Introduction to Database Systems

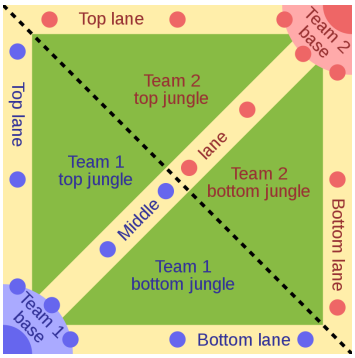
Individual Homework 1: SQL tasks in MySQL

1. Dataset

For briefly introduction, you can refer to these videos [1] [2]. League of Legends (LoL) is a multiplayer online battle arena (MOBA) game developed and published by Riot Games. In LoL, players, as known as “summoners”, can be the role of a “champion” with unique abilities and battle against a team of other player champions. Players compete in matches, lasting anywhere from 20 to 50 minutes on average. In each match, teams work together to achieve a victory condition, typically destroying the core building (called the Nexus) in the enemy team’s base after bypassing a line of defensive structures called turrets, or towers.

In all game modes, players control characters called champions, chosen or assigned every match. Champions begin every match at a low level, and then gain experience over the course of the match to achieve a maximum level of 18. Gaining champion levels in matches allow players to unlock their champion’s special abilities and augment them in a number of ways unique to each character. If a champion loses all their health, they are defeated but are automatically revived in their base after enough time passes. Players also begin each match with a low amount of gold, and can earn additional gold throughout the match in a variety of ways like killing minions and monsters, or killing or assisting to kill enemy players. All of these are specific to each match and can not carry over to subsequent matches. Thus, all players begin each match on a more-or-less equal footing relative to their opposing team.

Summoner’s Rift is the most popular map in LoL. On this map type, two teams of five players compete to destroy an enemy building called a Nexus, which is guarded by the enemy team and a number of defensive structures called turrets, or towers. One nexus is located in each enemy base on opposite sides of the map, in the lower-left and upper-right hand corners. These structures continually create weak non-player characters known as minions, which advance toward the enemy base along three paths: top, middle, and bottom lanes. Between lanes are neutral areas of the map known as the 'jungle', arrayed in four quadrants. A shallow river divides the map between the teams, but doesn't actually impede movement; all champions can wade through it no differently than dry land.



(origin PNG version by Raizin.)

In LoL, there are several game types. What we use here is called Draft Pick. Draft Pick allows each team to ban three champions each (a total of six champions banned), removing them from the match. Note that number of banned champions was former seasons instead of new seasons. Teams then take turns selecting their champions while being able to see the selections of the other team. It is available on Summoner's Rift for matchmaking games, and for all modes in custom games.

2. Tasks

A. Create Tables (20%)

Firstly, download the LoL data from here.

You can refer to this page to see the meaning of the columns.

Then you should create tables based on the following setting. Notice that you **must** make the detail of your tables the same as our description, including 'table name', 'attribute name', 'attribute type', 'primary key', 'foreign key', 'null'.

Please **paste the screenshot of your tables by using `describe` command to your report**, it will take **5%** of your grades in this homework.

Table Name	Attribute Name	Type	Primary Key	Foreign Key	NULL
champ	champion_name	varchar(15)			NO
	champion_id	int	YES		NO
match_info	match_id	int	YES		NO
	duration	int			
	version	varchar(15)			
participant	player_id	int	YES		NO

	match_id	int		match_info (match_id)	NO
	player	tinyint			
	champion_id	int			NO
	ss1	varchar(15)			
	ss2	varchar(15)			
	position	varchar(13)			NO
teamban	match_id	int	YES		NO
	team	char(1)			NO
	champion_id	int			NO
	banturn	tinyint	YES		NO
stat	player_id	int	YES		NO
	win	boolean			
	item1	smallint			
	item2	smallint			
	item3	smallint			
	item4	smallint			
	item5	smallint			
	item6	smallint			
	kills	tinyint			
	deaths	tinyint			
	assists	tinyint			
	longesttimespentliving	smallint			
	doublekills	tinyint			
	triplekills	tinyint			
	quadrakills	tinyint			
	pentakills	tinyint			
	legendarykills	tinyint			

	goldearned	mediumint			
	firstblood	boolean			

Please **answer following question in your report**, it will take **15%** of this homework.

