DATABASE SYSTEMS II (IS313P)

LECTURE NO. 09

06/12/2023

QUERY PROCESSING AND OPTIMIZATION

PART [2]

Consider the following schemas for next examples

EMPLOYEE

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno

DEPARTMENT

Dname	Dnumber	Mgr_ssn	Mgr_start_date

DEPT_LOCATIONS

Dnumber	Dlocation

PROJECT

Pname	Pnumber	Plocation	Dnum

WORKS_ON

Essn	Pno	Hours

DEPENDENT

Essn	Dependent_name	Sex	Bdate	Relationship

USING HEURISTICS IN QUERY OPTIMIZATION

- The main heuristic is to apply first the operations that reduce the size of intermediate results.
  - ✓ Ex. Apply SELECT and PROJECT operations before applying the JOIN or other binary operations.

USING HEURISTICS IN QUERY OPTIMIZATION

Example:

For every project located in 'Stafford', retrieve the project number, the controlling department number and the department manager's last name, address and birthdate.

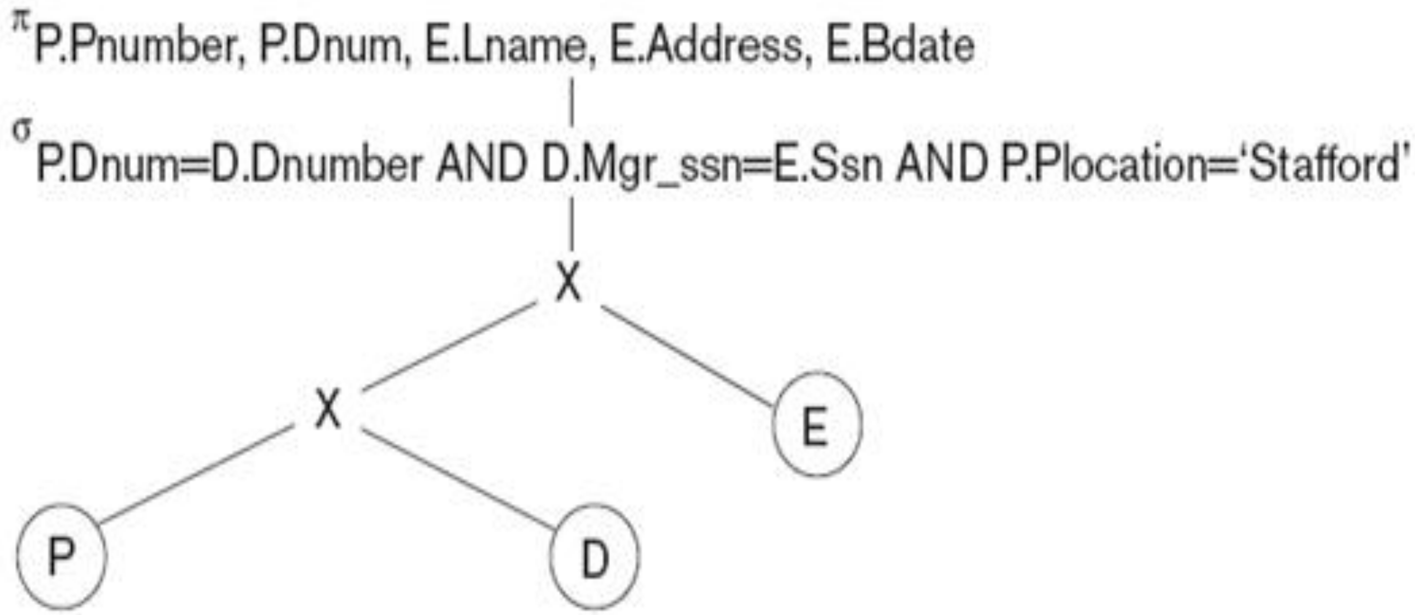
- **SQL query:**

```
SELECT  P.NUMBER,P.DNUM,E.LNAME, E.ADDRESS, E.BDATE
FROM    PROJECT AS P, DEPARTMENT AS D, EMPLOYEE AS E
WHERE   P.DNUM=D.DNUMBER AND D.MGR_SSN=E.SSN AND P.PLOCATION='STAFFORD';
```

