**Queries Equivalence Type**

1. **Swapping the order of nestedness level (change-nestedness)**

**“Swapping\_Nestedness”**

Example:

**Query:** SELECT customer\_id

FROM Orders

WHERE customer\_id IN (

SELECT customer\_id

FROM Customers

WHERE country = 'USA'

);

**Equivalent Query:** SELECT customer\_id

FROM Customers

WHERE country = 'USA'

AND customer\_id IN (

SELECT customer\_id

FROM Orders

);

1. **Change the join order (change-order)**

**"Join\_Order"**

Example query:

**Query:** SELECT first\_name, last\_name FROM employees, department WHERE department\_id = 10;

**Equivalent Query:** SELECT first\_name, last\_name FROM department, employees WHERE department\_id = 10;

1. **Converting JOIN(Explicit join to implicit join or vice versa)**

**“Explicit\_Implicit\_Joins”**

**Query:**  SELECT first\_name, last\_name FROM employees, department WHERE department\_id = 10;

**Equivalent Query:** SELECT e.first\_name, e.last\_name FROM employees e JOIN departments d ON e.department\_id = d.department\_id WHERE d.department\_id = 10;

1. **Equivalent Query Using a Subquery (changing from join to nestedness)**

**“Join\_Nested”**

**Query:**  SELECT e.EmployeeName, e.EmployeeID

FROM Employees e

JOIN Departments d ON e.DepartmentID = d.DepartmentID

WHERE d.DepartmentName = 'Sales';

**Equivalent Query:** SELECT EmployeeName, EmployeeID

FROM Employees

WHERE DepartmentID IN (

SELECT DepartmentID

FROM Departments

WHERE DepartmentName = 'Sales'

);

1. **(Using Subqueries)**

**"Condition\_Nested"**

**Query:**  SELECT first\_name, last\_name FROM employees, department WHERE department\_id = 10;

**Equivalent Query:** SELECT first\_name, last\_name FROM employees WHERE department\_id = (SELECT department\_id FROM departments WHERE department\_id = 10);

**7. Equivalent Query Using a join (changing from nestedness to join)**

**“Nested\_Join”**

**Query:**  SELECT first\_name, last\_name FROM employees WHERE department\_id = (SELECT department\_id FROM departments WHERE department\_id = 10);

**Equivalent Query:** SELECT first\_name, last\_name FROM employees, department WHERE department\_id = 10;

**8. Using Common Table Expressions (CTEs)**

**“CTEs”**

**Query:** SELECT text

FROM DBObjects

WHERE name = 'PhotoTag'

**Equivalent Query:** WITH FilteredObjects AS (

SELECT text

FROM DBObjects

WHERE name = 'PhotoTag'

)

SELECT text

FROM FilteredObjects

**9. Using UNION or INTERSECT ALL (convert and/or to union and intersect)**

**"Condition\_IntersectUnion"**

**Query:** SELECT employee\_id, name, department, salary

FROM employees

WHERE department = 'Sales' OR salary > 50000;

**Equivalent Query:** SELECT employee\_id, name, department, salary

FROM employees

WHERE department = 'Sales'

UNION

SELECT employee\_id, name, department, salary

FROM employees

WHERE salary > 50000;

The UNION operator selects only distinct values by default. If you want to include duplicates, you would use UNION ALL.

**10. Using EXISTS**

**“Condition\_EXISTs”**

**Query:** SELECT DISTINCT d.department\_name

FROM departments d

JOIN employees e ON d.department\_id = e.department\_id;

**Equivalent Query:** SELECT department\_name

FROM departments d

WHERE EXISTS (

SELECT 1

FROM employees e

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

);

**11. Using CASE Statements**

**"Case\_Statement"**

**Query:** SELECT p.product\_id,

p.price,

p.category,

p.price \* (1 - d.discount\_rate) AS sale\_price

FROM products p

LEFT JOIN discounts d

ON p.category = d.category;

**Equivalent Query:** SELECT product\_id,

price,

category,

CASE

WHEN category = 'Electronics' THEN price \* 0.90 -- 10% discount

WHEN category = 'Clothing' THEN price \* 0.85 -- 15% discount

WHEN category = 'Books' THEN price \* 0.80 -- 20% discount

ELSE price -- No discount

END AS sale\_price

FROM products;

