Conceptual Questions:

1. What is SQL Join?

A JOIN clause is used to combine rows from two or more tables, based on a related column between them.

1. The difference between Left Join and Right Join

* **(INNER) JOIN**: Returns records that have matching values in both tables
* **LEFT (OUTER) JOIN**: Return all records from the left table, and the matched records from the right table
* **RIGHT (OUTER) JOIN**: Return all records from the right table, and the matched records from the left table
* **FULL (OUTER) JOIN**: Return all records when there is a match in either left or right table

1. What is SQL Union?

The UNION operator is used to combine the result-set of two or more SELECT statements.

* Each SELECT statement within UNION must have the same number of columns
* The columns must also have similar data types
* The columns in each SELECT statement must also be in the same order

1. What is the difference between Union and Union All

UNION removes duplicate records (where all columns in the results are the same), UNION ALL does not.

**SQL:**

|  |  |  |
| --- | --- | --- |
| Product | Time | Visit |
| A | 103 | 10 |
| B | 103 | 11 |
| A | 104 | 12 |
| A | 105 | 10 |
| B | 105 | 13 |

How to use pivot to change this table to another type()

|  |  |  |
| --- | --- | --- |
| Time | A | B |
| 103 | 10 | 11 |
| 104 | 12 | Null |
| 105 | 10 | 13 |

Select Time,

sum(case when Product = 'A' then Visit else null end) as A,

sum(case when Product = 'B' then Visit else null end) as B

From Q1

group by Time;

1. How to transfer from table\_1 to table\_2（from table\_2 to table\_1）using Teradata SQL?

**table\_1**

|  |  |  |
| --- | --- | --- |
| date | item | quantity |
| 18/10/10 | a | 10 |
| 18/10/10 | b | 20 |
| 18/10/10 | c | 10 |
| 18/10/11 | a | 5 |
| 18/10/11 | b | 8 |
| 18/10/11 | c | 20 |
| 18/10/12 | a | 5 |

**table\_2**

|  |  |  |  |
| --- | --- | --- | --- |
| date | a | b | c |
| 18/10/10 | 10 | 20 | 10 |
| 18/10/11 | 5 | 8 | 20 |
| 18/10/12 | 5 | .. | .. |

select date,

sum(case when item = 'a' then quantity else null end) as a,

sum(case when item = 'b' then quantity else null end) as b,

sum(case when item = 'c' then quantity else null end) as c

from Q2

group by date;

1. **ANOTHER QUESTION Given the following table:**

**Items\_sold\_a**

|  |  |  |  |
| --- | --- | --- | --- |
| date\_sk | item\_a | item\_b | item\_c |
| 1/1/18 | 50 | 0 | 100 |
| 1/2/18 | 45 | 5 | 200 |
| 1/3/18 | 60 | 8 | 150 |

**Write a SQL script to transform it into the following table:**

**Items\_sold\_b**

|  |  |  |
| --- | --- | --- |
| date\_sk | item | qty |
| 1/1/18 | a | 50 |
| 1/1/18 | b | 0 |
| 1/1/18 | c | 100 |
| 1/2/18 | a | 45 |
| 1/2/18 | b | 5 |
| 1/2/18 | c | 200 |
| 1/3/18 | a | 60 |
| 1/3/18 | b | 8 |
| 1/3/18 | c | 150 |

create table item\_sold\_b as

(select data\_sk, 'a' item, item\_a qty

from item\_sold\_a

union all

select data\_sk, 'b' item, item\_b qty

from item\_sold\_a

union all

select data\_sk, 'c' item, item\_c qty

from item\_sold\_a

order by data\_sk, item);

1. Now given the table item\_sold\_b, recreate the table items\_sold\_a .

create table item\_sold\_a as

(select date\_sk,

sum(case when item = 'a' then qty else 0 end) as 'item\_a',

sum(case when item = 'b' then qty else 0 end) as 'item\_b',

sum(case when item = 'c' then qty else 0 end) as 'item\_c'

from items\_sold\_b);

1. For each user\_id, find the difference between the last action and the second last action. Action here is defined as visiting a page. If the user has just one action, you can either remove her from the final results or keep that user\_id and have NULL as time difference between the two actions.

