SQL for Dată Scientists / MLE / SDE

AppliedAICourse.com

InterviewPrep.AppliedCourse.com

Why ?

Data scientist → mostly one round

MLE → Programming + SQL

SDE → Programming + SQL

Data $\xrightarrow{\text{SQL}}$ Code

services / startups / product-based

Level 1 & 2

Level 2 & 3

rare

examples ahead!

Applied AI Course → SQL on IMDB dataset

[10 queries]

↓

Level 2 & 3

↓

nested-Queries

Level-1

①

schema: ⟨Id, Marks⟩ ← Students

Find the second highest marks

↳ very common in many companies

Sub-problems → sort descending order

→ pick the 2nd from top

edge-cases → 1, 89

2, 89

3, 86

[duplicates]

```
Select distinct Marks from Students
order by Marks desc
LIMIT 1 OFFSET 1
```

Alt:

ⓐ Select MAX(Marks) from Students

ⓑ Select max(marks) AS marks from
Students
WHERE marks < (select max(marks)
from students)

Level - 2

2

Points : $< x, y >$

Find the shortest distance between the points

[Amazon, Microsoft]

Python/Java/C++:

```
for each $P_i$ in points
    for each $P_j$ in points
        if $P_i \neq P_j$
            $d(P_i, P_j)$
```

```
SELECT   MIN (SQRT ( POW( P1.x - P2.x , 2 ) + POW ( P1.y -P2.y,2)
                                                              ))
                        as  minDist

         FROM  Points P1  JOIN  Points P2

                ON  P1.x != P2.x   OR  P1.y != P2.y
```

③

Level-2

orders: cust, item, date

Select customers who purchased atleast two items and on two different dates

→ e-commerce Companies

group-by COUNT → item Cnt, date Cnt
cust

↳ distinct

where → filtering (>,2)

```
Select Cust, COUNT(DISTINCT item) as itemCnt
             COUNT(DISTINCT date) as dateCnt

from Orders
GROUP BY cust
HAVING itemCnt >= 2 AND dateCnt >= 2
```

4

Level - 2.5

scores: player, country, goals

Find players who scored more than all Spanish players and more than atleast one german player

Find players who scored more than all Spanish players and more than atleast one german player

goals > ALL (spanish player goals)

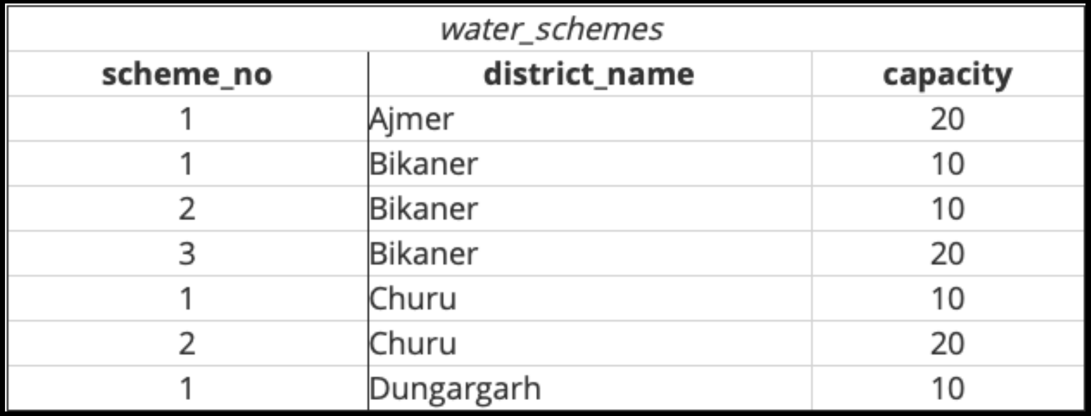
goals > ANY (german player goals)

```
Select t.player from scores As t
   where t.goals >ALL ( SELECT t1.goals
                        from scores As t1
                        where t1.country = 'Spain')

   AND

     t.goals >ANY ( SELECT t2.goals from
                    scores as t2 where
                      t2.country = 'germany')
```

5

Level - 2:5

[GATE 2016]



water_schemes

scheme_no	district_name	capacity
1	Ajmer	20
1	Bikaner	10
2	Bikaner	10
3	Bikaner	20
1	Churu	10
2	Churu	20
1	Dungargarh	10

Print names of districts whose total capacity $\geq$ average capacity of all districts total : district_name, capacity $\xleftarrow[\text{sum}]{\text{groupby}}$ water_schemes

total_avg : average_capacity $\xleftarrow{\text{avg}}$ total

result : district_name $\xleftarrow[X]{\text{where}}$ total, total_avg name of the sub-query block

```
with total(name, capacity) as

    select district_name, sum(capacity)
    from water_schemes
    group by district_name

with  total_avg(capacity) as
    select avg(capacity)
    from total

select name
    from total,  total_avg
    where total.capacity ≥ total_avg.capacity
```

} → sub-Query block

[sub-query -refactoring]

Key focus

→ Basic synlax

→ nesled Queries

→ self-joins [select ... from Emp E1,
Emp E2 ...]

→ Break a pooblem inlo Sub-pooblems

For more practice:

1. GATE CS https://www.geeksforgeeks.org/dbms-gq/sql-gq/

2. Online resources: "SQL interview Questions Amazon"

   ↳ SOLVE actual Query-related Questions

   ↳ Theory is easy!