

《資料庫應用》

- 一、在資料模型中，限制（Constraints）是資料庫狀態（Database States）任何時刻必須滿足的條件，請論述實體關聯模型（ER Models）有關實體間對應的兩個限制及關聯模型（Relational Models）的三個限制。另外，關聯資料庫模型如何表示上述實體關聯模型的兩個限制？（20分）（25分）

試題評析 名詞解釋背誦題。

答：

（一）

基數比：用來表示一個實體最多可以跟幾個對方實體發生此類關係。

參與度：用來表示一個實體最少可以跟幾個對方實體產生此類關係。

（二）

弱實體型態(Weak entity type)：關鍵屬性必須依賴某種關係型態的實體型態。

識別關係型態(Identifying relationship type)：所依賴的關係型態。

遞迴關係型態(Recursive relationship type)：有些二元關係型態的左右兩邊實體型態卻指向同一個，因此衍生遞迴關係型態。

（三）

基數比可以分成 1:1, 1:N, M:N

參與度則可分為完全參與(ERD 用實線表示)、部分參與(ERD 用單線表示)

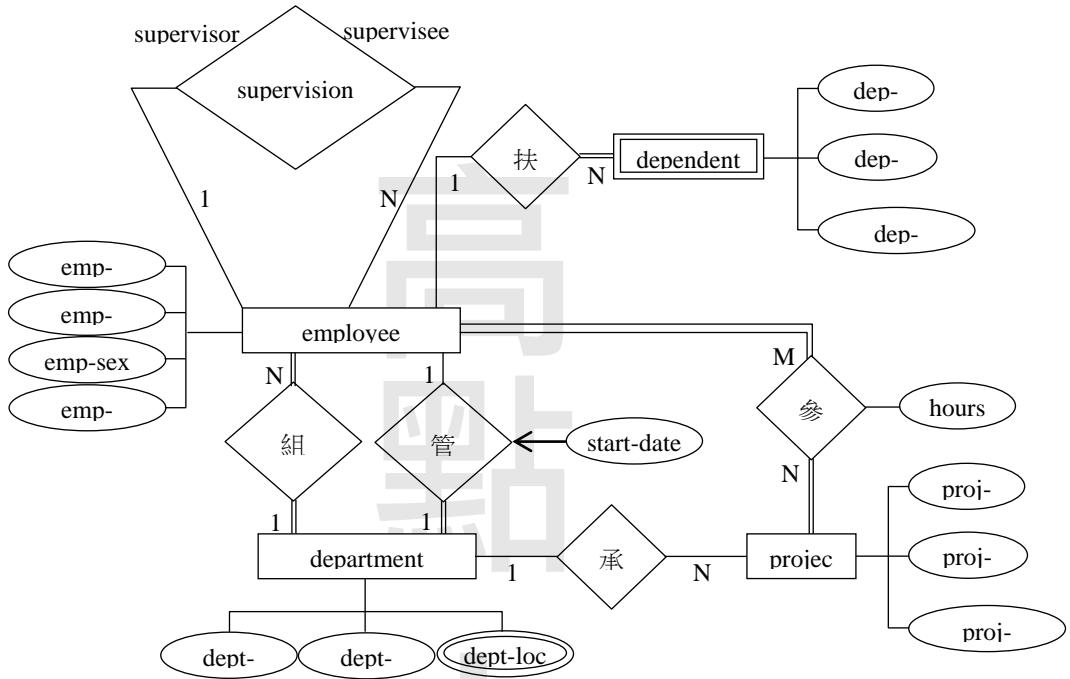
- 二、一家公司Company的資料，擬設計一資料庫來儲存，經系統分析後，得到如下需求：

- 1.Company 公司由許多部門（department）組成，每一部門有一部門代碼（dept-id）、一部門名稱（dept-name），一部門會有多個地點（dept-loc），每一部門有一員工擔任主管，並記錄開始擔任主管日期（start_date），部門代碼可區別不同部門；一個部門可以承接許多計畫（project），但一計畫僅由一部門承接，每一計畫有計畫代碼（proj-id）、名稱（proj-name）及計畫說明（proj-descript），計畫代碼可區別不同計畫。
- 2.一部門由許多員工（employee）組成，每一員工有員工代碼（emp-id）、員工姓名（emp-name）、員工性別（emp-sex）與員工生日（emp-birthdate），每一員工隸屬於一部門，每一員工代碼可區別不同員工。員工會互相監督（supervision），每位員工最多有一位直屬監督者，但可能會直接監督多位其他員工。
- 3.一員工可以扶養多位眷屬（dependent），每一眷屬有眷屬代碼（dep-id）、眷屬姓名（dep-name）與眷屬生日（dep-birthdate），但一眷屬只能由一員工來扶養。
- 4.一員工可以參與許多計畫，一計畫也可許多員工來參與，每一員工參與的計畫有一參與的時數（hours）。

