Datenbanken II
Übungsblatt 8 – WiSe 2015/16

1. This is exercise 15.7 taken from the book "*Database Management Systems (3rd Edition)*"written by *R. Ramakrishnan* and *J. Gehrke*. Helpful information can be found throughout all the slides of unit 4 (*Anfrageoptimierung*).

   Consider the following relational schema and SQL query. The schema captures information about employees, departments, and company finances (organized on a per department basis).

   Emp(eid, did, sal, hobby)
   Dept(did, dname, floor, phone)
   Finance(did, budget, sales, expenses)

   Consider the following query:

   ```
   SELECT D.dname, F.budget
   FROM Emp E, Dept D, Finance F
   WHERE E.did=D.did AND D.did=F.did AND D.floor=1
   AND E.sal >= 59000 AND E.hobby = 'yodeling'
   ```

   a) Identify a relational algebra tree that reflects the order of operations that results in a possibly efficient execution of the query.

   b) List the join orders which an optimizer should consider (i.e., orders in which pairs of relations can be joined together to compute the query result). We assume that the optimizer follows the heuristic of never considering plans that require the computation of cross-products. Briefly explain how you arrived at your list.

   c) Suppose that the following additional information is available: Unclustered B$^+$tree indexes exist on Emp.did, Emp.sal, Dept.floor, Dept.did, and Finance.did. The system's statistics indicate that employee salaries range from 10,000 to 60,000, employees enjoy 200 different hobbies, and the company owns two floors in the building. There are a total of 50,000 employees and 5,000 departments (each with corresponding financial information) in the database. The DBMS used by the company has just one join method available, namely, index nested loops.

      i. For each of the query's base relations (Emp, Dept, and Finance) estimate the number of tuples that would be initially selected from that relation if all of the non-join predicates on that relation were applied to it before any join processing begins.

      ii. Given your answer to the preceding question, which of the join orders has the least estimated cost?