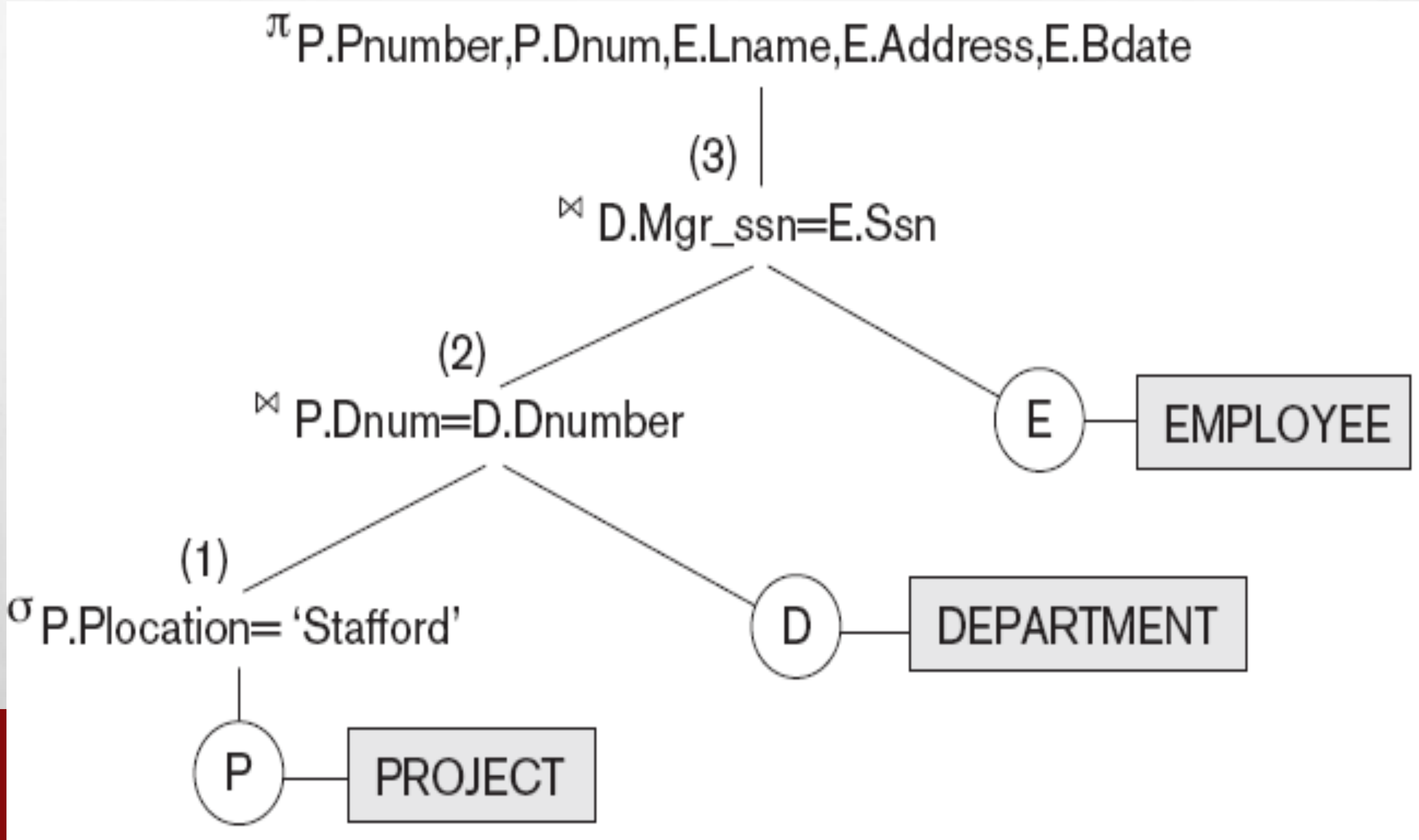
- **Relation algebra:**

π PNUMBER, DNUM, LNAME, ADDRESS, BDATE (((σ PLOCATION='STAFFORD'(PROJECT)) $\bowtie$ DNUM=DNUMBER (DEPARTMENT)) $\bowtie$ MGR_SSN=SSN(EMPLOYEE))

