

PERSONAL_ID	NAME	Geselle	Lehrling
1	Mark	null	Verkaeuffer
2	Pete	Lagerist	null

Table 1: Target representation of Data

1. Which kinds of data errors are possible and how can they be detected? Which errors are not detectable in which approach?
2. Detection of outliers: The values are given in table 3. Calculate the regression straight line for the given values. How can outliers be detected using the calculated values?
3. Extension of programming task for Schema-Integration. This task is prerequisite for the “Schein”: Extend the *Java* or *C++*-program!

This task will enhance the knowledge about schema integration.

Run the prepare SQL-Script in the database.¹

First transform the data from both tables (*Bierladen1*, *Bierladen2*) into schema of third table *IntegratedBierladen*. Do not just rely on SQL script but also write a Java or C++ program which will do the transformation of data.

For your help in this task, a small example “HelloJDBC.java” is provided. For its application actual oracle driver (ojdbc14-1.jar) must exist in Variable “CLASSPATH” .²

The **target**-Table and its relation is in Table 1 given.

The integration of schemata can cause tuple duplication (e.g. data from Mark and Pete).

Insert the tuple into the integration schema, that is defined by the Personal_ID '5' of table 2. The extended implementation shall recognize duplicates and shall delete them, e.g. the tuple defined by the Personal_ID '5' is a duplicate and should be deleted.

Discuss your results. Is it possible to parameterize your approach?

¹http://wwwiti.cs.uni-magdeburg.de/iti_db/lehre/dw/dwt1112/ueb04/ueb04_prepare.sql

²To download from the lecture website.

PERSONAL_ID	NAME	Geselle	Lehrling
1	Mark	null	Verkaeufuer
2	Pete	Lagerist	null
5	Pete	Lagerist	null

Table 2: Target table

X	Y
1	1
2	10
3	6
4	4
5	3

Table 3: Regression values