The table below shows for each user all the pages she visited and the corresponding timestamp.

Table Name: action

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Value** | **Description** |
| user\_id | 6684 | this is id of the user |
| page | home\_page | the page visited |
| unix\_timestamp | 1451640067 | unix timestamp in seconds |

Sample Output

|  |  |
| --- | --- |
| **user\_id** | **Delta\_SecondLast0ne\_LastOne** |
| 2 |  |
| 3 | 2 |
| 4 | 4 |
| 5 |  |
| 7 |  |

**SELECT DISTINCT**(**action**.**user\_id**), tmp3.diff **AS** time\_diff  
 **FROM action  
 LEFT JOIN** (  
 **SELECT** tmp1.**user\_id**, (tmp1.highest - tmp2.second\_highest) **AS** diff  
 **FROM** (  
 **SELECT** tmp1.**user\_id**, *MAX*(**unix\_timestamp**) **AS** highest  
 **FROM** (  
 **SELECT user\_id**, *COUNT*(**unix\_timestamp**) **AS** count1 **FROM action  
 GROUP BY user\_id  
 HAVING** count1>1) **AS** tmp1  
 **LEFT JOIN action ON** tmp1.**user\_id** = **action**.**user\_id  
 GROUP BY** tmp1.**user\_id**) **AS** tmp1  
  
 **LEFT JOIN** (  
 **SELECT user\_id**, *MAX*(**unix\_timestamp**) **AS** second\_highest  
 **FROM action  
 WHERE unix\_timestamp NOT IN** (**SELECT** *MAX*(**unix\_timestamp**) **FROM action  
 GROUP BY user\_id  
 HAVING** *COUNT*(**user\_id**) > 1  
 )  
 **GROUP BY user\_id**) **AS** tmp2 **ON** tmp2.**user\_id** = tmp1.**user\_id**) **AS** tmp3  
 **ON** tmp3.**user\_id** = **action**.**user\_id**;

6.

mobile

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Value** | **Description** |
| user\_id | 128 | this is id of the user who visited a given page on mobile |
| page | page\_5\_mobile | page visited by that user on mobile |

web

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Value** | **Description** |
| user\_id | 1210 | this is id of the user who visited a given page on web |
| page | page\_1\_web | page visited by that user on web |

Sample Output:

|  |  |  |
| --- | --- | --- |
| **WEB\_ONLY** | **MOBILE\_ONLY** | **BOTH** |
| 16 | 31 | 52 |

*###################################################***SELECT** *COUNT*(**user\_id**) **AS** mobile\_only  
**FROM** mobile  
**WHERE user\_id NOT IN** (**SELECT user\_id FROM** web);  
  
**SELECT** *COUNT*(**user\_id**) **AS** web\_only  
**FROM** web  
**WHERE user\_id NOT IN** (**SELECT user\_id FROM** mobile);  
  
**SELECT** *COUNT*(**user\_id**) **AS 'both'  
FROM** web  
**WHERE user\_id IN** (**SELECT user\_id FROM** mobile);

select \*

from (

select count(id) as mobile\_only

from `Q3.1`

where id NOT IN (select id from `Q3.2`)) as a,

(select count(id) as web\_only

from `Q3.2`

where id Not in (select id from `Q3.1`)) as b,

(select count(id) as `both`

from `Q3.1`

where id in (select id from `Q3.2`)) as c;  
*###################################################*

7.

Find the most population country from each continent

**Table Name: continent**

continent country population

asian        china 13 billion

asian        japan 3 billion

north\_amrica us     5 billion

north\_amrica can   2.5 billion

**Sample Output**

country             continent population

china                  asian 13 billion

north\_america    us 5 billion

SELECT DISTINCT(continent), country, population

FROM continent

WHERE (continent, population) IN

(

SELECT continent, MAX(population)

FROM continent

GROUP BY continent.

