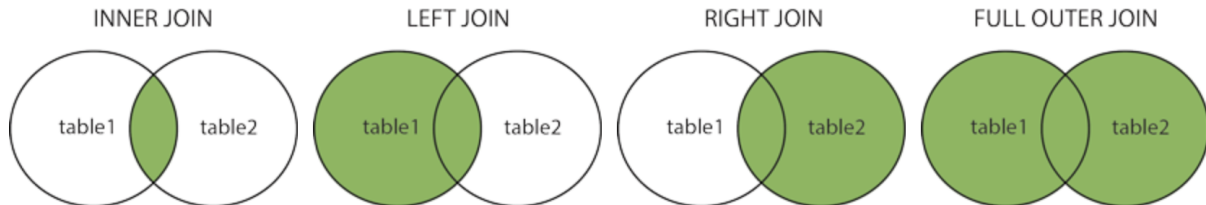


Week 2: SQL JOINS and Data Aggregation

LittleDatabase, Employee Self Join, and Parch and Posey Database

Types of SQL JOINS



[h]

Questions

Little Database

1. Create a PostgreSQL database called `littleDatabase` in pgAdmin.
2. Create two tables in `littleDatabase` called: `table_a` and `table_b`. Table A has two columns (`id` = [1,2,4], `A` = ['m', 'n', 'o']) and Table B has two columns (`id` = [2,3,4], `A` = ['p', 'q', 'r']).

Table 1: Table A

id	a
1	m
2	n
4	o

Table 2: Table B

id	b
2	p
3	q
5	r

3. Apply inner joins, outer joins (left, right, and full), natural join, cross join, and self join.

Table 3: Inner Join

id	a	id	b
2	n	2	p

Employee Database (SELF JOIN EXAMPLE)

This example is based on w3resources.com

Table 4: Left Join

id	a	id	b
1	m	NA	NA
2	n	2	p
4	o	NA	NA

Table 5: Right Join

id	a	id	b
2	n	2	p
NA	NA	3	q
NA	NA	5	r

Table 6: Full Join

id	a	id	b
1	m	NA	NA
2	n	2	p
4	o	NA	NA
NA	NA	5	r
NA	NA	3	q

Table 7: Natural Join

id	a	b
2	n	p

Table 8: Cross Join

id	a	id	b
1	m	2	p
1	m	3	q
1	m	5	r
2	n	2	p
2	n	3	q
2	n	5	r
4	o	2	p
4	o	3	q
4	o	5	r

