

DBMS 2023 Homework 4

- 繳交期限：04/16 23:59
- 繳交平台：Moodle
- 目的：練習 SQL 設計
- 參考資料：上課投影片 SQL.ppt.

- 題目：運用 MySQL 及 MySQL Workbench

0. 請先至 <https://reurl.cc/1ek3em> 下載 DBMSHW4.zip，解壓縮後共有 6 個.csv 檔 7 個.sql 檔

1. Create IMDB Database (Internet Movie DataBase)

IMDB Database Schema 包括下列 Tables

- TITLES (共 2,294,719 筆 records)
 - Title_id varchar(10)
 - Title_type varchar(100)
 - Primary_title varchar(500)
 - Original_title varchar(500)
 - Is_adult int
 - Premiered int
 - Ended int
 - Runtime_minutes int
 - Genres varchar(100)
 - PRIMARY KEY (Title_id),
- RATINGS (共 362,285 筆 records)
 - Title_id char(10)
 - Rating float
 - Votes int
- AKAS (共 1,437,745 筆 records)
 - Title_id varchar(10)
 - Title varchar(1000)
 - Region varchar(100)
 - Language varchar(100)
 - Title_types varchar(100)
 - Attributes varchar(100)
 - Is_original_title int
- EPISODES (共 593,366 筆 records)
 - Episode_title_id varchar(10)
 - Show_title_id varchar(10)

- Season_number int
- Episode_number int
- PEOPLE (共 3,571,829 筆 records)
 - Person_id varchar(10)
 - Name varchar(200)
 - Born int
 - Died int
 - PRIMARY KEY (Person_id),
- CREW (共 9,990,055 筆 records)
 - Title_id varchar(10)
 - Person_id varchar(10)
 - Category varchar(500)
 - Job varchar(500)

CreateIMDBSchema.sql 就是產生 IMDB Database Schema 的 SQL 檔。只要在 MySQL Workbench 執行 CreateIMDBSchema.sql 檔案就可產生 IMDB Database Schema。

2. Load Data Into IMDB Database

IMDB 的原始 Data 檔案是 6 個 CSV 檔：

- titles.csv
- ratings.csv
- akas.csv
- episodes.csv
- people.csv
- crew.csv

一般的方法是透過 Import 功能匯入已建立 Schema 的每個 Tables。MySQL Workbench 有 Import Wizard 可自動匯入，但費時許久，且未必可正確地匯入資料。

因此建議大家在 MySQL Workbench 以 SQL Script 運用 MySQL 的 LODA DATA LOCAL INFILE 指令，將 CSV 檔匯入 Table。

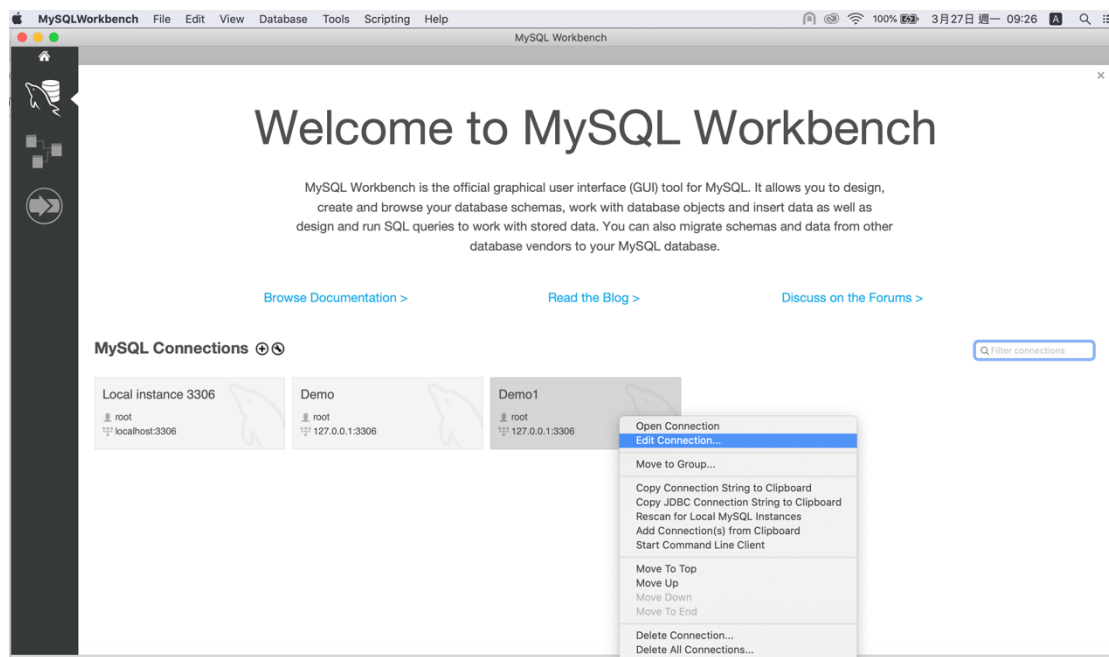
但 MySQL 考量到 Security，LODA DATA LOCAL INFILE 匯入外部資料時有些限制。

