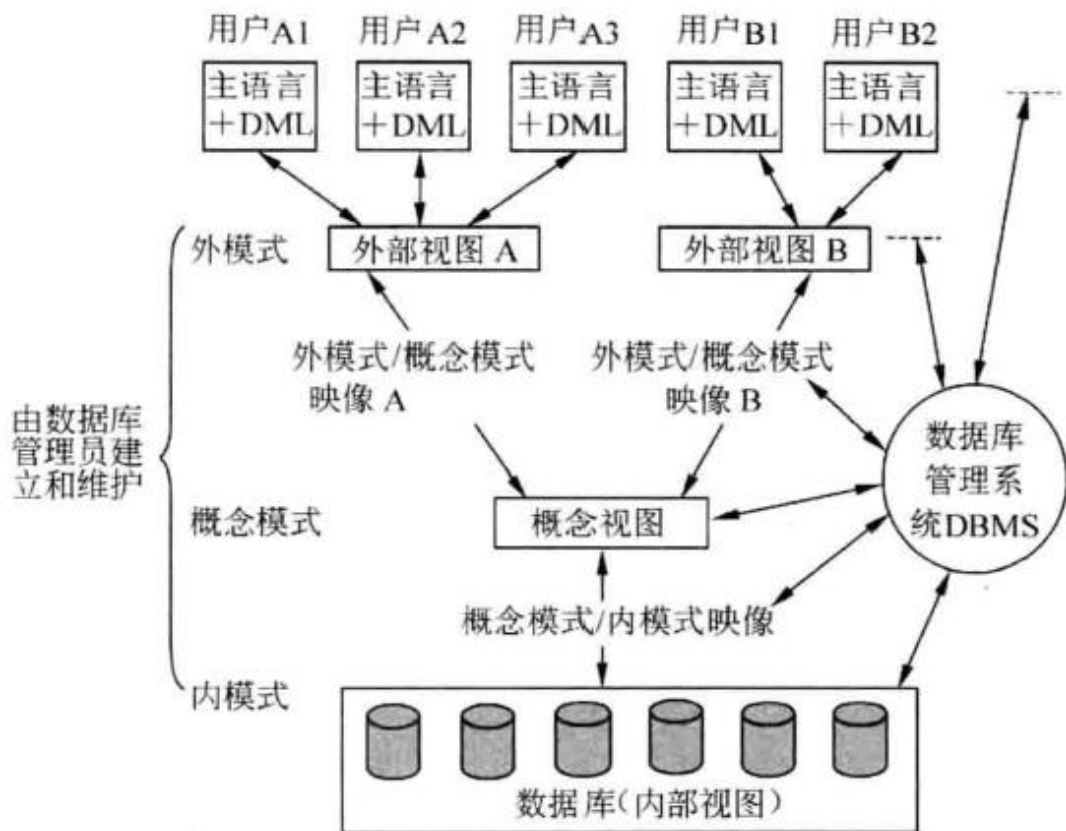
数据库模型：E-R模型、关系模型、关系代数（结合SQL语言）

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

三级模式-两级映射



◆ **内模式**：管理如何存储物理的数据，对应具体物理存储文件。

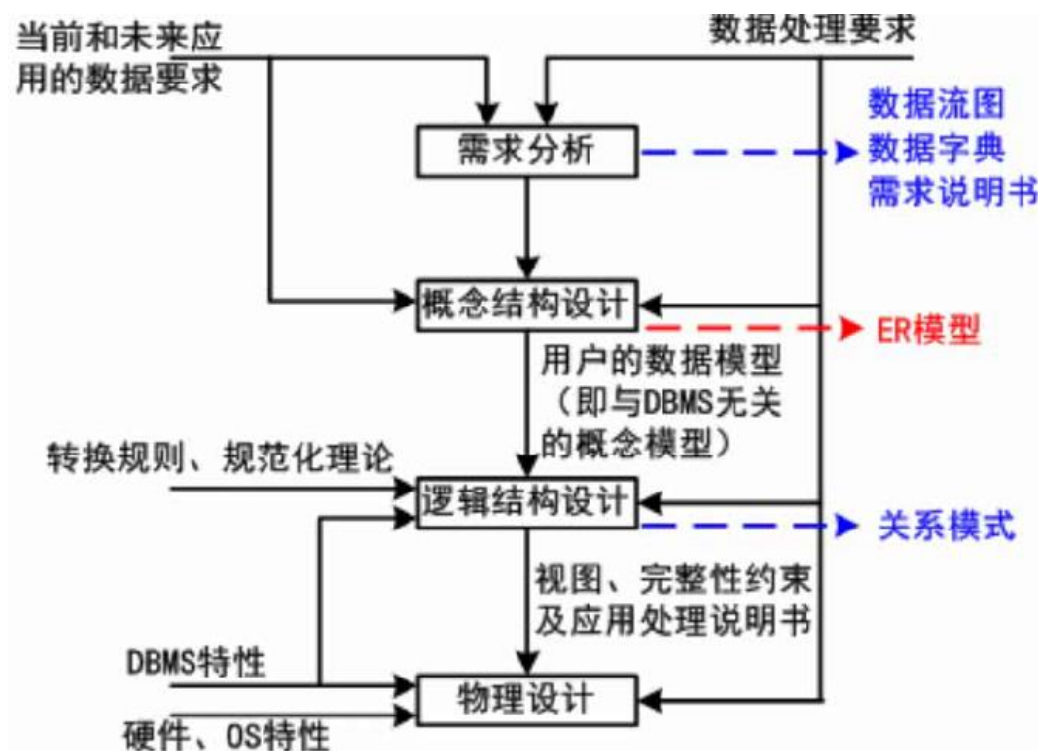
◆ **模式**：又称为概念模式，就是我们通常使用的基本表，根据应用、需求将物理数据划分成一张张表。

