University of El Oued — Department of Computer Science
2nd Y Master — Academic Year: 2023/2024

Correction of Final Exam of Distributed Databases

Exercise 01 (7 marks=2.5+3+1.5)

1) Given a table **T** stored on N pages having a B+tree index on the key **id** of high=4. What is the cost of the following queries?

a) Select * from T where id=5
   4 (search the key 5 in the index) + 1 read of data page = 5 IO
b) Select * from T where name='abc'
   Read all pages from the data file. Cost = N IO
c) Update T set name='def' where id=5
   4 (search the key 5 in the index) + 1 read of a data page + 1 write of data page = 6 IO

2) Consider the relation Student(id, name, age)having a B+-tree index on the attribute **name**. How to handle the following queries?

A. σ name="Sami" (Student)
   Search using the index on the attribute name
B. σname≠"Sami" (Student)
   Read all data pages (sequential scan)
C. σ id=2(Student)
   Read all data pages (sequential scan)
D. σ id=2 ∧ name="Sami"(Student)
   Search using the index on the attribute name the check the condition about id

3) Given a relation R that is horizontally fragmented into R1, R2, …, Rn. How to define completeness, disjointness and reconstruction of this fragmentation?

completeness R1 ∪ R2 … U Rn = R

disjointness R1 ∩ R2 … ∩ Rn = Ø

reconstruction R1 ∪ R2 … U Rn = R

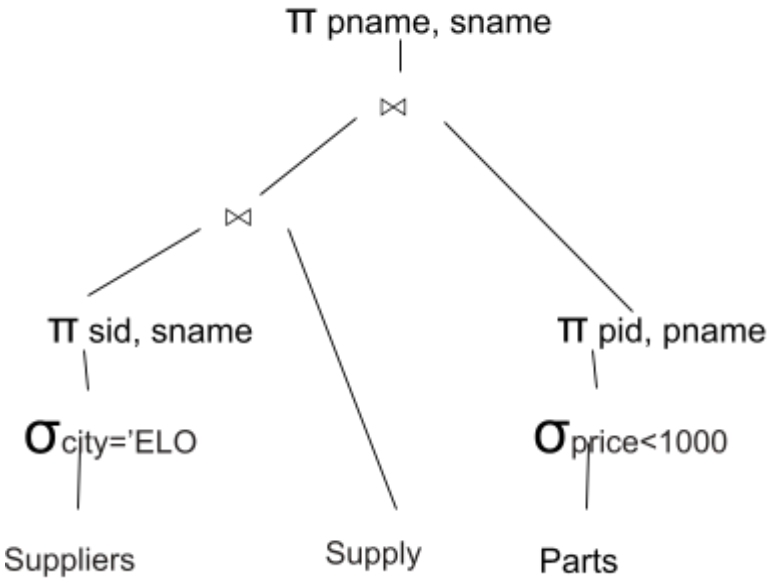
Exercise 02 (7 marks=1+3.5+2.5)

Consider the following relational schema and SQL query:

Suppliers(*sid:* `integer`, *sname:* `char(20)`, *city:* `char(20)`)
Supply(*sid:* `integer`, *pid:* `integer`)
Parts(*pid:* `integer`, *pname:* `char(20)`, *price:* `real`)

```
SELECT    S.sname, P.pname
FROM Suppliers S, Supply Y, Parts P
WHERE     Y.sid=S.sid AND Y.pid=P.pid AND S.city='ELO' AND P.prices<1000
```

1. Propose an optimized query execution plan for this query.



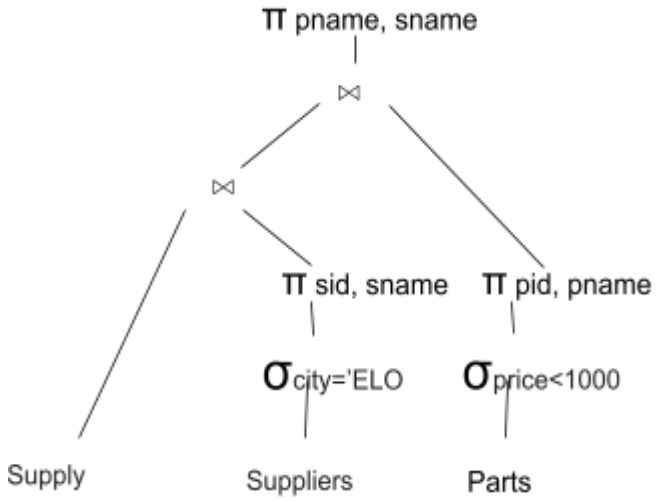
2. What information about these relations will the query optimizer need to estimate the cost of the query execution plan?

