

一、多选题

1. 文件系统管理数据的主要缺点有(**AC**)
 - A. 数据冗余与不一致
 - B. 数据更新具有原子性
 - C. 数据孤立
 - D. 高并发数据访问
2. DBA 的主要职责包括(**ABCD**)
 - A. 安全和完整性控制
 - B. 数据库性能监控、分析与优化
 - C. 数据库恢复
 - D. 数据库结构调整
 - E. 事务处理与并发控制
3. 下列哪些属于过程性查询语言? (**C**)
 - A. 元组关系演算
 - B. 域关系演算
 - C. 关系代数
 - D. SQL
4. 六个基本关系代数操作包括(**B**), 其他操作都可以由这些操作推导得到。
 - A. \cap , $-$, π , σ , \times , ρ
 - B. \cup , $-$, π , σ , \times , ρ
 - C. \cup , $-$, π , σ , \bowtie , ρ
 - D. \cap , \div , π , σ , \bowtie , ρ
5. 下列哪些可以确定一个关系中的所有属性? (**ACD**)
 - A. 超码
 - B. 外码
 - C. 候选码
 - D. 主码
 - E. 搜索码
6. 下列哪些操作可以在视图上进行? (**ABC**)
 - A. 查询
 - B. 定义新的视图
 - C. 更新视图
 - D. 定义新的关系
7. 给定关系 $R<U, F>$, $U=\{A, B, C, D\}$, $F=\{A \rightarrow B, B \rightarrow D, C \rightarrow D, D \rightarrow A\}$, 以及 R 的分解 $\rho=\{AB, BD, CD\}$, 则该分解是 (**B**)
 - A. 无损连接, 依赖保持
 - B. 无损连接, 不依赖保持
 - C. 有损连接, 依赖保持
 - D. 有损连接, 不依赖保持
8. 下列哪些协议可以避免出现死锁? (**BC**)
 - A. The Two-phase Locking Protocol
 - B. The Timestamp-based Protocol
 - C. The Graph-based Protocol
 - D. Strict Two-phase Locking
9. 查询处理主要包括哪些步骤? (**ACD**)
 - A. Parser and translator
 - B. Serialization
 - C. Optimization
 - D. Evaluation
10. 事务是一组访问或更新数据的操作, 事务具有下列属性(**ABCF**)
 - A. Consistency
 - B. Isolation
 - C. Atomicity
 - D. Recoverability
 - E. Serializability
 - F. Durability

二、数据库查询、设计与优化

1. 给定如下关系 R、W、D，计算关系代数表达式的结果。

a. $R_1 = \Pi_{Y,T}(R)$

b. $R_2 = \sigma_{P>5 \wedge T=e}(R)$

c. $R_3 = R \bowtie W$

d. $R_5 = \Pi_{T,Y}(R) - D$

e. $R_4 = \Pi_{[2],[1],[6]}(\sigma_{[3]=[5]}(R \times D))$ ，其中 $[x]$ 表示关系中第 x 个属性。例如，R 的第一个属性为 P。

关系 R

P	Q	T	Y
2	b	c	d
9	a	e	f
2	b	e	f
9	a	d	e
7	g	e	f
7	g	c	d

关系 W

T	Y	B
c	d	m
c	d	n
d	f	n

关系 D

T	Y
c	d
e	f

参考答案:

a.

YT

d c

f e

e d

b.

PQTY

9 a e f

7 g e f

c.

PQTYB

2 b c d m

2 b c d n

7 g c d m

7 g c d n

d.

TY

d e

e.

QPY

b 2 d

g 7 d

a 9 f

b 2 f

g 7 f

2. 数据库设计

在供应链管理中，有三个实体集：Product(商品), Store(商店),和 Supplier(供应商)，各自包含如下属性：

商品实体集：ProductID, ProductName, Specifications, and Unit Price

商店实体集：Store Number, Store Name, and City;

供应商实体集：Supplier Number, Supplier Name, and City.

