

《資料處理概要》

試題評析

第一題：此題為資料庫中並行控制的觀念，屬基本觀念，有準備此科之學生應可順利作答。

第二題：詢問正規化相關的議題，但並非以關聯表為基礎，以概念詢問為主要方向，難度不高。

第三題：資料結構的基本問題，並要求提出實例，需了解內容，死背不易拿分。

第四題：網路中傳輸層的觀念，難度較低，同學可輕易拿分。

第五題：時間複查度相關的問題，第二小題有陷阱，同學需謹慎作答。

一、假設資料庫運算的兩個交易，transaction_1與transaction_2，有共同的資料項目x跟y。也就是說，兩個交易都分別有讀寫共用資料的動作。例如，一個是{read(x), read(y), write(y), write(x)}，而另一個交易為{read(y), write(y), read(x), write(x)}。類此的交易運算表示方式，一個包含兩個交易的排程（schedule），在什麼條件情況下是不允許並行處理（concurrent execution）？又在什麼情況下是允許並行處理？除了敘述緣由外，請分別舉例說明。（20分）

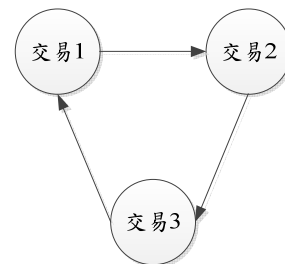
答：

排程可分為循序性排程(Serial Schedules)和可循序性排程(Serializable Schedules)。前者代表各交易間無並行性，即各單元操作之執行順序不會交錯(Interleaving)執行。相對的，後者則表排程中各交易之單元操作可交錯執行。此外，若交錯執行的排程與循序性排程之執行結果相同，便稱此二排程具等價關係(Equivalent)，並稱此交錯執行之排程滿足可循序性(Serializable)。

(一)不允許的情況

以左下表為例，T₁和T₂單元操作為交易1之Read b和交易2之Write b，存取標的皆為b，發生衝突。先行圖如下所示，此例具迴圈結構，不滿足可循序性。

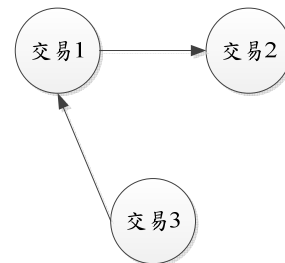
時間	交易1	交易2	交易3
T1	Read b		
T2		Write b	
T3			Read c
T4	Write c		
T5		Read a	
T6			Write a



(二)允許的情況

以左下表為例，T1和T2單元操作為交易1之Read b和交易2之Write b，存取標的皆為b，發生衝突。先行圖如下所示，此例不具迴圈結構，滿足可循序性。

時間	交易1	交易2	交易3
T1	Read b		
T2		Write b	
T3			Read c
T4	Write c		



【高分命中】石濟，《資料處理講義IV》，頁111-113。

二、關於資料庫系統的設計，我們常需先定義資料的綱要（schema），請問綱要的作用為何？在關聯式資料庫與物件導向式資料庫中，物件導向式資料庫比關聯式資料庫較具優勢之處為何？在考慮關聯式的資料庫設計時，我們常會因正規化的需要，將資料表作分解（decomposition）的動作。比如說，一個關於公司人員的資料表，假設該表原包含三個欄位（名字、職稱、部門），若資料庫設計者想將其分成兩個表格：（名字、職稱）與（職稱、部門），請問如此分解是好的正規化分解嗎？為什麼？請進一步說明藉由資料表分解，已達到正規化的目的為何？（20分）

答：

- (一)關聯是資料庫中，以ANSI/SPARC架構而言，綱要可分為外部綱要、概念綱要、以及內部綱要，各綱要間存在對應關係，可達成資料獨立性，亦即上層綱要並不會因下層綱要改變而受到影響，換言之，當下層綱要亦不會因上層綱要產生任何改變。
- (二)物件導向式資料庫模型統一了應用程式和資料庫開發的資料模型和語言環境，讓程式設計和資料庫統一使用物件導向方式進行設計。此模型亦可視作物件導向程式設計與資料庫技術的結合，減輕開發成本。
- (三)否，正規化分解是為消除相依性，但此題中假設名字不重複為主鍵，功能相依為：名字→職稱、名字→部門，職稱和部門間就常理而言不具功能相依，故此處不需對（名字、職稱、部門）做進一步的分解。
- (四)如同前述，正規化分解是為消除欄位間的相依性，避免新增異常、修改異常和刪除異常等，建立良好結構的關聯表。

