



Sheet7  
QUERY OPTIMIZATION

1) Consider the following queries.

- a) Draw at least two query trees that can represent each of these queries. Under what circumstances would you use each of your query trees?
- b) Draw the initial query tree for each of these queries; then show how the query tree is optimized by the query optimization heuristic algorithm.

[Q1] `SELECT Fname, Lname, Address  
FROM Employee, Department  
WHERE Dname='Research'  
AND Dnumber=Dno;`

[Q2] `SELECT E.Fname, E.Lname, S.Fname, S.Lname  
FROM Employee E, Employee S  
WHERE E.SuperSSN = S.Ssn;`

[Q3] `SELECT Pnumber, Pname, count(*)  
FROM Project, Works_on, Employee  
WHERE Pnumber = Pno  
AND Ssn = Essn  
AND Dno = 5  
GROUP BY Pnumber, Pname;`

2) Develop cost functions for the PROJECT, UNION, INTERSECTION, SET DIFFERENCE, and CARTESIAN PRODUCT algorithms.

3) Calculate the cost functions for different options of executing the JOIN operation  
`Department ⋈Mgr_ssn=Ssn Employee`

How to submit the homework assignments?

- Solve the sheet individually without looking up the solution on the Internet. The sheet is to practice; it is a learning tool not an exam.
- Assignments are to be **handwritten**.
- Papers are to be scanned (I like camscanner app). Put all images in a pdf file (camscanner does that for you)

# Sheet (7)

11

Name: Asmaa Gamal Abdel - Halim Habrouk Nagy.

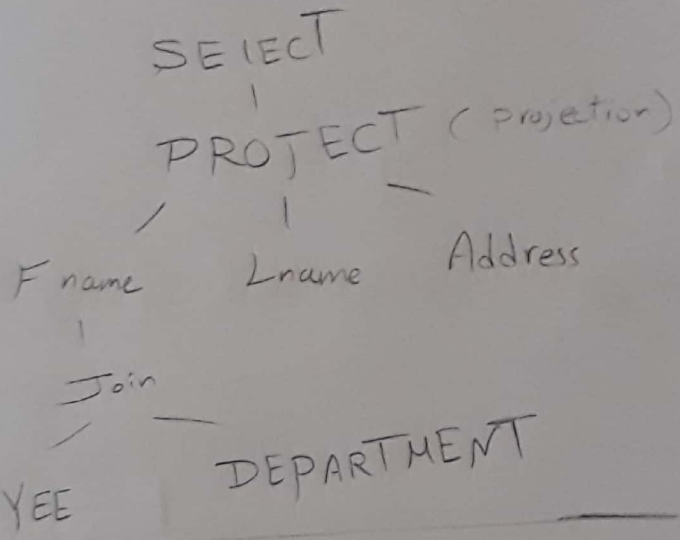
Course: "DB M S" - Audience Course

Department: Communications & Electronics

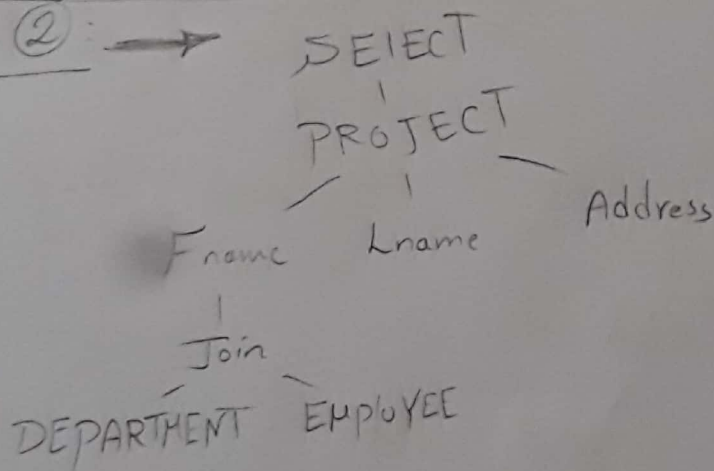
## Database

### Query optimization

① (a) \* query tree ①:  
[Q<sub>1</sub>]



\* Query Tree ②: →



→ This is suitable when we want to read info of employees working in the 'research' dep but we want to start with the Department table.

