Exercise 9 - Cost-based Optimization

1 Response Time vs. Total Time

Given the example from task 4 on exercise sheet 8 (multi-way join): compute the response and total time for the following two considered strategies:

1. Ship Whole

2. Semi-Join with join order: Employee $\rightarrow$ Assignment $\rightarrow$ Project

For this purpose, the following is known: each attribute value is 10 bytes in size, i.e. each tuple is $10 * numberOfcolumns$ and a join column is $10 * numberOfValues$ in size. The transfer of one Byte takes one millisecond, and each message has a minimum size of 300 Bytes for network packaging. Processor and I/O times can be neglected.

2 Query Optimization Search Space

We consider the following simple join query:

$$R \bowtie S \bowtie T \bowtie U$$

The relations are connected via primary/foreign key relationships along the given join order and are stored each on a separate node. The query is initiated on node N_R where the result must be returned. Furthermore, each join can be computed by one of the following algorithms implemented in the used DBMS:

- Nested Loop (NL) / Ship Whole (SW)

- NL / Fetch as Needed (FAN)

- Sort Merge (SM) / SW

- SM / FAN

- Semi-Join

- Bit Vector-Join

Joins can be processed on every node. The only considered constraint is that only Left Deep Trees are allowed for the join order. How many possible query plans result from that approach? How could the number of possible query plans be reduced by allowing only query plans without Cartesian products?

3 Selectivity Factor

How can the selectivity factor of the selection condition below be estimated? Which information are required for that?

```
SELECT *
FROM Student
WHERE (courseOfStudies='INF' OR courseOfStudies='WIF')
  AND NOT ( dateOfBirth < '1.1.1980' AND averageGrade > 1.5 )
```