INF3100: Databasesystemer – Oppgavesett 11

Oppgave 16.3.2: Replace the natural joins in the following expression by equivalent theta-joins and projections. Tell whether the resulting theta-joins form a commutative and associative group.

a) $(R(a,b) \bowtie S(b,c)) \bowtie (T(c,d) \bowtie U(d,e))$

b) $(R(a,b) \bowtie S(b,c)) \bowtie (T(c,d) \bowtie U(a,d))$

c) $(R(a,b) \bowtie S(b,c)) \bowtie_{R.a<T.c} T(c,d)$

Oppgave 16.6.4: Consider the join of relations $R(a,b)$, $S(b,c)$, $T(c,d)$, and $U(a,d)$, where R and U each have 1000 tuples, while S and T each have 200 tuples. Further, there are 200 values of all attributes of all relations, except for attribute c, where $V(S,c) = V(T,c) = 20$.

a) What is the order selected by the greedy algorithm? What is its cost?

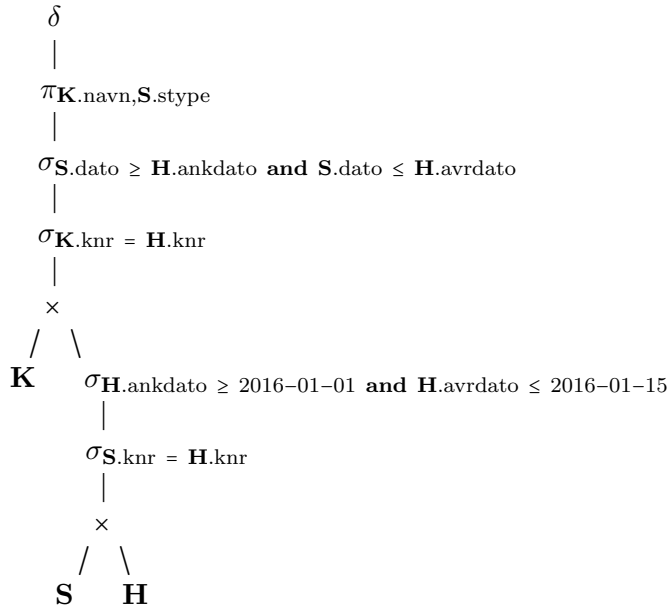
b) What is the optimum join ordering and its cost?

Oppgave 16.x.2: Gitt en hotelldatabase med relasjonene

Kunde(knr, navn, adr)
Hotellrom(knr, ankdato, avrdato, rom, pris)
Service(knr, dato, stype, pris)

Kunde og **Hotellrom** burde være rimelig selvforklarende. **Kunde** har primærnøkkel (knr) og **Hotellrom** primærnøkkel $(\text{knr}, \text{ankdato})$. **Service** inneholder informasjon om servicetilbud som hotellgjestene har benyttet seg av, som spa, frisør, restaurantbesøk mm., med prisinformasjon (som blir lagt til den endelige hotellregningen ved avreise). **Service** har primærnøkkel $(\text{knr}, \text{dato}, \text{stype})$.

I relasjonsalgebrauttrykket under er **K**, **H** og **S** forkortelser for henholdsvis **Kunde**, **Hotellrom** og **Service**.



a) Hva uttrykker spørringen?

b) Optimaliser uttrykket.

Oppgave 15.2.3: Give one-pass algorithms for each of the following join-like operators:

a) $R \overset{o}{\bowtie}_L S$, assuming R fits in memory (see Section 5.2.7 for definitions involving outerjoins).

b) $R \overset{o}{\bowtie}_L S$, assuming S fits in memory.

Oppgave 15.6.1: Suppose $B(R) = 10,000$ and $T(R) = 500,000$. Let there be an index on $R.a$, and let $V(R, a) = k$ for some number k. Give the cost of $\sigma_{a=0}(R)$ as a function of k, under the following circumstances. You may neglect disk I/O's needed to access the index itself.

a)The index is not clustering.

b) The index is clustering.

c) R is clustered, and the index is not used.

Oppgave 20.5.2: In this exercise, we need a notation for describing sequences of messages that can take place during a two-phase commit. Let (i, j, M) mean that site i sends the message M to site j, where the value of M and its meaning can be P (prepare), R (ready), D (don't

commit), C (commit), or A (abort). We shall discuss a simple situation in which site 0 is the coordinator, but not otherwise part of the transaction, and sites 1 and 2 are the components. For instance, the following is one possible sequence of messages that could take place during a successful commit of the transaction:

$$(0, 1, P), (0, 2, P), (2, 0, R), (1, 0, R), (0, 2, C), (0, 1, C)$$

a) Give an example of a sequence of messages that could occur if site 1 wants to commit and site 2 wants to abort.