Introduction to Database Systems

2020/21

1st exam, FAMNIT

You can use literature but not a notebook. The time available for the exam is 120 minutes.

Good luck!

NAME AND SURNAME:	
STUDENT NUMBER:	
PROGRAM:	
SIGNATURE:	

The company SPJ manages the orders of the parts from the suppliers for the projects that take part at other companies. A project is described with the project identifier, name, company where the project takes part, budget, start date, and the duration of the project. A part is presented with the part identifier, name, and price. A supplier is described by the supplier identifier, name, address, and phone number. For each order of a product from a given supplier for a given project, we store the number of parts ordered, price of the order, date, and comment.

The information system SPJ includes the following tables.

Products(pid, name, price);
Projects(jid, name, company, budget, start, duration);
Suppliers(sid, name, address, phone);
Orders(oid, pid, jid, sid, quantity, price, date, comment);

The following information describes the tables.

The page size is 4000 bytes!

|Products| = 25 bytes, 160 records/page, 50 pages, 8000 records
|Projects| = 100 bytes, 40 records/page, 100 pages, 4000 records
|Suppliers| = 50 bytes, 80 records/page, 100 pages, 8000 records
|Orders| = 100 bytes, 40 records/page, 25000 pages, 1000000 records

Exercise 1 (25%)

a) Write a query that prints the names of parts ordered from the supplier named »The supplier« for the projects that started after 1.1.2020. Use relational algebra and relational calculus!

b) Write an SQL query that prints the names and identifiers of the suppliers and the aggregate value of their orders for the suppliers that supplied in the year 2020 the products with the aggregate value grater than 500.000 EUR.

c) Describe in few sentences the relational completeness of query languages.

Exercise 2 (25%)

The following indexes are given:
- a B+ tree on the attribute `Orders.date`, and
- hash index on the attribute `Parts.pid`.

Two algorithms for the join operation are implemented:
- Index nested loops join, and
- Sort-merge join.

Suppose that the tables are ordered by keys.

```
Join[Parts.pid=Orders.pid](
    Select[date>=1.1.2019](Orders),
    Parts)
```

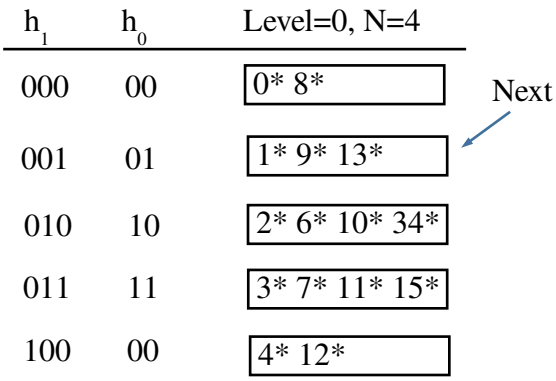
a) How many blocks do the query read from the disk if we use Index nested loops join? The relation `Orders` is the outer, and the relation `Parts` is the inner relation!

b) Which algorithm is faster? Give a brief explanation of the answer.

c) Describe the Hash join algorithm in three sentences.

Exercise 3 (25%)

a) An example of the linear hash index is given below. Every time a new overflow page is added, the bucket pointed by Next is split, and Next is incremented by one. Show the state of the index after adding the keys 18, 16, and 27.



b) Describe the extendible hash index in few sentences. You can describe it by relating it to the linear hash index.

Exercise 4 (25%)

a) Let $F_R = \{X \rightarrow Y, YZ \rightarrow W\}$ be a set of functional dependencies that hold in a given relation $R(X, Y, Z, W, Q)$.

- Using Armstrong's axioms prove that the functional dependency $XZ \rightarrow WY$ holds in R.
- In which normal form is the schema of the relation R?
- Decompose the schema of the relation R into the Boyce-Codd normal form. Does your solution retain functional dependencies?

b) Present a definition of the lossless join.