

《資料處理概要》

一、佇列 (queue) 和堆疊 (stack) 是二種常用的資料結構。請回答下列問題。

- (一) 若要用深度優先的方式 (depth-first search) 走訪一樹狀結構 (tree structure) 的所有節點 (node)，請問佇列和堆疊，何者較適合？並說明原因。(10分)
- (二) 若要用廣度優先的方式 (breadth-first search) 走訪一樹狀結構 (tree structure) 的所有節點 (node)，請問佇列和堆疊，何者較適合？並說明原因。(10分)

試題評析 本題考queue和stack及DFS/BFS應用上的基本觀念，是演算法和資料結構的基礎，屬於基本送分題。

答：

- (一) 堆疊較適合，因為Depth-first search本質上就是看到一條路能走，就先走該條路，所以跟堆疊LIFO的概念比較相近。而事實上如果用傳統遞迴的方式，去實作DFS，它背後實際上就是透過系統堆疊完成的，函數呼叫的狀態都是保存在系統堆疊中。
- (二) 佇列較適合，因為Breadth-first search跟Depth-first search不同，他是先從該節點往外看有哪些路可以走，再一層一層拓展出去，跟Depth-first search找到路就走的概念不太一樣。而佇列正好適合這個特性，因為佇列是FIFO，當我們碰到可以走的路，就先放進佇列，等到所有能走的路都放進來後，再從裡面取出一條路去走。

二、有一資料表如下圖，共有八筆資料，第一欄是鍵值 (key value) 的二進位表示法 (Binary representation)，第二欄是姓氏資料。今要用雜湊 (hash) 的方式來將資料分成若干籃子 (bucket)，請回答下列有關雜湊的問題。

鍵值 (二進位表示法)	姓
1000	Lee
0000	Chen
0010	Wang
1001	Hwang
1011	Lu
0100	Chao
1110	Liao
0111	Chang

- (一) 假設雜湊函數 (hashing function) 是 $key \bmod 4$ ，亦即將 key 的值除以 4 之餘數值，即為雜湊函數值。假設每個籃子的容量最多可存三筆資料，請以上表之資料為例，將鍵值之二進位值轉換為十進位值，並將這些資料按所給的雜湊函數，區分為應有的籃子，並以此例說明什麼是靜態雜湊 (static hashing)。(10分)
- (二) 假設每個籃子的容量最多可存二筆資料，請以上表鍵值之二進位表示法為例，用動態雜湊 (dynamic hashing) 的方法，將這八筆資料做 hash。(10分)

試題評析 這題主要在考Hash的基本觀念，以及碰撞的處理方式，難度比前一題高，但仍屬應把握分數的題型。

答：

(一) $Key \bmod 4$ ，其實就相當於只看二進位後兩位所以最後可以建出以下hash table

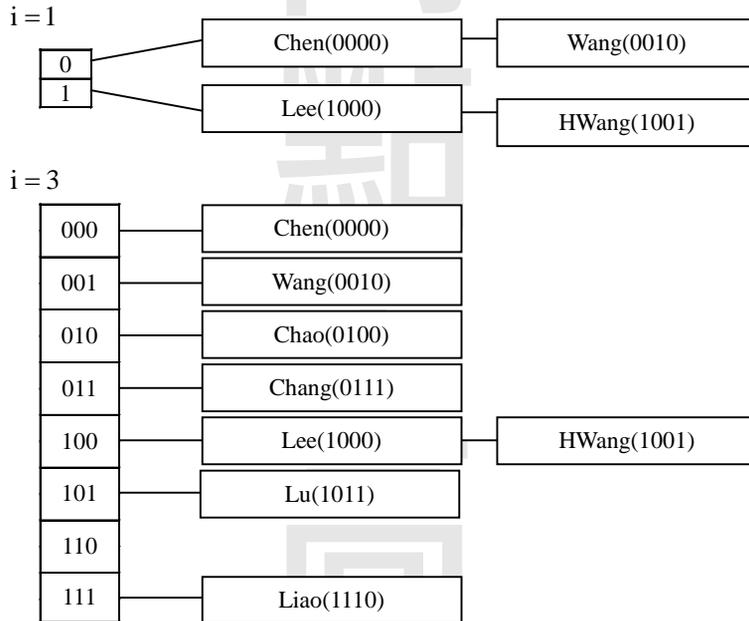
0 (00)	1 (01)	2 (10)	3 (11)
--------	--------	--------	--------

