University of El Oued — Department of Computer Science
1st Y Master IoT/AI — Academic Year: 2023/2024

Correction of Final Exam of Advanced Databases

Exercise 01 (6 marks=1.5+1.5+1+1+1)

What is the estimated cost of I/O accesses required in the following situations?

1. **Select * from S Order By id;** (order by uses the sorting algorithm) the table is stored on [S] pages with a buffer of size B pages such as B>=[S].
   The cost of this query tends to be the cost of sorting the table. While the buffer can read all pages of S at a time (B>=[S]) and the sorting will be done in memory, the cost is [S]
2. **Select * from S natural join R;** the tables are stored on [S] and [R] pages respectively. The used join algorithm is block nested loop with a buffer of size B pages such as B>[S]+2
   We can read all pages of [S] into the buffer and then read the pages of [R] one by one. Therefore, the cost is [S] + [R].
3. **Select * from S;** assume there is no index (estimate the worst case)
   Read all the pages, the cost is [S]
4. **Select * from S where id=10;** assume that there is no index (estimate the worst case)
   Read all the pages, the cost is [S]
5. **Select * from S where name='abc';** assume that there is a B+-tree index on the attribute name with high=3 levels
   Read 3 nodes (pages) of the index + 1 page of data. The cost is 4

Exercise 02 (6 marks=3+3)

Consider the B+ tree index of order d = 2 used in the course (root and intermediate nodes: 5 pointers and 4 keys. Leaf node: 4 entries)

1) Construct a B+-tree for the following set of values (2, 3, 5, 7, 11, 17, 19, 23, 29, 31, 35, 38, 42, 47, 51). Assume that the tree is initially empty and values are inserted in ascending order.
   Annex 01
2) Show the form of the B+-tree after each operation of the sequence: Insert 9; Insert 10; Insert 8; Insert 6; Insert 1; Insert 4
   Annex 02

Exercise 03 (8 marks)

Consider the following relational schema and SQL query:

$$\text{Suppliers}(\underline{sid:\ \texttt{integer}},\ sname:\ \texttt{char(20)},\ city:\ \texttt{char(20)})$$
$$\text{Supply}(\underline{sid:\ \texttt{integer},\ pid:\ \texttt{integer}})$$
$$\text{Parts}(\underline{pid:\ \texttt{integer}},\ pname:\ \texttt{char(20)},\ price:\ \texttt{real})$$

and the following 2 queries

```
SELECT *
FROM Supply Y, Suppliers S
WHERE Y.sid=S.sid AND S.city='ELO'
```

```
SELECT *
FROM Supply Y, Parts P
WHERE Y.pid=P.pid AND P.prices<1000
```

And assuming

- suppliers are 10000 tuples stored on 500 pages
- supply are 60000 tuples stored on 1500 pages
- parts are 14000 tuples stored on 700 pages
- 35% of parts have a price less than 1000 (SF(prices<1000)= 0.35)
- 7% of suppliers are from the city 'ELO' (SF(city='ELO')= 0.07)

Q) Estimate the cost of both queries using block nested loop join given buffer = 12 pages

```
SELECT *
FROM Supply Y, Suppliers S
WHERE Y.sid=S.sid AND S.city='ELO'
```

Given

- suppliers are 10000 tuples stored on 500 pages so 10000/500=20 tuples/page

The plan

$$(\sigma_{city='ELO'}(Suppliers)) \bowtie Supply$$

for the branch $(\sigma_{city='ELO'}(Suppliers))$
# of tuples = 10000*0.07=700
# of pages = 700/20 =35

The cost of $(\sigma_{city='ELO'}(Suppliers)) \bowtie Supply$
35 + 35/(12-2) * 1500 =
35 + 4 *1500= 6035

Query 02

```
SELECT *
FROM Supply Y, Parts P
WHERE Y.pid=P.pid AND P.prices<1000
```

Given

- parts are 14000 tuples stored on 700 pages, so 14000/700=20 tuples/page

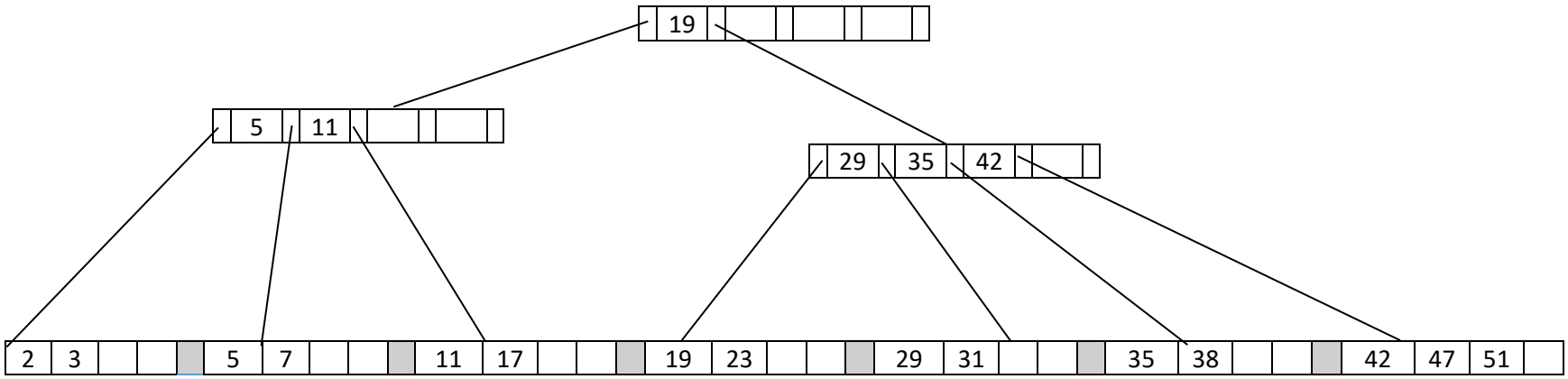
The plan

$$(\sigma_{prices<1000}(Parts)) \bowtie Supply$$

for the branch $(\sigma_{prices<1000}(Parts))$
# of tuples = 14000*0.35=4900
# of pages = 4900/20 =245

The cost of $(\sigma_{prices<1000}(Parts)) \bowtie Supply$
245+ 245/(12-2) * 1500 =
245+ 25 *1500= 37745

Annex 01



Annex 02