

# 統計

## 《資料處理》

### 試題評析

今年的資料處理題型範圍與去年雷同，完全著重於資料庫領域，並穿插物件導向的觀念。其中第一題的資料庫設計以往較少出現於本科目的命題中，第二題的EER是在ERD中擴充了物件觀念，第三題的SQL及relational algebra的對應屬基礎性觀念，第四題物件導向資料庫及關連式資料庫的差異，若細心作答應無問題。最後一題有關資料庫安全技術對於大多數的考生可能比較陌生。

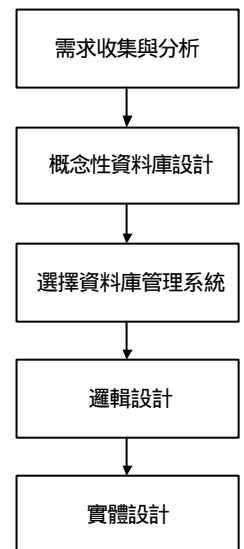
綜觀本份考題深淺適中，考生們平常如能留心近兩年的考題趨勢，掌握住資料庫及物件導向的領域，必能在本次考試中獲得佳績。預估一般考生分數約為60分，優秀考生可有80分。

### 一、請圖示說明資料庫設計的流程。(二十分)

答：

資料庫設計的流程，大致可以分為五個階段，如圖所示：

- (一)需求收集與分析(requirements collection & analysis)：包括辨識主要的應用程式與使用這些應用程式的使用者、對現有文件進行分析、分析工作環境與作業需求、進行問卷調查與訪談等。
- (二)概念性資料庫設計(conceptual database design)：本階段包括兩個步驟的行動，其一是概念性綱要設計，主要在檢查從第一階段得到的資料需求，以較高階的資料模式如ER模式或EER模式產生一個與DBMS無關的資料庫綱要(DBMS-independent conceptual database schema)，此外還可以改進此綱要的品質；其次就是交易設計(transaction design)，檢查從第一階段得到應用程式需求，產生較高階的交易描述，或者可以稱為功能綱要(functional schema)。
- (三)選擇資料庫管理系統：即是根據成本(包括軟體購置成本、維護成本、硬體購置成本、資料庫轉換成本、人事成本、訓練成本及操作成本)與需求(包括資料庫複雜度、應用程式共用程度、資料成長速度、資料多少及對控制的需求)綜合考量慎選適當的資料庫。
- (四)邏輯設計(logical design)：將第二階段所設計出來的高階概念性資料庫綱要對映成所選擇的資料模式(如關聯式資料模式..)，大致包括了與DBMS無關的對釋，及根據第三階段所選擇的資料庫管理系統對映出特別的資料模式，其設計結果就是邏輯綱要(logical schema)，包含ANSI/SPARC架構中的外層綱要(external schema)與概念綱要(conceptual schema)。
- (五)實體設計(physical design)：根據前所完成的資料庫綱要設實際的資料儲存架構，也就是ANSI/SPARC構中的內部綱要(internal schema)，而為提升整體的系統效率，此階段更應詳加考慮回應時間(response time)、空間使用率(space utilization)、異動產(transaction throughput)。



### 二、在實體 - 關係資料模型 (entity relationship data model) 中，如何描述實體 (entity) 和關係 (relationship) 這兩項基本觀念？在擴展式的實體 - 關係資料模型 (extended ER model, 簡稱EER模型)，加入特殊化 (specialization) 和一般化 (generalization)，請描述這兩種概念，並說明為何EER模型更能掌握真實世界的語意 (real world semantics)。(二十分)

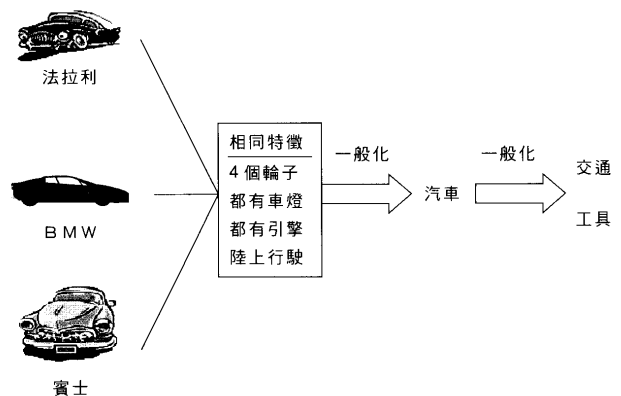
答：

(一)實體-關係資料模型(entity relationship data model)中：

1. 實體(entity)：是一些具有共同性質(properties)或特徵(characteristics)之實體案例(entity instance)或稱案例(instance)的集合。實體之種類很多，主要包括人、地方、物件、事件或使用者環境中之概念等，例如員工、客戶、訂單。
2. 關係(relationship)：指一個或多個實體間所發生的任何關聯(association)。關係的種類有1對1、1對多、多對多。

