

# 《資料庫應用》

一、某市政府希望透過網路來處理市民的反映問題，經過訪談和分析後，獲得以下需求：

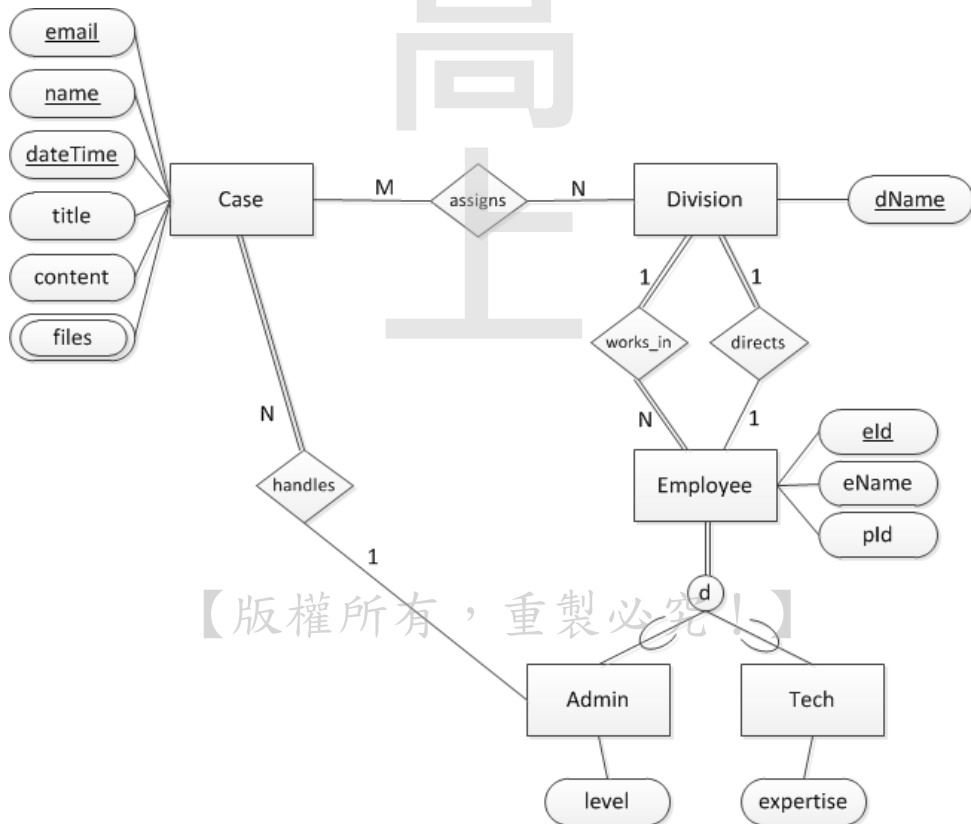
1. 陳情案 (Case)：必須記載陳情人的電子郵件信箱 (email) 和姓名 (name)、陳情時間 (dateTime)、主旨 (title)、內容 (content) 和附檔 (files)。其中陳情人的電子郵件信箱、姓名和陳情時間合起來為唯一，附檔可有多個。
2. 局處 (Division)：必須記載局處名稱 (dName) 和首長姓名 (director)。其中局處名稱為唯一，且每一陳情案可能會分配給數個相關局處來處理。
3. 員工 (Employee)：必須記載員工編號 (eId)、姓名 (eName) 和身分證字號 (pId)。其中員工編號和身分證字號均為唯一，且一位員工剛好屬於一個局處。員工一定是以下兩種身分中的一種：行政員工 (Admin) 和技術員工 (Tech)，行政員工要記載其職級 (level)，技術員工則要記載其技術能力 (expertise)。此外，一個陳情案會由一位行政員工來總負責 (Handles)。

(一)請依以上需求敘述，畫出其實體關係圖 (Entity Relationship Diagram)，員工的分類請使用繼承 (Inheritance) 來表示。(20分)

(二)請將(一)的實體關係圖轉成關聯網目並指出主鍵、外鍵，以及外鍵的參考關係。(10分)

試題評析	ER Model為一般資料庫常見題型。不過此題特別要求要以繼承方式作答，因此需採EER Model。EER Model轉為關聯有不只一種做法，下方參考解答為其中一種。
考點命中	《高點·高上資料庫講義》第三回，唐箏編撰，頁70-75、98-99。

答：  
(一)



假設：局處首長必為員工，且每一局處必至少有一局處首長。

(二)

Case (email, name, dateTime, title, content)

Casefiles (email, name, dateTime, files)

Division (dName, directorid)

Employee (eId, eName, pId, dName, employeetype, level, expertise)

