

Exercise 1
Dr. Tomáš Horváth, Rasoul Karimi
Information System and Machine Learning Lab (ISMLL)
University of Hildesheim

Deadline: 07.06.2012

Task 1: Dimensional Modelling I (20 points)

A renowned bookshop company from the Hanover area wants to support its newly established online selling system with a data warehouse (DWH), in order to be able, for example, to monitor new marketing channels and analyze the sales data. Assuming that you, as a new member of the consulting IT company UniConsult, are now in charge to create a data model for the bookshop.

The granularity level of the process should be based on the sale of each individual book. The CIO is however not sure what information he would like the DWH to record about this business process. After a meeting with him, you understand that the representation of the business process should, in any case, contain the following elements:

- information about each book that has been sold and its characteristic properties (title, ISBN, purchase price, sales price, Date of sale. . .)
- information about the publisher of the book
- information about each customer, whose data are recorded through the online platform (demographic data, e-mail address, credit card number. . .)
- information about the authors of the book

Data have to be loaded daily into the DWH. For the purposes of allowing possible adjustment to new requirements, the data model has to be as flexible as possible. The books have to be assigned in different categories, such as science, mystery, children's books, cookbooks, etc.

The CIO is also especially interested in further information about those collaborating companies that supply the books and those that deliver them to customers. The reason is that the CIO wants to be able to evaluate their performance.

Develop a dimensional modeling using a Star Schema in order to address the above requirements. Explain which dimensions and attributes you've chosen.

Task 2: Dimensional Modelling II (5 points)

After you develop the DWH in Task 1, you observe that not all data can become available each time. For example, at the moment time that the book is being sold online, it is not known which company will deliver the book to the customer. How do you plan to address such cases in your model?

Task 3: Dimensional Modeling III (5 points)

The recording of information about the time that each book sales happens is currently based on a daily level (i.e., only the day of sale is recorded). You want to extend the analysis from the level of day down to level of second (i.e., record also the hour, minute, and second of each sale).

A first approach would be to redefine the time dimension to granularity of seconds. Supposing that the DWH will operate for 10 years. How many records (in number of lines) would the time dimension require in this case? Can you propose a more efficient solution? How many record (in number of lines) does your solution require?