INITIAL (CANONICAL) QUERY TREE FOR SQL QUERY



QUERY TREE CORRESPONDING TO THE RELATIONAL ALGEBRA EXPRESSION



USING HEURISTICS IN QUERY OPTIMIZATION

- Heuristic Optimization of Query Trees:
  - The same query could correspond to many different relational algebra expressions and hence many different query trees.
  - The task of heuristic optimization of query trees is to find a final query tree that is **efficient to execute.**

- **Example:**

```
SELECT    LNAME
FROM      EMPLOYEE, WORKS_ON, PROJECT
WHERE     PNAME = 'AQUARIUS' AND PNMUBER=PNO
              AND ESSN=SSN AND BDATE > '1957-12-31';
```

USING HEURISTICS IN QUERY OPTIMIZATION

Steps in converting a query tree during heuristic optimization:

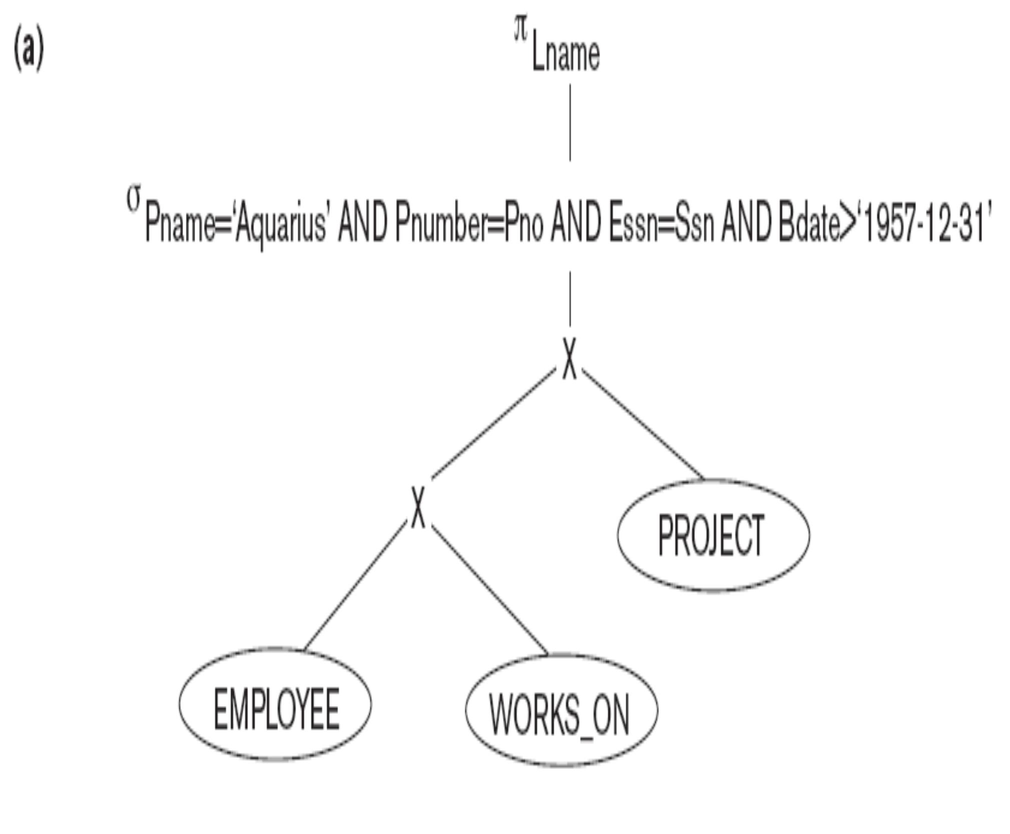
(a) Initial (canonical) query tree for SQL query Q.

