

Introduction to Database Systems

Individual Homework 1: SQL tasks in MySQL

陳光悅

10811020

Part A Create Tables

```
mysql> DESCRIBE games;
+-----+-----+-----+-----+-----+-----+
| Field      | Type      | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+-----+
| Game       | int       | NO   | PRI | NULL    |       |
| away       | char(3)   | NO   |     | NULL    |       |
| home       | char(3)   | NO   |     | NULL    |       |
| away_score | tinyint   | YES  |     | NULL    |       |
| home_score | tinyint   | YES  |     | NULL    |       |
| Date       | datetime  | NO   |     | NULL    |       |
+-----+-----+-----+-----+-----+-----+
6 rows in set (0.01 sec)
```

```
mysql> DESCRIBE inning;
+-----+-----+-----+-----+-----+-----+
| Field      | Type      | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+-----+
| Game       | int       | NO   | PRI | NULL    |       |
| Inning     | char(3)   | NO   | PRI | NULL    |       |
| Runs       | tinyint   | YES  |     | NULL    |       |
| Hits       | tinyint   | YES  |     | NULL    |       |
| Errors     | tinyint   | YES  |     | NULL    |       |
+-----+-----+-----+-----+-----+-----+
5 rows in set (0.00 sec)
```

```
mysql> DESCRIBE hitters;
```

Field	Type	Null	Key	Default	Extra
Game	int	NO	PRI	NULL	
Team	char(3)	NO		NULL	
AB	tinyint	YES		NULL	
R	tinyint	YES		NULL	
H	tinyint	YES		NULL	
RBI	tinyint	YES		NULL	
BB	tinyint	YES		NULL	
K	tinyint	YES		NULL	
num_P	tinyint	YES		NULL	
Position	varchar(20)	YES		NULL	
Hitter_Id	mediumint	NO	PRI	NULL	

11 rows in set (0.01 sec)

```
mysql> DESCRIBE pitchers;
```

Field	Type	Null	Key	Default	Extra
Game	int	NO	PRI	NULL	
Team	char(3)	NO		NULL	
IP	float	YES		NULL	
H	tinyint	YES		NULL	
R	tinyint	YES		NULL	
ER	tinyint	YES		NULL	
BB	tinyint	YES		NULL	
K	tinyint	YES		NULL	
HR	tinyint	YES		NULL	
PC_ST	varchar(10)	YES		NULL	
Pitcher_Id	mediumint	NO	PRI	NULL	

11 rows in set (0.00 sec)

```
mysql> DESCRIBE pitches;
+-----+-----+-----+-----+-----+-----+
| Field | Type      | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+-----+
| Pitch_Id | mediumint | NO   | PRI | NULL    |       |
| Game     | int       | NO   | MUL | NULL    |       |
| EventId  | smallint  | NO   |     | NULL    |       |
| Num      | tinyint   | NO   |     | NULL    |       |
| Inning   | char(3)   | YES  |     | NULL    |       |
| Pitcher  | varchar(35) | YES  |     | NULL    |       |
| Pitch    | varchar(50) | YES  |     | NULL    |       |
| _Type    | varchar(20) | YES  |     | NULL    |       |
| MPH      | smallint  | YES  |     | NULL    |       |
+-----+-----+-----+-----+-----+-----+
9 rows in set (0.00 sec)
```

```
mysql> DESCRIBE players;
+-----+-----+-----+-----+-----+-----+
| Field | Type      | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+-----+
| Id    | mediumint | NO   | PRI | NULL    |       |
| Name  | varchar(20) | YES  |     | NULL    |       |
+-----+-----+-----+-----+-----+-----+
2 rows in set (0.00 sec)
```

Q1.

CHAR 是用來儲存固定長度的字串；VARCHAR 則可以儲存長度不固定的字串。
例如：Team 的名字固定都是三個字元，其資料形態就可以是 CHAR(3)；而因為選手的名字長度都不一致，因此其資料形態則需宣告 VARCHAR(20)。

Q2.

Foreign key 是指向其他 table 的 primary key 的 column，用於確定兩個 table 的關聯性及資料完整性。Foreign key 中的資料，必須要出現在他指向的 primary key 中。

例如：inning 中的 Game，一定會出現在 games 中的 Game。這個限制可以避免inning 中出現一場不存在的比賽（無法在 games 中找到該場比賽）。

Q3.