);

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **第二轮（24hours take home FocusKPI）**  **SQL Exercise** | |  |  |  | |  |  |  |  | | | | |  |  |  |  |  |  |  |
|  |  |  |  |  |  | |  |  |  |  | | | | |  |  |  |  |  |  |  |
|  | LinkedIn uses a database to manage employee data. Your business partner asks you to answer  a few questions for an upcoming talent review. Table names and structures are described below. | | | | | | | | | | | | | | |  |  |  |  |  |  |
|  | Please write a sample query to answer each of the listed questions. | | | | | | | |  |  | | | | |  |  |  |  |  |  |  |
|  |  |  |  |  |  | |  |  |  |  | | | | |  |  |  |  |  |  |  |
|  | **Note:** You are welcome to write in whatever SQL style you are most familiar with  (ANSI, MySQL, PostgreSQL, etc.) | | | | | | | | | | | | | | | | |  |  |  |  |
|  |  |  |  |  |  | |  |  |  |  | | | | |  |  |  |  |  |  |  |
|  | **employee** |  |  |  | **level** | |  |  |  | **department** | | | | |  |  |  | **location** |  |  |  |
|  | employee\_id | | int |  | level\_sk | |  | int |  | dept\_sk | | | | |  | int |  | location\_sk |  | int |  |
|  | name |  | varchar |  | level\_name | |  | varchar |  | dept\_name | | | | |  | varchar |  | city |  | varchar |  |
|  | job\_title |  | varchar |  |  | |  |  |  |  | | | | |  |  |  | state |  | varchar |  |
|  | level\_sk |  | int |  |  | |  |  |  |  | | | | |  |  |  | country |  | varchar |  |
|  | dept\_sk |  | int |  |  | |  |  |  |  | | | | |  |  |  |  |  |  |  |
|  | manager\_id |  | int |  |  | |  |  |  |  | | | | |  |  |  |  |  |  |  |
|  | location\_sk |  | int |  |  | |  |  |  |  | | | | |  |  |  |  |  |  |  |
|  | salary |  | int |  |  | |  |  |  |  | | | | |  |  |  |  |  |  |  |
|  | start\_date |  | date |  |  | |  |  |  |  | | | | |  |  |  |  |  |  |  |
|  | term\_date |  | date |  |  | |  |  |  |  | | | | |  |  |  |  |  |  |  |
|  |  |  |  |  |  | |  |  |  |  | | | | |  |  |  |  |  |  |  |
|  |  |  |  |  |  | |  |  |  |  | | | | |  |  |  |  |  |  |  |
|  | 1. **Select the employee in each department with the highest salary in the US.**   **Include employee name, department name, and salary in the output.** | | | | | | | | | | | | | | | | | |  |  |  |
|  | SELECT e.salary, e.name, d.dept\_name  FROM employee e  INNER JOIN department d ON d.dept\_sk = e.dept\_sk  INNER JOIN location l ON l.location\_sk = e.location\_sk  WHERE l.country = 'US'  AND (e.dept\_sk,e.salary) IN    (    SELECT ee.dept\_sk, MAX(DISTINCT ee.salary)    FROM employee ee    WHERE ee.dept\_sk = d.dept\_sk    GROUP BY ee.dept\_sk    ); | | | | | | | | | | | | | | | | | | | |  |
|  |  |  |  |  | |  | | | | |  |  |  |  |  |  |  |  |  |  |  |
|  | **2. Select the top 5 employees in each department with the highest salaries in the US.**  **Include employee names, department names, and salaries in the output.** | | | | | | | | | | | | | | | | | |  |  |  |
|  | SELECT e.salary, e.name, d.dept\_name  FROM employee e  INNER JOIN department d ON d.dept\_sk = e.dept\_sk  INNER JOIN location l ON l.location\_sk = e.location\_sk  WHERE l.country = 'US'  AND 5 >    (    SELECT COUNT(DISTINCT (e2.salary))    FROM employee e2    WHERE e2.salary > e.salary    AND e2.dept\_sk = d.dept\_sk   );  SELECT  d.name as "Department"  ,e1.Name as "Employee"  , e1.Salary as "Salary"    FROM  Employee e1  JOIN Employee e2 JOIN Department d JOIN location L    WHERE  e1.DepartmentId = e2.DepartmentId  AND e1.Salary <= e2.Salary AND d.id = e2.DepartmentId  AND l.location\_sk = e.location\_sk    GROUP BY d.name,e1.id  HAVING COUNT(DISTINCT(e2.Salary)) <= 3  order by d.name , salary desc | | | | | | | | | | | | | | | | | | | |  |
|  |  |  |  |  | |  | | | | |  |  |  |  |  |  |  |  |  |  |  |
|  | **3. Create a flat table in the database with information from all tables for the department of BizOps.** | | | | | | | | | | | | | | |  |  |  |  |  |  |