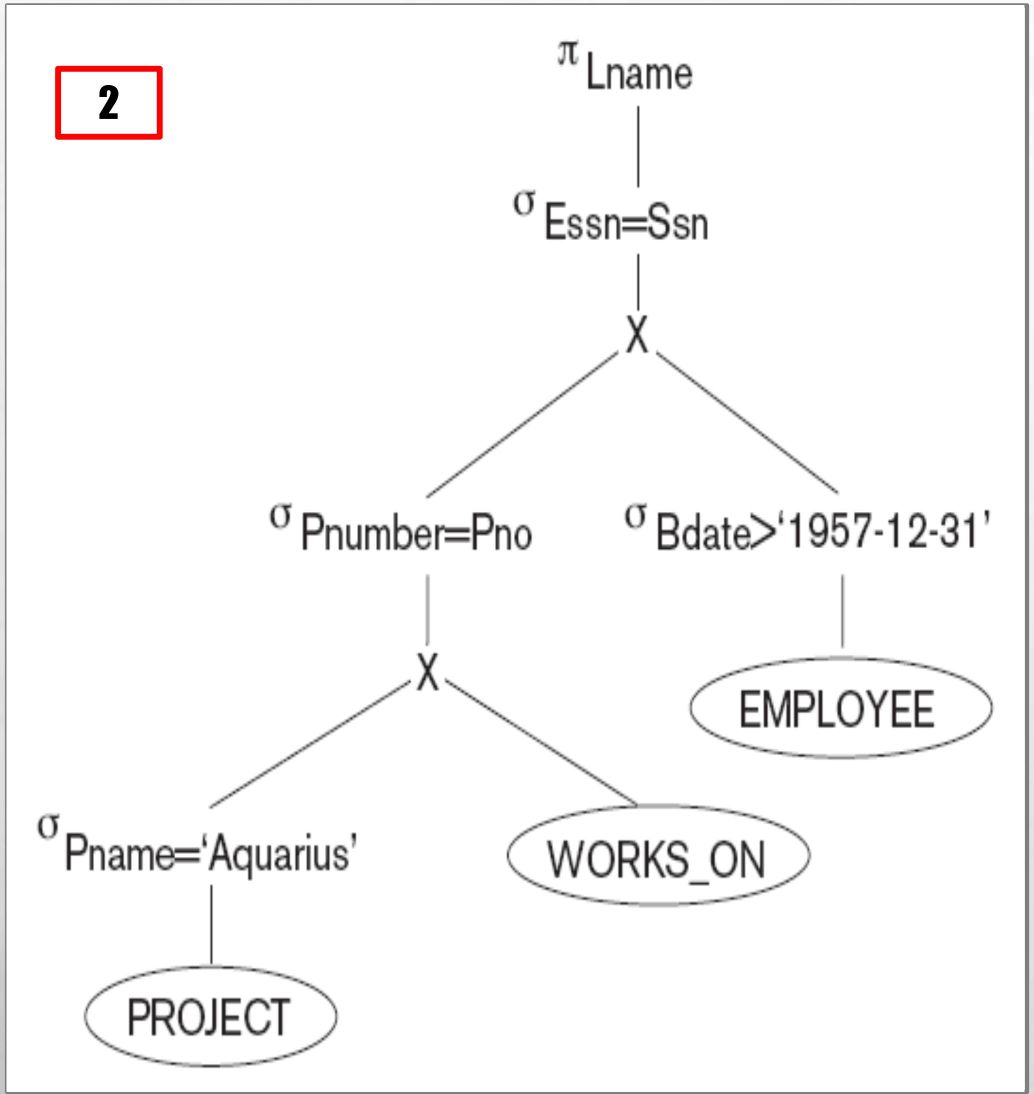
(b) Moving SELECT operations down the query tree.