註：

1. Casefiles的(email, name, dateTime)亦為外鍵，參考Case的(email, name, dateTime)。

2. Division的directorid為外鍵，參考Employee的eId欄位。

3. Employee的dName為外鍵，參考Division的dName。且employeetype為型態屬性，用以判斷該員工屬於Admin or Tech。例如：employeetype為A表Admin則level欄位有值expertise為空；反之employeetype為T表Tech則expertise欄位有值level為空。

二、考慮以下記載討論版文章的關聯：

R(發表人id, 發表人姓名, 發表時間, 主旨, 內容, 字數, 回應發表人id, 回應發表時間)

函數相依 (functional dependency) 以 $\rightarrow$ 表示，包含於 (is the subset of) 以 $\subseteq$ 表示：

發表人id $\rightarrow$ 發表人姓名

{發表人id, 發表時間} $\rightarrow$ {主旨, 內容, 回應發表人id, 回應發表時間}

內容 $\rightarrow$ 字數

{(回應發表人id, 回應發表時間)} $\subseteq$ {(發表人id, 發表時間)}

(一)請推論出以上關聯R的主鍵和外鍵，推導過程必須列出。(6分)

(二)請將關聯分解成第二正規式 (Second Normal Form, 2NF) 但不滿足第三正規式 (Third Normal Form, 3NF)，請指明主鍵、外鍵和其參考關係。(7分)

(三)請將關聯分解成第三正規式，請指明主鍵、外鍵和其參考關係。(7分)

試題評析	開放資料(Open Data)並不是一個新名詞，但在網際網路相當普及，網路上可取得資料越來越多，除了企業風行的Big Data外，世界各國由官方開始陸續推動Open data計劃。近幾年國內政府各單位也積極推動開放資料，雖然成效尚有相當成長空間，但開放資料儼然是電子化政府近年的新議題。但當然，開放資料由於是免費且容易取得的，維護資料的品質與個資法的把關也是在開放資料推行時責無旁貸的責任。
考點命中	《高點·高上資料庫講義》第二回，唐箏編撰，頁2-5、12-15。

答：

(一)推導主鍵如下：

發表人id $\rightarrow$ 發表人姓名，故{發表人id, 發表時間} $\rightarrow$ {發表人姓名, 發表時間}。

{發表人id, 發表時間} $\rightarrow$ 內容，且內容 $\rightarrow$ 字數，故{發表人id, 發表時間} $\rightarrow$ 字數。

{發表人id, 發表時間} $\rightarrow$ {發表人姓名, 發表時間}、{發表人id, 發表時間} $\rightarrow$ 字數、{發表人id, 發表時間} $\rightarrow$ {主旨, 內容, 回應發表人id, 回應發表時間}，故{發表人id, 發表時間} $\rightarrow$ {發表人姓名, 主旨, 內容, 字數, 回應發表人id, 回應發表時間}……(1)唯一性

此外，由於{發表人id, 發表時間} $\rightarrow$ {主旨, 內容, 回應發表人id, 回應發表時間}，故若{發表人id, 發表時間}缺少任一屬性則無法決定{主旨, 內容, 回應發表人id, 回應發表時間}……(2)最小性

由上述(1)(2)故候選鍵與主鍵為{發表人id, 發表時間}。

此外{(回應發表人id, 回應發表時間)} $\subseteq$ {(發表人id, 發表時間)}，故(回應發表人id, 回應發表時間)可作為外鍵參考同一關聯R的(發表人id, 發表時間)。

(二)發表人姓名部分功能相依於主鍵的發表人id欄位，故R非2NF。

2NF後：

R(發表人id, 發表時間, 主旨, 內容, 字數, 回應發表人id, 回應發表時間)

R1(發表人id, 發表人姓名)

其中R的發表人id為外鍵，參考R1的發表人id；R的(回應發表人id, 回應發表時間)為外鍵，參考R的(發表人id, 發表時間)。

(三)R中，字數欄位透過內容遞移相依於主鍵，故R非3NF；R1無任何遞移相依，為3NF。

3NF後：

R(發表人id, 發表時間, 主旨, 內容, 回應發表人id, 回應發表時間)

R1(發表人id, 發表人姓名)

R2(內容, 字數)

其中R的發表人id為外鍵，參考R1的發表人id；R的(回應發表人id, 回應發表時間)為外鍵，參考R的(發表人id, 發表時間)；R的內容為外鍵，參考R2的內容。

三、考慮以下記載討論版文章的關聯：

Member(mId, mName, email)

// 會員(編號, 姓名, 電子郵件信箱)

Post(pId, mId, dateTime, title, pContent)

// 文章(文章id, 作者編號, 日期時間, 主旨, 內容)

Reply(pId, rId, rId, dateTime, rContent)

// 回文(文章id, 回文id, 回文作者編號, 日期時間, 內容)

其中主鍵以底線表示，外鍵如下：

Post(mId)參考到Member(mId)

