

《資料庫應用》

試題評析

今年資料庫試題共有五題，份量稍稍偏重，但皆未脫離資料庫重點章節，所有內容亦皆包含在班內課程講義中，有好好準備的同學，各個題目應該都能回答得不錯。

第一題：ER Model幾乎每年都出現，對ER Model題型有多練習過的同學應該不成問題；

第二題：關聯式代數考古題出現頻率不是太頻繁，尤其第二小題的分群運算。總複習課程中有特別提到這個語法，如果有留心的同學應該記憶猶新。但若很少接觸關聯式代數的同學，可能就會回答得比較辛苦。

第三題：SQL語法雖然也每年都考，但SQL的查詢語法這幾年大多較為複雜，需要稍微思考才能回答正確。

第四題：主要關鍵在第一小題功能相依的簡化，如果第一小題回答正確，整題應該就不是問題；但若第一小題錯誤，整大題的分數便可能扣光光，要特別注意。

第五題：交易管理議題，較偏實務，需要思考一下再將課程中的理論拿進來套用。

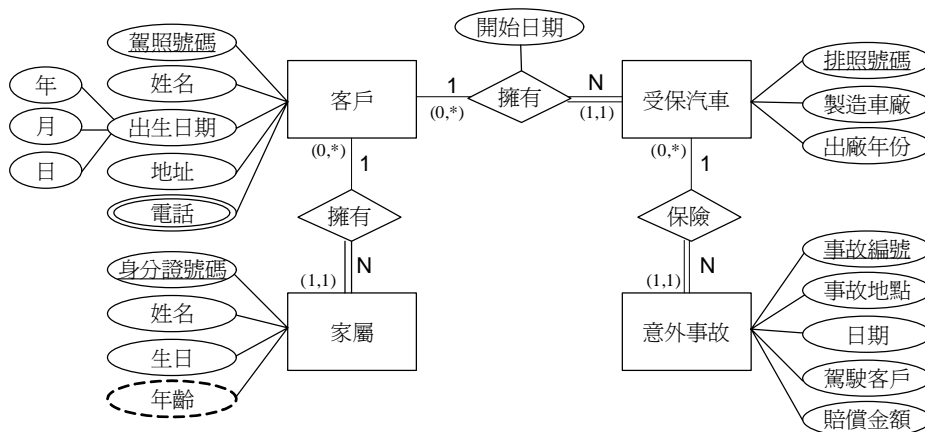
彙總上述說明，估計一般同學分數約在60~70分，程度較佳者可得到70~80分以上。

一、有一個汽車保險公司為他們的客戶和客戶的意外事故，要設計一個資料庫來記錄這些資料，已經得到的資料如下。

- 客戶資料，包括每個客戶的駕照號碼、姓名、出生日期、地址、電話，其中出生日期要分別記載年、月、日，每位客戶的電話號碼可能不只一個，受保客戶必須要有車子，否則不能購買保險。
- 汽車資料，包括受保的汽車牌照號碼、製造車廠、出廠年份。
- 每位客戶何時開始擁有受保的車子（開始日期），也要記錄下來
- 一位客戶可以擁有多輛受保的車子，但每輛車子只能隸屬於一位客戶。
- 每次意外事故要記載下來的內容，包括事故編號、事故地點、日期。
- 每次事故中，是由那位客戶駕駛，是那輛車發生事故，也要記載下來，事故的賠償金額也要記下。
- 因為事故有可能導致客戶喪生，賠償金此時要發給客戶的家屬。因此客戶的家屬資料也要記載下來，它包括家屬姓名、身分證號碼、年齡。
- 以上這些資料中，客戶的駕照號碼、汽車牌照號碼、事故編號、（客戶家屬的）身分證號碼，這些都是唯一的編號，也就是不會有二個相同的號碼。

請為這保險公司設計一個最佳的資料庫，用實體關係模型（Entity-Relationship model），把它設計出，並繪出其綱目圖（schema diagram），參與度（participation constraint）和基數比（cardinality ratio）亦須註明。（20分）

答：



- 假設：1.事故之車輛駕駛人，不一定為車主，故另由事故記錄的駕駛客戶欄位記錄事故時的駕駛人資料。
 2.系統中僅記錄家屬之生日，年齡則可由生日推導得到。
 3.客戶駕照號碼、家屬身分證號碼、受保汽車牌照號碼、意外事故編號皆為唯一不重複，故即使客戶不再投保亦可繼續記錄客戶及家屬資料。因此，四個個體皆為一般個體。