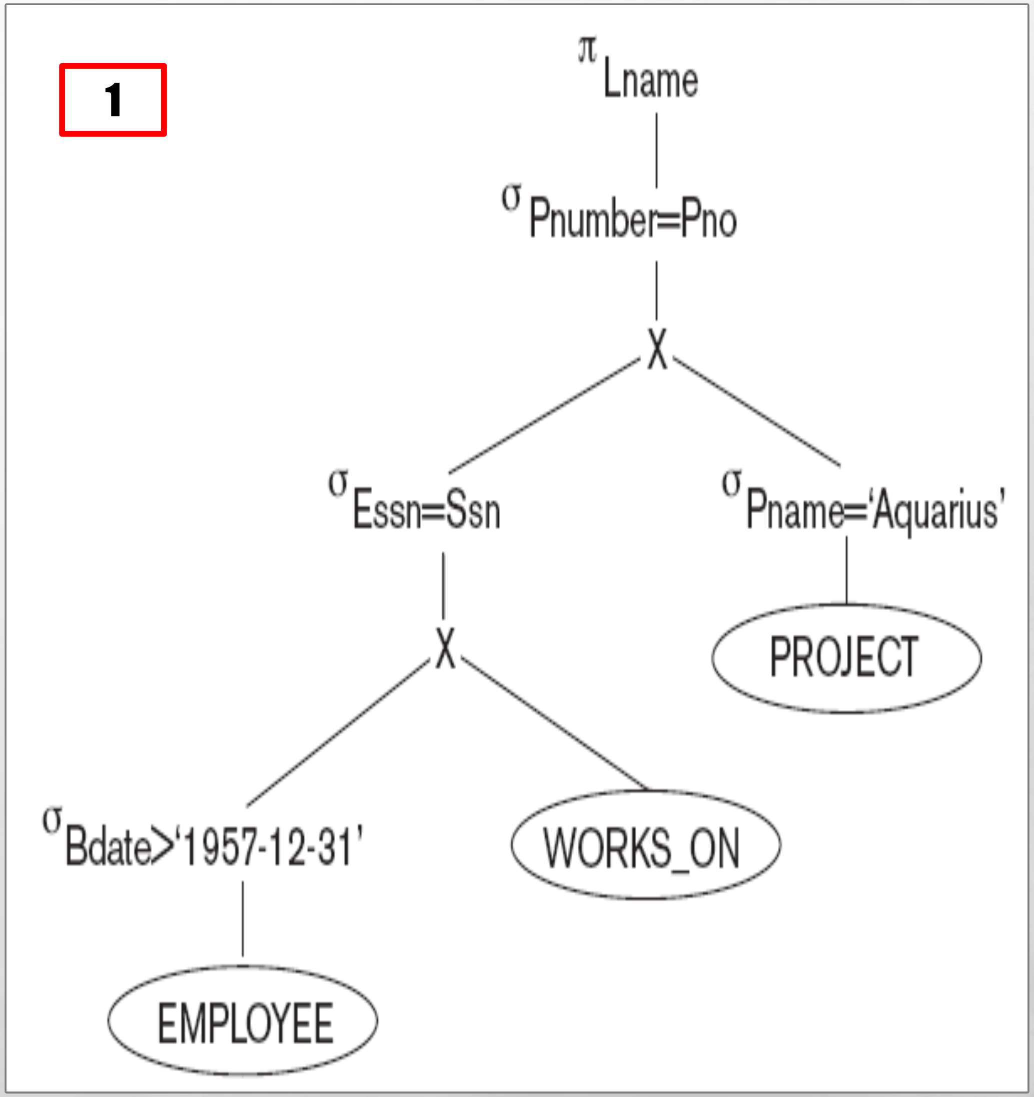
(c) Applying the more restrictive SELECT operation first.