Lee	Hwang	Wang	Lu
Chen		Liao	Chang
Chao			

由於題目說，每個Bucket最多放三個，所以這邊剛好夠放，不需要動態擴增Hash table 也就是所謂的Static Hashing

(二)Dynamic hashing就是可以動態新增Bucket的hashing方法

例如我下面的例子，當bucket只有2個時，剛好可以放前四個（用二進位來分），但當下一個要放進去時，就要強制分裂成更多Bucket才能放入



三、請說明物聯網（Internet of Things）、網際網路（Internet）、全球資訊網（World Wide Web）這三者不同之處。（10分）

試題評析

這題單純考名詞解釋，其中IoT屬於近年比較熱門的話題，比較偏時事題，這就考驗學生平常對資訊科技流行的敏銳度。

答：

- (一)物聯網：物聯網是由實際硬體，例如家用路由器、車輛、嵌入式應用等裝置，透過網際網路連結形成的訊息交換網路。
- (二)網際網路：網際網路是一個全球性的電腦網路，他將全球各個獨立的網路結合在一起，形成一個龐大的網路。
- (三)全球資訊網：常又稱為WWW，是一個透過網際網路存取，由網頁、超文字連結組成的一個系統。它是網際網路能提供的服務之一，所以不等同網際網路。

四、有一資料庫其中一個關連（relation）裡有 A, B, C, D, E 這些欄位（attribute），且他們之間存在一些功能相依（functional dependency）的關係： $\{CE \rightarrow ABD, A \rightarrow D, E \rightarrow B\}$ 。此關連已滿足第一正規化。

- (一)請將此關連作第二正規化處理，說明理由，並指出何者為關鍵欄位（key attribute）。（10分）

(二)請將此關連作第三正規化處理，說明理由，並指出何者為關鍵欄位。(10分)

試題評析 資料庫正規化以及功能相依的應用題目，內容略複雜，按照定義下去做即可。

答：

(一)E→B代表B部分相依於E，應該把B拆出去成一個表，然後E當成關鍵欄位

(二)A→D代表遞移相依，根據定義應該把D拆出去成一個表，然後A當成關鍵欄位

五、有一資料庫記載一些客戶資料，和他們下訂單訂購貨物的資訊如下：

客戶 (身分證字號，姓名，電話，地址，年齡)

貨品 (貨品編號，貨名，價格，庫存量)

訂單 (貨品編號，身分證字號，訂貨日期，訂購數量)

請用 SQL 回答下列查詢。

(一)找出所有年齡大於 60 歲 (年齡 > 60)，所下訂單之訂貨日期為 20180101 的客戶，列出他們的姓名和地址。(10分)

(二)找出年齡低於 20 歲 (年齡 < 20) 的客戶，列出他們的姓名，電話和在 2017 年 (20170101 ≤ 訂貨日期 ≤ 20171231) 裡下訂單的次數。(10分)

試題評析 常見的SQL題目，根據題目描述照著列出條件即可。

答：

(一)SELECT 客戶.姓名, 客戶.地址 FROM 客戶, 訂單 WHERE 客戶.年齡 > 60 AND 訂單.訂貨日期 = 20180101

(二)SELECT 客戶.姓名, 客戶.電話, COUNT(客戶.身分證字號) FROM 客戶, 訂單 WHERE 客戶.年齡 < 20 AND 訂單.訂貨日期 >= 20170101 AND 訂單.訂貨日期 <= 20171231

六、說明什麼是「區塊鏈」(Block Chain)。列舉四個區域鏈最重要的特色，並說明之。(10分)

試題評析 這題也是單純的名詞解釋，考的是這一兩年非常熱門的區塊鏈，所以也說明平常多關注新的資訊議題對於考試還是有幫助的。

答：

(一)區塊鏈：最早由中本聰白皮書提出區塊鏈概念，其基本原理是採用密碼學，將每個區塊中的內容做加密雜湊，每一個區塊包含前一個區塊的內容，環環相扣，使得區塊鏈上的節點難以偽造，適合拿來當作紀錄帳本等資訊的方式。

(二)特色：去中心化、難以竄改、運作採用Proof of Work、具有一定程度匿名性。

【版權所有，重製必究！】