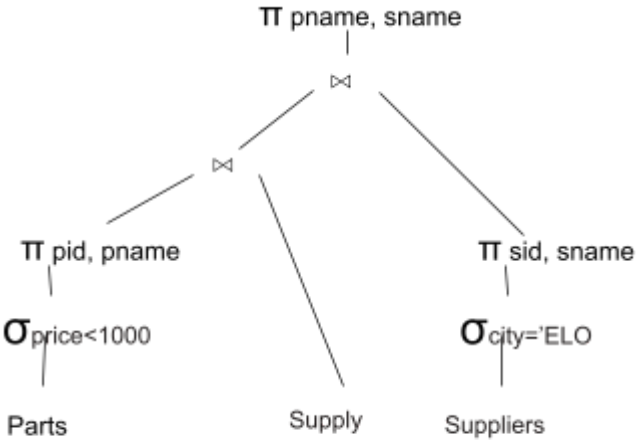
   The information needed

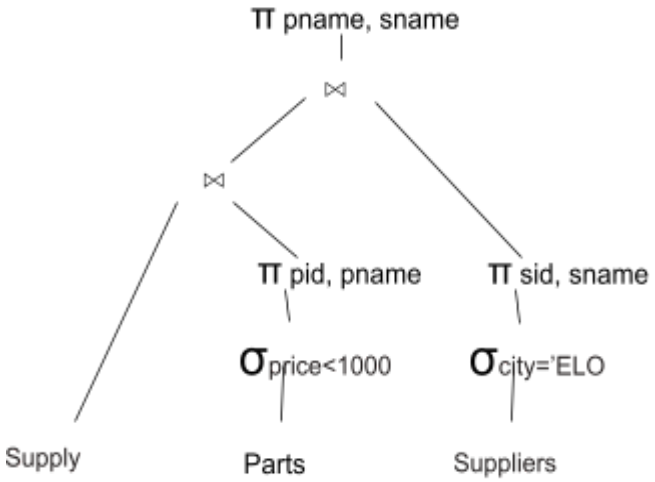
   - **Size of page and number of tuples per page**
   - **Number of tuples and pages of all tables (Suppliers, Supply, Parts)**
   - **Selectivity factors of all selection (SF(city='ELO), SF(price<1000)) or statistical information to calculate them.**
   - **The size of attributes used in projection (sid, sname, pid, pname)**
   - **The number of tuples resulting from the first join**

3. List various join orders that the query optimizer considers when estimating costs.

   The optimizer considers only left deep join







Exercise 03 (6 marks)

Given the following schema and SQL query:

Student (sid, name, age, address)
Book(bid, title, author, library)
Checkout(sid, bid, date)

```
SELECT   S.name
FROM     Student S, Book B, Checkout C
WHERE    S.sid = C.sid AND B.bid = C.bid AND   B.library = 'lib1' AND S.age > 12
```

Given that **Book** relation is Horizontally fragmented into 3 fragments according to the attribute library ('lib1', 'lin2', 'lib3'). **Student** and **checkout** are fragmented into 3 fragments according to student identifier (sid<1000,1000<=sid<=2000, side>2000).

Q) Write the optimized distributed query execution plan using algebraic optimization.

$$Q= ((\sigma_{library = 'lib1'}(Book))\bowtie Checkout)\bowtie(\sigma_{age > 12}(Student))=$$
$$((\sigma_{library = 'lib1'}(Book1 \cup Book2 \cup Book3))$$
$$\bowtie(Checkout1 \cup Checkout2 \cup Checkout3))\bowtie(\sigma_{age > 12}(Student1 \cup Student2 \cup Student3))$$

Given that Book relation is Horizentally fragmented into 3 fragments according to the attribute library ('lib1', 'lin2', 'lib3')

$$\text{So } (\sigma_{library = 'lib1'}(Book1 \cup Book2 \cup Book3)) = (\sigma_{library = 'lib1'}(Book1))$$

$$Q=(\sigma_{library = 'lib1'}(Book1))\bowtie(Checkout1 \cup Checkout2 \cup Checkout3))\bowtie$$
$$(\sigma_{age > 12}(Student1 \cup Student2 \cup Student3))$$

Given that **Student** and **checkout** are fragmented into 3 fragments according to student identifier (sid<1000,1000<=sid<=2000, side>2000) and by distribution of union ver join we find:

$$Q=(\sigma_{\text{library = 'lib1'}}(\text{Book1}))\bowtie(\text{Checkout1}\bowtie(\sigma_{\text{age > 12}}(\text{Student1}))\cup \text{Checkout2}\bowtie(\sigma_{\text{age > 12}}(\text{Student2}))\cup \text{Checkout3}\bowtie(\sigma_{\text{age > 12}}(\text{Student3})))$$