Homework Assignment 4

No submission is required.

table function

> ● Read the assignment carefully. *You will need to **write and execute several SQL queries**; and **submit the results of your queries***.
> ● You are **allowed to re-use any of the queries from the lecture slides** while developing solutions to the problems.
> ● This is an individual work; Please be clear with HGU CSEE Standard:
>   ○ Submitting assignments or program codes written by others or acquired from the internet without explicit approval of the professor is regarded as cheating.
>   ○ Showing or lending one's own homework to other student is also considered cheating that disturbs fair evaluation and hinders the academic achievement of the other student.
>   ○ It is regarded as cheating if two or more students conduct their homework together and submit it individually when the homework is not a group assignment.
> ● Posting any of the assignment on the Internet and asking for solutions to arbitrary human or non-human agents is prohibited.
> ● When finished, submit your work to *LMS*.

1. Read the textbook Chapters 5 and 17. Fill in the blanks with the correct answers.

(a) The (   ) function gives the same rank to all tuples that are equal on the ORDER BY attributes; whereas the (   ) function does not create gaps in the ordering.

(b) The use of the keyword (   ) in place of **ROWS** allows the windowing query to cover all tuples with a particular value rather than covering a specific number of typles. Thus, **ROWS CURRENT ROW** refers to eactly one tuple, while **RANGE CURRENT ROW** refers to all tuples whose value for the sort attribute is the same as that of the current tuple.

(c) A (   ) is a statement that the system executes automatically as a side effect of a modification to the database.

(d) The SQL standard supports functions returning tables as results: such functions are called (   ).

(e) (   ) refers to a collection of operations that form a single logical unit of work.

(f) To acquire fast random access to tuples in a file, one can use a/an (   ) structure.

(g) An (   ) stores the values of the search-keys in sorted order and associates each search-key with the records that contain it.

(h) A (   ) is an index whose search key also defines the sequential order of the file.

(i) In a (   ), an index entry appears for only some of the search-key values.

2. Answer the following questions.

(a) According to the textbook description from Section 4.2, what is the main difference between views and named subqueries defined by **WITH**?

(b) List the conditions that an SQL view is said to be updatable.

(c) According to the description from Section 17.1, explain what the ACID properties of a database are.
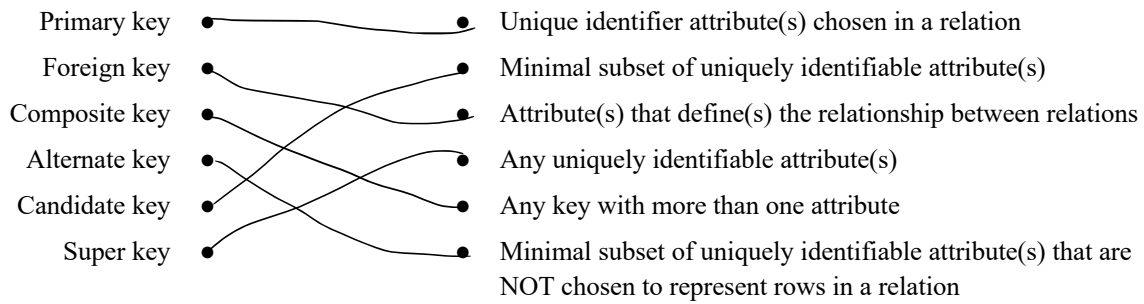
(d) Discuss the differences between the following two queries in terms of the computational efficiency.

- **SELECT** ID, **RANK() OVER (ORDER BY** GPA **DESC) AS** s_rank
  FROM student_grades;

- **SELECT** ID, (1 + (
      SELECT COUNT(*)
      FROM student_grades B
      WHERE B.GPA > A.GPA)) **AS** s_rank
  FROM student_grades A
  ORDER BY s_rank;

(e) Give at least three example items that metadata or system catalog stores.

(f) What are the main disadvantages of the index-sequential file organization?

3. Match each of the following key types to the corresponding definition.

Primary key	Unique identifier attribute(s) chosen in a relation
Foreign key	Minimal subset of uniquely identifiable attribute(s)
Composite key	Attribute(s) that define(s) the relationship between relations
Alternate key	Any uniquely identifiable attribute(s)
Candidate key	Any key with more than one attribute
Super key	Minimal subset of uniquely identifiable attribute(s) that are NOT chosen to represent rows in a relation

4. Structured Qeury Language. Consider the following database tables.

* Table: *artist*

id	name	networth	location
101	Marcus Miller	5,000,000	NY
102	Pat Metheny	10,000,000	MA
103	John Scofield	2,000,000	NY
104	Abraham Laboriel	1,500,000	CA
105	Kenny Garrett	1,500,000	NJ
106	Cory Wong	750,000	MN
107	Jacob Collier	20,000,000	England
108	Victor Wooten	5,000,000	TN
109	Earl Klugh	16,000,000	MI