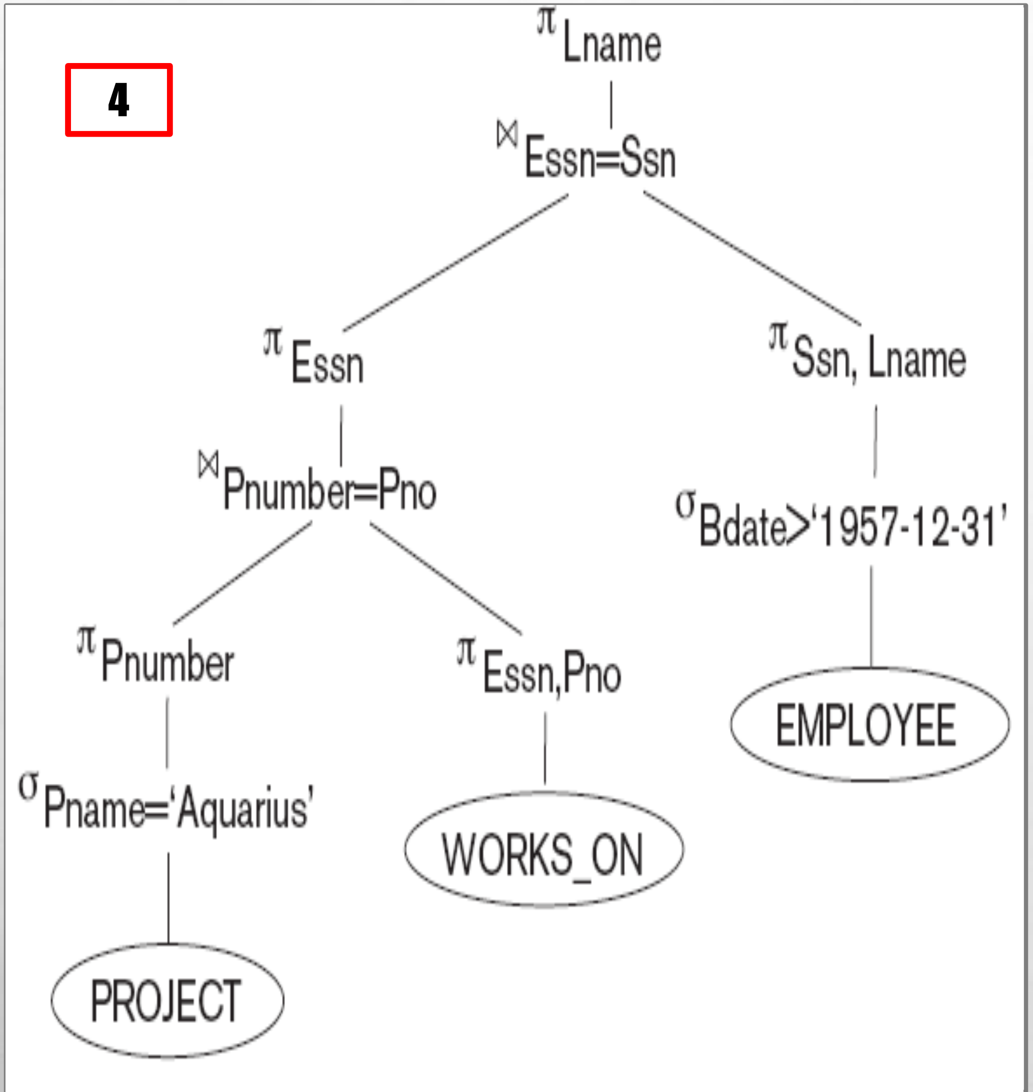
(d) Replacing CARTESIAN PRODUCT and SELECT with JOIN operations.

(e) Moving PROJECT operations down the query tree.