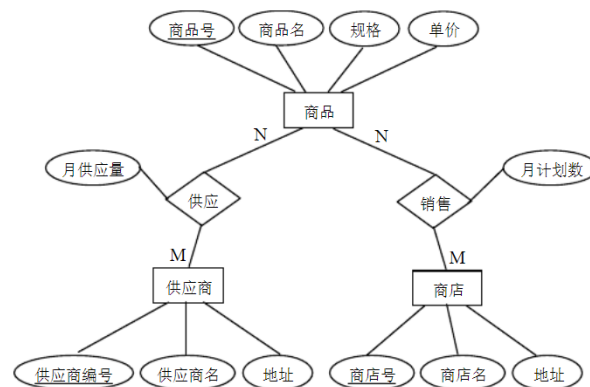
供应商和商品之间存在 Supply 关系，每个供应商可以提供多种商品，每种商品可以来自不同供应商，供应商一般按月提供商品；商店和商品之间存在 Sales 关系，每个商店可以出售多种商品，每个商品可以在多个商店销售，每个商店会针对每种商品制定月度采购计划。

- 画出 ER 图，创建对应关系模式，并指出每个关系的主码和外码；
- 用 SQL 语句定义 Product 表，并给出合理的完整性约束；
- 使用 SQL 查询：找出商店名称及其针对商品 tissue 的月度销售采购计划，要求该商品的供应商在 Shanghai；
- 画出小题 c 的查询对应的查询树。

参考答案：

(1)

ER 图及属性、联系图为：



这个 ER 图转换的关系模式如下：

商品 (商品号, 商品名, 规格, 单价)
供应商 (供应商编号, 供应商名, 地址)
商店 (商店号, 商店名, 地址)
供应 (商品号, 供应商编号, 月供应量)
销售 (商品号, 商店号, 月计划数)

b、c、d 所写 SQL 语句合理即可

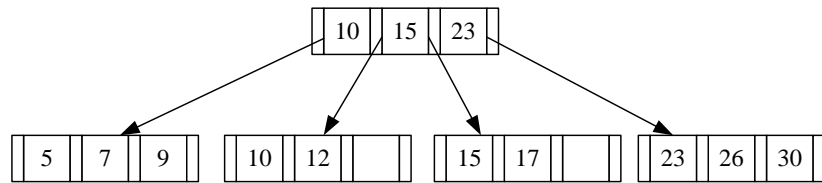
3. 规范化：给定关系模式 $R\langle U, F\rangle$, $U=\{A, B, C, D, E\}$, $F=\{AC\rightarrow BD, B\rightarrow C, C\rightarrow D, B\rightarrow E\}$;
- 使用 Armstrong axioms (公理)和相关的规则证明函数依赖 $AC\rightarrow E$;
 - 计算属性闭包 $(AB)^+$ 、 $(BC)^+$;
 - 计算 F 的正则覆盖 F_c ;
 - 找出 R 所有的候选码，并指出 R 满足第几范式;
 - 对 R 进行分解，使其满足 3NF，同时是无损分解，并依赖保持;
 - 证明上述分解为无损连接且保持依赖.

