

《資料處理》

試題評析

今年的資料處理考題全繞著資料庫的領域，並穿插物件導向的觀念。其中第一題的SQL語法以往較少出現於本科目的命題、第二題資料獨立性則屬基本觀念，第三題的實體關係圖則用來建立資料庫的概念資料模型，細心作答應無問題。至於第四題的資料庫挖掘(data mining)則見於去年普考考題，考生應已防範此命題，最後一題有關物件導向資料庫的名詞因屬較新的應用領域，所以大多數的考生比較陌生。綜觀本份考題對一般的考生較為困難，預估一般考生分數約為50分，優秀考生可有70分。

一、給予如下A公司關聯式資料庫 (relational database) 其中EMPLOYEE、DEPARTMENT和PROJECT分別代表員工、部門和計畫的資訊。

EMPLOYEE (ENAME, ID, SALARY, DNO)

ENAME：員工名字，ID：員工身分證字號，SALARY：員工薪資，DNO：員工所屬部門編號。

DEPARTMENT (DNAME, DNO, MGRID)

DNAME：部門名稱，DNO：部門編號，MGRID：部門經理身分證字號。

PROJECT (PNAME, PNO, DNO)

PNAME：計畫名稱，PNO：計畫編號，DNO：計畫歸屬部門。

請求SQL語言分別描述下列詢問 (query)。(每小題5分，共20)

- (一)列出D1這個部門所有員工的資料。
- (二)列出A公司所有部門名稱，以及各個部門員工人數。
- (三)列出D2這個部門所執行的所有計畫名稱。
- (四)列出P1這項計畫所屬部門的經理名字和薪資。

答：(一)SELECT *
FROM EMPLOYEE
WHERE DNO = 'D1';
(二)SELECT DNAME, COUNT(DISTINCT EMPLOYEE.ID)
FROM DEPARTMENT, EMPLOYEE
WHERE DEPARTMENT.DNO = EMPLOYEE.DNO
GROUP BY EMPLOYEE.DNO;
(三)SELECT PNAME
FROM PROJECT
WHERE DNO = 'D2';
(四)SELECT EMPLOYEE.ENAME, EMPLOYEE.SALARY
FROM EMPLOYEE, DEPARTMENT, PROJECT
WHERE PROJECT.PNO = 'P1'
AND PROJECT.DNO = DEPARTMENT.DNO
AND DEPARTMENT.MGRID = EMPLOYEE.ID;

二、請說明資料庫管理系統 (database management systems) 如何達到邏輯資料獨立性 (logical data independence)、實體資料獨立性 (physical data independence) 以及程式一運算獨立性 (program-operation independence)。(20分)

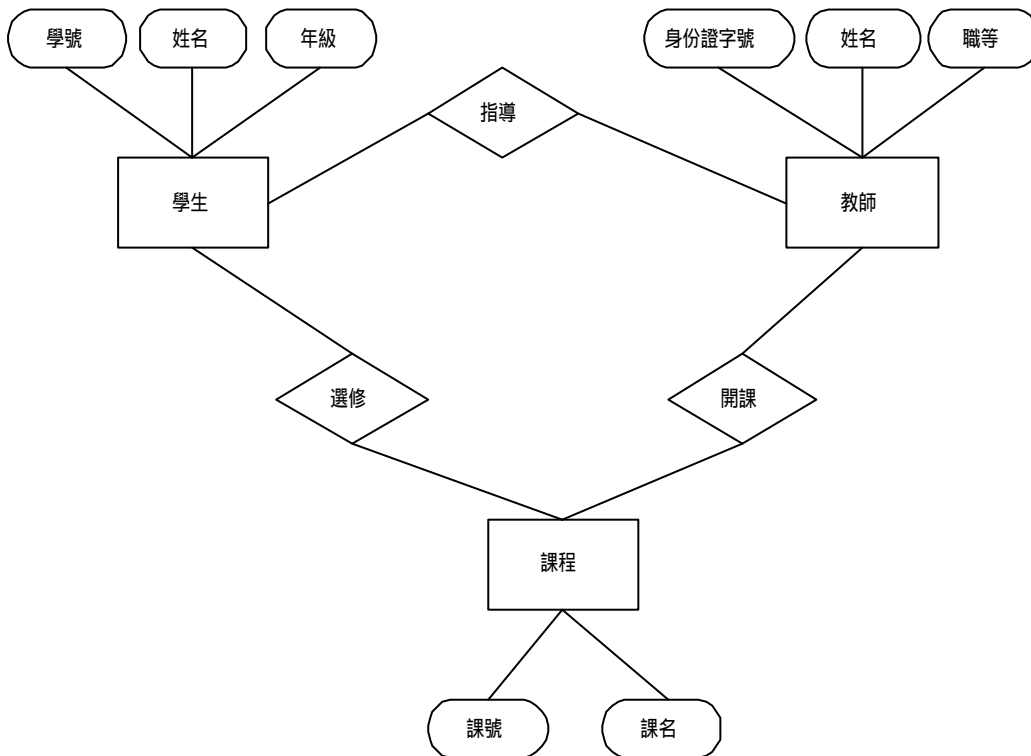
答：(一)邏輯資料獨立性(logical data independence)：指程式員或使用者可不必依賴特定的邏輯資料模式，而自由地發展其應用程式，而產生多種外部景觀(external view);也就是當資料格式有異動時，可透過資料庫管理系統(DBMS)來修改，而不必更動應用程式本身。概念層結構改變則只要改變概念層/外部層之間的映對(conceptual / external mapping)即可，外部層的綱要(external schema)可以維持不變。