1. (3%) What the difference between type “char” and type “varchar”?
   譯：變數型態“char”和“varchar”有什麼不同？
2. (3%) Type “boolean” would be stored as which type in MySQL?
   譯：“boolean”型別實際上是以什麼型別存在MySQL裡面的？
3. (4%) How many bytes it should take for “tinyint”, “smallint”, “mediumint”, “int”? (e.g. 8 bytes for “bigint”) And what’s the range they can express? (e.g. from −1000 to 1000)
   譯：“tinyint”, “smallint”, “mediumint”, “int”各需要多少bytes來儲存？(e.g. 8 bytes for “bigint”)
   還有他們的表示範圍可以從哪裡到哪裡？(e.g. from −1000 to 1000)
4. (5%) What do you think about this table schema? If you can change this table architecture, how would you modify it and why?
   譯：你對這架構有什麼想法？如果你可以更改這架構，你會怎麼改？為什麼？

B. Load csv Datas (0%)

You should load the data mentioned above into your database.
Here we don’t restrict the method you use, but your have to check the data is loaded successfully by yourself. The following number is the data records for each table.

Table Name	# of Data Records
champ	138
match_info	182527
participant	1825270
stat	1825270

teamban	1089969

C. Query Tasks (80%)

There will be 12 query tasks in this part.

For task 1 to 10, you are only allowed to use **one** query to find the answer, and you **don't** have to explain your SQL. But you should **paste the screenshot of your SQL and the query answer in your report**. Noted that the **column names** of your query answers should be **same as our examples**.

For task 11 and 12, you could use multiple queries to finish your task. Try to **explain** what're your queries doing, what's the meaning of the result, what's your conclusion etc.. And **paste the screenshot of your SQL and the query result in your report**.

For submission, also write every query task into a single `sql` file, named as "1.sql", "2.sql" ….

Note that this dataset is not clean, which means you may need to consider more things to get correct answer. Potential flaws like type of positions in participant.

1. (5%) Please list the number of all different champions. You must have **"COUNT"** syntax in usage of SQL.

列出有幾種不同的英雄(champions)，你的語法必須包含 **COUNT**。

cnt
123456

2. (5%) Please list the number of different versions. They are same version if the first two numbers of version are same. For example, "7.9.185.1051" and "7.9.186.8155" belong to same version, but different with "7.92.184.113". You must have **"DISTINCT"** syntax in usage of SQL.

資料中，遊戲總共經歷了幾次不同的版本(version)(只要版本號前兩位相同即為同一版本)(例："7.9.185.1051"與"7.9.186.8155"為同一版本，但與"7.92.184.113"為不同版本)，你的語法必須包含 **DISTINCT**。

cnt
65321

3. (5%) Please list the top 3 frequently use of the champion names and counts, which the position summoner choosing is JUNGLE. You must sort counts in decreasing order and have **"ORDER BY"** syntax in usage of SQL.

找出玩家在打JUNGLE位置時，前3名最常用的角色名稱與其資料筆數，並根據筆數由大到小排序，你的語法必須包含 **ORDER BY**。

champion_name	cnt
YOYOTV	66666

4. (5%) Please list the top 5 longest match id and how long the game is taken. You should transfer time format to hh:mm:ss.

列出遊戲時間前5高的場次id，並顯示其遊戲時間多久，請將遊戲時間換成"時:分:秒"的字串格式（例：00:34:07）

match_id	time
555	09:09:09

5. (5%) There are two teams in every match. Please list the number of winning teams and losing teams which average longest time spent living in each team greater than or equals to twenty minutes. You must output win or lose in string as following example. Note that longesttimespentliving only refers to one player's longest time spent living.

每場遊戲都是兩隊伍互相競賽。請查出隊伍平均最長存活時間（longesttimespentliving）超過20分鐘（包含20分鐘）的隊伍中，贏的隊伍和輸的隊伍筆數各有多少隊（注意longesttimespentliving只單指該玩家最長存活時間）（輸贏請用字串來輸出，如下方範例）。

win_lose	cnt
lose	123
win	456

6. (5%) In LoL, some teams will pick champions which have great ability to win matches in earlier or later period. Please list the most appear champions of each

position (TOP/MID/JUNGLE/DUO_CARRY/DUO_SUPPORT) which the matches end in forty to fifty minutes (including 40 and 50 minutes). You need to sort position in alphabetical order as following example, and you must have **"BETWEEN"** syntax in usage of SQL.

在LoL中，有時會刻意選擇前/後期英雄確保遊戲時間進行到前/後面能有很大的勝率。列出當比賽遊戲時間在40~50分鐘 (duration) (皆包含邊界時間) 時五個位置 (TOP/MID/JUNGLE/DUO_CARRY/DUO_SUPPORT) 分別出現最多次的英雄 (請依照位置字典順序輸出，如下方範例)，你的語法必須包含 **BETWEEN**。

position	champion_name
DUO_CARRY	A_champion
DUO_SUPPORT	B_champion
JUNGLE	C_champion
MID	A_champion
TOP	D_champion

7. (10%) Please list the champion names with highest KDA (KDA = (sum_of_Kills + sum_of_Assists) / sum_of_Deaths) and its corresponding KDA of each position. Note that you should not take into account if the total number of deaths of a champion is zero. You need to sort position in alphabetical order as following example. Hint: **GROUP BY**