参考答案：

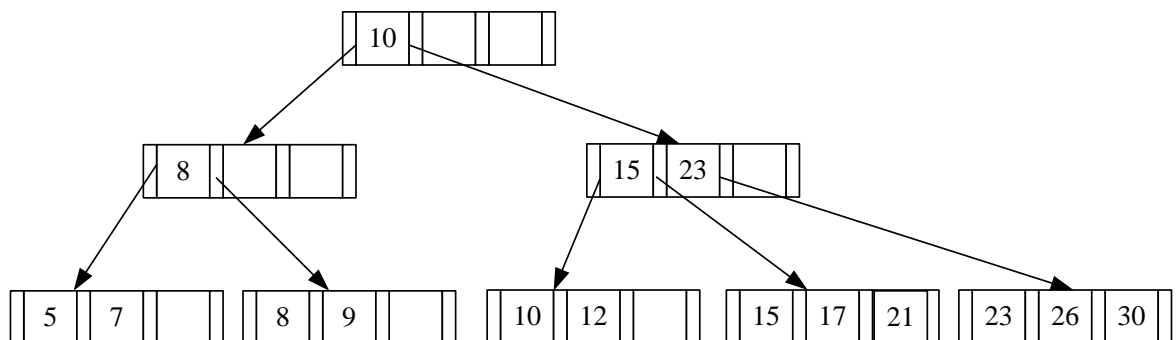
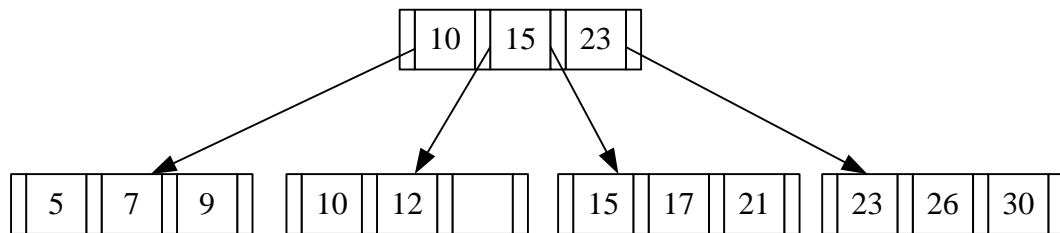
- 由 $AC\rightarrow BD$ 得 $AC\rightarrow B$ (分解规则); 再由 $B\rightarrow E$, 则有 $AC\rightarrow E$ (传递规则)
- $(AB)^+=ABCDE$, $(BC)^+=BCDE$
- 对 $R\langle U, F\rangle$ 中的函数依赖集 F 进行极小化处理，得最小依赖集 $F_c=\{AC\rightarrow B, B\rightarrow C, C\rightarrow D, B\rightarrow E\}$, 仍记为 F ; (可以有多种答案)
- 判定 R 属于第几范式:
 R 的候选码有: AC 、 AB ; 主属性为 A 、 B 、 C ;
 由 $C\rightarrow D$ 可见，非主属性 D 对码 AC 为部分函数依赖，故 $R\notin 2NF, R\in 1NF$.
- 将关系模式 R 分解为 3NF:
 全部属性均在 F 中出现了; 不存在 $X\rightarrow A\in F$, 且 $XA=U$.
 则对 F 按相同左部原则分组，有
 $U_1=\{A, B, C\}$, $F_1=\{AC\rightarrow B, B\rightarrow C\}$
 $U_2=\{B, C, E\}$, $F_2=\{B\rightarrow C, B\rightarrow E\}$
 $U_3=\{C, D\}$, $F_3=\{C\rightarrow D\}$
 $\rho=\{R_1\langle U_1, F_1\rangle, R_2\langle U_2, F_2\rangle, R_3\langle U_3, F_3\rangle\}$ 为保持函数依赖的分解。
 f. 由于码 AC 、 AB 都包含在 U_1 中，因此，由检测算法可以找到相应表中的一行可以成为 a_1, a_2, a_3, a_4, a_5 (用算法过程的表变化描述)，则 ρ 同时也具有无损连接性。

三、数据库管理系统

1. 针对下列 B⁺树索引，分别给出插入 21 和 8 之后的情况，并给出插入操作的具体步骤。



参考答案：分步骤画出。



2. 数据库并发操作可能产生丢失修改、不可重复读、读脏数据等数据不一致问题。请举例阐述这些问题的含义。基于锁的协议可以有效解决上述问题，但是可能出现饥饿（Starvation）现象。请解释什么是饥饿现象。

参考答案：

阐述清楚丢失修改、不可重复读、读脏数据和饥饿现象的含义即可，要求结合例子阐述。

3. 事务管理

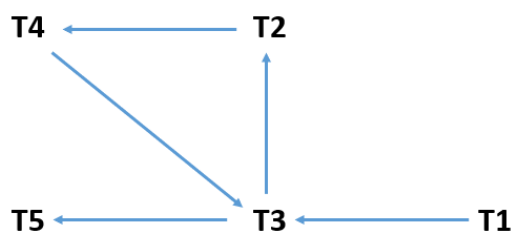
- 针对下列 5 个并发事务，是否满足两段锁协议？构建一个等待图，检查该调度是否会产生死锁。
- 在事务管理中，如何预防和检测死锁，以及死锁发生后如何恢复？

T ₁	T ₂	T ₃	T ₄	T ₅
lock-S(X) read(X)				
	lock-S(X) read(X)			
unlock(X)				lock-S(C) read(C)
		lock-X(Y) write(Y) lock-X(X)		
lock-S(Y)			lock-X(D)	
	lock-X(D)		lock-S(Y)	
		lock-x(C)		unlock(C)
unlock(Y)			unlock(D)	

