SQL Frequency Median Calculation

Asked in Visa SQL Round

Advanced Interview Question

> **Find median from a frequency distribution.** When data is stored as **value-frequency pairs**

Input Table

Value	Frequency
1	3
2	5
3	7
4	4
5	2

Key Insight:

Total Observations = 21 → Median position = 11th value

1 Calculate Total Frequency

Sum all frequencies: 3 + 5 + 7 + 4 + 2 = 21

2 Find Median Position

For odd total (21):

Median position = (N + 1) / 2 = (21 + 1) / 2 = 11

3 Calculate Cumulative Frequency

Value 1	3	3
Value 2	5	8
Value 3	7	15
Value 4	4	19
Value 5	2	21

4 Locate Median Group

Find where median position (11) falls:

- Cumulative frequency 8 (value 2) → position 9-?
- Cumulative frequency 15 (value 3) → position 9-15 ✓

Median is in value 3 group

5 Determine Median Value

Since median position (11) falls within value 3's frequency range (positions 9-15),

Median = 3

Cumulative Frequency Method

Universal SQL

```sql
-- Calculate cumulative frequency and total count
WITH cte AS (
    SELECT
        value,
        SUM(frequency) OVER (ORDER BY value) AS cum_freq,
        SUM(frequency) OVER () AS total_cnt
    FROM frequency_data
)
-- First value where cumulative frequency reaches the
median position
SELECT
    MIN(value) AS median
FROM cte
WHERE cum_freq ⩾ (total_cnt + 1) / 2;
```

🔍 HOW IT WORKS

Value	Frequency	Cumulative	Check: cum_freq >= 11
1	3	3	❌ 3 >= 11
2	5	8	❌ 8 >= 11
3	7	15	✅ 15 >= 11
4	4	19	✅ 19 >= 11
5	2	21	✅ 21 >= 11

- Selects the smallest value where cum_freq >= 11
- Returns **MIN(3, 4, 5)** = **3**

SQL Frequency Median

Clarity > Optimization

Recursive CTE Expansion + Median

Universal SQL

```sql
WITH RECURSIVE expanded AS (
    -- base case: first occurrence
    SELECT value, 1 AS seq
    FROM freq_table

    UNION ALL

    -- recursive case: keep expanding until frequency is
met
    SELECT e.value, e.seq + 1  FROM expanded e
    JOIN freq_table f ON e.value = f.value
    AND e.seq < f.frequency
),
ordered AS (
    SELECT value,
        ROW_NUMBER() OVER (ORDER BY value) AS rn,
        COUNT(*) OVER () AS cnt FROM expanded
)
SELECT AVG(value) AS median
FROM ordered
WHERE rn BETWEEN (cnt + 1) / 2 AND (cnt + 2) / 2;
```

Real-World Applications

Master frequency median calculations

Frequency
Median = 3

All SQL methods correctly return 3 as the median

🌍 Real-World Applications

Where frequency median is used:

1. **Survey Analysis:** Likert scale responses stored as frequency counts
2. **Retail Analytics:** Product ratings with frequency distributions
3. **Quality Control:** Defect counts per batch
4. **Web Analytics:** Page view counts by user segments
5. **Healthcare:** Symptom frequency in patient groups
6. **Education:** Test score distributions

Storage Benefit: Frequency tables reduce storage when many repeated values exist.

Master Advanced Statistical SQL

Frequency median calculation is essential for data compression and statistical analysis

✅ **Interview Essential:** Tests window functions and cumulative sums

✅ **Real Applications:** Survey analysis, rating systems, quality control

✅ **Performance:** Handle compressed data efficiently

✅ **SQL Mastery:** Advanced window functions and CTEs

🔁 **REPOST so your manager sees your growth.**