請劃出Company公司實體關聯網要（Entity-Relationship Schema），包括屬性（Attributes）、實體型態、關聯型態與相關限制。（20分）

試題評析 常見繪圖必考題，基本原則不外乎是看清楚限制條件和釐清資料屬性及資料間關係。。

答：



三、在資料庫的實作及環境中，同步控制（Concurrency Control）在交易（Transactions）處理是很重要的技術，請論述兩階段鎖定（Two-phase Locking）與時戳（Timestamps）技術及其不同處。（20分）

試題評析 基本的資料庫同步控制觀念題，考生把握基本概念答題應不是問題。

答：

(一)兩階段鎖定：

當某交易欲存取特定資料項目時，必須將此資料鎖住(Lock)，直到存取完畢才解除鎖定(Unlock)，若其它交易欲存取被鎖定的資料項目時，需等待至其解除鎖定。

(二)時間戳記：

每個交易依其進入資料庫系統的時間先後，給予一個時間戳記，以時間戳記的大小來判別交易之優先權(Priority)。若有交易欲存取某資料項目時，必須更新存取此資料項目的時間戳記。

四、現在的臉書與line有大量的資料，如應用人工智慧來發掘（Discover）相關關鍵詞，以便了解親友對產品的偏好，其中需要很大量資料來當作訓練與測試資料，以訓練出好的學習模型（Learning Models），請寫出資料庫設計過程（Database Design Process），設計資料庫以儲存臉書與line的大量資料（特別留意臉書與line資料庫特性），並論述每一過程（20分）

試題評析

結合現代化機器學習技術的時事考題，考生須了解基本機器學習特性以及資料庫對應之部署方式方可有效答題。

答：

資料庫設計主要可以分成以下幾個步驟：

1.Requirements specification and analysis

分析需求：一般的機器學習、深度學習中儲存資料、分析資料方式可以概略理解成預處理、資料探勘、模型訓練、套用模型預測結果等過程。

資料庫中則主要存放預處理前的大量Raw data，因此必須將資料間關係和類別做一個有效的規劃才能節省預處理、資料探勘的時間。

2. Conceptual design

將資料間的關係用結構化的方式呈現，建立資料庫邏輯的模型，歸類出要分析的資料特性，例如：喜好的产品是屬於哪種類別、欲分析的使用者是哪個年齡層等，並設計出對應的欄位以存放資料，才能讓後續預處理、資料探勘過程比較省時省力，此處常用ER-model等方式表示。

3. Logical design

透過 Relational model等方式，最後生成定義關聯網目(schema)主要由關聯表(Relation)，完整性條件限制(Integrity constrains)，關聯式代數等元素組成。

4. Physical design

在資料庫管理系統(例如：MySQL)上建立實際的資料庫結構即實作前述Conceptual design和Logical design之內容。

五、在關聯式資料庫模型，何謂關聯代數 (Relational Algebra)？為何需要關聯代數？何謂關聯完全 (Relationally Complete)？又商業化的資料庫管理系統會用到的資料庫查詢語言是什麼？與關聯代數有何關係？資料庫處理器 (Run-time Database Processor) 如何處理此查詢語言？(20分)

試題評析	資料庫理論名詞解釋題，屬於有背就有分的題型。
-------------	------------------------

答：

1.關聯代數：

是一種較低階的、程序性的、規範性之抽象的查詢語言，它是來描述如何產生查詢結果的步驟。我們可以想像成「演算法」，亦即描述解決問題的步驟。目前主流資料庫語言預作皆是以關聯代數作為理論基礎而建立起來的。

2.關聯完全：

關聯式模式的兩種運算模式：

關聯式代數

關聯式計算

任何一套資料庫查詢語言具備以上任一能力，皆可稱為關聯完全。

3.商業化常用的資料庫查詢語言一般是以 SQL 標準為基底建立的。例如：MySQL 就是一個以 SQL 為標準擴充的資料庫系統語言，其理論基礎便是關聯代數。

4.資料庫處理器：

optimizer 接收高階語言之後，會轉為較低階的語法並傳遞給資料庫處理器解析並執行。

【版權所有，重製必究！】