⑥ [Q<sub>2</sub>] • Initial Query tree :

SELECT

PROJECT

E. Fname

E. Lname

S. Fname

S. Lname

JOIN

EMPLOYEE

EMPLOYEE

• Optimized Query Tree :

SELECT

PROJECT

E. Fname

E. Lname

S. Fname

S. Lname

JOIN

EMPLOYEE

EMPLOYEE

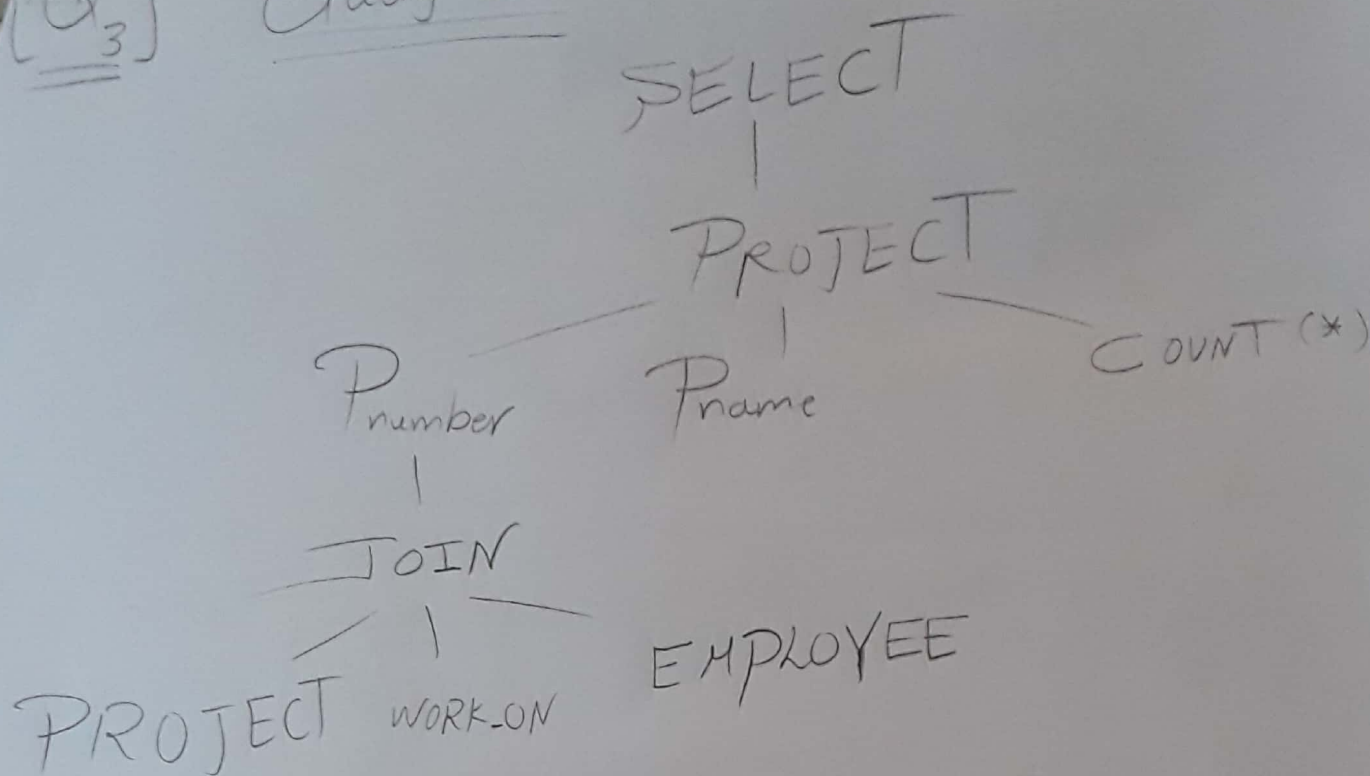
\* The query optimization heuristic algorithm (here) might not required any changes to the initial query since it is already optimized.



[Q<sub>3</sub>]

Query tree:

3



\* we can apply optimization using heuristic algorithm

by

- index usage
- Join re-ordering
- aggregation Push down
- and so on ...

$$\textcircled{2} * \text{Cost}_{\text{PROJECT}} = O(n_{\text{attributes}} * r_{\text{Size of relation}})$$

$$* \text{Cost}_{\text{UNION}} = O(r_1_{\text{Size}} + r_2_{\text{Size}} + \dots)$$

$$* \text{Cost}_{\text{INTERSECTION}} = O(\min(r_1, r_2, \dots))$$

$$* \text{Cost}_{\text{SET DIFFERENCE}} = O(r_1)$$

$$* \text{Cost}_{\text{CARTESIAN PRODUCT}} = O(r_1_{\text{Size}} * r_2_{\text{Size}})$$

4

$$\textcircled{3} \text{ Cost}_{\text{JOIN}} = O \left( r_1 \text{ size} * r_2 \text{ size} \right)$$

\* assuming: Simple nested loop Join algorithm  
 by matching each row from the 1<sup>st</sup> relation  
 with " " " " 2<sup>nd</sup> "

\* this will vary depending on  $\left\{ \begin{array}{l} \text{Join algorithm} \\ \text{Presence of indexes} \\ \text{DBMS optimization} \\ \text{technique} \end{array} \right.$