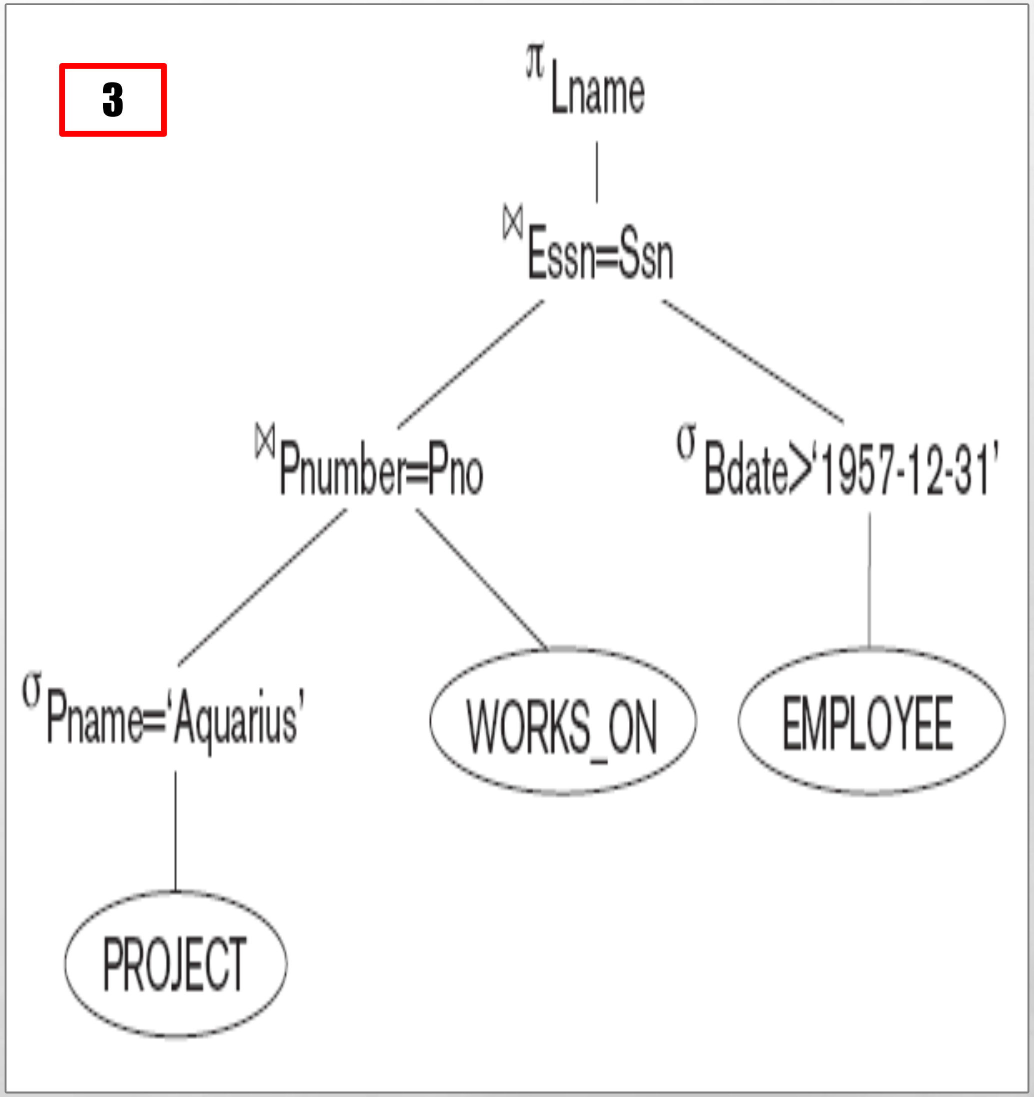



2
π Lname
σ Essn=Ssn
X
σ Pnumber=Pno
σ Bdate>'1957-12-31'
X
EMPLOYEE
σ Pname='Aquarius'
WORKS_ON
PROJECT


1
π Lname
σ Pnumber=Pno
X
σ Essn=Ssn
σ Pname='Aquarius'
X
PROJECT
σ Bdate>'1957-12-31'
WORKS_ON
EMPLOYEE



4
π Lname
⋈ Essn=Ssn
π Essn
π Ssn, Lname
⋈ Pnumber=Pno
σ Bdate>'1957-12-31'
π Pnumber
π Essn,Pno
EMPLOYEE
σ Pname='Aquarius'
WORKS_ON
PROJECT



3
π Lname
⋈ Essn=Ssn
⋈ Pnumber=Pno
σ Bdate>'1957-12-31'
σ Pname='Aquarius'
WORKS_ON
EMPLOYEE
PROJECT

USING HEURISTICS IN QUERY OPTIMIZATION

General Transformation Rules for Relational Algebra Operations:

1. Cascade of σ: A conjunctive selection condition can be broken up into a cascade (sequence) of individual **σ** operations:

- $\sigma_{c1 \text{ AND } c2 \text{ AND } \ldots \text{ AND } cn}(R) = \sigma_{c1}(\sigma_{c2}(\ldots(\sigma_{cn}(R))\ldots))$