(1) 匯入資料前，必須先在 MySQL Workbench Connect 至 MySQL Sever 前做些設定。

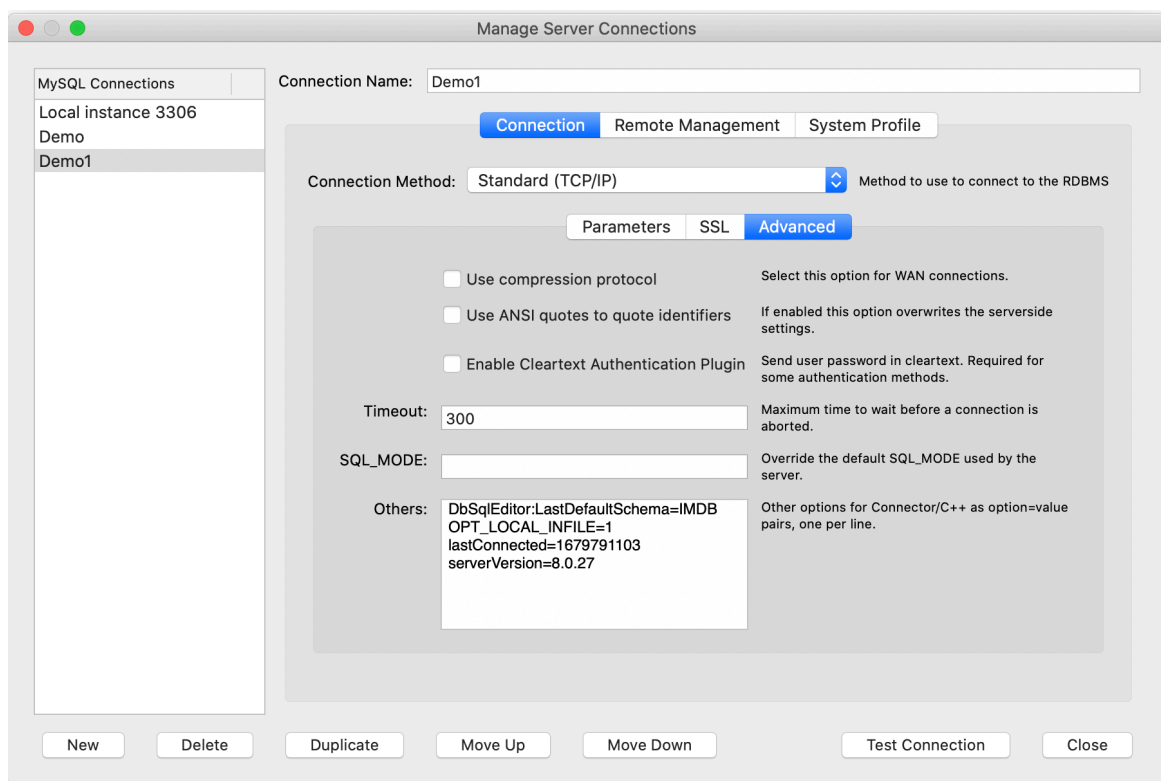
Step 1：首先在啟動 MySQL Workbench 後，在所要連線的 Connection，按右鍵點選 [Edit Connection]。(如圖一所示)