◆ **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。

考试试题

●在数据库设计的需求分析、概念结构设计、逻辑结构设计和物理结构设计的四个阶段中，基本E-R图是（40）。

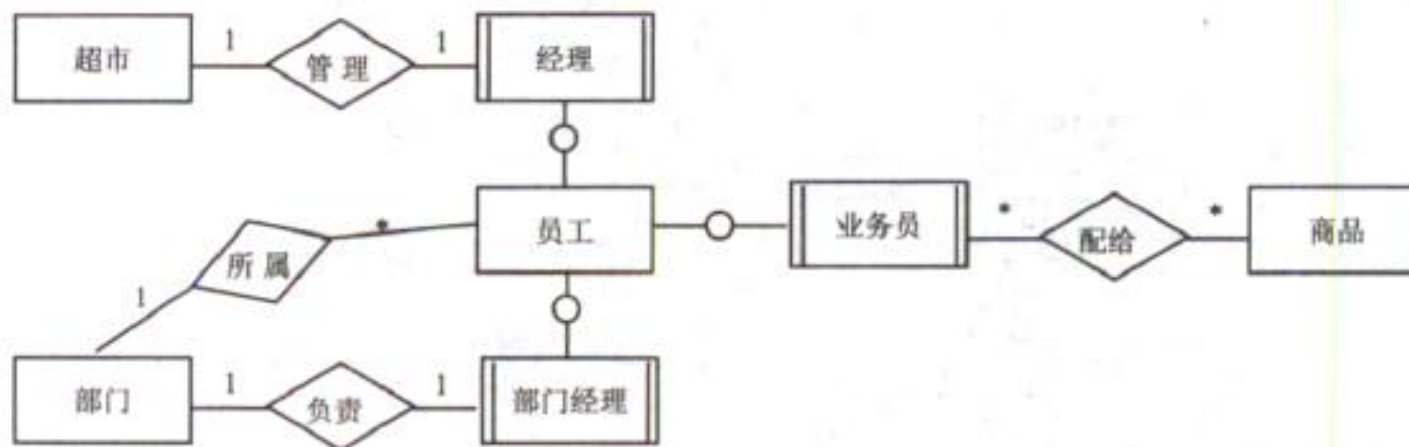
- A. 需求分析阶段形成的文档，并作为概念结构设计阶段的设计依据
- B. 逻辑结构设计阶段形成的文档，并作为概念结构设计阶段的设计依据
- C. 概念结构设计阶段形成的文档，并作为逻辑结构设计阶段的设计依据
- D. 概念结构设计阶段形成的文档，并作为物理设计阶段的设计依据