二、已知有一學校的資料庫綱目 (database schema) 如下：

Student (SID, SName, SAge, Address) //學生 (學號、姓名、年齡、地址)

Teacher (TID, TName, TAge, Office#) //老師 (老師編號、姓名、年齡、辦公室)

Course (CID, CTitle, Credit) //課程 (課程編號、課名、學分數)

Take (CID, TID, SID, Score) //修課 (課程編號、老師編號、學生編號、分數)

請用關連式代數 (relational algebra)，回答下列問題，關連 (relation) 和欄位 (attribute) 的名稱，請一律使用其英文名稱。

- (一)找出那些學生有修年齡是60歲 (含) 以上的老師 (TAge ≥ 60) 所教的任一門課程，列出這些學生的SName和教那些課程的老師的TName。(10分)
 (二)針對年齡是18歲 (含) 以下的學生 (i.e., SAge ≤ 18)，列出他們每個人的SID，和每個人所修所有課程的平均分數 (average score)。(10分)

答：

(一) $\pi_{SName, TName} (\sigma_{TAge \geq 60} (Teacher * Take * Student))$

(二) $SID \tau_{AVERAGE Score} (\sigma_{SAge \leq 18} (Student * Take))$

三、同上題的資料庫綱目 (database schema)，請用SQL來回答下列問題。

- (一)請找出有那些學生，他們修了TName="Harrison Ford" 這位老師所開設的所有課程，請列出這些學生的SName。(10分)
 (二)對於每一門學分數是3學分 (含) 以上 (i.e., Credit ≥ 3)，而且有50個 (含) 以上的學生來修的課程，列出這些課程的名稱、授課老師，和該門課程裡，學生成績超過95分 (含) 的人數，假設不會有二位老師開授同一門課程。(10分)

答：

(一)

```
SELECT S1.SName
FROM Student S
WHERE NOT EXISTS
  (SELECT *
   FROM Teacher TE, Take T1
   WHERE TE.TName='Harrison Ford' AND TE.TID=T1.TID
   AND NOT EXISTS
    (SELECT *
     FROM Take T2
     WHERE T2.SID=S.SID AND T2.CID=T1.CID))
```

(二)

```
SELECT CTitle, TName, COUNT(SID)
FROM Course C, Teacher TE, Take TA
WHERE C.CID=TA.CID AND TE.TID=TA.TID AND TA.Score >= 95 AND C.CID IN
  (SELECT C2.CID
   FROM Course C2, TAKE TA2
   WHERE C2.Credit >= 3 AND C2.CID=TA2.CID
   GROUP BY C2
   HAVING COUNT(SID) >= 50)
GROUP BY CTitle, TName
```

四、 $R(A, B, C, D, E)$ 是一個關連 (relation)，且它的欄位之間，存在有函數相依性 (functional dependency) 如下： $A \rightarrow B$ ， $B \rightarrow C$ ， $BC \rightarrow BD$ ， $D \rightarrow E$ ， $AB \rightarrow CD$ 。

(一)請將所給予的functional dependencies借由inference rules簡化到最簡形式。(5分)

(二)若僅將此關連作第二正規化 (second normal form)，會得到什麼關連？請註明每個關連的主鍵 (Key) 是何者？(5分)

(三)若將此關連作第三正規化 (third normal form)，會得到什麼關連？請註明每個關連的主鍵 (Key) 是何者？(10分)

答：

(一)簡化步驟：

1.每個功能相依僅有一個相依因素：

$A \rightarrow B$

$B \rightarrow C$

$BC \rightarrow BD$ 可分解為 $BC \rightarrow B$ 以及 $BC \rightarrow D$

$D \rightarrow E$

$AB \rightarrow CD$ 可分解為 $AB \rightarrow C$ 以及 $AB \rightarrow D$

整理後之功能相依： $A \rightarrow B$ ， $B \rightarrow C$ ， $BC \rightarrow B$ ， $BC \rightarrow D$ ， $D \rightarrow E$ ， $AB \rightarrow C$ ， $AB \rightarrow D$

