

《資料處理概要》

試題評析

資料處理無論對統計或資訊人員而言，一向都不是個簡單的科目，本試題雖然是普考考題，但是考題份量仍然偏重，每一題的內容都相當多，同學必須把握時間作答。

本次試題的部份考題相當偏重計算機底層技術，而且相當古老。例如第一、二題，便是告訴同學們，計算機概論的內容又廣又雜，但是同學們還是必須都要有所涉獵。

第三題的交易管理、第四題的資料結構、以及第五題資料庫規劃及查詢，一向是資料處理的重點，相信同學們都已有準備。本次試題估計一般同學應可在50~60分上下，程度較佳的同學可超過60~70分。

一、Email是網際網路傳統的應用層服務功能之一，請問在目前的IP網路中，Email是經由那個傳輸層的協定來達成信件的傳遞？若是採用SMTP來傳送資料，在發送端傳輸真正的信件資料之前，發送端會執行哪些SMTP的命令（command），以作為跟接收端的識別用途？假設你是公司的新進員工，需要利用你的個人電腦，以使用公司已架設好的私有郵件伺服器來進行收發信件。請問你需要那些具體的動作，包括在伺服器及個人電腦上的設定，才能成功的收發信件？假設公司尚未有專屬的郵件伺服器，而是先採用公眾網路的Web mail來收發信件，那麼你需要做那些處理呢？（20分）

答：

(一)目前IP網路多採TCP/IP協定，其中傳輸層採用者為TCP(Transmission Communication Protocol)。例如Email的SMTP亦採用TCP傳輸層協定。

(二)SMTP傳送資料主要指令：

- 1.HELO：與郵件伺服器建立連結，初始主機對話。
- 2.MAIL FROM：識別郵件寄件者Email Address。
- 3.RCPT TO：識別郵件收件者Email Address。
- 4.DATA：開始處理郵件內容，最後以單一句號結尾。
- 5.QUIT：結束與郵件伺服器之連結，結束工作。

故在傳送真正資料前，會使用HELO、MAIL FROM、RCPT TO之指令，至DATA指令後，開始傳輸資料內容。

(三)以公司使用Microsoft Exchange Server，個人電腦使用Outlook設定為例：

- 1.伺服器端：於Exchange Server上新增使用者帳號密碼，或新增帳號並與公司AD同步。
- 2.使用者端：
 - (1)開啓Outlook，選擇工具列的Email帳戶功能
 - (2)新增Email帳戶
 - (3)選擇Server Type，如Microsoft Exchange Server
 - (4)指定Microsoft Exchange Server位置以及使用者帳戶名稱，並確認密碼
 - (5)設定完畢儲存

(四)若使用公眾網路Web mail，則不需做任何設定，只要在Web mail註冊使用者email帳號後，使用瀏覽器輸入帳號密碼便可直接登入Web mail，收發郵件。

二、以電腦語言的發展來講，第二代語言與第三代語言主要的差異在那裡？一般所謂的高階語言指的是第幾代的語言？以管理電腦運作的角度來檢視，一套電腦的作業系統（例如Unix或Windows）包括那些單元？以電腦系統的發展來講，迷你電腦（mini computers）與微電腦（micro computers）在作業系統的使用上有何差異？（20分）

答：

(一)

	第二代語言	第三代語言
簡稱	2GL	3GL
語言	組合語言(Assembly Language)	程序性語言(Procedural Language)
執行速度	較快	較慢
易讀性	較難閱讀	較容易閱讀
程式長度	較冗長	較簡短
執行所需記憶體	較小	較大
特性	較接近機器語言，不同機器有不同的組合語言	較接近人類語言，可使用同一語言於不同機器

(二)一般高階語言指的便是第三代語言(3GL)。

(三)作業系統主要管理電腦運作包含下列三項：

1.資源分配

(1)CPU的管理：除了分配CPU的時間外，還根據使用者的權限，設定程式執行的優先權，以及調整分配的時間片段長短。

(2)記憶體的管理：管理的策略很多，例如將主記憶體分成多個區段，每個區段分給不同的程式使用。

(3)輸出入裝置的管理：藉由作業系統的管理，可以讓多個程式同時使用印表機、磁碟機等輸出入設備。

2.監督系統狀態

(1)系統效能監督：監督電腦系統上各種軟體執行的狀況，使用者提出工作要求的回應時間（Response Time），以及CPU的利用率。

(2)系統安全管理：可以設定使用者的名稱、密碼和使用者的使用權限等。

3.提供公用程式

作業系統中提供許多公用程式（Utilities），以方便使用者應用作業系統的資源。例如更改檔名、複製檔案、搬移檔案、刪除檔案、排序檔案等，這些經常會使用的作業系統附屬程式，稱為公用程式。

(四)1.迷你電腦(Mini Computer)：迷你電腦的設備和大型電腦相似，包括了主機、磁帶機、磁碟機、讀卡機和作為I/O用的終端機，例如王安電腦就是以迷你電腦為其主力產品。其計算機結構(Computer Architecture)已經設計成容許許多人同時使用(透過網路)，一般用於學校、中小型企業等，價格在數十萬到數百萬之間。

2.微電腦(Micro Computer)：所謂微電腦(micro computer)則是將微處理機接上必要的記憶裝置和輸入輸出設備，如此便是一部完整的微電腦了，如 IBM PC。所謂微處理機(Microprocessor)即一般所稱之 CPU；若把 ALU與 CU 的電路燒製在同一矽晶片上，即稱為微處理機，如Z-80就屬於8位處理機。是目前最普及、價格最大眾化的電腦。