TINYINT: 1 byte, from -128 to 127

SMALLINT: 2 bytes, from -32768 to 32767

MEDIUMINT: 3 bytes, from -8388608 to 8388607

INT: 4 bytes, from -2147483648 to 2147483647

Q4.

我覺得這個 schema 挺完整的，尤其是寫完後面的 query tasks 後，更是可以發現這麼這個 schema 如此拆分資料的方便性與完整性。

原本有在想為什麼 players 的那個 table 中，沒有一個 Team 的 column，但寫到後面的第 8 題才知道竟然還有交易選手這種東西的存在.....，如此就有可能一個 player 對應到多個 team，增加 table 的複雜度，因此才沒有在 players 的 table 中加入 team 的 column。

不過在 pitches 的 table 中，我覺得可以將 Pitcher 那一個 column 改成 Pitcher_id，因為 pitcher name 可能會重複，但 Pitcher_id 並不會重複。如此，想要知道 pitches 的 table 中 Pitcher 所對應到的 Pitcher_id，還要與 pitchers 與 players 的 table 做交叉比對，才能得到最終的答案。

Part C. Query Tasks

Q12.

考慮到欲選擇的後援投手應為還有在打比賽的選手，這樣才可以確保所選的選手仍能上場比賽。因此我只考慮在 2021 年有上過場，且先發次數小於等於 10 的投手，做為可以選擇的後援投手名單。

因為以後援來說，大部分最重要的就是救火的能力，能凍結跑者為佳，因此我選擇看三振率（9*K/IP）。

另外一個選擇看的是好球率（ST/PC）。這樣可以確保選擇的投手投出的球的穩定度，不會一直投出壞球而導致敵方保送。但考慮到好球率高也可能是因為容易被打，因此將他放在三振率後面來看。

下面兩個 table 為後援投手名單中，三振率與好球率最高的前 20 名投手：

Pitcher_id	Pitcher_name	avg_K/9
30412	C. Roe	90
31209	M. Freeman	90
32396	K. Quackenbush	90
42826	R. Gonzalez	90
41284	N. Maton	90
35222	D. Altavilla	49.5
36364	J. Cronenworth	45
30782	A. Rizzo	45
33724	K. Mella	45
40939	A. Munoz	45
36132	P. Weigel	44.1818
41455	D. Lee	34.2
40039	J. Weems	32.1429
32566	W. Harris	32.1429
41523	E. Paredes	32.1
41129	A. Vasquez	31.5
36500	J. Newberry	30.375
31756	K. Lobstein	30
42587	R. Rosso	29.5714
31601	Z. Rosscup	29.25

Pitcher_id	Pitcher_name	avg_ST/PC
31200	R. Torreyes	0.9311
33364	P. Valaika	0.9000
29607	J. Turner	0.9000
33979	J. Nogowski	0.8482
35222	D. Altavilla	0.8118
32201	P. Evans	0.8000
40393	D. Mendick	0.8000
34935	B. Lail	0.7565
33513	R. Bleier	0.7535
35076	S. Wilkerson	0.7500
31209	M. Freeman	0.7500
41129	A. Vasquez	0.7500
42826	R. Gonzalez	0.7500
41284	N. Maton	0.7500
33283	S. Alcantara	0.7500
28959	J. McGee	0.7454
29955	C. Culberson	0.7411
4888739	B. Sittinger	0.7398
33248	F. Mejia	0.7381
34862	M. Strahm	0.7347

從兩張 table 中，可以看到 N. Maton 與 D. Altavilla 的三振率與好球率皆在前 20 名內，代表兩人皆能在穩定的打出好球的同時，讓對手三振。然而兩位選

手的好球率相差不大，但三振率卻相差頗多，似乎 **N. Maton** 是更好的選擇。

此外，我將兩位選手分別抓出來，統計他們在 2021 年賽場上，會投出的球種數目，可以看到 **N. Maton** 會投四種球，而 **D. Altavilla** 只會投兩種。

Pitcher_id	Pitcher_name	num_type
35222	Altavilla	2
41284	Maton	4

因此綜觀下來，**N. Maton** 將會是我的首選目標。