【高分命中】石濟，《資料處理講義IV》，頁55-64。

三、請回答下列問題：（每小題5分，共20分）

- (一)資料結構中，樹（tree）的結構與圖（graph）的結構有何差別？
- (二)在一個有根節點（root node）的樹，其節點（node）的數目與邊（edge）的數目有何關係？
- (三)在沒方向性的連接圖（connected graph）中，一個節點的度數（degree）定義為該節點連接的邊的個數。請問其節點的度數與邊的數目有何關係？
- (四)請分別舉出兩個樹結構與圖結構在電腦應用中實際的例子。

答：

(一)比較如下：

	樹的結構	圖的結構
節點間是否具階層性	是，有父子關係	否，無特定階層
任意二節點間可否有鏈結	否，僅相鄰階層的點可連接，且一個子節點僅可連接一個父節點	是
是否具特殊節點	最高階層之點稱為樹根	否

(二)節點數=邊數+1

(三)

$$\sum_{i=1}^n d_i = 2 \times E, d_i \text{ 表第 } i \text{ 個點的度數, } n \text{ 表總點數, } E \text{ 表總邊數}$$

(四)

樹結構：

- (1)將資料建構為二元搜尋樹，依照中序走訪可由小至大完成排序
- (2)將資料建為堆積樹，可在 $O(\log^n)$ 的時間複雜度內找出極值

圖結構：

- (1)尋找最短路徑時，以圖結構表示每一步驟中候選的對象
- (2)佈建網路時，將每一個點視為部門，所有部門間全部連通成為圖，尋找最小佈建成本時便可用運最小成

本擴張樹演算法。

【高分命中】石濟，《資料處理補充講義》，頁2-4。

四、在通訊網路上使用的資料交換技術，以電路交換（circuit switch）與封包交換（packet switch）為主。請以線路的使用效率與資料傳輸的延遲因素，說明兩種技術間的差異性？在TCP/IP網路中，傳輸層的TCP與UDP的傳輸協定，何者可以稱為電路交換技術的應用？為什麼？（20分）

答：

(一)比較如下：

	電路交換	封包交換
線路使用效率	低，因每個用戶接分配一條線路，但大多數時間接閒置，故使用率低	高，因當用戶需要傳送資料時才分配實體線路，可達較高之線路使用率
資料傳輸延遲	低，因用戶已實體擁有線路，故傳輸品質高。	高，當尖峰時段多數用戶同時需傳輸資料時，便會發生擁塞現象，導致傳輸延遲，甚至封包遺失

(二)TCP可稱為電路交換技術的應用，雖然其下層是以封包交換技術為基礎，但此協定提供相對嚴謹的錯誤偵測與重送機制，能夠達到可靠傳輸的效果，雖然仍有傳輸延遲的問題，但透過重送可大幅降低收方無法收到封包的現象。

【高分命中】石濟，《資料處理講義IV》，頁51-54。

五、假設三種演算法的複雜度函數分別為 $f_1(n)=n^2$ ， $f_2(n)=2^n$ ， $f_3(n)=\log(n!)$ 。將此三個函數依複雜度的高低順序排序，其結果為何？請說明原因？假若 $g(n)=1+(1/2)+(1/2^2)+\dots+(1/2^{n-1})$ 。請問“ $g(n)=O(n)$ ”是否正確？為什麼？（註1：此題中 n 為正整數，註2： $O(n)$ 表示big-O of n 。）（20分）

答：

(一)排序如下：

$$f_2(n) > f_1(n) > f_3(n)$$

$f_2(n) > f_1(n)$ ：因變數位於指數之函數，其值較變數位於底數之函數高

$f_1(n) > f_3(n)$ ： $f_3(n)=\log(n*(n-1)*(n-2)*\dots*1) < \log(n*n*n*\dots*n)=\log(n^n)=n\log(n)$ ，故 $f_1(n) > f_3(n)$

(二)正確

$g(n)$ 是公比為 $1/2$ 之等比級數和，故： $g(n)=1-(1/2)^{n-1}$ ，但題目並未說明 $g(n)$ 為演算法複雜度函數，若 $g(n)$ 是欲解決之問題，而該問題以迴圈撰寫程式碼時，需執行 n 次，故複雜度為 $O(n)$ 無誤。

【高分命中】石濟，《資料處理補充講義》，頁5。