Reply(pId)參考到Post(pId)

Reply(rId)參考到Member(mId)

請用一個SQL查詢句(Queries)回答以下每一個查詢：

(一)列出2016年有PO過文章(即出現在Post關聯裡)的會員之會員編號和姓名。(5分)

(二)對於每一篇PO文，列出其pId, PO文時間, PO文作者姓名, 以及在2016年被回應的次數。(8分)

(三)列出有原作者回應的PO文之pId, title和作者姓名。(7分)

(四)找出PO文中(即出現在Post關聯裡)的內容有疑似西元日期的文章之pId, title, PO文作者姓名。所謂疑似西元日期是指出現連續四個數字, 且第一個數字為1或2, 或是(連續)數字前有西元或公元的字眼。(10分)

試題評析	今年同樣又出現SQL語法考題。其中第四小題的範圍模糊查詢語法平時使用機會較少，其餘幾題題目稍有變化但若時常練習SQL語法的考生小心作答應可應付。
考點命中	《高點·高上資料庫講義》第二回，唐箏編撰，第六章。

答：

(一)SELECT mId, mName

FROM Member

WHERE mId IN (SELECT mId FROM Post WHERE dateTime BETWEEN '2016/01/01' AND '2016/12/31')

(二)SELECT P.pId, P.dateTime, M.mName, R.cnt

FROM Post P, Member M, (SELECT pId, COUNT(\*) AS cnt FROM Reply WHERE dateTime BETWEEN

'2016/01/01' AND '2016/12/31' GROUP BY pId) R

WHERE P.mId=M.mId AND P.pId=R.pId

(三)SELECT P.pId, P.title, M.mName

FROM Post P, Member M

WHERE P.mId=M.mId AND EXISTS (SELECT \* FROM Reply R WHERE R.pId=P.pId AND R.rmId=P.mId)

(四)SELECT P.pId, P.title, M.mName

FROM Post P, Member M

WHERE P.mId=M.mId AND (pContent LIKE '%[1-2][0-9][0-9][0-9]%' OR pContent LIKE '%西元[0-9][0-9][0-9][0-9]%' OR pContent LIKE '%公元[0-9][0-9][0-9][0-9]%' )

四、MongoDB是目前使用愈來愈普遍的NoSQL資料庫管理系統，NoSQL的主要特色包括Schemaless, eventual consistency 和 horizontal scaling

(一)請說明Schemaless, eventual consistency 和 horizontal scaling的意義。(10分)

(二)假設你有一個order ( 訂單 ) 的collection，包含四個欄位：cust\_id ( 顧客id )，product ( 產品 )，amount ( 金額 )，type ( 產品型態 )，以下Mongo查詢句可用來找出衣服 ( clothing ) 類產品總購買金額超過10,000元的顧客之顧客id和衣服類產品購買總金額：

```
db.orders.aggregate([
  {$match:{type:"clothing"}}
  {
    $group:{
      _id:"cust_id",
      total:{$sum:"amount"}
    }
  }
  {$match:{total:{$gt:10000}}}
])
```

請仿以上MongoDB語法，找出購買金額超過1,000元的訂單中，各類產品的總購買金額。(10分)

試題評析	NoSQL考題不是第一年出現了，相信考生們已稍有準備。至於MongoDB語法，題目並不要求考生從無到有寫出，只要能看懂題目語法的邏輯，相信就有機會仿造出此題要求的語法。
考點命中	《高點·高上資料庫總複習講義》，唐箏編撰，頁20-22及補充筆記。

答：

(一)1.Schemaless：無綱要式。NoSQL不同於需要事先定義好資料欄位關係與結構的關聯式資料庫，NoSQL常需儲存結構化程度較低的資料內容，如網頁、圖片、影像等，且希望在資料量龐大時仍能保持一定的處理效率。故不採用關聯式資料庫預先定義schema的方式。而常採用如Key-Value、文件資料庫等方式。

2.Eventual consistency：最終一致性。不同於關聯式資料庫必須遵守的ACID，NoSQL採最終一致性，可允許資料有時間落差，非立即同步於所有分散點，但保證最終仍會同步於所有資料複本。

3.Horizontal scaling：水平擴充。只要增加新的伺服器節點，就可以不斷擴充資料庫系統的容量。而且可以利用低價的一般等級電腦就能進行水平擴充，不像關聯式資料庫的叢集系統往往需要效能和容量較大的伺服器才能勝任。

(二)以下找出購買金額超過1000元訂單中，各類產品的總購買金額(total)。

```
db.orders.aggregate([
  {$match : {amount : {$gt: 1000}}}
  {
    $group : {
      _type : "type",
      total : {$sum : "amount"}
    }
  }
])
```

【版權所有，重製必究！】