

EXERCISE 10

Andrea Crotti (#299466), Can Liu (#286105), Sebastian Roidl (#281941)

12 Gennaio 2010

Exercise 10.1[Query optimization]:

1.

```
SELECT gift.name, supplier.name, COUNT(*)
FROM gift, supplier, shipment
WHERE shipment.gift = gift.gift_id AND shipment.supplier = supplier.supplier_id
AND gift.price > 190 AND supplier.place IN
(SELECT DISTINCT place.place_id
FROM employee, place
WHERE employee.place = place.place_id AND employee.nationality = 'German' )
GROUP BY gift.name, supplier.name
```

We first count how many tuples we could have in total.

$$shipment = gift * employee * supplier = 6000 * 10000 * 500 = 3 * 10^{10}$$

the general join is $gift \bowtie supplier \bowtie shipment$

$$So \#tuples = 10^{10}$$

a) The selection (IN expression) on supplier.place

We know that there are 50 different nationalities, assuming that they are uniformly distributed we get:

With the line *SELECT DISTINCT place.place_id* we get a selectivity of 1/1000, which are the number of possible distinct places.

Then we filter on the nationality and assuming they are equally distributed we filter 1/50. So the selectivity of the IN expression is:

$$1/50 * 1/1000 = 1/50000$$

b) The selection on gift.price

Prices are equally distributed in the range (11, 210).

Choosing gift.price > 190 we get

$$S = (210 - 190)/(210 - 11) \approx 1/10$$

c) The join on shipment.gift and gift.gift_id

The general formula for join is $1/\max(V(S, y), V(R, y))$

So in this case $1/6000$ (number of distinct possible gifts).

d) The join on shipment.supplier and supplier.supplier_id

$$selectivity = 1/500$$

e) The join on employee.place and place.place_id

$$selectivity = 1/1000 \text{ (the number of employees)}$$

2. Sketch a query plan for this query that would result in the minimum amount of work given no indices and the selectivity estimates you gave above.

A query plan is a particular way to plan the query.

3. Recommend a set of indices that will improve the performance of this query at most