|  | SELECT \* FROM  (SELECT \* FROM  (SELECT \* FROM employee  INNER JOIN department USING (dept\_sk)  WHERE dept\_name = ‘BizOps’  ) AS tmp1  LEFT JOIN level USING (level\_sk)  ) AS tmp2  LEFT JOIN location USING (location\_sk); | | | | | | | | | | | | | | | | | | | |  |
|  |  |  |  |  | |  | | | | |  |  |  |  |  |  |  |  |  |  |  |
|  | **4. Pull a list of managers and their direct reports. Include IDs and names for both the**  **manager and direct reports in the output.** | | | | | | | | | | | | | | | | | |  |  |  |
|  | SELECT m.employee\_id AS ManagerID, m.name AS ManagerName,  e.employee\_id AS EmployeeID, e.name AS EmployeeName  FROM employee e  INNER JOIN employee m ON e.manager\_id = m.employee\_id; | | | | | | | | | | | | | | | | | | | |  |
|  |  |  |  |  | |  | | | | |  |  |  |  |  |  |  |  |  |  |  |
|  | **5. Find the number of employees that started at the company each quarter.** | | | | | | | | | | | | |  |  |  |  |  |  |  |  |
|  | SELECT  SUM( CASE WHEN Month(e.start\_date) < 4 THEN 1 ELSE 0 END) AS Q1,  SUM( CASE WHEN (4 <= Month(e.start\_date) AND Month(e.start\_date) < 7) THEN 1 ELSE 0 END) AS Q2,  SUM( CASE WHEN (7 <= Month(e.start\_date) AND Month(e.start\_date) < 10) THEN 1 ELSE 0 END) AS Q3,  SUM( CASE WHEN Month(e.start\_date) >= 10 THEN 1 ELSE 0 END) AS Q4 FROM employee e; | | | | | | | | | | | | | | | | | | | |  |
|  |  |  |  |  | |  | | | | |  |  |  |  |  |  |  |  |  |  |  |
|  | **6. Find the average tenure of all employees by level. If an employee is still at the company,**  **term\_date is null; use today's date to calculate tenure.** | | | | | | | | | | | | | | | | | |  |  |  |
|  | SELECT tmp1.level\_name,  AVG(tmp1.total\_time) AS average\_tenure FROM (SELECT e.level\_sk, l.level\_name, e.start\_date, e.term\_date,  IF(term\_date IS NOT NULL,  DATEDIFF(term\_date, start\_date),  DATEDIFF(NOW(), start\_date)  ) AS total\_time  FROM employee e  INNER JOIN level l  ON l.level\_sk = e.level\_sk ) AS tmp1 GROUP BY tmp1.level\_name; | | | | | | | | | | | | | | | | | | | |  |
|  |  |  |  |  | |  | | | | |  |  |  |  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Business Case |  | |  | |  | |
|  |  |  | |  | |  | |
|  | The sales team at The Sheinhardt Wig Company is struggling. In 2014, the team brought in $1.5 million in revenue versus a plan of $10 million.  The VP of Sales is your business partner, and you need to provide her with a diagnosis of what is causing the underperformance as well as  a high-level recommendation on how to address the team’s underperformance. | | | | | | |
|  |  |  | |  | |  | |
|  | Please present your recommendation using a combination of words and visuals. You may create slides, or write up an executive summary. | | | | | | |
|  |  |  | |  | |  | |
|  | **Widget Company Operating Metrics** | | **Industry Benchmark** | |  | |
|  | # of addressable companies: | 40,000 | |  | |  | |
|  | # of leads provided: | 20,000 | |  | |  | |
|  | # of calls made: | 2,500 | | 80% of leads are called | |  | |
|  | # of meetings set up: | 1,500 | | 50% of calls convert into meetings | |  | |
|  | # of deals closed: | 300 | | 25% of meetings led to closed deals | |  | |
|  | $ per deal: | $5,000 | | $7,500 | |  | |
|  | Total revenue: | $1,500,000 | |  | |  | |
|  |  |  | |  | |  | |
|  |  |  | |  | |  | |
|  | Use questions below as guidance. Make any assumptions you feel are needed,  and please state those assumptions. |  | |  | |  | |
|  |  |  | |  | |  | |
|  | **1. What is causing underperformance? What metrics help you identify the issue?** |  | |  | |  | |
|  |  |  | |  | |  | |
|  | 1. **What would you do to address these issues? (remember to tie back recommendations**   **to the underperformance)** |  | |  | |  | |
|  |  |  | |  | |  | |
|  | 1. **Assuming you have full authority to make the necessary changes,**   **what do you think would be a realistic plan for next year?** |  | |  | |  | |