2. Commutatively of σ : The s operation is commutative:

- $\sigma_{c1}(\sigma_{c2}(R)) = \sigma_{c2}(\sigma_{c1}(R))$

3. Cascade of π : In a cascade (sequence) of π operations, all but the last one can be ignored:

- $\pi_{List1}(\pi_{List2}(\ldots(\pi_{Listn}(R))\ldots)) = \pi_{List1}(R)$

4. Commuting σ with π : If the selection condition c involves only the attributes A1, ..., An in the projection list, the two operations can be commuted:

- $\pi_{A1, A2, \ldots, An}(\sigma_c(R)) = \sigma_c(\pi_{A1, A2, \ldots, An}(R))$

USING HEURISTICS IN QUERY OPTIMIZATION

5- Commutativity of $\bowtie$ (and x): The $\bowtie$ operation is commutative as is the x operation:

$$R \bowtie_c S = S \bowtie_c R; \quad R \times S = S \times R$$

6. Commuting σ with $\bowtie$ (or x): If all the attributes in the selection condition c involve only the attributes of one of the relations being joined, then the two operations can be commuted as follows:

$$\sigma_c (R \bowtie S) = (\sigma_c (R)) \bowtie S$$

- Alternatively, if the selection condition c can be written as (c1 and c2), where condition **c1 involves only the attributes of R** and condition **c2 involves only the attributes of S**, the operations commute as follows:

$$\sigma_c (R \bowtie S) = (\sigma_{c1} (R)) \bowtie (\sigma_{c2} (S))$$

USING HEURISTICS IN QUERY OPTIMIZATION

General Transformation Rules for Relational Algebra Operations

7. **Commuting π with $\bowtie$ (or x):** Suppose that the projection list is L = {A1, ..., An, B1, ..., Bm}, where A1, ..., An are attributes of R and B1, ..., Bm are attributes of S. If the join condition c involves only attributes in L, the two operations can be commuted as follows:

$$\pi_L (R \bowtie_c S) = (\pi_{A1, ..., An} (R)) \bowtie_c (\pi_{B1, ..., Bm} (S))$$

- **If the join condition C contains additional attributes not in L, these must be added to the projection list**, and a final $\bowtie$ operation is needed.

USING HEURISTICS IN QUERY OPTIMIZATION

General Transformation Rules for Relational Algebra Operations

8. **Commutativity of set operations**: The set operations $\cup$ and $\cap$ are commutative but **"–"** is not.
9. **Associativity of $\bowtie$, X, $\cup$, and $\cap$:** These four operations are individually associative; that is, if θ stands for any one of these four operations (throughout the expression), we have

$$(R \, \theta \, S) \, \theta \, T = R \, \theta \, (S \, \theta \, T)$$

10. **Commuting σ with set operations:** The σ operation commutes with $\cup$, $\cap$, and –.

   If θ stands for any one of these three operations, we have

$$\sigma_c (R \, \theta \, S) = (\sigma_c (R)) \, \theta \, (\sigma_c (S))$$

USING HEURISTICS IN QUERY OPTIMIZATION

General Transformation Rules for Relational Algebra Operations

11- The π operation commutes with $\cup$.

$$\pi_L (R \cup S) = (\pi_L (R)) \cup (\pi_L (S))$$

12- Converting a (σ, x) sequence into $\bowtie$:

If the condition c of a σ that follows a **x** Corresponds to a join condition, **convert the (σ, x)** sequence into a $\bowtie$ as follows:

$$(\sigma_c (R \times S)) = (R \bowtie_c S)$$

QUIZ

For the following schema,

a) Write SQL query to find the First names, Last names, and addresses for all employees works on Research department.

b) Write the Relational Algebra Expression corresponding to the query in part a.

c) Draw the query tree for the optimized solution.

EMPLOYEE

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno

DEPARTMENT

Dname	Dnumber	Mgr_ssn	Mgr_start_date

DEPT_LOCATIONS

Dnumber	Dlocation

PROJECT

Pname	Pnumber	Plocation	Dnum

WORKS_ON

Essn	Pno	Hours

DEPENDENT

Essn	Dependent_name	Sex	Bdate	Relationship

ANY QUESTIONS ??

THANK YOU