**12. Simplification and Direct Application**

**“Simplification”**

**Query:**  SELECT \* max(salary), min(salary), avg(salary), count(\*)

FROM (

SELECT Name, [Job Title] as job\_title, [2010 Gross Earnings] as salary

FROM [1314howe].[uw\_salaries\_2011.txt]

) x

WHERE job\_title LIKE '%RESEAR%SR'

AND job\_title NOT LIKE '%APL%'

AND salary > 100000

ORDER BY salary desc;

**Equivalent Query:** SELECT MAX(salary) AS max\_salary,

MIN(salary) AS min\_salary,

AVG(salary) AS avg\_salary,

COUNT(\*) AS total\_count

FROM [1314howe].[uw\_salaries\_2011.txt]

WHERE [Job Title] LIKE '%RESEAR%SR'

AND [Job Title] NOT LIKE '%APL%'

AND [2010 Gross Earnings] > 100000;

**13. Reordering the conditions**

**Query:** SELECT \* FROM Orders

WHERE CustomerID = 1234 AND OrderDate > '2020-01-01' AND Status = 'Shipped';

**Equivalent Query:** SELECT \* FROM Orders

WHERE Status = 'Shipped' AND OrderDate > '2020-01-01' AND CustomerID = 1234;

**Creating Not Equivalent Queries**

1. **Use Different Aggregate Functions**

**"Aggregate\_Function"**

**Query:** SELECT department, AVG(salary) FROM employees GROUP BY department;

**Non-equivalent:** SELECT department, SUM(salary) FROM employees GROUP BY department;

1. **Change Join Condition**

**“Change\_Join\_Condition”**

**Query:** SELECT \* FROM orders INNER JOIN customers ON orders.customer\_id = customers.id;

**Non-equivalent:** SELECT \* FROM orders LEFT JOIN customers ON orders.customer\_id = customers.id;

1. **Modify the Grouping Criteria**

**"Group\_by\_Criteria"**

**Query:**  SELECT city, COUNT(\*) FROM employees GROUP BY city;

**Non-equivalent:** SELECT state, city, COUNT(\*) FROM employees GROUP BY state, city;

1. **Introduce Subtle Changes in**

* **Logical conditions**

**"Logical\_Conditions"**

**Query:**  SELECT \* FROM products WHERE price < 20 AND quantity > 100;

**Non-equivalent:** SELECT \* FROM products WHERE price < 20 OR quantity > 100;

* **Changing values**

**"Value\_Change"**

**Query:**  SELECT \* FROM employees WHERE salary > 50000;

**Non-equivalent:** SELECT \* FROM employees WHERE salary > 500000;

* **Operators**

**“Operators”**

**Query:** SELECT \* FROM employees WHERE salary > 50000;

**Non-equivalent:** SELECT \* FROM employees WHERE salary >= 50000;

1. **Change Sorting and Limiting**

**“Change\_Sorting\_Limiting”**

**Query:**  SELECT \* FROM employees ORDER BY hire\_date DESC LIMIT 10;

**Non-equivalent:** SELECT \* FROM employees ORDER BY hire\_date ASC LIMIT 10;

1. **Remove a join/ a condition**
2. **Modify the Order By Criteria**

**“Order\_by\_Column”**

**Query:** SELECT city, COUNT(\*) FROM employees Order BY city;

**Non-equivalent:** SELECT state, city, COUNT(\*) FROM employees Order BY state, city;

1. **Introduce Ambiguous Column References**

**"Adding\_Removing\_Select\_Column"**

**Query:** SELECT EmployeeName, Department, Salary, HireDate, Position, Email

FROM Employees;

**Non Equivalent:** SELECT EmployeeName, Department, Salary, HireDate, Position

FROM Employees;

1. **Change Table Name**

**”Change\_Table”**

**Query:** SELECT EmployeeID, EmployeeName, Department, Salary

FROM Employees;

**Non Equivalent:** SELECT EmployeeID, EmployeeName, Department, Salary

FROM Employee;

**Syntax Error Type:**

We used 6 advanced syntax error types.