Step 2：出現[Management Server Connection]視窗後，在 [Connection]Tab 的 [Advanced] Tab 的 [Others] Box 填入 OPT_LOCAL_INFILE=1，同時也順便將 [Timeout]改成 300。(如圖二所示)

Step3：設定好後，點選 Close。接著就可直接建立 Connection。



圖一



圖二

(2) Connect 至 MySQL Server 後，檢查在步驟 1 執行 CreateIMDBSchema.sql 是否已產生 IMDB Schema 的 6 個 Tables。若都沒有問題，接下來就可一一執行我們提供的 6 個 Load Data 的 SQL Script，包括

- LoadTitles.sql

- LoadRatings.sql
- LoadAkas.sql
- LoadEpisodes.sql
- LoadPeople.sql
- LoadCrew.sql

* 執行前，請記得將其中 csv 檔的檔案路徑，修改為你的 csv 檔存放的路徑。例如如果你的 titles.csv 放在 C:/ProgramData/MySQL/MySQL Server 8.0/Uploads 目錄下，那麼 LoadTitles.sql 其中的 "Users/mkshan/Downloads/data/titles.csv" 就要改成 "C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/titles.csv"

* 資料匯入後，請注意回傳訊息。例如匯入 titles.csv 時，回傳訊息為 "2294719 row(s) affected Records: 2294719 Deleted: 0 Skipped 0 Warnings: 0" 才是成功地匯入。

* 有興趣的同學，可自學這 6 個 SQL script 中的指令，了解匯入資料過程中，可能會遇到的問題、可能的解決之道。

3. SQL Query Design: 請設計下列 10 題的 SQL query，並存檔成 Qx.sql, x 代表題號。例如第一題 Q1 存檔成 Q1.sql。每個 SQL Query 執行時間不能超過 3 分鐘。有時有些 SQL query 因執行過久，MySQL Workbench 與 MySQL Server 的 Connection 會 Timeout。解決的辦法是修改 MySQL Workbench 的 Preference 設定。在 MySQL Workbench 的左上方下拉式功能表中點選 Preferences，出現 Workbench Preferences 赤窗後，點選左方的 SQL Editor，就可修改右方 MySQL Session 的時間設定(圖三)。

Q1: List all distinct categories of crew ordered by category. (Hint: 12 records with 1 attribute)

Q2: List the average rating of movies acted by 'Robert Downey Jr.' (Hint: 1 record with 1 attribute)

Q3: List the year and the number of comedy movies in each year ordered by year. Note that year with zero comedy movie and year of null are excluded. (Note: Comedy movies 包含所有 genre 有出現 comedy 字眼的電影, Hint: 116 records with 2 attributes,)

Q4: List the name of actresses who have performed at least one action movie which has been released in Japan (region JP) during 1991~2000, ordered by name. (Note: action movies 包含所有 genre 有出現 action 字眼的電影, Hint: 58 records with 1 attribute)

Q5: List the number of actresses who have appeared in any title with 'Robert Downey Jr.' (Hint: 1 record with 1 attribute)