- (二)實體資料獨立性(physical data independence)：指資料庫管理員需依賴任何一特定的儲存結構，就可建立其邏輯資料模式，亦即當資料庫組織(physical organization)改變時，資料庫管理員不需改寫應用程式。當內層結構改變時，只要改變內部層/概念層之間的映對(internal/ conceptual mapping)即可，概念層的綱要(conceptual schema)不需任何異動。
- (三)程式運算獨立性(program operation independence)：因採用三層次的系統架構，因此當資料庫內部的儲存結構、存取方式(access Method)等運算有所改變時，也不會影響到原來的應用程式。

三、請求實體—關係資料模型(entity-relationship data model)描述下列大學資料庫：某一所大學有大學部及研究所，學生資料有姓名、學號、年級等。教師資料有名、身分證字號、職級等。課程資料有課號、課名等。教師每學期至少開授兩門課，學生每學期至少選修三門課，同一門課只有一位教師開授，每門課至少十位同學選修。一位教師可以指導多個研究生，但一個研究生只有一位指導教授。請根據以上描述，回答下面的問題。(每小題10分，共20分)

- (一)指出此資料庫中實體(entity)部分，以及各個實體帶有的屬性。
- (二)指出關係(relationship)部分，並判斷是那一種關係：1對1、1對多或多對多。

答：實體關係圖如下



(一)存在三個實體(Entity)如下：實體名稱(屬性1、屬性2、...、屬性n)

- 1.學生(學號、姓名、年級、...、(其他))
- 2.教師(身分證字號、姓名、職級、...、(其他))
- 3.課程(課號、課名、...、(其他))

(二)如圖中所示，存在的關係有：

- 1.在教師與學生的「指導」關係上為1對多
- 2.在教師與課程的「開課」關係上為1對多
- 3.在學生與課程的「選修」關係上為多對多

四、請說明如何用統計的方法，由大量的資料庫中挖掘(mining)出有用的規則，並說明其應用。(20分)

答：(一)資料庫挖掘(data mining)方法

- 1.特性規則(classification rule)：使用者想了解在資料庫中是否隱含了此類族群的特色或特性，希望能經由資料挖掘的方式，幫助了解。
- 2.區別規則(discrimination rule)：使用者希望能經由資料挖掘了解多類集群之間的差異性，以做比較之用。本法可找出特性規則。
- 3.關連規則(association rule)：考量顧客於同一次交易所購買的多項產品間是否有特定購買模式存在。這些模式可能由於顧客消費習性，或由於產品間的關連性甚或其它一些外在誘因(例如：某些特定商品大特價)等因素而造成。

4.分群規則(clustering rule)：使用者希望能經由資料挖掘的方式，將資料庫中之資料分群，造成群內差異小而群間差異大，使之能了解各群資料之分佈情況與特性。

5.演進規則(evaluation rule)：使用者欲了解資料在一段時間記錄下之變化趨勢，例如投資基金之價格變化情形等。

(二)資料庫挖掘(data mining)應用

1.共通性分析：找出高附加價值顧客、帳款風險與信用卡犯罪的共通性。

2.業務趨勢分析：找出高於(或低於)平均成長率的市場。

3.標的市場：找出促銷活動的客戶對象。

4.使用分析：找出產品與服務的使用模式。

5.活動效益：比較活動策略的效益。

6.產品相關性：找出同時購買的產品，或購買某類產品的顧客特徵。

五、請說明下列與物件導向資料庫 (object-oriented databases) 有關的概念：(每小題10分，共20分)

(一)永續性物件 (persistent object) 和短暫性物件 (transient object) 。

(二)方法 (method) 和特徵標記 (signature) 。

答：(一)永續性物件(persistent object)指該物件被啟用後即一直存在記憶體中，且擁有固定的記憶體空間，在任何時刻均可經由物件位址指標而取得該物件。短暫性物件(transient object)：此種物件於啟用後，只暫時存在記憶體中，隨著應用程結束而消失，並釋出原記憶體空間。

(二)方法(method)指在物件導向中，當物件接獲訊息(message)後，會由物件執行的程序。特徵標記(signature)則指在物件導向中，每一物件均有其唯一的識別代號，可用以區分各個不同的物件。