试题(45)数据的物理独立性和逻辑独立性分别是通过修改(45)来完成的。

- (45) A. 外模式与内模式之间的映像、模式与内模式之间的映像
- B. 外模式与内模式之间的映像、外模式与模式之间的映像
 - C. 外模式与模式之间的映像、模式与内模式之间的映像
 - D. 模式与内模式之间的映像、外模式与模式之间的映像

E-R模型

- ◆实体：**客观存在并可相互区别的事物**。可以是具体的人、事、物或抽象概念。
- ◆弱实体和强实体：弱实体依赖于强实体的存在而存在。
- ◆属性：**实体所具有的特性**。
- ◆属性分类：简单属性和复合属性；单值属性和多值属性；NULL属性；派生属性。
- ◆域：属性的取值范围称为该属性的域。
- ◆码（key）：唯一标识实体的属性集。
- ◆联系：在E-R图中反映为**实体内部的联系和实体之间的联系**。
- ◆联系类型：**一对一1:1、一对多1:N、多对多M:N**。



关系模型

◆关系模型中数据的逻辑结构是一张**二维表**，由行列组成：

学号	姓名	年龄	性别	系名	年级
2005004	王小明	19	女	社会学	2005
2005006	黄大鹏	20	男	商品学	2005
2005008	张文斌	18	女	法律	2005
...

◆E-R模型转换为关系模型：**每个实体都对应一个关系模式**；联系分为三种：

1:1联系中，联系可以**放到任意的两端实体中，作为一个属性**（要保证1:1的两端关联），也可以转换为一个单独的关系模式；

1:N的联系中，联系可以单独作为一个关系模式，也可以在**N端中加入1端实体的主键**；

M:N的联系中，联系**必须作为一个单独的关系模式**，其主键是M和N端的联合主键。

关系代数

- ◆笛卡尔积： $S1 \times S2$ ，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 $S1 \times S2$ 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

$S1 \times S2$ (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：

<i>A</i>	<i>B</i>	<i>C</i>
a	b	c
b	a	d
c	d	e
d	f	g

(a) 关系 *R*

<i>A</i>	<i>C</i>	<i>D</i>
a	c	d
d	f	g
b	d	g

(b) 关系 *S*

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>
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

◆授权Grant, with grant option允许授权他人

◆ 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

考试试题

42. 给定关系模式R(U, F), 其中U为属性集, $U = \{X, Y, Z\}$, F是U上的一组函数依赖。函数依赖的公理系统(Armstrong公理系统)中的()是指“若 $X \rightarrow Y, X \rightarrow Z$, 则 $X \rightarrow YZ$ 为F所蕴涵”。

- A. 自反率 B. 传递率 C. 增广律 D. 合并规则

43-44. 在关系R(A1, A2, A3)和S(A3, A4, A5)上进行关系运算的4个等价的表达式E1、E2、E3和E4如下所示:

$$E_1 = \pi_{A_1, A_4}(\sigma_{A_2 < '2022' \wedge A_4 = '95'}(R \bowtie S))$$

$$E_2 = \pi_{A_1, A_4}(\sigma_{A_2 < '2022'}(R) \bowtie \sigma_{A_4 = '95'}(S))$$

$$E_3 = \pi_{A_1, A_4}(\sigma_{A_2 < '2022' \wedge R.A_3 = S.A_3 \wedge A_4 = '95'}(R \times S))$$

$$E_4 = \pi_{A_1, A_4}(\sigma_{R.A_3 = S.A_3}(\sigma_{A_2 < '2022'}(R) \times \sigma_{A_4 = '95'}(S)))$$

