

資料庫管理（114-1）

作業三

作業設計：孔令傑

國立臺灣大學資訊管理學系

繳交作業時，請至 NTU COOL 下載本作業題目的地方上傳一個 PDF 檔。在生成這個 PDF 檔時，可以用打字的也可以用手寫的，但不管怎樣，請務必注意繳交的文件的专业程度（通常透過排版、文字圖片表格方程式的清晰程度、用字遣詞等面向呈現），如果專業程度不夠，也會被酌量扣分。每位學生都要上傳自己寫的解答。不接受紙本繳交；不接受遲交。可以用中文或英文作答。這份作業的截止時間是 **10 月 15 日早上 08:00:00**，遲交在 12 小時內者扣 10 分、在 12 到 24 小時內者扣 20 分、超過 24 小時的這份作業將得不到分數。

相關規定與提醒

1. 關於上網查詢與 AI 工具：

任何一份作業都可以被用任何方式完成，包括上網搜尋和使用各種 AI 工具。如果你想用，請留意以下幾件事。首先，抄襲還是不被允許的，如果我們發現抄襲（包括抄襲網路上的答案，或是抄襲同學的答案），都還是會給予嚴厲的懲罰（視情節輕重而定，通常是該份作業算零分，或者不予通過這門課）。只要沒有抄襲，那我們就只根據你交上來的答案的品質給分，不論你是自己想出來的，還是有利用 AI 工具。如果某甲善用了 AI 工具後寫得很好，某乙自己努力寫但寫得不好，那某甲會得到比較高分。其次，如大家所知，AI 工具給的答案可能會錯，也可能不合適。使用 AI 工具是學生的自由，但確認 AI 工具的答案是否合適、是否需要調整則是學生的任務。請務必自行確認答案的正確性與合適性。最後還是提醒大家，學到多少東西都是自己的，如果一時困難用 AI 工具度過難關那是無妨，但之後建議還是花時間把東西學起來，對自己比較好。

2. 關於「專業」：

在我們這門課及許多課程中，都需要繳交各種報告。一份報告如果要達成他的效果（例如成功募資、完成溝通的任務等等），除了需要好的內容，也需要「專業」，而顯得專業通常需要「長得好看」以及「看起來用心」，這在沒有標準答案的任務上更是如此。有鑑於此，在這門課的作業和專案，我們都會要求報告的格式和美觀，並且納入評分標準。為此，我們提供報告格式參考指南「DB_reportFormatGuideline.pdf」，上面列舉了一份格式良好的報告的最低標

準。在作業一我們會請助教就違反參考指南的地方標出來讓大家知道（我們理解那份指南並不是最完美的，但如果完全沒有標準，同學們容易無所適從，所以我們還是設計一份當標準），但不會扣分，只是提醒大家；從作業二起就會有部分分數是報告專業度分數。之所以要求這些不是想要找大家麻煩，而是大家離出社會也不是太遠了，確實應該要開始被要求報告的可讀性和易讀性，所以我們願意花一些時間要求大家，但不會刁難大家，也請大家理解和盡力嘗試了。

3. 關於「批改」：

如課程大綱所述，為了不要累死助教，每次作業可能只有部分題目會被批改和給予回饋，但每一題的參考解答都會在作業截止後公佈。如果有一題沒被批改，那所有有寫那一題的學生都會得到那一題的全部分數，但沒寫或期限前沒交作業的自然就不會拿到那一題的分數。最後，請注意是「可能」，換言之也有可能是所有題目都被批改。

題目

在本次作業中，有 10 分是依照大家交上來的報告的專業度給分（換言之，格式和美觀再怎麼糟糕，最多也只會扣 10 分）。其餘分數則按照以下題目的正確性和合適性給分。

1. (30 分；一題 5 分) 針對上課使用的 `RETAIL_ADV` 這個資料庫，請幫以下任務撰寫合適的 SQL 指令，每題的答案必須是恰好一句 SQL 指令（可以用 CTE、subquery），並附上執行該指令後得到的查詢結果。
 - (a) 列出所有沒有下屬的員工，列出其員工編號和姓名。
 - (b) 針對公司裡的所有員工，回傳其員工編號和 2025/1/5 所屬的分店編號。
 - (c) 針對公司裡的所有員工，回傳其員工編號、加入公司時的分店編號，以及第二個所屬之分店（如果有）的編號。如果有員工在加入公司後就沒有換過分店，第三個欄位就填 NULL。
 - (d) 將所有商品按照有史以來的被採購總數量由大到小排序，如果有商品的被採購總數量平手，則把商品編號小的排前面，然後取出排第四與第五的商品，列出其商品編號、商品名稱，以及這些商品個別在每間分店的被採購次數與被採購總數量。如果這些商品中有某個不曾被某間分店購買過，在搜尋結果中就不要列出該商品和該分店的資料列。

