MDS
Winter Term 2024/25
Basics of Query Optimization

This unit is based on content from Jens Dittrich's book "Patterns in Data Management: A Flipped Textbook" from 2016.

1. What is the task of the query optimizer in a database system?
   - ◯ Parsing queries
   - ◯ Answering queries
   - ◯ Defining a plan to efficiently answer a query
   - ◯ Gathering statistics about the database
2. What is the output of the query optimizer?
   - ◯ The result of the given query
   - ◯ The parse-tree of the query
   - ◯ A plan that tells you how a query should be executed efficiently
   - ◯ Statistics of the tables touched by the query
3. What kind of information do you think is useful to the query optimizer?
   - ◯ Distribution of attribute values
   - ◯ Record count for each table
   - ◯ The number of user accounts in the database
   - ◯ The available size of the main memory ("buffers" in database terms)
4. What is the main task of the rule-based optimizer?
   - ◯ To apply certain pre-defined rules that could optimize the query plan
   - ◯ To analyze the current workload and optimize the query plans based on that information
   - ◯ To efficiently parse the queries
5. Given two queries Q1 and Q2. Assume a rule-based optimizer is called on Q1 and then Q2. The result of the calls may lead to different plans, even though
   - ◯ the strings of Q1 and Q2 are equal
   - ◯ the strings of Q1 and Q2 are equal, except: some constants in the WHERE-clauses differ
   - ◯ Q1 is semantically equivalent to Q2, i.e. it returns the same set of results for any instance of the database that has the same schema.

6. Could the rule-based optimizer actually have a negative impact on query performance, rather than a positive one?

   ◯ Yes ◯ No

7. When performing a join operation in a relational database management system, does the order of the operands have anyimpact on the performance of the join in general?

   ◯ Yes ◯ No

8. When the query optimizer is trying to decide how to perform a join operation, what kind of operation fits better?

   ◯ Cost-based optimization

   ◯ Rule-based optimization

9. You are given the following SQL query:

```
SELECT DISTINCT A.b, B.c, C.d, D.e
FROM A, B, C, D
WHERE A.x=B.x AND A.a<24 AND A.c=C.d AND B.d=D.b
  AND D.a>24 AND C.d=D.c AND C.b>45 AND B.x=23
```

   Write down the canonical translation of this query into relational algebra.

   Use rule-based optimization to improve the logical plan. Explain what you did and draw the new plan as a relational algebra tree.

10. What is the rough idea of a cost-based query optimizer?

    ◯ Explore the search space to obtain a good way of performing the given operation under a given cost model.

    ◯ Apply certain predefined rules that are known to potentially give good performance.

11. What is the number of possible left-deep plans for n input relations?