如果严格按照表达式运算顺序, 则查询效率最高的是(), 将该查询转换价的SQL语句为: SELECT A1, A4 FROM R, S WHERE()。

- A. E1 B. E2 C. E3 D. E4

A. R. A2 <'2022' OR S. A4 <'95'

B. R. A2 <'2022' AND S. A4 <'95'

C. R. A2 <'2022' OR S. A4 <'95' OR R. A3=S. A3

D. R. A2 <'2022' AND S. A4 <'95' AND R. A3=S. A3

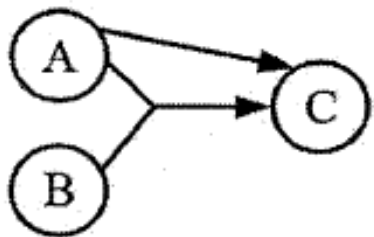
函数依赖

◆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。

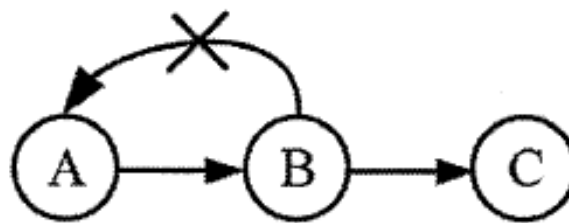
函数依赖又可扩展以下两种规则：

◆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。

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, 系名)

系(系名, 系主任)

选课(学号, 课程号, 成绩)

每张表都属于3NF。

考试试题

43-44. 给定关系模式 $R(U, F)$ ，其中：属性集 $U = \{A, B, C, D, E, G\}$ ，函数依赖集 $F = \{A \rightarrow BC, C \rightarrow D, A \rightarrow D, E \rightarrow G\}$ 。关系 R 中（ ），函数依赖集 F 中（ ）。

- A. 有1个候选码A
 - B. 有1个候选码AE
 - C. 有2个候选码AC和AE
 - D. 有2个候选码CE和AE
- A. 存在传递依赖，但不存在冗余函数依赖
- B. 既不存在传递依赖，也不存在冗余函数依赖
- C. 存在传递依赖，并且存在冗余函数依赖 $A \rightarrow D$
- D. 不存在传递依赖，但存在冗余函数依赖 $A \rightarrow D$

候选关键字的求法：根据依赖集，找出从未在右边出现过的属性，必然是候选键之一，以该属性为基础，根据依赖集依次扩展，看能否遍历所有属性，将无法遍历的加入候选键中。

并发控制

40. 在数据库的安全机制中，通过建立()使用户只能看到部分数据，从而保护了其它数据的安全性。

- A. 索引 B. 视图 C. 触发器 D. 存储过程

42. 在数据库系统中，视图实际上是一个()。

- A. 真实存在的表，并保存了待查询的数据
B. 真实存在的表，只有部分数据来源于基本表
C. 虚拟表，查询时只能从1个基本表中导出的表
D. 虚拟表，查询时可以从1个或多个基本表或视图中导出的表