提示：LIMIT 和 OFFSET 可能蠻有用的。

- (e) 同上題，但如果這些商品中有某個不曾被某間分店購買過，在搜尋結果中仍應該被列出，而其被採購次數與被採購總數量應為零。

提示：CROSS JOIN 和 COALESCE 可能蠻有用的。

- (f) 在每間分店幫所有會員計算他們在該店的總消費金額及排名，其中排名按照所有會員在該店的總消費金額由大到小排序。查詢結果應包含四個欄位，依序是分店編號、會員編號、該會員在該分店的總消費金額，以及該會員在該分店的總消費金額由大到小的排名。如果有複數個會員在一間分店的總消費金額相同，就讓他們有一樣的排名。

提示：RANK 和 WINDOW 可能蠻有用的。

2. (20 分；每小題 10 分) 針對課堂上使用的 ONLINE_LEARNING 資料庫，請在某種 general-purpose 程式語言（例如 Python）中讀取這四張表的資料，執行以下分析任務。

- (a) 請針對有被答題至少 1000 次且回答正確次數不大於 500 次的難題，找出在該題有回答正確且耗時最少的前三位學員；若有複數位學員答題時間相同，則以最先回答者優先。若一位學員對同個題目答題數次，以其首次回答正確所花的答題時間為準（因此每個難題的排行榜前三名不可能為同一人）。最後，請計算每位學員出現在這個「難題速解排行榜」上的次數，挑出有在至少兩題上榜的學員，以一張長條圖呈現這些學員的上榜題目數。

請複製你的完整程式碼（以在 Python 上用 DuckDB 為例，要包含匯入套件、讀取資料、對 DuckDB 的 connection 下 SQL、展示結果等等）到你的解答中，並且貼上你的長條圖。

- (b) 承上題，請針對這個「難題速解排行榜」，去計算每位榜上有名的學員的排行榜積分，並將學員依照排行榜積分由大到小排序，若遇到同分者則以學員編號小者排前面。積分算法是每位學員每登上某題的第一名可得 5 分、第二名 2 分、第三名 1 分。舉例來說，若某位學員分別在題目 1 和題目 2 獲得第一名和第三名，則其排行榜積分為 $4 + 1 = 5$ 。最後，請挑出積分至少 5 分以上的學員，以一張長條圖呈現這些學員的排行榜積分。

請複製你的完整程式碼（以在 Python 上用 DuckDB 為例，要包含匯入套件、讀取資料、對 DuckDB 的 connection 下 SQL、展示結果等等）到你的解答中，但跟第一小題重複的部份請不要繳交，並且貼上你的長條圖。

3. (20 分) 某租車公司在全國各地設有許多站點 (STATION)，把許多車輛 (VEHICLE) 投放在站點供會員 (MEMBER) 租借。由於會員租借車輛的需求不一定平衡，公司有許多員工 (EMPLOYEE) 負責在站點之間調度車輛（把一輛車從站點 A 開到站點

B，以便在站點 B 有預約的會員可以順利取車；就跟 YouBike 需要有人調度腳踏車類似，只是 YouBike 不用預約而已）。

- 每個站點會有其編號（`s_id`，長度最多為 10 的字串）、名稱（`s_name`，長度最多為 20 的字串）、郵遞區號（`postal_code`，恰好三碼的數字但本質上是字串）、地址（`address`，長度最多為 100 的字串）、狀態（`m_status`，一個字元，A 表示營運中，I 表示非營運中）；
- 每輛車會有其車牌號碼（`v_id`，長度最多為 10 的字串）、製造商編號（`manufacturer_name`，長度最多為 20 的字串，例如「Toyota」）、型號（`model_name`，長度最多為 20 的字串，例如「Camry」）、製造日期（`make_date`，包含年月日的日期）、狀態（`v_status`，一個字元，A 表示可供租借、M 表示正被維修保養中、R 表示正被租借中、I 表示已離開車隊；除了 A 以外的狀態都是不可被租借）。
- 每個員工會有其編號（`e_id`，整數）、姓名（`e_name`，長度最多為 20 的字串）、出生日期（`birthdate`，包含年月日的日期）、狀態（`e_status`，一個字元，A 表示在職中、L 表示已離職）。
- 每個會員會有其編號（`m_id`，長度最多為 10 的字串）、姓名（`m_name`，長度最多為 20 的字串）、加入會員日期（`m_date`，包含年月日的日期）、狀態（`m_status`，一個字元，A 表示會員資格生效中、I 表示會員資格已失效）。

