

Exercises for DW & DM

Sheet 4 (until 03.12.2010)

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 (2P)

Only for non-ITIS students: Please log in to our Homework Management System (HMS) at <https://www.ifis.cs.tu-bs.de:8443/hms/> using your y-number and password and sign in for this lecture. This will make grading and managing your homework easier for both of us. (Some of you have already been added to the lecture, so please check the accuracy of the information).

Exercise 2 (5P)

1. What kind of a schema is presented in Annex 1? (1P)
2. In which case should we consider migrating from a snowflake schema to a star schema? (1P)
3. Considering that the product dimension is subject to often change, how would you transform (draw) the schema in Annex 1, into a Starflake one? (3P)

Exercise 3 (3P)

1. What does linearization mean, in the case of multidimensional storage? (1P)
2. Explain why dimension order is important when storing multidimensional data in a linearized array. (2P)

Exercise 4 (18P)

1. Considering the R-Tree graphically represented through the MBR with a maximal node size of 3, in Annex 2, perform the following tasks:
 - a. Insert, in this order the following data (each of them will be represented as the small red squares): ("08 Qtr2", "b"), ("08 Qtr2", "c"), ("09 Qtr1", "c"). Represent each step graphically, evidencing the produced split. As split me-

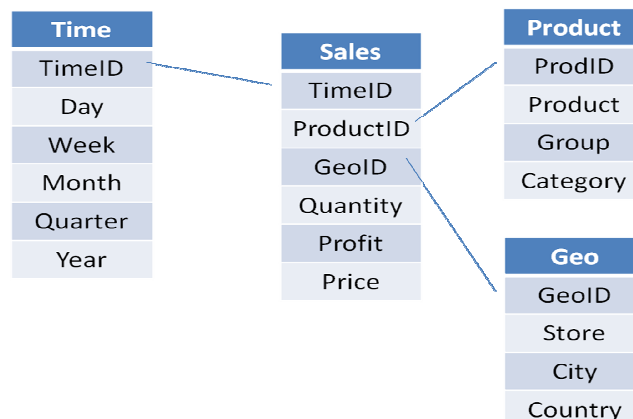
Method use the linear argument criterion. (12P)

- Draw the R-Tree according to the obtained graphical representation of the MBR, after performing exercise 4.a. (3P)
- Graphically represent (as in the lecture) the following search ($[08 Qtr_2, 08 Qtr_3]$, $[a,c]$) on both the MBR representation obtained from exercise 4.a, as well as on the R-Tree representation obtained from 4.b. (3P)

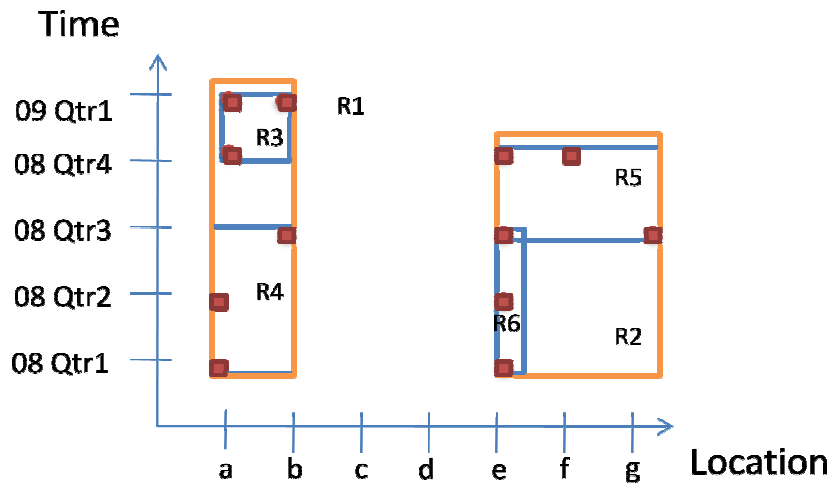
Exercise 5 (12P)

- Consider data presented in Annex 3. Construct an UB-Tree index on the Time and Product dimensions, which should serve queries as 'How many cell phones did we sell in Q1 and Q2?' For the UB-Tree, consider a size of 5 records for a HDD block. Present the following:
 - The Z-curve and the corresponding Z-regions; (6P)
 - A mapping between the sale ids and the UB index ids; (2P)
 - How many blocks do we need to access in order to answer the query on the UB Index and how many would we need without it? (4P)

Annex 1:



Annex 2:



Annex 3:

ID	Qty	ID_Prod	ID_Day
1	...	5	1
2		2	1
3		3	1
4		2	2
5		1	3
6		3	2
7		8	1
8		7	1
9		5	2
10		6	1
11		5	3
12		3	3
13		2	3
14		8	4
15		6	2
16		7	2
17		5	4
18		3	4
19		4	1
20		2	4
21		1	4

ID	Qtr	Year
1	Q1	2010
2	Q2	2010
3	Q3	2010
4	Q4	2010

ID	Product	Group	Category
1	Nokia N8	Cell Phones	Electronics
2	BlackBerry Bold	Cell Phones	Electronics
3	BlackBerry Storm	Cell Phones	Electronics
4	Apple Iphone	Cell Phones	Electronics
5	Samsung UE46	TV	Electronics
6	Panasonic TX50	TV	Electronics
7	Philips 46PFL	TV	Electronics
8	Panasonic TX46	TV	Electronics