

3. Blatt: SQL-queries, updates, referential integrity

For simplification, define a view *sym_borders* that contains the symmetric closure of *borders*. Don't write functions or procedures and pay attention to null values tampering with aggregations!

Exercise 3.1 (10 P.)

Find out the average coordinates of the geographical distribution of the world population.

- Calculate the average value of the coordinates of all cities at first.
- Now compute the average value of the average coordinates of all cities per country weighted by the country's population.
- Discuss the different results. Which one is closer to the point?

Exercise 3.2 (10 P.)

Select all cities which are less than 100km off from Freiburg and in the European Union. If using approximate solutions, give a short reason why you think it's appropriate.

Exercise 3.3 (20 P.)

The table *politics* contains a column with the independence date.

- Select all countries with an independence date between 1300 and 1600.
- Compute the arithmetic mean and the standard deviation of the independence date of the European countries.
- Specify the independence date for all countries in a table with two columns. The first column should contain the country code, the second the date in the form of the following example (without quotes): "18. Jan"
- Group all countries by the weekday of their independence date and compute the mode thereof (the weekday with the most independencies).

Exercise 3.4 (10 P.)

DECODE, CASE, REGEXP_REPLACE etc. are not allowed.

- Exchange the members of NATO and EU *without* introducing auxiliary constructs. Hint: compute the new value of the attribute *Organization* as a function from the original one.
- Restore the original state.

Exercise 3.5 (20 P.)

Find out if the infant mortality and the income per capita are correlated. Use the Pearson correlation coefficient r . Only the standard aggregation functions (MIN, MAX, SUM, COUNT, AVG) are allowed. The formula is:

$$r = \frac{\sum_{i=1}^n (x_i - m_x) \cdot (y_i - m_y)}{n \cdot s_x \cdot s_y}$$

Therein x_i and y_i are the values of the variables (here: infant mortality and income per capita), s_x and s_y their standard deviations plus m_x and m_y being their arithmetic means. There are positive and negative correlations. Regard an absolute value of r greater than 0.1 as small, greater than 0.3 as medium and greater than 0.5 as large correlation.

Exercise 3.6 (10 P.)

Compute all pairs of rivers that flow through the same countries. Don't consider rivers that flow just through one country.

Exercise 3.7 (10 P.)

In Exercise 1.3, the river network of the Zaire has been computed iteratively/recursively. Solve the same task in another way by using an auxiliary relation and referential actions.

Deadline: 11.6.2008, 11h