

114年專門職業及技術人員高等考試建築師、
25類科技師（含第二次食品技師）、大地工程
技師考試分階段考試（第二階段考試）
暨普通考試不動產經紀人、記帳士考試試題

等 別：高等考試
類 科：資訊技師
科 目：資料結構與資料庫及資料探勘
考試時間：2 小時

座號：_____

※注意：(一)禁止使用電子計算器。

(二)不必抄題，作答時請將試題題號及答案依照順序寫在試卷上，於本試題上作答者，不予計分。

(三)本科目除專門名詞或數理公式外，應使用本國文字作答。

一、假設有一筆長度為 $n=10^6$ 的資料，資料性質約 90% 已排序（近乎有序），其餘 10% 隨機亂序。每筆資料大小 100 bytes，儲存在主記憶體中。請回答：（每小題 10 分，共 20 分）

(一)在這種情況下，使用 Insertion Sort、Merge Sort、Quick Sort，請比較其平均與實際效能，並說明那一種最適合。

(二)在 Quick Sort 中實作上可採用多種 pivot 選擇策略（如：第一個元素、最後一個元素與中位數三取法 median-of-three），請說明這些策略對於上述資料進行反向（Reversed Order）排序的效能影響，並指出最推薦的選擇方法及理由。

二、給定一有向圖 $G=(V,E)$ ，其邊權重皆為正數。假設目前採用 Dijkstra 演算法計算單源（single-source）最短路徑，但在實際系統中，邊權重會隨時間動態變化（例如交通路網延遲）。請回答：（每小題 10 分，共 20 分）

(一)為何 Dijkstra 不適合在權重頻繁變動的情境下重複執行？試設計一種“增量式更新演算法”（incremental update approach），能在部分邊權更新後有效地維護最短路徑樹（SPT）。

(二)若圖使用 Fibonacci heap 實作 priority queue，請比較時間複雜度與一般 binary heap 的差異。

三、假設系統支援四種隔離等級：Read Uncommitted、Read Committed、Repeatable Read、Serializable。有兩個交易如下表，請回答下列問題：(每小題 10 分，共 20 分)

- (一)在四種不同隔離等級下，T1 是否可能讀到不同的 A 值，並說明理由？
(二)試分析異常現象 (dirty read、non-repeatable read、phantom read) 在何種隔離等級會發生？

Transaction T1	Transaction T2
Read (A)	
	Write (A=100)
Read (A)	Commit

四、假設一個 B+ Tree 索引的階 (order)=4，每個節點最多容納 4 個子節點。已存在以下鍵值：[5, 9, 12, 15, 18, 21, 25, 31, 33, 38]。(每小題 10 分，共 20 分)

- (一)請繪出最終的 B+ Tree 結構 (節點鍵值排列)。並執行範圍查詢 WHERE key BETWEEN 10 AND 30，請說明實際 I/O 步驟 (指明訪問那些節點)。
(二)若鍵值 25 被刪除，請說明重新平衡 (redistribution 或 merge) 的過程。

五、某資料集有屬性 A (二元)，以及目標屬性 Class(+/-)。資料如下表。請計算以下數值。(每小題 10 分，共 20 分)

- (一)整體熵 (Entropy of dataset)。
(二)以屬性 A 為分裂條件的資訊增益 (Information Gain)，並請修改其中一筆的 Class 值可以提高屬性 A 的資訊增益。

A	Class
T	+
T	+
T	-
F	-
F	-
F	+