

Exercises for DW & DM

Sheet 8 (until 11.06.2008)

Please note that you need **50%** of all exercise points to be admitted for the final exams. Exercises have to be turned in until **Thursday** before the next lecture and should be completed in teams of two students each. Write both names and “Matrikelnummer” on each page. If you have multiple pages, staple them together! Please hand in your solutions on **paper** into the mailbox at the IFIS floor or to our secretary (Mühlenpfordtstraße 23, 2nd floor). You may answer in either German or English.

Exercise 1 (4P)

1. What is a staging area? (1P)
2. Why are users not allowed to interact with the staging area? (1P)
3. When should we use flat files in the staging area? (1P)
4. What is ETL, and when should it be applied? (1P)

Exercise 2 (9P)

1. Install Eobjects Data Cleaner (<http://datacleaner.eobjects.org/downloads>). Perform the following tasks, by using the sample database provided with the software (by choosing it from the drop down menu as observed in the Annex1)
 - a. Compose a regular expression which validates only strings which contain letters only (no spaces or other characters than letters), start with only one capital letter, and continue with at least one, up to 20 small letters. See examples in Annex 2. (3P)
 - b. Use the regular expression from 2.a, and create a validation task, add as validation rule a “regex validation”, choose as data selections the CUSTOMER table, and as data subset the CONTACTLASTNAME and CONTACTFIRSTNAME attributes. Write the lastname and firstname of the clients which did not pass the validation. (If there are too many you did something wrong!!!) (3P)
 - c. Give three examples (of different patterns) of strings which pass the validation of the following regular expression, and one that doesn't:

$(\backslash+\backslash\{1,2\})?((\backslash\{1,4\}\backslash)|(\backslash\{3,5\}[-/?])(\backslash\{1,5\}))$ (3P)

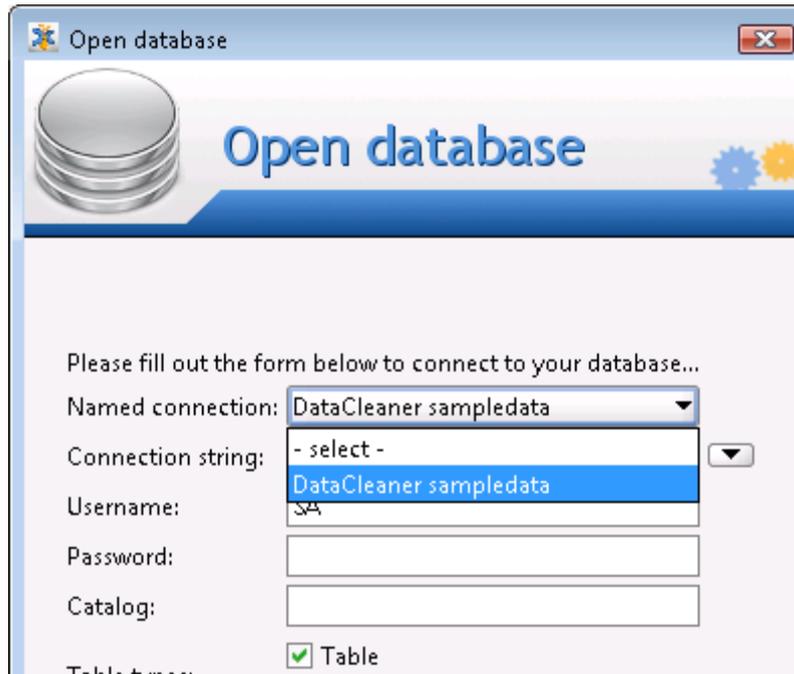
Exercise 3 (6P)

1. Briefly describe the basic steps in schema integration? (2P)

Exercises for DW & DM

2. Explain how schema mapping is performed in praxis. (2P)
3. When should we use bulk loading and why is it a good idea? (2P)

Annex 1



Annex 2

String	Evaluation
A	Bad
Aa	Good
AA	Bad
aa	Bad
A1	Very bad
Thomas	Good
Thomas Mann	Bad