

# Introduction to Database Systems

## Individual Homework 1 :SQL tasks in MySQL

### 1. Answer Questions (15%)

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

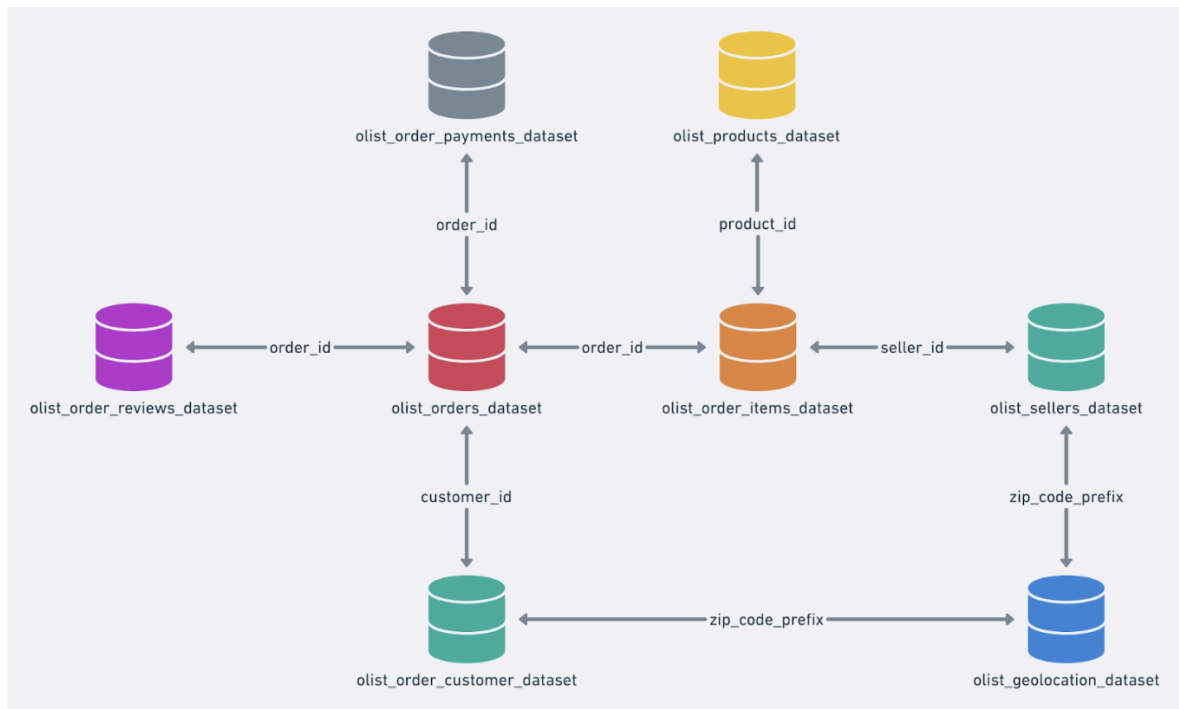
- a. (3%) What the difference between type “char” and type “varchar”?  
譯:變數型態 “char” 和 “varchar” 有什麼不同?
- b. (4%) Type “boolean” would be stored as which type in MySQL?  
譯:“boolean”型別實際上是以什麼型別存在MySQL裡面的?
- c. (8%) How many bytes it should take for “tinyint”, “smallint”, “mediumint”, “int”? (e.g. 8 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)

### 2. Dataset([link](#))

Target is a globally recognized brand and a leading retailer in the United States, known for offering exceptional value, inspiration, innovation, and a unique shopping experience.

This dataset focuses on Target's operations in Brazil, covering 100,000 orders placed between 2016 and 2018. It includes detailed information on order status, pricing, payment and shipping performance, customer locations, product attributes, and customer reviews.

The data is available in 8 csv files:



### 3. Tasks (85%)

#### A. Create Tables (5%)

Firstly, download the data from [here](#).

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.

'describe' command output example:

Field	Type	Null	Key	Default	Extra
seller_id	varchar(50)	NO	PRI	NULL	
seller_zip_code_prefix	int	YES		0	
seller_city	varchar(50)	YES		NULL	
seller_state	varchar(50)	YES		NULL	

4 rows in set (0.02 sec)

The following number is the data records for each table.