Table 9: Self Join

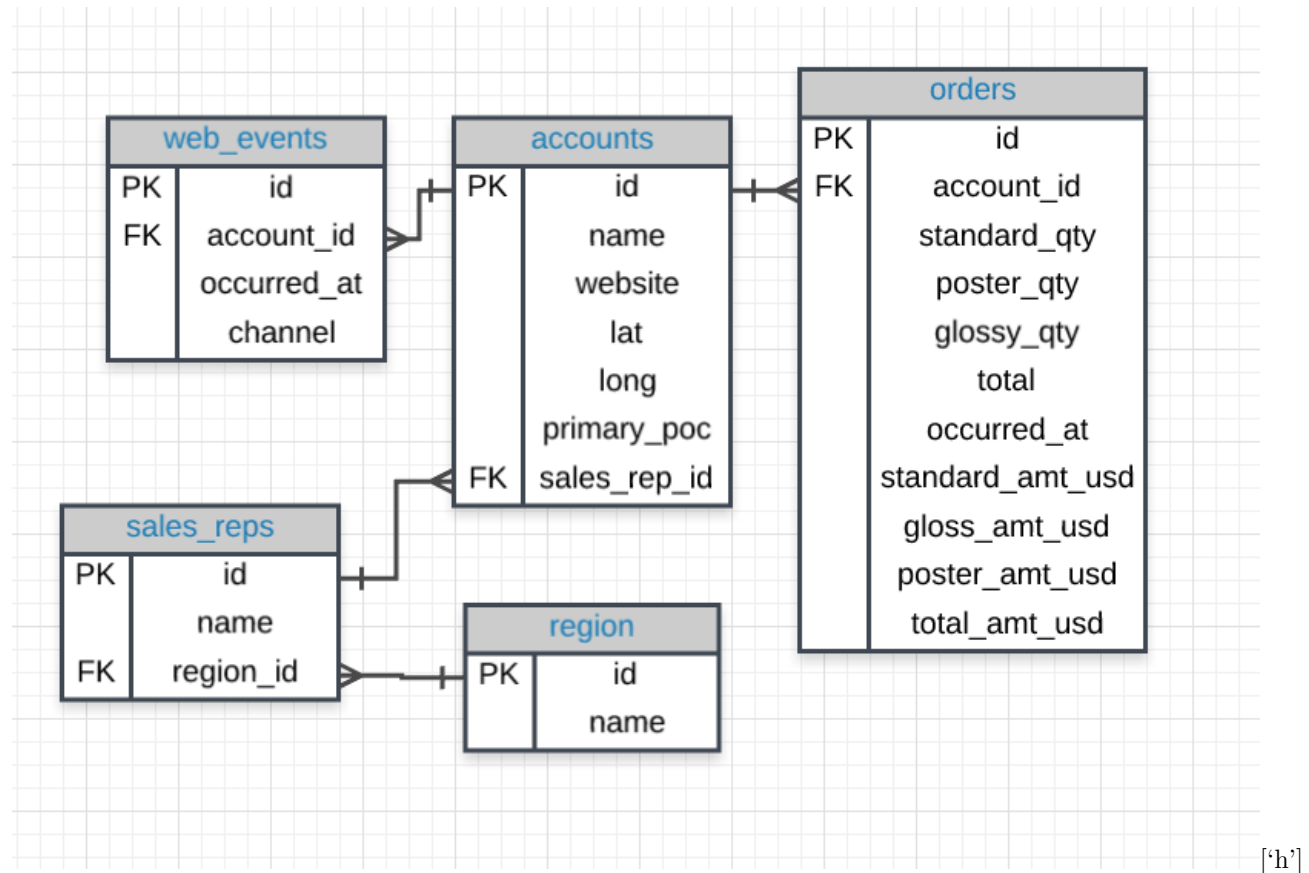
id	a	id	a
1	m	1	m
2	n	2	n
4	o	4	o

4. Create a PostgreSQL database called `employee_self_join` in pgAdmin.
5. Create the `employee` table using the `employee_self_join.sql` file.

6. What is the organization hierarchy structure of the company? i.e. List all the employees who work for each manager.

Parch and Posey Database

This example is based on Mode Analytics.



7. Write a query to return all orders by Walmart. Return the order id, account name, as well as the total paper quantity for each order.
8. **YOUR TURN** Create a table that has the different channels used by account id 1001. Your final table should have only 2 columns: account name and the different channels.
9. **YOUR TURN** Create a table that has the orders that occurred in 2015 (sorted by date). Your final table should have 4 columns: occurred_at, account name, order total, and order total_amt_usd.
10. Create a table that has the **region** for each **sales_rep** along with their associated accounts. Your final table should include three columns: the region name, the sales rep name, and the account name.
11. **YOUR TURN** Create a table that provides the region for each sales_rep along with their associated accounts. This time only for the Midwest region. Your final table should include three columns: the region name, the sales rep name, and the account name.
12. **YOUR TURN** Who was the primary contact associated with the earliest web_event?
13. **YOUR TURN** Create a table that has the name for each region for every order, as well as the account name and the unit price they paid ($\text{total_amt_usd}/\text{total}$) for the order. Your final table should have 3 columns: region name, account name, and unit price. A few accounts have 0 for total, so you might have to divide by $(\text{total} + 0.0001)$ to assure not dividing by zero.

Data Aggregation

14. Count the number of rows in the `accounts` table.
15. Find the total amount of `poster_qty` paper ordered in the `orders` table.
16. What is the min and max order quantity for each poster papers in the database?
17. **YOUR TURN** When was the earliest order ever placed?
18. **YOUR TURN** Find the mean (AVERAGE) amount spent per order on each paper type.
19. **YOUR TURN** What is the MEDIAN `total_usd` spent on all orders?

GROUP BY

20. Find the total sales in usd for each account. You should include two columns: the total sales for each company's orders in usd and the company name.
21. **YOUR TURN** Find the total number of times each type of channel from the `web_events` was used. Your final table should have two columns - the channel and the number of times the channel was used.
22. **YOUR TURN** What was the smallest order total value in USD placed by each account. Provide only two columns - the account name and the total usd. Order from smallest dollar amounts to largest.
23. **YOUR TURN** Find the number of sales reps in each region. Your final table should have two columns - the region and the number of `sales_reps`. Order from fewest reps to most reps.
24. **YOUR TURN** For each account, determine the average amount of each type of paper they purchased across their orders. Your result should have four columns - one for the account name and one for the average spent on each of the paper types.