(二)在擴展式實體-關係資料模型(extended ER model)中：

1. 一般化(generalization)：指找出某一概念比另一概念更普遍的行為或特徵，亦即辨別一類別比另一類別更具一般性。例如：因「汽車」是「交通工具」，所以「交通工具」比「汽車」更具一般性。
2. 特殊化(specialization)：是一般化的相反，指辨別一類別比另一類別更具有專門性，它是建立子類別的原則。例如：「汽車」是「交通工具」，「汽車」比「交通工具」更具專門性，所以「汽車」是子類別，「交通工具」是父類別。



3.因可利用一般化(generalization)與特殊化(specialization)的原則，萃取相關子類別的相同屬性和操作，並將之歸納為一個父類別來實現繼承(inheritance)的方式，增加更多可用語意，讓使用者更容易將真實世界的資料模式化，所以EER更能掌握真實世界的語意。

三、一般化的SQL詢問 (query) 包括那六部分？那些是一定要有的？那些是可有可無的？關連式代數 (relational algebra) 中的三個主要運算； $\sigma$  (selection, 選擇)、 $\pi$  (projection, 投影)、和 $\times$  (Cartesian product, 笛卡爾乘積)，在SQL詢問中如何呈現？(二十分)

**答：**

(一)SQL中包括的六部份如下所示，其中一定要有者為SELECT、及FROM兩部份，可有可無者為[ ]所包含的WHERE、GROUP、HAVING、及ORDER BY四部份。

```
SELECT ...
FROM ...
[ WHERE ... ]
[ GROUP ... ]
[ HAVING ... ]
[ ORDER BY ... ]
```

(二)本小題中關連式代數(relational algebra)運算及SQL語法的對應如下：

關連式代數運算	SQL語法
selection	WHERE ...
projection	SELECT ...
Cartesian product	FROM ...

四、請由物件結構 (object structure)、運算包裝 (encapsulation) 和型態繼承 (type inheritance) 等三方面，說明物件導向資料庫 (object oriented databases) 的主要概念，同時也比較與關連式資料庫 (relational databases) 的差異。(二十分)

**答：**

(一)本題中物件導向資料庫(object-oriented database)的三種概念如下：

- 物件結構(object structure)：物件(object)指一種實體(entity)(如車子、椅子、人、筆、等)或一種抽象的概念(concept, 如密碼、任務、思想、意見、等)。物件結構(object structure)包含描述屬性的資訊結構及描述物件行為的操作(或稱方法、method)兩部分。
- 運算封裝(encapsulation)：指物件內部所使用的資料結構、實作的細節都要受到保護，對外界是隱藏的，不能由外界直接來存取，所有的動作都需要透過訊息傳送(message passing)到該物件，由該物件本身的方法來存取及運算後，再傳回結果給使用者。可防止使用者對物件狀態及其運算方法的誤用，與不合法的存取。
- 型態繼承(type inheritance)：指可從已存在有的父類別(base class)衍生出其他新的子類別(derived class)，子類別除了自己本身的屬性與運算方法之外，所有父類別或父類別更上層所繼承而來的所有屬性與運算方法，都會由子類別來繼承。

(二)物件導向資料庫(object-oriented database)及關連式資料庫(relational database)的差異如下：

- 物件導向資料庫(object-oriented database)是將整個資料庫以一組「抽象化資料型態」(abstract data types)來表示，這組抽象化資料型態各有其所屬的運算(operations)以及屬於其該型態中的「物件實例」(instances)。這些物件實例都得透過定義於其資料型態上的運算來存取或做適當處理。此外，資料型態之間有繼承(inheritance)、組合(aggregation)的關係，在物件及物件之間是透過訊息(message)來溝通彼此。
- 關連式資料庫(relational database)是利用一些被定義的集合，以表格的形式組成各種關聯(relation)，然後將這些關聯組合起來，利用資料庫表格當中某些欄位的關連性，建立起整個資料庫的大架構，便構成一個關聯式資料庫。

五、請回答以下與統計資料庫安全 (statistical database security) 相關問題。(二十分)

- (一)何謂統計詢問 (statistical queries) ?
- (二)試提出一種統計資料庫安全技術。

**答：**

- (一)統計詢問(statistical queries)：提供使用者利用原始資料來獲得統計性的摘要資訊，可管理與製作各式報表與圖表及統計分析，並結合Data Mining 與資料倉儲(Data Warehouse)之觀念，提供使用者較大的彈性於管理查詢與各項應用。
- (二)在資訊安全技術上，可對資料採用加密、解密及壓縮技巧。