

《資料庫應用》

一、資料模型 (Data Models) 從階層式 (Hierarchical)、網路式 (Network)、關聯式 (Relational) 持續進展至最近的實體關係 (Entity-relationship)、物件導向 (Object-oriented)、與 NoSQL，請列表說明與比較後三種資料模型的優缺點。(20 分)

試題評析	「Big Data」，即巨量資料或稱海量資料，是這兩年來很熱門的話題；因而巨量資料的非關聯式儲存 NoSQL 也是考生們必須注意的。此題把實體關係模型、物件導向模型、以及 NoSQL 一起出題，只要三者觀念清楚，應可順利作答。
考點命中	1.《高點資料庫應用講義》第一回，唐箏編撰，第一章第四節。 2.《高點資料庫應用總複習講義》。

【擬答】

資料模型	優點	缺點
實體關係	<ul style="list-style-type: none"> 理論簡單、易於實行。 接近使用者角度，容易理解。 可達到資料獨立性，不針對特定資料庫設計。 普遍使用，可作為各資料庫概念層設計工具。 	<ul style="list-style-type: none"> 無法表達特定資料庫實作模式特性。 無法表達類別、特殊化/一般化等複雜關係，或非文字類型資料。 必須轉為特定實作資料模式才可實作。
物件導向	<ul style="list-style-type: none"> 支援非文字(text)資料，如：聲音、圖形、影像、時間序列等。 支援物件導向程式語言的許多特性，如：物件識別碼(Object Identifier)、繼承(inheritance)、封裝(encapsulation)、多型(polymorphism)、動態連結(dynamic binding)... 易與物件導向程式語言整合。 	<ul style="list-style-type: none"> 理論、標準尚未完全成熟。 實作產品不多，多為實驗性質。
NoSQL	<ul style="list-style-type: none"> 架構容易擴展/延伸。 對於巨量資料處理效率佳。 靈活的資料模型，可採 key-value 方式。 高可用性。 	<ul style="list-style-type: none"> 不需滿足 ACID，不適用對資料正確性要求非常高之應用。 不提供對 SQL 的支援。 支援的特性較關聯式資料庫弱，不夠豐富。 實作產品尚未為全成熟。

二、假設一所大學的資料庫包括下列相關關聯表 (Relation)：

名稱	綱要 (Schema) 與說明
學生 (STUDENT)	STUDENT(<u>S#</u> , S-NAME, DEGREE, T#, P#) 學號、姓名、學位 (大學部或研究所)、論文代號、專題代號
課程 (COURSE)	COURSE(<u>C#</u> , C-NAME, C-CREDIT) 課號、課名、學分數
註冊 (ENROLL)	ENROLL(<u>S#</u> , <u>C#</u> , SCORES) 學號、課號、成績

其中有底線的屬性 (Attribute) 代表主要鍵 (Primary Key)。

(一)請寫出下列查詢的 SQL 敘述 (Statement)：

- 學號 1123 學生的選修課名與成績 (顯示欄位：S#, S-NAME, C-NAME, SCORES)。(6 分)
- 學生修課平均成績前 5 名排行榜 (顯示欄位：S#, S-NAME, AVG(SCORES))。

- (二)假設 STUDENT 實體可再分為 GRADUATE (記錄論文資料) 與 UNDERGRAD (記錄專題資料) 兩大類, 請說明: (1)分離與重疊限制條件 (Disjoint and Overlapping Constraints), 及(2)特殊化與一般化 (Specialization and Generalization)。(4 分)
- (三)請說明擴充實體關係圖 (Extended Entity-Relationship Diagram; EERD) 與實體關係圖 (Entity-Relationship Diagram; ERD) 可能的主要差異。(4 分)

試題評析	此題兩個重點, 一為每年必考的 SQL 語法, 而另一則為 EER Model。今年考題第一、二兩題, 皆為 Data Model 考題, 也同時提及物件導向, 不過難度不高, 同樣考生們觀念清楚應可取分。
考點命中	《高點資料庫應用講義》第二回, 唐箏編撰, 第六章。 《高點資料庫應用講義》第三回, 唐箏編撰, 第九章第一節。

