

Introduction to Database Homework 1

110652021 龔大承

Task A:

```
kungtacheng@kungtacheng: ~/Desktop/DB/Homework 1
+-----+
| games |
| hitters |
| inning |
| pitchers |
| pitches |
| players |
+-----+
6 rows in set (0.00 sec)

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 (0.00 sec)

mysql> 
```

```
kungtacheng@kungtacheng: ~/Desktop/DB/Homework 1
+-----+-----+-----+-----+-----+-----+
| 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 (0.00 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> 
```

```
kungtacheng@kungtacheng: ~/Desktop/DB/Homework 1
| 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 (0.00 sec)

mysql>
```

```
kungtacheng@kungtacheng: ~/Desktop/DB/Homework 1
| Position | varchar(20) | YES | | NULL | |
| Hitter_Id | mediumint | NO | PRI | NULL | |
+-----+-----+-----+-----+-----+
11 rows in set (0.00 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>
```

```
kungtacheng@kungtacheng: ~/Desktop/DB/Homework 1
| 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>

kungtacheng@kungtacheng: ~/Desktop/DB/Homework 1
| 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)

mysql>
```

1. char 是固定長度的字串 varchar 則是最大長度為某個數字的字串
2. foreign key 可以保證每個局數都屬於某一場比賽而不會有某局不屬於任何一場比賽的狀況
以避免 inning 對應不上 games 的狀況
3.
 - a) tinyint -> 1 bytes, from 0 - 255
 - b) smallint -> 2 bytes, from -32768 - 32767
 - c) mediumint -> 3 bytes, from -8388608 - 8388607
 - d) int -> 4 bytes -> 4 bytes, from -2147483648 - 2147483647

4. 我認為 Game 中也應該要有 H, R, E 這些屬於每場比賽的欄位 另外對於 hitters 跟 pitchers 應該也可以和 players 整合在一起 可以減少合併這些表格需要的時間 畢竟大部分的查詢都會需要使用到姓名來區分資料會更為明顯

Task C:

1.

```
mysql> source 1.sql
+-----+
| cnt |
+-----+
| 558 |
+-----+
1 row in set (0.00 sec)
```

2.

```
mysql> source 2.sql
+-----+-----+
| Game      | num_innings |
+-----+-----+
| 360701114 | 19          |
| 370905102 | 19          |
| 401077087 | 19          |
+-----+-----+
3 rows in set (0.31 sec)
```

3.

```
mysql> source 3.sql
+-----+-----+-----+
| Pitcher_Id | Pitcher      | tol_innings |
+-----+-----+-----+
| 39251      | W. Buehler   | 222.5       |
| 5403       | A. Wainwright| 209.8       |
| 31267      | Z. Wheeler   | 209.6       |
+-----+-----+-----+
3 rows in set (0.19 sec)
```

4.

```
mysql> source 4.sql
+-----+-----+-----+-----+-----+-----+
| Hitter      | avg_P/PA | avg_AB | avg_BB | avg_K | tol_PA |
+-----+-----+-----+-----+-----+-----+
| E. Butler   | 4.2101   | 1.3913 | 0.0870 | 0.3478 | 42      |
| C. Quantrill| 4.1852   | 1.5000 | 0.0556 | 0.3333 | 34      |
| N. Margevicius| 3.9861   | 1.2500 | 0.0000 | 0.5833 | 22      |
+-----+-----+-----+-----+-----+-----+
3 rows in set (4.12 sec)
```

5.

空白

6.

```
mysql> source 6.sql
+-----+
| Type |
+-----+
| Eephus Pitch |
| Forkball |
| Intentional Ball |
| Knuckle Curve |
| Knuckleball |
| Screwball |
| Slow Curve |
+-----+
7 rows in set (18.79 sec)
```

7.

```
mysql> source 7.sql
+-----+-----+-----+-----+-----+
| Team | Hitter | avg_hit_rate | tol_hit | win_rate |
+-----+-----+-----+-----+-----+
| SF | B. Crawford | 0.2940 | 503 | 0.6527 |
| LAD | T. Turner | 0.3167 | 258 | 0.6437 |
| TB | W. Franco | 0.2860 | 300 | 0.6084 |
| HOU | Y. Gurriel | 0.3137 | 591 | 0.5843 |
| MIL | E. Escobar | 0.2965 | 189 | 0.5783 |
+-----+-----+-----+-----+-----+
5 rows in set (0.64 sec)
```

8.

```
mysql> source 8.sql
+-----+-----+-----+-----+-----+-----+
| Pitcher | cnt | 2020_avg_K/9 | 2021_avg_K/9 | 2020_PC-ST | 2021_PC-ST |
+-----+-----+-----+-----+-----+-----+
| Changed | 133 | 12.8063 | 12.1091 | 40.8364-25.6936 | 42.0674-25.6936 |
| Unchanged | 221 | 11.9625 | 11.8112 | 53.1891-33.6151 | 52.7626-33.6151 |
+-----+-----+-----+-----+-----+-----+
2 rows in set (0.37 sec)
```

9.

這題我跟助教給的 example 算出不一樣的結果 因為打擊率的計算方式跟助教的有點落差 但最後也不確定助教想要的是哪個計算方式 所以還是交出了這樣的 result 希望助教斟酌給分

```
mysql> source 9.sql
+-----+-----+
| hrd1 | win_rate |
+-----+-----+
| 0.11 | 0.9605 |
+-----+-----+
1 row in set (1.12 sec)
```

10.

空白

11.我是用打擊率 投手自責分率（防禦率）來和主場優勢進行比較 然而得出的結論是主場優勢帶來的勝率還是高於另外兩項數據 雖然沒辦法說明主場優勢一定是最重要的 但應該也是棒球戰績中相當重要的決定因素 這題希望我們用球隊的近況來分析 但很抱歉我對於 lag 相關的函數真的很不熟悉所以沒辦法達成題目的要求 希望助教斟酌給分

[illegible]

12.我使用了投手的自責分率及每局被上壘率來進行分析 最後取出他們交集的部份 得出以下的結果 為了確保成績的可信度 我限制每個投手的總投球局數必須超過 100 局 以避免過小的樣本帶來的錯誤數據 另外我有注意到此題是希望我們分析後援投手 但我很抱歉我所剩的時間讓我來不及對投手做分類 所以只能草草交出這樣的 result 先跟助教說對不起 之後會更早找別人討論 以避免這次作業自己硬寫寫不完的問題 這次的作業完成度不高還請助教見諒

```
mysql> source 12.sql
+-----+
| Pitcher |
+-----+
| C. Kershaw |
| J. deGrom |
| J. Hader |
| J. Verlander |
| K. Jansen |
| M. Scherzer |
| R. Osuna |
| W. Buehler |
+-----+
8 rows in set (0.29 sec)
```