2.去除多餘的決定因素：

$BC \rightarrow B$ 中，決定因素 B 即可決定 $B(B \rightarrow B)$ ，不需 C 屬性，且 $B \rightarrow B$ 為無價功能相依，故 $BC \rightarrow B$ 為多餘，可直接移除。

$BC \rightarrow D$ 中，由於 $B \rightarrow C$ 故 $B \rightarrow BC$ ，因此 $BC \rightarrow D$ 可簡化為 $B \rightarrow D$

$AB \rightarrow C$ 中，由於 $A \rightarrow B$ 故 $A \rightarrow AB$ ，因此 $AB \rightarrow C$ 可簡化為 $A \rightarrow C$ ，且 $AB \rightarrow D$ 可簡化為 $A \rightarrow D$

整理後之功能相依： $A \rightarrow B$ ， $B \rightarrow C$ ， $B \rightarrow D$ ， $D \rightarrow E$ ， $A \rightarrow C$ ， $A \rightarrow D$

3.去除多餘的功能相依：

$A \rightarrow B$ ， $B \rightarrow C$ ，則 $A \rightarrow C$ ，故 $A \rightarrow C$ 重複，可移除

$A \rightarrow B$ ， $B \rightarrow D$ ，則 $A \rightarrow D$ ，故 $A \rightarrow D$ 重複，可移除

故簡化後之最簡功能相依為 $A \rightarrow B$ ， $B \rightarrow C$ ， $B \rightarrow D$ ， $D \rightarrow E$

(二)由上述最簡功能相依得到，屬性 A 可決定所有非鍵屬性，為此關連唯一之候選鍵，即亦為主鍵。

該關連不存在任何部分功能相依於主鍵之屬性，故其已滿足第二正規化。

故2NF後之關連同原關連 $R(\underline{A}, B, C, D, E)$

(三)關連 R 中，

屬性 C, D 遞移相依於主鍵(透過 B)，屬性 E 亦遞移相依於主鍵(透過 B 及 D)，故3NF後關連如下：

$R1(\underline{A}, B)$ ，其中 $A \rightarrow B$

$R2(\underline{B}, C, D)$ ，其中 $B \rightarrow C$ ， $B \rightarrow D$

$R3(\underline{D}, E)$ ，其中 $D \rightarrow E$

五、回答下列問題：

(一)兩階段鎖定法 (Two-phase locking protocol) 是保證交易處理可序性

(Serializability) 的充分條件 (sufficient condition)，還是必要條件 (necessary condition)？為什麼？(10分)

(二)某銀行電腦系統於某日突然大當機。約數小時之後維修人員向總經理報告說『因為系統有軟體瑕疵 (software bugs)，以致於當機後在主記憶體 (main memory) 的資料都已喪失。為了重建當時主記憶體的資料，因此花了幾個小時』。依據你的專業判斷，所謂“重建當時主記憶體的資料”，指的是作了那些主要的動作，說明之。(10分)

答：

(一)兩階段鎖定，是保證交易處理可序列性的充分條件，非必要條件。

因為，兩階段鎖定可確保交易的序列性；但除此之外，亦有其他並行控制方式亦能確保交易的可序列性，例如：時間戳記同步控制法、多重版本同步控制法等。

(二)重建資料所需步驟，依照不同交易復原演算法有所不同，常見的便是延遲更新以及立即更新。

由於維修人員的說法為「重建當時主記憶體中的資料」，而非「自資料庫移除未完成之失敗交易的資料」，因此判斷該銀行採用者為延遲更新演算法。

但理論上，無論用何種演算法，主記憶體中所有重要更新操作，即系統當機後會影響復原動作者，皆應已寫入磁碟上的系統日誌，故「重建當時主記憶體的資料」應指重建(UNDO或REDO)系統日誌中的相關操作。

以延遲更新演算法為例，系統當機時，所有尚未Commit之交易一律不會寫入資料庫，故不需自資料庫移除任何不正確之資料(即不需UNDO)；但當機當時可能有部分已Commit之交易尚未被寫入至資料庫，故需至系統日誌中，找出當機當時已Commit之交易，將其寫入動作重新執行(REDO)，確認所有更新操作皆已反應至資料庫。待所有REDO操作皆已成功執行完畢後，才可重新啟動系統。