1. **Aggr-attribute:**

**Query:** SELECT DepartmentID, ManagerID, AVG(Salary) AS AverageSalary, MAX(HireDate) AS LatestHireDate

FROM Employees

WHERE HireDate > '2000-01-01'

GROUP BY DepartmentID, ManagerID;

**Modified Query:** SELECT DepartmentID, ManagerID, AVG(Salary), MAX(HireDate)

FROM Employees

WHERE HireDate > '2000-01-01';

Here, AVG(Salary) and MAX(HireDate) are aggregate functions, but DepartmentID and ManagerID are not part of any aggregation and there is no GROUP BY clause. This is incorrect because the SQL engine does not know how to group the non-aggregated columns relative to the aggregated ones.

1. **Aggr-having:**

Query: SELECT EmployeeID, COUNT(\*) AS NumProjects

FROM Projects

JOIN ProjectAssignments ON Projects.ProjectID = ProjectAssignments.ProjectID

WHERE Projects.Status = 'Active'

GROUP BY EmployeeID;

**Modified Query:** SELECT EmployeeID, COUNT(\*) AS NumProjects

FROM Projects

JOIN ProjectAssignments ON Projects.ProjectID = ProjectAssignments.ProjectID

GROUP BY EmployeeID

HAVING Projects.Status = 'Active';

Here, The HAVING clause is improperly used to filter on the Projects.Status column, which is neither an aggregated column nor included in the GROUP BY clause. HAVING should be used to filter aggregated results, not to perform row-level filtering that belongs in a WHERE clause.

1. **Type-Mismatch-nested:**

**Query:** SELECT s.fiberid, s.plate, s.mjd, s.run2d, s.class FROM PhotoObj AS p JOIN SpecObj AS s ON s.bestobjid = p.objid**;**

**Modified Query:** SELECT s.fiberid, s.plate, s.mjd, s.run2d, s.class

FROM PhotoObj AS p

JOIN SpecObj AS s ON s.bestobjid = (SELECT objid FROM PhotoObj WHERE ra > 180)

Here, the subquery will return more than one objid. SQL JOIN operations require that a subquery within the JOIN condition must return exactly one row (or a single value) to equate it to another field correctly. When multiple rows are returned, SQL will throw an error, typically something like "Subquery returns more than 1 row."

1. **Type-Mismatch-Condition:**

**Query:** SELECT EmployeeID, Name, Salary \* Rate -- Assuming 'Rate' is a numeric column or a numeric placeholder

FROM Employee;

**Modified Query:** SELECT EmployeeID, Name, Salary \* 'Rate'

FROM Employee;

Here, the query attempts to multiply Salary (presumably a numeric data type like integer or decimal) by a string literal 'Rate'. This is a type mismatch because SQL does not support direct mathematical operations between numbers and non-numeric strings.

1. **Alias-Undefined:**

**Query:** SELECT employeeName, DepartmentID

FROM Employees e

JOIN Departments d ON e.DeptID = d.DepartmentID;

**Modified Query:** SELECT employeeName, DepartmentID

FROM Employees e

JOIN Departments d ON e.DeptID = dept.DepartmentID;

In this query, dept is used as an alias for the Departments table in the JOIN condition, but it was never defined. The correct alias d was defined but not used consistently.

1. **Alias-Ambiguous:**

**Query:** SELECT e.EmployeeID as Employee\_EmployeeID, e.EmployeeName, m.EmployeeID as Manager\_EmployeeID, m.ManagerName

FROM Employees e

JOIN Managers m ON e.ManagerID = m.ManagerID;

**Modified Query:** SELECT EmployeeID, e.EmployeeName, ManagerName

FROM Employees e

JOIN Managers m ON e.ManagerID = m.ManagerID;

Here The column EmployeeID in the SELECT clause does not specify whether it comes from the **Employees** (e) or **Managers** (m) table. If both tables have an EmployeeID column, this would lead to an "Alias Ambiguous" error because the SQL engine cannot determine which table's EmployeeID to use based solely on the information provided.

**All Prompts:**

**For Syntax Error:** “Does this query have any syntax errors? If yes can you tell me what the problem is?”

**For Missing Type:** i will give you some sql queries, you need to answer some questions, 1. Does this query have any Syntax error or not(yes/no) 2. If yes then is there any missing word or not(yes/no)3. if yes then what is the missing type 4. if yes then what is the missing word 5. if yes then what is the missing position(word count). (You need to count the position of the missing word that is missing)'

**For Equivalence:** Do these queries equivalent to each other and produce the same results when executed on the same database schema? If yes then which method is used to make the second one equivalent?