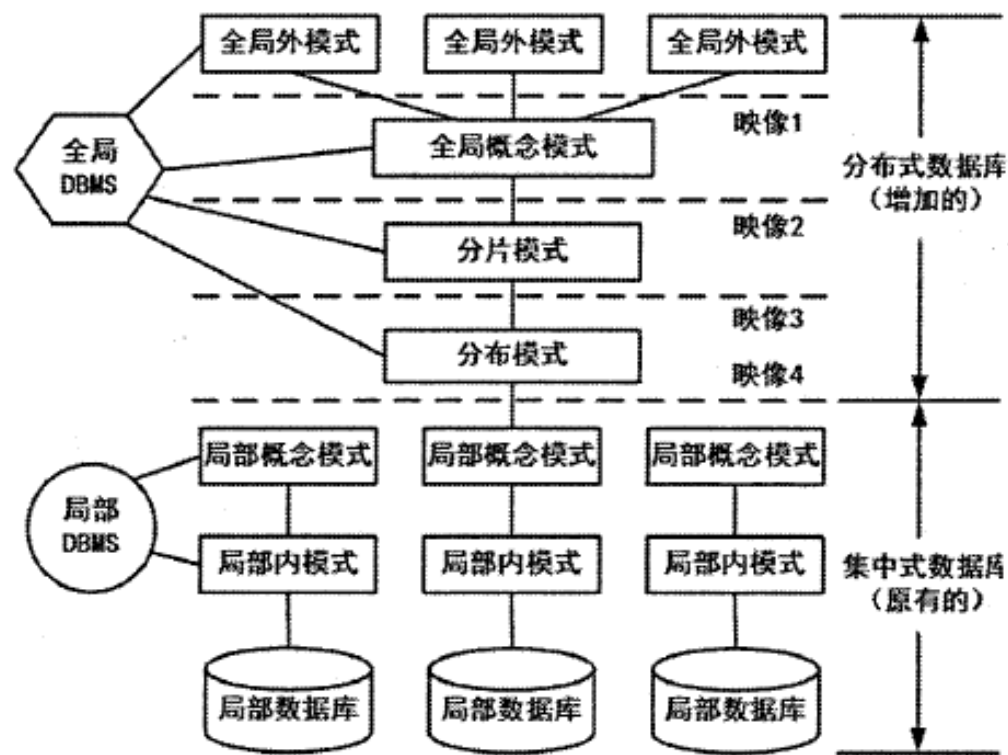
41. 事务的()是指“当多个事务并发执行时，任一事务的更新操作直到其成功提交的整个过程，对其他事务都是不可见的”。

- A. 原子性 B. 一致性 C. 隔离性 D. 持久性

40. 在数据库系统中，一般将事务的执行状态分为五种。若“事务的最后一语句自动执行后”，事务处于()状态。

- A. 活动 B. 部分提交 C. 提交 D. 失败

分布式数据库



◆局部数据库位于不同的物理位置，使用一个全局DBMS将所有局部数据库联网管理，这就是分布式数据库。

◆分片模式

水平分片：将表中**水平的记录**分别存放在不同的地方。

垂直分片：将表中的**垂直的列值**分别存放在不同的地方。

◆分布透明性

分片透明性：用户或应用程序**不需要知道逻辑上访问的表具体是如何分块存储的**。

位置透明性：应用程序**不关心数据存储物理位置的改变**。

逻辑透明性：用户或应用程序**无需知道局部使用的是哪种数据模型**。

复制透明性：用户或应用程序**不关心复制的数据从何而来**。

考试试题

某银行信息管理系统采用分布式数据库系统，以便对本地储户的存储业务能够在本地正常进行而不依赖于其他场地的数据库，这种情况称为分布式数据库的（ ）。

- A. 分布性 B. 共享性 C. 可用性 D. 自治性

41. 某证券公司股票交易系统采用分布式数据库，这样本地客户的交易业务能够在本地正常进行，而不需要依赖于其他场地数据库，这属于分布式数据库的（ ）特点。

- A. 共享性 B. 分布性 C. 可用性 D. 自治性

45. 分布式数据库系统中的两阶段提交协议（Two Phase Commit Protocol, 2PC 协议）包含协调者和参与者，通常有如下操作指令。满足2PC的正常序列是（ ）。

- ①协调者向参与者发prepare 消息
- ②参与者向协调者发回ready消息
- ③参与者向协调者发回abort 消息
- ④协调者向参与者发commit消息
- ⑤协调者向参与者发rollback 消息

- A. ①②④
B. ①②⑤
C. ②③④
D. ②③⑤



谢谢！