Uni Freiburg, Web Science Group
Prof. Peter Fischer
Systems Infrastructure for Data Science - Winter 2014/15

Exercise Sheet #6: Query Optimization

Dec 4, 2014

Exercise 6.1 : Dynamic Programming

Consider the following SQL query in a hypothetical flight reservation database:

```
SELECT A.name, F.airline, C.name
FROM Airport A, Flight F, Crew C
WHERE F.to = A.code
AND F.flight_no = C.flight_no
```

The database performs System R-style join optimization (physical dynamic programming). The optimizer generates left-deep plans only, avoids cross-products, and keeps plans with interesting order as discussed in the lecture

The available join algorithms are block-nested loop join (BNLJ) and index-nested loop join (INLJ). Assume there is a clustered B+tree on Airport.code. Assume that an INLJ is always faster than a BNLJ, if an index is available.

A. Show all the 1-relation plans considered by the optimizer in pass 1. Show which of these plans are retained and why.

B. Show all the 2-relation plans considered in pass 2. Show which of these plans are retained and why.

C. Show all the 3-relation plans considered in pass 3. You need not decide which of these is the best plan overall.

Exercise 6.2 : Query Planning

Consider the database containing the following tables:

$Courses(\underline{ID}, Name, Room, Time)$
$Exercises(\underline{ID}, C\_ID, A\_ID, Room, Time)$
$Assistants(\underline{ID}, Firstname, Lastname)$

and the query:

```
SELECT C.Name, A.Firstname, A.Lastname, E.Room, E.Time
FROM Courses C, Assistants A, Exercises E
WHERE C.ID = E.C_ID AND A.ID = E.A_ID AND C.Room like '10%' AND E.Room not like 'CAB%';
```

A. Show the logical canonical form for this query expressed in relational algebra. The logical canonical form is the straight-forward translation of the SQL query into a relational algebra query plan containing only selection, projection and cross product operators.

B. Perform Query Rewrite on part [A]. Which rules would you apply to get to this intermediate representation? Show the logical query plan of the resulting query.

C. Assume that there is a non-clustered index on $C.Room$ and a clustered, direct index on $E.A\_ID$. Based on part [B], show possible physical query execution plans (i.e., access paths, join implementations) that take advantage of the indexes. Discuss the efficiency of your plans based on the following cost metrics: disk I/O, intermediate result size.