**For Performance Prediction:** "Does this query take a long time to run: [query]?"

**For sqlToText**:"please use one statement to describe this query: [query]"

**All mock prompts:**

**Prompts for the Syntax Error:**

1. Examine this SQL query for syntax errors. If you find any, could you specify what the error is?
2. Does this query have any syntax errors? If yes can you tell me what the problem is?
3. Please check this query for any syntactical inaccuracies. If there are errors, what exactly are they?
4. Could you review this SQL query and identify if there are any syntax errors? If yes, please describe the specific issues.
5. Assess this SQL query for potential syntax errors. If errors are present, can you detail what they are?

**Prompts for Missing Token**

1. Is this query syntactically correct (Yes/No)? [If not] Is there a missing word in it (Yes or No)? [If yes] Identify the missing word. What is the type of the missing word (keyword, function, alias, table, column)? What is the position of the missing word (word offset)?
2. i will give you some sql queries, you need to answer some questions, 1. Does this query have any Syntax error or not(yes/no) 2. If yes then is there any missing word or not(yes/no)3. if yes then what is the missing type 4. if yes then what is the missing word 5. if yes then what is the missing position(word count). (You need to count the position of the missing word that is missing)'
3. I will present you with several SQL queries. For each query, please provide the following details: 1. Check if there is a syntax error (yes/no). 2. If there is an error, determine if a word is missing (yes/no). 3. If a word is missing, identify the type of the missing word. 4. Specify the exact word that is missing. 5. Indicate the position of the missing word based on word count.
4. Here are some SQL queries for you to analyze. Please answer the following questions for each query: 1. Is there a syntax error present? (yes/no) 2. If yes, is a word omitted? (yes/no) 3. If a word is omitted, what type of word is it? 4. What is the missing word? 5. At what position (word count) should this word appear in the query?
5. I'll provide a set of SQL queries for you to evaluate. You need to respond with: 1. Whether there's any syntax error (yes/no). 2. If there is, does it involve a missing word? (yes/no) 3. If a word is missing, what is the category of the word? 4. Identify the missing word. 5. State the word count position where the missing word should be located.

**Prompts for Equivalence Query:**

1. I will give you a pair of queries. You need to determine whether these queries are equivalent and if they produce the same results when executed against the same dataset and schema. If affirmative, what method is employed to make the second query equivalent?
2. Are the following two SQL queries equivalent, and do they produce the same results when executed on the same database schema? If the answer is yes, which method is used to make the second query equivalent?
3. Compare the outputs of these two SQL queries. Are the results identical across all possible inputs? If so, which technique is utilized to make the second query equivalent?
4. Do these queries equivalent to each other and produce the same results when executed on the same database schema? If yes then which method is used to make the second one equivalent?
5. Are these queries identical in output when run against the same database schema? If so, what method ensures the second query is equivalent?

**Prompts for Performance Prediction:**

1. "Does this query take a long time to run: [query]?"
2. "Is the runtime for this query slow: [query]?"
3. "Is this query expected to run fast: [query]?"
4. "Does this query take a significant amount of time: [query]?"
5. "Is this query expected to run slowly: [query]?"

**Prompts for SQLToText:**

1. Please use one statement to describe this query: [query]
2. Please provide a one-sentence description of this SQL query: [query].
3. Summarize this SQL query in a single sentence: [query].
4. Can you encapsulate the purpose of this SQL query in one sentence? [query].
5. Describe the functionality of this SQL query in just one statement: [query].

Modifeid Prompts:

* Syntax error: “Does this query contain any syntax errors? If so, can you explain what the issue is?”
* Missing token: Answer the following questions for this SQL query [query]:
  + Does this query have any syntax errors? (yes/no)
  + If yes, is there a missing word? (yes/no)
  + If yes, what is the type of the missing word?
  + If yes, what is the missing word?
  + If yes, what is the position of the missing word? (Provide the word count position where the word is missing.)
* Equivalent: Are these queries equivalent, and do they produce the same results when executed on the same database schema? If yes, why are they equivalent?
* Performance: Does this query take a long time to execute: [query]?
* Explanation: Please provide a single statement describing this query: [query]