1. What do you understand about the Differential Snapshot Problem and Record Linkage? What is the difference?

   The trivial approach is the comparison of every tuple against each tuple of the other table.

   (a) How the complexity of that comparison can be minimized? Which advantages and disadvantages are connected to them?

   (b) How is it possible to improve the data quality?

2. Compare the resource consumption concerning of MOLAP and ROLAP approaches to implement a DWH. Use the following data:

   (a) 1 Fact; 3 dimensions each having 1000 values; filling degree 20%; 1 attribute = 8byte

   (b) 1 Fact; 5 dimensions each having1000 values; filling degree 20%; 1 attribute = 8byte

   (c) 1 Fact; 3 dimensions each having 1000 values; filling degree 50%; 1 attribute = 8byte

   (d) 1 Fact; 5 dimensions each having 1000 values; filling degree 50%; 1 attribute = 8byte

3. MOLAP is the direct implementation of the multidimensional data cube using proprietary data structures. Because the amount of data enforces a persistent storage on the hard disk, the data structure has to be *linearized*. Implement a *Java* or *C++*-program, that is able to manage facts organized in two dimensions, that is accept and return them. The "mini-DWH" shall provide the following operations:

   - return a cell value (exact match),
   - return a row (partial match),
   - return a column

   The data management for that is only allowed using one dimensional arrays, similarly to the hard disk. Furthermore no nesting of arrays is allowed.
   A small test system shall demonstrate the functionality of the DWH. Which consequences are connected to the MOLAP implementation?

   ATTENTION: This task is prerequisite for the Schein.

4. Additional task: Extend the program of task 3 for the management of more than 2 dimensions. Are there any consequences for the query processing?