三、一般資料庫系統之交易（transaction）需滿足所謂的ACID性質，請解釋其中“I”代表著甚麼意義？當兩個交易同時執行時，為確保結果的一致性，可以藉由序列（serializability）架構的衝突（conflict）檢查，請說明何謂“衝突（conflict）”？又何謂一個平行執行的排程（schedule）可序列化（serializable）？為確保排程的序列性，我們可以鎖定（lock）協定為控制機制，請說明鎖定協定是如何運作的？（20分）

答：

(一)ACID的I代表「一致性(Consistency)」，即交易是從一個一致的狀態變更到另一個一致的狀態。也就是資料庫狀態在交易前後皆滿足資料庫所設定的限制，以及正確的結果。確保一致性通常是撰寫資料庫程式，或實現完整性限制的DBMS程式設計師的責任。

(二)當兩個以上的交易同時執行，交易操作之間發生的交互順序，可能產生交易間的衝突。例如下面兩個交易若存取相同資料項目，若有任一交易針對同一項目有寫入操作時，則會跟另一交易的相同項目的操作產生衝突。

T1 \ T2	讀(read)	寫(write)
讀(read)	否	是
寫(write)	是	是

- (三)一個n個交易的排程是可序列化的，假若此排程與相同的n個交易的某個序列排程(serial schedule)衝突等價(conflict equivalent)；即兩個排程發生衝突的順序相同。
- (四)鎖定(locking)是一個與資料庫中資料項目結合的變數，描述了資料項目的狀況，決定何種動作允許或不允許應用在此資料項目上。並行控制的鎖定技術主要為「二元鎖定(Binary locking)」、「共享互斥鎖定(Shared and Exclusive locking)」、與「兩階段鎖定(Two phase locking)」。

四、佇列(Queue)資料結構的特性為何？請舉出兩種電腦上使用堆疊的應用？以連結串列(Linked List)跟以連續陣列(Array)來實作佇列，在時間與空間的運作效率上有什麼差異？若允許有插隊加入的情況，在這兩種實作方式下如何完成？(20分)

答：

(一)佇列的特色是先進先出。即例如A、B、C進；A、B、C出。

(二)堆疊的應用：

- 1.處理遞迴式呼叫：在每次遞迴之前，須先將下一條指令的位址，暫存器及變數的值保存到堆疊中，當爾後遞迴回來時，能再從堆疊頂端取出暫存值，回到原來執行遞迴前狀況，並從下一指令繼續往下執行。
- 2.算術式之轉換：在編譯程式時，堆疊可幫助分析指令，以判斷某一個指令是否合法；此外，算術運算時，利用堆疊的特性將不同的算術式表示法轉換，以利算術式的執行。

(三)Linked List實作佇列時，Linked List的一側用以新增資料，另一側用以移出資料：因除了資料內容外，需另外耗費空間儲存資料間的指標(pointer)，故較浪費空間；但新增刪除資料時，Linked List內的資料不需挪動，較節省時間。

以Array實作佇列時，由於Array的index大小固定且連續，所以新增刪除資料時，需採push方式挪動資料位置，較浪費時間；但不需儲存指標，故較節省空間。

(四)若允許插隊加入情況，Linked List可直接插入資料至指定位置，僅需改變兩側指標指向即可；但Array必須將插隊之資料後方所有資料向後挪動一個位置，以空出位置存放插隊資料。

五、以學校學生的個人資料，修課資料與成績資料等來考慮，設計一關聯式資料庫系統時，請具體說明若將所有資料合併成一個大表格，可能會造成甚麼問題？以記錄一個人的名字為例，說明關聯式資料庫與物件導向式資料庫的優缺點？SQL查詢語言有關表格資料的運算中，請列出三種主要的運算？相對於SQL作為關聯式資料庫的查詢語言，請舉出一種針對XML資料的查詢語言？(20分)

答：

(一)學生(學號，姓名，地址，電話)

課程(課程代號，課程名稱，學分，老師代號)

老師(老師代號，姓名，職稱)

學生修課(學號，課目代號，成績)

(二)若合併為一個大表格，會造成資料正規化不足的異常現象。

例如：

- 1.新增異常：無法新增學生資料，因為必須同時新增課程資料
- 2.修改異常：多筆重複資料，造成資料更新時，有的改到、有的沒改到。
- 3.刪除異常：刪除一筆資料時，可能不小心一次刪除過多資訊，如刪除一門課程，卻同時刪除了所有只修這門課程的學生資料。

(三)記錄人名時，若採關聯式資料庫，可在各個表格單獨定義與設計名字資料格式，但可能導致系統定義較不一致，例如：客戶姓名、員工姓名在資料庫中有不同的長度定義，且客戶姓名為姓和名兩個欄位，員工姓名卻僅為一個欄位；物件導向資料庫則可將其包裝為一個可重複使用之物件，確保資料一致性。

(四)SQL的主要三種運算分別為：

- 1.SELECT：指定查詢所欲輸出的欄位。
- 2.FROM：指定查詢所牽涉到的關聯。
- 3.WHERE：指定查詢的條件。

(五)XML查詢語言XQuery(XML Query)，例如FLWOR，使用流程控制的方式從事XML查詢。FLWOR分別為查詢語法的五個關鍵字，For、Let、Where、Order by與Return，分別用以指定資料來源、指定變數、指定查詢條件、資料排序、以及回傳結果值。