【擬答】

- (一)
- 1.SELECT S.S#, S-NAME, C-NAME, SCORES
FROM STUDENT S, COURSE C, ENROLL E
WHERE S.S#=E.S# AND C.C#=E.C# AND S.S#='1123'
 - 2.假設使用 SQL Server 語法, 以 TOP N 顯示前 N 筆資料
SELECT TOP 5 S.S#, S-NAME, AVG(SCORES)
FROM STUDENT S, ENROLL E
WHERE S.S#=E.S#
GROUP BY S.S#, S-NAME
ORDER BY AVG(SCORES)
- (二)
- 1.假設一學生不可同時就讀於大學部以及研究所, 則為「分離(disjoint)」。若學生可同時就讀大學部與研究所, 則為「重疊(overlapping)」。
 - 2.學生可「特殊化(specialization)」為大學部學生(UNDERGRAD)與研究所學生(GRADUATE)。反之, 大學部學生與研究所學生可「一般化(generalization)」為學生。
- (三)擴充實體關係圖, 可表達類別(超類別/子類別)的特殊化與一般化關係, 也可記錄該關係的資料分離與重疊。但若採一般實體關係圖, 僅可使用一張學生(STUDENT)資料表記錄學生資訊, 且一名學生僅能儲存一筆資料, 因此, 無法記錄若學生同時在某科系就讀大學部且在另一科系就讀研究所之資料。且論文代號與專題代號, 多僅一欄位需作使用。
使用擴充實體關係圖, 可依照資料(學生)類型儲存其所需屬性於該子類別, 使用較為彈性且能記錄之資訊較完整。
- 三、假設關聯表綱要為 R (A, B, C, D, E, F, G), 具有下列功能相依 (Functional Dependencies; FDs) 關係: $RD=\{AB\rightarrow CDEFG, A\rightarrow C, B\rightarrow DEF, E\rightarrow F\}$ 。
- (一)列舉 FD 內的部分相依 (Partial Dependencies) 與遞移相依 (Transitive Dependencies)。(5 分)
- (二)依序分解關聯以符合第二正規化型式 (2nd Normal Form) 與第三正規化型式 (3rd Normal Form)。(10 分)
- (三)分析最後結果是否符合 Boyce – Codd 正規化型式。(5 分)

試題評析	第一小題, 應將功能相依先作化簡, 否則容易混淆。正規化則為資料庫常見考題, 考生們應皆有所準備。
考點命中	《高點資料庫應用講義》第二回, 唐箏編撰, 第四章。

【擬答】

- (一)
- 因 $A\rightarrow C$ 故 $AB\rightarrow C$ 為多餘, 應可化簡為 $A\rightarrow C$;
同理 $B\rightarrow DEF$ 故 $AB\rightarrow DEF$ 為多餘, 應可化簡為 $B\rightarrow DEF$;

故題目中的功能相依去除多餘的決定因素後，可化簡為：

$A \rightarrow C, B \rightarrow DEF, AB \rightarrow G, E \rightarrow F$

此外， $B \rightarrow E$ ，且 $E \rightarrow F$ 可推導出 $B \rightarrow F$ ，因此 $B \rightarrow F$ 為多餘的決定關係。

故此題的最簡功能相依為

$A \rightarrow C, B \rightarrow DE, AB \rightarrow G, E \rightarrow F$

由上，可列舉部分相依為：(1) C 部分相依於 A, (2) DEF 部分相依於 B

遞移相依為：F 遞移相依(透過 E)於 B

(二)

承(一)，關聯存在部分相依，故非 2NF。分解為 2NF 關聯如下：

R1 (A, B, G)

R2 (A, C)

R3 (B, D, E, F)

以上 R1, R2, R3 滿足 2NF。

承上，R3 中，F 遞移相依於主鍵 B，因此 R3 非 3NF。

R1, R2 不存在遞移相依，因此 R1, R2 已滿足 3NF。

分解 R3 為 3NF 關聯如下：

R3 (B, D, E)

R4 (E, F)

故題目關聯分解至 3NF 共產生四關聯如下：

R1 (A, B, G)

R2 (A, C)

R3 (B, D, E)

R4 (E, F)

(三)

第(二)小題之結果，各關聯所有決定因素皆已為各關聯之後選鍵，因此 R1, R2, R3, R4 皆符合 Boyce-Codd 正規化型式。

四、在共同的資料庫上同時執行交易，可能會產生許多問題，而需要透過同步控制 (Concurrency Control) 機制來確保交易的循序性。請說明：

- (一)多重使用者同時使用資料庫可能導致的問題。(3分)
- (二)以時間戳記 (Time Stamping) 法執行同步控制的方式。(7分)
- (三)樂觀式 (Optimistic) 同步控制的運作方式。(7分)
- (四)分析前述兩種同步控制的差別與適用狀況。(3分)

試題評析	本題為「交易管理」考題，同步控制不佳可能產生的問題，以及各種同步控制機制與其比較，此題為基本名詞與觀念題，應可輕鬆取分。
考點命中	《高點資料庫應用講義》第三回，唐箏編撰，第七章第二節。