* Table: *album*

id	a_id	title	year	label
10001	101	Afrodeezia	2015	Blue Note
10002	101	Renaissance	2012	Dreyfus Jazz
10003	102	The Unity Sessions	2016	Nonesuch
10004	102	Speaking of Now	2002	Warner Bros
10005	102	Letter from Home	1989	Geffen
10006	102	First Circle	1984	ECM
10007	103	Combo 66	2018	Verve
10008	103	Uberjam	2002	Verve
10009	103	Groove Elation	1995	Blue Note
10010	104	Guidum	1995	Wigwam
10011	104	Dear Friends	1993	101 South
10012	105	Sounds from the Ancestors	2021	Mack Avenue
10013	105	Beyond the Wall	2006	Nonesuch
10014	105	Songbook	1997	Warner Bros
10015	106	Motivational Music for the Syncopated Soul	2019	
10016	106	The Optimist	2018	
10017	107	In My Room	2016	Membran

Given the above tables, evaluate the following queries and write down the expected results.

(a)　SELECT location FROM artist
　　　GROUP BY location
　　　HAVING COUNT(*) = 1;

(b)　SELECT label, COUNT(*) CNT FROM album
　　　WHERE year < 2000
　　　GROUP BY label
　　　HAVING label LIKE 'W%';

(c)   SELECT name FROM artist
      WHERE id IN (
            SELECT a_id FROM album
            WHERE year BETWEEN 2000 AND 2010
)
      ORDER BY name;

(d)   SELECT name, location FROM artist a1
      WHERE EXISTS(
          SELECT * FROM artist a2
          WHERE a1.name <> a2.name AND a1.location = a2.location
      );

(e)   SELECT location, COUNT(DISTINCT album.title) AS CNT FROM artist, album
      WHERE artist.id=album.a_id
      GROUP BY location
      ORDER BY CNT;

5. More query exercises. Launch and access the MySQL databases distributed with the class virtual machine. Below uses the "*sakila*" database (DVD rental database), which consists of 16 tables regarding movie inventory, actors, customers, rental history, payment information, *etc*. For each of the following questions, **find the answer based on the information recorded in the database** and **write a query that shows how you obtained the answer**.

(a) How many distinct films rated 'PG' are available?

(b) How many active customers are living in the district of England?

(c) Considering the rental history (rental) and payment history (payment), who has paid the largest amount of money for renting movies? List the first and last name of the customer, the total number of movie rentals, and total amount of money s/he has paid.

(d) List three most frequent categories of film available at *store_id*=2 (if a store has multiple copies of the same film, consider each copy as an individual inventory).
*Tip: Use **LIMIT** 3 at the end of your query to limit the number of output tuples.*

(e) What is the title of the movie that has the longest description (film_text.description) among the rental store with store_id=2 has?

(f) Which of the films starred by "FRED COSTNER" rented the most? Write the title of the film.

(g) Using the *'customer_list'* view, list all names of people whose address is in the city of 'London'.

(h) Write a query that uses only tables (does not use any views) and returns the same information as in the previous problem (Problem (g)).

6. More query exercises. Launch and access the MySQL databases distributed with the class virtual machine. Below uses the "*sakila*" database (DVD rental database), which consists of 16 tables regarding movie inventory, actors, customers, rental history, payment information, *etc*.

For each of the following queries, evaluate and report the results. You may type in and execute each query to find the solution. Make sure you understand why you obtained the reported results.
(a) SELECT inventory_id,
        ROW_NUMBER() OVER (PARTITION BY inventory_id ORDER BY rental_date) AS ROW_NO,
        rental_id, customer_id, rental_date,
        LAG(rental_date, 1) OVER (PARTITION BY inventory_id ORDER BY rental_date) AS PREV_RENTAL
     FROM rental
     WHERE YEAR(rental_date) = 2005 AND MONTH(rental_date) = 8 AND inventory_id <= 5;

(b) SELECT payment_id, customer_id, rental_id, payment_date, amount,
        SUM(amount) OVER (ORDER BY payment_date
        RANGE BETWEEN INTERVAL 0.5 DAY PRECEDING
            AND INTERVAL 0.5 DAY FOLLOWING) AS DAILY_SUM
     FROM payment
     WHERE YEAR(payment_date) = 2005 AND MONTH(payment_date) = 5
        AND customer_id <= 5;

Write queries for the following questions and show the results of your queries.
(c) Who are the top five customers spending the most amount of money on movie rental in the database?
Hint: Use the `payment` table.

(d) Who are the top five customers renting the most movies in the database?
Hint: Use the `rental` table.

(e) What of the top seven movies (films) rented the most by the customers?
Hint: Use the `rental`, `inventory`, and `film` tables.

7. Indexes. Answer the following questions that are from the textbook exercise problem sets. You may refer to the Internet as well as the textbook for assistance; however, your solution should contain your own ideas in your own language.

(a) (Exercise 14.16) When is it preferable to use a dense index rather than a sparse index? Explain your answer.

(b) (Exercise 14.17) What is the difference between a clustering index and a secondary index?

(c) (Exercise 14.1) Indexes speed query processing, but it is usually a bad idea to create indexes on every attribute, and every combination of attributes, that are potentially search keys. Explain why.