Table Name	# of Data Records
customers.csv	99441
geolocation.csv	718104
order_items.csv	112650
orders.csv	99441
payments.csv	104k
products.csv	32951
sellers.csv	3095

## B. Query Tasks (80%)

There will be 10 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 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. (8%) Please list the number of all different clients. You must have “**COUNT**” syntax in usage of SQL.

請列出所有不同clients的數量。你必須在 SQL 中使用 "COUNT" 語法。

cnt
123

2. (8%) Please list the number of different customer\_unique\_id which start with “ff9”. You must have “**DISTINCT**” syntax in usage of SQL.

請列出以 "ff9" 開頭的不同 customer\_unique\_id 的數量。你必須在 SQL 中使用 "DISTINCT" 語法。

cnt
123

3. (8%) Please list which 3 cities have the most orders made in state 'RJ'. You must sort counts in decreasing order and have **"ORDER BY"** syntax in usage of SQL.

請列出在 'RJ' 州訂單最多的前三個城市。你必須在 SQL 中使用 "ORDER BY" 語法，並按訂單數量降序排列。

city
a
b
c

4. (8%) Please list the top 3 order\_id which cost the most days to deliver (the date of product arrival - the date the product starts to deliver). You must sort counts in decreasing order and have **"ORDER BY"** syntax in usage of SQL.

請列出花費最長配送時間的前三個 order\_id (計算方式為產品到達日期減去產品開始配送日期)。你必須在 SQL 中使用 "ORDER BY" 語法，並按天數降序排列。

order_id
a
b
c

5. (8%) Please list the numbers of each states' orders which are not delivered.

請列出每個州尚未送達訂單的數量。

state	cnt
a	123
b	123
c	123

6. (8%) There's a missing order in payment.csv, list its order\_id.

有一筆記錄在payment.csv中缺失，列出他的order\_id。

order_id
123

7. (8%) Please list the order\_id which cost the most days to deliver (the date of product arrival - the date the product starts to deliver ) in each state. You must sort states in decreasing order and have “**ORDER BY**” syntax in usage of SQL.

請列出每個州配送時間最長的 order\_id (計算方式為產品到達日期減去產品開始配送日期)。你必須在 SQL 中使用 "ORDER BY" 語法，並按州的名稱降序排列。

state	order_id
a	123
b	123
c	123

8. (8%) Please list the number of orders paid by debit card and purchase in 2017 in state 'RJ'.

請列出在 2017 年於 'RJ' 州使用debit card支付的訂單數量。

cnt
123

9. (8%) Please list the top three seller\_ids who sold the most items priced over 150 and the number of these items they sold (price over 150). Only consider sellers who sold more than 100 items. (Hint: Use **WITH**)

請列出前三賣出最多大於150元商品的seller\_id以及他們賣出大於150元的商品的數量，這邊只考慮售出商品數量大於100的銷售員。

seller_id	cnt
a	3
b	2
c	1

10. (8%) Please list the payment\_type that appears more than 100 times and the number they appear, **excluding the top two** most frequent payment methods. Order by the number the payment\_type appears.  
請列出出現次數大於100的支付方式以及他們的出現次數, 並且不包含出現前兩多的支付方式。依照支付方式的出現次數排列。

payment_type	cnt
a	3
b	2
c	1

## 4. 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.

## 5. Submission

1. The deadline of this homework is **10/7 (Mon.) 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. Submit file name `{student_id}_HW1.zip`(e.g. 110550001\_HW1.zip). Report file `{student_id}_HW1_report.pdf`. Only PDF is acceptable. Wrong file name will reduce 10 points.
4. Submit file format:
  - `{student_id}_{student_name}_HW1.zip`
    - 1.sql
    - 2.sql
    - 3.sql
    - 4.sql
    - 5.sql
    - 6.sql
    - 7.sql
    - 8.sql
    - 9.sql
    - 10.sql
    - `{student_id}_{student_name}_HW1_report.pdf`