【擬答】

- (一)多重使用者同時使用資料庫，若同步控制不佳，可能導致「遺失更新(Lost Update)」、「不一致分析(Inconsistent analysis / Incorrect summary)」、「未委任相依(Uncommitted dependency / Dirty reads / Temporary update)」等問題。
- (二)時間戳記(Timestamp)：每個交易依照時間先後給予其一時間戳記(Timestamp)，一但有交易欲存取某資料項目時，必須更新存取此資料項目之時間戳記；即可依照時間戳記的大小判斷交易之優先權(priority)。
- (三)樂觀同步控制(Optimistic concurrency control)：以一樂觀的想法假設所有交易皆可順利而正確地進行，不須在交易期間作任何檢查；直到交易完成才從事檢查。若交易違反可序列化(serializable)，則選擇影響較小之交易作撤回(abort)。
- (四)兩者主要差別在：時間戳記法，需在每次資料讀寫前，先檢查該操作是否可能違反同步控制限制，導致資料不一致，若有，則直接選擇交易作撤回，解決該同步控制問題。反之，樂觀同步控制，則在交易過程中不作

檢查，直到交易完成前再一次檢查。因此，樂觀同步控制適用於交易甚少發生不一致的狀況下，可提升執行效率，而時間戳記則適用於交易不一致機率較高之狀況。

五、假設您是 B2C (Business to Customer) 電子商務的經營業者，您有意善用資料庫資源，改善顧客滿意度與忠誠度，進而創造企業的持續性競爭優勢。在此背景下，請說明：

- (一)商業智慧 (Business Intelligence; BI)。(5 分)
- (二)資料倉儲 (Data Warehouse) 與其特色。(5 分)
- (三)巨量資料分析 (Big Data Analytics)。(5 分)
- (四)分析三者的關係與重要啟示。(5 分)

試題評析	此題前三小題考題令考生較無頭緒，題目沒描述清楚，希望考生"說明"「名詞定義」、「應用」、或其他。因此，擬答就稍微將定義與應用皆作些描述。
考點命中	《高點資料庫應用講義》第三回，唐箏編撰，第九章第三節。 《高點資料庫應用總複習講義》。

【擬答】

- (一)商業智慧：狹義的 BI 指利用 OLAP 與 Data Mining 擷取企業內部資料庫和資料倉儲的資料，快速提供決策所需的資料；廣義的 BI 指透過內部各種資訊系統，經過彙整、分類、統計，用以提供高階主管內部經營決策相關的資訊、知識、或關鍵績效指標(KPI)。商業智慧需要廣泛的資料來源作為分析基礎，例如：分析客戶、產品特性，以提升客戶滿意度。
- (二)資料倉儲：一個具有主題導向、整合性、時間變動性、資料不可揮發性特性的資料集合，用以支援管理決策。資料倉儲的資料可以自不同功能與性質的資料庫中取得，例如：銷售資料庫、客戶資料庫等。電子商務網站可利用資料倉儲，儲存與客戶相關分析所需資訊，並透過分析工具，以提升客戶滿意度與忠誠度。
- (三)巨量資料分析：相較於傳統資料庫或資料倉儲，巨量資料分析更為強調非結構化或是半結構化資料的分析，例如：Facebook、Twitter 的留言，此類資訊隨時間增加速度非常快，甚至超越一般資料庫能負載的程度。電子商務網站，可透過巨量資料分析，了解/分析市場資訊、趨勢、客戶對公司的評價等。
- (四)商業智慧，是電子商務經營業者希望洞悉市場取得的資訊目標；資料倉儲，是可用以取得商業智慧的資料來源；巨量資料分析，則是取得商業智慧使用的分析工具之一。
巨量資料的儲存、處理和分析，主要目的之一就是從中萃取出「商業智慧」：有組織、有系統地分析、彙整儲存於企業內外部的資料，從中汲取關鍵的概念、機制與活動，創造出有助於商業上各種決策的知識與洞見。其資料來源可能來自結構化的資料庫或資料倉儲，也可能來自儲存非結構化、巨量資料的儲存媒介。隨著資訊科技的進步、社群網站等資訊的快速累積，給我們的啟示是，企業經營者，已漸漸無法僅看企業內部資料庫儲存的資訊作決策，而更應著重企業外部廣泛而龐大的資料，將巨量資料化為商業智慧，進而嗅出趨勢、預測未來、創造商機。

【版權所有，重製必究！】