**CIS 422 DBMS**

**SQL Training and Exercises Part2**

**Section 1: Aggregate Functions and Grouping Data**

**GROUP BY:**

Grouping is one of the most important tasks that we have to deal with while working with data. To group rows into groups, you use the GROUP BY clause. The GROUP BY clause is an optional clause of the SELECT statement that combines rows into groups based on matching values in specified columns. One row is returned for each group. It’s worth to note that we often use the GROUP BY in conjunction with an aggregate function to calculate a measure that provides the information for each group. To sort the groups, you add the ORDER BY clause after the GROUP BY clause.

**HAVING:**

The HAVING clause is often used with the GROUP BY clause in the SELECT statement. If you use a HAVING clause without a GROUP BY clause, the HAVING clause behaves like the WHERE clause.

The WHERE clause applies the condition to individual rows before the rows are summarized into groups by the GROUP BY clause. However, the HAVING clause applies the condition to the groups after the rows are grouped into groups. Therefore, it is important to note that the HAVING clause is applied after whereas the WHERE clause is applied before the GROUP BY clause.

**Aggregate Functions:**

You will learn about the SQL aggregate functions including AVG(), COUNT(), MIN(), MAX(), and SUM(). An SQL aggregate function calculates on a set of values and returns a single value. For example, the average function (AVG) takes a list of values and returns the average.

Because an aggregate function operates on a set of values, it is often used with the GROUP BY clause of the SELECT statement. The GROUP BY clause divides the result set into groups of values and the aggregate function returns a single value for each group.

You can use aggregate functions as expressions only in the following:

* The select list of a SELECT statement, either a subquery or an outer query.
* A HAVING clause

1. AVG

The SQL AVG function is an aggregate function that calculates the average value of a set. The following illustrates the syntax of the SQL AVG function:

AVG([ALL|DISTINCT] expression)

If we use the ALL keyword, the AVG function takes all values in the calculation. By default, the AVG function uses ALL whether we specify it or not. If we specify the DISTINCT keyword explicitly, the AVG function will take the unique values only in the calculation.

Calculate the average salary of employees in the department id 5?

SELECT AVG (DISTINCT salary)

FROM employees

WHERE department\_id = 5

Returns department id, department name and the average salary of employees of each department and sort the result in descending order?

SELECT e.department\_id, department\_name, AVG(salary)

FROM employees e

INNER JOIN departments d ON d.department\_id = e.department\_id

GROUP BY e.department\_id

ORDER BY AVG(salary) DESC;

1. COUNT

The SQL COUNT function is an aggregate function that returns the number of rows returned by a query. The following illustrates the syntax of the SQL COUNT function:

COUNT([ALL | DISTINCT] expression);

The result of the COUNT function depends on the argument that you pass to it.

* By default, the COUNT function uses the ALL keyword whether you specify it or not. The ALL keyword means that all items in the group are considered including the duplicate values.
* If you specify the DISTINCT keyword explicitly, only unique non-null values are considered.

Another form of the COUNT function that accepts an asterisk (\*) as the argument is as follows:

COUNT(\*)

* The COUNT(\*) function returns the number of rows in a table including the rows that contain the NULL values.

Find the department id, department name and number of employees for each department in which the number of employees is greater than 5 sorted by the number of employees in descending order?

SELECT COUNT(\*) AS empCount, E.department\_id, department\_name

FROM departments D, employees E

WHERE D.department\_id = E.department\_id

GROUP BY department\_id

HAVING empCount > 5

ORDER BY empCount DESC;

1. SUM

The SQL SUM function is an aggregate function that returns the sum of all or distinct values. We can apply the SUM function to the numeric column only. The following illustrates the syntax of the SUM function.

SUM([ALL|DISTINCT] expression)

* The ALL operator allows you to apply the aggregate to all values. The SUM function uses the ALL operator by default.
* To calculate the sum of unique values, you use the DISTINCT.

Find department id, department name and the total salaries for each department in which the total salaries is greater than 3000 sorted in descending order?

SELECT E.department\_id, D.department\_name, SUM(E.salary) AS totalSal

FROM departments D, employees E

WHERE D.department\_id = E.department\_id

GROUP BY E.department\_id

HAVING totalSal > 30000

ORDER BY totalSal DESC;

SQL provides the MAX function that allows you to find the maximum value in a set of values. The following illustrates the syntax of the MAX function:

MAX(expression)

* The MAX function ignores NULL values.
* Unlike the SUM, COUNT, and AVG functions, the DISTINCT option is not applicable to the MAX function.

Find the employee(s) id, first name, last name, and salary who have the highest salary? Note: try to use the subquery in this question.

SELECT employee\_id, first\_name, last\_name, salary

FROM employees

WHERE salary = (

SELECT MAX(salary)

FROM employees

);

Find department id , department name for each department that has employee whose highest salary is greater than 12000?

SELECT E.department\_id, D.department\_name, MAX(E.salary) AS biggestSal

FROM departments D, employees E

WHERE D.department\_id = E.department\_id

GROUP BY E.department\_id

HAVING biggestSal > 12000;

1. MIN

The SQL MIN function returns the minimum value in a set of values. The following demonstrates the syntax of the MIN function:

MIN(expression)

* Like the MAX function, the MIN function also ignores NULL values, and the DISTINCT option is not applicable to the MIN function.

Retrieve the employee who has the lowest salary in each department, then sorts these departments by the salary in ascending order?

SELECT d.department\_id, d.department\_name, MIN(salary)

FROM employees e, departments d

WHERE e.department\_id = d.department\_id

GROUP BY department\_id

ORDER BY MIN(salary);