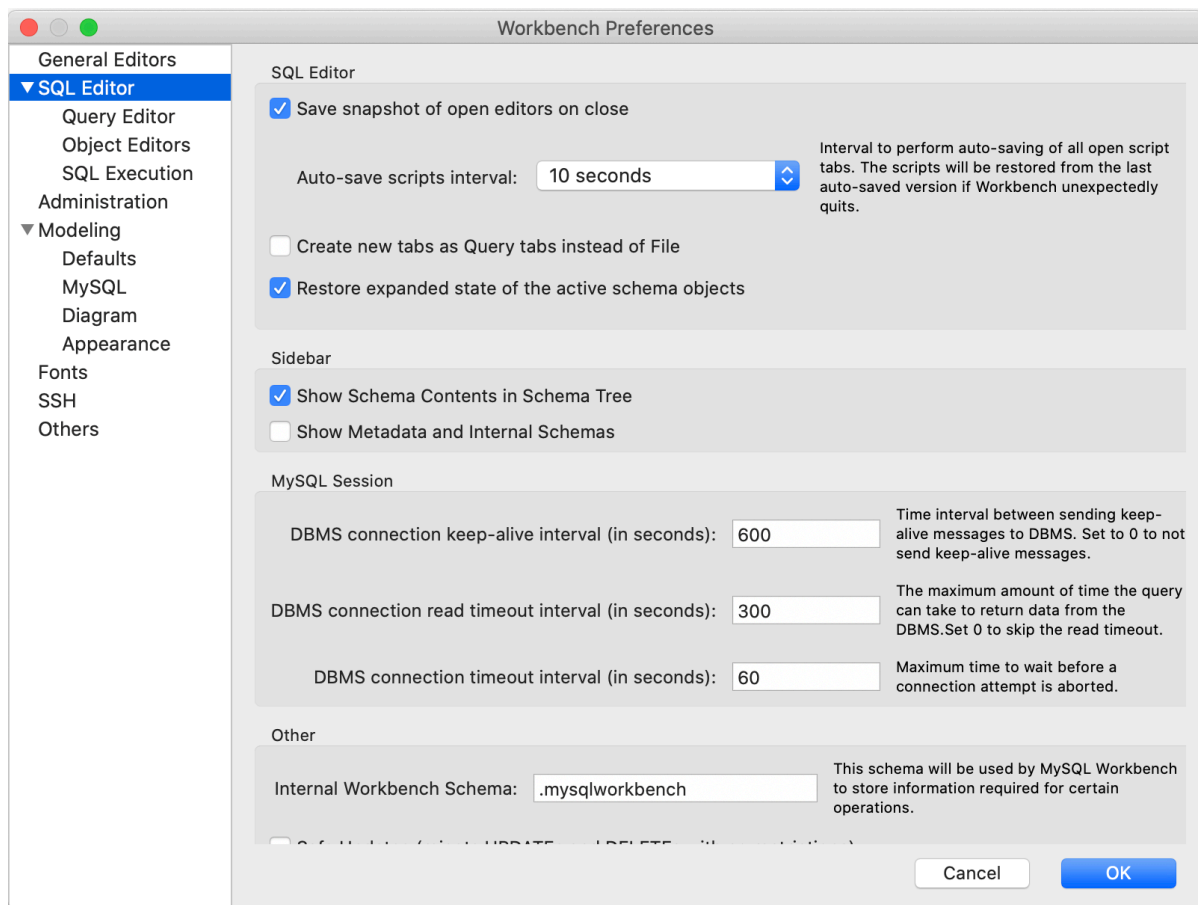
Q6: List the name, and the birth year of actors who appeared in any title before he was twenty (≤ 20) ordered by name. (Note: people are born at age zero, Hint: 4264 records with 2 attributes)

Q7: List the name of composer who appeared in more than 10 high rating titles ordered by name. A title is high rating if the rating is above 8. (Hint: 137 records with 1 attribute)

Q8: List the birth year, both names of actress and actor who were born on the same year and have participated in the same title ordered by the birth year, actress names, and actor name. (Note: the same time denotes the same title_ID, Hint: 4731 records with 3 attributes)

Q9: List the name of people who appeared in any titles with 'Jennifer Lawrence' including herself ordered by their name. (Hint: 7 records with 1 attribute)

Q10: List the name of people who appeared in all the titles with 'Jennifer Lawrence' including herself ordered by their name. (此題為加分題， 4 records with 1 attribute)



圖三

- 繳交時，請將 10 個.sql 檔壓縮成 zip 檔。Zip 檔的檔名為[學號]+HW4.zip。例如學號若是 111753122，則將 Q1.sql, Q2.sql, ..., Q10.sql，一起壓縮成 111753122HW4.zip。
- 不同版本的 MySQL 及 MySQL Workbench (甚至 Mac 與 Windows)，功能或操作難免有些許差異。例如 MySQL 8.0.31 之後有 Except 的功能，但之前的版本則沒有。因此操作過程中，若遇到與此說明或截圖畫面不同的情形，有時是版本間的差異。同學可先透過 Google 或 ChatGPT 查詢錯誤訊息及解決之道。若還是無法解決，歡迎與助教討論。