|  |  |  | |  | |  | |
|  |  |  | |  | |  | |
|  | 1.  There are three metrics performed under the industry benchmark that caused the underperformance: 1) Percentage of leads are called = Number of # of calls made / # of leads provided = 2500/ 20000 = 12.5%; it is significant lower than industry benchmark, which is 80%. This metric is the most important issue causing underperformance. 2) Percentage of meetings led to closed deals = # of deals closed / # of meetings set up = 300 / 1500 = 20%; it is a little bit smaller than industry benchmark, which is 25%. It is less effective than percentage of leads are called. 3) $ per deal is 5000 which is less than industry benchmark, $7500. Also, this metric is less effective than percentage of leads are called too.  Percentage of calls convert into meetings = # of meetings set up / # of calls made = 60%, which is larger than industry benchmark, 50%; it does not cause underperformance. |  | |  | |  | |
|  | 2. I would use the following three ways to address these issues. 1) If the industry benchmark is similar to other competitive companies’, I would check if 20,000 leads provided maybe the wrong target group of our products. For example, if our products are mainly focusing on female market but the leads provided are mainly male; this situation would decrease the # of calls made. Make sure the leads provided are on our target group is very necessary. 2) If the leads provided are mainly on target group, I would check the benchmark made by other companies who have similar product with us. I would analyze the industry benchmark to see if it needs to be adjust based on our situation, which making it easier to achieve the goal in the coming period. 3) If the benchmark is industry standard and the leads provided are target group, it also has other things could cause the underperformance, such as: a) The product maybe restricted by policies announced by government. b) The quality of the product may need to be improved to reach the industry benchmark of $ per deal. |  | |  | |  | |
|  |  |  | |  | |  | |
|  | 3. First of all, I will make plans to conduct researches on the product’s target consumer. Use marketing vehicles according to the target audience’s age, gender or other characteristics. These market vehicles include audio advertisements, Email Direct Marketing, Social Media Marketing, Print advertisements and etc. Use these market vehicles as long as these strategies work to promote sales and revenue.  Secondly, in order to find the more efficient channel to sell products, I would compare the efficiency for each channel, metric such as cost per $1 of closed business. Even low-efficiency market vehicles might promote sales because as the number of channels through which your products can be exposed to, so does their importance in evaluating the products’ purchasing value. When people who are not target consumers see the advertisements, they still have impression on the product that increase the probability of purchasing the product. According to Fractional Attribution, it is a good way of customer nurturing. It is also important to make more efficient phone calls that actually convert into meeting or sales. I will change the benchmark for the number of phone calls that we make to target the leads who are more likely to be the target lead. I will be sure not to send repeating and frequent emails to prevent customers’ dislike toward the product.  In order to decrease customer defections, I will make surveys that can help explore customers’ opinions and interests on the product. I will also provide discounts or promotions for customers who have purchase histories. These customer retention activities help companies retain customers, which will lead to increase of sales. |  | |  | |  | |
|  |  |  | |  | |  | |