根據以上說明，請回答以下問題。

- (a) 該公司也要在資料庫中留存車輛調度的記錄，意即記錄針對每次調度，記錄是哪個員工從哪個站點把哪一輛車移動到哪個站點，以及相對應的出發及抵達時間。在調度車輛時，有時會有複數個員工一起做同一趟調度，此時所有參與的員工都應該被記錄下來。

你現在要幫公司在資料庫中設計相對應的資料表，讓公司可以詳實地留存這些記錄：

- (5 分) 假設你要建立一張資料表，其中每一筆資料代表一次調度紀錄（當然，視情況而定你可能會建立更多其它資料表），請問你在設定 primary key 時，你會用 Surrogate Key 還是 Natural Key？如果是 Surrogate Key，你會用流水號還是 UUID？如果是 Natural Key，你會用哪些（個）欄位？請說明你的作法以及你那麼做的原因。
- (5 分) 你會不會幫儲存調度紀錄的資料表設計欄位以註記一筆資料是否「已被刪除」？為什麼？

iii. (5 分) 請設計一個關聯式資料庫綱要，去記錄站點、車輛、員工、會員與調度紀錄，並且以 relational schema diagram 的方式呈現。在你的圖中不需要呈現資料型態資訊，只要如往常一樣展示 table、column、primary key、foreign key 等即可。你可以假設對應到站點、車輛、員工、會員的資料表都使用 natural key，但你得自己選擇合適的欄位；對於儲存調度紀錄的資料表，請按照你在第 3(a)i 和 3(a)ii 小題的設計理念規劃欄位。你的設計必須滿足 1NF、2NF、3NF、BCNF 和 4NF 的條件，但你不用說明你的設計是如何滿足這些正規化條件的。

(b) (5 分) 承上題，該公司也要在資料庫中留存會員租車的記錄，意即針對每筆交易，記錄是哪個會員在何時從哪個站點把哪輛車取走，又在何時到哪個站點把該輛車還了。在會員租車時，也可能會有複數個會員一起租車（意即在這次租賃中這些會員都可以駕車），但會有恰好一個會是主要租車的交易對象。一個會員可以在一筆交易中租複數輛車。

你現在要幫公司在資料庫中設計相對應的資料表，讓公司可以詳實地留存這些記錄。請設計一個關聯式資料庫綱要，去記錄站點、車輛、員工、會員、調度紀錄與交易紀錄，並且以 relational schema diagram 的方式呈現。在你的圖中不需要呈現資料型態資訊，只要如往常一樣展示 table、column、primary key、foreign key 等即可。你必須呈現一個完整的 schema diagram，但當然你可以複製貼上你在第 3(a)iii 小題的 schema diagram 再做擴充。你可以假設對應到站點、車輛、員工、會員的資料表都使用 natural key，但你得自己選擇合適的欄位。請直接決定你用來儲存交易紀錄的資料表要使用 Surrogate Key 還是 Natural Key，以及是否幫儲存交易紀錄的資料表設計欄位以註記一筆資料是否「已被刪除」，不用附上說明，但如果你的設計顯得不夠合理，還是會被酌予扣分。你的設計必須滿足 1NF、2NF、3NF、BCNF 和 4NF，但你不用說明你的設計是如何滿足這些正規化條件的。

4. (20 分；每小題 10 分) 承第 3 題，請針對以下兩小題獨立回答：

(a) 假設你朋友設計了以下一張資料表去儲存調度紀錄

<pre>RELOCATION(v_id, from_datetime, from_s_id, to_datetime, to_s_id, e_ids, make_date)</pre>

其中前兩個欄位一起做為 primary key。如果有複數個員工一起調度同一輛車，就在 e_ids 欄位把這些員工的編號用逗點隔開。

請問這張資料表的欄位設計違反了 1NF、2NF、3NF、BCNF 和 4NF 中的哪個或哪些，還是都不違反？請說明你的理由。

(b) 假設你朋友設計了以下兩張資料表去儲存調度紀錄

```
RELOCATION(relocation_id, v_id, from_datetime,  
          from_s_id, to_datetime, to_s_id, make_date)
```

```
RELOCATION_EMPLOYEE(relocation_id, e_id)
```

在 RELOCATION 表中，relocation_id 是流水號性質的 primary key。在 RELOCATION_EMPLOYEE 表中，兩個欄位一起做為 primary key，如果有複數個員工一起調度同一輛車，就在 RELOCATION_EMPLOYEE 中新增數列資料。

請問這兩張資料表的欄位設計違反了 1NF、2NF、3NF、BCNF 和 4NF 中的哪個或哪些，還是都不違反？請說明你的理由。