参考答案：

a. No

wait-for graph:



The schedule will lead to a deadlock because there is a circle among T₂, T₃, and T₄.

b. 预防死锁方法：graph-based protocol, timestamps-based protocol

检测死锁：wait-for graph, kill 导致死锁 transaction 或回滚部分操作

4. 恢复系统

下图中包含 4 个并发事务 T₁, T₂, T₃, T₄, 其中数据 A, B, C 的初始值为:
A=800, B=500, C=300.

- 如果日志采用“**Immediate database modification**”，且在事务 T₂ 开始时创建了一个检查点（checkpoint），给出上述调度的日志记录以及 A, B, C 的最终结果.
- 在事务 T₂ 提交前发生了故障，应当采取什么样的恢复操作，并解释原因？

T ₁	T ₂	T ₃	T ₄
read(A)			
A:=A+100		read(B)	
write(A)			
	read(C)	B:=B+100	
		write(B)	read(A)
			write(A)
	C:=C+100		
read(A)	write(C)		
read(B)			
B:=B+A			
write(B)			
		read(C)	
		C:=100	
		write(C)	

参考答案:

a. The log records for the above transaction schedule:

<T₁ start>

<T₃ start>

<T₁, A, 800, 900>

<checkpoint{T₁, T₃}>

<T₂ start>

<T₃, B, 500, 600>

<T₄ start>

<T₄ commit>

<T₂, C, 300, 400>

<T₂ commit>

<T₁, B, 600, 1500>

<T₁ commit>

<T₃, C, 400, 100>

<T₃ commit>

A=900, B=1500, C=100

因为建立了检查点，发生故障的时候，T4 已经 commit，T1、T3、T2 已经 start 而尚未 commit，因此：

T4 需要 redo，T1、T3、T2 需要 undo

四、附加题

1. 对数据进行排序时，如果数据规模较大，则无法将所有数据加载到内存进行排序，需要在内存和磁盘之间进行多次数据交互。归并排序是实现这类排序的一种有效方法。假设每个数据块只能容纳一个元组，内存可以保留 3 个数据块。给出针对下列元组中第一个属性的归并排序的过程。

$t_1=(j,17); t_2=(u,21); t_3=(e,1); t_4=(w,13); t_5=(p,3); t_6=(m,8); t_7=(v,4); t_8=(z,11);$
 $t_9=(n,6); t_{10}=(i,9); t_{11}=(h,2); t_{12}=(b,12).$

2. AI4DB 和 DB4AI 是目前数据库领域的前沿热点，相关研究极大地提升了传统数据库的性能。请选择一个与 AI4DB 或 DB4AI 相关的研究方向，从研究问题、主要挑战、相关技术和个人看法等方面进行阐述。
3. 图（Graph）数据是一种重要的数据类型，广泛应用于社交网络、金融风控、智能交通等领域。因此，实现对图数据的有效管理具有重要意义。请从存储模式、查询处理和主流系统等角度阐述图数据管理相关技术与平台。

1. 参考答案：

Using tuple numbers t_1 through t_{12} . We refer to the j -th run used by the i -th pass as r_{ij} . The initial sorted runs have three blocks each. They are:

$r_{11} = \{t_3, t_1, t_2\}$
 $r_{12} = \{t_6, t_5, t_4\}$
 $r_{13} = \{t_9, t_7, t_8\}$
 $r_{14} = \{t_{12}, t_{11}, t_{10}\}$

Each pass merges two runs. Therefore, the runs after the end of the first pass are:

$r_{21} = \{t_3, t_1, t_6, t_5, t_2, t_4\}$
 $r_{22} = \{t_{12}, t_{11}, t_{10}, t_9, t_7, t_8\}$

At the end of the second pass, the tuples are completely sorted into one run:

$r_{31} = \{t_{12}, t_3, t_{11}, t_{10}, t_1, t_6, t_9, t_5, t_2, t_7, t_4, t_8\}$

2 和 3 阐述合理即可，要求涵盖指定的几方面的内容。