列出五個位置 (TOP/MID/JUNGLE/DUO_CARRY/DUO_SUPPORT) 中，且總擊殺參與率 (總KDA = (總Kill + 總Assist) / 總Death) 最高的英雄與對應的總KDA值 (不考慮總Death為0的英雄)，請依照位置字典順序輸出，如下方範例。提示：**GROUP BY**

position	champion_name	kda
DUO_CARRY	A_champion	5.0000
DUO_SUPPORT	B_champion	4.3210
JUNGLE	C_champion	3.2130
MID	C_champion	2.2222

TOP	D_champion	1.1234

8. (5%) Please list the champion names which are not banned in version 7.7. You need to sort champion names in in alphabetical order, and you must have **"NOT IN"** syntax in usage of SQL.

列出在7.7版本中，沒有被禁用過的英雄名稱，請依照英雄名稱字典順序由小到大排序，你的語法必須包含 **NOT IN**。

champion_name
YS Hero

9. (10%) There is a slogan said by a famous streamer: If you pick Lee Sin, I will pick Teemo. Please list the number of win, lose counts and its winning ratio (#win / #(win+lose)) in each version which definition is same as Q2 when Lee Sin and Teemo are in same teams in the match. You need to sort version in alphabetical order as following example like 4.1, 4.10, 4.2, 4.3.

某知名實況主曾言：你李星 (Lee Sin) 我提摩 (Teemo)。請列出在各個版本 (version) 中 (小數點前二位相同即為同一版本，反之則不同)，當李星 (Lee Sin) 與提摩 (Teemo) 在同一隊時，勝場、敗場與勝率 (勝場 / (勝場+敗場)) 分別為何，請依照 version "字典順序" 輸出，例：4.1, 4.10, 4.2, 4.3。

version	win_cnt	lose_cnt	win_ratio
4.10	5	4	0.5555

10. (10%) In LoL, every champion may be more easily defeated by specific champions, which is called "counter". Please list the top 5 winning ratio of champion names, KDA which is defined as Q9 and average gold earned (goldearned) of both sides and battle records when summoners select TOP position and the opposite champion is Gragas. Note that you only need to consider the number of matches of each champion facing Gragas on TOP larger than 100.

在LoL角色中，有些角色會比較容易打贏或打輸特定角色。依序列出當上路(TOP)對手為古拉格斯(Gragas)時，上路(TOP)角色勝率（參考第九題定義）最高的前五隻英雄，並列出上路(TOP)雙方的總擊殺參與率(總KDA，參考第七題定義)、平均經濟(goldearned)與對戰場數。這邊我們只在乎在上路與古拉格斯(Gragas)對峙的資料大於100筆的英雄。

self_champ_name	win_ratio	self_kda	self_avg_gold	enemy_champ_name	enemy_kda	enemy_avg_gold	battle_record
OAO!	0.9999	2.3456	12.4567	Gragas	2.2222	456.4561	486
OwO/	0.9988	2.3633	98.6666	Gragas	2.3111	457.1234	321

11. (10%) If you want to play the "TOP" position and get the maximum win rate, how will you choose your summoner spells (ss1 and ss2)? (Answer by your own view)

當上路(TOP)玩家想要獲得最大的勝率時，應該要如何選擇的召喚師技能(summoner spell, ss1 and ss2)？(此題為開放式答案，請利用 SQL 找出的結果來闡述你的觀點)

12. (5%) Feel free to think any valuable observation with explanation.
請自由發想 SQL，找出有趣的資訊並解釋他的涵義

3. Discussion

TAs had opened a channel **HW1 討論區** on E3 of the course, you can ask questions about the homework in the channel. TAs will answer questions in the channel as soon as possible.

Discussion rules:

1. Do not ask for answer of the homework.
2. Check if someone had asked the question you have before asking.
3. We encourage you to answer other students' questions, but again, do not give the answer of the homework. Reply the messages to answer questions.
4. Since we have this discussion channel, do not send email to ask questions about the homework unless the questions are personal and you do not want to ask publicly.

4. Submission

1. The deadline of this homework is **11/2 (Thr.) 23:59:59**.
2. Paste your screenshot and write down your answer for Task A and Task C. See the above description for more information. You can use Chinese or English in this homework.
3. You should put your `pdf` and `sql` files into one folder, each should named as "**HW1_XXXXXXX.pdf**", "**1.sql**", "**2.sql**" …. And the folder should named as "**HW1_XXXXXXX**" where XXXXXXX is your student ID.

   Then compress your folder into one `zip` file. Submit it to New E3 System with the format **HW1_XXXXXXX.zip** where XXXXXXX is your student ID.

   We **only accept one zip file**, **wrong format or naming format cause -10 points** to your score (after considering late submission penalty).

4. Late submission lead to score of (original score)*0.7.
5. If there is anything you are not sure about submission, ask in the forum.