

Exercises for DW & DM

Sheet 6 (until 17.12.2010)

You may hand in your solutions into the mailbox at the IFIS floor (Mühlenpfordtstraße 23, 2nd floor). For the ITIS students only, please send your solutions to silviu@ifis.cs.tu-bs.de. The deadline is Friday, after the next lecture (date is also mentioned above). You may answer in either German or English. **You are encouraged to work in teams of 2 students** (not more than 2), and send your solution as a team. Please mention the **name of both students** together with the corresponding **inmatriculation numbers**.

Exercise 1 (5P)

1. Consider a client table as in Annex 1. Knowing that most of the queries project only the birth date, income and zip code, how would you partition the client table? Please draw the physically partitioned table. (3P)
2. What are mini-dimensions, and how are they useful? How would you modify the table in Annex 1 to benefit from the advantage of mini-dimensions? (2P)

Exercise 2 (15P)

1. In which cases, is using a cross product a good idea in optimizing star-joins? (1P)
2. Consider a star schema with a fact table for sales, and 3 dimensions, the Geo, Time and Product dimension. The fact table has 5 integer (4 byte) attributes (1 primary key, 3 foreign keys, and the sales amount). The Geo table has 4 attributes: 1 primary key (of 4 bytes), and 3 string attributes (each of 20 bytes). The Product table has 4 attributes: 1 primary key (of 4 bytes), and 3 string attributes (each of 20 bytes). The Time table has 4 attributes: 1 primary key (of 4 bytes), one date (of 4 bytes) one string for the month (10 bytes) and the year (2 bytes). Consider the case of a low selectivity percentage of about 50%, for a fact table of 10 million records, Geo dimension with 2000 records, Product dimension with 2000 records and Time dimension with 260 records. (Express all the intermediate results in MB, GB, or TB by case).
 - a. How big are the temporary results in the case of a dimensional cross product join strategy? (explain how you have reached these results). (4P)
 - b. How big are the intermediate results in the case of the (IBM) semi-join strategy? (explain how you have reached these results). Consider here a 32 bit based architecture with pointers to the data of 4 bytes. (10P)

Exercise 3 (12P)

1. Fill in with records the tables described in the schema from Annex 2. (2P)
2. Write an SQL select query, which applies the Cube operator on all 3 dimensions, Cube(Week, Store, Group) of the database in Annex 2 with the records you have filled at Exercise 3.1, and then present the result of the query as a table. (4P)
3. Write an SQL query which retrieves the top 3 selling product groups in February, in Braunschweig. (3P)
4. Write an MDX query performing the same task as exercise 1.3 (the previous exercise). (3P)

Annex 1:

ID	Lastname	Firstname	Birth Date	ZIP code	Phone	Income
1	A	X	01.02.78	38100	61432	35K
2	B	Y	06.03.75	38108	61433	60K
3	C	Z	09.02.88	38106	51434	55K
...

Annex 2:

